

IIII

30 August – 3 September 2021 INTERSPECH 2021 BRNO | CZECHIA Speech everywhere!

ABSTRACT BOOK



Customer-obsessed science

Amazon fundamentally believes that scientific innovation is essential to being the most customer centric company in the world.

Learn more at amazon.science

Global research locations

Aachen Atlanta Austin Barcelona Bellevue Bengaluru Berkeley Berlin Boston Cambridge Chennai Culver City Cupertino Dublin Edinburgh Gdansk Graz Haifa Hyderabad Irvine London Luxembourg Arlington Manhattan Beach New York Newark North Reading Palo Alto Pasadena Pittsburgh San Diego San Francisco Santa Clara Santa Monica Seattle Shanghai Sunnyvale Sydney Tel Aviv Tokyo Toronto Tübingen Turin Vancouver Westborough



Microsoft

Learn more about Microsoft Research at www.microsoft.com/research/

TABLE OF CONTENTS

Sponsors	4
Exhibitors	5
Welcome Message from the ISCA President	6
Welcome Message From The Interspeech 2021 General Chairs	8
Message from the Technical Program Chairs	9
Welcome to Brno, Czech Republic	13
INTERSPEECH 2021 General information	15
ISCA Ethics for Authors and Attendees	16
Program at a glance	18
Social Program	22
Organizers: International Speech Communication Association (ISCA)	23
Organizers: Brno University of Technology (BUT)	23

INTERSPEECH 2021 Organizing Committee	24
Awards and Grants	25
Keynote Speakers	27
Survey Talks	29
Tutorials - Tutorial chairs	31
Special Sessions and Challenges - Chairs	35
ISCA Student Advisor Committee (SAC)	41
Area Chairs	42
Satellite Events - Satellite committee	43
Scientific Review Committee	44
Future INTERSPEECH Conferences	54
Session Index	56
Abstracts	60
Author Index	262

SPONSORS

We would like to thank to all below mentioned partners for their contribution on INTERSPEECH 2021.

FOUNDING SPONSOR



EXHIBITORS AND INSTITUTIONAL SUPPORTERS

EXHIBITORS



MEDIA PARTNER



BRNO | CZECHIA

WELCOME MESSAGE FROM THE ISCA PRESIDENT

Welcome to Brno, Czech Republic, where we hold the 22nd INTERSPEECH. INTERSPEECH began in 2000/2001 with Beijing, China in 2000, and Aalborg, Denmark in 2001, after the merging of two international organizations, ESCA: the European Speech Communication Assoc., and PC-ICSLP: the Permanent Council of Inter. Conf. on Spoken Language Processing. The merger of these two organizations unified ISCA as a global organization, dedicated to supporting and promoting the field of speech communications in both scientific foundational advancements as well as technology innovations. Over these past +21 years, ISCA has seen remarkable growth, in terms of our membership, educational and research impact in terms of conferences, workshops, Special Interest Groups (SIGs), Distinguished Lecturers (DL) programs, support for diversity across all domains, and community outreach.

When INTERSPEECH-2021 was awarded to Brno, Czech Republic, it could never have been imagined the added challenges this Organizing Team would face with the current world pandemic. While the INTERSPEECH-2020 team successfully accomplished a major success in Shanghai, China holding a fully remote/virtual conference, new challenges existed for the IS-2021 team in transitioning to our society's first hybrid conference – with both in-person and remote participation. The main challenges clearly involve world time-zones and not having all attendees able to participate on the same timeline. The entire INTERSPEECH-2021 Organizing Team skillfully navigated one of the most challenging conference organizational paths - ultimately moving to a hybrid physical in-person and virtual/remote access format. INTERSPEECH has always embraced human communication interaction as our core, and the ability to meet physically to exchange ideas, see the latest scientific and technological advancements in our field, and to re-connect with colleagues as well as establish new collaboration opportunities has always been paramount to successful INTERSPEECH conferences. The INTERSPEECH-2021 Organizing Committee worked closely with the ISCA Board, and specially our Conference Board Members (Sebastian Möller and Margaret Zellers) as the team navigated the challenging decisions during the past year. This conference represents our first hybrid conference, at a time when our membership has reached an all time high (over 2500). The program established here balances the goals of what we have always cherished in our previous INTERSPEECH events: an extremely strong technical program, world-class Keynote Speakers, expert Tutorials, a wide encompassing range of special sessions, and the corresponding social and networking options necessary for our students and young researchers in the field. The IS-2021 Organizing Committee is to be commended for an outstanding job in planning, executing and delivering a world-class experience for our membership.



ISCA MEDAL for SCIENTIFIC ACHIEVEMENT: The ISCA Medal for Scientific Achievement recognizes and honors an individual each year who has made extraordinary contributions to the field of speech communication science and technology. We are pleased to announce that at this year's INTERSPEECH-2021, we will honor **Prof**. **Dr.-Ing. Hermann NEY**, "for pioneering and seminal contributions to data-driven methods for automatic speech recognition and machine translation." His extensive publications and impact have been internationally recognized in automatic speech recognition, statistical classification and machine learning, and statistical machine translation.



ISCA SERVICE MEDAL: The ISCA Service Medal recognizes and honors an individual who has contributed to the speech communication community by advancing the goals and mission of ISCA. Since our inception in 2000, the ISCA Service Medal has only been awarded eight times. This year, the ISCA Board has moved forward with recognizing an individual who has demonstrated remarkable service for our society. We are pleased to recognize **Prof. Martin Cooke**, "for sustained service to ISCA and the Speech Communication community, Publications and Community Support."

ISCA FELLOWS: We will also celebrate our members' achievements by recognizing eight ISCA Fellows for 2021, who include: Mary BECKMAN, Karen LIVESCU, Eric FOSLER-LUSSIER, Maxine ESKENAZI, Jean-Francois BONASTRE, Yang LIU, DeLiang WANG, and Dong YU. Please join me in extending our warm congratulations to these pioneers who

6











have contributed both scientific and technical advancements in speech communication!

INTERSPEECH 2021 marks the 22nd Annual Conference of ISCA, which continues the success of recent events. It is a privilege for me to be a part of the conference preparation as ISCA President. This year's INTERSPEECH Organizing Committee was truly diverse and comprehensive in their scope, depth, and attention to detail – overseeing all the logistics of paper submission, reviews, and final program preparation leading to our first hybrid conference experience, supporting both in-person and remote/ virtual participation. ISCA and myself extend a sincere thanks to all of our colleagues for completing your reviews, and the Technical Program Committee and Area Chairs for ensuring the highest quality technical program.

As ISCA President, I also wish to extend a sincere thanks to the all those organizations who provided financial sponsorship for INTERSPEECH-2021. These contributions have gone to provide many ISCA Travel Fellowships for deserving students and young scientists to attend INTERSPEECH-2021 (both in-person and virtual), as well as to strengthen and enrich the diversity of the conference with many special events. MANY THANKS to those sponsoring companies and organizations!!

Organizing an INTERSPEECH event takes enormous courage, endurance and dedication, I would like to express my gratitude and appreciation to the Conference General Chairs Honza Černocký and Hynek Heřmanský, as well as the Technical Co-Chairs Lukáš Burget, Lori Lamel, Odette Scharenborg, and Petr Motlíček, who led an experienced organization team to bring INTERSPEECH to Brno for the first time. On behalf of the ISCA Membership and Speech Communication community as a whole, we express our sincere thanks and gratitude to the entire INTERSPEECH-2021 Organizing Committee, who went beyond what any of us could expect in adapting to these new conference challenges during a world pandemic, and delivering an in-person and virtual/ remote experience which all future INTERSPEECH conferences will benefit.

Finally, my term as ISCA President will be ending this year, and during this INTERSPEECH-2021 the ISCA Board will elect a new ISCA Leadership Team including President, Vice-President, Treasurer, and Secretary. I have thoroughly enjoyed serving as your President over the past four years, and seeing the tremendous growth, advancement, and support for our community from our ISCA Board. One of my goals when I took on this role was to expand diversity including worldwide participation - which, thanks to many on our ISCA Board, especially our ISCA Diversity Committee, we are making effective steps to provide better opportunities and support for all in our field. Another goal of mine was to ensure ISCA events including INTERSPEECH would provide an opportunity for those not able to travel to the conference to be able to participate. This is a critical problem for many students studying for PhDs in countries where visa's to either enter or leave are very restrictive. In addition, a number of our members have at times health constraints and cannot travel. Remote participation would support those with visa constraints, health/ travel restrictions, as well as open up opportunities for many in countries where ISCA membership is low (Africa, South America, etc.) to build a broader membership worldwide. IS-2020 and IS-2021 have shown us that supporting remote participation for our community is both viable, and provides excellent outreach opportunities to parts for the world where we have not been able to recruit members. It is my hope that this pandemic is able to be addressed as we move forward, and that we continue to provide remote access to our landmark conference in the field of speech communication. I express a sincere thanks to all those who have helped me as well as made contributions to ISCA during my time as President, and I look forward to continuing to serve our society in other ways going forward.

With my best wishes for a successful conference!



John H.L. Hansen ISCA President

WELCOME MESSAGE FROM THE INTERSPEECH 2021 GENERAL CHAIRS

Dear colleagues

Welcome to Czechia, the country with one of the highest number of speech-related activities per capita in the world, and to Brno where every fifth inhabitant of the city is a student. The topic of the Interspeech 2021 is "Speech Everywhere". Indeed, the topic is very appropriate. Some of us remember days when research in speech communication was done by few research groups, most of us knowing each other, and papers in speech-related conferences could be gone over in a day or two. Now, speech technology is in every phone, our teleconferences are instantaneously transcribed, and our machines are being controlled by users instantaneously verified from their voices. However, it doesn't mean that research in communication stops. Just as the first designs of flying machines heavier than air spurred more research in aeronautics, the introduction of commercially successful speech technology only shows how little we still know about communication by speech. So, our Interspeech conferences are only growing, with numbers of participants never seen before. And rightly so. Together with many other human capabilities, the human ability to communicate by speech is a challenge still to be conquered. We welcome you to being a part of this endeavor.

Since early 2020, Covid 19 pandemics have made the organization of all scientific events difficult. Still, it was our hope all along that we could host most of you in person here. This is unfortunately not possible yet. However, the data from author registration allowed us to declare this Interspeech to be a mix of in-person and remote presentations. We made this decision with joy and fear at the same time, knowing we're running into a risky business and unproven terrain. Unlike conferences in the past, where most of the things were known well beforehand and fixed for many months, we had to accept the fact that we were operating in a fuzzy and probabilistic environment. Down-scaling the physical side of Interspeech had also some positive effects - we could move the venue for the conference to Hotel International, just 100 meters from the vibrant historic center of our city. For those of you who must participate in this conference remotely, we tried our best to provide all available tools for the efficient dissemination of your work in the community. For those whose situation allows for live participation in person, we hope that your experience will match and exceed what you expected.

This Interspeech would not happen without the help of many - let us thank our organizing committee, and among them, our Technical chairs that took the greatest share of work. Thanks must also go to ISCA, Brno University of Technology, and to Guarant International. We are also grateful to all sponsors and exhibitors as well as the city of Brno and numerous organizations supporting us. The biggest thanks go however to those contributing to our excellent technical program - Keynote and Survey talk speakers, Tutorial speakers, Special and Challenge Session organizers, Show & Tell contributors, but primarily - all of you who submitted quality papers.

We wish you an excellent Interspeech

Hynek Heřmanský and Honza Černocký

General Chairs





MESSAGE FROM THE TECHNICAL PROGRAM CHAIRS









Lukáš Burget

Lori Lamel

Odette Scharenborg

Petr Motlicek

It is certainly an honor to serve as the Technical Program Chairs of INTERSPEECH 2021 in Brno. INTERSPEECH is the flagship ISCA conference, providing an annual opportunity to meet with colleagues for formal technical and less formal individual exchanges. When due to Covid-19 the decision was taken to hold INTERSPEECH 2020 (Shanghai) virtually, we were expecting a large number of submissions and attendees in Brno, but we were still surprised by the number we received!

Dealing with the uncertainty with respect to the format of the conference and the ever-changing context has been an additional element of stress and required substantial adaptability for all parties.

The overall technical program is similar to those of previous years, with different types of technical activities (tutorials, keynotes, regular sessions, special sessions, and similar to Hyderabad and Graz, there will also be 4 survey talks on selected topics. We have tried to maintain the interdisciplinary flavor of Interspeech, gathering researchers across many domains including speech sciences, linguistics, hearing and perception, spoken languages technologies and their applications.

We hope that you will enjoy the program, and look forward to seeing you either in-person for those of you who are able to travel to Brno, or virtually at INTERSPEECH 2021.

For the past many years we have seen an increase in the interest in speech and speech technology everywhere around us. This is also reflected by the steadily growing attendance at the largest speech science and speech technology conference in the world, INTERSPEECH. We again received a record number of 2277 submissions to Interspeech 2021 from 67 countries. The distribution of paper submissions by country is shown in the figure below. Roughly half come from Asia, and a quarter each from Europe and North America. We were truly in awe and happy as well as slightly worried because of the enormous task we, and our 52 area chairs, were facing. The papers were submitted to 13 different areas, each with a large number of subareas. New this year is area 13 which was created to reflect the increase in research on voice, hearing, and speech disorders both from the speech science and speech technology side, for which paper submissions were previously distributed over different areas. We were very happy to see a whopping 103 submissions submitted to this new area! Each area was overseen by 2 to 6 area chairs, who did an outstanding task to ensure that all submissions were reviewed by at least 3 reviewers. Area Chairs were chosen based on their expertise. We tried to balance for seniority, inviting several new mid-career researchers to act as Area Chair, for gender, for industry vs. academia, and for geographical region. We are thankful for the help we obtained from the Diversity and Inclusion Chairs (Heidi Christensen and Julia Hirschberg) in this selection.





1990 submissions were sent out to review to 1600 reviewers, from 66 countries. The next figure shows the distribution of reviewers by country. About 42% of the reviewers are from Europe, 30% from North America and 25% from Asia. The percentage of reviewers from Asia is low compared to the proportion of submissions. This may in part be due to email communication from the START system being blocked by some authorities. We thank all reviewers for their support of Interspeech, without whom the conference would not be possible! Unfortunately the number of reviewers is actually lower than the number of submissions that were sent out to review. This meant that each reviewer had quite a high review load, with some people reviewing up to 12 submissions, and many people reviewing more submissions than they signed up for. We would therefore also like to use this opportunity to ask you, if you hold a PhD or an MA/MSc with 7 years of research experience to please sign up as reviewer for Interspeech, if you are not yet a reviewer. This can be done via the ISCA website.



Once all of the reviews were in, each review was read by the Area Chair and a preliminary decision was made for each submission based on the reviews. This decision was then checked by a second Area Chair, and in case of a disagreement, a discussion ensued to resolve the question of acceptance and rejection. This year, for the first time, the Area Chairs were asked to write a meta-review (not visible to the authors) so that we could keep better track of the discussions and reasoning for acceptance and rejection of individual papers. Moreover, also new this year was that reviewers were invited to discuss with the Area Chairs about their and the other reviews of the submission and the Area Chair's decision for the paper. In total, 963 submissions were accepted for publication and presentation at Interspeech 2021 Brno.

After the accept/reject decisions were made, all accepted submissions were checked using a plagiarism checker. Moreover, during the review procedure, several Area Chairs and reviewers pointed out ethical issues with several submissions. In total, 23 submissions with suspected ethical problems were forwarded to the ISCA Ethical Committee. Each submission was checked by all members of the ISCA Ethical Committee, and after discussion, a final assessment was provided for each submission, and a level of violation was determined:

- Level 0: No violation there are elements which caused concern but none rose to the level needed to continue or notify the authors.
- Level 1: Violation with a warning elements did rise to a violation of ethics rules. The level of deliberate intent might be mild to moderate, so an official sanction was not applied but authors are notified that they did violate rules.
- Level 2: Serious violation it has been determined that the authors deliberately violated ISCA Code of Ethics in publication. Papers are automatically rejected if in the review process, or officially removed and not included in the official ISCA published conferences, workshops, etc. It is also the case that if authors who are determined to have violated publication ethics rules, either must remove their names from other accepted papers in that meeting, or those papers would also be removed from the program/publication (e.g., so as to not penalize co-authors from other papers who were not involved in a violation case, these secondary papers could still appear if the sanctioned authors are removed). In this case, an official sanction period is applied to those authors deemed responsible with a ban on both/either not being able to publish in ISCA events, and/or attend/participate in ISCA Events.

Unfortunately, several submissions (both accepted and already rejected submissions) rose to Level 1 or Level 2 violations. The ISCA Ethics Committee spent a lot of time reviewing and assessing each case. We urge every (potential) author to read and adhere to the code of conduct for authors on the ISCA website. A special note about ArXiv papers: although ISCA allows ArXiv papers to be submitted to Interspeech and other ISCA supported workshops, we noticed that several manuscripts on ArXiv mentioned different co-authors than the version submitted to Interspeech 2021 Brno, moreover several times the ArXiv manuscript did not have a note saying that the manuscript was submitted to Interspeech. This is not good practice, against ISCA's code for authors, and automatically raises potential ethical concerns. It also creates significant work for the TPC and the ISCA Ethics Committee.

In the final step, all Area Chairs, together with the Technical Program Chairs, combined all accepted papers into sessions. An interesting challenge for this year was the hybrid format of Interspeech 2021 Brno, which means that some sessions will be held fully on-site in Brno while others will be held fully virtual. In order to be able to make these different types of sessions, which each required a different number of papers per session, the paper presenters needed to indicate very early whether they were going to present on-site or online. We believe that the area chairs did an amazing job in putting together coherent sessions both on-site and virtually, and we hope you will have a great experience!

The session planning has also been quite dynamic, with numerous changes due to pandemic related travel restrictions, imposing changes in the paper presentation format. This in turn led to many moves of presentations from one session to another. Even at the time of this writing, about a month before the conference, we have daily requests to change the presentation format (unfortunately mostly from in-person to virtual).

As at every Interspeech, this year we again have the best student paper awards that will be announced and handed out at the closing ceremony of Interspeech 2021 Brno. During the review procedure, reviewers were asked to indicate if they thought a student paper was particularly good and should be considered for the best paper award. Subsequently, the Area Chairs of each area were asked to propose one student paper from their area for consideration for the best student paper award. The recommendations from the reviewers were taken into account but did not necessarily lead to a nomination. All nominated papers were then handed over to the chair of the Best student paper award committee led by Torbjørn Svendsen. The nominated students will be asked to present their work in a talk. After reading the paper, the reviews, and watching the talk, the committee will decide the Best student paper awards during the conference.

No Interspeech would be complete without keynote speakers who inspire you and make you think beyond your current research fields! We are therefore very excited that four amazing keynote speakers (overseen by Mark Gales and Jan Hajič) have accepted our invitation: Hermann Ney, the ISCA-medallist, Mounya Elhilali, Pascale Fung, and Thomáš Mikolov. Their talks cover topics that are directly related to the topics at Interspeech and extend those topics into the wider area of speech communication. Let yourself be inspired by these outstanding speakers! Following the new tradition started at Interspeech 2018 Hyderabad, we also offer you four exciting survey talks, one of which is focused on the new Area 13 on voice, hearing, and speech disorders to further highlight this exciting and important new(ish) research direction. We are delighted that four of our excellent, midcareer colleagues accepted our invitation to present an overview of ongoing research being pursued in one of their preferred areas at Interspeech 2021. For both the keynote speakers and the survey talks we tried to ensure to cover the wide range of topics that our Interspeech community covers so that there is something to enjoy for everyone. In addition to the diversity of topics, we tried to ensure a diversity in gender and have a regional balance. We are thankful for the diversity chairs for their help with this! You will find the keynote and survey talks interleaved through the technical program because we wanted to ensure that as many people from across the globe would be able to attend these talks.

In addition to the accepted papers, the keynote talks and the survey talks, the program contains six highly interesting tutorials on a wide range of topics (overseen by Yenda Trmal and Anil Kumar Vuppala), 14 exciting special sessions (overseen by Sakriani Sakti and Pavel Ircing) and 4 Show & Tell sessions (overseen by Reinhold Haeb-Umbach and Najim Dehak). In total, 6 tutorials were submitted and after the reviewing procedure, all were accepted. We received 25 special session and challenge proposals on a wide range of highly interesting topics, including 2 on the timely (sadly) topic of COVID-19 detection from audio. After



an initial selection taking into account the fit of the proposed special session or challenge with Interspeech and ensuring a wide range of different topics with not too much overlap between the topics, 20 special sessions and challenges were provisionally accepted. Finally, 5 special sessions and 9 challenges received enough accepted papers that they found their way into the exciting technical program of Interspeech 2021. Following Interspeech 2019 in Graz, applicants for Show & Tell demonstrations had to submit a short video of their intended demonstration in addition to a 2 page extended summary of their work. We received 33 demonstrations, which were all reviewed. In total, 29 were accepted and will be presented at Interspeech for you to enjoy. Finally, we have 9 ISCA supported Satellite Workshops (overseen by S. Umesh and Tomohiro Nakatani) that will be held surrounding the main Interspeech conference, enriching the technical program.

Organising any Interspeech is a collaborative effort. Organising the technical program of Interspeech is definitely a collaborative effort too. We therefore have many people to thank for their help in putting together this technical program that we are very excited about. We especially would like to point out that organising an Interspeech in the face of a global pandemic and a hybrid conference in addition is particularly challenging, and we are therefore even more grateful for all the support and help we received from (in random order):

- All Area Chairs, without you this massive effort would not have been possible. Thank you for your prompt actions and answers to our requests, your (last minute) availability for meetings (in spite of the early morning or late night times for some of you), and all the hard work you put in, despite in many cases also dealing with exhaustion, home-schooling, illness, or some (nasty) effects of getting the Covid-19 vaccination.
- All reviewers who provided us with the necessary reviews, and particularly those who were able to take on additional papers or were available for last minute reviews (e.g., when suddenly several reviewers fell ill or had family emergencies).
- Our always cheerful and on-top-of-things General Chairs Honza and Hynek. Your support and help was instrumental in making everything come together.
- The ISCA board, and particularly Sebastian Möller and Meg Zellers, for all things related to organising Interspeech, Hema Murthy and Jianhao Tao for all things related to the technical part of organising Interspeech, and John Hansen, our president, who was always available for anything and everything.
- The ISCA Ethics Committee, for their extremely fast and thorough evaluation and assessment of the ethical issues that were raised.
- The special sessions, satellite workshops, plenary talks, tutorials, and show & tell chairs for enriching the technical program in such an exciting way.
- The diversity chairs, for their help in diversifying Interspeech at many levels.
- The authors, for submitting their interesting work to Interspeech. Without you there would be no conference!
- The people from Guarant for all the issues that deal with the logistical, financial, and registration side of organising Interspeech.

Interspeech 2021 Brno is a hybrid Interspeech, the first of its kind. We hope that you will have a highly enjoyable experience either virtually through the conference platform or physically in Brno, and we hope you will be inspired by the highly diverse program, special sessions and challenges, survey talks, show & tell, tutorials, satellite workshops, and keynote speakers! Although as a community we have not been able to gather together due to the ongoing pandemic for what feels like too long a time, we hope that this hybrid conference will allow you to mingle, build new networks, and "see" again your friends and colleagues. Enjoy!

Lukáš Burget Lori Lamel Odette Scharenborg Petr Motlicek

WELCOME TO BRNO, CZECH REPUBLIC

Brno – Central Europe's Hidden Gem

Welcome to Brno, Czech Republic, the second largest city in the country, known to the backpackers for its friendliness, to the entrepreneurs for being a technology and innovation hub and to the scientists for its 12 universities and research centres.

Once you get over the remnants of the more recent past, you can see Brno's true beauty, which lies in the combo of friendly laid-back locals and historic and technological sights worth visiting. Be it functionalist villas, historic cathedral, planetarium, museum of technology or modern art in the streets, you can enjoy everything within a few minutes ride in our awardwinning public transport or on foot, with a cup of delicious coffee or a chilled drink in your hand.

If you are not a sightseeing type, Brno is a paradise for people enjoying great cuisine. Just stroll around in the city centre and blend with the locals who love to spend their afternoons or evenings in the streets, and explore whatever



food and drinks would suit you best. As a bonus, while you can hear English here quite often thanks to our many expats, you will probably be the only tourist around.

Technological highlights of the region

Brno has become an innovation hub thanks to the unique cooperation among the public, private and academic sectors. With the support of the public institutions, the universities and local entrepreneurs laid a strong foundation for todays vibrant IT and technology sphere, resulting in companies such as Avast and Kiwi, firms developing electron microscopes – around a third of the world's production comes from Brno – and world-famous game design studios Bohemia Interactive and Amanita Design. Brno also hosts a number of scientific sites, for example CEITEC, a research centre for life science, advanced materials and nanotechnology, founded by four of Brno's biggest universities and two research institutes.

3,1% of the region's GDP is invested into research and development and 3,9% of people working in the region are employed in the R&D sphere. Both numbers are high above the European average and the highest in the Czech Republic. This, with the purposeful support of entrepreneurship and innovation from both the city and the region's officials creates an ideal ground for technological start-ups and projects. Moreover, thanks to Brno's qualities as a place to study, work and live, there is a high concentration of skilled and experienced workers and scientists who are the backbone of an innovative business.

TOURIST HIGHLIGHTS

Tugendhat Villa

The UNESCO heritage site, villa of Greta and Fritz Tugendhat, was designed by Ludwig Mies van der Rohe and was constructed between 1929 and 1930. It was the first private house of its kind in Czechoslovakia which employed a steel loadbearing structure with columns on a cruciform floor plan. Onyx from Northern Morocco, Italian travertine, and veneer from exotic woods such as rosewood, zebrawood, and Makassar ebony decorate the interior. The technology used in the house, such as hot-air heating or electrically-operated windows make it absolutely unique. Book your tour as soon as possible (or just visit the beautiful gardens).



Old Town Hall

Enter its arched passage through the portal beneath the legendary Late Gothic turret by sculptor Anton Pilgram. In the passage, you can see two items from other famous Brno legends: a crocodile, also called the Brno dragon, and a wheel. The courtyard beyond, with Renaissance arcades from the end of the 16th century, was built by Italian designers, and later modifications are the result of its Early Baroque renovations. Also, you can climb the 63-metre tower to enjoy an impressive view of Brno from the top.

Špilberk Castle

Špilberk Castle was established in the 13th century to protect both the Czech lands and the town of Brno. An occasional residence of Moravian margraves, the castle became a huge military fortress in the 17th and 18th centuries, to be





Underground reservoirs in Žlutý kopec

For centuries the city had drawn its water from the Svratka River, but the quantity and quality of water soon proved insufficient. So, in 1863 a competition was announced to construct a new water supply network. The solution based on a design by the London builder Thomas Docwra created the structure without recourse to steel, wood or reinforced concrete, which meant employing various types of arch, giving us today the striking impression of an underground labyrinth. These underground halls powerfully evoke the remnants of some long-dead civilization from another world, or a computer game.

Cathedral of St Peter and Paul

The cathedral dates back to the 11th and 12th centuries. The church was rebuilt in the early Gothic style and made a provost church and a collegiate chapter. The artist behind the 18th-century Baroque interior was architect Mořic Grimm. The bells on the cathedral towers ring at 11 am instead of at noon in remembrance of a legendary trick that Jean-Louis Raduit de Souches played on the Swedish army as it lay siege to Brno. As the story goes, Swedish general Torstenson claimed he would abandon the siege if his army failed to conquer the city before the bells

> started to ring at noon. For this reason, de Souches decided to have the bells ring one hour earlier.

Vegetable Market









INTERSPEECH 2021 GENERAL INFORMATION ON-SITE PARTICIPATION

Conference Venue	BEST WESTERN PREMIER HOTEL INTERNATIONAL BRNO Husova 16/200 602 00 Brno-střed					
Tutorials Venue	Brno University of Tech Faculty of Information Božetěchova 2 612 66 Brno	nnology Technology				
Registration desk	Monday Tuesday Wednesday Thursday Friday	30 August 31 August 1 September 2 September 3 September	7:30 - 18:00 7:00- 21:00 8:30-21:00 8:30 -21:00 8:30 -19:00			
Official Language	The official language o	f the conference is En	glish.			
Time zone	The Czech Republic is in the Central European Summer Time (CEST), which is two hours ahead of Greenwich Mean Time (GMT+2).					
Local Currency	The Czech Republic has the Czech crown (CZK, Kč) as its sole currency. One euro represents approximately 25.475 Czech crowns.					
Name Badge	Please wear your conf events are dependent	erence name badge on a visibly displayed	at all times. Access to the conference and all the social badge.			
Electricity	In the Czech Republic	the electrical voltage	used is 220/230V.			
Wifi	Free WiFi will be provid	ded at the conference	venue.			
Emergency	Please notify any of the emergency call numbe	NTERSPEECH 2021 s er in the Czech Repub	taff members if you need medical assistance. The general .ic is 155.			
Insurance	The organizers do not accept any responsibility for individual medical, travel or personal insurance. Delegates are strongly advised to have their own travel insurance policies to cover risks including (but not limited to) loss, cancellation, medical costs and injury. The INTERSPEECH 2021 organizers will not accept any responsibility for any delegate failing to take out adequate insurance.					
Disclaimer	The organizers are no any other individuals a the conference. The or situations during or as for their own safety and	It liable for any loss of companying them, b ganizers also cannot s a consequence of d belongings.	or damage incurred by the conference delegates or by both during the official activities as well as going to/from accept liability for injuries arising from accidents or other the conference attendance. Delegates are responsible			

ISCA ETHICS FOR AUTHORS AND ATTENDEES

Code-of-Ethics for Authors

ISCA is committed to publishing high-quality journals and conference proceedings. To this end, all authors are requested to ensure they adhere to ethical standards, specifically (but not limited to):

The work does NOT include fabrication, falsification or any kind of data breach. Authors should keep their code and log data that produced the results in the paper. Authors are also encouraged to open their code and dataset.

The work does NOT include plagiarism or significant self-plagiarism. The work is NOT ALLOWED to be submitted to any other conference, workshop or journal during the review process. ISCA (and conference organizers or journal editors) may use tools to detect plagiarism and reject papers without review.

The work does NOT use figures, photographs or any other kind of content whose copyright is not owned by or granted to the authors, except for proper quotations allowed by the copyright law. ISCA (and conference organizers or journal editors) may request authors to provide evidence of permission to use the content for their work.

The work does NOT include inappropriate content in terms of human rights. ISCA (and conference organizers or journal editors) may request authors to provide evidence of approval from the host Ethics Committee (Institutional Review Board or equivalent) that the work meets their Institution's ethical requirements, and/or explicit consent from the human subjects used in the work.

All (co-)authors must be responsible and accountable for the work, and consent to its submission.

If any concerns relating to this code are raised or reported, ISCA (and conference organizers or journal editors) will convene their Ethics Committee to investigate the matter and decide on appropriate action including withdrawal of the paper and suspension of future submissions by the authors.

ISCA also enforces the NO-show policy for conference papers. Any paper accepted into the technical program, but not presented on-site, may be withdrawn from the official proceedings. Please refer to the ISCA Conferences Policy point 2) and 8).

Code-of-Conduct for Conference and Workshop Attendees

ISCA is committed to providing a pleasant conference experience without harassment and discrimination for anyone, regardless of gender, sexual orientation, race, religion, disability and physical appearance. We do not tolerate any verbal or non-verbal expressions of harassment or discrimination. Please note that it matters if a person feels harassed or discriminated regardless of the original intent of the expressions. In particular, sexual language and imagery are not appropriate in any conference venue. Conference participants who engage in inappropriate behavior may be expelled from the conference without a refund at the discretion of the conference organizer. These persons may be included in a watchlist for future ISCA-sponsored events.

If you are troubled by the behavior of another attendee at the conference, or notice someone is in trouble, please speak immediately to a member of conference staff or send a message to <ethics@isca-speech.org>.

Your concern will be heard in confidence and taken seriously to solve the problem.



PROGRAM

MONDAY, 30.08.2021 - TUTORIALS

TIME (all times are CEST)	Faculty of Information Technology of Brno University of Technology									
10:00 - 11:00	Warm-up Coffee									
11:00 - 14:00	Intonation Transcription and Modelling in Research and Speech Technology Applications - Amalia Arvaniti et al.	Neural target speech extraction - Marc Delcroix et al.	Speech Recognition with Next-Generation Kaldi (K2, Lhotse, Icefall) - Piotr Żelasko et al.							
14:00 - 15:00		Lunch Break								
15:00 - 18:00	An Introduction to Automatic Differentiation with Weighted Finite-State Automata - Awni Hannun	SpeechBrain: Unifying Speech Technologies and Deep Learning With an Open Source Toolkit - Mirco Ravanelli et al.	Concept to Code: Semi-Supervised End-To- End Approaches For Speech Recognition - Omprakash Sonie et al.							

TUESDAY, 31.08.2021 - Program at a Glance

TIME (all times are CEST)	Hall A+B	Hall C Hall D		Lacina			
9:30 - 11:30	Tue-M-O-1 Speech Synthesis: Other topics	Tue-M-O-2 Disordered speech	Tue-M-O-3 Speech signal analysis and representation II	Tue-M-SS-1 The INTERSPEECH 2021 Computational Paralinguistics Challenge (ComParE) - COVID-19 Cough, COVID-19 Speech, Escalation & Primates	Tue-M-V-1 Feature, Embedding and Neural Architecture for Speaker Recognition	Tue-M-V-2 Speech Synthesis: Toward End- to-End Synthesis II	
11:30 - 12:30	Survey Talk						
12:30 - 13:30				Lunch Break			
13:30 - 15:30	Tue-A-O-1 Embedding and Network Architecture for Speaker Recognition	Tue-A-SS-2 The First DiCOVA Challenge: Diagnosis of COVid-19 using Acoustics	Tue-A-O-2 Speech perception I		Tue-A-V-1 Acoustic event detection and acoustic scene classification	Tue-A-V-2 Diverse modes of speech acquisition and processing	
15:30 - 16:00				Coffee Break			
16:00 - 17:00	Opening Ceremony						
17:00 - 18:00	Keynote - ISCA medalist						
18:00 - 19:00			Coff	ee Break / Light Dinner			
19:00 - 21:00	Tue-E-O-1 ASR Technologies and systems	Tue-E-O-2 Phonation and voicing	Tue-E-O-3 Health and Affect I	Tue-E-SS-1 Voice quality characterization for clinical voice assessment: Voice production, acoustics, and auditory perception	Tue-E-V-1 Robust Speaker Recognition	Tue-E-V-2 Source separation, dereverberation and echo cancellation	

WEDNESDAY, 01.09.2021 - Program at a Glance

TIME (all times are CEST)	Hall A+B	Hall C	Hall D	Lacina						
10:30 -11:00			Warm-up Coffee							
11:00 - 13:00	Wed-M-O-1 Miscellaneous topics in ASR	Wed-M-O-2 Phonetics I	Wed-M-O-3 Target speaker detection, localization and separation	Wed-M-SS-1 ConferencingSpeech 2021 challenge: Far-field Multi- Channel Speech Enhancement for Video Conferencing	Wed-M-V-1 Language and Accent Recognition	Wed-M-V-2 Low- resource speech recognition				
13:00 - 14:00	Survey Talk									
14:00 - 15:00				Lunch Break						
15:00 - 16:00	Keynote									
16:00 - 18:00	Wed-A-O-1 Language Modeling and Text- based Innovations for ASR	Wed-A-O-2 Speaker, Language, and Privacy		Wed-A-O-3 Assessment of pathological speech and language I	Wed-A-V-1 Communication and interaction, multimodality	Wed-A-V-2 Language and Lexical Modeling for ASR				
18:00 - 19:00			Coff	ee Break / Light Dinner						
19:00 - 21:00	Wed-E-O-1 Graph and End-to-End Learning for Speaker Recognition	Wed-E-O-2 Spoken Language Processing II	Wed-E-O-3 Speech and audio analysis	Wed-E-SS-1 INTERSPEECH 2021 Deep Noise Suppression Challenge	Wed-E-V-1 Cross/ multi-lingual and code- switched ASR	Wed-E-V-2 Health and Affect II				

AREAS

Area Number	Area Name and Color
1	Speech Perception, Production and Acquisition by Human Listeners
2	Phonetics, Phonology, and Prosody
3	Paralinguistics in Speech and Language: Human and Automatic Analysis and Processing
4	Speaker and Language Identification
5	Analysis of Speech and Audio Signals
6	Speech Coding and Enhancement
7	Speech Synthesis and Spoken Language Generation

8	Speech Recognition: Signal Processing, Acoustic Modeling, Robustness, Adaptation
9	Speech Recognition: Architecture, Search, and Linguistic Components
10	Speech Recognition: Technologies and Systems for New Applications
11	Spoken dialog systems and conversational analysis
12	Spoken Language Processing: Translation, Information Retrieval, Summarization, Resources and Evaluation
13	Speech, voice, and hearing disorders
14	Special sessions challenges
15	Show and Tell

TUESDAY, 31.08.2021 - Program at a Glance

	OLSDAI,	24.0	O'EVE		Ugi u	III at a	atunec	
	Unified Virtu	ial Ses	sions		-			
Tue-M-V-3 Speech enhancement and intelligibility	Tue-M-V-4 Spoken Dialogue Systems I	Tue-M- ASR: I feature e far-	V-5 Topics in Robustness, extraction, and field ASR	Tue-M-V-6 activity detec keyword s	Voice ction and potting	Tue-M-V-7 Voice voicing	and	
				Lunch Bre	ak			
Tue-A-V-3 Multi- channel speech enhancement and hearing aids	Tue-A-V-4 Self- supervision and semi- supervision for neural ASR training	Tue-A Languag	-V-5 Spoken Je Processing I	Tue-A-V-6 Conversic Adaptat	Voice on and ion II	Tue-A-SS-1 Privo preserving Mach Learning for Aud Speech Process	icy- ine io & ing	Tue-A-S&T-1 Show and Tell 1
				Coffee Bre	ak			
			Cof	fee Break / Lig	ht Dinner			
Tue-E-V-3 Speech signal analysis and representation I	Tue-E-V-4 Spoken Language Understanding I	Tue-E- ASR: transt children low-reso	V-5 Topics in Adaptation, fer learning, 's speech, and purce settings	Tue-E-V-6 Conversic Adaptat	Voice on and ion I			
		/ 01	00.20) 24 - [Drog	ram at	a Glanco	
VV E)21 - 1	rog	i alli al	a Glance	
	Onmed Virtu	lat Ses	sions	Warm-up Co	ffee			
Wed-M-V-3 Speech Synthesis: Singing, Multimodal, Crosslingu Synthesis	Wed-M-V-4 Speecl ual and privac	n coding /	Wed-M-V- percep	5 Speech tion II	Wed-M-V ASR/RN	/-6 Streaming for IN Transducers		
				Lunch Bred	ak			
Wed-A-V-3 Novel neur network architectures ASR	ral Wed-A-V-4 Spe for Localization, Enhar and Quality Asse:	eech acement, ssment	Wed-A-V-5 Synthesis Waveform C	5 Speech s: Neural Generation	Wed- machi	A-V-6 Spoken ne translation	Wed-A-S&T-1 Show and Tell 2	Wed-A-SS-1 SdSV Challenge 2021: Analysis and Exploration of New Ideas on Short- Duration Speaker Verification
			Coff	fee Break / Lig	ht Dinner			
Wed-E-V-3 Neural netw training methods for A	rork Wed-E-V-4 Pros SR features and stru	odic ucture	Wed-E-V-5 Sir speech enh	ngle-channel nancement	Wed- Synthe e	E-V-6 Speech sis: tools, data, valuation		
		II	ITERSPEECH 2	2021 19	BRNO) CZECHIA		

	THURSDAY, 02.09.2021 - Program at a Glance										
TIME (all times are CEST)	Hall A+B	Hall C	Hall D	Lacina	Virtual Hall						
10:30 -11:00				Warm-up (Coffee						
11:00 - 13:00	Thu-M-O-1 Neural Network Training Methods and Architectures for ASR	Thu-M-O-2 Emotion and Sentiment Analysis I	Thu-M-O-3 Linguistic Components in end-to-end ASR	Thu-M-SS-2 Automatic Speech Recognition in Air Traffic Management	Thu-M-SS-1 Oriental Language Recognition	Thu-M-V-1 Assessment of pathological speech and language II	Thu-M-V-2 Multimodal systems				
13:00 - 14:00	Survey Talk										
14:00 - 15:00				Lunch B	reak						
15:00 - 16:00	Keynote										
16:00 - 18:00	Thu-A-O-2 Speech enhancement and coding	Thu-A-O-1 Speech production I	Thu-A-SS-1 Non- Autoregressive Sequential Modeling for Speech Processing			Thu-A-V-1 Emotion and Sentiment Analysis II	Thu-A-V-2 Multi- and cross-lingual ASR, other topics in ASR				
18:00 - 19:00				Coffee Break / L	ight Dinner						
19:00 - 21:00	ISCA General Assembly										

FRIDAY, 03.09.2021 - Program at a Glance

TIME (all times are CEST)	Hall A+B	Hall C	Hall D	Lacina			
10:30 -11:00				Warm-up Coffee			
11:00 - 13:00	Fri-M-O-1 Robust and Far-field ASR	Fri-M-O-2 Speech Synthesis: Prosody Modeling II		Fri-M-SS-1 OpenASR20 and Low Resource ASR Development	Fri-M-V-1 Non-native speech	Fri-M-V-2 Phonetics II	
13:00 - 14:00	Survey Talk						
14:00 - 15:00				Lunch Break			
15:00 - 16:00	Keynote						
16:00 - 18:00	Keynote Fri-A-O-1 Voice activity detection	Fri-A-SS-2 Speech Recognition of Atypical Speech	Fri-A-O-2 Keyword search and spoken language processing	Fri-A-SS-1 INTERSPEECH 2021 Acoustic Echo Cancellation Challenge	Fri-A-V-1 Applications in transcription, education and learning	Fri-A-V-2 Emotion and Sentiment Analysis III	

THURS	DAY. 02.0	9.2021 - F	Program a	t a Glanc	e	
U	nified Virtual Se	ssions	3		-	
		Warm-up (Coffee			
Thu-M-V-3 Source Separation I	Thu-M-V-4 Speaker Diarization I	Thu-M-V-5 Speech Synthesis: Prosody Modeling	Thu-M-V-6 Speech production II	Thu-M-V-7 Spoken Dialogue Systems II		Thu-M-S&T-1 Show and Tell 3
		Lunch B	reak			
Thu-A-V-3 Source Separation II	Thu-A-V-4 Speaker Diarization II	Thu-A-V-5 Speech Synthesis:Toward End- to-End Synthesis I	Thu-A-V-6 Tools, corpora and resources	Thu-A-SS-2 The ADReSSo Challenge: Detecting cognitive decline using speech only		
		Coffee Break / L	ight Dinner			
 FDID	AV 02 00	2021 - Dr	oram at :	a Glanco		
	ind Virtual Sossi		gramat	a Glance		
Olim	ieu vii tuat sessi	Warm up /	20#00			
		wann-up v	Jonee			
Fri-M-V-3 Search/ decoding techniques and confidence measures for ASR	Fri-M-V-4 Speech Synthesis: Linguistic processing, paradigms and other topics	Fri-M-V-5 Speech type classification and diagnosis	Fri-M-V-6 Spoken Term Detection & Voice Search	Fri-M-V-7 Voice Anti-Spoofing and Countermeasure		
		Lunch B	reak			
Fri-A-V-3 Resource- constrained ASR	Fri-A-V-4 Speaker Recognition: Applications	Fri-A-V-5 Speech Synthesis: Speaking Style and Emotion	Fri-A-V-6 Spoken Language Understanding II	Fri-A-S&T-1 Show and Tell 4		

AREAS		
Area Number	Area Name and Color	
1	Speech Perception, Production and Acquisition by Human Listeners	
2	Phonetics, Phonology, and Prosody	
3	Paralinguistics in Speech and Language: Human and Automatic Analysis and Processing	
4	Speaker and Language Identification	
5	Analysis of Speech and Audio Signals	
6	Speech Coding and Enhancement	
7	Speech Synthesis and Spoken Language Generation	

8	Speech Recognition: Signal Processing, Acoustic Modeling, Robustness, Adaptation
9	Speech Recognition: Architecture, Search, and Linguistic Components
10	Speech Recognition: Technologies and Systems for New Applications
11	Spoken dialog systems and conversational analysis
12	Spoken Language Processing: Translation, Information Retrieval, Summarization, Resources and Evaluation
13	Speech, voice, and hearing disorders
14	Special sessions challenges
15	Show and Tell

SOCIAL PROGRAM

Welcome Reception

(all welcome!)

- Tuesday, 31 August, 21:30 to 23:00
- Best Western Premier Hotel International Brno, Conference venue

The traditional conference social opening will take place at the International Hotel right after the conference evening session. We will have plenty of space in the lobby bar and foyer. We will enjoy a few good drinks and some finger food to have a pleasant start to the week.





Student Party (students only!)

- Wednesday, 1 September, 21:30 to 00:00
- Brewery house Poupě, Dominikánská 342/15, Brno

Brno is a student city and with its many cafes, bars and restaurants, it was not easy to choose the right place for a student party. In the end, however, we found the best place for a mutual meeting - the Poupě brewery house, which is located a few minutes walk from the conference venue. This is the best place to enjoy Czech cuisine, taste Czech craft beer and, if restrictions allow, for dancing and partying.

Reviewers' Reception (reviewers only!)

- Wednesday, 1 September, 21:30 to 00:00
- Old Town Hall Brno, Radnická 8, Brno

Many old myths and legends are associated with the old town hall. One of them tells the story of the author of the Gothic portal, sculptor and architect Anton Pilgram. Reviewers can learn this and other stories during a special evening and can also visit the Town Hall Tower. In addition, of course, they can look forward to refreshments and pleasant musical performances.





Conference banquet

(all welcome!)

- Thursday, 2 September, 21:00 to 01:00
- Governor´s Palace, Moravské náměstí 1a, Brno

The history of the Governor's Palace, this remarkable Baroque complex in Brno, dates back to the middle of the 14th century. Today's appearance of the former Augustinian monastery is in fact the result of the Baroque reconstruction of Moritz Grimm from the middle of the 18th century. Following the reforms introduced by Emperor Joseph II. the monks were replaced by local government and state officers who resided here until the end of World War II.

In addition to the halls, there is also a courtyard in this beautiful historic building, where the dance group Javorník Brno will perform at the beginning of the evening and show you a sample of Czech folk dances. The Morgal Café Gallery will prepare for you, among other things, drinks from traditional Czech suppliers. Where else could you experience an evening full of Czech culture and traditions like at the Interspeech 2021 banquet? Don't miss this unique experience!

ORGANIZERS: INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION (ISCA)

ISCA is a non-profit organization. Its original statutes (statutes in French or their translation in English) were deposited on February 23rd at the Prefecture of Grenoble in France by René CARRÉ and registered on March 27th, 1988.

The association started as ESCA (European Speech Communication Association) and, since its foundation, has been steadily expanding and consolidating its activities. It has offered an increasing range of services and benefits to its members and also it has put its financial and administrative functions on a firm professional footing. Indeed, over the ten years of its existence, ESCA has evolved from a small EEC-supported European organization to a fully-independent and self-supporting international association.

At the General Assembly that took place during the Eurospeech conference in Budapest (September 1999), ESCA became a truly international association in the global field of speech science and technology, changing its name to ISCA (International Speech Communication Association) and modifying its statutes accordingly.

The purpose of the association is to promote, in an international world-wide context, activities and exchanges in all fields related to speech communication science and technology. The association is aimed at all persons and institutions interested in fundamental research and technological development that aims at describing, explaining and reproducing the various aspects of human communication by speech, that is, without assuming this enumeration to be exhaustive, phonetics, linguistics, computer speech recognition and synthesis, speech compression, speaker recognition, aids to medical diagnosis of voice pathologies

ORGANIZERS: BRNO UNIVERSITY OF TECHNOLOGY

Brno University of Technology

Established in 1899 by Emperor Franz Joseph I, Brno University of Technology (BUT) is the city's oldest university. Today it offers high-quality studies in engineering, scientific, economic and artistic fields. With its 20 thousand students and 8 faculties, BUT is the nation's largest technical university. Focusing on science and research, the University has five of its own research centres: ADMAS (Advanced Materials, Structures and Technologies), NETME (New Technologies for Mechanical Engineering), CVVOZE (Centre for Research and Utilization of Renewable Energy), SIX (Centre for Sensor, Information and Communication Systems and CMV (Materials Research Centre). It is engaged in two centres of excellence: CEITEC (Central European Institute of Technology) and IT4Innovations National Supercomputer Centre. The university focuses on cooperation with both EU and non-EU partners, has over 600 partnership agreements. BUT and JIC (the South-Moravian Innovation Center, joint venture with other Brno Universities) have been the backbone of regional transformation from heavy industry and agriculture to 21st century technologies. BUT actively participates in the definition of regional and national innovation strategies. BUT took an active part in the formulation of National Artificial Intelligence Strategy of the Czech Republic, and contributes to the Czech Republic position in the planning of the Digital Europe programme.

Brno speech valley

Brno is Czech and European hub of speech R&D and production technologies. BUT Speech@FIT group was formed in 1997 at the Faculty of Electrical Engineering and Computer science at BUT, and joined the newly established Faculty of Information Technology (FIT) in January 2002. The group is advised by Prof. Hermansky, managed by Dr. Jan "Honza" Cernocky, and its research director is Dr. Lukas Burget - all involved in Interspeech 2021 as general chairs and lead technical chair respectively. From a local player, BUT Speech@FIT evolved into an internationally recognized group participating in European, US and local projects. The R&D side is complemented by industrial success: Phonexia Phonexia Ltd. (est. 2006), delivers speech analytics solutions to customers in commercial and security/defense sectors, and ReplayWell Ltd. (est. 2011) commercializes BUT's lecture indexing and browsing technology. Brno speech valley has also attracted interest of established Czech companies such as Lingea (electronic and printed dictionaries), TOVEK (information mining) and Optimsys (dialog systems), branches of foreign companies (SpeechMatics and Omilia) as well as new ventures such as Vocalls (voice chat-bots).

INTERSPEECH 2021 ORGANIZING COMMITTEE

General chairs	Hynek Heřmanský (Johns Hopkins University and BUT) Honza Černocký (Brno University of Technology)
Technical chairs	Lukáš Burget (Brno University of Technology) - Lead chair Lori Lamel (LIMSI/LISN) Odette Scharenborg (TU Delft) Petr Motlicek (IDIAP and BUT)
Publications	Ralf Schlüter (RWTH Aachen University) Jindřich Matoušek (University of West Bohemia Plzeň)
Local Organization	Renata Kohlová (Brno University of Technology) Katka Zemanová (Phonexia) Barbara Schuppler (TU Graz)
Plenaries	Mark Gales (University of Cambridge) Jan Hajič (Charles University Prague)
Special Sessions & Challenges	Sakriani Sakti (NAIST / RIKEN AIP) Pavel Ircing (UWB Plzen)
Tutorials	Yenda Trmal (JHU) Anil Kumar Vuppala (IIIT-Hyderabad)
Satellite Workshops	S Umesh (IIT-Madras) Tomohiro Nakatani (NTT)
Industry Liaison	Petr Schwarz (Phonexia) Bhuvana Ramabhadran (Google)
Sponsoring	Ilya Oparin (Apple) Sanjeev Khudanpur (Johns Hopkins University)
Exhibits	Petr Pollák (CTU Prague) Igor Szöke (Brno University of Technology)
Show and Tell	Najim Dehak (Johns Hopkins University) Reinhold Haeb-Umbach (Paderborn University)
Social events	Oldřich Plchot (Brno University of Technology) Ondřej Glembek (Brno University of Technology)
Finances	Sylva Otáhalová (Brno University of Technology) Gernot Kubin (TU Graz)
Publicity	Jan Kleindienst (IBM Prague) Kai Yu (Shanghai Jiao Tong University) Florian Metze (Facebook)
Grants	Esther Klabbers (ReadSpeaker) Mireia Diez Sanchez (Brno University of Technology)
Diversity and inclusion	Julia Hirschberg (Columbia University) Heidi Christensen (University of Sheffield)
Student coordinator	Kateřina Žmolíková (Brno University of Technology) Karel Beneš (Brno University of Technology)

ISCA MEDALIST

The ISCA Medal for Scientific Achievement recognizes and honors an individual each year who has made extraordinary contributions to the field of speech communication science and technology. The award of this ISCA Medal has been made since 1989. This year, ISCA has selected Prof. Dr.-Ing. Hermann Ney as the recipient of the ISCA Medal for Scientific Achievement 2021 "for pioneering and seminal contributions to data-driven methods for automatic speech recognition and machine translation."

Congratulations to Prof. Hermann Ney!

ISCA BEST STUDENT PAPER AWARD

The following 16 papers are shortlisted for the ISCA Best Student Paper Award 2021. Note that there may be changes to some of the paper sessions in the final program.

Yinghao Li, Ali Zare and Nima Mesgarani: StarGANv2-VC: A Diverse, Unsupervised, Non-parallel Framework for Natural-Sounding Voice Conversion

Tue-E-V-6 Tuesday, August 31, 19:00-21:00 Virtual: Voice Conversion and Adaptation I

Christiaan Jacobs and Herman Kamper: Multilingual transfer of acoustic word embeddings improves when training on languages related to the target zero-resource language

Tue-M-V-2 Tuesday, August 31, 09:30-11:30 Virtual: Speech Synthesis: Toward End-to-End Synthesis II

Yuya Chiba and Ryuichiro Higashinaka: Dialogue Situation Recognition for Everyday Conversation Using Multimodal Information

Tue-M-V-4 Tuesday, August 31, 09:30-11:30 Virtual: Spoken Dialogue Systems I

Piyush Vyas, Anastasia Kuznetsova and Donald Williamson: Optimally Encoding Inductive Biases into the Transformer Improves End-to-End Speech Translation

Tue-A-V-1 Tuesday, August 31, 13:30-15:30 Virtual: Acoustic event detection and acoustic scene classification

Tanya Talkar, Nancy Solomon, Douglas Brungart, Stefanie Kuchinsky, Megan Eitel, Sara Lippa, Tracey Brickell, Louis French, Rael Lange and Thomas Quatieri: Acoustic Indicators of Speech Motor Coordination in Adults With and Without Traumatic Brain Injury

Tue-M-O-2 Tuesday, August 31, 09:30-11:30 In-person Oral: Disordered speech

Sarah Li, Colin Annand, Sarah Dugan, Sarah Schwab, Kathryn Eary, Michael Swearengen, Sarah Stack, Suzanne Boyce, Michael Riley and T. Mast: An automatic, simple ultrasound biofeedback parameter for distinguishing accurate and misarticulated rhotic syllables

Tue-A-V-2 Tuesday, August 31, 13:30-15:30 Virtual: Diverse modes of speech acquisition and processing

Anupama Chingacham, Vera Demberg and Dietrich Klakow: Exploring the Potential of Lexical Paraphrases for Mitigating Noise-Induced Comprehension Errors

Wed-M-V-5 Wednesday, September 1, 11:00-13:00 Virtual: Speech perception II

Mani Kumar Tellamekala, Enrique Sanchez, Georgios Tzimiropoulos, Timo Giesbrecht and Michel Valstar: Stochastic Process Regression for Cross-Cultural Speech Emotion RecognitionThu-A-V-1 Thursday, September 2, 16:00-18:00 Virtual: Emotion and Sentiment Analysis II

Junyi Peng, Xiaoyang Qu, Rongzhi Gu, Jianzong Wang, Jing Xiao, Lukas Burget and Jan Černocký: Effective Phase Encoding for End-to-end Speaker Verification

Wed-E-O-1 Wednesday, September 1, 19:00-21:00 In-person Oral: Graph and End-to-End Learning for Speaker Recognition

Andreea-Maria Oncescu, A. Sophia Koepke, João Henriques, Zeynep Akata and Samuel Albanie: Audio Retrieval with Natural Language Queries Wed-E-O-3 Wednesday, September 1, 19:00-21:00 In-person Oral: Speech and audio analysis

Kexun Zhang, Yi Ren, Changliang Xu and Zhou Zhao: WSRGlow: A Glow-based Waveform Generative Model for Audio Super-Resolution

Wed-M-V-4 Wednesday, September 1, 11:00-13:00 Virtual: Speech coding and privacy

Mandana Saebi, Ernest Pusateri, Aaksha Meghawat and Christophe Van Gysel: A Discriminative Entity-Aware Language Model for Virtual Assistants

Wed-A-V-2 Wednesday, September 1, 16:00-18:00 Virtual: Language and Lexical Modeling for ASR

Baptiste Pouthier, Laurent Pilati, Leela Gudupudi, Charles Bouveyron and Frederic Precioso: Active Speaker Detection as a Multi-Objective Optimization with Uncertainty-based Multimodal Fusion

Wed-E-O-2 Wednesday, September 1, 19:00-21:00 In-person Oral: Spoken Language Processing II

Einari Vaaras, Sari Ahlqvist-Björkroth, Konstantinos Drossos and Okko Räsänen: Automatic Analysis of the Emotional Content of Speech in Daylong Child-Centered Recordings from a Neonatal Intensive Care Unit

Thu-A-V-1 Thursday, September 2, 16:00-18:00 Virtual: Emotion and Sentiment Analysis II

Miran Oh, Dani Byrd and Shrikanth Narayanan: Leveraging Real-time MRI for Illuminating Linguistic Velum Action Fri-M-V-2 Friday, September 3, 11:00-13:00 Virtual: Phonetics II

BRNO | CZECHIA

TRAVEL GRANTS

A total of 109 ISCA Travel Grants for virtual or in-person participation in Interspeech 2021 have been awarded. Congratulations to all recipients!

Joglekar Aditya Ranya Aloufi Ivry Amir Tejaswini Ananthanarayana Or Haim Anidjar Siddhant Arora Abhijeet Awasthi Tanuka Bhattacharjee Xiaoyu Bie Deboshree Bose Samuel Broughton Heng-Jui Chang Pan Changjie Gunvant Chaudhari Eunbi Choi Thomas Coy Helena Cuesta Xia Cui Xudong Dai Nilaksh Das Hira Dhamyal

Anuj Diwan Phat Do Qingyun Dou Denis Dresvyanskiy Alfredo Esquivel Jaramillo Yuan Gao Yang Gao Maria Lara Gauder Marc-Antoine Georges Medda Giacomo Xun Gong Pengcheng Guo Zhe-Chen Guo Pascal Hecker Zhenhou Hong Wangrui Hou Wenxin Hou Jing Huang Yu-Lin Huang Waldemar Jęśko Madhav Kashyap

Lee Keon Thanancahi Kongthaworn Mari Ganesh Kumar **Dipesh Kumar Singh** Young Dae Kwon Bodur Kübra Jatin Lamba Tingle Li Xiang Li Jiachen Lian Jheng-Hao Lin Jinjiang Liu Jiawang Liu Jian Luo Haoxin Ma Noa Mansbach Courtney Mansfield Saddler Mark Xu Min Oh Miran Ritika Nandi

Shamila Nasreen Kumar Neeraj Natalia Nessler Huyen Nguyen **Eng Nicholas** Andreea-Maria Oncescu Kapoor Parul Soumava Paul Leonardo Pepino Vyas Piyush Desh Raj Anton Ratnaraiah Nicolae Catalin Ristea Georgios Rizos Mahdin Rohmatillah Amrit Romana **Billington Rosey** Magdalena Rybicka Elena Ryumina Mandana Saebi

CHRISTIAN BENOÎT AWARD

The laureate of the 11th Christian Benoît Award is Daniel Michelsanti. He is an Industrial Postdoc fellow at Demant Enterprises and Aalborg University in Denmark.

Daniel Michelsanti's project is: "Audio-Visual Speech Enhancement - From Pioneering Studies To Recent Advances".

It will be presented by the recipient at the INTERSPEECH 2021 closing ceremony.

The Christian Benoît Award is sponsored jointly by both ISCA and AFCP. It is awarded through a competitive nomination and review process to promising young scientists in the domain of speech communication to further their career in the field. The Award provides the elected scientist with financial support for the development of a personal short-term research project. The Award is valued at 7,500 EUR.

Applications were evaluated by an international committee including experts in the field of speech communication and representatives of the institutions supporting the award. The jury of this 11th edition was composed by:

- Gérard Bailly, chair
- Torbjørn Svendsen, chair of ISCA awards
- Tatsuya Kawahara, ISCA representative
- Hema Murthy, ISCA representative
- Véronique Delvaux, AFCP representative

- Martine Adda-Decker, AFCP representative
- Sankar Mukherjee, winner of the 10th Christian Benoît award
- Marcin Włodarczak, winner of the 9th Christian Benoît award
- Mathilde Fort, winner of the 8th Christian Benoît award

INTERSPEECH 2021 (26 BRNO | CZECHIA

KEYNOTE SPEAKERS



Hermann Ney: Forty years of speech and language processing: from Bayes decision rule to deep learning

Tue 31 Aug 17:00 CEST - Room A+B

We are proud to announce that one of the keynote speeches will be delivered by Hermenn Ney.

Hermann Ney is a professor of computer science at RWTH Aachen University, Germany. His main research interests lie in the area of statistical classification, machine learning and neural networks with specific applications to speech recognition, handwriting recognition, machine translation and other tasks in natural language processing.

He and his team participated in a large number of large-scale joint projects like the German project VERBMOBIL, the European projects TC-STAR, QUAERO, TRANSLECTURES, EU-BRIDGE and US-American projects GALE, BOLT, BABEL. His work has resulted in more than 700 conference and journal papers with an h index of 100+ and 60000+ citations (based on Google scholar). More than 50 of his former PhD students work for IT companies on speech and language technology.

The results of his research contributed to various operational research prototypes and commercial systems. In 1993 Philips Dictation Systems Vienna introduced a large-vocabulary continuous-speech recognition product for medical applications. In 1997 Philips Dialogue Systems Aachen introduced a spoken dialogue system for train table information via the telephone. In VERBMOBIL, his team introduced the phrase-based approach to data-driven machine translation, which in 2008 was used by his former PhD students at Google as a starting point for the service Google Translate. In TC-STAR, his team built the first research prototype system for spoken language translation of real-life domains.

Abstract: When research on automatic speech recognition started, the statistical (or data-driven) approach was associated with methods like Bayes decision rule, hidden Markov models, Gaussian models and expectation-maximization algorithm. Later extensions included discriminative training and hybrid hidden Markov models using multi-layer perceptrons and recurrent neural networks. Some of the methods originally developed for speech recognition turned out to be seminal for other language processing tasks like machine translation, handwritten character recognition and sign language processing. Today's research on speech and language processing is dominated by deep learning, which is typically identified with methods like attention modelling, sequence-to-sequence processing and end-to-end processing.

In this talk, I will present my personal view of the historical developments of research on speech and language processing. I will put particular emphasis on the framework of Bayes decision rule and on the question of how the various approaches developed fit into this framework.

Awards: 2005 Technical Achievement Award of the IEEE Signal Processing Society; 2013 Award of Honour of the International Association for Machine Translation; 2019 IEEE James L. Flanagan Speech and Audio Processing Award; 2021 ISCA Medal for Scientific Achievements.



Pascale Fung: Ethical and Technological Challenges of Conversational AI

Wed 1 Sep 15:00 CEST, Room A+B

Pascale Fung is a Professor at the Department of Electronic & Computer Engineering and Department of Computer Science & Engineering at The Hong Kong University of Science & Technology (HKUST), and a visiting professor at the Central Academy of Fine Arts in Beijing. She is an elected Fellow of the Association for Computational Linguistics (ACL) for her "significant contributions towards statistical NLP, comparable corpora, and building intelligent systems that can understand and empathize with humans". She is a Fellow of the Institute of Electrical and Electronic Engineers (IEEE) for her "contributions to human-machine interactions", and an elected Fellow

of the International Speech Communication Association for "fundamental contributions to the interdisciplinary area of spoken language human-machine interactions". She is the Director of HKUST Centre for AI Research (CAiRE), an interdisciplinary research center on top of all four schools at HKUST. She is the founding chair of the Women Faculty Association at HKUST. She is an expert on the Global Future Council, a think tank for the World Economic Forum where she started to advocate for AI ethics issues since 2015. She represents HKUST on Partnership on AI to Benefit People and Society. She was invited as an AI expert to the UN panel on Lethal Autonomous Weapons, the UN Economic and Social Council, and various EU official panels. She is a member of the IEEE Working Group to develop an IEEE standard - Recommended Practice for Organizational Governance of Artificial Intelligence. She is on the Board of Governors of the IEEE Signal Processing Society. Her research team has won several best and outstanding paper awards at ACL, ACL and NeurIPS conferences and workshops. She is currently the Editor-in-Chief of the ACL Rolling Review system and the Diversity and Inclusion Chair of Neurips 2021.

Abstract: Conversational AI (ConvAI) systems have applications ranging from personal assistance, health assistance to customer services. They have been in place since the first call centre agent went live in the late 1990s. More recently, smart speakers and smartphones are powered with conversational AI with similar architecture as those from the 90s. On the other hand, research on ConvAI systems has made leaps and bounds in recent years with sequence-to-sequence, generation-based models. Thanks to the advent of large scale pre-trained language models, state-of-the-art ConvAI systems can generate surprisingly human-like responses to user queries in open domain conversations, known as chit-chat. However, these generation based ConvAI systems are difficult to control and can lead to inappropriate, biased and sometimes even toxic responses. In addition, unlike previous modular conversational AI systems, it is also challenging to incorporate external knowledge into these models for task-oriented dialog scenarios such as personal assistance and customer services, and to maintain consistency.

With great power comes great responsibility. We must address the many ethical and technical challenges of generation based conversational AI systems to control for bias and safety, consistency, style, knowledge incorporation, etc. In this talk, I will introduce state-of-the-art generation based conversational AI approaches, and will point out remaining challenges of conversational AI and possible directions for future research, including how to mitigate inappropriate responses. I will also present some ethical guidelines that conversational AI systems can follow.



Mounya Elhilali: Adaptive listening to everyday soundscapes

Thu 2 Sep 15:00 CEST, Room A+B

Mounya Elhilali is a professor of Electrical and Computer Engineering at the Johns Hopkins University with a joint appointment in the department of Psychology and Brain Sciences. She directs the Laboratory for Computational Audio Perception and is affiliated with the Center for Speech and Language Processing and the Center for Hearing and Balance. Her research examines sound processing by humans and machines in noisy soundscapes, and investigates reverse engineering intelligent processing of sounds by brain networks with applications to speech and audio technologies and medical systems. She received her Ph.D. degree in Electrical and Computer

Engineering from the University of Maryland, College Park. Dr. Elhilali was named the Charles Renn faculty scholar in 2015, received a Johns Hopkins catalyst award in 2017 and was recognized as outstanding women innovator in 2020. She is the recipient of the National Science Foundation career award and the Office of Naval Research young investigator award.

Abstract: As we navigate our everyday life, we are continuously parsing through a cacophony of sounds that are constantly impinging on our senses. This ability to sieve through everyday sounds and pick-out signals of interest may seem intuitive and effortless, but it is a real feat that involves complex brain networks that balance the sensory signal with our goals, expectations, attentional state and prior knowledge (what we hear, what we want to hear, what we expect to hear, what we know). A similar challenge faces computer systems that need to adapt to dynamic inputs, evolving objectives and novel surrounds. A growing body of work in neuroscience has been amending our views of processing in the brain; replacing the conventional view of 'static' processing with a more 'active' and malleable mapping that rapidly adapts to the task at hand and listening conditions. After all, humans and most animals are not specialists, but generalists whose perception is shaped by experience, context and changing behavioral demands. The talk will discuss theoretical formulations of these adaptive processes and lessons to leverage attentional feedback in algorithms for detecting and separating sounds of interest (e.g. speech, music) amidst competing distractors.



Tomáš Mikolov: Language Modeling and Artificial Intelligence

Fri 3 Sep 15:00 CEST, Room A+B

Tomas Mikolov is a researcher at CIIRC, Prague. Currently he leads a research team focusing on development of novel techniques within the area of complex systems, artificial life and evolution. Previously, he did work at Facebook AI and Google Brain, where he led development of popular machine learning tools such as word2vec and fastText. He obtained a PhD at the Brno University of Technology in 2012 for his work on neural language models (the RNNLM project). His main research interest is to understand intelligence, and to create artificial intelligence that can help people to solve complex problems.

Abstract: Statistical language modeling has been labeled as an AI-complete problem by many famous researchers of the past. However, despite all the progress made in the last decade, it remains unclear how much progress towards truly intelligent language models we made.

In this talk, I will present my view on what has been accomplished so far, and what scientific challenges are still in front of us. We need to focus more on developing new mathematical models with certain properties, such as the ability to learn continually and without explicit supervision, generalize to novel tasks from limited amounts of data, and the ability to form non-trivial long-term memory. I will describe some of our attempts to develop such models within the framework of complex systems.

SURVEY TALKS



Heidi Christensen: Towards automatic speech recognition for people with atypical speech

Tue 31 Aug 11:30 CEST, Room A+B

Abstract: In the last decade we have seen how speech technologies for typical speech have matured and thus enabled the advancement of a multitude of services and technologies including voice-enabled conversational interfaces, dictation and successfully underpinning the use of state-of-the-art NLP techniques. This ever more pervasive offering allows for an often far more convenient and natural way of interacting with machines and systems. However it also represents

an ever-growing gap experienced by people with atypical (dysarthric) voices: people with even just mild-to-moderate speech disorders cannot achieve satisfactory performance with current automatic speech recognition (ASR) systems and hence they are falling further and further behind in terms of their ability to use modern devices and interfaces. This talk will present the major challenges in porting mainstream ASR methodologies to work for atypical speech, discuss recent advances and present thoughts on where the research effort should be focusing to have real impact in this community of potential users. Being able to speak a query or dictate an email offers a lot of convenience to most of us but for this group of people can have significant implications on ability to fully take part in society and life quality.

Bio Dr Heidi Christensen is a Senior Lecturer in Computer Science at the University of Sheffield, United Kingdom. Her research interests are on the application of Al-based voice technologies to healthcare and focus on two main areas: i) the automatic recognition of atypical speech and ii) the detection and monitoring of people's physical and mental health including verbal and non-verbal traits for expressions of emotion, anxiety, depression and neurodegenerative conditions in e.g., therapeutic or diagnostic settings.



Siriam Ganapathy: Uncovering the acoustic cues of COVID-19 infection

Wed 1 Sep 13:00 CEST, Room A+B

Abstract: The investigation of acoustic biomarkers of respiratory diseases has societal and public health impact following the onset of COVID-19 pandemic. The efforts in the pre-pandemic period focused on developing smartphone friendly diagnostic tools for the detection of chronic pulmonary diseases, Tuberculosis and asthmatic conditions using cough sounds. During the past two years, several research works of varying scales have been undertaken by the speech and signal processing community for analyzing the acoustic symptoms of COVID. The motivation for the development of acoustic-based tools for COVID diagnostics arises from the key limitations of

cost, time, and safety of the current gold standard in COVID testing, namely the reverse transcription polymerase chain reaction (RT-PCR) testing.

In this talk, I will survey the major efforts undertaken by groups across the world in i) developing data resources of acoustic signals for COVID-19 diagnostics, and ii) designing models and learning algorithms for tool development. The landscape of data resources ranges from controlled hospital recordings to crowdsourced smartphone-based data. While the primary signal modality recorded is the cough data, the impact of COVID on other modalities like breathing, speech and symptom data are also studied. In the talk, I will also discuss the considerations in designing data resources and machine learning models for COVID detection from acoustic data. The pointers to open-source data resources and tools will be highlighted with the aim of encouraging budding researchers to pursue this important direction.

The talk will conclude by remarking about the progress made by our group, Coswara, where a multi-modal combination of information from several modalities shows the potential to surpass regulatory requirements needed for a rapid acoustic-based point of care testing (POCT) tool.

Bio: Sriram Ganapathy is a faculty member at the Electrical Engineering, Indian Institute of Science, Bangalore, where he heads the activities of the learning and extraction of acoustic patterns (LEAP) lab. Prior to joining the Indian Institute of Science, he was a research staff member at the IBM Watson Research Center, Yorktown Heights, USA. He received his Doctor of Philosophy from the Center for Language and Speech Processing, Johns Hopkins University. He obtained his Bachelor of Technology from College of Engineering, Trivandrum, India and Master of Engineering from the Indian Institute of Science, Bangalore. He has also worked as a Research Assistant in Idiap Research Institute, Switzerland.

At the LEAP lab, his research interests include signal processing, digital health, machine learning methodologies for speech analytics and auditory neuroscience. He is a subject editor for the Speech Communications journal, member of ISCA and a senior member of IEEE. He is the recipient of young scientist awards from Department of Science and Technology (DST), India, Department of Atomic Energy (DAE), India and the Pratiksha Trust, Indian Institute of Science, Bangalore. Over the past 10 years, he has published more than 100 peer-reviewed journals/conference publications in the areas of deep learning, and speech/ audio processing.

29

INTERSPEECH 2021

BRNO CZECHIA



Karen Livescu: Learning speech models from multi-modal data

Thu 2 Sep 13:00 CEST, Room A+B

Abstract: Speech is usually recorded as an acoustic signal, but it often appears in context with other signals. In addition to the acoustic signal, we may have available a corresponding visual scene, the video of the speaker, physiological signals such as the speaker's movements or neural recordings, or other related signals. It is often possible to learn a better speech model or representation by considering the context provided by these additional signals, or to learn with less training data. Typical approaches to training from multi-modal data are based on the idea that models or representations of each modality should be in some sense predictive of the other

modalities. Multi-modal approaches can also take advantage of the fact that the sources of noise or nuisance variables are different in different measurement modalities, so an additional (non-acoustic) modality can help learn a speech representation that suppresses such noise. This talk will survey several lines of work in this area, both older and newer. It will cover some basic techniques from machine learning and statistics, as well as specific models and applications for speech.

Bio: Karen Livescu is an Associate Professor at TTI-Chicago. She completed her PhD in electrical engineering and computer science at MIT. Her main research interests are in speech and language processing, as well as related problems in machine learning. Some specific interests include multi-view representation learning, visually grounded speech models, acoustic word embeddings, new models for speech recognition and understanding, unsupervised and weakly supervised models for speech and text, and sign language recognition from video. Her professional activities include serving as a program chair of ICLR 2019, ASRU 2015/2017/2019, and Interspeech 2022, and on the editorial boards of IEEE OJ-SP and IEEE TPAMI. She is an ISCA fellow and an IEEE SPS Distinguished Lecturer.



Alejandrina Cristia: Child Language Acquisition studied with Wearables

Fri 3 September, 13:00 CEST, Room A+B

Abstract: In recent years, the ease with which we can collect audio (and to a lesser extent visual information) with wearables has improved dramatically. These allow unprecedented access to the speech that children produce, and that which they hear. Although many conclusions drawn from short observations seem to generalize to these naturalistic datasets, others appear questionable based on human annotations of data collected with wearables. Making the best of such recordings also requires unique tool development.

Alejandrina Cristia is a senior researcher at the Centre National de la Recherche Scientifique (CNRS), leader of the Language Acquisition Across Cultures team, and director of the Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP) cohosted by the Ecole Normale Supérieure, EHESS, and PSL. In 2021, she is an invited researcher in the Foundations of Learning Program of the Abdul Latif Jameel Poverty Action Lab (J-PAL), and a guest researcher at the Max Planck Institute for Evolutionary Anthropology. Her long-term aim is to answer the following.

Bio: Alejandrina Cristia - questions: What are the linguistic representations that infants and adults have? Why and how are they formed? How may learnability biases shape the world's languages? To answer these questions, she combines multiple methodologies including spoken corpora analyses, behavioral studies, neuroimaging (NIRS), and computational modeling. This interdisciplinary approach has resulted in over 100 publications in psychology, linguistics, and development journals as well as IEEE and similar conferences. With an interest in cumulative, collaborative, and transparent science, she contributed to the creation of the first meta-meta-analysis platform (metalab.stanford.edu) and several international networks, including saliently the LangVIEW consortium that is leading I/L+/, the First truly global summer/winter school on language acquisition!(https://www.dpss.unipd.it/summer-school-2021/home). She received the 2017 John S. McDonnell Scholar Award in Understanding Human Cognition, the 2020 Médaille de Bronze CNRS Section Linguistique, and an ERC Consolidator Award (2021-2026) for the [ExELang](exelang.fr) project.

TUTORIALS

Monday August 30, 11:00-14:00 CEST

Intonation Transcription and Modelling in Research and Speech Technology Applications -Amalia Arvaniti, Kathleen (Katie) Jepson, Cong Zhang, and Katherine Marcoux

This tutorial covers the theory and practical applications of intonation research. The following three topics will be introduced to speech technology engineers and researchers new to the field of intonation and prosody: a. the fundamentals of the autosegmental-metrical theory of intonational phonology (AM), a widely accepted phonological framework of intonation; b. a range of automatic and manual annotation methods that can create fast or detailed transcriptions of prosody; c. state-of-the-art modelling techniques for explaining intonation.



Amalia Arvaniti is the Chair of English Language and Linguistics at Radboud University, Netherlands. She received her Ph.D. from the University of Cambridge (1991) and has since then held appointments at the University of Kent (2012-2020), the University of California, San Diego (2002-2012), the University of Cyprus (1995-2001), and the University of Oxford (1991-1994). She has published extensively on prosody, particularly on the phonetics and phonology of intonation, and the nature and measurement of speech rhythm. Her research is currently supported by an ERC Advanced grant titled SPRINT which investigates the role of variation in the phonetics and phonology of the intonation systems of English and Greek.

Amalia was co-editor and then editor of the Journal of the International Phonetic Association (2014-2015 and 2015-2019 respectively). She also serves on the editorial board of the Journal of Phonetics, Journal of Greek Linguistics, and the Studies in Laboratory Phonology series of Language Science Press; from 2000 to 2020 she was also on the editorial board of Phonology. She is currently the President of the Executive Permanent Council for the Organisation of the International Congress of Phonetic Sciences (2019-2023).



Kathleen Jepson is a postdoctoral researcher on Amalia Arvaniti's ERC-funded SPRINT project, based at Radboud University. She received her Bachelor's degree (Honours) from the Australian National University in 2013, and her PhD in Linguistics from the University of Melbourne in 2019. Kathleen's doctoral research, supervised by Prof. Janet Fletcher, Dr. Ruth Singer, and Dr. Hywel Stoakes, was a description of aspects of the prosodic system of Djambarrpuyŋu, an Australian Indigenous language. She has experience in conducting data collection for prosodic analysis in remote locations, and developing analyses of under-described languages. Kathleen's research interests include the production and perception of prosody, particularly intonation, as well as language description of under-resourced languages in Australia and the Pacific region.



Cong Zhang is a postdoctoral researcher on the ERC-funded SPRINT project at Radboud University. She is in charge of collecting, analysing, modelling, and interpreting the English intonation data. Cong received her DPhil degree from the University of Oxford in 2018, with a thesis examining the interaction of tone and intonation in Tianjin Mandarin. Following her DPhil, she worked as a TTS Linguistics Engineer at A-Lab, Rokid Inc., where she led a project for developing a Singing Voice Synthesis system. Cong's research covers various aspects of speech prosody; she is also interested in bridging the gap between linguistics theories and speech technology.



Katherine Marcoux is the lab manager of SPRINT. She assists with various aspects of the research process, mainly focusing on data analysis. Marcoux completed her MSc at the Universitat Pompeu Fabra, after which she began her PhD thesis at Radboud University investigating the production and perception of native and non-native Lombard speech. She is currently finalizing her doctoral manuscript.

Neural target speech extraction - Marc Delcroix and Kateřina Žmolíková

Dealing with overlapping speech remains one of the great challenges of speech processing. Target speech extraction consists of directly estimating speech of a desired speaker in a speech mixture, given clues about that speaker, such as a short enrollment utterance or video of the speaker. It is an emergent field of research that has gained increased attention since it provides a practical alternative to blind source separation for processing overlapping speech. Indeed, by focusing on extracting only one speaker, target speech extraction can relax some of the limitations of blind source separation, such as the necessity of knowing the number of speakers, and the speaker permutation ambiguity.

In this tutorial, we will present an in-depth review of neural target speech extraction including audio, visual, and multi-channel approaches, covering the basic concepts up to the most recent developments in the field. We will provide a uniformed presentation of the different approaches to emphasize their similarities and differences. We will also discuss extensions to other tasks such as speech recognition or voice activity detection and diarization.



Marc Delcroix received the M.Eng. degree from the Free University of Brussels, Brussels, Belgium, and the École Centrale Paris, Paris, France, in 2003, and the Ph.D. degree from Hokkaido University, Sapporo, Japan, in 2007. He was a Research Associate with NTT Communication Science Laboratories (CS labs), Kyoto, Japan, from 2007 to 2008 and 2010 to 2012, where he then became a Permanent Research Scientist in 2012. He was a Visiting Lecturer with Waseda University, Tokyo, Japan, from 2015 to 2018. He is currently a Distinguished Researcher with CS labs.

His research interests cover various aspects of speech signal processing such as robust speech recognition, speech enhancement, target speech extraction, model adaptation, etc. Together with Kateřina Žmolíková, they pioneered the field of neural network-based target speech extraction and he has been actively pursuing research on that direction, publishing also early works on target speaker-ASR, audio-visual target speech extraction, and presenting a show-and-tell on the topic at ICASSP'19.

Dr. Delcroix is a member of the IEEE Signal Processing Society Speech and Language Processing Technical Committee (SLTC). He was one of the organizers of the REVERB Challenge 2014 and the ASRU 2017. He was also a senior affiliate at the Jelinek workshop on speech and language technology (JSALT) in 2015 and 2020.



Kateřina Žmolíková received a B.Sc. degree in information technology in 2014 and an Ing. degree in mathematical methods in information technology in 2016 from the Faculty of Information Technology, Brno University of Technology (BUT), Czech Republic, where she is currently working towards her Ph.D. degree. Since 2013, she has been part of the Speech@FIT research group at BUT. She took part in an internship in the Toshiba Research Laboratory in Cambridge in 2014 and in the Signal Processing Research Group in NTT in Kyoto in 2017. She also took part in the Jelinek workshop on speech and language technology in 2015 and 2020.

During her Ph.D. degree, she focuses on the topic of target speech extraction using neural networks. Her research covers the general design of the target-informed neural networks and their integration with multi-channel and automatic speech recognition systems. She has experience with teaching courses in signal processing and speech processing courses at the Brno University of Technology. She also took part in programs outreaching to high-school students, presenting lectures introducing speech technology.

Speech recognition with Next-Generation Kaldi (k2, Lhotce, Icefall) - Piotr Zelasko, Daniel **Povey and Sanjeev Khudanpur**

This tutorial introduces k2, the cutting-edge successor to Kaldi speech processing, which consists of several Python-centric modules to enable building speech recognition systems, along with its enabling counterparts, Lhotse and Icefall. The participants will learn how to perform swift data manipulation with Lhotse; how to build and leverage auto-differentiable weighted finite state transducers with k2; and how these two can be combined to create Pytorch-based, state-of-the-art hybrid ASR system recipes from Snowfall, the precursor to Icefall.



Dr. Daniel Povey is an expert in ASR, best known as the lead author of the Kaldi toolkit and also for popularizing discriminative training (now known as "sequence training" in the form of MMI and MPE). He has worked in various research positions at IBM, Microsoft and Johns Hopkins University, and is now Chief Speech Scientist of Xiaomi Corporation in Beijing, China.

Dr. Piotr Żelasko is an expert in ASR and spoken language understanding, with extensive experience in developing practical and scalable ASR solutions for industrial-strength use. He worked with successful speech processing start-ups - Techmo (Poland) and IntelligentWire (USA, acquired by Avaya). At present, he is a research scientist at Johns Hopkins University.



Prof. Sanjeev Khudanpur has 25+ years of experience working on almost all aspects of human language technology, including ASR, machine translation, and information retrieval. He has lead a number of research projects from NSF, DARPA, IARPA, and industry sponsors, and published extensively. He has trained more than 40 PhD and Masters students to use Kaldi for their dissertation work.

Monday August 30, 15:00-18:00 CEST

An Introduction to Automatic Differentiation with Weighted Finite-State Automata - Awni Hannun

Weighted finite-state automata (WFSAs) have been a critical building block in modern automatic speech recognition. However, their use in conjunction with "end-to-end" deep learning systems is limited by the lack of efficient frameworks with support for automatic differentiation. This limitation is being overcome with the advent of new frameworks like GTN and K2. This tutorial will cover the basics of WFSAs and review their application in speech recognition. We will then explain the core concepts of automatic differentiation and show how to use it with WFSAs to rapidly experiment with new and existing algorithms. We will conclude with a discussion of the open challenges and opportunities for WFSAs to grow as a central component in automatic speech recognition and related applications.



Awni Hannun is a research scientist at the Facebook AI Research (FAIR) lab, focusing on low-resource machine learning, speech recognition, and privacy. He earned a Ph.D. in computer science from Stanford University. Prior to Facebook, he worked as a research scientist in Baidu's Silicon Valley AI Lab, where he co-led the Deep Speech projects.

SpeechBrain: Unifying Speech Technologies and Deep Learning With an Open Source Toolkit - Mirco Ravanelli and Titouan Parcollet

SpeechBrain is a novel open-source speech toolkit natively designed to support various speech and audio processing applications. It currently supports a large variety of tasks, such as speech recognition, speaker recognition, speech enhancement, speech separation, multi-microphone signal processing, just to name a few. This toolkit is very flexible, modular, easy-to-use, well-document, and can be used to quickly develop speech technologies. With this tutorial, we would like to present, for the first time, SpeechBrain to the INTERSPEECH attenders. First, the design and the general architecture of SpeechBrain will be discussed. Then, its flexibility and simplicity will be shown through practical examples on different speech tasks.



Mirco Ravanelli is currently a postdoc researcher at Mila (Université de Montréal) working under the supervision of Prof. Yoshua Bengio. His main research interests are deep learning, speech recognition, far-field speech recognition, cooperative learning, and self-supervised learning. He is the author or coauthor of more than 40 papers on these research topics. He received his PhD (with cum laude distinction) from the University of Trento in December 2017. Mirco is an active member of the speech and machine learning communities. He is founder and leader of the SpeechBrain project.



Titouan Parcollet is an associate professor in computer science at the Laboratoire Informatique d'Avignon (LIA), from Avignon University (FR) and a visiting scholar at the Cambridge Machine Learning Systems Lab from the University of Cambridge (UK). Previously, he was a senior research associate at the University of Oxford (UK) within the Oxford Machine Learning Systems group. He received his PhD in computer science from the University of Avignon (France) and in partnership with Orkis focusing on quaternion neural networks, automatic speech recognition, and representation learning. His current work involves efficient speech recognition, federated learning and self-supervised learning. He is also currently collaborating with the Mila-Quebec AI institute on the SpeechBrain project.

Concept to Code: Semi-Supervised End-To-End Approaches For Speech Recognition -Omprakash Sonie and Venkateshan Kannan

Training Automatic Speech Recognition (ASR) models usually requires transcribing large quantities of audio, which is both expensive and time-consuming. To overcome this limitation, many semi-supervised training approaches have been proposed to take advantage of abundant unpaired audio and text data. In this tutorial we describe the conceptual understanding and implementation of semi-supervised speech applications - Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) applications. We begin the tutorial with concepts for core building blocks which include Speech pre-processing, Transformer, Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN). We also describe the state-of-the-art approaches in this domain, and the key ideas underlying them.

We walk through the code for implementations. We provide details for installation prerequisites and code using Jupyter notebooks with comments on concepts, key steps, visualization and results.

We believe that a self-contained tutorial giving a good overview of the core techniques with sufficient mathematical background along with actual code will be of immense help to participants.



Omprakash "Om" Sonie is a data scientist at Flipkart who has been working on Speech Recognition Systems, Recommender Systems and Natural Language Processing. Om is passionate about providing guidance to budding data scientists for quality machine learning, deep learning and reinforcement learning using DeepThinking.AI platform. Om is the organiser of local Deep Learning meetup. Om plans to write books on Code to Concept for Machine. Om (as primary author) has presented tutorials and conducted hands-on workshops at KDD, WWW (TheWeb), RecSys (2018, 2019), ECIR, IJCAI, GTC-Nvidia and various meet-ups.



Venkateshan Kannan is a data scientist at Flipkart who is presently working in the domain of speech recognition. In the past, he has worked on diverse problems related to complex networks, information theory, disease modeling, dynamic assignment algorithms, vehicle route optimization, etc. He has a PhD. in theoretical physics.

SPECIAL SESSIONS AND CHALLENGES

The Organizing Committee of INTERSPEECH 2021 is proudly announcing the following special sessions and challenges for INTERSPEECH 2021.

Special sessions and challenges focus on relevant 'special' topics which may not be covered in regular conference sessions.

Papers have to be submitted following the same schedule and procedure as regular papers; the papers undergo the same review process by anonymous and independent reviewers.

Speech Recognition of Atypical Speech

While speech recognition systems generally work well on the average population with typical speech characteristics, performance on subgroups with unique speaking patterns is usually significantly worse.

Speech that contains non-standard speech patterns (acoustic-phonetic phonotactic, lexical and prosodic patterns) is particularly challenging, both because of the small population with these speech patterns, and because of the generally higher variance of speech patterns. In the case of dysarthric speech, which is often correlated with mobility or other accessibility limitations, accuracy of existing speech recognition systems is often particularly poor, rendering the technology unusable for many speakers who could benefit the most.

In this oral session, we seek to promote interdisciplinary collaborations between researchers and practitioners addressing this problem, to build community and stimulate research. We invite papers analyzing and improving systems dealing with atypical speech.

Topics of interest include, but are not limited to: Automatic Speech Recognition (ASR) of atypical speech, Speech-to-Speech conversion/normalization (e.g. from atypical to typical), Voice enhancement and convergence to improve intelligibility of spoken content of atypical speech, Automated classification of atypical speech conditions, Robustness of speech processing systems for atypical speech in common application scenarios, Data augmentation techniques to deal with data sparsity, Aspects of creating, managing data quality, and sharing of data sets of atypical speech, and Multi-modal integration (e.g. video and voice) and its application

Organizers Jordan R. Green, MGH Institute of Health Professions, Harvard University

Michael P. Brenner, Harvard University, Google

- Fadi Biadsy, Google
- Bob MacDonald, Google
- Katrin Tomanek, Google

Oriental Language Recognition

Oriental languages are rich and complex. With the great diversity in terms of both acoustics and linguistics, oriental language is a treasure for multilingual research. The Oriental Language Recognition (OLR) challenge has been conducted for 5 years with big success, and demonstrated many novel and interesting techniques devised by the participants.

The main goal of this special session is to summarize the technical advance of OLR 2020, but it will welcome all submissions related to language recognition and multilingual speech processing.

- Organizers Dong Wang (Tsinghua University)
 - Qingyang Hong (Xiamen University)

Xiaolei Zhang (Northwestern Polytechnical University)

Ming Li (Duke Kunshan University)

Yufeng Hao (Speechocean)

Far-field Multi-Channel Speech Enhancement Challenge for Video Conferencing (ConferencingSpeech 2021)

The ConferencingSpeech 2021 challenge is proposed to stimulate research in multi-channel speech enhancement and aims for processing the far-field speech from microphone arrays in the video conferencing rooms. Targeting the real video conferencing room application, the ConferencingSpeech 2021 challenge database is recorded from real speakers. The number of speakers and distances between speakers and microphone arrays vary according to the sizes of meeting rooms. Multiple microphone arrays from three different types of geometric topology are allocated in each recording environment.

The challenge will have two tasks:

Task 1 is multi-channel speech enhancement with a single microphone array and focusing on practical application with real-time requirement.

Task 2 is multi-channel speech enhancement with multiple distributed microphone arrays, which is non-real-time tracking and does not have any constraints so that participants could explore any algorithms to obtain high speech quality.

To focus on the development of algorithms, the challenge requires the close - training condition. Only provided lists of open source clean speech datasets and noise dataset could be used for training. In addition, the challenge will provide the development set, scripts for simulating the training data, baseline systems for participants to develop their systems. The final ranking of the challenge will be decided by subjective evaluation. The subjective evaluation will be performed using Absolute Category Ratings (ACR) to estimate a Mean Opinion Score (MOS) through the Tencent Online Media Subjective Evaluation platform.

- Organizers Wei Rao, Tencent Ethereal Audio Lab, China
 - Lei Xie, Northwestern Polytechnical University, China
 - Yannan Wang, Tencent Ethereal Audio Lab, China
 - Tao Yu, Tencent Ethereal Audio Lab, USA
 - Shinji Watanabe, Associate Professor, Carnegie Mellon University / Johns Hopkins University, USA
 - Zheng-Hua Tan, Aalborg University, Denmark
 - Hui Bu, AISHELL foundation, China
 - Shidong Shang, Tencent Ethereal Audio Lab, China

Voice quality characterization for clinical voice assessment: Voice production, acoustics, and auditory perception

The appraisal of voice quality is relevant to the clinical care of disordered voices. It contributes to the selection and optimization of clinical treatment as well as to the assessment of the outcome of the treatment. Levels of description of voice quality include the biomechanics of the vocal folds and their kinematics, temporal and spectral acoustic features, as well as the auditory scoring of hoarseness, hyper- and hypo-functionality, creakiness, diplophonia, harshness, etc. Broad and fuzzy definitions of terms regarding voice quality are in use, which impede scientific and clinical communication.

The aim of the special session is to contribute to the improvement of the clinical assessment of voice quality via a translational approach, which focuses on quantifying and explaining relationships between several levels of description. The objective is to gather new insights, advancement of knowledge and practical tools to assist researchers and clinicians in obtaining effective descriptions of voice quality and reliable measures of its acoustic correlates. Topics of interest include, but are not limited to , (i) the statistical analysis and automatic classification, possibly relying on state-of-the-art machine learning approaches, of distinct types of voice quality via non-obtrusively recorded features, (ii) the analysis and simulation of vocal fold vibrations by means of analytical, kinematic or mechanical modelling, (iii) the interpretation and modeling of both acoustic emission and/or high-speed video recordings such as videolaryngoscopy and videokymography, (iv) the synthesis of disordered voices jointly with auditory experimentation involving synthetic and natural disordered voice stimuli.

Organizers Philipp Aichinger (philipp.aichinger@meduniwien.ac.at)

Abeer Alwan (alwan@ee.ucla.edu)

Carlo Drioli (carlo.drioli@uniud.it)

Jody Kreiman (jkreiman@ucla.edu)

Jean Schoentgen (jschoent@ulb.ac.be)

Automatic Speech Recognition in Air Traffic Management (ASR-ATM)

Air-traffic management is a dedicated domain where in addition to using the voice signal, other contextual information (i.e. air traffic surveillance data, meteorological data, etc.) plays an important role. Automatic speech recognition is the first challenge in the whole chain. Further processing usually requires transforming the recognized word sequence into the conceptual form, a more important application in ATM. This also means that the usual metrics for evaluating ASR systems (e.g. word error rate) are less important, and other performance criteria (i.e. objective such as command recognition error rate, callsign detection accuracy, overall algorithmic delay, real-time factor, or reduced flight times, or subjective such as decrease of a workload of the users) are employed.

The main objective of the special session is to bring together ATM players (both academic and industrial) interested in ASR and ASR researchers looking for new challenges. This can accelerate near future R&D plans to enable an integration of speech technologies to the challenging, but highly safety oriented air-traffic management domain.

- Organizers Hartmut Helmke (DLR)
 - Pavel Kolcarek (Honeywell)
 - Petr Motlicek (Idiap Research Institute)
Alzheimer's Dementia Recognition through Spontaneous Speech: The ADReSS Challenge

Dementia is a category of neurodegenerative diseases that entails a long-term and usually gradual decrease of cognitive functioning. The main risk factor for dementia is age, and therefore its greatest incidence is amongst the elderly. Due to the severity of the situation worldwide, institutions and researchers are investing considerably on dementia prevention and early detection, focusing on disease progression. There is a need for cost-effective and scalable methods for detection of dementia from its most subtle forms, such as the preclinical stage of Subjective Memory Loss (SML), to more severe conditions like Mild Cognitive Impairment (MCI) and Alzheimer's Dementia (AD) itself.

The ADReSSo (ADReSS, speech only) targets a difficult automatic prediction problem of societal and medical relevance, namely, the detection of Alzheimer's Dementia (AD). The challenge builds on the success of the ADReSS Challenge (Luz et Al, 2020), the first such shared-task event focused on AD, which attracted 34 teams from across the world. While a number of researchers have proposed speech processing and natural language processing approaches to AD recognition through speech, their studies have used different, often unbalanced and acoustically varied data sets, consequently hindering reproducibility and comparability of approaches. The ADReSSo Challenge will provide a forum for those different research groups to test their existing methods (or develop novel approaches) on a new shared standardized dataset. The approaches that performed best on the original ADReSS dataset employed features extracted from manual transcripts, which were provided. The ADReSSo challenge provides a more challenging and improved spontaneous speech dataset, and requires the creation of models straight from speech, without manual transcription. In keeping with the objectives of AD prediction evaluation, the ADReSSo challenge's dataset will be statistically balanced so as to mitigate common biases often overlooked in evaluations of AD detection methods, including repeated occurrences of speech from the same participant (common in longitudinal datasets), variations in audio quality, and imbalances of gender and age distribution. This task focuses AD recognition using spontaneous speech, which marks a departure from neuropsychological and clinical evaluation approaches. Spontaneous speech analysis has the potential to enable novel applications for speech technology in longitudinal, unobtrusive monitoring of cognitive health, in line with the theme of this year's INTERSPEECH, "Speech Everywhere!".

Organizers Saturnino Luz, Usher Institute, University of Edinburgh

Fasih Haider, University of Edinburgh

Sofia de la Fuente, University of Edinburgh

Davida Fromm, Carnegie Mellon University

Brian MacWhinney, Carnegie Mellon University

SdSV Challenge 2021: Analysis and Exploration of New Ideas on Short-Duration Speaker Verification

Are you searching for new challenges in speaker recognition? Join SdSV Challenge 2021 which focuses on the analysis and exploration of new ideas for short duration speaker verification.

Following the success of the SdSV Challenge 2020, the SdSV Challenge 2021 focuses on systematic benchmark and analysis on varying degrees of phonetic variability on short-duration speaker recognition. The challenge consists of two tasks.

Task 1 is defined as speaker verification in text-dependent mode where the lexical content (in both English and Persian) of the test utterances is also taken into consideration.

Task 2 is defined as speaker verification in text-independent mode with same- and cross-language trials.

The main purpose of this challenge is to encourage participants on building single but competitive systems, to perform analysis as well as to explore new ideas, such as multi-task learning, unsupervised/self-supervised learning, single-shot learning, disentangled representation learning and so on, for short-duration speaker verification. The participating teams will get access to a train set and the test set drawn from the DeepMine corpus which is the largest public corpus designed for short-duration speaker verification with voice recordings of 1800 speakers. The challenge leaderboard is hosted at CodaLab.

Organizers Hossein Zeinali (Amirkabir University of Technology, Iran)

Kong Aik Lee (I2R, A*STAR, Singapore)

Jahangir Alam (CRIM, Canada)

Lukáš Burget (Brno University of Technology, Czech Republic)

Acoustic Echo Cancellation (AEC) Challenge

The INTERSPEECH 2021 Acoustic Echo Cancellation (AEC) challenge is designed to stimulate research in the AEC domain by open sourcing a large training dataset, test set, and subjective evaluation framework. We provide two new open source datasets for training AEC models. The first is a real dataset captured using a large-scale crowdsourcing effort. This dataset consists of real recordings that have been collected from over 5,000 diverse audio devices and environments. The second is a synthetic dataset with added room impulse responses and background noise derived from the INTERSPEECH 2020 DNS Challenge. An initial test set will be released for the researchers to use during development and a blind test near the end which will be used to decide the final competition winners. We believe these datasets are large enough to facilitate deep learning and representative enough for practical usage in shipping telecommunication products.

Organizers Ross Cutler, Microsoft Corp, USA Ando Saabas, Microsoft Corp, Tallinn Tanel Parnamaa, Microsoft Corp, Tallinn Markus Loide, Microsoft Corp, Tallinn Sten Sootla, Microsoft Corp, Tallinn Hannes Gamper, Microsoft Corp, USA Sebastian Braun, Microsoft Corp, USA Karsten Sorensen, Microsoft Corp, USA Robert Aichner, Microsoft Corp, USA Sriram Srinivasan, Microsoft Corp, USA

Non-Autoregressive Sequential Modeling for Speech Processing

Non-autoregressive modeling is a new direction in speech processing research that has recently emerged. One advantage of non-autoregressive models is their decoding speed: decoding is only composed of forward propagation through a neural network, hence complicated left-to-right beam search is not necessary. In addition, they do not assume a left-to-right generation order and thus represent a paradigm shift in speech processing, where left-to-right, autoregressive models have been believed to be legitimate. This special session aims to facilitate knowledge sharing between researchers involved in non-autoregressive modeling across various speech processing fields, including, but not limited to , automatic speech recognition, speech translation, and text to speech, via panel discussions with leading researchers followed by a poster session.

Organizers Katrin Kirchhoff (Amazon)

Shinji Watanabe (Carnegie Mellon University)

Yuya Fujita (Yahoo Japan Corporation)

DiCOVA: Diagnosis of COVID-19 using Acoustics

The COVID-19 pandemic has resulted in more than 93 million infections, and more than 2 million casualties. Large scale testing, social distancing, and face masks have been critical measures to help contain the spread of the infection. While the list of symptoms is regularly updated, it is established that in symptomatic cases COVID-19 seriously impairs normal functioning of the respiratory system. Does this alter the acoustic characteristics of breathe, cough, and speech sounds produced through the respiratory system? This is an open question waiting for answers. A COVID-19 diagnosis methodology based on acoustic signal analysis, if successful, can provide a remote, scalable, and economical means for testing of individuals. This can supplement the existing nucleotides based COVID-19 testing methods, such as RT-PCR and RAT.

The DiCOVA Challenge is designed to find answers to the question by enabling participants to analyze an acoustic dataset gathered from COVID-19 positive and non-COVID-19 individuals. The findings will be presented in a special session at Interspeech 2021. The timeliness, and the global societal importance of the challenge warrants focussed effort from researchers across the globe, including from the fields of medical and respiratory sciences, mathematical sciences, and machine learning engineers.

Organizers Neeraj Sharma (Indian Institute of Science, Bangalore, India)

Prasanta Kumar Ghosh (Indian Institute of Science, Bangalore, India)

Srikanth Raj Chetupalli (Indian Institute of Science, Bangalore, India)

Sriram Ganapathy (Indian Institute of Science, Bangalore, India)

Deep Noise Suppression Challenge – INTERSPEECH 2021

The Deep Noise Suppression (DNS) challenge is designed to foster innovation in the area of noise suppression to achieve superior perceptual speech quality. We recently organized a DNS challenge special session at INTERSPEECH 2020 and ICASSP 2020. We open-sourced training and test datasets for the wideband scenario. We also open-sourced a subjective evaluation framework based on ITU-T standard P.808, which was used to evaluate challenge submissions. Many researchers from academia and industry made significant contributions to push the field forward, yet even the best noise suppressor was far from achieving superior speech quality in challenging scenarios. In this version of the challenge organized at INTERSPEECH 2021, we are expanding both our training and test datasets to accommodate full band scenarios. The two tracks in this challenge will focus on real-time denoising for (i) wide band, and (ii) full band scenarios. We are also making available a reliable non-intrusive objective speech quality metric for wide band called DNSMOS for the participants to use during their development phase. The final evaluation will be based on ITU-T P.835 subjective evaluation framework that gives the quality of speech and noise in addition to the overall quality of the speech.

We will have two tracks in this challenge:

Track 1: Real-Time Denoising track for wide band scenario

The noise suppressor must take less than the stride time Ts (in ms) to process a frame of size T (in ms) on an Intel Core i5 quadcore machine clocked at 2.4 GHz or equivalent processor. For example, Ts = T/2 for 50% overlap between frames. The total algorithmic latency allowed including the frame size T, stride time Ts, and any look ahead must be less than or equal to 40ms. For example, for a real-time system that receives 20ms audio chunks, if you use a frame length of 20ms with a stride of 10ms resulting in an algorithmic latency of 30ms, then you satisfy the latency requirements. If you use a frame of size 32ms with a stride of 16ms resulting in an algorithmic latency of 48ms, then your method does not satisfy the latency requirements as the total algorithmic latency exceeds 40ms. If your frame size plus stride T1=T+Ts is less than 40ms, then you can use up to (40-T1) ms future information.

Track 2: Real-Time Denoising track for full band scenario

Satisfy Track 1 requirements but at 48 kHz.

More details about the datasets and the challenge are available in the paper and the challenge github page. Participants must adhere to the rules of the challenge.

Organizers Chandan K A Reddy (Microsoft Corp, USA) Hari Dubey (Microsoft Corp, USA) Kazuhito Koishada (Microsoft Corp, USA) Arun Nair (Johns Hopkins University, USA) Vishak Gopal (Microsoft Corp, USA) Ross Cutler (Microsoft Corp, USA) Robert Aichner (Microsoft Corp, USA) Sebastian Braun (Microsoft Research, USA) Hannes Gamper (Microsoft Research, USA) Sriram Srinivasan (Microsoft Corp, USA)

Privacy-preserving Machine Learning for Audio, Speech and Language Processing

This special session focuses on privacy-preserving machine learning (PPML) techniques in speech, language and audio processing, including centralized, distributed and on-device processing approaches. Novel contributions and overviews on the theory and applications of PPML in speech, language and audio are invited. We encourage submissions related to ethical and regulatory aspects of PPML in this context. Sending speech, language or audio data to a cloud server exposes private information. One approach called anonymization is to preprocess the data so as to hide information which could identify the user by disentangling it from other useful attributes. PPML is a different approach, which solves this problem by moving computation near the clients. Due to recent advances in Edge Computing and Neural Processing Units on mobile devices, PPML is now a feasible technology for most speech, language and audio applications that enables companies to train on customer data without needing them to share the data. With PPML, data can sit on a customer's device where it is used for model training. During the training process, models from several clients are often shared with aggregator nodes that perform model averaging and sync the new models to each client. Next, the new averaged model is used for training on each client. This process continues and enables each client to benefit from training data on all other clients. Such processes were not possible in conventional audio/speech ML. On top of that, high-quality synthetic data can also be used for training thanks to advances in speech, text, and audio synthesis.

Organizers Harishchandra Dubey (Microsoft)

Amin Fazel (Amazon, Alexa)

Mirco Ravanelli (MILA, Université de Montréal)

Emmanuel Vincent (Inria)

Computational Paralinguistics ChallengE (ComParE) - COVID-19 Cough, COVID-19 Speech, Escalation & Primates

Interspeech ComParE is an open Challenge dealing with states and traits of speakers as manifested in their speech signal's properties. In this 13th edition, we introduce four new tasks and Sub-Challenges:

COVID-19 Cough based recognition,

COVID-19 Speech based recognition,

Escalation level assessment in spoken dialogues,

Primates classification based on their vocalisations.

Sub-Challenges allow contributors to find their own features with their own machine learning algorithm. However, a standard feature set and tools including recent deep learning approaches are provided that may be used. Participants have five trials on the test set per Sub-Challenge. Participation has to be accompanied by a paper presenting the results that undergoes the Interspeech peer-review.

Contributions using the provided or equivalent data are sought for (but not limited to):

Participation in a Sub-Challenge

Contributions around the Challenge topics

Organizers Björn Schuller (University of Augsburg, Germany / Imperial College, UK)

Anton Batliner (University of Augsburg, Germany)

Christian Bergler (FAU, Germany)

Cecilia Mascolo (University of Cambridge, UK)

Jing Han (University of Cambridge, UK)

Iulia Lefter (Delft University of Technology, The Netherlands)

Heysem Kaya (Utrecht University, The Netherlands)

OpenASR20 and Low Resource ASR Development

The goal of the OpenASR (Open Automatic Speech Recognition) Challenge is to assess the state of the art of ASR technologies for low-resource languages.

The OpenASR Challenge is an open challenge created out of the IARPA (Intelligence Advanced Research Projects Activity) MATERIAL (Machine Translation for English Retrieval of Information in Any Language) program that encompasses more tasks, including CLIR (cross-language information retrieval), domain classification, and summarization. For every year of MATERIAL, NIST supports a simplified, smaller scale evaluation open to all, focusing on a particular technology aspect of MATERIAL. The capabilities tested in the open challenges are expected to ultimately support the MATERIAL task of effective triage and analysis of large volumes of data, in a variety of less-studied languages.

The special session aims to bring together researchers from all sectors working on ASR for low-resource languages to discuss the state of the art and future directions. It will allow for fruitful exchanges between OpenASR20 Challenge participants and other researchers working on low-resource ASR. We invite contributions from OpenASR20 participants, MATERIAL performers, as well as any other researchers with relevant work in the low-resource ASR problem space.

Topics: OpenASR20 Challenge reports, including

Cross-lingual training techniques to compensate for ten-hour training condition

Factors influencing ASR performance on low resource languages by gender and dialect

Resource conditions used for unconstrained development condition

IARPA MATERIAL performer reports on low-resource ASR, including

Low Resource ASR tailored to MATERIAL's Cross Language Information Retrieval Evaluation

Genre mismatch condition between speech training data and evaluation

Other topics focused on low-resource ASR challenges and solutions

Organizers Peter Bell, University of Edinburgh

Jayadev Billa, University of Southern California Information Sciences Institute

William Hartmann, Raytheon BBN Technologies

Kay Peterson, National Institute of Standards and Technology

ISCA STUDENT ADVISORY COMMITTEE (SAC)

The aim of ISCA SAC is to put forward ideas for the expansion of student activities within ISCA and to implement them. ISCA SAC Board members are Jasper Ooster (General Coordinator), Catarina Botelho (Event Coordinator), and Francisco Teixeira (Media Coordinator). This year's local student volunteer is Kateřina Žmolíková. We are always looking for volunteers! If you are interested, come to our events, listen to our podcast "Speech Pitch", and feel free to get in touch with us.

This year, we are pleased to announce that ISCA-SAC will host three student events at INTERSPEECH 2021 in Brno, Czech Republic: Students Meet Experts, Doctoral Consortium as well as the Mentoring event, which is co-organized with ISCA's Diversity Committee.

7th Doctoral Consortium

Sunday, August 29th, 2021, Online

The ISCA-SACis pleased to announce its 7th Doctoral Consortium. This event gives doctoral candidates the opportunity to present and discuss their research with a panel of experts. The discussion includes feedback on the evolution and progress of the students, in order to help them identify a road-map towards refining their thesis. Similar to last year's edition, this year, the Doctoral Consortium will be held online just before the beginning of Interspeech 2021.

Participants were selected based on their submitted abstracts. If you missed this year's submission deadline for the Doctoral Consortium please consider submitting an abstract next year. We are looking forward to it!

3rd Mentoring by ISCA-SAC and ISCA's Diversity Committee

Wednesday, September 1st, 16.00-18.00 CEST, Online

After two successful editions of the mentoring event, at INTERSPEECH 2019 in Graz and INTERSPEECH 2020 in Shanghai (online), the SCA-SAC joined forces with ISCA's Diversity Committee to organize the third Mentoring Event at INTERSPEECH 2021 in Brno (online).

In previous editions, PhD students were given the opportunity to engage in a discussion with early-career and senior researchers from academia and industry, in a warm environment. This year, we will extend the Mentoring event to the entire community, i.e., we invite all researchers in any stage of their career. We will have two types of mentoring: round tables and one-on-one mentoring.

Round table discussions:

To cater for potential different needs and wishes, we will have round tables only for PhD students and round tables for a mix of people at different career stages. The PhD round tables will have two mentors and 6-8 PhD students; the mixed career round tables will have one discussion leader and 6-8 participants. Each table will have an assigned topic and the mentors/discussion leaders will be chosen and invited accordingly.

The topics we propose for discussion are presented below. This list is only tentative and we may change it depending on the interests of the participants and mentors/discussion leaders. We are open to suggestions for other topics, during the application process!

Topics:

- 1.) Successes, failures and imposter syndrome
 - 2.) Time management: work-life balance, combining academic career with side-projects or jobs, ...
 - 3.) Starting a family during my career
 - 4.) Doing research in academia vs doing research in industry
 - 5.) Professional development: planning ahead
 - 6.) Essentials of publishing
 - 7.) Competitive academic environment, publishing pressure and slow science
 - 8.) Being the "odd one out": only woman, only person of color, only person in your research field
 - 9.) Plagiarism, harassment and other unethical behaviour
 - 10.) Creating/maintaining your own research line/group

One-on-one mentoring:

Parallel to this event, we will start enabling one-on-one mentoring sessions. If you would feel more comfortable discussing certain topics in a one-on-one session, you can also register for it in the application form, and we will do our best to find a mentor matching your needs. We will then introduce mentor and participant pairs, and let you arrange the best time/place for your conversation.

For the application please visit http://www.isca-students.org/ or the students section on the Interspeech website.



8th Students meet Experts at Interspeech 2021

Thursday, September 2nd, 16.00-18.00 CEST, Online

After successful editions in Lyon (2013), Singapore (2014), San Francisco (2016), Stockholm (2017), Hyderabad (2018), Graz (2019), and virtually in Shanghai (2020) we are excited to announce that the Students Meet Experts event is now coming to INTERSPEECH 2021 in Brno.

We will have a panel discussion with experts from academia and industry, where the experts respond to questions submitted by students. To submit questions please visit http://www.isca-students.org/ or the students section on the INTERSPEECH website. All students are welcome to participate!

AREA CHAIRS

1. Speech Perception, Production and Acquisition

- Martin Cooke, Ikerbasque
- Jeesun Kim, Western Sydney University
- Fanny Meunier, University of Cote d'Azur

2. Phonetics, Phonology and Prosody

- Margaret Zellers, University of Kiel
- Priyankoo Sarmah, Indian Institute of Technology Guwahati
- Prasanta Ghosh, Indian Institute of Science

3. Analysis of Paralinguistics in Speech and Language

- Khiet Truong, University of Twente
- Shrikanth Narayanan, University of Southern California
- Chloé Clavel, Télécom Paris
- Carol Espy-Wilson, University of Maryland

4. Speaker and Language Identification

- Kong Aik Lee, NEC Corporation
- Niko Brummer, Phonexia
- Luciana Ferrer, National Council of Scientific and Technical Research, Argentina
- Alicia Lozano, Universidad Autonoma de Madrid
- Srikanth Madikeri, Idiap Research Institute

5. Analysis of Speech and Audio Signals

- Mark Hasegawa-Johnson, University of Illinois
- Jingdong Chen, Northwestern Polytechnical University
- Mounya Elhilali, Johns Hopkins University
- Zbyněk Koldovský, Technical University of Liberec
- Koichi Shinoda, Tokyo Institute of Technology

6. Speech Coding and Enhancement

- Tom Bäckström, Aalto University
- Marc Delcroix, NTT Communication Science Laboratories
- Ina Kodrasi, Idiap Research Institute
- Vladimir Malenovsky, University of Sherbrooke
- John Hansen, University of Taxes at Dallas, USA

7. Speech Synthesis and Spoken Language Generation

- Hema Murthy, Indian Institute of Technology Madras
- Esther Klabbers-Judd, ReadSpeaker
- Jindrich Matousek, University of West Bohemia
- Berrak Sisman, National University of Singapore
- Heiga Zen, Google Inc.

8. Speech Recognition – Signal Processing, Acoustic Modeling, Robustness and Adaptation

- Thomas Hain, University of Sheffield
- Martin Karafiat, Brno University of Technology
- Penny Karanasou, Amazon
- Karen Livescu, Toyota Technological Institute at Chicago
- Alexandros Potamianos, National Technical University of Athens
- Umesh Srinivasan, Indian Institute of Technology Madras

9. Speech Recognition – Architecture, Search, and Linguistic Components

- Preethi Jyothi, Indian Institute of Technology Bombay
- Murat Saraclar, Boğaziçi University
- Ngoc Thang Vu, University of Stuttgart
- Bhuvana Ramabhadran, Google

10. Speech Recognition – Technologies and Systems for New Applications

- Torbjørn Svendsen, Norwegian University of Science and Technology
- Eric Fossler Lussier, The Ohio State University
- Florian Metze, Carnegie Mellon University

11 Spoken dialog systems and conversational analysis

- Dilek Hakkani-Tur, Amazon
- Kristiina Jokinen, AIST Tokyo Waterfront
- Catharine Oertel, TU Delft

12 Spoken Language Processing: Translation, Information Retrieval, Summarization, Resources and Evaluation

- Bin Ma, Alibaba Inc.
- Jan Trmal, Johns Hopkins University
- Kate Knill, University of Cambridge

13 Speech, voice, and hearing disorders

- Mathew Magimai Doss, Idiap Research Institute
- Heidi Christensen, University of Sheffield
- Bernd T. Meyer, University of Oldenburg

SATELLITE EVENTS

The 11th ISCA Speech Synthesis Workshop (SSW11)

Extended submission deadline: April 26, 2021 23:59 (AoE) Date of the event: August 26-28, 2021 Venue: hybrid (online and Vital Hotel Nautis, Gárdony, Hungary)

First Shared Task on Automatic Minuting (AutoMin'2021)

Submission deadline: June 1, 2021 (System Submission) & July 10, 2021 (System Report Submission) Date of the event: September 4, 2021 Venue: TBD (check updates on the event webpage)

Workshop on Machine Learning in Speech and Language Processing 2021 (MLSLP 2021)

Submission deadline: June 10, 2021 Date of the event: September 6, 2021 Venue: Online Event (please see the event webpage for information)

The 1st International Clarity Workshop on Machine Learning Challenges for Hearing Aids (Clarity-2021)

Submission deadline: June 15, 2021 (challenge and regular papers) Date of the event: September 16-17, 2021 Venue: Online

Automatic Speaker Verification Spoofing And Countermeasures Challenge (ASVspoof 2021)

Submission deadline: July 9, 2021 Date of the event: September 16, 2021 Venue: Fully online event

Speech, Music and Mind 2021 (SMM21)

Submission deadline: June 25, 2021 Date of the event: August 27, 2021 Venue: Virtual format

24th International Conference on Text, Speech and Dialogue (TSD2021)

Submission deadline: April 18, 2021 Date of the event: September 6-9 2021 Venue: Olomouc, Czech Republic

4th International Workshop on the History of Speech Communication Research (HSCR 2021)

Submission deadline: May 16, 2021 Date of the event: August 27-28, 2021 Venue: Hybrid (Online and Institute of Phonetics, Charles University, n. J. Palacha 2, Prague, Czech Republic)

Workshop for Young Female Researchers in Speech Science & Technology (YFRSW 2021)

Submission deadline: June 4 2021 (abstract up to 300 words) Date of the event: August 29 2021 Venue: online

Automatic Speech Recognition in Air Traffic Management (ASR-ATM)

Date of the event: August 30, 2021 14:00 CEST Venue: Faculty of Information, Technology, Brno University of Technology + virtual

https://www.interspeech2021.org/satellites

SCIENTIFIC REVIEW COMMITTEE

Prathosh A P Alberto Abad Solomon Teferra Abate Mourad Abbas Ossama Abdel-Hamid Mohammed Abdelwahab Åsa Abelin Ajish Kuriakose Abraham Vinayak Abrol Alex Acero Sivanand Achanta MP Actlin Jeeva Sharath Adavanne Gilles Adda Martine Adda-Decker Farah Adeeba Yossi Adi Nagaraj Adiga Mohamed Afify Sachin Agarwal Shyam Agrawal Purvi Agrawal Carla Agurto Mohsin Ahmed Masato Akagi Yuya Akita Maryam Al Dabel Oumayma Al Dakkak Mohammed Salah Al-Radhi Firoj Alam Md Jahangir Alam Aaron Albin Felix Albu Yahya Aldholmi Jan Alexandersson Anastasios Alexandridis Ghada AlHarbi Ahmed Ali Paavo Alku Alexandre Allauzen Sarah AlShareef

Tanel Alumäe Abeer Alwan Gilbert Ambrazaitis Angélique Amelot Noam Amir Shahin Amiriparian Tasos Anastasakos Asger Heidemann Andersen Bistra Andreeva Jesús Andrés-Ferrer Walter Andrews Jorn Anemuller Pongtep Angkititrakul Xavier Anguera Vijendra Apsingekar Takayuki Arai Shoko Araki Dhany Arifianto Yasuo Ariki Yoshiko Arimoto Ebru Arisoy Marc Arnela Hagai Aronowitz Levent Arslan Michael Ashby Eva Liina Asu Bishnu Atal Kartik Audhkhasi Nicolas Audibert Carlos Avendano Bharathi B Harald Baayen Molly Babel Deepak Baby Tom Bäckström Sameer Badaskar Pierre Badin Paolo Baggia Ye Bai Gérard Bailly Pallavi Baljek

Andrea Bandini Yoshiaki Bando Jorge Baptista Plinio Barbosa Nelly Barbot Jon Barker Dante Barone Roberto Barra-Chicote Vincent Barriac Murali Karthick Baskar Fernando Batista Anton Batliner Timo Baumann Ali Orkan Bayer Frederic Bechet Tilman Becker Steve Beet Homayoon Beigi Julie Belião Peter Bell Jerome Bellegarda Ashwin Bellur Sofia Ben Jebara Atef Ben Youssef Štefan Beňuš Mohamed Faouzi Benzeghiba Christina Bergmann Nicole Beringer Jared Bernstein Frederic Berthommier Lauren Besacier Simon Betz Sujeeth Bharadwaj Jayashree Bhat Sreekar Bhaviripudi Peter Birkholz Judith Bishop Maria Paola Bissiri Arijit Biswas Cynthia Blanco

Jose Luis Blanco Murillo Ronald Böck Tobias Bocklet Bajibabu Bollepalli Jean-Francois Bonastre Zinny Bond Nandini Bondale Anne Bonneau Jonas Borgstrom Hynek Boril Tomáš Bořil Michal Borsky Mariem Bouafif Gilles Boulianne Rachel F. Bouserhal Pierre-Michel Bousquet Suzanne Boyce Anna Bradlow David Braude Bettina Braun NorbertBraunschweiler Hervé Bredin Andrew Breen John Bridle Markus Brückl **Raymond Brueckner** Laurence Bruggeman Alejna Brugos Niko Brummer Pablo Brusco Alessio Brutti H Timothy Bunnell Susanne Burger Felix Burkhardt Carlos Busso Dani Byrd Vera Cabarrão Joao Cabral Peter Cahill Weicheng Cai Luis Caldas de Oliveira

44 📕

Zoraida Calleias Jose Ramon Calvo de Lara Nathalie Camelin Yuan Cao Valentín Cardeñoso-Payo Joe Caroselli Julie Carson-Berndsen Diamantino Caseiro **Diego** Castan Angel Mario Castro Martinez Benjamin Cauchi Chiara Celata Christophe Cerisara Milos Cernak Jan Černocký Petr Cerva Damian Cetnarowicz Soumitro Chakrabarty William Chan Senthil Chandramohan Guan-Lin Chao Eric Charton Theodora Chaspari Sourish Chaudhuri Ciprian Chelba Kuan-Yu Chen **Qiang Chen** Tongzhou Chen Ling-Hui Chen I-Fan Chen Zhehuai Chen Yun-Nung Chen Nancy Chen Wenda Chen Xie Chen Berlin Chen Fei Chen Liping Chen Hongjie Chen Zhuo Chen Jingdong Chen Jian Cheng Gaofeng Cheng Rathinavelu Chengalvarayan

Nivedita Chennupati Chen-Yu CHIANG Yuya Chiba Jen-Tzung Chien Ioana Chitoran **Fleanor Chodroff** Gérard Chollet Jan Chorowski Khalid Choukri Shammur Absar Chowdhury Mads Christensen Heidi Christensen Yu-An Chung Minhwa Chung Joon Son Chung Vataya Chunwijitra Robert Clark Chloé Clavel Martin Cmejrek Jennifer Cole Claudia Roberta Combei Martin Cooke Erica Cooper Hugo Cordeiro Ricardo Cordoba Piero Cosi Marius Cotescu Christophe Coupe Andre Coy Olivier Crouzet Marisa Cruz Tamás Gábor Csapó Jia Cui Xiaodong Cui Sandro Cumani Nicholas Cummins Ross Cutler Francesco Cutugno Christophe d'Alessandro Luis Fernando D'Haro Mariapaola D'Imperio Deborah Dahl Rasmus Dall Siddharth Dalmia

Géraldine Damnati Jianwu Dang Falavigna Daniele Khalid Daoudi Amit Das Mandal S K Das Rohan Kumar Das Tran-Huy Dat Nauman Dawalatabad José Mario De Martino Renato De Mori Febe De Wet Carme de-la-Mota Gilles Degottex Najim Dehak Reda Dehak Ali Dehghan Firoozabadi Michael Deisher Marc Delcroix Héctor Delgado Symeon Delikaris Manias Andrea DeMarco Andrea Deme Cenk Demiroglu Yasuharu Den Huiqun Deng Ratnadeep Deshmukh Giuseppe Di Fabbrizio Ivana Didirkova Mireia Diez Leena Dihingia Erinç Dikici **Dimitrios Dimitriadis** Hongwei Ding Amazouz Djegdjiga Cong-Thanh Do Simon Dobnik Simon Dobrišek Gerry Docherty Laura Docio-Fernandez Rama Sanand Doddipatla Marion Dohen Minghui Dong Mortaza (Morrie) Doulaty

Saad Dr. Maged Szymon Drgas Carlo Drioli Jasha Droppo Konstantinos Drossos Jun Du Zhiyao Duan **Richeng Duan** Harishchandra Dubev Jacques Duchateau Sophie Dufour **Richard Dufour** Fwan Dunbar Daniel Duran Ondřej Dušek Thierry Dutoit Sushanta Kabir Dutta Indranil Dutta Bernardt Duvenhage Jens Edlund Vaclav Eksler Mounya Elhilali Wendy Elvira-García Mahsa Elyasi Dimitra Emmanouilidou Julien Epps Hakan Erdogan Iwagami Eri Donna Erickson Daniel Erro Engin Erzin David Escudero-Mancebo Felipe Espic Carol Espy-Wilson Yannick Estève Nicholas Evans Vera Evdokimova Florian Eyben Mauro Falcone Isabel Falé Tiago Falk Cunhang Fan Fuming Fang Shiwei Fang

Jérôme Farinas Mojtaba Farmani Kevin Farrell Mireia Farrús Friedrich Faubel Camille Fauth **Benoit Fauve** Haytham Fayek Marcello Federico Ingo Feldhausen Junlan Feng Raul Fernandez Fernando Fernández Martínez Isabelle Ferrané Marc Ferras Luciana Ferrer Carlos Ferrer **Tim Fingscheidt** Volker Fischer Anouschka Foltz Lionel Fontan Eric Fosler-Lussier Cécile Fougeron Petr Fousek Thiago Fraga da Silva Horacio Franco Corinne Fredouille Sónia Frota Ruibo Fu Susanne Fuchs **Guillaume Fuchs Robert Fuchs** Mark Fuhs Masakiyo Fujimoto Yuya Fujita Yusuke Fujita Takashi Fukuda Takahiro Fukumori Keiichi Funaki Adamantios Gafos Olivier Galibert Ascension Gallardo-Antolin Sriram Ganapathy Ankur Gandhe

Suryakanth V Gangashetty Siva Reddy Gangireddy Sharon Gannot Jian Gao **Yingming Gao** Fernando García Leibny Paola Garcia Perera Sri Garimella Philip N. Garner Maeva Garnier Harinath Garudadri Nikolay Gaubitch Yashesh Gaur Roberto Gemello Nithin George Munir Georges Panaviotis Georgiou Branislav Gerazov Timo Gerkmann Francois Germain Omid Ghahabi Sahar Ghannay Samujjwal Ghosh Sucheta Ghosh Prasanta Ghosh Arnab Ghoshal Daniele Giacobello Dafydd Gibbon Emer Gilmartin Laurent Girin K V Vijay Girish Mircea Giurgiu Ondrej Glembek Juan Ignacio Godino llorente Rahul Goel Nagendra Goel Stefan Goetze Pavel Golik Christian Gollan Angel Gomez Pedro Gómez-Vilda Jian Gong Yuan Gong Yifan Gong

Rosa Gonzalez Hautamäki Jose A. Gonzalez-Lopez Arseniy Gorin Mária Gósy Gábor Gosztolya Martijn Goudbeek Philippe Gournay Evandro Gouvea D Govind Dhananjaya Gowda Martin Graciarena Simone Graetzer Calbert Graham Bénédicte Grandon David Grayden Phil Green Adrien Gresse Frantisek Grezl David Griol Peter Grosche Tamás Grósz Wentao Gu Michele Gubian Jon Gudnason David Guennec Tanaya Guha Rodrigo Guido Camille Guinaudeau Batuhan Gundogdu Radha Krishna Guntur Chitralekha Gupta Shikha Gupta Vishwa Gupta Sanjeev Gurugopinath Joakim Gustafson Ricardo Gutierrez-Osuna Jayanna H S Emanuël Habets Tania Habib Hossein Hadian Reinhold Haeb-Umbach Christina Hagedorn Parisa Haghani Martin Hagmüller

Seonaiun Hahm Stefan Hahn Fasih Haider Thomas Hain Dilek Hakkani-Tur Pierre Halle Simon Hammond **Jiqing Han** Kyu Han Jing Han Abualsoud Hanani Cemal Hanilci Mirko Hannemann Eric Hansen John H.L. Hansen Philip Harding Jonathan Harrington William Hartmann David Harwath Madina Hasan Taufiq Hasan Mark Hasegawa-Johnson Ville Hautamaki Weipeng He Lei He Yanzhang He Larry Heck Peter Heeman Masoumeh Heidari Kapourchali Paul Heisterkamp Hartmut Helmke John Henderson Hee-Soo Heo **Christian Herbst** Enno Hermann Anne Hermes Inma Hernaez Luis Hernandez-Gomez Gabriel Hernandez-Sierra Javier Hernando Sanjika Hewavitharana James Hieronymus Stefan Hillmann Ivan Himawan

🤇 46 🍃

Andrew Hines Yusuke Hioka Takuya Hiraoka Keikichi Hirose Ruediger Hoffmann **Bjorn Hoffmeister** Volker Hohmann Kiyoshi Honda **Qingyang Hong** Florian Hönig **Pierre-Edouard Honnet** Ron Hoory Chiori Hori Takaaki Hori Merle Horne David House lan Howard **Christine Howes** Wei-Ning Hsu Fang Hu Ke Hu Shoukang Hu Qiong Hu Che-Wei Huang Yiheng Huang **Qiang Huang** Jing Huang **Rongqing Huang** Yinghui Huang Chien-Lin Huang Gongping Huang Mark Huckvale Thomas Hueber David Huggins Daines Vincent Hughes Lluís-F. Hurtado Brian Hutchinson Mathilde Hutin Yusuke liima Aravind Illa Satoshi Imaizumi David Imseng Hirofumi Inaguma Koji Inoue

Pavel Ircing Toshio Irino Markus Iseli Masato Ishizaki Koii Iwano Mahaveer Jain Mimansa Jaiswal Artur Janicki Adam Janin David Janiszek **Stephen Jannetts Esther Janse** Adèle Jatteau Peter Jax Jesper Jensen Anna Jespersen Alexandra Jesse Michael Jessen Luis M.T. Jesus Minho Jin Qin Jin Cheolwoo Jo Aditya Joglekar Michael Johnson Michael Johnston Kristiina Jokinen Oliver Jokisch Arnaud Joly Gareth Jones Shrikant Joshi Vikas Joshi Denis Jouvet **Tim Juergens** Ante Jukić Preethi Jyothi Deepak K T Tokihiko Kaburagi Abdellah Kacha Tatiana Kachkovskaia Zdravko Kacic Stanisław Kacprzak Sudarsana Reddy Kadiri George Kafentzis Takayuki Kagomiya

Alexander Kain Sofoklis Kakouros Kaustubh Kalgaonkar Ozlem Kalinli Sishir Kalita Yutaka Kamamoto Madhu Kamble Hirokazu Kameoka Svetlana Kaminskaia Herman Kamper Ahilan Kanagasundaram Naoyuki Kanda John Kane Shiyin Kang Hong-Goo Kang Chieh-Chi Kao Martin Karafiat Penny Karanasou Fredrik Karlsson Alexey Karpov James Kates Hemant Kathania Hiroaki Kato Tsuneo Kato Nassos Katsamanis Kouichi Katsurada Hideki Kawahara Tatsuya Kawahara Heysem Kaya Simon Keizer Amelia Kelly Niamh Kelly Tom Kenter Seppo Keronen Santosh Kesiraju Aparna Khare Shreya Khare Yerbolat Khassanov Khaled KHELIF Kevin Kilgour Hoi Rin Kim Wooil Kim Hyung Soon Kim Kwangyoun Kim

Jona-mi Kim Seokhwan Kim Hong Kook Kim Sanghun Kim Samuel Kim Sunhee Kim Eugene Kim Minje Kim Yeon-Jun Kim Jaebok Kim Jeesun Kim Brian King Simon King Brian Kingsbury Tomi Kinnunen Keisuke Kinoshita Irina Kipyatkova Andreas Søeborg Kirkedal Norihide Kitaoka Esther Klabbers W. Kleijn Thomas Kleinbauer Stefan Kleiner Ondrej Klejch Katarzyna Klessa Viacheslav Klimkov Nicola Klingler Kate Knill Hanseok Ko Satoshi Kobashikawa Akio Kobayashi Kazuhiro Kobayashi Takao Kobayashi Daniil Kocharov Marcel Kockmann Ina Kodrasi Sri Rama Murty Kodukula Tina Kohler Daniel Kohlsdorf Yuma Koizumi Jachym Kolar Zbyněk Koldovský Thomas Kollar Dorothea Kolossa

Kazunori Komatani Mariko Kondo Miloslav Konopik Myoung-Wan Koo Sunil Kumar Kopparapu Jacques Koreman Tomoki Koriyama Srikanth Korse Daniel Korzekwa Takafumi Koshinaka Sandeep Reddy Kothinti Maria Koutsogiannaki György Kovács Konrad Kowalczyk Pavel Kral Ivan Kraljevski Kelsey Kraus Jody Kreiman Sreenivasa Rao Krothapalli Roland Kuhn Vivek Kumar Varun Kumar Anurag Kumar Manoj Kumar Naveen Kumar Shankar Kumar R Kumaraswamy Jimmy Kunzmann Gakuto Kurata Mikko Kurimo Frank Kurth Pietro Laface Catherine Lai Ying-Hui Lai Divesh Lala Pierre Lanchantin Adrian Lancucki lan Lane Itshak Lapidot **Yves** Laprie Anthony Larcher Kornel Laskowski Lukas Latacz Siddique Latif

Javier Latorre Antoine Laurent Galina Lavrentyeva Aaron Lawson Duc Le Viet Bac Le Phule Gaël Le Lan Sébastien Le Maguer Jonathan Le Roux **Benjamin Lecouteux** Jaewon Lee Lin-shan Lee Kong Aik Lee Shi-wook Lee Chin-Hui Lee Yoonjeong Lee Chi-Chun Lee Hung-yi Lee Bowon Lee Roch Lefebvre Milan Legát Yun Lei Manon Lelandais Cheung-Chi LEUNG Gary Leung Michael Levit Rivka Levitan Gina-Anne Levow Haifeng Li Sheng Li Junfeng Li Aijun Li Ruirui Li Ruizhi Li Yongwei Li Lantian Li Kehuang Li Bo Li Na Li Ming Li Jinyu Li Haizhou Li Haoqi Li

Shan Liang Paul Pu Liang Yuan-Fu Liao Hank Liao Christopher Liberatore Tatiana Likhomanenko Rasa Lileikyte Jason Lilley Carlos Lima Amaro Lima **QIGUANG LIN** Ju Lin Baihan Lin Kin Wah Edward Lin Georges Linares Zhenhua Ling Pärtel Lippus **Pierre Lison Jianming Liu** Yang Liu Li Liu Pengfei Liu Xiaoyu Liu Chunxi Liu Yi-Wen Liu Gang Liu Yuzong Liu Xunying Liu Bin Liu Rui Liu Karen Livescu Eduardo Lleida Solano Deborah Loakes Anders Lofqvist Yanhua Long David Looney José Lopes Iván López-Espejo Paula Lopez-Otero Teresa Lopez-Soto Jaime Lorenzo-Trueba Anastassia Loukina Frfan Loweimi Alicia Lozano-Diez

Heng Lu Xugang Lu Liang Lu Bogdan Ludusan Loren Lugosch Beata Lukaszewicz Steven Lulich Patrick Lumban Tobing Elena Lyakso Gurunath Reddy M Changxue Ma Jeff Ma Min Ma Ning Ma Jianbo Ma Bin Ma **Roland Maas** Ewen MacDonald Javier Macias-Guarasa Maulik Madhavi Srikanth Madikeri Hari Krishna Maganti Mathew Magimai Doss Paul Magron Ranniery Maia Soumi Maiti Manwai Mak Brian Mak Tomasz Maka Vladimir Malenovsky Zofia Malisz Sri Harish Mallidi Nursadul Mamun Michael Mandel Kazunori Mano Vimal Manohar Hongda Mao Boháč Marek Soroosh Mariooryad Mirko Marras Philippe MARTIN Rainer Martin David Martínez González Carlos-D. Martínez-Hinarejos

Ricard Marxer Hinako Masuda Takashi Masuko Ryo Masumura Ana Isabel Mata Marco Matassoni Pavel Matejka Jindrich Matousek Spyros Matsoukas Tomoko Matsui Yuri Matveev julie mauclair Olga Maxwell Juan Manuel Mayor Torres Tara McAllister Alan McCree Erik McDermott Mitchell McLaren Ian McLoughlin Michael McTear Ivan Medennikov Sylvain Meignier Alexsandro Meireles Geoffrey Meltzner Chiara Meluzzi Joram Meron Thomas Merritt Angeliki Metallinou Marie Meteer Florian Metze **Christine Meunier** Fanny Meunier Bernd T. Meyer Lei Miao Jan Michalsky Michel Hoen Michel Hoen Antonio Miguel Peter Mihajlik Ben Milner Rosanna Milner Masato Mimura Yasuhiro Minami Nobuaki Minematsu Bahman Mirheidari

Seyedmahdad Mirsamadi Jagabandhu Mishra Ananya Misra Abhinav Misra Hemant Misra Vikramjit Mitra Yuki Mitsufuji Gabriel Mittag Dr. Vinay Kumar Mittal Hansjörg Mixdorff Bernd Möbius Takemi Mochida Abdelrahman Mohamed Alexis Moinet Peggy Mok Sebastian Möller Helena Moniz Vahid Montazeri Pratibha Moogi Elliot Moore Roger Moore Tine Mooshammer Mohamed Morchid Hiroki Mori Maho Morimoto Niko Moritz Laureano Moro Velazquez Athanasios Mouchtaris Emily Mower Provost Pejman Mowlaee losif Mporas Hannah Muckenhirn Manjunath Mulimani Ludek Muller Markus Müller Hema Murthy Xingyu Na Babak Naderi Tohru Nagano Yasuko Nagano-Madsen Venki Nagesha Kathleen Nagle Kazuhiro Nakadai Seiichi Nakagawa

Satoshi Nakamura Toru Nakashika Tomohiro Nakatani Arivudai Nambi Mahesh Kumar Nandwana Praveen Narayanan Shrikanth Narayanan Arun Narayanan Md Nasir Karan Nathwani Andreas Nautsch Eva Navas Shekhar Nayak Matteo Negri Jana Neitsch Géza Németh Shahan Nercessian Friedrich Neubarth Graham Neubig Michael Neumann Albert Newton Hermann Ney Noel Nguyen Trung Hieu Nguyen Patrick Nguyen Jinfu Ni Chongjia Ni Mauro Nicolao Oliver Niebuhr Kuniko Nielsen Thomas Niesler Lasquido Nio Masafumi Nishida Masayuki Nishiguchi Masafumi Nishimura Takanobu Nishiura Hiromitsu Nishizaki Nobuyuki Nishizawa Raquel Norel Elmar Nöth Jan Nouza David Novick Sergey Novoselov Stavros Ntalampiras

Aditya Arie Nugraha Markus Nussbaum-Thom Douglas O'Shaughnessy Nicolas Obin Keiko Ochi Tsubasa Ochiai Catharine Oertel Atsunori Ogawa Tetsuji Ogawa Yoo Rhee Oh Yamato Ohtani Lucas Ondel Dan Oneată Ilya Oparin Irit Opher Brian Ore Juan Rafael Orozco-Arroyave Alfonso Ortega Mari Ostendorf Nele Ots Youssef Oualil Slim Ouni Win Pa Pa Trideba Padhi Sarala Padi Mukund Padmanabhan Gangamohan Paidi Serguei Pakhomov Shruti Palaskar Yi-Cheng Pan Ho-hsien Pan Sankaran Panchapagesan Ashish Panda Prem C. Pandey Alexandros Papangelis Emilia Parada-Cabaleiro Titouan parcollet José Pardo Douglas Parham Naveen Parihar Daniel Park Soo Jin Park Sangwook Park Patrick Paroubek

Srinivas Parthasarathv Sarangarajan Parthasarathy Ankita Pasad Tanvina Patel Dr. Vaishali V. Patil Jose Patino Dipjyoti Paul Rahul Pawar Vijavaditva Peddinti Antonio M. Peinado Carmen Peláez-Moreno Jason Pelecanos Thomas Pellegrini Xavier Pelorson Mikel Penagarikano Gerald Penn Joshua Penney Fernando Perdigão Javier Perez Rubén Pérez Ramón José L. Pérez-Córdoba Franz Pernkopf Pascal Perrier Olivier Perrotin Pablo Peso Parada Stavros Petridis Caterina Petrone Huy Phan Michael Picheny Roberto Pieraccini Marv Pietrowicz Claire Pillot-Loiseau Julien Pinguier Shadi Pirhosseinloo Ramin Pishehvar Christian Plahl Oldrich Plchot Jouni Pohjalainen Florian Pokorny Joseph Polifroni Archontis Politis Petr Pollak Tim Polzehl Anna Pompili

Abhimanyu Popli François Portet **Brechtie Post** Alexandros Potamianos Gerasimos Potamianos Marianne Pouplier Dan Povey Kishore Prahallad Jeena Prakash Kristin Precoda Patti Price Michael Proctor Jorge Proenca Emily Prud'hommeaux Tarun Pruthi Michael Pucher Ernest Pusateri Felix Putze Xinyuan Qian Kun Qian Thomas Quatieri Azam Rabiee Ganna Raboshchuk Martin Radfar Zeynab Raeesy Anton Ragni Md Hafizur Rahman Tuomo Raitio Padmanabhan Rajan Avni Rajpal Nitendra Rajput Anirudh Raju Dhananjay Ram Bhuvana Ramabhadran Vikram Ramanarayanan V Ramasubramanian Anssi Rämö Shreyas Ramoji Daniel Ramos Vivek Kumar Rangarajan Sridhar Shivesh Ranjan Wei Rao Kanishka Rao Preeti Rao

Okko Räsänen Ramya Rasipuram Anabela Rato Mirco Ravanelli Christian Raymond Manny Rayner Melissa A. Redford Henning Reetz Klaus Reinhard Margaret Renwick Fernando Gil Vianna **Resende Junior** Dayana Ribas Ricardo Ribeiro Manuel Sam Ribeiro Giuseppe Riccardi Korin Richmond Korbinian Riedhammer Luca Rigazio Michael Riley Albert Rilliard Fabien Ringeval Christian Ritz Matthew Roddy Luis Javier Rodriguez-Fuentes Rosalia Rodriguez-Vazguez Axel Roebel Rungpat Roengpitya Johan Rohdin Mikael Roll Srikanth Ronanki Andrey Ronzhin **Richard Rose** Ralph Rose Andrew Rosenberg Solange Rossato Sophie Rosset Jean-Luc Rouas Mickael Rouvier Somnath Roy Alexander Rudnicky Frank Rudzicz Nicholas Ruiz Vesa Ruoppila

Martin Russell Michela Russo David Rybach Omid Sadjadi Rahim Saeidi Sylvie Saget Goutam Saha Lakshmi Saheer Md Sahidullah Saurabh Sahu Hardik Sailor Tara Sainath Daisuke Saito Yuki Saito Shinsuke Sakai Ken-Ichi Sakakibara Sakriani Sakti Elizabeth Salesky Gláucia Laís Salomão Elliot Saltzman Stan Salvador Giampiero Salvi Eugenia San Segundo Fernández Rubén San Segundo Hernández Jon Sanchez Victoria Sanchez Germán Sanchis-Trilles **Bonny Sands** Ananth Sankar Joao Felipe Santos George Saon Ashtosh Sapru Murat Saraclar Ibon Saratxaga Seyyed Saeed Sarfjoo Leda Sari Ruhi Sarikaya Achintya kumar Sarkar Mousmita Sarma Priyankoo Sarmah Milton Sarria-Paja Akira Sasou Antonio Satue Vilar

Michelina Savino Oscar Saz Thomas Schaaf Ellika Schalling Robin Scheibler Henning Schepker Ralf Schlüter Sven Schmeier Maximilian Schmitt Jean Schoentgen Tanja Schultz Barbara Schuppler Petr Schwarz Antje Schweitzer Jilt Sebastian João Sedoc Chandra Sekhar Seelamantula Carlos Segura Jose Segura Ethan Selfridge Gregory Sell Michael Seltzer **Gregory Senay** Cheol Jae Seong Dmitriy Serdyuk Romain Serizel Joan Serrà Christophe Servan Guruprasad Seshadri Vidhyasaharan Sethu Lee Seunghun Nirmesh Shah Milind Shah Mostafa Shahin Mahaboob Ali Basha Shaik Ravi Shankar Kalluri Shareef Babu Hamid Sharifzadeh Neeraj Sharma Bidisha Sharma Hershev Shawn Slava Shechtman Imran Sheikh Peng Shen

Sven Shepstone Tetsuya Shimamura Hiroshi Shimodaira Jiyoung Shin Koichi Shinoda Takahiro Shinozaki Takahito Shinya Sayaka Shiota Suwon Shon Elizabeth Shriberg Stephen Shum Ingo Siegert Stephan Sigg Siddharth Sigtia Vered Silber-Varod Jan Silovsky Fabrice Silva Khe Chai Sim Juraj Šimko Flavio Simoes Konstantin Simonchik Madhusudan Singh Mittul Singh Sabato Marco Siniscalchi Olivier siohan Berrak Sisman Dinkar Sitaram Sunayana Sitaram Ganesh Siyaraman Sunit Sivasankaran Gabriel Skantze Jan Skoglund Pavel Skrelin Malcolm Slaney Raymond Slyh Kamel Smaili Flena Sokolova Rubén Solera-Ureña Hagen Soltau Mitchell Sommers Peng Song Eunwoo Song Meet Soni Mee SONU

Frank Soong Alexander Sorin Victor Sorokin Laura Spinu Stefano Squartini Ajay Srinivasamurthy Sridhar Srinivasan Sundararaian Srinivasan Anuroop Sriram Kaavya Sriskandaraja Siddharth Srivastava Themos Stafylakis Adriana Stan **Tiomkin Stas** Mark Steedman Sabrina Stehwien John Steinberg Georg Stemmer **Evgeny Stepanov Richard Stern** George Sterpu Hywel Stoakes Andreas Stolcke Simon Stone Svetlana Stoyanchev Helmer Strik Nikko Strom Sebastian Stüker Yannis Stylianou Yi Su Ming-Hsiang Su Aswin Shanmugam Subramanian Ondrej Šuch Youngjoo Suh Hanwu Sun Sining Sun Ming Sun Harshavardhan Sundar Shiva Sundaram Loredana Sundberg Cerrato Gajan Suthokumar Masayuki Suzuki Jan Švec Torbjørn Svendsen

Ali Raza Sved György Szaszák Eva Szekely Igor Szoke Dávid Sztahó Martha Yifiru Tachbelie Marie Tahon Naoya Takahashi Toru Takahashi Shinji Takaki Aniruddha Tammewar Tian Tan Zheng-Hua Tan Tien Ping Tan Hiroki Tanaka Yun Tang Zhiyuan Tang **Qingming Tang** Hao Tang Christina Tånnander Jianhua Tao Fei Tao Naohiro Tawara António Teixeira João Paulo Teixeira Carlos Teixeira Louis ten Bosch Joseph Tepperman Pr. Sten Ternström Veena Thenkanidiyoor Barry-John Theobald **Oliver Thiergart** Mohan Thippeswamy Samuel Thomas Jill Thorson Xiaohai Tian Mark Tiede Daniel Tihelka Noé Tits Tomoki Toda Massimiliano Todisco Roberto Togneri Shinichi Tokuma Doroteo T. Toledano

Vikrant Singh Tomar Kanako Tomaru Natalia Tomashenko Laura Tomokiyo Renjie Tong Pedro Torres-Carrasquillo László Tóth Asterios Toutios Tien Dung Tran Isabel Trancoso Ruchir Travadi Jan Trmal Juergen Trouvain Khiet Truong Yu Tsao Pirros Tsiakoulis Andreas Tsiartas Nikos Tsourakis Kimiko Tsukada Ming Tu Hitesh Tulsiani Máté Ákos Tündik Gokhan Tur Oytun Turk Markku Turunen Zoltán Tüske Michael Tyler Stefan Ultes Srinivasan Umesh Masashi Unoki Maria Uther Vishnu Vidyadhara Raju V Martti Vainio Cassia Valentini-Botinhao Jean-Marc Valin Dirk Van Compernolle Kristin Van Engen Hugo Van hamme Charl van Heerden David van Leeuwen Maarten Van Segbroeck Rob van Son Carlos Vaquero Amparo Varona

Adriana Vasilache Ioana Vasilescu Juan Camilo Vásquez Correa Christophe Veaux Karjigi Veena Nanette Veilleux Rajbabu Velmurugan Jithendra Vepa Dimitra Vergyri Lyan Verwimp Karel Veselý Klara Vicsi Bianca Vieru Marina Vigario Samudra Vijaya P Vijayalakshmi Deepu Vijavasenan Coriandre Vilain Jesus Villalba Fernando Villavicencio **Emmanuel Vincent** Ravichander Vipperla Tuomas Virtanen Srikanth Vishnubhotla Bogdan Vlasenko Carl Vogel Jan Volín Stephen Voran Ngoc Thang Vu Vedran Vukotic Anil Kumar Vuppala Hari Krishna Vydana Petra Wagner Mary Ann Walter Patrick Wambacq Vincent Wan Michael Wand Yongqiang Wang Yiming Wang **Qiongqiong Wang** Hsin-Min Wang Hao Wang Xiaofei Wang Jun Wang

Jun Wang Haipeng Wang Shuai Wang Xin Wang Dongmei Wang Nigel Ward Shinii Watanabe Catherine Watson Ratree Wavland James Wayman Philip Weber Jianguo Wei Lenka Weingartova Eugene Weinstein **Benjamin Weiss** Ron Weiss Liang Wen Chao Weng Kurt Werner Gordon Wichern Kyle Williams Donald Williamson Colin Wilson Jonathan Wintrode Alan Wisler Marcin Wlodarczak Matthias Wolff Krzysztof Wolk Maria Wolters Phil Woodland Chuck Wooters Jane Wottawa Chunyang Wu Zhizheng Wu Mengyue Wu Junru Wu Xixin Wu Chung-Hsien Wu Zhiyong Wu Yaru WU Chai Wutiwiwatchai Hilary Wynne Xianjun Xia Wei Xia

Bo Xiao Xiong Xiao Feng-Long Xie Yanlu Xie Xurong Xie Chenglin Xu Jiaming Xu Yi Xu Kele Xu Haihua Xu Mingiang Xu Yong Xu Hainan Xu Shaofei Xue Yawen Xue Wei Xue Nelson Yalta Junichi Yamagishi Hitoshi Yamamoto Kazumasa Yamamoto Yonghong Yan Yike Yang Lei Yang Bo Yang **Xuesong Yang** Longfei Yang Zhaojun Yang Seung Hee Yang Umit Yapanel Mahsa Yarmohammadi Chiranjeevi Yarra Keiichi Yasu Guoli Ye Bayya Yegnanarayana Raymond Yeh Hani Camille Yehia Jiangyan Yi Emre Yilmaz Selma Yılmazyıldız Kayaarma Sevinj Yolchuyeva Chang Yoo Dongsuk YOOK Su-Youn Yoon Tsukasa Yoshinaga

🤇 52 🖕

Koichiro Yoshino	Guanlong Zhao
Changhuai You	Baigong Zheng
Dong Yu	Cong Zhou
Dong Yu	Xinhui Zhou
Kai Yu	Jian Zhu
Chengzhu Yu	Xuan Zhu
Ha-Jin Yu	Xiaodan Zhuang
Yougen Yuan	Bartosz Ziolko
Jiahong Yuan	Imed Zitouni
Young-Sun Yun	Hatice Zora
François Yvon	Catalin Zorila
Stephen Zahorian	Enrico Zovato
Zbynek Zajic	Geoffrey Zweig
Alexander Zatvornitskiy	Marzena Zyg
Vicky Zayats	
Hossein Zeinali	
Piotr Żelasko	
Milos Zelezny	
Margaret Zellers	
Heiga Zen	
Biao Zeng	
Abraham Woubie Zewoudie	
Andrej Zgank	
Chong Zhang	
Jinsong Zhang	
Mingyang Zhang	
Hui Zhang	
Chi (Leo) Zhang	
Pengyuan Zhang	
Xiaohui Zhang	
ShiLiang Zhang	
Chao Zhang	
Yu Zhang	
Xueliang Zhang	
Chunlei Zhang	
Zhengchen Zhang	
Yang Zhang	
Shi-Xiong Zhang	
Tao Zhang	
Xiao-Lei Zhang	
Frank Zhang	
Rui Zhao	
Yunxin Zhao	
Yuaniun Zhao	

INTERSPEECH 2022 Sep 18 – 22 • Incheon Korea www.interspeech2022.org

Human and Humanizing Speech Technology











Incheon Tourism Organization



ABSTRACTS

Session Index

Tuesday 31 A	ugust 2021	
Tue-M-O-1	Speech Synthesis: Other Topics	60
Tue-M-O-2	Disordered Speech	60
Tue-M-O-3	Speech Signal Analysis and Representation II	62
Tue-M-V-1	Feature, Embedding and Neural Architecture for Speaker Recognition	62
Tue-M-V-2	Speech Synthesis: Toward End-to-End Synthesis II	64
Tue-M-V-3	Speech Enhancement and Intelligibility	66
Tue-M-V-4	Spoken Dialogue Systems I	69
Tue-M-V-5	Topics in ASR: Robustness, Feature Extraction, and Far-Field ASR	70
Tue-M-V-6	Voice Activity Detection and Keyword Spotting	73
Tue-M-V-7	Voice and Voicing	75
Tue-M-SS-1	The INTERSPEECH 2021 Computational Paralinguistics Challenge (ComParE) — COVID-19 Cough, COVID-19 Speech, Escalation & Primates	77
Tue-Survey	Survey Talk 1: Heidi Christensen	79
Tue-A-O-1	Embedding and Network Architecture for Speaker Recognition	80
Tue-A-O-2	Speech Perception I	81
Tue-A-V-1	Acoustic Event Detection and Acoustic Scene Classification	82
Tue-A-V-2	Diverse Modes of Speech Acquisition and Processing	84
Tue-A-V-3	Multi-Channel Speech Enhancement and Hearing Aids	86
Tue-A-V-4	Self-Supervision and Semi-Supervision for Neural ASR Training	88
Tue-A-V-5	Spoken Language Processing I	90
Tue-A-V-6	Voice Conversion and Adaptation II	92
Tue-A-SS-1	Privacy-Preserving Machine Learning for Audio & Speech Processing	94
Tue-A-SS-2	The First DiCOVA Challenge: Diagnosis of COVID-19 Using Acoustics	96
Tue-A-S&T-1	Show and Tell 1	98
Tue-Keynote	Keynote 1: Hermann Ney	100
Tue-E-O-1	ASR Technologies and Systems	100
Tue-E-O-2	Phonation and Voicing	101
Tue-E-O-3	Health and Affect I	102
Tue-E-V-1	Robust Speaker Recognition	103
Tue-E-V-2	Source Separation, Dereverberation and Echo Cancellation	105
Tue-E-V-3	Speech Signal Analysis and Representation I	107
Tue-E-V-4	Spoken Language Understanding I	110
Tue-E-V-5	Topics in ASR: Adaptation, Transfer Learning, Children's Speech, and Low-Resource Settings	5 112
Tue-E-V-6	Voice Conversion and Adaptation I	114
Tue-E-SS-1	Voice Quality Characterization for Clinical Voice Assessment: Voice Production, Acoustics, and Auditory Perception	117

Wednesday 1 September 2021

Wed-M-O-1	Miscellaneous Topics in ASR 119
Wed-M-O-2	Phonetics I
Wed-M-O-3	Target Speaker Detection, Localization and Separation121
Wed-M-V-1	Language and Accent Recognition 122
Wed-M-V-2	Low-Resource Speech Recognition 123
Wed-M-V-3	Speech Synthesis: Singing, Multimodal, Crosslingual Synthesis 126
Wed-M-V-4	Speech Coding and Privacy 128
Wed-M-V-5	Speech Perception II
Wed-M-V-6	Streaming for ASR/RNN Transducers
Wed-M-SS-1	ConferencingSpeech 2021 Challenge: Far-Field Multi-Channel Speech Enhancement for Video Conferencing
Wed-Survey	Survey Talk 2: Sriram Ganapathy 137
Wed-Keynote	Keynote 2: Pascale Fung
Wed-A-O-1	Language Modeling and Text-Based Innovations for ASR
Wed-A-O-2	Speaker, Language, and Privacy
Wed-A-O-3	Assessment of Pathological Speech and Language I
Wed-A-V-1	Communication and Interaction, Multimodality
Wed-A-V-2	Language and Lexical Modeling for ASR
Wed-A-V-3	Novel Neural Network Architectures for ASR
Wed-A-V-4	Speech Localization, Enhancement, and Quality Assessment
Wed-A-V-5	Speech Synthesis: Neural Waveform Generation
Wed-A-V-6	Spoken Machine Translation 152
Wed-A-SS-1	SdSV Challenge 2021: Analysis and Exploration of New Ideas on Short-Duration Speaker Verification
Wed-A-S&T-1	Show and Tell 2
Wed-E-O-1	Graph and End-to-End Learning for Speaker Recognition 157
Wed-E-O-2	Spoken Language Processing II
Wed-E-O-3	Speech and Audio Analysis
Wed-E-V-1	Cross/Multi-Lingual and Code-Switched ASR 160
Wed-E-V-2	Health and Affect II
Wed-E-V-3	Neural Network Training Methods for ASR
Wed-E-V-4	Prosodic Features and Structure
Wed-E-V-5	Single-Channel Speech Enhancement 170
Wed-E-V-6	Speech Synthesis: Tools, Data, Evaluation
Wed-E-SS-1	INTERSPEECH 2021 Deep Noise Suppression Challenge 175

Thursday 2 September 2021

Thu-M-O-1	Neural Network Training Methods and Architectures for ASR	177
Thu-M-O-2	Emotion and Sentiment Analysis I	178
Thu-M-O-3	Linguistic Components in End-to-End ASR	178
Thu-M-V-1	Assessment of Pathological Speech and Language II	180
Thu-M-V-2	Multimodal Systems	182
Thu-M-V-3	Source Separation I	184
Thu-M-V-4	Speaker Diarization I	187
Thu-M-V-5	Speech Synthesis: Prosody Modeling I	188
Thu-M-V-6	Speech Production II	190
Thu-M-V-7	Spoken Dialogue Systems II	192
Thu-M-SS-1	Oriental Language Recognition	193
Thu-M-SS-2	Automatic Speech Recognition in Air Traffic Management	195
Thu-M-S&T-1	Show and Tell 3	196
Thu-Survey	Survey Talk 3: Karen Livescu	197
Thu-Keynote	Keynote 3: Mounya Elhilali	198
Thu-A-O-1	Speech Production I	198
Thu-A-O-2	Speech Enhancement and Coding	199
Thu-A-V-1	Emotion and Sentiment Analysis II	200
Thu-A-V-2	Multi- and Cross-Lingual ASR, Other Topics in ASR	202
Thu-A-V-3	Source Separation II	204
Thu-A-V-4	Speaker Diarization II	207
Thu-A-V-5	Speech Synthesis: Toward End-to-End Synthesis I	209
Thu-A-V-6	Tools, Corpora and Resources	211
Thu-A-SS-1	Non-Autoregressive Sequential Modeling for Speech Processing	214
Thu-A-SS-2	The ADReSSo Challenge: Detecting Cognitive Decline Using Speech Only	216

Friday 3 September 2021

Fri-M-O-1	Robust and Far-Field ASR	219
Fri-M-O-2	Speech Synthesis: Prosody Modeling II	. 220
Fri-M-O-3	Source Separation III	. 221
Fri-M-V-1	Non-Native Speech	. 222
Fri-M-V-2	Phonetics II	224
Fri-M-V-3	Search/Decoding Techniques and Confidence Measures for ASR	. 227
Fri-M-V-4	Speech Synthesis: Linguistic Processing, Paradigms and Other Topics	. 229
Fri-M-V-5	Speech Type Classification and Diagnosis	. 231
Fri-M-V-6	Spoken Term Detection & Voice Search	234
Fri-M-V-7	Voice Anti-Spoofing and Countermeasure	. 236
Fri-M-SS-1	OpenASR20 and Low Resource ASR Development	. 239
Fri-Survey	Survey Talk 4: Alejandrina Cristia	. 240
Fri-Keynote	Keynote 4: Tomáš Mikolov	. 241
Fri-Keynote Fri-A-O-1	Keynote 4: Tomáš Mikolov Voice Activity Detection	. 241 . 241
Fri-Keynote Fri-A-O-1 Fri-A-O-2	Keynote 4: Tomáš MikolovVoice Activity DetectionKeyword Search and Spoken Language Processing	. 241 . 241 . 242
Fri-Keynote Fri-A-O-1 Fri-A-O-2 Fri-A-V-1	Keynote 4: Tomáš Mikolov Voice Activity Detection Keyword Search and Spoken Language Processing Applications in Transcription, Education and Learning	. 241 . 241 . 242 . 243
Fri-Keynote Fri-A-O-1 Fri-A-O-2 Fri-A-V-1 Fri-A-V-2	Keynote 4: Tomáš MikolovVoice Activity DetectionKeyword Search and Spoken Language ProcessingApplications in Transcription, Education and LearningEmotion and Sentiment Analysis III	. 241 . 241 242 243 243
Fri-Keynote Fri-A-O-1 Fri-A-O-2 Fri-A-V-1 Fri-A-V-2 Fri-A-V-3	Keynote 4: Tomáš Mikolov Voice Activity Detection Keyword Search and Spoken Language Processing Applications in Transcription, Education and Learning Emotion and Sentiment Analysis III Resource-Constrained ASR	. 241 . 241 . 242 . 243 . 243 . 245 . 247
Fri-Keynote Fri-A-O-1 Fri-A-O-2 Fri-A-V-1 Fri-A-V-2 Fri-A-V-3 Fri-A-V-4	Keynote 4: Tomáš Mikolov Voice Activity Detection Keyword Search and Spoken Language Processing Applications in Transcription, Education and Learning Emotion and Sentiment Analysis III Resource-Constrained ASR Speaker Recognition: Applications	. 241 . 241 242 243 243 . 245 . 247 249
Fri-Keynote Fri-A-O-1 Fri-A-O-2 Fri-A-V-1 Fri-A-V-2 Fri-A-V-3 Fri-A-V-4 Fri-A-V-5	Keynote 4: Tomáš Mikolov Voice Activity Detection Keyword Search and Spoken Language Processing Applications in Transcription, Education and Learning Emotion and Sentiment Analysis III Resource-Constrained ASR Speaker Recognition: Applications Speech Synthesis: Speaking Style and Emotion	. 241 . 241 . 242 . 243 . 243 . 245 . 247 . 249 . 251
Fri-Keynote Fri-A-O-1 Fri-A-O-2 Fri-A-V-1 Fri-A-V-2 Fri-A-V-3 Fri-A-V-4 Fri-A-V-5 Fri-A-V-6	Keynote 4: Tomáš Mikolov Voice Activity Detection Keyword Search and Spoken Language Processing Applications in Transcription, Education and Learning Emotion and Sentiment Analysis III Resource-Constrained ASR Speaker Recognition: Applications Speech Synthesis: Speaking Style and Emotion Spoken Language Understanding II	. 241 . 241 242 243 245 . 247 249 251 . 254
Fri-Keynote Fri-A-O-1 Fri-A-O-2 Fri-A-V-1 Fri-A-V-2 Fri-A-V-3 Fri-A-V-4 Fri-A-V-5 Fri-A-V-6 Fri-A-SS-1	Keynote 4: Tomáš Mikolov Voice Activity Detection Keyword Search and Spoken Language Processing Applications in Transcription, Education and Learning Emotion and Sentiment Analysis III Resource-Constrained ASR Speaker Recognition: Applications Speech Synthesis: Speaking Style and Emotion Spoken Language Understanding II INTERSPEECH 2021 Acoustic Echo Cancellation Challenge	. 241 . 241 . 242 . 243 . 245 . 245 . 247 . 249 . 251 . 254 . 255
Fri-Keynote Fri-A-O-1 Fri-A-O-2 Fri-A-V-1 Fri-A-V-2 Fri-A-V-3 Fri-A-V-4 Fri-A-V-5 Fri-A-V-5 Fri-A-V-6 Fri-A-SS-1 Fri-A-SS-2	Keynote 4: Tomáš Mikolov Voice Activity Detection Keyword Search and Spoken Language Processing Applications in Transcription, Education and Learning Emotion and Sentiment Analysis III Resource-Constrained ASR Speaker Recognition: Applications Speaker Recognition: Applications Speech Synthesis: Speaking Style and Emotion Spoken Language Understanding II INTERSPEECH 2021 Acoustic Echo Cancellation Challenge Speech Recognition of Atypical Speech	. 241 . 241 . 242 . 243 . 245 . 247 . 247 . 247 . 251 . 251 . 255 . 257

Abstracts

Tue-M-O-1: Speech Synthesis: Other Topics

Room A+B, 09:30-11:30, Tuesday 31 August 2021 Chairs: Jindřich Matoušek and Michael Pucher

Conversion of Airborne to Bone-Conducted Speech with Deep Neural Networks

*Michael Pucher*¹, *Thomas Woltron*²; ¹*Austrian Academy of Sciences, Austria*; ²*FH Wiener Neustadt, Austria* Tue-M-O-1-1, Time: 09:30

It is a common experience of most speakers that the playback of one's own voice sounds strange. This can be mainly attributed to the missing bone-conducted speech signal that is not present in the playback signal. It was also shown that some phonemes have a high bone-conducted relative to air-conducted sound transmission, which means that the bone-conduction filter is phone-dependent. To achieve such a phone-dependent modeling we train different speaker dependent and speaker adaptive speech conversion systems using airborne and bone-conducted speech data from 8 speakers (5 male, 3 female), which allow for the conversion of airborne speech to bone-conducted speech. The systems are based on Long Short-Term Memory (LSTM) deep neural networks, where the speaker adaptive versions with speaker embedding can be used without bone-conduction signals from the target speaker. Additionally we also used models that apply a global filtering. The different models are then evaluated by an objective error metric and a subjective listening experiment, which show that the LSTM based models outperform the global filters.

T5G2P: Using Text-to-Text Transfer Transformer for Grapheme-to-Phoneme Conversion

Markéta Řezáčková, Jan Švec, Daniel Tihelka; University of West Bohemia, Czechia

Tue-M-O-1-2, Time: 09:50

Despite the increasing popularity of end-to-end text-to-speech (TTS) systems, the correct grapheme-to-phoneme (G2P) module is still a crucial part of those relying on a phonetic input. In this paper, we, therefore, introduce a T5G2P model, a Text-to-Text Transfer Transformer (T5) neural network model which is able to convert an input text sentence into a phoneme sequence with a high accuracy. The evaluation of our trained T5 model is carried out on English and Czech, since there are different specific properties of G2P, including homograph disambiguation, cross-word assimilation and irregular pronunciation of loanwords. The paper also contains an analysis of a homographs issue in English and offers another approach to Czech phonetic transcription using the detection of pronunciation exceptions.

Evaluating the Extrapolation Capabilities of Neural Vocoders to Extreme Pitch Values

Olivier Perrotin, Hussein El Amouri, Gérard Bailly, Thomas Hueber; GIPSA-lab (UMR 5216), France Tue-M-O-1-3, Time: 10:10

Neural vocoders are systematically evaluated on homogeneous train and test databases. This kind of evaluation is efficient to compare neural vocoders in their "comfort zone", yet it hardly reveals their limits towards unseen data during training. To compare their extrapolation capabilities, we introduce a methodology that aims at quantifying the robustness of neural vocoders in synthesising unseen data, by precisely controlling the ranges of seen/unseen data

NOTES

in the training database. By focusing in this study on the pitch (F_0) parameter, our methodology involves a careful splitting of a dataset to control which F_0 values are seen/unseen during training, followed by both global (utterance) and local (frame) evaluation of vocoders. Comparison of four types of vocoders (autoregressive, sourcefilter, flows, GAN) displays a wide range of behaviour towards unseen input pitch values, including excellent extrapolation (WaveGlow); widely-spread F_0 errors (WaveRNN); and systematic generation of the training set median F_0 (LPCNet, Parallel WaveGAN). In contrast, fewer differences between vocoders were observed when using homogeneous train and test sets, thus demonstrating the potential and need for such evaluation to better discriminate the neural vocoders abilities to generate out-of-training-range data.

A Systematic Review and Analysis of Multilingual Data Strategies in Text-to-Speech for Low-Resource Languages

Phat Do¹, Matt Coler¹, Jelske Dijkstra¹, Esther Klabbers²; ¹Rijksuniversiteit Groningen, The Netherlands; ²ReadSpeaker, The Netherlands Tue-M-O-1-4, Time: 10:30

We provide a systematic review of past studies that use multilingual data for text-to-speech (TTS) of low-resource languages (LRLs). We focus on the strategies used by these studies for incorporating multilingual data and how they affect output speech quality. To investigate the difference in output quality between corresponding monolingual and multilingual models, we propose a novel measure to compare this difference across the included studies and their various evaluation metrics. This measure, called the Multilingual Model Effect (MLME), is found to be affected by: acoustic model architecture, the difference ratio of target language data between corresponding multilingual and monolingual experiments, the balance ratio of target language data to total data, and the amount of target language data used. These findings can act as reference for data strategies in future experiments with multilingual TTS models for LRLs. Language family classification, despite being widely used, is not found to be an effective criterion for selecting source languages.

Tue-M-O-2: Disordered Speech

Room C, 09:30–11:30, Tuesday 31 August 2021 Chairs: Rob van Son and Gábor Gosztolya

Acoustic Indicators of Speech Motor Coordination in Adults With and Without Traumatic Brain Injury

Tanya Talkar¹, Nancy Pearl Solomon², Douglas S. Brungart², Stefanie E. Kuchinsky², Megan M. Eitel², Sara M. Lippa², Tracey A. Brickell², Louis M. French², Rael T. Lange², Thomas F. Quatieri¹; ¹Harvard University, USA; ²Walter Reed National Military Medical Center, USA

Tue-M-O-2-1, Time: 09:30

A traumatic brain injury (TBI) can lead to various long-term effects on memory, attention, and mood, as well as the occurrence of headaches, speech, and hearing problems. There is a need to better understand the long-term effects of a TBI for objective tracking of an individual's recovery, which could be used to determine intervention trajectories. This study utilizes acoustic features derived from recordings of speech tasks completed by active-duty service members and veterans (SMVs) enrolled in the Defense and Veterans Brain Injury (DVBIC)/Traumatic Brain Injury Center of Excellence (TBICOE) 15-Year Longitudinal TBI Study. We hypothesize that the individuals diagnosed with moderate to severe TBI would demonstrate motor speech impairments through decreased coordination of the speech production subsystems as compared to individuals with no history of TBI. Speech motor coordination is measured through correlations of acoustic feature time series representing speech subsystems. Eigenspectra derived from these correlations are utilized in machine learning models to discriminate between the two groups. The fusion of correlation features derived from the recordings achieves an AUC of 0.78. This suggests that residual motor impairments from moderate to severe TBI could be detectable through objective measures of speech motor coordination.

On Modeling Glottal Source Information for Phonation Assessment in Parkinson's Disease

J.C. Vásquez-Correa¹, Julian Fritsch², J.R. Orozco-Arroyave¹, Elmar Nöth¹, Mathew Magimai-Doss²; ¹FAU Erlangen-Nürnberg, Germany; ²Idiap Research Institute, Switzerland Tue-M-O-2-2, Time: 09:50

Parkinson's disease produces several motor symptoms, including different speech impairments that are known as hypokinetic dysarthria. Symptoms associated to dysarthria affect different dimensions of speech such as phonation, articulation, prosody, and intelligibility. Studies in the literature have mainly focused on the analysis of articulation and prosody because they seem to be the most prominent symptoms associated to dysarthria severity. However, phonation impairments also play a significant role to evaluate the global speech severity of Parkinson's patients. This paper proposes an extensive comparison of different methods to automatically evaluate the severity of specific phonation impairments in Parkinson's patients. The considered models include the computation of perturbation and glottal-based features, in addition to features extracted from a zero frequency filtered signals. We consider as well end-to-end models based on 1D CNNs, which are trained to learn features from the raw speech waveform, reconstructed glottal signals, and zero-frequency filtered signals. The results indicate that it is possible to automatically classify between speakers with low versus high phonation severity due to the presence of dysarthria and at the same time to evaluate the severity of the phonation impairments on a continuous scale, posed as a regression problem.

Distortion of Voiced Obstruents for Differential Diagnosis Between Parkinson's Disease and Multiple System Atrophy

Khalid Daoudi¹, Biswajit Das¹, Solange Milhé de Saint Victor², Alexandra Foubert-Samier², Anne Pavy-Le Traon³, Olivier Rascol³, Wassilios G. Meissner², Virginie Woisard³; ¹Inria, France; ²CHU de Bordeaux, France; ³CHU de Toulouse, France Tue-M-O-2-3, Time: 10:10

Parkinson's disease (PD) and the parkinsonian variant of Multiple System Atrophy (MSA-P) are two neurodegenerative diseases which share similar clinical features, particularly in early disease stages. The differential diagnosis can be thus very challenging. Dysarthria is known to be a frequent and early clinical feature of PD and MSA. It can be thus used as a vehicle to provide a vocal biomarker which could help in the differential diagnosis. In particular, distortion of consonants is known to be a frequent impairment in these diseases. The aim of this study is to investigate distinctive patterns in the distortion of voiced obstruents (plosives and fricatives). It is the first study which attempts to examine such distortions in the French language for the purpose of the differential diagnosis between PD

NOTES

and MSA-P (and among the very few studies if we consider all languages). We carry out a perceptual and objective analysis of voiced obstruents extracted from isolated pseudo-words initials. We first show that devoicing is a significant impairment which predominates in MSA-P. We then show that voice onset time (VOT) of voiced plosives (prevoicing duration) can be a complementary feature to improve the accuracy in discrimination between PD and MSA-P.

A Study into Pre-Training Strategies for Spoken Language Understanding on Dysarthric Speech

*Pu Wang*¹, *Bagher BabaAli*², *Hugo Van hamme*¹; ¹*KU Leuven, Belgium;* ²*University of Tehran, Iran* Tue-M-O-2-4, Time: 10:30

End-to-end (E2E) spoken language understanding (SLU) systems avoid an intermediate textual representation by mapping speech directly into intents with slot values. This approach requires considerable domain-specific training data. In low-resource scenarios this is a major concern, e.g., in the present study dealing with SLU for dysarthric speech. Pretraining part of the SLU model for automatic speech recognition targets helps but no research has shown to which extent SLU on dysarthric speech benefits from knowledge transferred from other dysarthric speech tasks. This paper investigates the efficiency of pre-training strategies for SLU tasks on dysarthric speech. The designed SLU system consists of a TDNN acoustic model for feature encoding and a capsule network for intent and slot decoding. The acoustic model is pre-trained in two stages: initialization with a corpus of normal speech and finetuning on a mixture of dysarthric and normal speech. By introducing the intelligibility score as a metric of the impairment severity, this paper quantitatively analyzes the relation between generalization and pathology severity for dysarthric speech.

EasyCall Corpus: A Dysarthric Speech Dataset

Rosanna Turrisi¹, Arianna Braccia², Marco Emanuele¹, Simone Giulietti², Maura Pugliatti², Mariachiara Sensi², Luciano Fadiga¹, Leonardo Badino³; ¹IIT, Italy; ²Università di Ferrara, Italy; ³PerVoice, Italy Tue-M-O-2-5, Time: 10:50

This paper introduces a new dysarthric speech command dataset in Italian, called EasyCall corpus. The dataset consists of 21386 audio recordings from 24 healthy and 31 dysarthric speakers, whose individual degree of speech impairment was assessed by neurologists through the Therapy Outcome Measure. The corpus aims at providing a resource for the development of ASR-based assistive technologies for patients with dysarthria. In particular, it may be exploited to develop a voice-controlled contact application for commercial smartphones, aiming at improving dysarthric patients' ability to communicate with their family and caregivers. Before recording the dataset, participants were administered a survey to evaluate which commands are more likely to be employed by dysarthric individuals in a voice-controlled contact application. In addition, the dataset includes a list of non-commands (i.e., words near/inside commands or phonetically close to commands) that can be leveraged to build a more robust command recognition system. At present commercial ASR systems perform poorly on the EasyCall Corpus as we report in this paper. This result corroborates the need for dysarthric speech corpora for developing effective assistive technologies. To the best of our knowledge, this database represents the richest corpus of dysarthric speech to date.

Tue-M-O-3: Speech Signal Analysis and Representation II

Room D, 09:30–11:30, Tuesday 31 August 2021 Chairs: Torbjørn Svendsen and Benjamin Elie

A Benchmark of Dynamical Variational Autoencoders Applied to Speech Spectrogram Modeling

*Xiaoyu Bie*¹, *Laurent Girin*², *Simon Leglaive*³, *Thomas Hueber*², *Xavier Alameda-Pineda*¹; ¹LJK (UMR 5224), *France*; ²*GIPSA-lab* (UMR 5216), *France*; ³IETR (UMR 6164), *France*

Tue-M-O-3-1, Time: 09:30

The Variational Autoencoder (VAE) is a powerful deep generative model that is now extensively used to represent high-dimensional complex data via a low-dimensional latent space learned in an unsupervised manner. In the original VAE model, input data vectors are processed independently. In recent years, a series of papers have presented different extensions of the VAE to process sequential data, that not only model the latent space, but also model the temporal dependencies within a sequence of data vectors and corresponding latent vectors, relying on recurrent neural networks. We recently performed a comprehensive review of those models and unified them into a general class called Dynamical Variational Autoencoders (DVAEs). In the present paper, we present the results of an experimental benchmark comparing six of those DVAE models on the speech analysis-resynthesis task, as an illustration of the high potential of DVAEs for speech modeling.

Fricative Phoneme Detection Using Deep Neural Networks and its Comparison to Traditional Methods

Metehan Yurt¹, Pavan Kantharaju¹, Sascha Disch¹, Andreas Niedermeier¹, Alberto N. Escalante-B.², Veniamin I. Morgenshtern³; ¹Fraunhofer IIS, Germany; ²WS Audiology, Germany; ³FAU Erlangen-Nürnberg, Germany

Tue-M-O-3-2, Time: 09:50

Accurate phoneme detection and processing can enhance speech intelligibility in hearing aids and audio & speech codecs. As fricative phonemes have an important part of their energy concentrated in high frequency bands, frequency lowering algorithms are used in hearing aids to improve fricative intelligibility for people with high-frequency hearing loss. In traditional audio codecs, while processing speech in blocks, spectral smearing around fricative phoneme borders results in pre and post echo artifacts. Hence, detecting the fricative borders and adapting the processing accordingly could enhance the quality of speech. Until recently, phoneme detection and analysis were mostly done by extracting features specific to the class of phonemes. In this paper, we present a deep learning based fricative phoneme detection algorithm that exceeds the state-of-the-art fricative phoneme detection accuracy on the TIMIT speech corpus. Moreover, we compare our method to other approaches that employ classical signal processing for fricative detection and also evaluate it on the TIMIT files coded with AAC codec followed by bandwidth limitation. Reported results of our deep learning approach on original TIMIT files are reproducible and come with an easy to use code that could serve as a baseline for any future research on this topic.

Identification of F1 and F2 in Speech Using Modified Zero Frequency Filtering

RaviShankar Prasad, Mathew Magimai-Doss; Idiap Research Institute, Switzerland Tue-M-O-3-3, Time: 10:10

Formants are major resonances in the vocal tract system. Identification of formants is important for study of speech. In the literature, formants are typically identified by first deriving formant frequency candidates (e.g., using linear prediction) and then applying a tracking mechanism. In this paper, we propose a simple tracking-free formant identification approach based on zero frequency filtering. More precisely, formants F1-F2 are identified by modifying the trend removal operation in zero frequency filtering and picking simply the dominant peak in the short-term discrete Fourier transform spectra. We demonstrate the potential of the approach by comparing it against state-of-the-art formant identification approaches on a typical speech data set (TIMIT-VTR) and an atypical speech data set (PC-GITA).

Phoneme-to-Audio Alignment with Recurrent Neural Networks for Speaking and Singing Voice

Yann Teytaut, Axel Roebel; STMS (UMR 9912), France Tue-M-O-3-4, Time: 10:30

Phoneme-to-audio alignment is the task of synchronizing voice recordings and their related phonetic transcripts. In this work, we introduce a new system to forced phonetic alignment with Recurrent Neural Networks (RNN). With the Connectionist Temporal Classification (CTC) loss as training objective, and an additional reconstruction cost, we learn to infer relevant per-frame phoneme probabilities from which alignment is derived. The core of the neural architecture is a context-aware attention mechanism between mel-spectrograms and side information. We investigate two contexts given by either phoneme sequences (model PHATT) or spectrograms themselves (model SPATT). Evaluations show that these models produce precise alignments for both speaking and singing voice. Best results are obtained with the model PHATT, which outperforms baseline reference with an average imprecision of 16.3ms and 29.8ms on speech and singing, respectively. The model SPATT also appears as an interesting alternative, capable of aligning longer audio files without requiring phoneme sequences on small audio segments.

Tue-M-V-1: Feature, Embedding and Neural Architecture for Speaker Recognition

09:30-11:30, Tuesday 31 August 2021 Chairs: Hervé Bredin and Niko Brummer

Adaptive Convolutional Neural Network for Text-Independent Speaker Recognition

Seong-Hu Kim, Yong-Hwa Park; KAIST, Korea Tue-M-V-1-1, Time: 09:30

In text-independent speaker recognition, each speech is composed of different phonemes depending on spoken text. The conventional neural networks for speaker recognition are static models, so they do not reflect this phoneme-varying characteristic well. To tackle this limitation, we propose an adaptive convolutional neural network (ACNN) for text-independent speaker recognition. The utterance is divided along the time axis into short segments with small fluctuating phonemes. Frame-level features are extracted by applying input-dependent kernels adaptive to each segment. By applying time average pooling and linear layers, utterance-level embeddings extraction and speaker recognition are performed. Adaptive VGG-M

using 0.356 seconds segmentation shows better speaker recognition performance than baseline models, with a Top-1 of 86.51% and an EER of 5.68%. It extracts more accurate frame-level embeddings for vowel and nasal phonemes compared to the conventional method without overfitting and large parameters. This framework for textindependent speaker recognition effectively utilizes phonemes and text-varying characteristic of speech.

Bidirectional Multiscale Feature Aggregation for Speaker Verification

Jiajun Qi, Wu Guo, Bin Gu; USTC, China Tue-M-V-1-2, Time: 09:30

In this paper, we propose a novel bidirectional multiscale feature aggregation (BMFA) network with attentional fusion modules for textindependent speaker verification. The feature maps from different stages of the backbone network are iteratively combined and refined in both a bottom-up and top-down manner. Furthermore, instead of simple concatenation or elementwise addition of feature maps from different stages, an attentional fusion module is designed to compute the fusion weights. Experiments are conducted on the NIST SRE16 and VoxCeleb1 datasets. The experimental results demonstrate the effectiveness of the bidirectional aggregation strategy and show that the proposed attentional fusion module can further improve the performance.

Improving Time Delay Neural Network Based Speaker **Recognition with Convolutional Block and Feature Aggregation Methods**

Yu-Jia Zhang¹, Yih-Wen Wang¹, Chia-Ping Chen¹, Chung-Li Lu², Bo-Cheng Chan²; ¹National Sun Yat-sen University, Taiwan; ²Chunghwa Telecom Laboratories, Taiwan

Tue-M-V-1-3, Time: 09:30

In this paper, we develop a system that integrates multiple ideas and techniques inspired by the convolutional block and feature aggregation methods. We begin with the state-of-the-art speakerembedding model for speaker recognition, namely the model of Emphasized Channel Attention, Propagation, and Aggregation in Time Delay Neural Network, and then gradually experiment with the proposed network modules, including bottleneck residual blocks, attention mechanisms, and feature aggregation methods. In our final model, we replace the Res2Block with SC-Block and we use a hierarchical architecture for feature aggregation. We evaluate the performance of our model on the VoxCeleb1 test set and the 2020 VoxCeleb Speaker Recognition Challenge (VoxSRC20) validation set. The relative improvement of the proposed models over ECAPA-TDNN is 22.8% on VoxCeleb1 and 18.2% on VoxSRC20.

Improving Deep CNN Architectures with Variable-Length Training Samples for **Text-Independent Speaker Verification**

Yanfeng Wu, Junan Zhao, Chenkai Guo, Jing Xu; Nankai University, China

Tue-M-V-1-4, Time: 09:30

Deep Convolutional Neural Network (CNN) based speaker embeddings, such as r-vectors, have shown great success in textindependent speaker verification (TI-SV) task. However, previous deep CNN models usually use fixed-length samples for training and employ variable-length utterances for speaker embeddings, which generates a mismatch between training and embedding. To address this issue, we investigate the effect of employing variable-length training samples on CNN-based TI-SV systems and explore two

NOTES

approaches to improve the performance of deep CNN architectures on TI-SV through capturing variable-term contexts. Firstly, we present an improved selective kernel convolution which allows the networks to adaptively switch between short-term and long-term contexts based on variable-length utterances. Secondly, we propose a multi-scale statistics pooling method to aggregate multiple timescale features from different layers of the networks. We build a novel ResNet34 based architecture with two proposed approaches. Experiments are conducted on the VoxCeleb datasets. The results demonstrate that the effect of using variable-length samples is diverse in different networks and the architecture with two proposed approaches achieves significant improvement over r-vectors baseline system.

Binary Neural Network for Speaker Verification

Tinglong Zhu, Xiaoyi Qin, Ming Li; Duke Kunshan University, China

Tue-M-V-1-5, Time: 09:30

Although deep neural networks are successful for many tasks in the speech domain, the high computational and memory costs of deep neural networks make it difficult to directly deploy high-performance Neural Network systems on low-resource embedded devices. There are several mechanisms to reduce the size of the neural networks i.e. parameter pruning, parameter quantization, etc. This paper focuses on how to apply binary neural networks to the task of speaker verification. The proposed binarization of training parameters can largely maintain the performance while significantly reducing storage space requirements and computational costs. Experiment results show that, after binarizing the Convolutional Neural Network, the ResNet34-based network achieves an EER of around 5% on the Voxceleb1 testing dataset and even outperforms the traditional real number network on the text-dependent dataset: Xiaole while having a $32 \times$ memory saving.

Mutual Information Enhanced Training for Speaker Embedding

Youzhi Tu, Man-Wai Mak; PolyU, China Tue-M-V-1-6, Time: 09:30

Mutual information (MI) is useful in unsupervised and self-supervised learning. Maximizing the MI between the low-level features and the learned embeddings can preserve meaningful information in the embeddings, which can contribute to performance gains. This strategy is called deep InfoMax (DIM) in representation learning. In this paper, we follow the DIM framework so that the speaker embeddings can capture more information from the frame-level features. However, a straightforward implementation of DIM may pose a dimensionality imbalance problem because the dimensionality of the frame-level features is much larger than that of the speaker embeddings. This problem can lead to unreliable MI estimation and can even cause detrimental effects on speaker verification. To overcome this problem, we propose to squeeze the frame-level features before MI estimation through some global pooling methods. We call the proposed method squeeze-DIM. Although the squeeze operation inevitably introduces some information loss, we empirically show that the squeeze-DIM can achieve performance gains on both Voxceleb1 and VOiCES-19 tasks. This suggests that the squeeze operation facilitates the MI estimation and maximization in a balanced dimensional space, which helps learn more informative speaker embeddings.

Y-Vector: Multiscale Waveform Encoder for Speaker Embedding

Ge Zhu, Fei Jiang, Zhiyao Duan; University of Rochester, USA

Tue-M-V-1-7, Time: 09:30

State-of-the-art text-independent speaker verification systems typically use cepstral features or filter bank energies as speech features. Recent studies attempted to extract speaker embeddings directly from raw waveforms and have shown competitive results. In this paper, we propose a novel multi-scale waveform encoder that uses three convolution branches with different time scales to compute speech features from the waveform. These features are then processed by squeeze-and-excitation blocks, a multi-level feature aggregator, and a time delayed neural network (TDNN) to compute speaker embedding. We show that the proposed embeddings outperform existing raw-waveform-based speaker embeddings on speaker verification by a large margin. A further analysis of the learned filters shows that the multi-scale encoder attends to different frequency bands at its different scales while resulting in a more flat overall frequency response than any of the single-scale counterparts.

Phoneme-Aware and Channel-Wise Attentive Learning for Text Dependent Speaker Verification

Yan Liu, Zheng Li, Lin Li, Qingyang Hong; Xiamen University, China

Tue-M-V-1-8, Time: 09:30

This paper proposes a multi-task learning network with phonemeaware and channel-wise attentive learning strategies for textdependent Speaker Verification (SV). In the proposed structure, the frame-level multi-task learning along with the segment-level adversarial learning is adopted for speaker embedding extraction. The phoneme-aware attentive pooling is exploited on frame-level features in the main network for speaker classifier, with the corresponding posterior probability for the phoneme distribution in the auxiliary subnet. Further, the introduction of Squeeze and Excitation (SE-block) performs dynamic channel-wise feature recalibration, which improves the representational ability. The proposed method exploits speaker idiosyncrasies associated with pass-phrases, and is further improved by the phoneme-aware attentive pooling and SE-block from temporal and channel-wise aspects, respectively. The experiments conducted on RSR2015 Part 1 database confirm that the proposed system achieves outstanding results for text-dependent SV.

Serialized Multi-Layer Multi-Head Attention for Neural Speaker Embedding

*Hongning Zhu*¹, *Kong Aik Lee*², *Haizhou Li*¹; ¹*NUS, Singapore*; ²*A***STAR, Singapore* Tue-M-V-1-9, Time: 09:30

This paper proposes a serialized multi-layer multi-head attention for neural speaker embedding in text-independent speaker verification. In prior works, frame-level features from one layer are aggregated to form an utterance-level representation. Inspired by the Transformer network, our proposed method utilizes the hierarchical architecture of stacked self-attention mechanisms to derive refined features that are more correlated with speakers. Serialized attention mechanism contains a stack of self-attention modules to create fixed-dimensional representations of speakers. Instead of utilizing multi-head attention in parallel, the proposed serialized multi-layer multi-head attention is designed to aggregate and propagate attentive statistics from one layer to the next in a serialized manner. In addition, we employ an input-aware query for each utterance with the statistics pooling. With more layers stacked, the neural network can learn

NOTES

more discriminative speaker embeddings. Experiment results on VoxCeleb1 dataset and SITW dataset show that our proposed method outperforms other baseline methods, including x-vectors and other x-vectors + conventional attentive pooling approaches by 9.7% in EER and 8.1% in DCF10⁻².

Tue-M-V-2: Speech Synthesis: Toward End-to-End Synthesis II

09:30–11:30, Tuesday 31 August 2021 Chairs: Heiga Zen and Cassia Valentini

TacoLPCNet: Fast and Stable TTS by Conditioning LPCNet on Mel Spectrogram Predictions

*Cheng Gong*¹, *Longbiao Wang*¹, *Ju Zhang*², *Shaotong Guo*¹, *Yuguang Wang*², *Jianwu Dang*¹; ¹*Tianjin University, China*; ²*Huiyan Technology, China* Tue-M-V-2-1, Time: 09:30

The combination of the recently proposed LPCNet vocoder and a seq-to-seq acoustic model, i.e., Tacotron, has successfully achieved lightweight speech synthesis systems. However, the quality of synthesized speech is often unstable because the precision of the pitch parameters predicted by acoustic models is insufficient, especially for some tonal languages like Chinese and Japanese. In this paper, we propose an end-to-end speech synthesis system, TacoLPCNet, by conditioning LPCNet on Mel spectrogram predictions. First, we extend LPCNet for the Mel spectrogram instead of using explicit pitch information and pitch-related network. Furthermore, we optimize the system by model pruning, multi-frame inference, and increasing frame length, to enable it to meet the conditions required for realtime applications. The objective and subjective evaluation results for various languages show that the proposed system is more stable for tonal languages within the proposed optimization strategies. The experimental results also verify that our model improves synthesis runtime by 3.12 times than that of the baseline on a standard CPU while maintaining naturalness.

FastPitchFormant: Source-Filter Based Decomposed Modeling for Speech Synthesis

Taejun Bak, Jae-Sung Bae, Hanbin Bae, Young-Ik Kim, Hoon-Young Cho; NCSOFT, Korea Tue-M-V-2-2, Time: 09:30

Methods for modeling and controlling prosody with acoustic features have been proposed for neural text-to-speech (TTS) models. Prosodic speech can be generated by conditioning acoustic features. However, synthesized speech with a large pitch-shift scale suffers from audio quality degradation, and speaker characteristics deformation. To address this problem, we propose a feed-forward Transformer based TTS model that is designed based on the source-filter theory. This model, called *FastPitchFormant*, has a unique structure that handles text and acoustic features in parallel. With modeling each feature separately, the tendency that the model learns the relationship between two features can be mitigated. Owing to its structural characteristics, FastPitchFormant is robust and accurate for pitch control and generates prosodic speech preserving speaker characteristics. The experimental results show that proposed model outperforms the baseline FastPitch.

Sequence-to-Sequence Learning for Deep Gaussian Process Based Speech Synthesis Using Self-Attention GP Layer

Taiki Nakamura, Tomoki Koriyama, Hiroshi Saruwatari; University of Tokyo, Japan

Tue-M-V-2-3, Time: 09:30

This paper presents a speech synthesis method based on deep Gaussian process (DGP) and sequence-to-sequence (Seq2Seq) learning toward high-quality end-to-end speech synthesis. Feed-forward and recurrent models using DGP are known to produce more natural synthetic speech than deep neural networks (DNNs) because of Bayesian learning and kernel regression. However, such DGP models consist of a pipeline architecture of independent models, acoustic and duration models, and require a high level of expertise in text processing. The proposed model is based on Seq2Seq learning, which enables a unified training of acoustic and duration models. The encoder and decoder layers are represented by Gaussian process regressions (GPRs) and the parameters are trained as a Bayesian model. We also propose a self-attention mechanism with Gaussian processes to effectively model character-level input in the encoder. The subjective evaluation results show that the proposed Seq2Seq-SA-DGP can synthesize more natural speech than DNNs with self-attention and recurrent structures. Besides, Seq2Seq-SA-DGP reduces the smoothing problems of recurrent structures and is effective when a simple input for an end-to-end system is given.

Phonetic and Prosodic Information Estimation from Texts for Genuine Japanese End-to-End Text-to-Speech

Naoto Kakegawa¹, Sunao Hara¹, Masanobu Abe¹, Yusuke Ijima²; ¹Okayama University, Japan; ²NTT, Japan

Tue-M-V-2-4, Time: 09:30

The biggest obstacle to develop end-to-end Japanese text-to-speech (TTS) systems is to estimate phonetic and prosodic information (PPI) from Japanese texts. The following are the reasons: (1) the Kanji characters of the Japanese writing system have multiple corresponding pronunciations, (2) there is no separation mark between words, and (3) an accent nucleus must be assigned at appropriate positions. In this paper, we propose to solve the problems by neural machine translation (NMT) on the basis of encoder-decoder models, and compare NMT models of recurrent neural networks and the Transformer architecture. The proposed model handles texts on token (character) basis, although conventional systems handle them on word basis. To ensure the potential of the proposed approach, NMT models are trained using pairs of sentences and their PPIs that are generated by a conventional Japanese TTS system from 5 million sentences. Evaluation experiments were performed using PPIs that are manually annotated for 5,142 sentences. The experimental results showed that the Transformer architecture has the best performance, with 98.0% accuracy for phonetic information estimation and 95.0% accuracy for PPI estimation. Judging from the results, NMT models are promising toward end-to-end Japanese TTS.

Information Sieve: Content Leakage Reduction in End-to-End Prosody Transfer for Expressive Speech Synthesis

Xudong Dai, Cheng Gong, Longbiao Wang, Kaili Zhang; Tianjin University, China Tue-M-V-2-5, Time: 09:30

Expressive neural text-to-speech (TTS) systems incorporate a style encoder to learn a latent embedding as the style information. However, this embedding process may encode redundant textual information. This phenomenon is called content leakage. Researchers have attempted to resolve this problem by adding an ASR or other auxiliary supervision loss functions. In this study, we propose an unsupervised method called the "information sieve" to reduce the effect of content leakage in prosody transfer. The rationale of this approach is that the style encoder can be forced to focus on style information rather than on textual information contained in the reference speech by a well-designed downsample-upsample filter, i.e., the extracted style embeddings can be downsampled at a certain interval and then upsampled by duplication. Furthermore, we used instance normalization in convolution layers to help the system learn a better latent style space. Objective metrics such as the significantly lower word error rate (WER) demonstrate the effectiveness of this model in mitigating content leakage. Listening tests indicate that the model retains its prosody transferability compared with the baseline models such as the original GST-Tacotron and ASR-guided Tacotron.

Deliberation-Based Multi-Pass Speech Synthesis

Qingyun Dou, Xixin Wu, Moquan Wan, Yiting Lu, Mark J.F. Gales; University of Cambridge, UK Tue-M-V-2-6, Time: 09:30

Sequence-to-sequence (seq2seq) models have achieved state-of-theart performance in a wide range of tasks including Neural Machine Translation (NMT) and Text-To-Speech (TTS). These models are usually trained with teacher forcing, where the reference back-history is used to predict the next token. This makes training efficient, but limits performance, because during inference the free-running back-history must be used. To address this problem, deliberationbased multi-pass seq2seq has been used in NMT. Here the output sequence is generated in multiple passes, each one conditioned on the initial input and the free-running output of the previous pass. This paper investigates, and compares, deliberation-based multi-pass seq2seq for TTS and NMT. For NMT the simplest form of multi-pass approaches, where the free-running first-pass output is combined with the initial input, improves performance. However, applying this scheme to TTS is challenging: the multi-pass model tends to converge to the standard single-pass model, ignoring the previous output. To tackle this issue, a guided attention loss is added, enabling the system to make more extensive use of the free-running output. Experimental results confirm the above analysis and demonstrate that the proposed TTS model outperforms a strong baseline.

Parallel Tacotron 2: A Non-Autoregressive Neural TTS Model with Differentiable Duration Modeling

Isaac Elias¹, Heiga Zen², Jonathan Shen³, Yu Zhang³, Ye Jia³, R.J. Skerry-Ryan³, Yonghui Wu³; ¹Google, Israel; ²Google, Japan; ³Google, USA Tue-M-V-2-7, Time: 09:30

This paper introduces *Parallel Tacotron 2*, a non-autoregressive neural text-to-speech model with a fully differentiable duration model which does not require supervised duration signals. The duration model is based on a novel attention mechanism and an iterative reconstruction loss based on Soft Dynamic TimeWarping, this model can learn token-frame alignments as well as token durations automatically. Experimental results show that Parallel Tacotron 2 outperforms baselines in subjective naturalness in several diverse multi speaker evaluations.

Transformer-Based Acoustic Modeling for Streaming Speech Synthesis

Chunyang Wu, Zhiping Xiu, Yangyang Shi, Ozlem Kalinli, Christian Fuegen, Thilo Koehler, Oing He; Facebook, USA

Tue-M-V-2-8, Time: 09:30

Transformer models have shown promising results in neural speech synthesis due to their superior ability to model long-term dependencies compared to recurrent networks. The computation complexity of transformers increases quadratically with sequence length, making it impractical for many real-time applications. To address the complexity issue in speech synthesis domain, this paper proposes an efficient transformer-based acoustic model that is constant-speed regardless of input sequence length, making it ideal for streaming speech synthesis applications. The proposed model uses a transformer network that predicts the prosody features at phone rate and then an Emformer network to predict the frame-rate spectral features in a streaming manner. Both the transformer and Emformer in the proposed architecture use a self-attention mechanism that involves explicit long-term information, thus providing improved speech naturalness for long utterances. In our experiments, we use a WaveRNN neural vocoder that takes in the predicted spectral features and generates the final audio. The overall architecture achieves human-like speech quality both on short and long utterances while maintaining a low latency and low real-time factor. Our mean opinion score (MOS) evaluation shows that for short utterances, the proposed model achieves a MOS of 4.213 compared to ground-truth with MOS of 4.307; and for long utterances, it also produces high-quality speech with a MOS of 4.201 compared to ground-truth with MOS of 4.360.

PnG BERT: Augmented BERT on Phonemes and Graphemes for Neural TTS

Ye Jia¹, Heiga Zen², Jonathan Shen¹, Yu Zhang¹, Yonghui Wu¹; ¹Google, USA; ²Google, Japan Tue-M-V-2-9, Time: 09:30

This paper introduces PnG BERT, a new encoder model for neural TTS. This model is augmented from the original BERT model, by taking both phoneme and grapheme representations of text as input, as well as the word-level alignment between them. It can be pre-trained on a large text corpus in a self-supervised manner, and fine-tuned in a TTS task. Experimental results show that a neural TTS model using a pre-trained PnG BERT as its encoder yields more natural prosody and more accurate pronunciation than a baseline model using only phoneme input with no pre-training. Subjective side-by-side preference evaluations show that raters have no statistically significant preference between the speech synthesized using a PnG BERT and ground truth recordings from professional speakers.

Speed up Training with Variable Length Inputs by **Efficient Batching Strategies**

Zhenhao Ge, Lakshmish Kaushik, Masanori Omote, Saket Kumar; Sony, USA Tue-M-V-2-10, Time: 09:30

In the model training with neural networks, although the model performance is always the first priority to optimize, training efficiency also plays an important role in model deployment. There are many ways to speed up training with minimal performance loss, such as training with more GPUs, or with mixed precisions, optimizing training parameters, or making features more compact but more representable. Since mini-batch training is now the go-to approach for many machine learning tasks, minimizing the zero-padding to incorporate samples of different lengths into one batch, is an alter-

NOTES

native approach to save training time. Here we propose a batching strategy based on semi-sorted samples, with dynamic batch sizes and batch randomization. By replacing the random batching with the proposed batching strategies, it saves more than 40% training time without compromising performance in training seq2seq neural text-to-speech models based on the Tacotron framework. We also compare it with two other batching strategies and show it performs similarly in terms of saving time and maintaining performance, but with a simpler concept and a smoother tuning parameter to balance between zero-padding and randomness level.

Tue-M-V-3: Speech Enhancement and Intelligibility

09:30-11:30, Tuesday 31 August 2021 Chairs: Sharon Gannot and Tao Zhang

Funnel Deep Complex U-Net for Phase-Aware Speech Enhancement

Yuhang Sun, Linju Yang, Huifeng Zhu, Jie Hao; OPPO, China

Tue-M-V-3-1, Time: 09:30

The emergence of deep neural networks has made speech enhancement well developed. Most of the early models focused on estimating the magnitude of spectrum while ignoring the phase, this gives the evaluation result a certain upper limit. Some recent researches proposed deep complex network, which can handle complex inputs, and realize joint estimation of magnitude spectrum and phase spectrum by outputting real and imaginary parts respectively. The encoder-decoder structure in Deep Complex U-net (DCU) has been proven to be effective for complex-valued data. To further improve the performance, in this paper, we design a new network called Funnel Deep Complex U-net (FDCU), which could process magnitude information and phase information separately through one-encodertwo-decoders structure. Moreover, in order to achieve better training effect, we define negative stretched-SI-SNR as the loss function to avoid errors caused by the negative vector angle. Experimental results show that our FDCU model outperforms state-of-the-art approaches in all evaluation metrics.

Temporal Convolutional Network with Frequency Dimension Adaptive Attention for Speech Enhancement

Qiquan Zhang¹, Qi Song², Aaron Nicolson³, Tian Lan², Haizhou Li¹; ¹NUS, Singapore; ²Alibaba, China; ³CSIRO, Australia

Tue-M-V-3-2, Time: 09:30

Despite much progress, most temporal convolutional networks (TCN) based speech enhancement models are mainly focused on modeling the long-term temporal contextual dependencies of speech frames, without taking into account the distribution information of speech signal in frequency dimension. In this study, we propose a frequency dimension adaptive attention (FAA) mechanism to improve TCNs, which guides the model selectively emphasize the frequency-wise features with important speech information and also improves the representation capability of network. Our extensive experimental investigation demonstrates that the proposed FAA mechanism is able to consistently provide significant improvements in terms of speech quality (PESQ), intelligibility (STOI) and three other composite metrics. More promisingly, it has better generalization ability to real-world noisy environment.

Perceptual Contributions of Vowels and Consonant-Vowel Transitions in Understanding Time-Compressed Mandarin Sentences

*Changjie Pan*¹, *Feng Yang*², *Fei Chen*¹; ¹*SUSTech, China*; ²*Shenzhen Second People's Hospital, China* Tue-M-V-3-3, Time: 09:30

Many early studies reported the importance of vowels and vowelconsonant transitions to speech intelligibility. The present work assessed their perceptual impacts to the understanding of timecompressed sentences, which could be used to measure the temporal acuity during speech understanding. Mandarin sentences were edited to selectively preserve vowel centers or vowel-consonant transitional segments, and compress the rest regions with equipment time compression rates (TCRs) up to 3, including conditions only preserving vowel centers or vowel-consonant transitions. The processed stimuli were presented to normal-hearing listeners to recognize. Results showed that, consistent with the segmental contributions in understanding uncompressed speech, the vowel-only time-compressed stimuli were highly intelligible (i.e., intelligibility score >85%) at a TCR around 3, and vowel-consonant transitions carried important intelligibility information in understanding time-compressed sentences. The time-compression conditions in the present work provided higher intelligibility scores than their counterparties in understanding the PSOLA-processed time-compressed sentences with TCRs around 3. The findings in this work suggested that the design of time compression processing could be guided towards selectively preserving perceptually important speech segments (e.g., vowels) in the future.

Transfer Learning for Speech Intelligibility Improvement in Noisy Environments

*Ritujoy Biswas*¹, *Karan Nathwani*¹, *Vinayak Abrol*²; ¹*IIT Jammu, India*; ²*IIIT Delhi, India* Tue-M-V-3-4, Time: 09:30

In a recent work [1], a novel Delta Function-based Formant Shifting approach was proposed for speech intelligibility improvement. The underlying principle is to dynamically relocate the formants based on their occurrence in the spectrum away from the region of noise. The manner in which the formants are shifted is decided by the parameters of the Delta Function, the optimal values of which are evaluated using Comprehensive Learning Particle Swarm Optimization (CLPSO). Although effective, CLPSO is computationally expensive to the extent that it overshadows its merits in intelligibility improvement. As a solution to this, the current work aims to improve the Short-Time Objective Intelligibility (STOI) of (target) speech using a Delta Function that has been generated using a different (source) language. This transfer learning is based upon the relative positioning of the formant frequencies and pitch values of the source & target language datasets. The proposed approach is demonstrated and validated by subjecting it to experimentation with three different languages under variable noisy conditions.

Comparison of Remote Experiments Using Crowdsourcing and Laboratory Experiments on Speech Intelligibility

Ayako Yamamoto¹, Toshio Irino¹, Kenichi Arai², Shoko Araki², Atsunori Ogawa², Keisuke Kinoshita², Tomohiro Nakatani²; ¹Wakayama University, Japan; ²NTT, Japan Tue-M-V-3-5, Time: 09:30

Many subjective experiments have been performed to develop objective speech intelligibility measures, but the novel coronavirus outbreak has made it difficult to conduct experiments in a laboratory. One solution is to perform remote testing using crowdsourcing; however, because we cannot control the listening conditions, it is unclear whether the results are entirely reliable. In this study, we compared the speech intelligibility scores obtained from remote and laboratory experiments. The results showed that the mean and standard deviation (SD) of the remote experiments' speech reception threshold (SRT) were higher than those of the laboratory experiments. However, the variance in the SRTs across the speech-enhancement conditions revealed similarities, implying that remote testing results may be as useful as laboratory experiments to develop an objective measure. We also show that practice session scores are correlated with SRT values. This is a priori information before performing the main tests and would be useful for data screening to reduce the variability of the SRT distribution.

Know Your Enemy, Know Yourself: A Unified Two-Stage Framework for Speech Enhancement

Wenzhe Liu, Andong Li, Yuxuan Ke, Chengshi Zheng, Xiaodong Li; CAS, China Tue-M-V-3-6, Time: 09:30

Traditional spectral subtraction-type single channel speech enhancement (SE) algorithms often need to estimate interference components including noise and/or reverberation before subtracting them while deep neural network-based SE methods often aim to realize the end-to-end target mapping. In this paper, we show that both denoising and dereverberation can be unified into a common problem by introducing a two-stage paradigm, namely for interference components estimation and speech recovery. In the first stage, we propose to explicitly extract the magnitude of interference components, which serves as the prior information. In the second stage, with the guidance of this estimated magnitude prior, we can expect to better recover the target speech. In addition, we propose a transform module to facilitate the interaction between interference components and the desired speech modalities. Meanwhile, a temporal fusion module is designed to model long-term dependencies without ignoring short-term details. We conduct the experiments on the WSJ0-SI84 corpus and the results on both denoising and dereverberation tasks show that our approach outperforms previous advanced systems and achieves state-of-the-art performance in terms of many objective metrics.

Speech Enhancement with Weakly Labelled Data from AudioSet

Qiuqiang Kong, Haohe Liu, Xingjian Du, Li Chen, Rui Xia, Yuxuan Wang; ByteDance, China Tue-M-V-3-7, Time: 09:30

Speech enhancement is a task to improve the intelligibility and perceptual quality of degraded speech signals. Recently, neural network-based methods have been applied to speech enhancement. However, many neural network-based methods require users to collect clean speech and background noise for training, which can be time-consuming. In addition, speech enhancement systems trained on particular types of background noise may not generalize well to a wide range of noise. To tackle those problems, we propose a speech enhancement framework trained on weakly labelled data. We first apply a pretrained sound event detection system to detect anchor segments that contain sound events in audio clips. Then, we randomly mix two detected anchor segments as a mixture. We build a conditional source separation network using the mixture and a conditional vector as input. The conditional vector is obtained from the audio tagging predictions on the anchor segments. In inference, we input a noisy speech signal with the one-hot encoding of "Speech" as a condition to the trained system to predict enhanced speech. Our system achieves a PESQ of 2.28 and an SSNR of 8.75 dB on the

VoiceBank-DEMAND dataset, outperforming the previous SEGAN system of 2.16 and 7.73 dB respectively.

Improving Perceptual Quality by Phone-Fortified Perceptual Loss Using Wasserstein Distance for Speech Enhancement

*Tsun-An Hsieh*¹, *Cheng Yu*¹, *Szu-Wei Fu*¹, *Xugang Lu*², *Yu Tsao*¹; ¹*Academia Sinica, Taiwan;* ²*NICT, Japan* Tue-M-V-3-8, Time: 09:30

Speech enhancement (SE) aims to improve speech quality and intelligibility, which are both related to a smooth transition in speech segments that may carry linguistic information, e.g. phones and syllables. In this study, we propose a novel phone-fortified perceptual loss (PFPL) that takes phonetic information into account for training SE models. To effectively incorporate the phonetic information, the PFPL is computed based on latent representations of the wav2vec model, a powerful self-supervised encoder that renders rich phonetic information. To more accurately measure the distribution distances of the latent representations, the PFPL adopts the Wasserstein distance as the distance measure. Our experimental results first reveal that the PFPL is more correlated with the perceptual evaluation metrics, as compared to signal-level losses. Moreover, the results showed that the PFPL can enable a deep complex U-Net SE model to achieve highly competitive performance in terms of standardized quality and intelligibility evaluations on the Voice Bank-DEMAND dataset.

MetricGAN+: An Improved Version of MetricGAN for Speech Enhancement

Szu-Wei Fu¹, Cheng Yu¹, Tsun-An Hsieh¹, Peter Plantinga², Mirco Ravanelli³, Xugang Lu⁴, Yu Tsao¹; ¹Academia Sinica, Taiwan; ²Ohio State University, USA; ³Mila, Canada; ⁴NICT, Japan Tue-M-V-3-9, Time: 09:30

The discrepancy between the cost function used for training a speech enhancement model and human auditory perception usually makes the quality of enhanced speech unsatisfactory. Objective evaluation metrics which consider human perception can hence serve as a bridge to reduce the gap. Our previously proposed MetricGAN was designed to optimize objective metrics by connecting the metric with a discriminator. Because only the scores of the target evaluation functions are needed during training, the metrics can even be non-differentiable. In this study, we propose a MetricGAN+ in which three training techniques incorporating domain-knowledge of speech processing are proposed. With these techniques, experimental results on the VoiceBank-DEMAND dataset show that MetricGAN+ can increase PESQ score by 0.3 compared to the previous MetricGAN and achieve state-of-the-art results (PESQ score = 3.15).

A Spectro-Temporal Glimpsing Index (STGI) for Speech Intelligibility Prediction

Amin Edraki¹, Wai-Yip Chan¹, Jesper Jensen², Daniel Fogerty³; ¹Queen's University, Canada; ²Aalborg University, Denmark; ³University of Illinois at Urbana-Champaign, USA Tue-M-V-3-10, Time: 09:30

We propose a monaural intrusive speech intelligibility prediction (SIP) algorithm called STGI based on detecting *glimpses* in short-time segments in a spectro-temporal modulation decomposition of the input speech signals. Unlike existing glimpse-based SIP methods, the application of STGI is not limited to additive uncorrelated noise; STGI can be employed in a broad range of degradation conditions.

Our results show that STGI performs consistently well across 15 datasets covering degradation conditions including modulated noise, noise reduction processing, reverberation, near-end listening enhancement, checkerboard noise, and gated noise.

Self-Supervised Learning Based Phone-Fortified Speech Enhancement

Yuanhang Qiu, Ruili Wang, Satwinder Singh, Zhizhong Ma, Feng Hou; Massey University, New Zealand Tue-M-V-3-11, Time: 09:30

For speech enhancement, deep complex network based methods have shown promising performance due to their effectiveness in dealing with complex-valued spectrums. Recent speech enhancement methods focus on further optimization of network structures and hyperparameters, however, ignore inherent speech characteristics (e.g., phonetic characteristics), which are important for networks to learn and reconstruct speech information. In this paper, we propose a novel self-supervised learning based phone-fortified (SSPF) method for speech enhancement. Our method explicitly imports phonetic characteristics into a deep complex convolutional network via a Contrastive Predictive Coding (CPC) model pre-trained with self-supervised learning. This operation can greatly improve speech representation learning and speech enhancement performance. Moreover, we also apply the self-attention mechanism to our model for learning long-range dependencies of a speech sequence, which further improves the performance of speech enhancement. The experimental results demonstrate that our SSPF method outperforms existing methods and achieves state-of-the-art performance in terms of speech quality and intelligibility.

Incorporating Embedding Vectors from a Human Mean-Opinion Score Prediction Model for Monaural Speech Enhancement

Khandokar Md. Nayem, Donald S. Williamson; Indiana University, USA

Tue-M-V-3-12, Time: 09:30

Objective measures of success, such as the perceptual evaluation of speech quality (PESQ), signal-to-distortion ratio (SDR), and short-time objective intelligibility (STOI), have recently been used to optimize deep-learning based speech enhancement algorithms, in an effort to incorporate perceptual constraints into the learning process. Optimizing with these measures, however, may be sub-optimal, since the objective scores do not always strongly correlate with a listener's evaluation. This motivates the need for approaches that either are optimized with scores that are strongly correlated with human assessments or that use alternative strategies for incorporating perceptual constraints. In this work, we propose an attention-based approach that uses learned speech embedding vectors from a meanopinion score (MOS) prediction model and a speech enhancement module to jointly enhance noisy speech. Our loss function is jointly optimized with signal approximation and MOS prediction loss terms. We train the model using real-world noisy speech data that has been captured in everyday environments. The results show that our proposed model significantly outperforms other approaches that are optimized with objective measures.

Restoring Degraded Speech via a Modified Diffusion Model

Jianwei Zhang, Suren Jayasuriya, Visar Berisha; Arizona State University, USA

Tue-M-V-3-13, Time: 09:30

There are many deterministic mathematical operations (e.g. compression, clipping, downsampling) that degrade speech quality

considerably. In this paper we introduce a neural network architecture, based on a modification of the DiffWave model, that aims to restore the original speech signal. DiffWave, a recently published diffusion-based vocoder, has shown state-of-the-art synthesized speech quality and relatively shorter waveform generation times, with only a small set of parameters. We replace the mel-spectrum upsampler in DiffWave with a deep CNN upsampler, which is trained to alter the degraded speech mel-spectrum to match that of the original speech. The model is trained using the original speech waveform, but conditioned on the degraded speech mel-spectrum. Post-training, only the degraded mel-spectrum is used as input and the model generates an estimate of the original speech. Our model results in improved speech quality (original DiffWave model as baseline) on several different experiments. These include improving the quality of speech degraded by LPC-10 compression, AMR-NB compression, and signal clipping. Compared to the original DiffWave architecture, our scheme achieves better performance on several objective perceptual metrics and in subjective comparisons. Improvements over baseline are further amplified in a out-of-corpus evaluation setting.

Tue-M-V-4: Spoken Dialogue Systems I

09:30-11:30, Tuesday 31 August 2021 Chairs: Sakriani Sakti and Evelyne Tzoukerman

User-Initiated Repetition-Based Recovery in Multi-Utterance Dialogue Systems

Hoang Long Nguyen, Vincent Renkens, Joris Pelemans, Srividya Pranavi Potharaju, Anil Kumar Nalamalapu, Murat Akbacak; Apple, USA Tue-M-V-4-1, Time: 09:30

Recognition errors are common in human communication. Similar errors often lead to unwanted behaviour in dialogue systems or virtual assistants. In human communication, we can recover from them by repeating misrecognized words or phrases; however in human-machine communication this recovery mechanism is not available. In this paper, we attempt to bridge this gap and present a system that allows a user to correct speech recognition errors in a virtual assistant by repeating misunderstood words. When a user repeats part of the phrase the system rewrites the original query to incorporate the correction. This rewrite allows the virtual assistant to understand the original query successfully. We present an end-to-end 2-step attention pointer network that can generate the the rewritten query by merging together the incorrectly understood utterance with the correction follow-up. We evaluate the model on data collected for this task and compare the proposed model to a rule-based baseline and a standard pointer network. We show that rewriting the original query is an effective way to handle repetitionbased recovery and that the proposed model outperforms the rule based baseline, reducing Word Error Rate by 19% relative at 2% False Alarm Rate on annotated data.

Self-Supervised Dialogue Learning for Spoken Conversational Question Answering

*Nuo Chen*¹, *Chenyu You*², *Yuexian Zou*¹; ¹*Peking University, China*; ²*Yale University, USA* Tue-M-V-4-2, Time: 09:30

In spoken conversational question answering (SCQA), the answer to the corresponding question is generated by retrieving and then analyzing a fixed spoken document, including multi-part conversations. Most SCQA systems have considered only retrieving information from ordered utterances. However, the sequential order of dialogue is important to build a robust spoken conversational question answering system, and the changes of utterances order may severely result in low-quality and incoherent corpora. To this end, we introduce a self-supervised learning approach, including *incoherence discrimination, insertion detection,* and *question prediction,* to explicitly capture the coreference resolution and dialogue coherence among spoken documents. Specifically, we design a joint learning framework where the auxiliary self-supervised tasks can enable the pre-trained SCQA systems towards more coherent and meaningful spoken dialogue learning. We also utilize the proposed self-supervised learning tasks to capture intra-sentence coherence. Experimental results demonstrate that our proposed method provides more coherent, meaningful, and appropriate responses, yielding superior performance gains compared to the original pre-trained language models. Our method achieves state-of-the-art results on the Spoken-CoQA dataset.

Act-Aware Slot-Value Predicting in Multi-Domain Dialogue State Tracking

Ruolin Su, Ting-Wei Wu, Biing-Hwang Juang; Georgia Tech, USA

Tue-M-V-4-3, Time: 09:30

As an essential component in task-oriented dialogue systems, dialogue state tracking (DST) aims to track human-machine interactions and generate state representations for managing the dialogue. Representations of dialogue states are dependent on the domain ontology and the user's goals. In several task-oriented dialogues with a limited scope of objectives, dialogue states can be represented as a set of slot-value pairs. As the capabilities of dialogue systems expand to support increasing naturalness in communication, incorporating dialogue act processing into dialogue model design becomes essential. The lack of such consideration limits the scalability of dialogue state tracking models for dialogues having specific objectives and ontology. To address this issue, we formulate and incorporate dialogue acts, and leverage recent advances in machine reading comprehension to predict both categorical and non-categorical types of slots for multi-domain dialogue state tracking. Experimental results show that our models can improve the overall accuracy of dialogue state tracking on the MultiWOZ 2.1 dataset, and demonstrate that incorporating dialogue acts can guide dialogue state design for future task-oriented dialogue systems.

Dialogue Situation Recognition for Everyday Conversation Using Multimodal Information

*Yuya Chiba*¹, *Ryuichiro Higashinaka*²; ¹*NTT*, *Japan*; ²*Nagoya University, Japan* Tue-M-V-4-4, Time: 09:30

In recent years, dialogue systems have been applied to daily living. Such systems should be able to associate conversations with dialogue situations, such as a place where a dialogue occurs and the relationship between participants. In this study, we propose a dialogue situation recognition method that understands the perspective of dialogue scenes. The target dialogue situations contain dialogue styles, places, activities, and relations between participants. We used the Corpus of Everyday Japanese Conversation (CEJC), which records natural everyday conversations in various situations for experiments. We experimentally verified the effectiveness of our proposed method using multimodal information for situation recognition.

Neural Spoken-Response Generation Using Prosodic and Linguistic Context for Conversational Systems

Yoshihiro Yamazaki¹, Yuya Chiba², Takashi Nose¹, Akinori Ito¹; ¹Tohoku University, Japan; ²NTT, Japan Tue-M-V-4-5, Time: 09:30

Spoken dialogue systems have become widely used in daily life. Such a system must interact with the user socially to truly operate

as a partner with humans. In studies of recent dialogue systems, neural response generation led to natural response generation. However, these studies have not considered the acoustic aspects of conversational phenomena, such as the adaptation of prosody. We propose a spoken-response generation model that extends a neural conversational model to deal with pitch control signals. Our proposed model is trained using multimodal dialogue between humans. The generated pitch control signals are input to a speech synthesis system to control the pitch of synthesized speech. Our experiment shows that the proposed system can generate synthesized speech with an appropriate F0 contour as an utterance in context compared to the output of a system without pitch control, although language generation remains an issue.

Semantic Transportation Prototypical Network for Few-Shot Intent Detection

*Weiyuan Xu*¹, *Peilin Zhou*¹, *Chenyu You*², *Yuexian Zou*¹; ¹*Peking University, China*; ²*Yale University, USA* Tue-M-V-4-6, Time: 09:30

Few-shot intent detection is a problem that only a few annotated examples are available for unseen intents, and deep models could suffer from the overfitting problem because of scarce data. Existing state-of-the-art few-shot model, Prototypical Network (PN), mainly focus on computing the similarity between examples in a metric space by leveraging sentence-level instance representations. However, sentence-level representations may incorporate highly noisy signals from unrelated words which leads to performance degradation. In this paper, we propose Semantic Transportation Prototypical Network (STPN) to alleviate this issue. Different from the original PN, our approach takes word-level representation as input and uses a new distance metric to obtain better sample matching result. And we reformulate the few-shot classification task into an instance of optimal matching, in which the key word semantic information between examples are expected to be matched and the matching cost is treated as similarity. Specifically, we design Mutual-Semantic mechanism to generate word semantic information, which could reduce the unrelated word noise and enrich key word information. Then, Earth Mover's Distance (EMD) is applied to find an optimal matching solution. Comprehensive experiments on two benchmark datasets are conducted to validate the effectiveness and generalization of our proposed model.

Domain-Specific Multi-Agent Dialog Policy Learning in Multi-Domain Task-Oriented Scenarios

Li Tang, Yuke Si, Longbiao Wang, Jianwu Dang; Tianjin University, China

Tue-M-V-4-7, Time: 09:30

Traditional dialog policy learning methods train a generic dialog agent to address all situations. However, when the dialog agent encounters a complicated task that involves more than one domain, it becomes difficult to perform concordant actions due to the hybrid information in the multi-domain ontology. Inspired by a real-life scenario at a bank, there are always several specialized departments that deal with different businesses. In this paper, we propose Domain-Specific Multi-Agent Dialog Policy Learning (DSMADPL), in which the dialog system is composed of a set of agents where each agent represents a specialized skill in a particular domain. Every domain-specific agent is first pretrained with supervised learning using a dialog corpus, and then they are jointly improved with multiagent reinforcement learning. When the dialog system interacts with the user, in each turn the system action is decided by the actions of relevant agents. Experiments conducted on the commonly used MultiWOZ dataset prove the effectiveness of the proposed method, in which dialog success rate increases from 55.0% for the traditional method to 67.2% for our method in multi-domain scenarios.

NOTES

Leveraging ASR N-Best in Deep Entity Retrieval

Haoyu Wang¹, John Chen², Majid Laali³, Kevin Durda³, Jeff King³, William Campbell¹, Yang Liu¹; ¹Amazon, USA; ²University of Toronto, Canada; ³Amazon, Canada Tue-M-V-4-8, Time: 09:30

Entity Retrieval (ER) in spoken dialog systems is a task that retrieves entities in a catalog for the entity mentions in user utterances. ER systems are susceptible to upstream errors, with Automatic Speech Recognition (ASR) errors being particularly troublesome. In this work, we propose a robust deep learning based ER system by leveraging ASR N-best hypotheses. Specifically, we evaluate different neural architectures to infuse ASR N-best through an attention mechanism. On 750 hours of audio data taken from live traffic, our best model achieves 11.07% relative error reduction while maintaining the same performance on rejecting out-of-domain ER requests.

Tue-M-V-5: Topics in ASR: Robustness, Feature Extraction, and Far-Field ASR 09:30-11:30, Tuesday 31 August 2021

Chairs: Sriram Ganapathy and Seyyed Saeed Sarfjoo

End-to-End Spelling Correction Conditioned on Acoustic Feature for Code-Switching Speech Recognition

Shuai Zhang¹, Jiangyan Yi², Zhengkun Tian¹, Ye Bai¹, Jianhua Tao¹, Xuefei Liu², Zhengqi Wen²; ¹UCAS, China; ²CAS, China Tue-M-V-5-1, Time: 09:30

In this work, we propose a new end-to-end (E2E) spelling correction method for post-processing of code-switching automatic speech recognition (ASR). Existing E2E spelling correction models take the hypotheses of ASR as inputs and annotated text as the targets. Due to the powerful modeling capabilities of the E2E model, the training of the correction system is extremely prone to over-fitting. It usually requires sufficient data diversity for reliable training. Therefore, it is difficult to apply the E2E correction models to the code-switching ASR task because of the data shortage. In this paper, we introduce the acoustic features into the spelling correction model. Our method can alleviate the problem of over-fitting and has better performance. Meanwhile, because the acoustic features are encode-free, our proposed model can be applied to the ASR model without significantly increasing the computational cost. The experimental results on ASRU 2019 Mandarin-English Code-switching Challenge data set show that the proposed method achieves 11.14% relative error rate reduction compared with baseline.

Phoneme Recognition Through Fine Tuning of Phonetic Representations: A Case Study on Luhya Language Varieties

Kathleen Siminyu¹, Xinjian Li², Antonios Anastasopoulos³, David R. Mortensen², Michael R. Marlo⁴, Graham Neubig²; ¹Georgia Tech, USA; ²Carnegie Mellon University, USA; ³George Mason University, USA; ⁴Mizzou, USA Tue-M-V-5-2, Time: 09:30

Models pre-trained on multiple languages have shown significant promise for improving speech recognition, particularly for lowresource languages. In this work, we focus on phoneme recognition using Allosaurus, a method for multilingual recognition based on phonetic annotation, which incorporates phonological knowledge through a language-dependent allophone layer that associates a universal narrow phone-set with the phonemes that appear in each language. To evaluate in a challenging real-world scenario, we curate phone recognition datasets for Bukusu and Saamia, two varieties of the Luhya language cluster of western Kenya and eastern Uganda. To our knowledge, these datasets are the first of their kind. We carry out similar experiments on the dataset of an endangered Tangkhulic language, East Tusom, a Tibeto-Burman language variety spoken mostly in India. We explore both zero-shot and few-shot recognition by fine-tuning using datasets of varying sizes (10 to 1000 utterances). We find that fine-tuning of Allosaurus, even with just 100 utterances, leads to significant improvements in phone error rates.

Speech Acoustic Modelling Using Raw Source and Filter Components

*Erfan Loweimi*¹, *Zoran Cvetkovic*², *Peter Bell*¹, *Steve Renals*¹; ¹*University of Edinburgh, UK*; ²*King's College London, UK*

Tue-M-V-5-3, Time: 09:30

Source-filter modelling is among the fundamental techniques in speech processing with a wide range of applications. In acoustic modelling, features such as MFCC and PLP which parametrise the filter component are widely employed. In this paper, we investigate the efficacy of building acoustic models from the raw filter and source components. The raw magnitude spectrum, as the primary information stream, is decomposed into the excitation and vocal tract information streams via cepstral liftering. Then, acoustic models are built via multi-head CNNs which, among others, allow for processing each individual stream via a sequence of bespoke transforms and fusing them at an optimal level of abstraction. We discuss the possible advantages of such information factorisation and recombination, investigate the dynamics of these models and explore the optimal fusion level. Furthermore, we illustrate the CNN's learned filters and provide some interpretation for the captured patterns. The proposed approach with optimal fusion scheme results in up to 14% and 7% relative WER reduction in WSJ and Aurora-4 tasks.

Noise Robust Acoustic Modeling for Single-Channel Speech Recognition Based on a Stream-Wise Transformer Architecture

Masakiyo Fujimoto, Hisashi Kawai; NICT, Japan Tue-M-V-5-4, Time: 09:30

This paper addresses a noise-robust automatic speech recognition (ASR) method under the constraints of real-time, one-pass, and single-channel processing. Under such strong constraints, singlechannel speech enhancement becomes a key technology because methods with multiple-passes or batch processing, such as acoustic model adaptation, are not suitable for use. However, single-channel speech enhancement often degrades ASR performance due to speech distortion. To overcome this problem, we propose a noise robust acoustic modeling method based on the stream-wise transformer model. The proposed method accepts multi-stream features obtained by multiple single-channel speech enhancement methods as input and selectively uses an appropriate feature stream according to the noise environment by paying attention to the noteworthy stream on the basis of multi-head attention. The proposed method considers the attention for the stream direction instead of the time series direction, and it is thus capable of real-time and low-latency processing. Comparative evaluations reveal that the proposed method successfully improves the accuracy of ASR in noisy environments and reduces the number of model parameters even under strong constraints.

IR-GAN: Room Impulse Response Generator for Far-Field Speech Recognition

Anton Ratnarajah, Zhenyu Tang, Dinesh Manocha; University of Maryland, USA Tue-M-V-5-5, Time: 09:30

We present a Generative Adversarial Network (GAN) based room impulse response generator (IR-GAN) for generating realistic synthetic room impulse responses (RIRs). IR-GAN extracts acoustic parameters from captured real-world RIRs and uses these parameters to generate new synthetic RIRs. We use these generated synthetic RIRs to improve far-field automatic speech recognition in new environments that are different from the ones used in training datasets. In particular, we augment the far-field speech training set by convolving our synthesized RIRs with a clean LibriSpeech dataset [1]. We evaluate the quality of our synthetic RIRs on the far-field LibriSpeech test set created using real-world RIRs from the BUT ReverbDB [2] and AIR [3] datasets. Our IR-GAN reports up to an 8.95% lower error rate than Geometric Acoustic Simulator (GAS) in far-field speech recognition benchmarks. We further improve the performance when we combine our synthetic RIRs with synthetic impulse responses generated using GAS. This combination can reduce the word error rate by up to 14.3% in far-field speech recognition benchmarks.

Scaling Sparsemax Based Channel Selection for Speech Recognition with ad-hoc Microphone Arrays

Junqi Chen, Xiao-Lei Zhang; Northwestern Polytechnical

University, China Tue-M-V-5-6, Time: 09:30

Recently, speech recognition with ad-hoc microphone arrays has received much attention. It is known that channel selection is an important problem of ad-hoc microphone arrays, however, this topic seems far from explored in speech recognition yet, particularly with a large-scale ad-hoc microphone array. To address this problem, we propose a Scaling Sparsemax algorithm for the channel selection problem of the speech recognition with large-scale ad-hoc microphone arrays. Specifically, we first replace the conventional Softmax operator in the stream attention mechanism of a multichannel endto-end speech recognition system with Sparsemax, which conducts channel selection by forcing the channel weights of noisy channels to zero. Because Sparsemax punishes the weights of many channels to zero harshly, we propose Scaling Sparsemax which punishes the channels mildly by setting the weights of very noisy channels to zero only. Experimental results with ad-hoc microphone arrays of over 30 channels under the conformer speech recognition architecture show that the proposed Scaling Sparsemax yields a word error rate of over 30% lower than Softmax on simulation data sets, and over 20% lower on semi-real data sets, in test scenarios with both matched and mismatched channel numbers.

Multi-Channel Transformer Transducer for Speech Recognition

Feng-Ju Chang, Martin Radfar, Athanasios Mouchtaris, Maurizio Omologo; Amazon, USA Tue-M-V-5-7, Time: 09:30

Multi-channel inputs offer several advantages over single-channel, to improve the robustness of on-device speech recognition systems. Recent work on multi-channel transformer, has proposed a way to incorporate such inputs into end-to-end ASR for improved accuracy. However, this approach is characterized by a high computational complexity, which prevents it from being deployed in on-device systems. In this paper, we present a novel speech recognition model, *Multi-Channel Transformer Transducer (MCTT)*, which features end-to-end multi-channel training, low computation cost, and low

latency so that it is suitable for streaming decoding in on-device speech recognition. In a far-field in-house dataset, our MCTT outperforms stagewise multi-channel models with transformer-transducer up to 6.01% relative WER improvement (WERR). In addition, MCTT outperforms the multi-channel transformer up to 11.62% WERR, and is 15.8 times faster in terms of inference speed. We further show that we can improve the computational cost of MCTT by constraining the future and previous context in attention computations.

Data Augmentation Methods for End-to-End Speech Recognition on Distant-Talk Scenarios

*Emiru Tsunoo*¹, *Kentaro Shibata*¹, *Chaitanya Narisetty*², *Yosuke Kashiwagi*¹, *Shinji Watanabe*²; ¹*Sony, Japan;* ²*Carnegie Mellon University, USA* Tue-M-V-5-8, Time: 09:30

Although end-to-end automatic speech recognition (E2E ASR) has achieved great performance in tasks that have numerous paired data, it is still challenging to make E2E ASR robust against noisy and low-resource conditions. In this study, we investigated data augmentation methods for E2E ASR in distant-talk scenarios. E2E ASR models are trained on the series of CHiME challenge datasets, which are suitable tasks for studying robustness against noisy and spontaneous speech. We propose to use three augmentation methods and their combinations: 1) data augmentation using text-tospeech (TTS) data, 2) cycle-consistent generative adversarial network (Cycle-GAN) augmentation trained to map two different audio characteristics, the one of clean speech and of noisy recordings, to match the testing condition, and 3) pseudo-label augmentation provided by the pretrained ASR module for smoothing label distributions. Experimental results using the CHiME-6/CHiME-4 datasets show that each augmentation method individually improves the accuracy on top of the conventional SpecAugment; further improvements are obtained by combining these approaches. We achieved 4.3% word error rate (WER) reduction, which was more significant than that of the SpecAugment, when we combine all three augmentations for the CHiME-6 task.

Leveraging Phone Mask Training for Phonetic-Reduction-Robust E2E Uyghur Speech Recognition

Guodong Ma¹, Pengfei Hu², Jian Kang², Shen Huang², Hao Huang¹; ¹Xinjiang University, China; ²Tencent, China

Tue-M-V-5-9, Time: 09:30

In Uyghur speech, consonant and vowel reduction are often encountered, especially in spontaneous speech with high speech rate, which will cause a degradation of speech recognition performance. To solve this problem, we propose an effective phone mask training method for Conformer-based Uyghur end-to-end (E2E) speech recognition. The idea is to randomly mask off a certain percentage features of phones during model training, which simulates the above verbal phenomena and facilitates E2E model to learn more contextual information. According to experiments, the above issues can be greatly alleviated. In addition, deep investigations are carried out into different units in masking, which shows the effectiveness of our proposed masking unit. We also further study the masking method and optimize filling strategy of phone mask. Finally, compared with Conformer-based E2E baseline without mask training, our model demonstrates about 5.51% relative Word Error Rate (WER) reduction on reading speech and 12.92% on spontaneous speech, respectively. The above approach has also been verified on test-set of open-source data THUYG-20, which shows 20% relative improvements.

NOTES

Rethinking Evaluation in ASR: Are Our Models Robust Enough?

Tatiana Likhomanenko¹, Qiantong Xu¹, Vineel Pratap¹, Paden Tomasello¹, Jacob Kahn¹, Gilad Avidov¹, Ronan Collobert¹, Gabriel Synnaeve²; ¹Facebook, USA; ²Facebook, France

Tue-M-V-5-10, Time: 09:30

Is pushing numbers on a single benchmark valuable in automatic speech recognition? Research results in acoustic modeling are typically evaluated based on performance on a single dataset. While the research community has coalesced around various benchmarks, we set out to understand generalization performance in acoustic modeling across datasets — in particular, if models trained on a single dataset transfer to other (possibly out-of-domain) datasets. Further, we demonstrate that when a large enough set of benchmarks is used, average word error rate (WER) performance over them provides a good proxy for performance on real-world data. Finally, we show that training a single acoustic model on the most widely-used datasets — combined — reaches competitive performance on both research and real-world benchmarks.

Raw Waveform Encoder with Multi-Scale Globally Attentive Locally Recurrent Networks for End-to-End Speech Recognition

*Max W.Y. Lam*¹, *Jun Wang*¹, *Chao Weng*¹, *Dan Su*¹, *Dong Yu*²; ¹*Tencent, China;* ²*Tencent, USA* Tue-M-V-5-11, Time: 09:30

End-to-end speech recognition generally uses hand-engineered acoustic features as input and excludes the feature extraction module from its joint optimization. To extract learnable and adaptive features and mitigate information loss, we propose a new encoder that adopts globally attentive locally recurrent (GALR) networks and directly takes raw waveform as input. We observe improved ASR performance and robustness by applying GALR on different window lengths to aggregate fine-grain temporal information into multi-scale acoustic features. Experiments are conducted on a benchmark dataset AISHELL-2 and two large-scale Mandarin speech corpus of 5,000 hours and 21,000 hours. With faster speed and comparable model size, our proposed multi-scale GALR waveform encoder achieved consistent character error rate reductions (CERRs) from 7.9% to 28.1% relative over strong baselines, including Conformer and TDNN-Conformer. In particular, our approach demonstrated notable robustness than the traditional handcrafted features and outperformed the baseline MFCC-based TDNN-Conformer model by a 15.2% CERR on a music-mixed real-world speech test set.
Tue-M-V-6: Voice Activity Detection and Keyword Spotting

09:30–11:30, Tuesday 31 August 2021 Chairs: Haizhou Li and Tanel Alumäe

Attention-Based Cross-Modal Fusion for Audio-Visual Voice Activity Detection in Musical Video Streams

Yuanbo Hou¹, Zhesong Yu², Xia Liang², Xingjian Du², Bilei Zhu², Zejun Ma², Dick Botteldooren¹; ¹Ghent University, Belgium; ²ByteDance, China Tue-M-V-6-1, Time: 09:30

Many previous audio-visual voice-related works focus on speech, ignoring the singing voice in the growing number of musical video streams on the Internet. For processing diverse musical video data, voice activity detection is a necessary step. This paper attempts to detect the speech and singing voices of target performers in musical video streams using audio-visual information. To integrate information of audio and visual modalities, a multi-branch network is proposed to learn audio and image representations, and the representations are fused by attention based on semantic similarity to shape the acoustic representations through the probability of anchor vocalization. Experiments show the proposed audio-visual multi-branch network far outperforms the audio-only model in challenging acoustic environments, indicating the cross-modal information fusion based on semantic correlation is sensible and successful.

Noise-Tolerant Self-Supervised Learning for Audio-Visual Voice Activity Detection

Ui-Hyun Kim; Toshiba, Japan

Tue-M-V-6-2, Time: 09:30

Recent audio-visual voice activity detectors based on supervised learning require large amounts of labeled training data with manual mouth-region cropping in videos, and the performance is sensitive to a mismatch between the training and testing noise conditions. This paper introduces contrastive self-supervised learning for audio-visual voice activity detection as a possible solution to such problems. In addition, a novel self-supervised learning framework is proposed to improve overall training efficiency and testing performance on noise-corrupted datasets, as in real-world scenarios. This framework includes a branched audio encoder and a noise-tolerant loss function to cope with the uncertainty of speech and noise feature separation in a self-supervised manner. Experimental results, particularly under mismatched noise conditions, demonstrate the improved performance compared with a self-supervised learning baseline and a supervised learning framework.

Noisy Student-Teacher Training for Robust Keyword Spotting

Hyun-Jin Park, Pai Zhu, Ignacio Lopez Moreno, Niranjan Subrahmanya; Google, USA Tue-M-V-6-3, Time: 09:30

We propose self-training with noisy student-teacher approach for streaming keyword spotting, that can utilize large-scale unlabeled data and aggressive data augmentation. The proposed method applies aggressive data augmentation (spectral augmentation) on the input of both student and teacher and utilize unlabeled data at scale, which significantly boosts the accuracy of student against challenging conditions. Such aggressive augmentation usually degrades model performance when used with supervised training with hard-labeled data. Experiments show that aggressive spec

NOTES

augmentation on baseline supervised training method degrades accuracy, while the proposed self-training with noisy student-teacher training improves accuracy of some difficult-conditioned test sets by as much as 60%.

Multi-Channel VAD for Transcription of Group Discussion

*Osamu Ichikawa*¹, *Kaito Nakano*¹, *Takahiro Nakayama*², *Hajime Shirouzu*³; ¹*Shiga University, Japan*; ²*University of Tokyo, Japan*; ³*NIER, Japan* Tue-M-V-6-4, Time: 09:30

Attempts are being made to visualize the learning process by attaching microphones to students participating in group works conducted in classrooms, and subsequently, their speech using an automatic speech recognition (ASR) system. However, the voices of nearby students frequently become mixed with the output speech data, even when using close-talk microphones with noise robustness. To resolve this challenge, in this paper, we propose using multi-channel voice activity detection (VAD) to determine the speech segments of a target speaker while also referencing the output speech from the microphones attached to the other speakers in the group. The conducted evaluation experiments using the actual speech of middle school students during group work lessons showed that our proposed method significantly improves the frame error rate (38.7%) compared to that of the conventional technology, single-channel VAD (49.5%). In our view, conventional approaches, such as distributed microphone arrays and deep learning, are somewhat dependent on the temporal stationarity of the speakers' positions. However, the proposed method is essentially a VAD process and thus works robustly. It is the practical and proven solution in a real classroom environment.

Audio-Visual Information Fusion Using Cross-Modal Teacher-Student Learning for Voice Activity Detection in Realistic Environments

Hengshun Zhou¹, Jun Du¹, Hang Chen¹, Zijun Jing², Shifu Xiong², Chin-Hui Lee³; ¹USTC, China; ²iFLYTEK, China; ³Georgia Tech, USA Tue-M-V-6-5, Time: 09:30

We propose an information fusion approach to audio-visual voice activity detection (AV-VAD) based on cross-modal teacher-student learning leveraging on factorized bilinear pooling (FBP) and Kullback-Leibler (KL) regularization. First, we design an audio-visual network by using FBP fusion to fully utilize the interaction between audio and video modalities. Next, to transfer the rich information in audio-based VAD (A-VAD) model trained with a massive audio-only dataset to AV-VAD model built with relatively limited multi-modal data, a cross-modal teacher-student learning framework is then proposed based on cross entropy with regulated KL-divergence. Finally, evaluated on an in-house dataset recorded in realistic conditions using standard VAD metrics, the proposed approach yields consistent and significant improvements over other state-of-the-art techniques. Moreover, by applying our AV-VAD technique to an audio-visual Chinese speech recognition task, the character error rate is reduced by 24.15% and 8.66% from A-VAD and the baseline AV-VAD systems, respectively.

Enrollment-Less Training for Personalized Voice Activity Detection

Naoki Makishima, Mana Ihori, Tomohiro Tanaka, Akihiko Takashima, Shota Orihashi, Ryo Masumura; NTT, Japan

Tue-M-V-6-6, Time: 09:30

We present a novel personalized voice activity detection (PVAD) learning method that does not require enrollment data during training. PVAD is a task to detect the speech segments of a specific target speaker at the frame level using enrollment speech of the target speaker. Since PVAD must learn speakers' speech variations to clarify the boundary between speakers, studies on PVAD used large-scale datasets that contain many utterances for each speaker. However, the datasets to train a PVAD model are often limited because substantial cost is needed to prepare such a dataset. In addition, we cannot utilize the datasets used to train the standard VAD because they often lack speaker labels. To solve these problems, our key idea is to use one utterance as both a kind of enrollment speech and an input to the PVAD during training, which enables PVAD training without enrollment speech. In our proposed method, called enrollment-less training, we augment one utterance so as to create variability between the input and the enrollment speech while keeping the speaker identity, which avoids the mismatch between training and inference. Our experimental results demonstrate the efficacy of the method.

Voice Activity Detection for Live Speech of Baseball Game Based on Tandem Connection with Speech/Noise Separation Model

Yuto Nonaka¹, Chee Siang Leow¹, Akio Kobayashi², Takehito Utsuro³, Hiromitsu Nishizaki¹; ¹University of Yamanashi, Japan; ²NTUT, Japan; ³University of Tsukuba, Japan

Tue-M-V-6-7, Time: 09:30

When applying voice activity detection (VAD) to a noisy sound, in general, noise reduction (speech separation) and VAD are performed separately. In this case, the noise reduction may suppress the speech, and the VAD may not work well for the speech after the noise reduction. This study proposes a VAD model through the tandem connection of neural network-based noise separation and a VAD model. By training the two models simultaneously, the noise separation model is expected to be trained to consider the VAD results, and thus effective noise separation can be achieved. Moreover, the improved speech/noise separation model will improve the accuracy of the VAD model. In this research, we deal with real-live speeches from baseball games, which have a very poor signal-to-noise ratio. The VAD experiments showed that the VAD performance at the frame level achieved 4.2 points improvement in F1-score by tandemly connecting the speech/noise separation model and the VAD model.

FastICARL: Fast Incremental Classifier and Representation Learning with Efficient Budget Allocation in Audio Sensing Applications

Young D. Kwon, Jagmohan Chauhan, Cecilia Mascolo; University of Cambridge, UK

Tue-M-V-6-8, Time: 09:30

Various incremental learning (IL) approaches have been proposed to help deep learning models learn new tasks/classes continuously without forgetting what was learned previously (i.e., avoid catastrophic forgetting). With the growing number of deployed audio sensing applications that need to dynamically incorporate new tasks and changing input distribution from users, the ability of IL on-device becomes essential for both efficiency and user privacy.

However, prior works suffer from high computational costs and storage demands which hinders the deployment of IL on-device. In this work, to overcome these limitations, we develop an end-to-end and on-device IL framework, FastICARL, that incorporates an exemplarbased IL and quantization in the context of audio-based applications. We first employ k-nearest-neighbor to reduce the latency of IL. Then, we jointly utilize a quantization technique to decrease the storage requirements of IL. We implement FastICARL on two types of mobile devices and demonstrate that FastICARL remarkably decreases the IL time up to 78–92% and the storage requirements by 2–4 times without sacrificing its performance. FastICARL enables complete on-device IL, ensuring user privacy as the user data does not need to leave the device.

End-to-End Transformer-Based Open-Vocabulary Keyword Spotting with Location-Guided Local Attention

Bo Wei¹, Meirong Yang¹, Tao Zhang¹, Xiao Tang¹, Xing Huang¹, Kyuhong Kim², Jaeyun Lee², Kiho Cho², Sung-Un Park²; ¹Samsung, China; ²Samsung, Korea Tue-M-V-6-9, Time: 09:30

Open-vocabulary keyword spotting (KWS) aims to detect arbitrary keywords from continuous speech, which allows users to define their personal keywords. In this paper, we propose a novel location guided end-to-end (E2E) keyword spotting system. Firstly, we predict endpoints of keyword in the entire speech based on attention mechanism. Secondly, we calculate the existence probability of keyword by fusing the located keyword speech segment and text with local attention. The results on Librispeech dataset and Google speech commands dataset show our proposed method significantly outperforms the baseline method and the latest small-footprint E2E KWS method.

Segmental Contrastive Predictive Coding for Unsupervised Word Segmentation

Saurabhchand Bhati, Jesús Villalba, Piotr Żelasko, Laureano Moro-Velázquez, Najim Dehak; Johns Hopkins University, USA

Tue-M-V-6-10, Time: 09:30

Automatic detection of phoneme or word-like units is one of the core objectives in zero-resource speech processing. Recent attempts employ self-supervised training methods, such as contrastive predictive coding (CPC), where the next frame is predicted given past context. However, CPC only looks at the audio signal's frame-level structure. We overcome this limitation with a segmental contrastive predictive coding (SCPC) framework that can model the signal structure at a higher level e.g. at the phoneme level. In this framework, a convolutional neural network learns frame-level representation from the raw waveform via noise-contrastive estimation (NCE). A differentiable boundary detector finds variable-length segments, which are then used to optimize a segment encoder via NCE to learn segment representations. The differentiable boundary detector allows us to train frame-level and segment-level encoders jointly. Typically, phoneme and word segmentation are treated as separate tasks. We unify them and experimentally show that our single model outperforms existing phoneme and word segmentation methods on TIMIT and Buckeye datasets. We analyze the impact of boundary threshold and when is the right time to include the segmental loss in the learning process.

A Lightweight Framework for Online Voice Activity Detection in the Wild

*Xuenan Xu*¹, *Heinrich Dinkel*², *Mengyue Wu*¹, *Kai Yu*¹; ¹*SJTU, China*; ²*Xiaomi, China* Tue-M-V-6-11, Time: 09:30

Voice activity detection (VAD) is an essential pre-processing component for speech-related tasks such as automatic speech recognition (ASR). Traditional VAD systems require strong frame-level supervision for training, inhibiting their performance in real-world test scenarios. Previously, the general-purpose VAD (GPVAD) framework has been proposed to enhance noise robustness significantly. However, GPVAD models are comparatively large and only work for offline evaluation. This work proposes the use of a knowledge distillation framework, where a (large, offline) teacher model provides frame-level supervision to a (light, online) student model. Our experiments verify that our proposed lightweight student models outperform GPVAD on all test sets, including clean, synthetic and real-world scenarios. Our smallest student model only uses 2.2% of the parameters and 15.9% duration cost of our teacher model for inference when evaluated on a Raspberry Pi.

Tue-M-V-7: Voice and Voicing

09:30-11:30, Tuesday 31 August 2021 Chairs: Hongwei Ding and Tan Lee

"See what I mean, huh?" Evaluating Visual Inspection of F₀ Tracking in Nasal Grunts

Aurélie Chlébowski, Nicolas Ballier; CLILLAC-ARP (EA 3967), France

Tue-M-V-7-1, Time: 09:30

This paper proposes to evaluate the method used in Chlébowski and Ballier [1] for the annotation of F_0 variations in nasal grunts. We discuss and test issues raised by this kind of approach exclusively based on visual inspection of the F_0 tracking in *Praat* [2]. Results tend to show that consistency in the annotation depends on acoustic features intrinsic to the grunts such as F_0 slope and duration that are sensitive to display settings. We nonetheless acknowledge the potential benefits of such a method for automation and implementation in IA and in this respect, we introduce *Prosogram* [3] as an alternative material-maker.

System Performance as a Function of Calibration Methods, Sample Size and Sampling Variability in Likelihood Ratio-Based Forensic Voice Comparison

Bruce Xiao Wang, Vincent Hughes; University of York, UK Tue-M-V-7-2, Time: 09:30

In data-driven forensic voice comparison, sample size is an issue which can have substantial effects on system output. Numerous calibration methods have been developed and some have been proposed as solutions to sample size issues. In this paper, we test four calibration methods (i.e. logistic regression, regularised logistic regression, Bayesian model, ELUB) under different conditions of sampling variability and sample size. Training and test scores were simulated from skewed distributions derived from real experiments, increasing sample sizes from 20 to 100 speakers for both the training and test sets. For each sample size, the experiments were replicated 100 times to test the susceptibility of different calibration methods to sampling variability. The $C_{\rm llr}$ mean and range across replications were used for evaluation. The Bayesian model and regularized logistic regression produced the most stable $C_{\rm llr}$ values when the sample size is small (i.e. 20 speakers), although mean $C_{\rm llr}$ is consis-

NOTES

tently lowest using logistic regression. The ELUB calibration method generally is the least preferred as it is the most sensitive to sample size and sampling variability (mean = 0.66, range = 0.21-0.59).

Voicing Assimilations by French Speakers of German in Stop-Fricative Sequences

Anne Bonneau; Loria (UMR 7503), France Tue-M-V-7-3, Time: 09:30

Voicing assimilations inside groups of obstruents occur in opposite directions in French and German, where they are respectively regressive and progressive. The aim of the study is to investigate (1) whether non native speakers (here French learners of German) are apt to acquire subtle L2 specificities like assimilation direction, although they are not aware of their very existence, or (2) whether their productions depend essentially upon other factors, in particular consonant place of articulation. To that purpose, a corpus made up of groups of obstruents (/t/ followed by /z/, /v/ or /f/) embedded into sentences has been recorded by 16 French learners of German (beginners and advanced speakers). The consonants are separated by a word or a syllable boundary. Results, derived from the analysis of consonant periodicity and duration, do not stand for an acquisition of progressive assimilation, even by advanced speakers, and do not show differences between the productions of advanced speakers and beginners. On the contrary the boundary type and the consonant place of articulation play an important role in the presence or absence of voicing inside obstruent groups. The role of phonetic, universal mechanisms against linguistic specific rules is discussed to interpret the data.

The Four-Way Classification of Stops with Voicing and Aspiration for Non-Native Speech Evaluation

*Titas Chakraborty*¹, *Vaishali Patil*², *Preeti Rao*¹; ¹*IIT Bombay, India*; ²*IIIT Pune, India* Tue-M-V-7-4, Time: 09:30

The four-way distinction of plosives in terms of voicing and aspiration is rare in the world's languages, but is an important characteristic of the Indo-Aryan language family. Both perception and production pose challenges to the language learner whose native tongue does not afford the specific distinctions. A study of the acoustic-phonetics of the sounds and their possible dependence on speaker characteristics, such as gender or native tongue, can inform methods for accurate feedback on the quality of the phones produced by a non-native learner. We present a system for the four-way classification of stops building on features previously proposed for aspiration detection in unvoiced and voiced plosives. Trained on an available dataset of Hindi speech by native speakers, the system works reliably on production data comprising Bangla words uttered by native Bangla and non-native (American English L1) speakers. The latter display a variety of articulation patterns for the given target contrasts, providing useful insights related to L1 influence on the voicing-aspiration production in word-initial CV contexts.

Acoustic and Prosodic Correlates of Emotions in Urdu Speech

Saba Urooj¹, Benazir Mumtaz², Sarmad Hussain¹, Ehsan ul Haq¹; ¹UET Lahore, Pakistan; ²Universität Konstanz, Germany Tue-M-V-7-5, Time: 09:30

Emotional speech corpora exhibit differences in duration, intensity and fundamental frequency. We investigated acoustic as well as prosodic correlates of emotional speech in Urdu. We recorded a corpus of 23 sentences from four speakers of Urdu covering four emotional states. Main results show that: a) sadness exhibits lowest utterance rate, lowest intensity and narrow pitch range, b) anger exhibits highest utterance rate, highest intensity and wider pitch range, and c) happiness exhibits higher utterance rate and wider pitch range as compared to neutral and sadness; but no significant differences are found between the intensity and pitch range of anger and happiness. The analysis also shows differences in terms of pitch or phrase accents and boundary tones.

Voicing Contrasts in the Singleton Stops of Palestinian Arabic: Production and Perception

Nour Tamim, Silke Hamann; University of Amsterdam, The Netherlands

Tue-M-V-7-6, Time: 09:30

This study investigates the stop voicing contrast in Palestinian Arabic (PA) by examining Voice Onset Time (VOT) in both production and perception. An acoustic analysis of the recordings of 8 speakers showed that word-initial voiced stops in sentence context have an average VOT of -93 msec, and word-initial voiceless stops one of 29 msec. PA thus belongs, like most dialects of Arabic, to true voicing languages, i.e., languages with a contrast between voicing lead and short lag VOT.

We furthermore tested whether the phoneme /b/, without voiceless counterpart /p/ in PA, has similar VOT values to /d, d^{ς} /, which have voiceless counterparts /t, t^{ς} /. Similarly, we compared /k/, without counterpart /g/ in the PA dialect we investigated, to /t, t^{ς} /. For /b/ we found very similar VOT values to /d, d^{ς} /, while for /k/ we found a difference to /t, t^{ς} /, attributable to a general tendency of velars to have longer VOT than denti-alveolars. We thus found no evidence for a less contrastive realization of unpaired plosives in PA.

In a categorization experiment of the denti-alveolar phoneme pairs with the same 8 speakers, VOT proved sufficient as a perceptual cue, though f0 of the following vowel also influenced the categorization.

A Comparison of the Accuracy of Dissen and Keshet's (2016) DeepFormants and Traditional LPC Methods for Semi-Automatic Speaker Recognition

Thomas Coy, Vincent Hughes, Philip Harrison, Amelia J. Gully; University of York, UK Tue-M-V-7-7, Time: 09:30

There is a growing trend in the field of forensic speech science towards integrating the vanguard of speech technology with traditional linguistic methods in pursuit of both scalable (i.e. automatable) and accurate evidential methods. To this end, this paper investigates DeepFormants, a DNN formant estimator which its creators, Dissen and Keshet [1], claim constitutes an accurate tool ready for use by linguists. In the present paper, DeepFormants is integrated into semiautomatic speaker recognition systems using long-term formant distributions and compared against systems using traditional linear predictive coding. The readiness of the tool is assessed on overall speaker recognition performance, measured using equal error rates (EER) and the log LR cost functions (C_{llr}). In high-quality conditions, DeepFormants outperforms the best performing LPC systems. Much poorer overall performance is found in channel mismatch conditions for DeepFormants, suggesting it is not adaptable to conditions it was not originally trained on. However, this is also true of LPC methods, raising questions over the validity of using formant analysis at all in such cases. A major benefit of DeepFormants over LPC is that the analyst does not need to specify settings. We discuss the implications of this with regard to results for individual speakers.

MAP Adaptation Characteristics in Forensic Long-Term Formant Analysis

Michael Jessen; Bundeskriminalamt, Germany Tue-M-V-7-8, Time: 09:30

Forensic data from long-term formant analysis were used as input to the GMM-UBM approach, which is a way of deriving Likelihood Ratios. Tests were performed running 22 same-speaker comparisons and 462 different-speaker comparisons from a corpus of anonymized casework data involving telephone-intercepted speech. In a first series of tests, the number of Gaussian modules for GMM-modeling was increased from 1 to 32. In a second series of tests the duration of formant input in the compared files was reduced from 10 seconds to 5 and then to 2.5. All tests were performed both without and with the use of MAP adaptation. Results were evaluated in terms of overall performance characteristics EER and Cllr and in terms of score distributions visualized as Tippett plots. The main goal of the study was to compare the use and non-use of MAP and to look at the practical forensic implications of the difference. Results show that in terms of overall performance characteristics there is little difference between the selection and de-selection of MAP. Tippett plot patterns however reveal strong differences. Application of MAP allows for more symmetric same- and different-speaker distributions and shows more robustness against duration reductions, both of which are forensically important.

Cross-Linguistic Speaker Individuality of Long-Term Formant Distributions: Phonetic and Forensic Perspectives

Justin J.H. Lo; University of York, UK Tue-M-V-7-9, Time: 09:30

This study considers issues of language- and speaker-specificity in long-term formant distributions (LTFDs) from phonetic and forensic perspectives and examines their potential value in cases of crosslanguage forensic voice comparison. Acoustic analysis of 60 male English-French bilinguals revealed systematic differences in LTFDs between the two languages, with higher LTF2-4 in French than in English. Cross-linguistic differences in the shapes of LTFDs were also found. These differences are argued to reflect not only vowel inventories of each language but also language-specific phonetic settings. At the same time, a high degree of within-speaker consistency was found across languages. Likelihood ratio based testing was carried out to examine the effect of language mismatch on the utility of LTFDs as speaker discriminants. Results showed that while the performance of LTFDs was worse in cross-language comparisons than in same-language comparisons, they were still capable of providing speaker-specific information. These findings demonstrate that, in spite of deteriorated performance, LTFDs are still potentially useful speaker discriminants in cases of language mismatch. These findings thus call for further empirical investigation into the use of linguistic-phonetic features in cross-language comparisons.

Sound Change in Spontaneous Bilingual Speech: A Corpus Study on the Cantonese n-l Merger in Cantonese-English Bilinguals

Rachel Soo, Khia A. Johnson, Molly Babel; University of British Columbia, Canada Tue-M-V-7-10, Time: 09:30

In Cantonese and several other Chinese languages, /n/ is merging with /l/. The Cantonese merger appears categorical, with /n/ becoming /l/ word-initially. This project aims to describe the status of /n/ and /l/ in bilingual Cantonese and English speech to better understand individual differences at the interface of crosslinguistic influence and sound change. We examine bilingual speech using the

SpiCE corpus, composed of speech from 34 early Cantonese-English bilinguals. Acoustic measures were collected on pre-vocalic nasal and lateral onsets in both languages. If bilinguals maintain separate representations for corresponding segments across languages, smaller differences between /n/ and /l/ are predicted in Cantonese compared to English. Measures of mid-frequency spectral tilt suggest that the /n/ and /l/ contrast is robustly maintained in English, but not Cantonese. The spacing of F2-F1 suggests small differences between Cantonese (n/ and /l/, and robust differences in English. While cross-language categories appear independent, substantial individual differences exist in the data. These data contribute to the understanding of the /n/ and /l/ merger in Cantonese and other Chinese languages, in addition to providing empirical and theoretical insights into crosslinguistic influence in early bilinguals.

Characterizing Voiced and Voiceless Nasals in Mizo

Wendy Lalhminghlui, Priyankoo Sarmah; IIT Guwahati, India

Tue-M-V-7-11, Time: 09:30

Mizo has voicing contrasts in nasals. This study investigates the acoustic properties of Mizo voiced and voiceless nasals using nasometric measurements. The dual channel data obtained for Mizo nasals is separated into oral and nasal channels and nasalance is calculated at every 10% of the duration of the nasals. Apart from that, the amount of voicing and duration of the nasals are also measured. The results show that nasalance is affected by the place of articulation of the nasals. Additionally, the voiceless nasals are found to be significantly longer than the voiced nasals.

Tue-M-SS-1: The INTERSPEECH 2021 Computational Paralinguistics Challenge (ComParE) — COVID-19 Cough, COVID-19 Speech, Escalation & Primates

Room Lacina, 09:30-11:30, Tuesday 31 August 2021 Chairs: Björn Schuller and Anton Batliner

The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19 Cough, COVID-19 Speech, Escalation & Primates

Björn W. Schuller¹, Anton Batliner², Christian Bergler³, Cecilia Mascolo⁴, Jing Han⁴, Iulia Lefter⁵, Heysem Kaya⁶, Shahin Amiriparian², Alice Baird², Lukas Stappen², Sandra Ottl², Maurice Gerczuk², Panagiotis Tzirakis¹, Chloë Brown⁴, Jagmohan Chauhan⁴, Andreas Grammenos⁴, Apinan Hasthanasombat⁴, Dimitris Spathis⁴, Tong Xia⁴, Pietro Cicuta⁴, Leon J.M. Rothkrantz⁵, Joeri A. Zwerts⁶, Jelle Treep⁶, Casper S. Kaandorp⁶; ¹Imperial College London, UK; ²Universität Augsburg, Germany; ³FAU Erlangen-Nürnberg, Germany; ⁴University of Cambridge, UK; ⁵Technische Universiteit Delft, The Netherlands; ⁶Universiteit Utrecht, The Netherlands

Tue-M-SS-1-1, Time: 09:30

The INTERSPEECH 2021 Computational Paralinguistics Challenge addresses four different problems for the first time in a research competition under well-defined conditions: In the *COVID-19 Cough* and *COVID-19 Speech* Sub-Challenges, a binary classification on COVID-19 infection has to be made based on coughing sounds and speech; in the *Escalation* Sub-Challenge, a three-way assessment of

the level of escalation in a dialogue is featured; and in the *Primates* Sub-Challenge, four species vs background need to be classified. We describe the Sub-Challenges, baseline feature extraction, and classifiers based on the 'usual' COMPARE and BoAW features as well as deep unsupervised representation learning using the AUDEEP toolkit, and deep feature extraction from pre-trained CNNs using the DEEP SPECTRUM toolkit; in addition, we add deep end-to-end sequential modelling, and partially linguistic analysis.

Transfer Learning-Based Cough Representations for Automatic Detection of COVID-19

Rubén Solera-Ureña, Catarina Botelho, Francisco Teixeira, Thomas Rolland, Alberto Abad, Isabel Trancoso; INESC-ID Lisboa, Portugal Tue-M-SS-1-2, Time: 09:40

In the last months, there has been an increasing interest in developing reliable, cost-effective, immediate and easy to use machine learning based tools that can help health care operators, institutions, companies, etc. to optimize their screening campaigns. In this line, several initiatives emerged aimed at the automatic detection of COVID-19 from speech, breathing and coughs, with inconclusive preliminary results. The ComParE 2021 COVID-19 Cough Sub-challenge provides researchers from all over the world a suitable test-bed for the evaluation and comparison of their work. In this paper, we present the INESC-ID contribution to the ComParE 2021 COVID-19 Cough Sub-challenge. We leverage transfer learning to develop a set of three expert classifiers based on deep cough representation extractors. A calibrated decision-level fusion system provides the final classification of coughs recordings as either COVID-19 positive or negative. Results show unweighted average recalls of 72.3% and 69.3% in the development and test sets, respectively. Overall, the experimental assessment shows the potential of this approach although much more research on extended respiratory sounds datasets is needed.

The Phonetic Footprint of Covid-19?

*P. Klumpp*¹, *T. Bocklet*², *T. Arias-Vergara*¹, *J.C. Vásquez-Correa*¹, *P.A. Pérez-Toro*¹, *S.P. Bayerl*², *J.R. Orozco-Arroyave*¹, *Elmar Nöth*¹; ¹*FAU Erlangen-Nürnberg, Germany*; ²*TH Nürnberg, Germany* Tue-M-SS-1-3, Time: 09:49

Against the background of the ongoing pandemic, this year's Computational Paralinguistics Challenge featured a classification problem to detect Covid-19 from speech recordings. The presented approach is based on a phonetic analysis of speech samples, thus it enabled us not only to discriminate between Covid and non-Covid samples, but also to better understand how the condition influenced an individual's speech signal.

Our deep acoustic model was trained with datasets collected exclusively from healthy speakers. It served as a tool for segmentation and feature extraction on the samples from the challenge dataset. Distinct patterns were found in the embeddings of phonetic classes that have their place of articulation deep inside the vocal tract. We observed profound differences in classification results for development and test splits, similar to the baseline method.

We concluded that, based on our phonetic findings, it was safe to assume that our classifier was able to reliably detect a pathological condition located in the respiratory tract. However, we found no evidence to claim that the system was able to discriminate between Covid-19 and other respiratory diseases.

Transfer Learning and Data Augmentation Techniques to the COVID-19 Identification Tasks in ComParE 2021

Edresson Casanova¹, Arnaldo Candido Jr.², Ricardo Corso Fernandes Jr.², Marcelo Finger¹, Lucas Rafael Stefanel Gris², Moacir Antonelli Ponti¹, Daniel Peixoto Pinto da Silva²; ¹Universidade de São Paulo, Brazil; ²Universidade Tecnológica Federal do Paraná, Brazil

Tue-M-SS-1-4, Time: 09:58

In this work, we propose several techniques to address data scarceness in ComParE 2021 COVID-19 identification tasks for the application of deep models such as Convolutional Neural Networks. Data is initially preprocessed into spectrogram or MFCC-gram formats. After preprocessing, we combine three different data augmentation techniques to be applied in model training. Then we employ transfer learning techniques from pretrained audio neural networks. Those techniques are applied to several distinct neural architectures. For COVID-19 identification in speech segments, we obtained competitive results. On the other hand, in the identification task based on cough data, we succeeded in producing a noticeable improvement on existing baselines, reaching 75.9% unweighted average recall (UAR).

Visual Transformers for Primates Classification and Covid Detection

Steffen Illium, Robert Müller, Andreas Sedlmeier, Claudia-Linnhoff Popien; LMU München, Germany Tue-M-SS-1-5, Time: 10:07

We apply the vision transformer, a deep machine learning model build around the attention mechanism, on mel-spectrogram representations of raw audio recordings. When adding mel-based data augmentation techniques and sample-weighting, we achieve comparable performance on both (PRS and CCS challenge) tasks of ComParE21, outperforming most single model baselines. We further introduce overlapping vertical patching and evaluate the influence of parameter configurations.

Deep-Learning-Based Central African Primate Species Classification with MixUp and SpecAugment

Thomas Pellegrini; IRIT (UMR 5505), France Tue-M-SS-1-6, Time: 10:16

In this paper, we report experiments in which we aim to automatically classify primate vocalizations according to four primate species of interest, plus a background category with forest sound events. We compare several standard deep neural networks architectures: standard deep convolutional neural networks (CNNs), MobileNets and ResNets. To tackle the small size of the training dataset, less than seven thousand audio files, the data augmentation techniques SpecAugment and MixUp proved to be very useful. Against the very unbalanced classes of the dataset, we used a balanced data sampler that showed to be efficient. An exponential moving average of the model weights allowed to get slight further gains. The best model was a standard 10-layer CNN, comprised of about five million parameters. It achieved a 93.6% Unweighted Average Recall (UAR) on the development set, and generalized well on the test set with a 92.5% UAR, outperforming an official baseline of 86.6%. We quantify the performance gains brought by the augmentations and training tricks, and report fusion and classification experiments based on embeddings that did not bring better results.

A Deep and Recurrent Architecture for Primate Vocalization Classification

Robert Müller, Steffen Illium, Claudia Linnhoff-Popien; LMU München, Germany Tue-M-SS-1-7, Time: 10:25

Wildlife monitoring is an essential part of most conservation efforts where one of the many building blocks is acoustic monitoring. Acoustic monitoring has the advantage of being non-invasive and applicable in areas of high vegetation. In this work, we present a deep and recurrent architecture for the classification of primate vocalizations that is based upon well proven modules such as bidirectional Long Short-Term Memory neural networks, pooling, normalized softmax and focal loss. Additionally, we apply Bayesian optimization to obtain a suitable set of hyperparameters. We test our approach on a recently published dataset of primate vocalizations that were recorded in an African wildlife sanctuary. Using an ensemble of the best five models found during hyperparameter optimization on the development set, we achieve a Unweighted Average Recall of 89.3% on the test set. Our approach outperforms the best baseline, an ensemble of various deep and shallow classifiers, which achieves a UAR of 87.5%.

Introducing a Central African Primate Vocalisation Dataset for Automated Species Classification

Joeri A. Zwerts, Jelle Treep, Casper S. Kaandorp, Floor Meewis, Amparo C. Koot, Heysem Kaya; Universiteit Utrecht, The Netherlands

Tue-M-SS-1-8, Time: 10:34

Automated classification of animal vocalisations is a potentially powerful wildlife monitoring tool. Training robust classifiers requires sizable annotated datasets, which are not easily recorded in the wild. To circumvent this problem, we recorded four primate species under semi-natural conditions in a wildlife sanctuary in Cameroon with the objective to train a classifier capable of detecting species in the wild. Here, we introduce the collected dataset, describe our approach and initial results of classifier development. To increase the efficiency of the annotation process, we condensed the recordings with an energy/change based automatic vocalisation detection. Segmenting the annotated chunks into training, validation and test sets, initial results reveal up to 82% unweighted average recall test set performance in four-class primate species classification.

Multi-Attentive Detection of the Spider Monkey Whinny in the (Actual) Wild

Georgios Rizos, Jenna Lawson, Zhuoda Han, Duncan Butler, James Rosindell, Krystian Mikolajczyk, Cristina Banks-Leite, Björn W. Schuller; Imperial College London, UK

Tue-M-SS-1-9, Time: 10:43

We study deep bioacoustic event detection through multi-head attention based pooling, exemplified by wildlife monitoring. In the multiple instance learning framework, a core deep neural network learns a projection of the input acoustic signal into a sequence of embeddings, each representing a segment of the input. Sequence pooling is then required to aggregate the information present in the sequence such that we have a single clip-wise representation. We propose an improvement based on Squeeze-and-Excitation mechanisms upon a recently proposed audio tagging ResNet, and show that it performs significantly better than the baseline, as well as a collection of other recent audio models. We then further enhance our model, by performing an extensive comparative study of recent sequence pooling mechanisms, and achieve our best result using multi-head self-attention followed by concatenation of the head-specific pooled

embeddings — better than prediction pooling methods, as well as compared to other recent sequence pooling tricks. We perform these experiments on a novel dataset of spider monkey whinny calls we introduce here, recorded in a rainforest in the South-Pacific coast of Costa Rica, with a promising outlook pertaining to minimally invasive wildlife monitoring.

Identifying Conflict Escalation and Primates by Using Ensemble X-Vectors and Fisher Vector Features

José Vicente Egas-López, Mercedes Vetráb, László Tóth, Gábor Gosztolya; University of Szeged, Hungary Tue-M-SS-1-10, Time: 10:52

Computational paralinguistics is concerned with the automatic identification of non-verbal information in human speech. The Interspeech ComParE challenge features new paralinguistic tasks each year; this time, among others, a cross-corpus conflict escalation task and the identification of primates based solely on audio are the actual problems set. In our entry to ComParE 2021, we utilize x-vectors and Fisher vectors as features. To improve the robustness of the predictions, we also experiment with building an ensemble of classifiers from the x-vectors. Lastly, we exploit the fact that the Escalation Sub-Challenge is a conflict detection task, and incorporate the SSPNet Conflict Corpus in our training workflow. Using these approaches, at the time of writing, we had already surpassed the official Challenge baselines on both tasks, which demonstrates the efficiency of the employed techniques.

Ensemble-Within-Ensemble Classification for Escalation Prediction from Speech

Oxana Verkholyak¹, Denis Dresvyanskiy², Anastasia Dvoynikova¹, Denis Kotov², Elena Ryumina¹, Alena Velichko¹, Danila Mamontov², Wolfgang Minker², Alexey Karpov¹; ¹RAS, Russia; ²Universität Ulm, Germany

Tue-M-SS-1-11, Time: 11:01

Conflict situations arise frequently in our daily life and often require timely response to resolve the issues. In order to automatically classify conflict (also referred to as escalation) speech utterances we propose ensemble learning as it improves prediction performance by combining several heterogeneous models that compensate for each other's weaknesses. However, the effectiveness of the classification ensemble greatly depends on its constituents and their fusion strategy. This paper provides experimental evidence for effectiveness of different prediction-level fusion strategies and demonstrates the performance of each proposed ensemble on the Escalation Sub-Challenge (ESS) in the framework of the Computational Paralinguistics Challenge (ComParE-2021). The ensembles comprise various machine learning approaches based on acoustic and linguistic characteristics of speech. The training strategy is specifically designed to increase the generalization performance on the unseen data, while the diverse nature of ensemble candidates ensures high prediction power and accurate classification.

Analysis by Synthesis: Using an Expressive TTS Model as Feature Extractor for Paralinguistic Speech Classification

Dominik Schiller¹, Silvan Mertes¹, Pol van Rijn², Elisabeth André¹; ¹Universität Augsburg, Germany; ²MPI for Empirical Aesthetics, Germany Tue-M-SS-1-12, Time: 11:10

Modeling adequate features of speech prosody is one key factor to good performance in affective speech classification. However, the distinction between the prosody that is induced by 'how' something is said (i.e., affective prosody) and the prosody that is induced by 'what' is being said (i.e., linguistic prosody) is neglected in state-ofthe-art feature extraction systems. This results in high variability of the calculated feature values for different sentences that are spoken with the same affective intent, which might negatively impact the performance of the classification. While this distinction between different prosody types is mostly neglected in affective speech recognition, it is explicitly modeled in expressive speech synthesis to create controlled prosodic variation. In this work, we use the expressive Text-To-Speech model Global Style Token Tacotron to extract features for a speech analysis task. We show that the learned prosodic representations outperform state-of-the-art feature extraction systems in the exemplary use case of Escalation Level Classification.

Discussion

Time: 11:19

Tue-Survey: Survey Talk 1: Heidi Christensen Room A+B, 11:30-12:30, Tuesday 31 August 2021 Chairs: TBD

Towards Automatic Speech Recognition for People with Atypical Speech

Heidi Christensen; University of Sheffield, UK Tue-Survey, Time: 11:30

In the last decade we have seen how speech technologies for typical speech have matured and thus enabled the advancement of a multitude of services and technologies including voice-enabled conversational interfaces, dictation and successfully underpinning the use of state-of-the-art NLP techniques. This ever more pervasive offering allows for an often far more convenient and natural way of interacting with machines and systems. However it also represents an ever-growing gap experienced by people with atypical (dysarthric) voices: people with even just mild-to-moderate speech disorders cannot achieve satisfactory performance with current automatic speech recognition (ASR) systems and hence they are falling further and further behind in terms of their ability to use modern devices and interfaces. This talk will present the major challenges in porting mainstream ASR methodologies to work for atypical speech, discuss recent advances and present thoughts on where the research effort should be focusing to have real impact in this community of potential users. Being able to speak a query or dictate an email offers a lot of convenience to most of us but for this group of people can have significant implications on ability to fully take part in society and life quality.

Tue-A-O-1: Embedding and Network Architecture for Speaker Recognition

Room A+B, 13:30–15:30, Tuesday 31 August 2021 Chairs: Oldřich Plchot and Themos Stafylakis

Leveraging Speaker Attribute Information Using Multi Task Learning for Speaker Verification and Diarization

Chau Luu, Peter Bell, Steve Renals; University of Edinburgh, UK Tue-A-O-1-1, Time: 13:30

Deep speaker embeddings have become the leading method for encoding speaker identity in speaker recognition tasks. The embedding space should ideally capture the variations between all possible speakers, encoding the multiple acoustic aspects that make up a speaker's identity, whilst being robust to non-speaker acoustic variation. Deep speaker embeddings are normally trained discriminatively, predicting speaker identity labels on the training data. We hypothesise that additionally predicting speaker-related auxiliary variables — such as age and nationality — may yield representations that are better able to generalise to unseen speakers. We propose a framework for making use of auxiliary label information, even when it is only available for speech corpora mismatched to the target application. On a test set of US Supreme Court recordings, we show that by leveraging two additional forms of speaker attribute information derived respectively from the matched training data, and VoxCeleb corpus, we improve the performance of our deep speaker embeddings for both verification and diarization tasks, achieving a relative improvement of 26.2% in DER and 6.7% in EER compared to baselines using speaker labels only. This improvement is obtained despite the auxiliary labels having been scraped from the web and being potentially noisy.

Spine2Net: SpineNet with Res2Net and Time-Squeeze-and-Excitation Blocks for Speaker Recognition

Magdalena Rybicka¹, Jesús Villalba², Piotr Żelasko², Najim Dehak², Konrad Kowalczyk¹; ¹AGH UST, Poland; ²Johns Hopkins University, USA Tue-A-Q-1-2, Time: 13:50

Modeling speaker embeddings using deep neural networks is currently state-of-the-art in speaker recognition. Recently, ResNet-based structures have gained a broader interest, slowly becoming the baseline along with the deep-rooted Time Delay Neural Network based models. However, the scale-decreased design of the ResNet models may not preserve all of the speaker information. In this paper, we investigate the SpineNet structure with scale-permuted design to tackle this problem, in which feature size either increases or decreases depending on the processing stage in the network. Apart from the presented adjustments of the SpineNet model for the speaker recognition task, we also incorporate popular modules dedicated to the residual-like structures, namely the Res2Net and Squeeze-and-Excitation blocks, and modify them to work effectively in the presented neural network architectures. The final proposed model, i.e., the SpineNet architecture with Res2Net and Time-Squeeze-and-Excitation blocks, achieves remarkable Equal Error Rates of 0.99 and 0.92 for the Extended and Original trial lists of the well-known VoxCeleb1 dataset.

Speaker Embeddings by Modeling Channel-Wise Correlations

*Themos Stafylakis*¹, *Johan Rohdin*¹, *Lukáš Burget*²; ¹*Omilia, Greece;* ²*Brno University of Technology,*

Czechia

Tue-A-O-1-3, Time: 14:10

Speaker embeddings extracted with deep 2D convolutional neural networks are typically modeled as projections of first and second order statistics of channel-frequency pairs onto a linear layer, using either average or attentive pooling along the time axis. In this paper we examine an alternative pooling method, where pairwise correlations between channels for given frequencies are used as statistics. The method is inspired by style-transfer methods in computer vision, where the style of an image, modeled by the matrix of channel-wise correlations, is transferred to another image, in order to produce a new image having the style of the first and the content of the second. By drawing analogies between image style and speaker characteristics, and between image content and phonetic sequence, we explore the use of such channel-wise correlations features to train a ResNet architecture in an end-to-end fashion. Our experiments on VoxCeleb demonstrate the effectiveness of the proposed pooling method in speaker recognition.

Multi-Task Neural Network for Robust Multiple Speaker Embedding Extraction

Weipeng He, Petr Motlicek, Jean-Marc Odobez; Idiap Research Institute, Switzerland Tue-A-O-1-4, Time: 14:30

This paper introduces a novel approach for extracting speaker embeddings from audio mixtures of multiple overlapping voices. This approach is based on a multi-task neural network. The network first extracts a latent feature for each direction. This feature is used for detecting sound sources as well as identifying speakers. In contrast to traditional approaches, the proposed method does not rely on explicit sound source separation. The neural network model learns from data to extract the most suitable features of the sounds at different directions. The experiments using audio recordings of overlapping sound sources show that the proposed approach outperforms a beamforming-based traditional method.

ICSpk: Interpretable Complex Speaker Embedding Extractor from Raw Waveform

Junyi Peng¹, Xiaoyang Qu¹, Jianzong Wang¹, Rongzhi Gu², Jing Xiao¹, Lukáš Burget³, Jan Černocký³; ¹Ping An Technology, China; ²Peking University, China; ³Brno University of Technology, Czechia Tue-A-O-1-5, Time: 14:50

Recently, extracting speaker embedding directly from raw waveform has drawn increasing attention in the field of speaker verification. Parametric real-valued filters in the first convolutional layer are learned to transform the waveform into time-frequency representations. However, these methods only focus on the magnitude spectrum and the poor interpretability of the learned filters limits the performance. In this paper, we propose a complex speaker embedding extractor, named ICSpk, with higher interpretability and fewer parameters. Specifically, at first, to quantify the speaker-related frequency response of waveform, we modify the original short-term Fourier transform filters into a family of complex exponential filters, named interpretable complex (IC) filters. Each IC filter is confined by a complex exponential filter parameterized by frequency. Then, a deep complex-valued speaker embedding extractor is designed to operate on the complex-valued output of IC filters. The proposed

ICSpk is evaluated on VoxCeleb and CNCeleb databases. Experimental results demonstrate the IC filters-based system exhibits a significant improvement over the complex spectrogram based systems. Furthermore, the proposed ICSpk outperforms existing raw waveform based systems by a large margin.

Tue-A-O-2: Speech Perception I

Room D, 13:30–15:30, Tuesday 31 August 2021 Chairs: Georgia Zellou and Josiane Riverin-Coutlée

Prosodic Disambiguation Using Chironomic Stylization of Intonation with Native and Non-Native Speakers

Xiao Xiao¹, Nicolas Audibert¹, Grégoire Locqueville², Christophe d'Alessandro², Barbara Kuhnert¹, Claire Pillot-Loiseau¹; ¹LPP (UMR 7018), France; ²∂'Alembert (UMR 7190), France Tue-A-O-2-1, Time: 13:30

This paper introduces an interface that enables the real-time gestural control of intonation in phrases produced by a vocal synthesizer. The melody and timing of a target phrase can be modified by tracing melodic contours on the touch-screen of a mobile tablet. Envisioning this interface as a means for non-native speakers to practice the intonation of a foreign language, we present a pilot study where native and non-native speakers imitated the pronunciation of French phrases using their voice and the interface, with a visual guide and without. Comparison of resulting F0 curves against the reference contour and a preliminary perceptual assessment of synthesized utterances suggest that for both non-native and native speakers, imitation with the help of a visual guide is comparable in accuracy to vocal imitation, and that timing control was a source of difficulty.

Variation in Perceptual Sensitivity and Compensation for Coarticulation Across Adult and Child Naturally-Produced and TTS Voices

Aleese Block, Michelle Cohn, Georgia Zellou; University of California at Davis, USA Tue-A-0-2-2, Time: 13:50

The current study explores whether perception of coarticulatory vowel nasalization differs by speaker age (adult vs. child) and type of voice (naturally produced vs. synthetic speech). Listeners completed a 4IAX discrimination task between pairs containing acoustically identical (both nasal or oral) vowels and acoustically distinct (one oral, one nasal) vowels. Vowels occurred in either the same consonant contexts or different contexts across pairs. Listeners completed the experiment with either naturally produced speech or text-to-speech (TTS). For same-context trials, listeners were better at discriminating between oral and nasal vowels for child speech in the synthetic voices but adult speech in the natural voices. Meanwhile, in different-context trials, listeners were less able to discriminate, indicating more perceptual compensation for synthetic voices. There was no difference in different-context discrimination across talker ages, indicating that listeners did not compensate differently if the speaker was a child or adult. Findings are relevant for models of compensation, computer personification theories, and speaker-indexical perception accounts.

Extracting Different Levels of Speech Information from EEG Using an LSTM-Based Model

Mohammad Jalilpour Monesi, Bernd Accou, Tom Francart, Hugo Van hamme; KU Leuven, Belgium Tue-A-O-2-3, Time: 14:10

Decoding the speech signal that a person is listening to from the human brain via electroencephalography (EEG) can help us understand how our auditory system works. Linear models have been used to reconstruct the EEG from speech or vice versa. Recently, Artificial Neural Networks (ANNs) such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) based architectures have outperformed linear models in modeling the relation between EEG and speech. Before attempting to use these models in real-world applications such as hearing tests or (second) language comprehension assessment we need to know what level of speech information is being utilized by these models. In this study, we aim to analyze the performance of an LSTM-based model using different levels of speech features. The task of the model is to determine which of two given speech segments is matched with the recorded EEG. We used low- and high-level speech features including: envelope, mel spectrogram, voice activity, phoneme identity, and word embedding. Our results suggest that the model exploits information about silences, intensity, and broad phonetic classes from the EEG. Furthermore, the mel spectrogram, which contains all this information, yields the highest accuracy (84%) among all the features.

Word Competition: An Entropy-Based Approach in the DIANA Model of Human Word Comprehension

Louis ten Bosch, Lou Boves; Radboud Universiteit, The Netherlands

Tue-A-O-2-4, Time: 14:30

We discuss the role of entropy of the set of unfolding word candidates in the context of DIANA, a computational model of human auditory speech comprehension. DIANA consists of three major interacting components: Activation, Decision and Execution. The Activation component computes activations of word candidates that change over time as a function of the unfolding audio input. The resulting set of word candidate activations can be associated with an entropy that is related to difficulty of the decision when one of these candidates must be selected at time T. The paper presents the close relation between entropy measures and the between-word competition during the unfolding of the auditory stimuli, and at the end of the stimuli if no decision could be made before stimulus offset. We present a way for computing the entropy that takes into account linguistic-phonetic constraints that play a role in speech comprehension and in lexical decision experiments. Using the BALDEY data set and linear mixed effects regression models for RT, we show that entropy measures explain differences between RTs of words with different morphological structure.

Time-to-Event Models for Analyzing Reaction Time Sequences

Louis ten Bosch, Lou Boves; Radboud Universiteit, The Netherlands

Tue-A-O-2-5, Time: 14:50

We investigate reaction time (RT) sequences obtained from lexical decision experiments by applying Time-to-Event modelling (Survival Analysis). This is a branch of statistics for analyzing the expected duration until one or more events happen, associated with a set of potential 'causes' (in our case the decision for a 'word' judgment as a function of conventional predictors such as lexical frequency, stimulus duration, reduction, etc.). In this analysis, RTs are considered a

by-product of an (unobservable) cumulative incidence function that results in a decision when it exceeds a certain threshold.

We show that Survival Analysis can be effectively used to narrow the gap between data-oriented models and process-oriented models for RT data from lexical decision experiments. Results of this analysis technique are presented for two different RT data sets. The analysis reveals time-varying patterns of predictors that reflect the differences in cognitive processes during the presentation of auditory stimuli.

Models of Reaction Times in Auditory Lexical Decision: RTonset versus RToffset

Sophie Brand¹, Kimberley Mulder², Louis ten Bosch³, Lou Boves³; ¹Zuyd Hogeschool, The Netherlands; ²Universiteit Utrecht, The Netherlands; ³Radboud Universiteit, The Netherlands Tue-A-O-2-6, Time: 15:10

We investigate how the role of predictors in models of reaction times in auditory lexical decision experiments depends on the operational definition of RT: whether the time is measured from stimulus onset or from stimulus offset. In a large body of literature, RTs are measured from the onset of the stimulus to the start of the response (often a button press or an oral response). The rationale behind this choice is that information about the stimulus becomes available to the listener starting at onset. Alternatively, the RT from offset is less dependent on stimulus duration and is assumed to focus on those cognitive processes that play a role late(r) in the word and after word offset, when all information is available.

The paper presents RT-onset and RT-offset-based linear mixed effects models for three different lexical decision-based data sets and explains the significant differences between these models, showing to what extent both definitions of reaction time reveal different roles for predictors and how early and later contributions to the overall RT can be differentiated.

Tue-A-V-1: Acoustic Event Detection and Acoustic Scene Classification

13:30–15:30, Tuesday 31 August 2021 Chairs: Thomas Pellegrini and Yaniv Zigel

SpecMix : A Mixed Sample Data Augmentation Method for Training with Time-Frequency Domain Features

*Gwantae Kim*¹, *David K. Han*², *Hanseok Ko*¹; ¹*Korea University, Korea*; ²*Drexel University, USA* Tue-A-V-1-1, Time: 13:30

A mixed sample data augmentation strategy is proposed to enhance the performance of models on audio scene classification, sound event classification, and speech enhancement tasks. While there have been several augmentation methods shown to be effective in improving image classification performance, their efficacy toward time-frequency domain features of audio is not assured. We propose a novel audio data augmentation approach named "Specinix" specifically designed for dealing with time-frequency domain features. The augmentation method consists of mixing two different data samples by applying time-frequency masks effective in preserving the spectral correlation of each audio sample. Our experiments on acoustic scene classification, sound event classification, and speech enhancement tasks show that the proposed Specmix improves the performance of various neural network architectures by a maximum of 2.7%.

SpecAugment++: A Hidden Space Data Augmentation Method for Acoustic Scene Classification

*Helin Wang*¹, *Yuexian Zou*¹, *Wenwu Wang*²; ¹*Peking University, China*; ²*University of Surrey, UK* Tue-A-V-1-2, Time: 13:30

In this paper, we present SpecAugment++, a novel data augmentation method for deep neural networks based acoustic scene classification (ASC). Different from other popular data augmentation methods such as SpecAugment and mixup that only work on the input space, SpecAugment++ is applied to both the input space and the hidden space of the deep neural networks to enhance the input and the intermediate feature representations. For an intermediate hidden state, the augmentation techniques consist of masking blocks of frequency channels and masking blocks of time frames, which improve generalization by enabling a model to attend not only to the most discriminative parts of the feature, but also the entire parts. Apart from using zeros for masking, we also examine two approaches for masking based on the use of other samples within the mini-batch, which helps introduce noises to the networks to make them more discriminative for classification. The experimental results on the DCASE 2018 Task1 dataset and DCASE 2019 Task1 dataset show that our proposed method can obtain 3.6% and 4.7% accuracy gains over a strong baseline without augmentation (i.e. CP-ResNet) respectively, and outperforms other previous data augmentation methods.

An Effective Mutual Mean Teaching Based Domain Adaptation Method for Sound Event Detection

*Xu Zheng*¹, *Yan Song*¹, *Li-Rong Dai*¹, *Ian McLoughlin*¹, *Lin Liu*²; ¹*USTC, China*; ²*iFLYTEK, China* Tue-A-V-1-3, Time: 13:30

In this paper, we present a novel mutual mean teaching based domain adaptation (MMT-DA) method for sound event detection (SED) task, which can effectively exploit synthetic data to improve the SED performance. Existing methods simply treat the synthetic data as strongly-labeled data in semi-supervised learning (SSL) framework. Benefiting from the strong labels of synthetic data, superior SED performance can be achieved. However, a distribution mismatch between synthetic and real data raises an evident challenge for domain adaptation (DA). In MMT-DA, convolutional recurrent neural networks (CRNN) learned from different datasets (i.e. total data:real+synthetic, and real data) are exploited for DA. Specifically, mean teacher method using CRNN is employed for utilizing the unlabeled real data. To compensate the domain diversity, an additional domain classifier with gradient reverse layer(GRL) is used for training a mean teacher for total data. The student CRNNs are mutually taught using the soft predictions of unlabeled data obtained from different teachers. Furthermore, a strip pooling based attention module is exploited to model the inter-dependencies between channels and time-frequency dimensions to exploit the structure information. Experimental results on Task4 of DCASE2020 demonstrate the ability of the proposed method, achieving 52.0% F1-score on the validation dataset, which outperforms the winning system's 50.6%.

Acoustic Scene Classification Using Kervolution-Based SubSpectralNet

Ritika Nandi, Shashank Shekhar, Manjunath Mulimani; MAHE, India

<u>Tue-A</u>-V-1-5, Time: 13:30

In this paper, a Kervolution-based SubSpectralNet model is proposed for Acoustic Scene Classification (ASC). SubSpectralNet is a competitive model which divides the mel spectrogram into horizontal slices termed as sub-spectrograms that are considered as input to

the Convolutional Neural Network (CNN). In this work, the linear convolutional operation of SubSpectralNet is replaced with a non-linear operation using the kernel trick. This is also known as kervolution (kernel convolution)-based SubSpectralNet. The performance of the proposed methodology is evaluated on the DCASE (Detection and Classification of Acoustic Scenes and Events) 2018 development dataset. The proposed method achieves 73.52% and 75.76% accuracy with Polynomial and Gaussian Kernels respectively.

Event Specific Attention for Polyphonic Sound Event Detection

Harshavardhan Sundar, Ming Sun, Chao Wang; Amazon, USA

Tue-A-V-1-6, Time: 13:30

The concept of multi-headed self attention (MHSA) introduced as a critical building block of a Transformer Encoder/Decoder Module has made a significant impact in the areas of natural language processing (NLP), automatic speech recognition (ASR) and recently in the area of sound event detection (SED). The current state-of-the-art approaches to SED employ a shared attention mechanism achieved through a stack of MHSA blocks to detect multiple sound events. Consequently, in a multi-label SED task, a common attention mechanism would be responsible for generating relevant feature representations for each of the events to be detected. In this paper, we show through empirical evaluation that having more MHSA blocks dedicated specifically for individual events, rather than having a stack of shared MHSA blocks, improves the overall detection performance. Interestingly, this improvement in performance comes about because the event-specific attention blocks help in resolving confusions in the case of co-occurring events. The proposed "Event-specific Attention Network" (ESA-Net) can be trained in an end-to-end manner. On the DCASE 2020 Task 4 data set, we show that with ESA-Net, the best single model achieves an event-based F1 score of 52.1% on the public validation data set improving over the existing state of the art result.

AST: Audio Spectrogram Transformer

Yuan Gong, Yu-An Chung, James Glass; MIT, USA Tue-A-V-1-7, Time: 13:30

In the past decade, convolutional neural networks (CNNs) have been widely adopted as the main building block for end-to-end audio classification models, which aim to learn a direct mapping from audio spectrograms to corresponding labels. To better capture long-range global context, a recent trend is to add a self-attention mechanism on top of the CNN, forming a CNN-attention hybrid model. However, it is unclear whether the reliance on a CNN is necessary, and if neural networks purely based on attention are sufficient to obtain good performance in audio classification. In this paper, we answer the question by introducing the *Audio Spectrogram Transformer* (AST), the first convolution-free, purely attention-based model for audio classification. We evaluate AST on various audio classification benchmarks, where it achieves new state-of-the-art results of 0.485 mAP on AudioSet, 95.6% accuracy on ESC-50, and 98.1% accuracy on Speech Commands V2.

Shallow Convolution-Augmented Transformer with Differentiable Neural Computer for Low-Complexity Classification of Variable-Length Acoustic Scene

Soonshin Seo, Donghyun Lee, Ji-Hwan Kim; Sogang University, Korea

Tue-A-V-1-8, Time: 13:30

Convolutional neural networks (CNNs) exhibit good performance in low-complexity classification with fixed-length acoustic scenes. How-

ever, previous studies have not considered variable-length acoustic scenes in which performance degradation is prevalent. In this regard, we investigate two novel architectures - convolution-augmented transformer (Conformer) and differentiable neural computer (DNC). Both the models show desirable performance for variable-length data but require a large amount of data. In other words, small amounts of data, such as those from acoustic scenes, lead to overfitting in these models. In this paper, we propose a shallow convolutionaugmented Transformer with a differentiable neural computer (shallow Conformer-DNC) for the low-complexity classification of variable-length acoustic scenes. The shallow Conformer-DNC is enabled to converge with small amounts of data. Short-term and long-term contexts of variable-length acoustic scenes are trained by using the shallow Conformer and shallow DNC, respectively. The experiments were conducted for variable-length conditions using the TAU Urban Acoustic Scenes 2020 Mobile dataset. As a result, a peak accuracy of 61.25% was confirmed for shallow Conformer-DNC with a model parameter of 34 K. It is comparable performance to state-of-the-art CNNs.

An Evaluation of Data Augmentation Methods for Sound Scene Geotagging

Helen L. Bear, Veronica Morfi, Emmanouil Benetos; Queen Mary University of London, UK Tue-A-V-1-9, Time: 13:30

Sound scene geotagging is a new topic of research which has evolved from acoustic scene classification. It is motivated by the idea of audio surveillance. Not content with only describing a scene in a recording, a machine which can locate where the recording was captured would be of use to many. In this paper we explore a series of common audio data augmentation methods to evaluate which best improves the accuracy of audio geotagging classifiers.

Our work improves on the state-of-the-art city geotagging method by 23% in terms of classification accuracy.

Optimizing Latency for Online Video Captioning Using Audio-Visual Transformers

Chiori Hori, Takaaki Hori, Jonathan Le Roux; MERL, USA Tue-A-V-1-10, Time: 13:30

Video captioning is an essential technology to understand scenes and describe events in natural language. To apply it to real-time monitoring, a system needs not only to describe events accurately but also to produce the captions as soon as possible. Low-latency captioning is needed to realize such functionality, but this research area for online video captioning has not been pursued yet. This paper proposes a novel approach to optimize each caption's output timing based on a trade-off between latency and caption quality. An audio-visual Transformer is trained to generate ground-truth captions using only a small portion of all video frames, and to mimic outputs of a pre-trained Transformer to which all the frames are given. A CNN-based timing detector is also trained to detect a proper output timing, where the captions generated by the two Transformers become sufficiently close to each other. With the jointly trained Transformer and timing detector, a caption can be generated in the early stages of an event-triggered video clip, as soon as an event happens or when it can be forecasted. Experiments with the ActivityNet Captions dataset show that our approach achieves 94% of the caption quality of the upper bound given by the pre-trained Transformer using the entire video clips, using only 28% of frames from the beginning.

Variational Information Bottleneck for Effective Low-Resource Audio Classification

Shijing Si¹, Jianzong Wang¹, Huiming Sun¹, Jianhan Wu², Chuanyao Zhang², Xiaoyang Qu¹, Ning Cheng¹, Lei Chen³, Jing Xiao¹; ¹Ping An Technology, China; ²USTC, China; ³HKUST, China Tue-A-V-1-11, Time: 13:30

Large-scale deep neural networks (DNNs) such as convolutional neural networks (CNNs) have achieved impressive performance in audio classification for their powerful capacity and strong generalization ability. However, when training a DNN model on low-resource tasks, it is usually prone to overfitting the small data and learning too much redundant information. To address this issue, we propose to use variational information bottleneck (VIB) to mitigate overfitting and suppress irrelevant information. In this work, we conduct experiments on a 4-layer CNN. However, the VIB framework is ready-to-use and could be easily utilized with many other state-of-the-art network architectures. Evaluation on a few audio datasets shows that our approach significantly outperforms baseline methods, yielding $\geq 5.0\%$ improvement in terms of classification accuracy in some low-source settings.

Improving Weakly Supervised Sound Event Detection with Self-Supervised Auxiliary Tasks

Soham Deshmukh¹, Bhiksha Raj², Rita Singh²; ¹Microsoft, USA; ²Carnegie Mellon University, USA Tue-A-V-1-12, Time: 13:30

While multitask and transfer learning has shown to improve the performance of neural networks in limited data settings, they require pretraining of the model on large datasets beforehand. In this paper, we focus on improving the performance of weakly supervised sound event detection in low data and noisy settings simultaneously without requiring any pretraining task. To that extent, we propose a shared encoder architecture with sound event detection as a primary task and an additional secondary decoder for a self-supervised auxiliary task. We empirically evaluate the proposed framework for weakly supervised sound event detection on a remix dataset of the DCASE 2019 task 1 acoustic scene data with DCASE 2018 Task 2 sounds event data under 0, 10 and 20 dB SNR. To ensure we retain the localisation information of multiple sound events, we propose a two-step attention pooling mechanism that provides a time-frequency localisation of multiple audio events in the clip. The proposed framework with two-step attention outperforms existing benchmark models by 22.3%, 12.8%, 5.9% on 0, 10 and 20 dB SNR respectively. We carry out an ablation study to determine the contribution of the auxiliary task and two-step attention pooling to the SED performance improvement.

Acoustic Event Detection with Classifier Chains

Tatsuya Komatsu¹, Shinji Watanabe², Koichi Miyazaki³, Tomoki Hayashi³; ¹LINE, Japan; ²Carnegie Mellon University, USA; ³Nagoya University, Japan Tue-A-V-1-13, Time: 13:30

This paper proposes acoustic event detection (AED) with classifier chains, a new classifier based on the probabilistic chain rule. The proposed AED with classifier chains consists of a gated recurrent unit and performs iterative binary detection of each event one by one. In each iteration, the event's activity is estimated and used to condition the next output based on the probabilistic chain rule to form classifier chains. Therefore, the proposed method can handle the interdependence among events upon classification, while the conventional AED methods with multiple binary classifiers with a linear

NOTES

layer and sigmoid function have placed an assumption of conditional independence. In the experiments with a real-recording dataset, the proposed method demonstrates its superior AED performance to a relative 14.80% improvement compared to a convolutional recurrent neural network baseline system with the multiple binary classifiers.

Tue-A-V-2: Diverse Modes of Speech Acquisition and Processing

13:30–15:30, Tuesday 31 August 2021 Chairs: Milos Cernak and Vikramjit Mitra

Segment and Tone Production in Continuous Speech of Hearing and Hearing-Impaired Children

*Shu-Chuan Tseng*¹, *Yi-Fen Liu*²; ¹*Academia Sinica, Taiwan*; ²*Feng Chia University, Taiwan* Tue-A-V-2-1, Time: 13:30

Verbal communication in daily use is conducted in the form of continuous speech that theoretically is the ideal data format for assessing oral language ability in educational and clinical domains. But as phonetic reduction and particularly lexical tones in Chinese are greatly affected by discourse context, it is a challenging task for automatic systems to evaluate continuous speech only by acoustic features. This study analyzed repetitive and storytelling speech produced by selected Chinese-speaking hearing and hearing-impaired children with distinctively high and low speech intelligibility levels. Word-based reduction types are derived by phonological properties that characterize contraction degrees of automatically generated surface forms of disyllabic words. F0-based tonal contours are visualized using the centroid-nearest data points in the major clusters computed for tonal syllables. Our results show that primary speech characteristics across different groups of children can be differentiated by means of reduction type and tone production.

Effect of Carrier Bandwidth on Understanding Mandarin Sentences in Simulated Electric-Acoustic Hearing

*Feng Wang*¹, *Jing Chen*², *Fei Chen*¹; ¹*SUSTech, China*; ²*Peking University, China* Tue-A-V-2-2, Time: 13:30

For patients suffering with high-frequency hearing loss and preserving low-frequency hearing, combined electric-acoustic stimulation (EAS) may significantly improve their speech perception compared with cochlear implants (CIs). In combined EAS, a hearing aid provides low-frequency information via acoustic (A) stimulation and a CI evokes high-frequency sound sensation via electrical (E) stimulation. The present work investigated the EAS advantage when only a small number (i.e., 1 or 2) of channels were provided for electrical stimulation in a CI, and the effect of carrier bandwidth on understanding Mandarin sentences in a simulation of combined EAS experiment. The A-portion was extracted via low-pass filtering processing and the E-portion was generated with a vocoder model preserving multi-channel temporal envelope waveforms, whereas a noise-vocoder and a tone-vocoder were used to simulate the effect of carrier bandwidth. The synthesized stimuli were presented to normal-hearing listeners to recognize. Experimental results showed that while low-pass filtered Mandarin speech was not very intelligible, adding one or two E channels could significantly improve the intelligibility score to above 86.0%. Under the condition with one E channel, using a large carrier bandwidth in noise-vocoder processing provided a better intelligibility performance than using a narrow carrier bandwidth in tone-vocoder processing.

A Comparative Study of Different EMG Features for Acoustics-to-EMG Mapping

Manthan Sharma, Navaneetha Gaddam, Tejas Umesh, Aditya Murthy, Prasanta Kumar Ghosh; Indian Institute of Science, India

Tue-A-V-2-3, Time: 13:30

Electromyography (EMG) signals have been extensively used to capture facial muscle movements while speaking since they are one of the most closely related bio-signals generated during speech production. In this work, we focus on speech acoustics to EMG prediction. We present a comparative study of ten different EMG signal-based features including Time Domain (TD) features existing in the literature to examine their effectiveness in speech acoustics to EMG inverse (AEI) mapping. We propose a novel feature based on the Hilbert envelope of the filtered EMG signal. The raw EMG signal is reconstructed from these features as well. For the AEI mapping. we use a bi-directional long short-term memory (BLSTM) network in a session-dependent manner. To estimate the raw EMG signal from the EMG features, we use a CNN-BLSTM model comprising of a convolution neural network (CNN) followed by BLSTM layers. AEI mapping performance using the BLSTM network reveals that the Hilbert envelope based feature is predicted from speech with the highest accuracy, among all the features. Therefore, it could be the most representative feature of the underlying muscle activation during speech production. The proposed Hilbert envelope feature, when used together with the existing TD features, improves the raw EMG signal reconstruction performance compared to using the TD features alone.

Image-Based Assessment of Jaw Parameters and Jaw Kinematics for Articulatory Simulation: Preliminary Results

Ajish K. Abraham, V. Sivaramakrishnan, N. Swapna, N. Manohar; AIISH, India Tue-A-V-2-4, Time: 13:30

Correcting the deficits in jaw movements have often been ignored in assessment and treatment of speech disorders. A robotic simulation is being developed to facilitate Speech Language Pathologists to demonstrate the movement of jaw, tongue and teeth during production of speech sounds, as a part of a larger study. Profiling of jaw movement is an important aspect of articulatory simulation. The present study attempts to develop a simple and efficient technique for deriving the jaw parameters and using them to simulate jaw movements through inverse kinematics.

Three Kannada speaking male participants in the age range of 26 to 33 years were instructed to produce selected speech sounds. The image of the final position of the jaw during production of each speech sound was recorded through CT scan and video camera. Angle of ramus and angle of body of mandible were simulated through inverse kinematics using RoboAnalyzer software. The variables for inverse kinematics were derived through kinematic analysis. The Denavit-Hartenberg (D-H) parameters required for kinematic analysis were obtained from still image. Angles simulated were compared with the angles obtained from CT scan images. No significant difference was observed.

An Attention Self-Supervised Contrastive Learning Based Three-Stage Model for Hand Shape Feature Representation in Cued Speech

*Jianrong Wang*¹, *Nan Gu*¹, *Mei Yu*¹, *Xuewei Li*¹, *Qiang Fang*², *Li Liu*³; ¹*Tianjin University, China*; ²*CASS, China*; ³*CUHK, China* Tue-A-V-2-5, Time: 13:30

Cued Speech (CS) is a communication system for deaf people or hearing impaired people, in which a speaker uses it to aid a lipreader in phonetic level by clarifying potentially ambiguous mouth movements with hand shape and positions. Feature extraction of multi-modal CS is a key step in CS recognition. Recent supervised deep learning based methods suffer from noisy CS data annotations especially for hand shape modality. In this work, we first propose a self-supervised contrastive learning method to learn the feature representation of image without using labels. Secondly, a small amount of manually annotated CS data are used to fine-tune the first module. Thirdly, we present a module, which combines Bi-LSTM and self-attention networks to further learn sequential features with temporal and contextual information. Besides, to enlarge the volume and the diversity of the current limited CS datasets, we build a new British English dataset containing 5 native CS speakers. Evaluation results on both French and British English datasets show that our model achieves over 90% accuracy in hand shape recognition. Significant improvements of 8.75% (for French) and 10.09% (for British English) are achieved in CS phoneme recognition correctness compared with the state-of-the-art.

Remote Smartphone-Based Speech Collection: Acceptance and Barriers in Individuals with Major Depressive Disorder

Judith Dineley¹, Grace Lavelle², Daniel Leightley², Faith Matcham², Sara Siddi³, Maria Teresa Peñarrubia-María⁴, Katie M. White², Alina Ivan², Carolin Oetzmann², Sara Simblett², Erin Dawe-Lane², Stuart Bruce², Daniel Stahl², Yatharth Ranjan², Zulqarnain Rashid², Pauline Conde², Amos A. Folarin², Josep Maria Haro³, Til Wykes², Richard J.B. Dobson², Vaibhav A. Narayan⁵, Matthew Hotopf², Björn W. Schuller¹, Nicholas Cummins¹, The RADAR-CNS Consortium; ¹Universität Augsburg, Germany; ²King's College London, UK; ³CIBERSAM, Spain; ⁴IDIAP Jordi Gol, Spain; ⁵Janssen, USA

Tue-A-V-2-6, Time: 13:30

The ease of in-the-wild speech recording using smartphones has sparked considerable interest in the combined application of speech, remote measurement technology (RMT) and advanced analytics as a research and healthcare tool. For this to be realised, the acceptability of remote speech collection to the user must be established, in addition to feasibility from an analytical perspective. To understand the acceptance, facilitators, and barriers of smartphone-based speech recording, we invited 384 individuals with major depressive disorder (MDD) from the Remote Assessment of Disease and Relapse Central Nervous System (RADAR-CNS) research programme in Spain and the UK to complete a survey on their experiences recording their speech. In this analysis, we demonstrate that study participants were more comfortable completing a scripted speech task than a free speech task. For both speech tasks, we found depression severity and country to be significant predictors of comfort. Not seeing smartphone notifications of the scheduled speech tasks, low mood and forgetfulness were the most commonly reported obstacles to providing speech recordings.

An Automatic, Simple Ultrasound Biofeedback Parameter for Distinguishing Accurate and Misarticulated Rhotic Syllables

Sarah R. Li, Colin T. Annand, Sarah Dugan, Sarah M. Schwab, Kathryn J. Eary, Michael Swearengen, Sarah Stack, Suzanne Boyce, Michael A. Riley, T. Douglas Mast; University of Cincinnati, USA

Tue-A-V-2-7, Time: 13:30

Characterizing accurate vs. misarticulated patterns of tongue movement using ultrasound can be challenging in real time because of the fast, independent movement of tongue regions. The usefulness of ultrasound for biofeedback speech therapy is limited because speakers must mentally track and compare differences between their tongue movement and available models. It is desirable to automate this interpretive task using a single parameter representing deviation from known accurate tongue movements. In this study, displacements recorded automatically by ultrasound image tracking were transformed into a single biofeedback parameter (time-dependent difference between blade and dorsum displacements). Receiver operating characteristic (ROC) curve analysis was used to evaluate this parameter as a predictor of production accuracy over a range of different vowel contexts with initial and final /r/ in American English. Areas under ROC curves were 0.8 or above, indicating that this simple parameter may provide useful real-time biofeedback on /r/ accuracy within a range of rhotic contexts.

Silent versus Modal Multi-Speaker Speech Recognition from Ultrasound and Video

*Manuel Sam Ribeiro*¹, *Aciel Eshky*², *Korin Richmond*³, *Steve Renals*³; ¹*Amazon, Poland*; ²*Rasa Technologies, UK*; ³*University of Edinburgh, UK* Tue-A-V-2-8, Time: 13:30

We investigate multi-speaker speech recognition from ultrasound images of the tongue and video images of the lips. We train our systems on imaging data from modal speech, and evaluate on matched test sets of two speaking modes: silent and modal speech. We observe that silent speech recognition from imaging data underperforms compared to modal speech recognition, likely due to a speaking-mode mismatch between training and testing. We improve silent speech recognition performance using techniques that address the domain mismatch, such as fMLLR and unsupervised model adaptation. We also analyse the properties of silent and modal speech in terms of utterance duration and the size of the articulatory space. To estimate the articulatory space, we compute the convex hull of tongue splines, extracted from ultrasound tongue images. Overall, we observe that the duration of silent speech is longer than that of modal speech, and that silent speech covers a smaller articulatory space than modal speech. Although these two properties are statistically significant across speaking modes, they do not directly correlate with word error rates from speech recognition.

RaSSpeR: Radar-Based Silent Speech Recognition

David Ferreira, Samuel Silva, Francisco Curado, António Teixeira; Universidade de Aveiro, Portugal Tue-A-V-2-9, Time: 13:30

Speech is our most natural and efficient way of communication and offers a strong potential to improve how we interact with machines. However, speech communication can sometimes be limited by environmental (e.g., ambient noise), contextual (e.g., need for privacy in a public place), or health conditions (e.g., laryngectomy), hindering the consideration of audible speech. In this regard, silent

speech interfaces (SSI) have been proposed (e.g., considering video, electromyography), however, many technologies still face limitations regarding their everyday use, e.g., the need to place equipment in contact with the speaker (e.g., electrodes/ultrasound probe), and raise technical (e.g., lighting conditions for video) or privacy concerns. In this context, the consideration of technologies that can help tackle these issues, e.g, by being contactless and/or placed in the environment, can foster the widespread use of SSI. In this article, continuous-wave radar is explored to assess its potential for SSI. To this end, a corpus of 13 words was acquired, for 3 speakers, and different classifiers were tested on the resulting data. The best results, obtained using Bagging classifier, trained for each speaker, with 5-fold cross-validation, yielded an average accuracy of 0.826, an encouraging result that establishes promising grounds for further exploration of this technology for silent speech recognition.

Investigating Speech Reconstruction for Laryngectomees for Silent Speech Interfaces

Beiming Cao¹, Nordine Sebkhi², Arpan Bhavsar², Omer T. Inan², Robin Samlan³, Ted Mau⁴, Jun Wang¹; ¹University of Texas at Austin, USA; ²Georgia Tech, USA; ³University of Arizona, USA; ⁴UT Southwestern Medical Center, USA

Tue-A-V-2-10, Time: 13:30

Silent speech interfaces (SSIs) are devices that convert non-audio bio-signals to speech, which hold the potential of recovering quality speech for laryngectomees (people who have undergone laryngectomy). Although significant progress has been made, most of the recent SSI works focused on data collected from healthy speakers. SSIs for laryngectomees have rarely been investigated. In this study, we investigated the reconstruction of speech for two laryngectomees who either use tracheoesophageal puncture (TEP) or electro-larynx (EL) speech as their post-surgery communication mode. We reconstructed their speech using two SSI designs (1) realtime recognition-and-synthesis and (2) directly articulation-to-speech synthesis (ATS). The reconstructed speech samples were measured in subjective evaluation by 20 listeners in terms of naturalness and intelligibility. The results indicated that both designs increased the naturalness of alaryngeal speech. The real-time recognitionand-synthesis design obtained higher intelligibility in electrolarynx speech as well, while the ATS did not. These preliminary results suggest the real-time recognition-and-synthesis design may have a better potential for clinical applications (for laryngectomees) than ATS.

Tue-A-V-3: Multi-Channel Speech Enhancement and Hearing Aids

13:30-15:30, Tuesday 31 August 2021 Chairs: Masahito Togami and Ina Kodrasi

LACOPE: Latency-Constrained Pitch Estimation for Speech Enhancement

*Hendrik Schröter*¹, *Tobias Rosenkranz*², *Alberto N. Escalante-B.*², *Andreas Maier*¹; ¹*FAU Erlangen-Nürnberg, Germany*; ²*Sivantos, Germany* Tue-A-V-3-1, Time: 13:30

Fundamental frequency (f_0) estimation, also known as pitch tracking, has been a long-standing research topic in the speech and signal processing community. Many pitch estimation algorithms, however, fail in noisy conditions or introduce large delays due to their frame size or Viterbi decoding.

In this study, we propose a deep learning-based pitch estimation algorithm, LACOPE, which was trained in a joint pitch estimation and speech enhancement framework. In contrast to previous work, this algorithm allows for a configurable latency down to an algorithmic delay of 0. This is achieved by exploiting the smoothness properties of the pitch trajectory. That is, a recurrent neural network compensates delay introduced by the feature computation by predicting the pitch for a desired point, allowing a trade-off between pitch accuracy and latency.

We integrate the pitch estimation in a speech enhancement framework for hearing aids. For this application, we allow a delay on the analysis side of approx. 5ms. The pitch estimate is then used for constructing a comb filter in frequency domain as post-processing step to remove intra-harmonic noise.

Our pitch estimation performance is on par with SOTA algorithms like PYIN or CREPE for spoken speech in all noise conditions while introducing minimal latency.

Alpha-Stable Autoregressive Fast Multichannel Nonnegative Matrix Factorization for Joint Speech Enhancement and Dereverberation

Mathieu Fontaine¹, Kouhei Sekiguchi¹, Aditya Arie Nugraha¹, Yoshiaki Bando², Kazuyoshi Yoshii³; ¹RIKEN, Japan; ²AIST, Japan; ³Kyoto University, Japan Tue-A-V-3-2, Time: 13:30

This paper proposes α -stable autoregressive fast multichannel nonnegative matrix factorization (α -AR-FastMNMF), a robust joint blind speech enhancement and dereverberation method for improved automatic speech recognition in a realistic adverse environment. The state-of-the-art versatile blind source separation method called FastMNMF that assumes the short-time Fourier transform (STFT) coefficients of a direct sound to follow a circular complex Gaussian distribution with jointly-diagonalizable full-rank spatial covariance matrices was extended to AR-FastMNMF with an autoregressive reverberation model. Instead of the light-tailed Gaussian distribution. we use the heavy-tailed α -stable distribution, which also has the reproductive property useful for the additive source modeling, to better deal with the large dynamic range of the direct sound. The experimental results demonstrate that the proposed α -AR-FastMNMF works well as a front-end of an automatic speech recognition system. It outperforms α -AR-ILRMA, which is a special case of α -AR-FastMNMF, and their Gaussian counterparts, i.e., AR-FastMNMF and AR-ILRMA, in terms of the speech signal quality metrics and word error rate.

Microphone Array Generalization for Multichannel Narrowband Deep Speech Enhancement

Siyuan Zhang, Xiaofei Li; Westlake University, China Tue-A-V-3-3, Time: 13:30

This paper addresses the problem of microphone array generalization for deep-learning-based end-to-end multichannel speech enhancement. We aim to train a unique deep neural network (DNN) potentially performing well on unseen microphone arrays. The microphone array geometry shapes the network's parameters when training on a fixed microphone array, and thus restricts the generalization of the trained network to another microphone array. To resolve this problem, a single network is trained using data recorded by various microphone arrays of different geometries. We design three variants of our recently proposed narrowband network to cope with the agnostic number of microphones. Overall, the goal is to make the network learn the universal information for speech enhancement that is available for any array geometry, rather than learn the one-array-dedicated characteristics. The experiments on

NOTES

both simulated and real room impulse responses (RIR) demonstrate the excellent across-array generalization capability of the proposed networks, in the sense that their performance measures are very close to, or even exceed the network trained with test arrays. Moreover, they notably outperform various beamforming methods and other advanced deep-learning-based methods.

Multiple Sound Source Localization Based on Interchannel Phase Differences in All Frequencies with Spectral Masks

Hyungchan Song, Jong Won Shin; GIST, Korea Tue-A-V-3-4, Time: 13:30

One of the most widely used cues for sound source localization is the interchannel phase differences (IPDs) in the frequency domain. However, the spatial aliasing makes the utilization of the IPDs in the high frequencies difficult, especially when the distance between the microphones is high. Recently, the phase replication method which considers the direction-of-arrival (DoA) candidates corresponding to all the possible unwrapped phase differences in all frequency bins was proposed. However, high frequency bins with possible spatial aliasing contribute more when constructing initial DoA histograms compared with low frequency bins, which may not be desirable for source localization. In this paper, we propose to utilize the IPDs in all the frequency bins with equal weights regardless of maximum number of phase wrapping in that frequency for dual microphone sound source localization. We applied spectral masks based on local signal-to-noise ratios and coherences between microphone signals to exclude time-frequency bins without directional audio signal from the DoA histogram construction. Experimental results show that the proposed method results in more distinct peaks in the DoA histogram and outperforms the conventional method in various noisy and reverberant environments.

Cancellation of Local Competing Speaker with Near-Field Localization for Distributed ad-hoc Sensor Network

Pablo Pérez Zarazaga¹, Mariem Bouafif Mansali², Tom Bäckström¹, Zied Lachiri²; ¹Aalto University, Finland; ²Université de Tunis El Manar, Tunisia Tue-A-V-3-5, Time: 13:30

In scenarios such as remote work, open offices and call centers, multiple people may simultaneously have independent spoken interactions with their devices in the same room. The speech of competing speakers will however be picked up by all microphones, both reducing the quality of audio and exposing speakers to breaches in privacy. We propose a cooperative cross-talk cancellation solution breaking the single active speaker assumption employed by most telecommunication systems. The proposed method applies source separation on the microphone signals of independent devices, to extract the dominant speaker in each device. It is realized using a localization estimator based on a deep neural network, followed by a time-frequency mask to separate the target speech from the interfering one at each time-frequency unit referring to its orientation. By experimental evaluation, we confirm that the proposed method effectively reduces crosstalk and exceeds the baseline expectation maximization method by 10 dB in terms of interference rejection. This performance makes the proposed method a viable solution for cross-talk cancellation in near-field conditions, thus protecting the privacy of external speakers in the same acoustic space.

A Deep Learning Method to Multi-Channel Active Noise Control

Hao Zhang, DeLiang Wang; Ohio State University, USA Tue-A-V-3-6, Time: 13:30

This paper addresses multi-channel active noise control (MCANC) on the basis of deep ANC, which performs active noise control by employing deep learning to encode the optimal control parameters corresponding to different noises and environments. The proposed method trains a convolutional recurrent network (CRN) to estimate the real and imaginary spectrograms of all the canceling signals simultaneously from the reference signals so that the corresponding anti-noises cancel or attenuate the primary noises in an MCANC system. We evaluate the proposed method under multiple MCANC setups and investigate the impact of the number of canceling loudspeakers and error microphones on the overall performance. Experimental results show that deep ANC is effective for MCANC in various scenarios. Moreover, the proposed method is robust against untrained noises and works well in the presence of loudspeaker nonlinearity.

Clarity-2021 Challenges: Machine Learning Challenges for Advancing Hearing Aid Processing

Simone Graetzer¹, Jon Barker², Trevor J. Cox¹, Michael Akeroyd³, John F. Culling⁴, Graham Naylor³, Eszter Porter³, Rhoddy Viveros Muñoz⁴; ¹University of Salford, UK; ²University of Sheffield, UK; ³University of Nottingham, UK; ⁴Cardiff University, UK Tue-A-V-3-7, Time: 13:30

In recent years, rapid advances in speech technology have been made possible by machine learning challenges such as CHiME, REVERB, Blizzard, and Hurricane. In the Clarity project, the machine learning approach is applied to the problem of hearing aid processing of speech-in-noise, where current technology in enhancing the speech signal for the hearing aid wearer is often ineffective. The scenario is a (simulated) cuboid-shaped living room in which there is a single listener, a single target speaker and a single interferer, which is either a competing talker or domestic noise. All sources are static, the target is always within $\pm 30^{\circ}$ azimuth of the listener and at the same elevation, and the interferer is an omnidirectional point source at the same elevation. The target speech comes from an open source 40-speaker British English speech database collected for this purpose. This paper provides a baseline description of the round one Clarity challenges for both enhancement (CEC1) and prediction (CPC1). To the authors' knowledge, these are the first machine learning challenges to consider the problem of hearing aid speech signal processing.

Optimising Hearing Aid Fittings for Speech in Noise with a Differentiable Hearing Loss Model

Zehai Tu, Ning Ma, Jon Barker; University of Sheffield, UK

Tue-A-V-3-8, Time: 13:30

Current hearing aids normally provide amplification based on a general prescriptive fitting, and the benefits provided by the hearing aids vary among different listening environments despite the inclusion of noise suppression feature. Motivated by this fact, this paper proposes a data-driven machine learning technique to develop hearing aid fittings that are customised to speech in different noisy environments. A differentiable hearing loss model is proposed and used to optimise fittings with back-propagation. The customisation is reflected on the data of speech in different noise with also the consideration of noise suppression. The objective evaluation shows

NOTES

the advantages of optimised custom fittings over general prescriptive fittings.

Explaining Deep Learning Models for Speech Enhancement

Sunit Sivasankaran¹, Emmanuel Vincent², Dominique Fohr²; ¹Microsoft, USA; ²Loria (UMR 7503), France Tue-A-V-3-9, Time: 13:30

We consider the problem of explaining the robustness of neural networks used to compute time-frequency masks for speech enhancement to mismatched noise conditions. We employ the Deep SHapley Additive exPlanations (DeepSHAP) feature attribution method to quantify the contribution of every time-frequency bin in the input noisy speech signal to every time-frequency bin in the output time-frequency mask. We define an objective metric - referred to as the speech relevance score - that summarizes the obtained SHAP values and show that it correlates with the enhancement performance, as measured by the word error rate on the CHiME-4 real evaluation dataset. We use the speech relevance score to explain the generalization ability of three speech enhancement models trained using synthetically generated speech-shaped noise, noise from a professional sound effects library, or real CHiME-4 noise. To the best of our knowledge, this is the first study on neural network explainability in the context of speech enhancement.

Minimum-Norm Differential Beamforming for Linear Array with Directional Microphones

*Weilong Huang*¹, *Jinwei Feng*²; ¹*Alibaba, China;* ²*Alibaba, USA*

Tue-A-V-3-10, Time: 13:30

Among different differential beamforming approaches, the minimumnorm one has received much attention as it maximizes the white noise gain(WNG). WNG measures the robustness of beamformer. But in practice, the conventional minimum-norm differential beamforming with omnidirectional elements still suffers in low white-noise-gain at the low frequencies. The major contributions of this paper are as follows: First, we extend the existing work by presenting a new solution with the use of the directional microphone elements, and show clearly the connection between the conventional beamforming and the proposed beamforming. Second, through the derivation as well as simulations, we show the proposed solution brings noticeable improvement in WNG at the low frequencies when the null positions of the directional elements coincide with the null-constraints of minimum norm solution.

Tue-A-V-4: Self-Supervision and Semi-Supervision for Neural ASR Training

13:30-15:30, Tuesday 31 August 2021 Chairs: Tara Sainath and Herman Kamper

Improving Streaming Transformer Based ASR Under a Framework of Self-Supervised Learning

Songjun Cao, Yueteng Kang, Yanzhe Fu, Xiaoshuo Xu, Sining Sun, Yike Zhang, Long Ma; Tencent, China Tue-A-V-4-1, Time: 13:30

Recently self-supervised learning has emerged as an effective approach to improve the performance of automatic speech recognition (ASR). Under such a framework, the neural network is usually pre-trained with massive unlabeled data and then fine-tuned with limited labeled data. However, the non-streaming architecture like

bidirectional transformer is usually adopted by the neural network to achieve competitive results, which cannot be used in streaming scenarios. In this paper, we mainly focus on improving the performance of streaming transformer under the self-supervised learning framework. Specifically, we propose a novel two-stage training method during fine-tuning, which combines knowledge distilling and self-training. The proposed training method achieves 16.3% relative word error rate (WER) reduction on Librispeech noisy test set. Finally, by only using the 100h clean subset of Librispeech as the labeled data and the rest (860h) as the unlabeled data, our streaming transformer based model obtains competitive WERs 3.5/8.7 on Librispeech clean/noisy test sets.

wav2vec-C: A Self-Supervised Model for Speech Representation Learning

Samik Sadhu¹, Di He², Che-Wei Huang², Sri Harish Mallidi², Minhua Wu², Ariya Rastrow², Andreas Stolcke², Jasha Droppo², Roland Maas²; ¹Johns Hopkins University, USA; ²Amazon, USA Tue-A-V-4-2, Time: 13:30

wav2vec-C introduces a novel representation learning technique combining elements from wav2vec 2.0 and VQ-VAE. Our model learns to reproduce quantized representations from partially masked speech encoding using a contrastive loss in a way similar to wav2vec 2.0. However, the quantization process is regularized by an additional consistency network that learns to reconstruct the input features to the wav2vec 2.0 network from the quantized representations in a way similar to a VQ-VAE model. The proposed self-supervised model is trained on 10k hours of unlabeled data and subsequently used as the speech encoder in a RNN-T ASR model and fine-tuned with 1k hours of labeled data. This work is one of the very few studies of self-supervised learning on speech tasks with a large volume of real far-field labeled data. The wav2vec-C encoded representations achieve, on average, twice the error reduction over baseline and a higher codebook utilization in comparison to wav2vec 2.0.

On the Learning Dynamics of Semi-Supervised Training for ASR

Electra Wallington, Benji Kershenbaum, Ondřej Klejch, Peter Bell; University of Edinburgh, UK Tue-A-V-4-3, Time: 13:30

The use of semi-supervised training (SST) has become an increasingly popular way of increasing the performance of ASR acoustic models without the need for further transcribed speech data. However, the performance of the technique can be very sensitive to the quality of the initial ASR system. This paper undertakes a comprehensive study of the improvements gained with respect to variation in the initial systems, the quantity of untranscribed data used, and the learning schedules. We postulate that the reason SST can be effective even when the initial model is poor is because it enables utterance-level information to be propagated to the frame level, and hence hypothesise that the quality of the language model plays a much larger role than the quality of the acoustic model. In experiments on Tagalog data from the IARPA MATERIAL programme, we find that indeed this is the case, and show that with an appropriately chosen recipe it is possible to achieve over 50% relative WER reductions from SST, even when the WER of the initial system is more than 80%.

Robust wav2vec 2.0: Analyzing Domain Shift in Self-Supervised Pre-Training

Wei-Ning Hsu¹, Anuroop Sriram¹, Alexei Baevski¹, Tatiana Likhomanenko¹, Qiantong Xu¹, Vineel Pratap¹, Jacob Kahn¹, Ann Lee¹, Ronan Collobert¹, Gabriel Synnaeve², Michael Auli¹; ¹Facebook, USA; ²Facebook, France

Tue-A-V-4-4, Time: 13:30

Self-supervised learning of speech representations has been a very active research area but most work is focused on a single domain such as read audio books for which there exist large quantities of labeled and unlabeled data. In this paper, we explore more general setups where the domain of the unlabeled data for pre-training data differs from the domain of the labeled data for fine-tuning, which in turn may differ from the test data domain. Our experiments show that using target domain data during pre-training leads to large performance improvements across a variety of setups. With no access to in-domain labeled data, pre-training on unlabeled in-domain data closes 66-73% of the performance gap between the ideal setting of in-domain labeled data and a competitive supervised out-of-domain model. This has obvious practical implications since it is much easier to obtain unlabeled target domain data than labeled data. Moreover, we find that pre-training on multiple domains improves generalization performance on domains not seen during training. We will release pre-trained models.

Momentum Pseudo-Labeling for Semi-Supervised Speech Recognition

Yosuke Higuchi, Niko Moritz, Jonathan Le Roux, Takaaki Hori; MERL, USA

Tue-A-V-4-5, Time: 13:30

Pseudo-labeling (PL) has been shown to be effective in semisupervised automatic speech recognition (ASR), where a base model is self-trained with pseudo-labels generated from unlabeled data. While PL can be further improved by iteratively updating pseudolabels as the model evolves, most of the previous approaches involve inefficient retraining of the model or intricate control of the label update. We present momentum pseudo-labeling (MPL), a simple yet effective strategy for semi-supervised ASR. MPL consists of a pair of online and offline models that interact and learn from each other, inspired by the mean teacher method. The online model is trained to predict pseudo-labels generated on the fly by the offline model. The offline model maintains a momentum-based moving average of the online model. MPL is performed in a single training process and the interaction between the two models effectively helps them reinforce each other to improve the ASR performance. We apply MPL to an end-to-end ASR model based on the connectionist temporal classification. The experimental results demonstrate that MPL effectively improves over the base model and is scalable to different semi-supervised scenarios with varying amounts of data or domain mismatch.

A Comparison of Supervised and Unsupervised Pre-Training of End-to-End Models

Ananya Misra, Dongseong Hwang, Zhouyuan Huo, Shefali Garg, Nikhil Siddhartha, Arun Narayanan, Khe Chai Sim; Google, USA Tue-A-V-4-6, Time: 13:30

In the absence of large-scale in-domain supervised training data, ASR models can achieve reasonable performance through pre-training on additional data that is unlabeled, mismatched or both. Given such data constraints, we compare pre-training end-to-end models

on matched but unlabeled data (unsupervised) and on labeled but mismatched data (supervised), where the labeled data is mismatched in either domain or language. Across encoder architectures, pretraining methods and languages, our experiments indicate that both types of pre-training improve performance, with relative WER reductions of 15–30% in the domain mismatch case and up to 15% in the language mismatch condition. We further find that the advantage from unsupervised pre-training is most prominent when there is no matched and labeled fine-tuning data, provided that a sufficient amount of mismatched data is still available for supervised fine-tuning.

Semi-Supervision in ASR: Sequential MixMatch and Factorized TTS-Based Augmentation

Zhehuai Chen¹, Andrew Rosenberg¹, Yu Zhang¹, Heiga Zen², Mohammadreza Ghodsi¹, Yinghui Huang¹, Jesse Emond¹, Gary Wang¹, Bhuvana Ramabhadran¹, Pedro J. Moreno¹; ¹Google, USA; ²Google, Japan Tue-A-V-4-7, Time: 13:30

Semi and self-supervised training techniques have the potential to improve performance of speech recognition systems without additional transcribed speech data. In this work, we demonstrate the efficacy of two approaches to semi-supervision for automated speech recognition. The two approaches leverage vast amounts of available unspoken text and untranscribed audio. First, we present *factorized multilingual speech synthesis* to improve data augmentation on unspoken text. Next, we propose the *Sequential MixMatch* algorithm with *iterative learning* to learn from untranscribed speech. The algorithm is built on top of our online implementation of Noisy Student Training. We demonstrate the compatibility of these techniques yielding an overall relative reduction of word error rate of up to 14.4% on the voice search tasks on 4 Indic languages.

slimIPL: Language-Model-Free Iterative Pseudo-Labeling

Tatiana Likhomanenko¹, Qiantong Xu¹, Jacob Kahn¹, Gabriel Synnaeve², Ronan Collobert¹; ¹Facebook, USA; ²Facebook, France____

Tue-A-V-4-8, Time: 13:30

Recent results in end-to-end automatic speech recognition have demonstrated the efficacy of pseudo-labeling for semi-supervised models trained both with Connectionist Temporal Classification (CTC) and Sequence-to-Sequence (seq2seq) losses. Iterative Pseudo-Labeling (IPL), which continuously trains a single model using pseudo-labels iteratively re-generated as the model learns, has been shown to further improve performance in ASR. We improve upon the IPL algorithm: as the model learns, we propose to iteratively re-generate transcriptions with hard labels (the most probable tokens), that is, without a language model. We call this approach Language-Model-Free IPL (slimIPL) and give a resultant training setup for low-resource settings with CTC-based models. slimIPL features a dynamic cache for pseudo-labels which reduces sensitivity to changes in relabeling hyperparameters and results in improved training stability. slimIPL is also highly-efficient and requires $3.5-4 \times$ fewer computational resources to converge than other state-of-theart semi/self-supervised approaches. With only 10 hours of labeled audio, slimIPL is competitive with self-supervised approaches, and is state-of-the-art with 100 hours of labeled audio without the use of a language model both at test time and during pseudo-label generation.

Phonetically Motivated Self-Supervised Speech Representation Learning

Xianghu Yue, Haizhou Li; NUS, Singapore Tue-A-V-4-9, Time: 13:30

Self-supervised representation learning has seen remarkable success in encoding high-level semantic information from unlabelled speech data. The studies have been focused on exploring new pretext tasks to improve the learned speech representation and various masking schemes with reference to speech frames. We consider effective latent speech representation should be phonetically informed. In this work, we propose a novel phonetically motivated masking scheme. Specifically, we select the masked speech frames according to the phonetic segmentation in an utterance. The phonetically motivated self-supervised representation learns the speech representation that benefits downstream speech processing tasks. We evaluate the proposed learning algorithm on phoneme classification, speech recognition, and speaker recognition, and show that it consistently outperforms competitive baselines.

Improving RNN-T for Domain Scaling Using Semi-Supervised Training with Neural TTS

*Yan Deng*¹, *Rui Zhao*², *Zhong Meng*², *Xie Chen*², *Bing Liu*¹, *Jinyu Li*², *Yifan Gong*², *Lei He*¹; ¹*Microsoft, China;* ²*Microsoft, USA*

Tue-A-V-4-10, Time: 13:30

Recurrent neural network transducer (RNN-T) has shown to be comparable with conventional hybrid model for speech recognition. However, there is still a challenge in out-of-domain scenarios with context or words different from training data. In this paper, we explore the semi-supervised training which optimizes RNN-T jointly with neural text-to-speech (TTS) to better generalize to new domains using domain-specific text data. We apply the method to two tasks: one with out-of-domain context and the other with significant out-of-vocabulary (OOV) words. The results show that the proposed method significantly improves the recognition accuracy in both tasks, resulting in 61.4% and 53.8% relative word error rate (WER) reductions respectively, from a well-trained RNN-T with 65 thousand hours of training data. We do further study on the semi-supervised training methodology: 1) which modules of RNN-T model to be updated; 2) the impact of using different neural TTS models; 3) the performance of using text with different relevancy to target domain. Finally, we compare several RNN-T customization methods, and conclude that semi-supervised training with neural TTS is comparable and complementary with Internal Language Model Estimation (ILME) or biasing.

Tue-A-V-5: Spoken Language Processing I

13:30-15:30, Tuesday 31 August 2021 Chairs: Jan Švec and Roger Moore

Speaker-Conversation Factorial Designs for Diarization Error Analysis

Scott Seyfarth, Sundararajan Srinivasan, Katrin Kirchhoff; Amazon, USA Tue-A-V-5-1, Time: 13:30

Speaker diarization accuracy can be affected by both acoustics and conversation characteristics. Determining the cause of diarization errors is difficult because speaker voice acoustics and conversation structure co-vary, and the interactions between acoustics, conversational structure, and diarization accuracy are complex. This paper proposes a methodology that can distinguish independent marginal effects of acoustic and conversation characteristics on diarization

accuracy by remixing conversations in a factorial design. As an illustration, this approach is used to investigate gender-related and language-related accuracy differences with three diarization systems: a baseline system using subsegment x-vector clustering, a variant of it with shorter subsegments, and a third system based on a Bayesian hidden Markov model. Our analysis shows large accuracy disparities for the baseline system primarily due to conversational structure, which are partially mitigated in the other two systems. The illustration thus demonstrates how the methodology can be used to identify and guide diarization model improvements.

SmallER: Scaling Neural Entity Resolution for Edge Devices

Ross McGowan, Jinru Su, Vince DiCocco, Thejaswi Muniyappa, Grant P. Strimel; Amazon, USA Tue-A-V-5-2, Time: 13:30

In this paper we introduce SmallER, a scalable neural entity resolution system capable of running directly on edge devices. SmallER addresses constraints imposed by the on-device setting such as bounded memory consumption for both model and catalog storage, limited compute resources, and related latency challenges introduced by those restrictions. Our model includes distinct modules to learn syntactic and semantic information and is trained to handle multiple domains within one compact architecture. We use compressed tries to reduce the space required to store catalogs and a novel implementation of spatial partitioning trees to strike a balance between reducing runtime latency and preserving recall relative to full catalog search. Our final model consumes only 3MB of memory at inference time with classification accuracy surpassing that of previously established, domain-specific baseline models on live customer utterances. For the largest catalogs we consider (300 or more entries), our proxy metric for runtime latency is reduced by more than 90%

Disfluency Detection with Unlabeled Data and Small BERT Models

Johann C. Rocholl, Vicky Zayats, Daniel D. Walker, Noah B. Murad, Aaron Schneider, Daniel J. Liebling; Google, USA

Tue-A-V-5-3, Time: 13:30

Disfluency detection models now approach high accuracy on English text. However, little exploration has been done in improving the size and inference time of the model. At the same time, Automatic Speech Recognition (ASR) models are moving from server-side inference to local, on-device inference. Supporting models in the transcription pipeline (like disfluency detection) must follow suit. In this work we concentrate on the disfluency detection task, focusing on small, fast, on-device models based on the BERT architecture. We demonstrate it is possible to train disfluency detection models as small as 1.3 MiB, while retaining high performance. We build on previous work that showed the benefit of data augmentation approaches such as self-training. Then, we evaluate the effect of domain mismatch between conversational and written text on model performance. We find that domain adaptation and data augmentation strategies have a more pronounced effect on these smaller models, as compared to conventional BERT models.

Discriminative Self-Training for Punctuation Prediction

Qian Chen, Wen Wang, Mengzhe Chen, Qinglin Zhang; Alibaba, China

Tue-A-V-5-4, Time: 13:30

Punctuation prediction for automatic speech recognition (ASR) output transcripts plays a crucial role for improving the readability

NOTES

of the ASR transcripts and for improving the performance of downstream natural language processing applications. However, achieving good performance on punctuation prediction often requires large amounts of labeled speech transcripts, which is expensive and laborious. In this paper, we propose a Discriminative Self-Training approach with weighted loss and discriminative label smoothing to exploit unlabeled speech transcripts. Experimental results on the English IWSLT2011 benchmark test set and an internal Chinese spoken language dataset demonstrate that the proposed approach achieves significant improvement on punctuation prediction accuracy over strong baselines including BERT, RoBERTa, and ELECTRA models. The proposed Discriminative Self-Training approach outperforms the vanilla self-training approach. We establish a new state-of-the-art (SOTA) on the IWSLT2011 test set, outperforming the current SOTA model by 1.3% absolute gain on F₁.

Zero-Shot Joint Modeling of Multiple Spoken-Text-Style Conversion Tasks Using Switching Tokens

Mana Ihori, Naoki Makishima, Tomohiro Tanaka, Akihiko Takashima, Shota Orihashi, Ryo Masumura; NTT, Japan

Tue-A-V-5-5, Time: 13:30

In this paper, we propose a novel spoken-text-style conversion method that can simultaneously execute multiple style conversion modules such as punctuation restoration and disfluency deletion without preparing matched datasets. In practice, transcriptions generated by automatic speech recognition systems are not highly readable because they often include many disfluencies and do not include punctuation marks. To improve their readability, multiple spoken-text-style conversion modules that individually model a single conversion task are cascaded because matched datasets that simultaneously handle multiple conversion tasks are often unavailable. However, the cascading is unstable against the order of tasks because of the chain of conversion errors. Besides, the computation cost of the cascading must be higher than the single conversion. To execute multiple conversion tasks simultaneously without preparing matched datasets, our key idea is to distinguish individual conversion tasks using the on-off switch. In our proposed zero-shot joint modeling, we switch the individual tasks using multiple switching tokens, enabling us to utilize a zero-shot learning approach to executing simultaneous conversions. Our experiments on joint modeling of disfluency deletion and punctuation restoration demonstrate the effectiveness of our method.

A Noise Robust Method for Word-Level Pronunciation Assessment

Binghuai Lin, Liyuan Wang; Tencent, China Tue-A-V-5-6, Time: 13:30

The common approach for pronunciation evaluation is based on Goodness of pronunciation (GOP). It has been found that GOP may perform worse under noise conditions. Traditional methods compensate pronunciation features to improve the performance of pronunciation assessment in noise situations. This paper proposed a noise robust model for word-level pronunciation assessment based on a domain adversarial training (DAT) method. We treat the pronunciation assessment in the clean and noise situations as the source and target domains. The network is optimized by incorporating both the pronunciation assessment and noise domain discrimination. The domain labels are generated from unsupervised methods to adapt to various noise situations. We evaluate the model performance based on English words recorded by Chinese English learners and labeled by three experts. Experimental results show on average the proposed model outperforms the baseline by 3% in Pearson correlation coefficients (PCC) and 4% in accuracy under different noise conditions.

Targeted Keyword Filtering for Accelerated Spoken Topic Identification

Jonathan Wintrode; Raytheon, USA Tue-A-V-5-7, Time: 13:30

We present a novel framework for spoken topic identification that simultaneously learns both topic-specific keywords and acoustic keyword filters from only document-level topic labels. At inference time, only audio segments likely to contain topic-salient keywords are fully decoded, reducing the system's overall computation cost. We show that this filtering allows for effective topic classification while decoding only 50% of ASR output word lattices, and achieves error rates within 1.2% and precision within 2.6% of an unfiltered baseline system.

Multimodal Speech Summarization Through Semantic Concept Learning

Shruti Palaskar, Ruslan Salakhutdinov, Alan W. Black, Florian Metze; Carnegie Mellon University, USA Tue-A-V-5-8, Time: 13:30

We propose a cascaded multimodal abstractive speech summarization model that generates semantic concepts as an intermediate step towards summarization. We describe a method to leverage existing multimodal dataset annotations to curate groundtruth labels for such intermediate concept modeling. In addition to cascaded training, the concept labels also provide an interpretable intermediate output level that helps improve performance on the downstream summarization task. On the open-domain How2 data, we conduct utterance-level and video-level experiments for two granularities of concepts: Specific and Abstract. We compare various multimodal fusion models for concept generation based on the respective input modalities. We observe consistent improvements in concept modeling by using multimodal adaptation models over unimodal Using the cascaded multimodal speech summarization models. model, we see a significant improvement of 7.5 METEOR points and 5.1 ROUGE-L points compared to previous methods of speech summarization. Finally, we show the benefits of scalability of the proposed approaches on 2000 h of video data.

Enhancing Semantic Understanding with Self-Supervised Methods for Abstractive Dialogue Summarization

Hyunjae Lee, Jaewoong Yun, Hyunjin Choi, Seongho Joe, Youngjune L. Gwon; Samsung, Korea Tue-A-V-5-9, Time: 13:30

Contextualized word embeddings can lead to state-of-the-art performances in natural language understanding. Recently, a pre-trained deep contextualized text encoder such as BERT has shown its potential in improving natural language tasks including abstractive summarization. Existing approaches in dialogue summarization focus on incorporating a large language model into summarization task trained on large-scale corpora consisting of news articles rather than dialogues of multiple speakers. In this paper, we introduce self-supervised methods to compensate shortcomings to train a dialogue summarization model. Our principle is to detect incoherent information flows using pretext dialogue text to enhance BERT's ability to contextualize the dialogue text representations. We build and fine-tune an abstractive dialogue summarization model on a shared encoder-decoder architecture using the enhanced BERT. We empirically evaluate our abstractive dialogue summarizer with the

NOTES

SAMSum corpus, a recently introduced dataset with abstractive dialogue summaries. All of our methods have contributed improvements to abstractive summary measured in ROUGE scores. Through an extensive ablation study, we also present a sensitivity analysis to critical model hyperparameters, probabilities of switching utterances and masking interlocutors.

Speaker Transition Patterns in Three-Party Conversation: Evidence from English, Estonian and Swedish

*Marcin Włodarczak*¹, *Emer Gilmartin*²; ¹*Stockholm University, Sweden;* ²*Trinity College Dublin, Ireland* Tue-A-V-5-10, Time: 13:30

During conversation, speakers hold and relinquish the floor, resulting in turn yield and retention. We examine these phenomena in three-party conversations in English, Swedish, and Estonian. We define within- and between-speaker transitions in terms of shorter intervals of speech, silence and overlap bounded by stretches of one-party speech longer than 1 second by the same or different speakers. This method gives us insights into how turn change and retention proceed, revealing that the majority of speaker transitions are more complex and involve more intermediate activity than a single silence or overlap. We examine the composition of within and between transitions in terms of number of speakers involved, incidence and proportion of solo speech, silence and overlap. We derive the most common within- and between-speaker transitions in the three languages, finding evidence of striking commonalities in how the floor is managed. Our findings suggest that current models of turn-taking used in dialogue technology could be extended using these results to more accurately reflect the realities of human-human dialogue.

Tue-A-V-6: Voice Conversion and Adaptation II

13:30–15:30, Tuesday 31 August 2021 Chairs: Tomoki Toda and Jithendra Vepa

Investigating Deep Neural Structures and their Interpretability in the Domain of Voice Conversion

*Samuel J. Broughton*¹, *Md. Asif Jalal*², *Roger K. Moore*²; ¹*NUS, Singapore;* ²*University of Sheffield, UK* Tue-A-V-6-1, Time: 13:30

Generative Adversarial Networks (GANs) are machine learning networks based around creating synthetic data. Voice Conversion (VC) is a subset of voice translation that involves translating the paralinguistic features of a source speaker to a target speaker while preserving the linguistic information. The aim of non-parallel conditional GANs for VC is to translate an acoustic speech feature sequence from one domain to another without the use of paired data. In the study reported here, we investigated the interpretability of state-of-the-art implementations of non-parallel GANs in the domain of VC. We show that the learned representations in the repeating layers of a particular GAN architecture remain close to their original random initialised parameters, demonstrating that it is the number of repeating layers that is more responsible for the quality of the output. We also analysed the learned representations of a model trained on one particular dataset when used during transfer learning on another dataset. This also showed high levels of similarity in the repeating layers. Together, these results provide new insight into how the learned representations of deep generative networks change during learning and the importance of the number of layers, which would help build better GAN-based speech conversion models.

Limited Data Emotional Voice Conversion Leveraging Text-to-Speech: Two-Stage Sequence-to-Sequence Training

*Kun Zhou*¹, *Berrak Sisman*², *Haizhou Li*¹; ¹*NUS*, *Singapore*; ²*SUTD*, *Singapore* Tue-A-V-6-2, Time: 13:30

Emotional voice conversion (EVC) aims to change the emotional state of an utterance while preserving the linguistic content and speaker identity. In this paper, we propose a novel 2-stage training strategy for sequence-to-sequence emotional voice conversion with a limited amount of emotional speech data. We note that the proposed EVC framework leverages text-to-speech (TTS) as they share a common goal that is to generate high-quality expressive voice. In stage 1, we perform style initialization with a multi-speaker TTS corpus, to disentangle speaking style and linguistic content. In stage 2, we perform emotion training with a limited amount of emotional speech data, to learn how to disentangle emotional style and linguistic information from the speech. The proposed framework can perform both spectrum and prosody conversion and achieves significant improvement over the state-of-the-art baselines in both objective and subjective evaluation.

Adversarial Voice Conversion Against Neural Spoofing Detectors

*Yi-Yang Ding*¹, *Li-Juan Liu*², *Yu Hu*¹, *Zhen-Hua Ling*¹; ¹*USTC, China*; ²*iFLYTEK, China* Tue-A-V-6-3, Time: 13:30

The naturalness and similarity of voice conversion have been significantly improved in recent years with the development of deeplearning-based conversion models and neural vocoders. Accordingly, the task of detecting spoofing speech also attracts research attention. In the latest ASVspoof 2019 challenge, the best spoofing detection model can distinguish most artificial utterances from natural ones. Inspired by recent progress of adversarial example generation, this paper proposes an adversarial post-processing network (APN) which generates adversarial examples against a neural-network-based spoofing detector by white-box attack. The APN model post-processes the speech waveforms generated by a baseline voice conversion system. An adversarial loss derived from the spoofing detector together with two regularization losses are applied to optimize the parameters of APN. In our experiments, using the logical access (LA) dataset of ASVspoof 2019, results show that our proposed method can improve the adversarial ability of converted speech against the spoofing detectors based on light convolution neural networks (LCNNs) effectively without degrading its subjective quality.

An Improved StarGAN for Emotional Voice Conversion: Enhancing Voice Quality and Data Augmentation

Xiangheng He¹, Junjie Chen², Georgios Rizos¹, Björn W. Schuller¹; ¹Imperial College London, UK; ²University of Tokyo, Japan

Tue-A-V-6-4, Time: 13:30

Emotional Voice Conversion (EVC) aims to convert the emotional style of a source speech signal to a target style while preserving its content and speaker identity information. Previous emotional conversion studies do not disentangle emotional information from emotion-independent information that should be preserved, thus transforming it all in a monolithic manner and generating audio of low quality, with linguistic distortions. To address this distortion problem, we propose a novel StarGAN framework along with a

NOTES

two-stage training process that separates emotional features from those independent of emotion by using an autoencoder with two encoders as the generator of the Generative Adversarial Network (GAN). The proposed model achieves favourable results in both the objective evaluation and the subjective evaluation in terms of distortion, which reveals that the proposed model can effectively reduce distortion. Furthermore, in data augmentation experiments for end-to-end speech emotion recognition, the proposed StarGAN model achieves an increase of 2% in Micro-F1 and 5% in Macro-F1 compared to the baseline StarGAN model, which indicates that the proposed model is more valuable for data augmentation.

TVQVC: Transformer Based Vector Quantized Variational Autoencoder with CTC Loss for Voice Conversion

Ziyi Chen, Pengyuan Zhang; CAS, China Tue-A-V-6-5, Time: 13:30

Techniques of voice conversion (VC) aim to modify the speaker identity and style of an utterance while preserving the linguistic content. Although there are lots of VC methods, the state of the art of VC is still cascading automatic speech recognition (ASR) and text-to-speech (TTS). This paper presents a new structure of vector-quantized autoencoder based on transformer with CTC loss for non-parallel VC, which inspired by cascading ASR and TTS VC method. Our proposed method combines CTC loss and vector quantization to get high-level linguistic information without speaker information. Objective and subjective evaluations on the mandarin datasets show that the converted speech of our proposed model is better than baselines on naturalness, rhythm and speaker similarity.

Enriching Source Style Transfer in Recognition-Synthesis Based Non-Parallel Voice Conversion

*Zhichao Wang*¹, *Xinyong Zhou*¹, *Fengyu Yang*¹, *Tao Li*¹, *Hongqiang Du*¹, *Lei Xie*¹, *Wendong Gan*², *Haitao Chen*², *Hai Li*²; ¹*Northwestern Polytechnical University, China*; ²*iQIYI, China* Tue-A-V-6-6, Time: 13:30

Current voice conversion (VC) methods can successfully convert timbre of the audio. As modeling source audio's prosody effectively is a challenging task, there are still limitations of transferring source style to the converted speech. This study proposes a source style transfer Previously method based on recognition-synthesis framework. in speech generation task, prosody can be modeled explicitly with prosodic features or implicitly with a latent prosody extractor. In this paper, taking advantages of both, we model the prosody in a hybrid manner, which effectively combines explicit and implicit methods in a proposed prosody module. Specifically, prosodic features are used to explicit model prosody, while VAE and reference encoder are used to implicitly model prosody, which take Mel spectrum and bottleneck feature as input respectively. Furthermore, adversarial training is introduced to remove speaker-related information from the VAE outputs, avoiding leaking source speaker information while transferring style. Finally, we use a modified self-attention based encoder to extract sentential context from bottleneck features, which also implicitly aggregates the prosodic aspects of source speech from the layered representations. Experiments show that our approach is superior to the baseline and a competitive system in terms of style transfer; meanwhile, the speech quality and speaker similarity are well maintained.

S2VC: A Framework for Any-to-Any Voice Conversion with Self-Supervised Pretrained Representations

Jheng-hao Lin, Yist Y. Lin, Chung-Ming Chien, Hung-yi Lee; National Taiwan University, Taiwan Tue-A-V-6-7, Time: 13:30

Any-to-any voice conversion (VC) aims to convert the timbre of utterances from and to any speakers seen or unseen during training. Various any-to-any VC approaches have been proposed like AUTOVC, AdaINVC, and FragmentVC. AUTOVC, and AdaINVC utilize source and target encoders to disentangle the content and speaker information of the features. FragmentVC utilizes two encoders to encode source and target information and adopts cross attention to align the source and target features with similar phonetic content. Moreover, pretrained features are adopted. AUTOVC used d-vector to extract speaker information, and self-supervised learning (SSL) features like wav2vec 2.0 is used in FragmentVC to extract the phonetic content information. Different from previous works, we proposed S2VC that utilizes Self-Supervised features as both source and target features for the VC model. Supervised phoneme posteriorgram (PPG), which is believed to be speaker-independent and widely used in VC to extract content information, is chosen as a strong baseline for SSL features. The objective evaluation and subjective evaluation both show models taking SSL feature CPC as both source and target features outperforms that taking PPG as source feature, suggesting that SSL features have great potential in improving VC.

An Exemplar Selection Algorithm for Native-Nonnative Voice Conversion

Christopher Liberatore, Ricardo Gutierrez-Osuna; Texas A&M University, USA

Tue-A-V-6-8, Time: 13:30

We present an algorithm for selecting exemplars for native-tononnative voice conversion (VC) using a Sparse, Anchor-Based Representation of speech (SABR). The algorithm uses phoneme labels and clustering to learn optimal exemplars when source and target speakers are affected by poor time alignment, as is common in in native-to-nonnative voice conversion. We evaluate the method on speech from the ARCTIC and L2-ARCTIC corpora and compare it to a baseline exemplar-based VC algorithm. The proposed algorithm significantly improves synthesis quality and more than doubles that of a baseline exemplar-based VC system while using two orders of magnitude fewer atoms. Additionally, the proposed algorithm significantly reduces the VC error and improves the synthesis quality as compared to unoptimized SABR models. We discuss the implications of both optimization algorithms for SABR and broader exemplar-based VC systems.Index terms should be included as shown below.

Adversarially Learning Disentangled Speech Representations for Robust Multi-Factor Voice Conversion

*Jie Wang*¹, *Jingbei Li*¹, *Xintao Zhao*¹, *Zhiyong Wu*¹, *Shiyin Kang*², *Helen Meng*¹; ¹*Tsinghua University, China*; ²*Huya, China* Tue-A-V-6-9, Time: 13:30

Factorizing speech as disentangled speech representations is vital to achieve highly controllable style transfer in voice conversion (VC). Conventional speech representation learning methods in VC only factorize speech as speaker and content, lacking controllability on other prosody-related factors. State-of-the-art speech representation learning methods for more speech factors are using primary disentangle algorithms such as random resampling and ad-hoc bottleneck layer size adjustment, which however is hard to ensure robust speech representation disentanglement. To increase the robustness of highly controllable style transfer on multiple factors in VC, we propose a disentangled speech representation learning framework based on adversarial learning. Four speech representations characterizing content, timbre, rhythm and pitch are extracted, and further disentangled by an adversarial Mask-And-Predict (MAP) network inspired by BERT. The adversarial network is used to minimize the correlations between the speech representations, by randomly masking and predicting one of the representations from the others. Experimental results show that the proposed framework significantly improves the robustness of VC on multiple factors by increasing the speech quality MOS from 2.79 to 3.30 and decreasing the MCD from 3.89 to 3.58.

Many-to-Many Voice Conversion Based Feature Disentanglement Using Variational Autoencoder

*Manh Luong*¹, *Viet Anh Tran*²; ¹*VinAI Research, Vietnam;* ²*Deezer, France* Tue-A-V-6-10, Time: 13:30

Voice conversion is a challenging task which transforms the voice characteristics of a source speaker to a target speaker without changing linguistic content. Recently, there have been many works on many-to-many Voice Conversion (VC) based on Variational Autoencoder (VAEs) achieving good results, however, these methods lack the ability to disentangle speaker identity and linguistic content to achieve good performance on unseen speaker's scenarios. In this paper, we propose a new method based on feature disentanglement to tackle many-to-many voice conversion. The method has the capability to disentangle speaker identity and linguistic content from utterances, it can convert from many source speakers to many target speakers with a single autoencoder network. Moreover, it naturally deals with the unseen target speaker's scenarios. We perform both objective and subjective evaluations to show the competitive performance of our proposed method compared with other state-of-the-art models in terms of naturalness and target speaker similarity.

Tue-A-SS-1: Privacy-Preserving Machine Learning for Audio & Speech Processing

13:30-15:30, Tuesday 31 August 2021 Chairs: Emmanuel Vincent and Harishchandra Dubey

Privacy-Preserving Voice Anti-Spoofing Using Secure Multi-Party Computation

Oubaïda Chouchane, Baptiste Brossier, Jorge Esteban Gamboa Gamboa, Thomas Lardy, Hemlata Tak, Orhan Ermis, Madhu R. Kamble, Jose Patino, Nicholas Evans, Melek Önen, Massimiliano Todisco; EURECOM, France Tue-A-SS-1-1, Time: 13:30

In recent years the automatic speaker verification (ASV) community has grappled with vulnerabilities to spoofing attacks whereby fraudsters masquerade as enrolled subjects to provoke illegitimate accepts. Countermeasures have hence been developed to protect ASV systems from such attacks. Given that recordings of speech contain potentially sensitive information, any system operating upon them, including spoofing countermeasures, must have provisions for privacy preservation. While privacy enhancing technologies such as Homomorphic Encryption or Secure Multi-Party Computation (MPC) are effective in preserving privacy, these tend to impact upon computational capacity and computational precision, while no available spoofing countermeasures preserve privacy. This paper reports the first solution based upon the combination of shallow

neural networks with secure MPC. Experiments performed using the ASVspoof 2019 logical access database show that the proposed solution is not only computationally efficient, but that it also improves upon the performance of the ASVspoof baseline countermeasure, all while preserving privacy.

Configurable Privacy-Preserving Automatic Speech Recognition

Ranya Aloufi, Hamed Haddadi, David Boyle; Imperial College London, UK

Tue-A-SS-1-2, Time: 13:30

Voice assistive technologies have given rise to far-reaching privacy and security concerns. In this paper we investigate whether modular automatic speech recognition (ASR) can improve privacy in voice assistive systems by combining independently trained separation, recognition, and discretization modules to design configurable privacy-preserving ASR systems. We evaluate privacy concerns and the effects of applying various state-of-the-art techniques at each stage of the system, and report results using task-specific metrics (i.e., WER, ABX, and accuracy). We show that overlapping speech inputs to ASR systems present further privacy concerns, and how these may be mitigated using speech separation and optimization techniques. Our discretization module is shown to minimize paralinguistics privacy leakage from ASR acoustic models to levels commensurate with random guessing. We show that voice privacy can be configurable, and argue this presents new opportunities for privacy-preserving applications incorporating ASR.

Adjunct-Emeritus Distillation for Semi-Supervised Language Model Adaptation

Scott Novotney, Yile Gu, Ivan Bulyko; Amazon, USA Tue-A-SS-1-3, Time: 13:30

To improve customer privacy, commercial speech applications are reducing human transcription of customer data. This has a negative impact on language model training due to a smaller amount of in-domain transcripts. Prior work demonstrated that training on automated transcripts alone provides modest gains due to reinforcement of recognition errors. We consider a new condition, where a model trained on historical human transcripts, but not the transcripts themselves, are available to us. To overcome temporal drift in vocabulary and topics, we propose a novel extension of knowledge distillation, adjunct-emeritus distillation where two imperfect teachers jointly train a student model. We conduct experiments on an English voice assistant domain and simulate a one year gap in human transcription. Unlike fine-tuning, our approach is architecture agnostic and achieves a 14% relative reduction in perplexity over the baseline approach of freezing model development and improves over the baseline of knowledge distillation.

Communication-Efficient Agnostic Federated Averaging

Jae Ro, Mingqing Chen, Rajiv Mathews, Mehryar Mohri, Ananda Theertha Suresh; Google, USA Tue-A-SS-1-4, Time: 13:30

In distributed learning settings such as federated learning, the training algorithm can be potentially biased towards different clients. [1] proposed a domain-agnostic learning algorithm, where the model is optimized for any target distribution formed by a mixture of the client distributions in order to overcome this bias. They further proposed an algorithm for the cross-silo federated learning setting, where the number of clients is small. We consider this problem in the cross-device setting, where the number of clients is much larger. We propose a communication-efficient distributed algorithm called AGNOSTIC FEDERATED AVERAGING (or AGNOSTICFEDAVG) to minimize the domain-agnostic objective proposed in [1], which is amenable to other private mechanisms such as secure aggregation. We highlight two types of naturally occurring domains in federated learning and argue that AGNOSTICFEDAVG performs well on both. To demonstrate the practical effectiveness of AGNOSTICFEDAVG, we report positive results for large-scale language modeling tasks in both simulation and live experiments, where the latter involves training language models for Spanish virtual keyboard for millions of user devices.

Privacy-Preserving Feature Extraction for Cloud-Based Wake Word Verification

Timm Koppelmann, Alexandru Nelus, Lea Schönherr, Dorothea Kolossa, Rainer Martin; Ruhr-Universität Bochum, Germany Tue-A-SS-1-5, Time: 13:30

Wake word detection and verification systems often involve a local, on-device wake word detector and a cloud-based verification node. In such systems, the audio representation sent to the cloud-based server may exhibit sensitive information that might be intercepted by an eavesdropper. To improve privacy of cloud-based wake word verification (WWV) systems, we propose to use a privacy-preserving feature representation that minimizes the automatic speech recognition (ASR) capability of a potential attacker. The proposed approach employs an adversarial training schedule that aims to minimize an attacker's word error rate (WER) while maintaining a high WWV performance. To this end, we apply an adaptive weighting factor in the combined loss function to control the balance between minimizing the WWV loss and maximizing the ASR loss. We show that the proposed training method significantly reduces possible privacy risks while maintaining a strong WWV performance.

PATE-AAE: Incorporating Adversarial Autoencoder into Private Aggregation of Teacher Ensembles for **Spoken Command Classification**

Chao-Han Huck Yang, Sabato Marco Siniscalchi, Chin-Hui Lee; Georgia Tech, USA Tue-A-SS-1-6, Time: 13:30

We propose using an adversarial autoencoder (AAE) to replace generative adversarial network (GAN) in private aggregation of teacher ensembles (PATE), a solution for ensuring differential privacy in speech applications. The AAE architecture allows us to obtain good synthetic speech leveraging upon a discriminative training of latent vectors. Such synthetic speech is used to build a privacy-preserving classifier when non-sensitive data is not sufficiently available in the public domain. This classifier follows the PATE scheme that uses an ensemble of noisy outputs to label the synthetic samples and guarantee ε -differential privacy (DP) on its derived classifiers. Our proposed framework thus consists of an AAE-based generator and a PATE-based classifier (PATE-AAE). Evaluated on the Google Speech Commands Dataset Version II, the proposed PATE-AAE improves the average classification accuracy by +2.11% and +6.60%, respectively, when compared with alternative privacy-preserving solutions, namely PATE-GAN and DP-GAN, while maintaining a strong level of privacy target at $\varepsilon = 0.01$ with a fixed $\delta = 10^{-5}$.

Continual Learning for Fake Audio Detection

Haoxin Ma, Jiangyan Yi, Jianhua Tao, Ye Bai, Zhengkun Tian, Chenglong Wang; CAS, China Tue-A-SS-1-7, Time: 13:30

Fake audio attack becomes a major threat to the speaker verification Although current detection approaches have achieved system.

promising results on dataset-specific scenarios, they encounter difficulties on unseen spoofing data. Fine-tuning and retraining from scratch have been applied to incorporate new data. However, fine-tuning leads to performance degradation on previous data. Retraining takes a lot of time and computation resources. Besides, previous data are unavailable due to privacy in some situations. To solve the above problems, this paper proposes detecting fake without forgetting, a continual-learning-based method, to make the model learn new spoofing attacks incrementally. A knowledge distillation loss is introduced to loss function to preserve the memory of original model. Supposing the distribution of genuine voice is consistent among different scenarios, an extra embedding similarity loss is used as another constraint to further do a positive sample alignment. Experiments are conducted on the ASVspoof2019 dataset. The results show that our proposed method outperforms fine-tuning by the relative reduction of average equal error rate up to 81.62%.

Evaluating the Vulnerability of End-to-End Automatic Speech Recognition Models to Membership Inference Attacks

Muhammad A. Shah, Joseph Szurley, Markus Mueller, Athanasios Mouchtaris, Jasha Droppo; Amazon, USA Tue-A-SS-1-8, Time: 13:30

Recent studies have shown that it may be possible to determine if a machine learning model was trained on a given data sample, using Membership Inference Attacks (MIA). In this paper we evaluate the vulnerability of state-of-the-art speech recognition models to MIA under black-box access. Using models trained with standard methods and public datasets, we demonstrate that without any knowledge of the target model's parameters or training data a MIA can successfully infer membership with precision and recall more than 60%. Furthermore, for utterances from about 39% of the speakers the precision is more than 75%, indicating that training data membership can be inferred more precisely for some speakers than others. While strong regularization reduces the overall accuracy of MIA to almost 50%, the attacker can still infer membership for utterances from 25% of the speakers with high precision. These results indicate that (1) speaker-level MIA success should be reported, along with overall accuracy, to provide a holistic view of the model's vulnerability and (2) conventional regularization is an inadequate defense against MIA.We believe that the insights gleaned from this study can direct future work towards more effective defenses.

SynthASR: Unlocking Synthetic Data for Speech Recognition

Amin Fazel¹, Wei Yang¹, Yulan Liu², Roberto Barra-Chicote², Yixiong Meng¹, Roland Maas¹, Jasha Droppo¹; ¹Amazon, USA; ²Amazon, UK Tue-A-SS-1-9, Time: 13:30

End-to-end (E2E) automatic speech recognition (ASR) models have recently demonstrated superior performance over the traditional hybrid ASR models. Training an E2E ASR model requires a large amount of data which is not only expensive but may also raise dependency on production data. At the same time, synthetic speech generated by the state-of-the-art text-to-speech (TTS) engines has advanced to near-human naturalness. In this work, we propose to utilize synthetic speech for ASR training (SynthASR) in applications where data is sparse or hard to get for ASR model training. In addition, we apply continual learning with a novel multi-stage training strategy to address catastrophic forgetting, achieved by a mix of weighted multistyle training, data augmentation, encoder freezing, and parameter regularization. In our experiments conducted on in-house datasets for a new application of recognizing medication names, training ASR

NOTES

RNN-T models with synthetic audio via the proposed multi-stage training improved the recognition performance on new application by more than 65% relative, without degradation on existing general applications. Our observations show that SynthASR holds great promise in training the state-of-the-art large-scale E2E ASR models for new applications while reducing the costs and dependency on production data.

Tue-A-SS-2: The First DiCOVA Challenge: Diagnosis of COVID-19 Using Acoustics

Room C, 13:30–15:30, Tuesday 31 August 2021 Chairs: Mark Hasegawa-Johnson and Neeraj Sharma

Introduction

Time: 13:30

Short Presentations of Papers Time: 13:35

DiCOVA Challenge: Dataset, Task, and Baseline System for COVID-19 Diagnosis Using Acoustics

Ananya Muguli¹, Lancelot Pinto², Nirmala R.¹, Neeraj Sharma¹, Prashant Krishnan¹, Prasanta Kumar Ghosh¹, Rohit Kumar¹, Shrirama Bhat³, Srikanth Raj Chetupalli¹, Sriram Ganapathy¹, Shreyas Ramoji¹, Viral Nanda²; ¹Indian Institute of Science, India; ²P.D. Hinduja Hospital, India; ³KMC Hospital, India Tue-A-SS-2-1, Time: 14:30

The DiCOVA challenge aims at accelerating research in diagnosing COVID-19 using acoustics (DiCOVA), a topic at the intersection of speech and audio processing, respiratory health diagnosis, and machine learning. This challenge is an open call for researchers to analyze a dataset of sound recordings, collected from COVID-19 infected and non-COVID-19 individuals, for a two-class classification. These recordings were collected via crowdsourcing from multiple countries, through a website application. The challenge features two tracks, one focusing on cough sounds, and the other on using a collection of breath, sustained vowel phonation, and number counting speech recordings. In this paper, we introduce the challenge and provide a detailed description of the task, and present a baseline system for the task.

PANACEA Cough Sound-Based Diagnosis of COVID-19 for the DiCOVA 2021 Challenge

Madhu R. Kamble¹, Jose A. Gonzalez-Lopez², Teresa Grau³, Juan M. Espin³, Lorenzo Cascioli¹, Yiqing Huang¹, Alejandro Gomez-Alanis², Jose Patino¹, Roberto Font³, Antonio M. Peinado², Angel M. Gomez², Nicholas Evans¹, Maria A. Zuluaga¹, Massimiliano Todisco¹; ¹EURECOM, France; ²Universidad de Granada, Spain; ³Biometric Vox, Spain Tue-A-SS-2-2, Time: 14:30

The COVID-19 pandemic has led to the saturation of public health services worldwide. In this scenario, the early diagnosis of SARS-Cov-2 infections can help to stop or slow the spread of the virus and to manage the demand upon health services. This is especially important when resources are also being stretched by heightened demand linked to other seasonal diseases, such as the flu. In this context, the organisers of the DiCOVA 2021 challenge have collected a database with the aim of diagnosing COVID-19 through the use of coughing audio samples. This work presents the details of the automatic system for COVID-19 detection from cough recordings presented by team PANACEA. This team consists of researchers from two European academic institutions and one company: EURECOM (France), University of Granada (Spain), and Biometric Vox S.L. (Spain). We developed several systems based on established signal processing and machine learning methods. Our best system employs a Teager energy operator cepstral coefficients (TECCs) based front-end and Light gradient boosting machine (LightGBM) back-end. The AUC obtained by this system on the test set is 76.31% which corresponds to a 10% improvement over the official baseline.

Recognising Covid-19 from Coughing Using Ensembles of SVMs and LSTMs with Handcrafted and Deep Audio Features

Vincent Karas, Björn W. Schuller; Universität Augsburg, Germany

Tue-A-SS-2-3, Time: 14:30

As the Covid-19 pandemic continues, digital health solutions can provide valuable insights and assist in diagnosis and prevention. Since the disease affects the respiratory system, it is hypothesised that sound formation is changed, and thus, an infection can be automatically recognised through audio analysis. We present an ensemble learning approach used in our entry to Track 1 of the Di-COVA 2021 Challenge, which aims at binary classification of Covid-19 infection on a crowd-sourced dataset of 1 040 cough sounds. Our system is based on a combination of handcrafted features for paralinguistics with deep feature extraction from spectrograms using pre-trained CNNs. We extract features both at segment level and with a sliding window approach, and process them with SVMs and LSTMs, respectively. We then perform least-squares weighted late fusion of our classifiers. Our system surpasses the challenge baseline, with a ROC-AUC on the test set of 78.18%.

Detecting COVID-19 from Audio Recording of Coughs Using Random Forests and Support Vector Machines

Isabella Södergren, Maryam Pahlavan Nodeh, Prakash Chandra Chhipa, Konstantina Nikolaidou, György Kovács; Luleå University of Technology, Sweden Tue-A-SS-2-4, Time: 14:30

The detection of COVID-19 is and will remain in the foreseeable future a crucial challenge, making the development of tools for the task important. One possible approach, on the confines of speech and audio processing, is detecting potential COVID-19 cases based on cough sounds. We propose a simple, yet robust method based on the well-known ComParE 2016 feature set, and two classical machine learning models, namely Random Forests, and Support Vector Machines (SVMs). Furthermore, we combine the two methods, by calculating the weighted average of their predictions. Our results in the DiCOVA challenge show that this simple approach leads to a robust solution while producing competitive results. Based on the Area Under the Receiver Operating Characteristic Curve (AUC ROC) score, both classical machine learning methods we applied markedly outperform the baseline provided by the challenge organisers. Moreover, their combination attains an AUC ROC score of 85.21, positioning us at fourth place on the leaderboard (where the second team attained a similar, 85.43 score). Here, we would describe this system in more detail, and analyse the resulting models, drawing conclusions, and determining future work directions.

Diagnosis of COVID-19 Using Auditory Acoustic Cues

Rohan Kumar Das, Maulik Madhavi, Haizhou Li; NUS,

Singapore

Tue-A-SS-2-5, Time: 14:30

COVID-19 can be pre-screened based on symptoms and confirmed using other laboratory tests. The cough or speech from patients are also studied in the recent time for detection of COVID-19 as they are indicators of change in anatomy and physiology of the respiratory system. Along this direction, the diagnosis of COVID-19 using acoustics (DiCOVA) challenge aims to promote such research by releasing publicly available cough/speech corpus. We participated in the Track-1 of the challenge, which deals with COVID-19 detection using cough sounds from individuals. In this challenge, we use a few novel auditory acoustic cues based on long-term transform, equivalent rectangular bandwidth spectrum and gammatone filterbank. We evaluate these representations using logistic regression, random forest and multilayer perceptron classifiers for detection of COVID-19. On the blind test set, we obtain an area under the ROC curve (AUC) of 83.49% for the best system submitted to the challenge. It is worth noting that the submitted system ranked among the top few systems on the leaderboard and outperformed the challenge baseline by a large margin.

Classification of COVID-19 from Cough Using Autoregressive Predictive Coding Pretraining and Spectral Data Augmentation

John Harvill¹, Yash R. Wani², Mark Hasegawa-Johnson¹, Narendra Ahuja¹, David Beiser², David Chestek³; ¹University of Illinois at Urbana-Champaign, USA; ²University of Chicago, USA; ³University of Illinois at Chicago, USA Tue-A-SS-2-6, Time: 14:30

Serum and saliva-based testing methods have been crucial to slowing the COVID-19 pandemic, yet have been limited by slow throughput and cost. A system able to determine COVID-19 status from cough sounds alone would provide a low cost, rapid, and remote alternative to current testing methods. We explore the applicability of recent techniques such as pre-training and spectral augmentation in improving the performance of a neural cough classification system. We use Autoregressive Predictive Coding (APC) to pre-train a unidirectional LSTM on the COUGHVID dataset. We then generate our final model by fine-tuning added BLSTM layers on the DiCOVA challenge dataset. We perform various ablation studies to see how each component impacts performance and improves generalization with a small dataset. Our final system achieves an AUC of 85.35 and places third out of 29 entries in the DiCOVA challenge.

The DiCOVA 2021 Challenge — An Encoder-Decoder Approach for COVID-19 Recognition from Coughing Audio

*Gauri Deshpande*¹, *Björn W. Schuller*²; ¹*TCS, India;* ²*Universität Augsburg, Germany* Tue-A-SS-2-7, Time: 14:30

This paper presents the automatic recognition of COVID-19 from coughing. In particular, it describes our contribution to the DiCOVA challenge — Track 1, which addresses such cough sound analysis for COVID-19 detection. Pathologically, the effects of a COVID-19 infection on the respiratory system and on breathing patterns are known. We demonstrate the use of breathing patterns of the cough audio signal in identifying the COVID-19 status. Breathing patterns of the subset of the UCL Speech Breath Monitoring (UCL-SBM) database.

This database provides speech recordings of the participants while their breathing values are captured by a respiratory belt. We use an encoder-decoder architecture. The encoder encodes the audio signal into breathing patterns and the decoder decodes the COVID-19 status for the corresponding breathing patterns using an attention mechanism. The encoder uses a pre-trained model which predicts breathing patterns from the speech signal, and transfers the learned patterns to cough audio signals.

With this architecture, we achieve an AUC of 64.42% on the evaluation set of Track 1.

COVID-19 Detection from Spectral Features on the DiCOVA Dataset

Kotra Venkata Sai Ritwik, Shareef Babu Kalluri, Deepu Vijayasenan; NITK Surathkal, India Tue-A-SS-2-8, Time: 14:30

In this paper we investigate the cues of COVID-19 on sustained phonation of Vowel-/i/, deep breathing and number counting data of the DiCOVA dataset. We use an ensemble of classifiers trained on different features, namely, super-vectors, formants, harmonics and MFCC features. We fit a two-class Weighted SVM classifier to separate the COVID-19 audio from Non-COVID-19 audio. Weighted penalties help mitigate the challenge of class imbalance in the dataset. The results are reported on the stationary (breathing, Vowel-/i/) and non-stationary (counting data) data using individual and combination of features on each type of utterance. We find that the Formant information plays a crucial role in classification. The proposed system resulted in an AUC score of 0.734 for cross validation, and 0.717 for evaluation dataset.

Cough-Based COVID-19 Detection with Contextual Attention Convolutional Neural Networks and Gender Information

Adria Mallol-Ragolta¹, Helena Cuesta², Emilia Gómez², Björn W. Schuller¹; ¹Universität Augsburg, Germany; ²Universitat Pompeu Fabra, Spain Tue-A-SS-2-9, Time: 14:30

The aim of this contribution is to automatically detect COVID-19 patients by analysing the acoustic information embedded in coughs. COVID-19 affects the respiratory system, and, consequently, respiratory-related signals have the potential to contain salient information for the task at hand. We focus on analysing the spectrogram representations of cough samples with the aim to investigate whether COVID-19 alters the frequency content of these signals. Furthermore, this work also assesses the impact of gender in the automatic detection of COVID-19. To extract deep-learnt representations of the spectrograms, we compare the performance of a cough-specific, and a Resnet18 pre-trained Convolutional Neural Network (CNN). Additionally, our approach explores the use of contextual attention, so the model can learn to highlight the most relevant deep-learnt features extracted by the CNN. We conduct our experiments on the dataset released for the Cough Sound Track of the DICOVA 2021 Challenge. The best performance on the test set is obtained using the Resnet18 pre-trained CNN with contextual attention, which scored an Area Under the Curve (AUC) of 70.91% at 80% sensitivity.

Contrastive Learning of Cough Descriptors for Automatic COVID-19 Preliminary Diagnosis

Swapnil Bhosale, Upasana Tiwari, Rupayan Chakraborty, Sunil Kumar Kopparapu; TCS, India Tue-A-SS-2-10, Time: 14:30

Cough sounds as a descriptor have been used for detecting various respiratory ailments based on its intensity, duration of intermediate phase between two cough sounds, repetitions, dryness etc. However, COVID-19 diagnosis using only cough sounds is challenging because of cough being a common symptom among many non COVID-19 health diseases and inherent data imbalance within the available datasets. As one of the approach in this direction, we explore the robustness of multi-domain representation by performing the early fusion over a wide set of temporal, spectral and tempo-spectral handcrafted features, followed by training a Support Vector Machine (SVM) classifier. In our second approach, using a contrastive loss function we learn a latent space from Mel Filter Cepstral Coefficients (MFCCs) where representations belonging to samples having similar cough characteristics are closer. This helps learn representations for the highly varied COVID-negative class (healthy and symptomatic COVID-negative), by learning multiple smaller clusters. Using only the DiCOVA data, multi-domain features yields an absolute improvement of 0.74% and 1.07%, whereas our second approach shows an improvement of 2.09% and 3.98%, over the blind test and validation set, respectively, when compared with challenge baseline.

Investigating Feature Selection and Explainability for COVID-19 Diagnostics from Cough Sounds

*Flavio Avila*¹, *Amir H. Poorjam*¹, *Deepak Mittal*¹, *Charles Dognin*¹, *Ananya Muguli*², *Rohit Kumar*², *Srikanth Raj Chetupalli*², *Sriram Ganapathy*², *Maneesh Singh*¹; ¹*Verisk Analytics, USA*; ²*Indian Institute of Science, India*

Tue-A-SS-2-11, Time: 14:30

In this paper, we propose an approach to automatically classify COVID-19 and non-COVID-19 cough samples based on the combination of both feature engineering and deep learning models. In the feature engineering approach, we develop a support vector machine classifier over high dimensional (6373D) space of acoustic features. In the deep learning-based approach, on the other hand, we apply a convolutional neural network trained on the log-mel spectrograms. These two methodologically diverse models are then combined by fusing the probability scores of the models. The proposed system, which ranked 9th on the 2021 Diagnosing COVID-19 using Acoustics (DiCOVA) challenge leaderboard, obtained an area under the receiver operating characteristic curve (AUC) of 0.81 on the blind test data set, which is a 10.9% absolute improvement compared to the baseline. Moreover, we analyze the explainability of the deep learning-based model when detecting COVID-19 from cough signals.

Tue-A-S&T-1 : Show and Tell 1

13:30-15:30, Tuesday 31 August 2021

Application for Detecting Depression, Parkinson's Disease and Dysphonic Speech

Gábor Kiss, Dávid Sztahó, Miklós Gábriel Tulics; BME, Hungary

Tue-A-S&T-1-1, Time: 13:30

In this Show&Tell presentation we demonstrate an application that is able to assess a voice sample according to three different voice disorders: depression, Parkinson's disease and dysphonic speech. Affection probability of each disorder is analyzed along with their severity estimation. Although the acoustic models (support vector machine and regression models) are trained on Hungarian voice samples, English samples can also be utilized for assessment. The results are displayed by as pie chart for probabilities and separate severity scores. The input of the application is a read text with a fixed linguistic content. It is possible to load a pre-recorded voice sample or create a live recording. The developed system could evaluate a speaker's voice sample, assisting medical staff.

Beey: More Than a Speech-to-Text Editor

Lenka Weingartová, Veronika Volná, Ewa Balejová; NEWTON Technologies, Czechia Tue-A-S&T-1-2, Time: 13:30

We present Beey, a newly developed web-based multimedia platform for producing Automatic Speech Recognition (ASR) and editing its output. In addition to ASR, Beey employs modules for speaker diarization and identification, text formatting, automatic punctuation insertion, subtitling, automatic translation, transcription of stream and more.

The platform and its development are focused on user experience and fast document creation. Our aim is to transfer research results in the field of speech recognition and signal processing into practice and enable Beey's users to make their production processes faster and cheaper by minimizing human effort and costs.

Downsizing of Vocal-Tract Models to Line up Variations and Reduce Manufacturing Costs

Takayuki Arai; Sophia University, Japan Tue-A-S&T-1-3, Time: 13:30

Demonstrating vowel production with physical models of the human vocal tract is a part of intuitive education in speech science. The adult male vocal tract was most often used as a model in the past because of the limited availability of physical models, but discussions on different vocal tract sizes were ongoing. Therefore, we focused on downsizing the vocal-tract models in this study, especially the straight models. We reduced the cross-sectional area function for the sliding three-tube model (including the total length) to female adult and child sizes. Furthermore, we created fixed straight models of similar dimensions for the five Japanese vowels. We found that the intelligibility of each model was preserved as long as the ratios of the cross-sectional areas were maintained even if the cross-sections were less than the average human sizes. This indicates that we can reduce the cost of manufacturing the models, as cost is typically a barrier when the models are used for pedagogical purposes.

ROXANNE Research Platform: Automate Criminal Investigations

Maël Fabien¹, Shantipriya Parida¹, Petr Motlicek¹, Dawei Zhu², Aravind Krishnan², Hoang H. Nguyen³; ¹Idiap Research Institute, Switzerland; ²Universität des Saarlandes, Germany; ³Leibniz Universität Hannover, Germany

Tue-A-S&T-1-4, Time: 13:30

Criminal investigations require manual intervention of several investigators and translators. However, the amount and the diversity of the data collected raises many challenges, and cross-border investigations against organized crime can quickly impossible to handle. We developed ROXANNE Research platform, an all-in-one platform which processes intercepted phone calls, runs state-of-the-art components such as speaker identification, automatic speech recognition or named entity detection, and builds a knowledge graph of the extracted information. Our aim for this work is to do a first step in the direction of an open research platform combining speech, text, and video processing algorithms with criminal network analysis for combating organized crime.

The LIUM Human Active Correction Platform for Speaker Diarization

Alexandre Flucha, Anthony Larcher, Ambuj Mehrish, Sylvain Meignier, Florian Plaut, Nicolas Poupon, Yevhenii Prokopalo, Adrien Puertolas, Meysam Shamsi, Marie Tahon; LIUM (EA 4023), France

Tue-A-S&T-1-5, Time: 13:30

We developed a human assisted speaker diarization platform that enables a human annotator to correct the output of any speaker diarization system by providing a graphical view of the diarization segmentation and clustering steps while guiding the human annotator to optimize the correction process and easily improve the resulting diarization.

On-Device Streaming Transformer-Based End-to-End Speech Recognition

Yoo Rhee Oh, Kiyoung Park; ETRI, Korea Tue-A-S&T-1-6, Time: 13:30

This work is the first attempt to run streaming Transformer-based end-to-end speech recognition on embedded scale IoT systems. Recently there are many researches on online Transformer-based speech recognition such as a contextual block encoder [1] and a block-wise synchronous beam search [2]. Based on them we designed a novel fully-streaming end-to-end speech recognition method using Transformer. By efficiently utilizing a connectionist temporal classification network to detect symbol and sentence boundaries, we make decoder in streaming manner. Moreover, by using the optimized model structure, the proposed method could be deployed on a low-power edge device such as Raspberry Pi 4B with the high accuracy and the small latency. With the experiments with Librispeech corpus, the methods achieved word error rates of 3.76% and 9.25% respectively. Also the recognition speed is measured in two aspects; the real-time factor and the user perceived latency. The system is evaluated to have 0.84 xRT and the average latency of 0.75±0.62 seconds on Raspberry Pi 4B.

Advanced Semi-Blind Speaker Extraction and Tracking Implemented in Experimental Device with Revolving Dense Microphone Array

J. Čmejla, T. Kounovský, J. Janský, Jiri Malek, M. Rozkovec, Z. Koldovský; Technical University of Liberec, Czechia

Tue-A-S&T-1-7, Time: 13:30

We present a new device for speaker extraction and physical tracking and demonstrate its use in real conditions. The device is equipped with a dense planar array consisting of 64 microphones mounted on a rotating platform. State-of-the-art blind source extraction algorithms controlled by x-vector piloting are used to extract the desired speaker, which is being tracked by the rotating microphone array. The audience will experience the functionality of the device and the potential of the blind algorithms to extract the speaker from multi-source noisy recordings in a live situation.

Tue-Keynote: Keynote 1: Hermann Ney

Room A+B, 17:00–18:00, Tuesday 31 August 2021 Chairs: TBD

Forty Years of Speech and Language Processing: From Bayes Decision Rule to Deep Learning

Hermann Ney; RWTH Aachen University, Germany Tue-Keynote, Time: 17:00

When research on automatic speech recognition started, the statistical (or data-driven) approach was associated with methods like Bayes decision rule, hidden Markov models, Gaussian models and expectation-maximization algorithm. Later extensions included discriminative training and hybrid hidden Markov models using multi-layer perceptrons and recurrent neural networks. Some of the methods originally developed for speech recognition turned out to be seminal for other language processing tasks like machine translation, handwritten character recognition and sign language processing. Today's research on speech and language processing is dominated by deep learning, which is typically identified with methods like attention modelling, sequence-to-sequence processing and end-to-end processing.

In this talk, I will present my personal view of the historical developments of research on speech and language processing. I will put particular emphasis on the framework of Bayes decision rule and on the question of how the various approaches developed fit into this framework.

Tue-E-O-1: ASR Technologies and Systems

Room A+B, 19:00-21:00, Tuesday 31 August 2021 Chairs: Louis ten Bosch and Petr Červa

Information Retrieval for ZeroSpeech 2021: The Submission by University of Wroclaw

Jan Chorowski¹, Grzegorz Ciesielski¹, Jarosław Dzikowski¹, Adrian Łańcucki², Ricard Marxer³, Mateusz Opala¹, Piotr Pusz¹, Paweł Rychlikowski¹, Michał Stypułkowski¹; ¹University of Wrocław, Poland; ²NVIDIA, Poland; ³LIS (UMR 7020), France Tue-E-O-1-1, Time: 19:00

We present a number of low-resource approaches to the tasks of the Zero Resource Speech Challenge 2021. We build on the unsupervised representations of speech proposed by the organizers as a baseline, derived from CPC and clustered with the k-means algorithm. We demonstrate that simple methods of refining those representations can narrow the gap, or even improve upon the solutions which use a high computational budget. The results lead to the conclusion that the CPC-derived representations are still too noisy for training language models, but stable enough for simpler forms of pattern matching and retrieval.

Aligned Contrastive Predictive Coding

Jan Chorowski¹, Grzegorz Ciesielski¹, Jarosław Dzikowski¹, Adrian Łańcucki², Ricard Marxer³, Mateusz Opala¹, Piotr Pusz¹, Paweł Rychlikowski¹, Michał Stypułkowski¹; ¹University of Wrocław, Poland; ²NVIDIA, Poland; ³LIS (UMR 7020), France Tue-E-O-1-2, Time: 19:20

We investigate the possibility of forcing a self-supervised model

NOTES

trained using a contrastive predictive loss, to extract slowly varying latent representations. Rather than producing individual predictions for each of the future representations, the model emits a sequence of predictions shorter than the sequence of upcoming representations to which they will be aligned. In this way, the prediction network solves a simpler task of predicting the next symbols, but not their exact timing, while the encoding network is trained to produce piece-wise constant latent codes. We evaluate the model on a speech coding task and demonstrate that the proposed Aligned Contrastive Predictive Coding (ACPC) leads to higher linear phone prediction accuracy and lower ABX error rates, while being slightly faster to train due to the reduced number of prediction heads.

Neural Text Denormalization for Speech Transcripts

Benjamin Suter, Josef Novak; Spitch, Switzerland Tue-E-O-1-3, Time: 19:40

This paper presents a simple sequence-to-sequence approach to restore standard orthography in raw, normalized speech transcripts, including insertion of punctuation marks, prediction of capitalization, restoration of numeric forms, formatting of dates and times, and other, fully data-driven adjustments. We further describe our method to generate synthetic parallel training data, and explore suitable performance metrics, which we align with human judgment through subjective MOS-like evaluations.

Our models for English, Russian, and German have a word error rate of 6.36%, 4.88%, and 5.23%, respectively. We focus on simplicity and reproducibility, make our framework available under a BSD license, and share our base models for English and Russian.

Fearless Steps Challenge Phase-3 (FSC P3): Advancing SLT for Unseen Channel and Mission Data Across NASA Apollo Audio

Aditya Joglekar¹, Seyed Omid Sadjadi², Meena Chandra-Shekar¹, Christopher Cieri³, John H.L. Hansen¹; ¹University of Texas at Dallas, USA; ²NIST, USA; ³University of Pennsylvania, USA Tue-E-O-1-4, Time: 20:00

The Fearless Steps Challenge (FSC) initiative was designed to host a series of progressively complex tasks to promote advanced speech research across naturalistic "Big Data" corpora. The Center for Robust Speech Systems at UT-Dallas in collaboration with the National Institute of Standards and Technology (NIST) and Linguistic Data Consortium (LDC) conducted Phase-3 of the FSC series (FSC P3), with a focus on motivating speech and language technology (SLT) system generalizability across channel and mission diversity under the same training conditions as in Phase-2. The FSC P3 introduced 10 hours of previously unseen channel audio from Apollo-11 and 5 hours of novel audio from Apollo-13 to be evaluated over both previously established and newly introduced SLT tasks with streamlined tracks. This paper presents an overview of the newly introduced conversational analysis tracks, Apollo-13 data, and analysis of system performance for matched and mismatched challenge conditions. We also discuss the Phase-3 challenge results, evolution of system performance across the three Phases, and next steps in the Challenge Series

Tue-E-O-2: Phonation and Voicing

Room C, 19:00-21:00, Tuesday 31 August 2021 Chairs: Bogdan Ludusan and Lenka Weingartová

Voice Quality in Verbal Irony: Electroglottographic Analyses of Ironic Utterances in Standard Austrian German

Hannah Leykum; Austrian Academy of Sciences, Austria Tue-E-O-2-1, Time: 19:00

When using verbal irony in interpersonal communication, paraverbal cues can reduce the risk of misunderstandings. Besides fundamental frequency, intensity and duration, speakers could use voice quality parameters to disambiguate between ironic and literal utterances. How these paraverbal cues are used to mark irony appears to be language- and/or culture-specific. Since the role of voice quality in ironic utterances has not yet been investigated in Austrian German, the present study addresses this issue. In addition to the acoustic signal, the vocal fold vibration is recorded via electroglottography (EGG). The detailed analysis of the EGG data as well as the acoustic data, provides insight into voice quality characteristics of ironic and literal realisations of short utterances. The analyses reveal that, in Standard Austrian German, some differences in voice quality exist between ironic and literal realisations of utterances: When being ironic, speakers' voices tend to be breathier, creakier or rougher. Differences are more pronounced in the older age group and in male speakers.

Synchronic Fortition in Five Romance Languages? A Large Corpus-Based Study of Word-Initial Devoicing

Mathilde Hutin¹, Yaru Wu¹, Adèle Jatteau², Ioana Vasilescu¹, Lori Lamel¹, Martine Adda-Decker¹; ¹LISN (UMR 9015), France; ²STL (UMR 8163), France Tue-E-O-2-2, Time: 19:20

Devoicing is a process whereby a voiced consonant such as /bdg/ is realized as voiceless [ptk]. Some theorists [1,2] propose that this phenomenon is an instance of fortition, or consonant strengthening, especially when it occurs word-initially. This study proposes an in-depth exploration of voicing alternations in word-initial position in five Romance languages (Portuguese, Spanish, French, Italian, Romanian) using large corpora (ca. 1000h of speech) and automatic alignment. Our results show that (i) there is initial devoicing in all languages, and (ii) this devoicing is conditioned by the preceding context. This allows the languages to be divided into those displaying (a) only phrase-initial fortition (Spanish), (b) phrase-initial and post-obstruent fortition (French, Romanian and possibly Italian) and (c) generalized word-initial fortition (Portuguese).

Glottal Stops in Upper Sorbian: A Data-Driven Approach

Ivan Kraljevski¹, Maria Paola Bissiri², Frank Duckhorn¹, Constanze Tschoepe¹, Matthias Wolff³; ¹Fraunhofer IKTS, Germany; ²Università dell'Insubria, Italy; ³Brandenburgische Technische Universität, Germany Tue-E-O-2-3, Time: 19:40

We present a data-driven approach for the quantitative analysis of glottal stops before word-initial vowels in Upper Sorbian, a West Slavic minority language spoken in Germany. Glottal stops are wordboundary markers and their detection can improve the performance of automatic speech recognition and speech synthesis systems.

We employed cross-language transfer using an acoustic model in

German to develop a forced-alignment method for the phonetic segmentation of a read-speech corpus in Upper Sorbian. The missing phonemic units were created by combining the existing phoneme models. In the forced-alignment procedure, the glottal stops were considered optional in front of word-initial vowels.

To investigate the influence of speaker type (males, females, and children) and vowel on the occurrence of glottal stops, binomial regression analysis with a generalized linear mixed model was performed. Results show that children glottalize word-initial vowels more frequently than adults, and that glottal stop occurrences are influenced by vowel quality.

Cue Interaction in the Perception of Prosodic Prominence: The Role of Voice Quality

Bogdan Ludusan¹, Petra Wagner¹, Marcin Włodarczak²; ¹Universität Bielefeld, Germany; ²Stockholm University,

Sweden

Tue-E-O-2-4, Time: 20:00

Voice quality is an important dimension in human communication, used to mark a variety of phenomena in speech, including prosodic prominence. Even though numerous studies have shown that speakers modify their voice quality parameters for marking prosodic prominence, the impact of these modifications on perceived prominence is less studied. Our investigation looks at the effect of a well-known measure of voice quality, cepstral peak prominence (CPP), on syllabic prominence ratings given by both naive and expert listeners. Employing read speech materials in German, we quantify the role of CPP alone and in combination with other acoustic cues marking prominence, namely intensity, duration and fundamental frequency. While CPP, by itself, had a significant effect on the perceived prominence for most of the listeners, when used in conjunction with the other cues, its impact was reduced. Moreover, when assessing the importance of each of these four cues for determining the perceived prominence score we found important individual variation, as well as differences between naive and expert listeners

Glottal Sounds in Korebaju

Jenifer Vega Rodriguez, Nathalie Vallée; GIPSA-lab (UMR 5216), France

Tue-E-O-2-5, Time: 20:20

Korebaju (ISO639-3: coe) [kòrèβàhí] is a tonal language spoken in the foothills of the Colombian Amazon. Three field surveys carried out between 2017 and 2019 with six native speakers (3 females and 3 males) from the same village provide a set of glottal productions at both phonetic and phonological levels. This study focuses on the four types of glottal units we have found in this language: A set of vowels /a²/, /e²/, /o²/, [i²] and [i²] including 3 phonemes; the glottal stop [?] and the consonant [*] transcribed and described as a *creaky voiced glottal approximant* by [1]. Both consonants occurred in intervocalic contexts and can be analyzed as a suprasegmental feature [constricted glottis] which marks the syllable onset. Finally, we have also found a clear and systematic burst which accompanies the release of the nasal consonants [m², n², μ^2]. No change was found in the EGG signal for these consonants suggesting an abrupt release of the aeroacoustic pressure.

Automatic Classification of Phonation Types in Spontaneous Speech: Towards a New Workflow for the Characterization of Speakers' Voice Quality

Anaïs Chanclu¹, Imen Ben Amor¹, Cédric Gendrot², Emmanuel Ferragne², Jean-François Bonastre¹; ¹LIA (EA 4128), France; ²LPP (UMR 7018), France Tue-E-O-2-6, Time: 20:40

Voice quality is known to be an important factor for the characterization of a speaker's voice, both in terms of physiological features (mainly laryngeal and supralaryngeal) and of the speaker's habits (sociolinguistic factors). This paper is devoted to one of the main components of voice quality: phonation type. It proposes neural representations of speech followed by a cascade of two binary neural network-based classifiers, one dedicated to the detection of modal and nonmodal vowels, and one for the classification of nonmodal vowels into creaky and breathy types. This approach is evaluated on the spontaneous part of the PTSVOX database, following an expert manual labelling of the data by phonation type. The results of the proposed classifiers reaches on average 85% accuracy at the frame-level and up to 95% accuracy at the segment-level. Further research is planned to generalize the classifiers on more contexts and speakers, and thus pave the way for a new workflow aimed at characterizing phonation types.

Tue-E-O-3: Health and Affect I

Room D, 19:00–21:00, Tuesday 31 August 2021 Chairs: László Tóth and Tanya Talkar

Measuring Voice Quality Parameters After Speaker Pseudonymization

Rob J.J.H. van Son; Netherlands Cancer Institute, The Netherlands

Tue-E-O-3-1, Time: 19:00

Collecting and sharing speech resources is important for progress in speech science and technology. Often, speech resources cannot be shared because of concerns over the privacy of the speakers, e.g., minors or people with medical conditions. Current technologies for pseudonymizing speech have only been tested on "standard" speech for which pseudonymization methods are evaluated on speaker identification risk, intelligibility, and naturalness. For many applications, the important characteristics are para-linguistic aspects of the speech, e.g., voice quality, emotion, or disease progression. Little information is available about the extent to which speaker pseudonymization methods preserve such paralinguistic information. The current study investigates how well voice quality parameters are preserved by an example speech pseudonymization application. Correlations prove to be high between original and pseudonymized recordings for seven acoustic parameters and a composite measure of dysphonia, the AVQI. Root mean square errors for these parameters were reasonably small. A linear mixed effect model shows a link between the difference between source and target speaker and the size of the absolute difference in the AVQI. It is argued that new measures of quality are needed for pseudonymized non-standard speech before wide-spread application of pseudonymized speech can be considered in research and clinical practise.

Audio-Visual Recognition of Emotional Engagement of People with Dementia

Lars Steinert, Felix Putze, Dennis Küster, Tanja Schultz; Universität Bremen, Germany Tue-E-O-3-2, Time: 19:20

Dementia places an immeasurable burden on affected individuals and caregivers. In addition to general cognitive decline, dementia has a negative impact on communication. Technical activation systems are thus in high demand, as cognitive activation may help to moderate the decline. However, effective activation requires sustained engagement — which, in turn, first needs to be reliably recognized. In this study, we examine emotional engagement recognition for People with Dementia (PwD) using non-intrusive biosignals resulting from speech communication and facial expressions. PwD suffering from mild to severe dementia used a tablet-based activation system over multiple sessions. We demonstrate that they retained their ability to verbally express emotional engagement even at severe stages of the disease. For recognition of emotional engagement, we propose an architecture of Bidirectional Long-Short-Term-Memory Networks that combines video information with up to three speech-based feature sets (eGeMAPS, ComParE'13, DeepSpectrum). Using data of 24 PwD, we show that adding speech improves recognition performance significantly compared to a video-only model. Interestingly, disease-progression did not appear to have a substantial impact on recognition performance in this sample. We further discuss the opportunities and challenges of detecting emotional engagement from speech in PwD.

Speaking Corona? Human and Machine Recognition of COVID-19 from Voice

Pascal Hecker¹, Florian B. Pokorny², Katrin D. Bartl-Pokorny², Uwe Reichel¹, Zhao Ren², Simone Hantke¹, Florian Eyben¹, Dagmar M. Schuller¹, Bert Arnrich³, Björn W. Schuller¹; ¹audEERING, Germany; ²Universität Augsburg, Germany; ³Universität Potsdam, Germany

Tue-E-O-3-3, Time: 19:40

With the COVID-19 pandemic, several research teams have reported successful advances in automated recognition of COVID-19 by voice. Resulting voice-based screening tools for COVID-19 could support large-scale testing efforts. While capabilities of machines on this task are progressing, we approach the so far unexplored aspect whether human raters can distinguish COVID-19 positive and negative tested speakers from voice samples, and compare their performance to a machine learning baseline. To account for the challenging symptom similarity between COVID-19 and other respiratory diseases, we use a carefully balanced dataset of voice samples, in which COVID-19 positive and negative tested speakers are matched by their symptoms alongside COVID-19 negative speakers without symptoms. Roth human raters and the machine struggle to reliably identify COVID-19 positive speakers in our dataset. These results indicate that particular attention should be paid to the distribution of symptoms across all speakers of a dataset when assessing the capabilities of existing systems. The identification of acoustic aspects of COVID-19-related symptom manifestations might be the key for a reliable voice-based COVID-19 detection in the future by both trained human raters and machine learning models.

Acoustic-Prosodic, Lexical and Demographic Cues to Persuasiveness in Competitive Debate Speeches

Huyen Nguyen¹, Ralph Vente², David Lupea³, Sarah Ita Levitan², Julia Hirschberg⁴; ¹Universität Hamburg, Germany; ²CUNY Hunter College, USA; ³NYU, USA; ⁴Columbia University, USA Tue-E-O-3-4, Time: 20:00

We analyze the acoustic-prosodic and lexical correlates of persuasiveness, taking into account speaker, judge and debate characteristics in a novel data set of 674 audio profiles, transcripts, evaluation scores and demographic data from professional debate tournament By conducting 10-fold cross validation experiments speeches. with linear, LASSO and random forest regression, we predict how different feature combinations contribute toward speech scores (i.e. persuasiveness) between men and women. Overall, lexical features, i.e. word complexity, nouns, fillers and hedges, are the most predictive features of speech evaluation scores; in addition to the gender composition of judge panels and opponents. In a combined lexical and demographic feature model, we achieve an \mathbb{R}^2 of 0.40. Different lexical features predict speech evaluation scores for male vs. female speakers, and further investigation is necessary to understand whether differential evaluation standards applied across genders. This work contributes a larger-scale debate data set in a democratically relevant, competitive format with high external relevance to persuasive speech education in other competitive settings.

Tue-E-V-1: Robust Speaker Recognition

19:00-21:00, Tuesday 31 August 2021 Chairs: Jason Pelecanos and Sandro Cumani

Unsupervised Bayesian Adaptation of PLDA for Speaker Verification

Bengt J. Borgström; MIT Lincoln Laboratory, USA Tue-E-V-1-1, Time: 19:00

This paper presents a Bayesian framework for unsupervised domain adaptation of Probabilistic Linear Discriminant Analysis (PLDA). By interpreting class labels as latent random variables, Variational Bayes (VB) is used to derive a maximum a posterior (MAP) solution of the adapted PLDA model when labels are missing, referred to as VB-MAP. The VB solution iteratively infers class labels and updates PLDA hyperparameters, offering a systematic framework for dealing with unlabeled data. While presented as a general solution, this paper includes experimental results for domain adaptation in speaker verification. VB-MAP estimation is applied to the 2016 and 2018 NIST Speaker Recognition Evaluations (SREs), both of which included small and unlabeled in-domain data sets, and is shown to provide performance improvements over a variety of state-of-the-art domain adaptation methods. Additionally, VB-MAP estimation is used to train a fully unsupervised PLDA model, suffering only minor performance degradation relative to conventional supervised training, offering promise for training PLDA models when no relevant labeled data exists.

The DKU-Duke-Lenovo System Description for the Fearless Steps Challenge Phase III

*Weiqing Wang*¹, Danwei Cai¹, Jin Wang², Qingjian Lin², Xuyang Wang², Mi Hong², Ming Li¹; ¹Duke Kunshan University, China; ²Lenovo, China Tue-E-V-1-2, Time: 19:00

This paper describes the systems developed by the DKU-Duke-Lenovo team for the Fearless Steps Challenge Phase III. For the speech activity detection (SAD) task, we employ the U-Net-based model which has not been used for SAD before, observing a DCF of 1.915% on the eval set. For the speaker identification (SID) task, we adopt the ResNet-SE and ECAPA-TDNN model, and we obtain a Top-5 accuracy of 86.21%. For the speaker diarization (SD) task, we employ several different clustering methods. Besides, domain adaptation, system fusion, and Target-Speaker Voice Activity Detection (TS-VAD) significantly improve the SD performance. We obtain a DER of 12.32% on track 2, and the major contribution is from our ResNet-based TS-VAD model. We finally achieve a first-place ranking for SD and SID and a second-place for SAD in the challenge.

Improved Meta-Learning Training for Speaker Verification

Yafeng Chen, Wu Guo, Bin Gu; USTC, China Tue-E-V-1-3, Time: 19:00

Meta-learning (ML) has recently become a research hotspot in speaker verification (SV). We introduce two methods to improve the meta-learning training for SV in this paper. For the first method, a backbone embedding network is first jointly trained with the conventional cross entropy loss and prototypical networks (PN) loss. Then, inspired by speaker adaptive training in speech recognition, additional transformation coefficients are trained with only the PN loss. The transformation coefficients are used to modify the original backbone embedding network in the x-vector extraction process. Furthermore, the random erasing (RE) data augmentation technique is applied to all support samples in each episode to construct positive pairs, and a contrastive loss between the augmented and the original support samples is added to the objective in model training. Experiments are carried out on the Speaker in the Wild (SITW) and VOiCES databases. Both of the methods can obtain consistent improvements over existing meta-learning training frameworks. By combining these two methods, we can observe further improvements on these two databases.

Variational Information Bottleneck Based Regularization for Speaker Recognition

Dan Wang, Yuanjie Dong, Yaxing Li, Yunfei Zi, Zhihui Zhang, Xiaoqi Li, Shengwu Xiong; WHUT, China Tue-E-V-1-4, Time: 19:00

Speaker recognition (SR) is inevitably affected by noise in real-life scenarios, resulting in decreased recognition accuracy. In this paper, we introduce a novel regularization method, variable information bottleneck (VIB), in speaker recognition to extract robust speaker embeddings. VIB prompts the neural network to ignore as much speaker-identity irrelevant information as possible. We also propose a more effective network, VovNet with an ultra-lightweight subspace attention module (ULSAM), as a feature extractor. ULSAM infers different attention maps for each feature map subspace, enabling efficient learning of cross-channel information along with multi-scale and multi-frequency feature representation. The experimental results demonstrate that our proposed framework outperforms the ResNet-based baseline by 11.4% in terms of equal error rate (EER). The VIB regularization method gives a further performance boost with an 18.9% EER decrease.

Out of a Hundred Trials, How Many Errors Does Your Speaker Verifier Make?

*Niko Brümmer*¹, *Luciana Ferrer*², *Albert Swart*¹; ¹*Phonexia, South Africa;* ²*UBA-CONICET ICC, Argentina* Tue-E-V-1-5, Time: 19:00

Out of a hundred trials, how many errors does your speaker verifier make? For the user this is an important, practical question, but researchers and vendors typically sidestep it and supply instead the

conditional error-rates that are given by the ROC/DET curve. We posit that the user's question is answered by the Bayes error-rate. We present a tutorial to show how to compute the error-rate that results when making Bayes decisions with calibrated likelihood ratios, supplied by the verifier, and an hypothesis prior, supplied by the user. For perfect calibration, the Bayes error-rate is upper bounded by min(EER,P,1-P), where EER is the equal-error-rate and P, 1-P are the prior probabilities of the competing hypotheses. The EER represents the accuracy of the verifier, while min(P,1-P) represents the hardness of the classification problem. We further show how the Bayes error-rate can be computed also for non-perfect calibration and how to generalize from error-rate to expected cost. We offer some criticism of decisions made by direct score thresholding. Finally, we demonstrate by analyzing error-rates of the recently published DCA-PLDA speaker verifier.

SpeakerStew: Scaling to Many Languages with a Triaged Multilingual Text-Dependent and Text-Independent Speaker Verification System

Roza Chojnacka, Jason Pelecanos, Quan Wang, Ignacio Lopez Moreno; Google, USA Tue-E-V-1-6, Time: 19:00

In this paper, we describe *SpeakerStew* — a hybrid system to perform speaker verification on 46 languages. Two core ideas were explored in this system: (1) Pooling training data of different languages together for multilingual generalization and reducing development cycles; (2) A novel triage mechanism between text-dependent and text-independent models to reduce runtime cost and expected latency. To the best of our knowledge, this is the first study of speaker verification systems at the scale of 46 languages. The problem is framed from the perspective of using a smart speaker device with interactions consisting of a wake-up keyword (text-dependent) followed by a speech query (text-independent). Experimental evidence suggests that training on multiple languages can generalize to unseen varieties while maintaining performance on seen varieties. We also found that it can reduce computational requirements for training models by an order of magnitude. Furthermore, during model inference on English data, we observe that leveraging a triage framework can reduce the number of calls to the more computationally expensive text-independent system by 73% (and reduce latency by 59%) while maintaining an EER no worse than the text-independent setup.

AntVoice Neural Speaker Embedding System for FFSVC 2020

Zhiming Wang, Furong Xu, Kaisheng Yao, Yuan Cheng, Tao Xiong, Huijia Zhu; Ant, China Tue-E-V-1-7, Time: 19:00

This paper presents a comprehensive description of the AntVoice system for the first two tracks of far-field speaker verification from single microphone array in FFSVC 2020 [1]. The system is based on neural speaker embeddings from deep neural network-based encoder networks. These encoder networks for acoustic modeling include 2D convolutional residual-like networks that are shown to be effective on the tasks. Specifically, we apply the Squeeze-and-Excitation residual network (SE-ResNet) [2] to model cross-channel inter-dependency information. On short utterances, we observe that SE-ResNet outperforms alternative methods in the text-dependent verification task. The system adopts a joint loss function that combines the additive cosine margin softmax loss [3] with the equidistant triplet-based loss[4]. This loss function results in performance gains with more discriminative speaker embeddings from enhanced intraclass similarity and increased inter-class variances. We also apply speech enhancement and data augmentation to improve data quality

and diversity. Even without using model ensembles, the proposed system significantly outperforms the baselines [1] in both tracks of the speaker verification challenge. With fusion of several encoder neural networks, this system is able to achieve further performance improvements consistently. In the end, the AntVoice system achieves the third place in the text-dependent verification task.

Gradient Regularization for Noise-Robust Speaker Verification

Jianchen Li, Jiqing Han, Hongwei Song; Harbin Institute of Technology, China Tue-E-V-1-8, Time: 19:00

Noise robustness is a challenge for speaker recognition systems. To solve this problem, one of the most common approaches is to joint-train a model by using both clean and noisy utterances. However, the gradients calculated on noisy utterances generally contain speaker-irrelevant noisy components, resulting in overfitting for the seen noisy data and poor generalization for the unseen noisy environments. To alleviate this problem, we propose the gradient regularization method to reduce the speaker-irrelevant noisy components by aligning the gradients among the noisy utterances and their clean counterparts. Specifically, the gradients on noisy utterances are forced to follow the directions of the gradients calculated on their clean counterparts, and the gradients across different types of noisy utterances are also aligned to point in similar directions. Since the noise-related components of the gradients can be reduced by the above alignment, the speaker model can be prevented from encoding irrelevant noisy information. To achieve the gradient regularization goals, a novel sequential inner training strategy is also proposed. Experiments on the VoxCeleb1 dataset indicate that our method achieves the best performance in seen and unseen noisy environments.

Deep Feature CycleGANs: Speaker Identity Preserving Non-Parallel Microphone-Telephone Domain Adaptation for Speaker Verification

Saurabh Kataria, Jesús Villalba, Piotr Żelasko, Laureano Moro-Velázquez, Najim Dehak; Johns Hopkins University, USA Tue-E-V-1-9, Time: 19:00

With the increase in the availability of speech from varied domains, it is imperative to use such out-of-domain data to improve existing speech systems. Domain adaptation is a prominent pre-processing approach for this. We investigate it to adapt microphone speech to the telephone domain. Specifically, we explore CycleGANbased unpaired translation of microphone data to improve the x-vector/speaker embedding network for Telephony Speaker Verification. We first demonstrate the efficacy of this on real challenging data and then, to improve further, we modify the CycleGAN formulation to make the adaptation task-specific. We modify CycleGAN's identity loss, cycle-consistency loss, and adversarial loss to operate in the *deep feature* space. *Deep features* of a signal are extracted from an auxiliary (speaker embedding) network and, hence, preserves speaker identity. Our 3D convolution-based Deep Feature Discriminators (DFD) show relative improvements of 5-10% in terms of equal error rate. To dive deeper, we study a challenging scenario of pooling (adapted) microphone and telephone data with data augmentations and telephone codecs. Finally, we highlight the sensitivity of CycleGAN hyper-parameters and introduce a parameter called probability of adaptation.

Scaling Effect of Self-Supervised Speech Models

*Jie Pu*¹, *Yuguang Yang*², *Ruirui Li*², *Oguz Elibol*², *Jasha Droppo*²; ¹*University of Cambridge, UK*; ²*Amazon, USA* Tue-E-V-1-10, Time: 19:00

The success of modern deep learning systems is built on two cornerstones, massive amount of annotated training data and advanced computational infrastructure to support large-scale computation. In recent years, the model size of state-of-the-art deep learning systems has rapidly increased and sometimes reached to billions of parameters. Herein we take a close look into this phenomenon and present an empirical study on the scaling effect of model size for self-supervised speech models. In particular, we investigate the quantitative relationship between the model size and the loss/accuracy performance on speech tasks. First, the power-law scaling property between the number of parameters and the L_1 self-supervised loss is verified for speech models. Then the advantage of large speech models in learning effective speech representations is demonstrated in two downstream tasks: i) speaker recognition and ii) phoneme classification. Moreover, it has been shown that the model size of self-supervised speech networks is able to compensate the lack of annotation when there is insufficient training data.

Joint Feature Enhancement and Speaker Recognition with Multi-Objective Task-Oriented Network

*Yibo Wu*¹, Longbiao Wang¹, Kong Aik Lee², Meng Liu¹, Jianwu Dang¹; ¹Tianjin University, China; ²A*STAR, Singapore

Tue-E-V-1-11, Time: 19:00

Recently, increasing attention has been paid to the joint training of upstream and downstream tasks, and to address the challenge of how to synchronize various loss functions in a multi-objective scenario. In this paper, to address the competing gradient directions between the speaker classification loss and the feature enhancement loss, we propose an asynchronous subregion optimization approach for the joint training of feature enhancement and speaker embedding neural networks. For the asynchronous subregion optimization, the squeeze and excitation (SE) method is introduced in the enhancement network to adaptively select important channels for speaker embedding. Furthermore, channel-wise feature concatenation is applied between the input feature and the enhanced feature to address the distortion of speaker information that is caused by enhancement loss. By using the proposed joint training network with asynchronous subregion optimization and channel-wise feature concatenation, we obtained relative gains of 11.95% and 6.43% in equal error rate on a noisy version of Voxceleb1 and VOiCES corpus, respectively.

Multi-Level Transfer Learning from Near-Field to Far-Field Speaker Verification

*Li Zhang*¹, *Qing Wang*¹, *Kong Aik Lee*², *Lei Xie*¹, *Haizhou Li*³; ¹*Northwestern Polytechnical University, China*; ²*A***STAR, Singapore*; ³*NUS, Singapore* Tue-E-V-1-12, Time: 19:00

In far-field speaker verification, the performance of speaker embeddings is susceptible to degradation when there is a mismatch between the conditions of enrollment and test speech. To solve this problem, we propose the feature-level and instance-level transfer learning in the teacher-student framework to learn a domain-invariant embedding space. For the feature-level knowledge transfer, we develop the contrastive loss to transfer knowledge from teacher model to student model, which not only decrease the intra-class distance, but also enlarge the inter-class distance. Moreover, we propose

NOTES

the instance-level pairwise distance transfer method to force the student model to preserve pairwise instances distance from the well optimized embedding space of the teacher model. On FFSVC 2020 evaluation set, our EER on Full-eval trials is relatively reduced by 13.9% compared with the fusion system result on Partial-eval trials of Task2. On Task1, compared with the winner's DenseNet result on Partial-eval trials, our minDCF on Full-eval trials is relatively reduced by 6.3%. On Task3, the EER and minDCF of our proposed method on Full-eval trials are very close to the result of the fusion system on Partial-eval trials. Our results also outperform other competitive domain adaptation methods.

Speaker Anonymisation Using the McAdams Coefficient

Jose Patino¹, Natalia Tomashenko², Massimiliano Todisco¹, Andreas Nautsch¹, Nicholas Evans¹; ¹EURECOM, France; ²LIA (EA 4128), France Tue-E-V-1-13, Time: 19:00

Anonymisation has the goal of manipulating speech signals in order to degrade the reliability of automatic approaches to speaker recognition, while preserving other aspects of speech, such as those relating to intelligibility and naturalness. This paper reports an approach to anonymisation that, unlike other current approaches, requires no training data, is based upon well-known signal processing techniques and is both efficient and effective. The proposed solution uses the McAdams coefficient to transform the spectral envelope of speech signals. Results derived using common VoicePrivacy 2020 databases and protocols show that random, optimised transformations can outperform competing solutions in terms of anonymisation while causing only modest, additional degradations to intelligibility, even in the case of a semi-informed privacy adversary.

Tue-E-V-2: Source Separation, Dereverberation and Echo Cancellation

19:00–21:00, Tuesday 31 August 2021 Chairs: Franz Pernkopf and Jonathan Le Roux

Multi-Stream Gated and Pyramidal Temporal Convolutional Neural Networks for Audio-Visual Speech Separation in Multi-Talker Environments

*Yiyu Luo*¹, *Jing Wang*¹, *Liang Xu*¹, *Lidong Yang*²; ¹*BIT*, *China*; ²*IMUST*, *China*

Tue-E-V-2-1, Time: 19:00

Speech separation is the task of extracting target speech from noisy mixture. In applications like video telephones or video conferencing, lip movements of the target speaker are accessible, which can be leveraged for speech separation. This paper proposes a time-domain audio-visual speech separation model under multi-talker environments. The model receives audio-visual inputs including noisy mixture and speaker lip embedding, and reconstructs clean speech waveform for the target speaker. Once trained, the model can be flexibly applied to unknown number of total speakers. This paper introduces and investigates the multi-stream gating mechanism and pyramidal convolution in temporal convolutional neural networks for audio-visual speech separation task. Speaker- and noise-independent multi-talker separation experiments are conducted on GRID benchmark dataset. The experimental results demonstrate the proposed method achieves 3.9 dB and 1.0 dB SI-SNRi improvement when compared with audio-only and audio-visual baselines respectively, showing effectiveness of the proposed method.

TeCANet: Temporal-Contextual Attention Network for Environment-Aware Speech Dereverberation

Helin Wang¹, Bo Wu², Lianwu Chen², Meng Yu³, Jianwei Yu², Yong Xu³, Shi-Xiong Zhang³, Chao Weng², Dan Su², Dong Yu³; ¹Peking University, China; ²Tencent, China; ³Tencent, USA Tue-E-V-2-2, Time: 19:00

In this paper, we exploit the effective way to leverage contextual information to improve the speech dereverberation performance in real-world reverberant environments. We propose a temporalcontextual attention approach on the deep neural network (DNN) for environment-aware speech dereverberation, which can adaptively attend to the contextual information. More specifically, a FullBand based Temporal Attention approach (FTA) is proposed, which models the correlations between the fullband information of the context frames. In addition, considering the difference between the attenuation of high frequency bands and low frequency bands (high frequency bands attenuate faster than low frequency bands) in the room impulse response (RIR), we also propose a SubBand based Temporal Attention approach (STA). In order to guide the network to be more aware of the reverberant environments, we jointly optimize the dereverberation network and the reverberation time (RT60) estimator in a multi-task manner. Our experimental results indicate that the proposed method outperforms our previously proposed reverberation-time-aware DNN and the learned attention weights are fully physical consistent. We also report a preliminary yet promising dereverberation and recognition experiment on real test data.

Residual Echo and Noise Cancellation with Feature Attention Module and Multi-Domain Loss Function

Jianjun Gu, Longbiao Cheng, Xingwei Sun, Junfeng Li, Yonghong Yan; CAS, China Tue-E-V-2-3, Time: 19:00

For real-time acoustic echo cancellation in noisy environments, the classical linear adaptive filters (LAFs) can only remove the linear components of acoustic echo. To further attenuate the non-linear echo components and background noise, this paper proposes a deep learning-based residual echo and noise cancellation (RENC) model, where multiple inputs are utilized and weighted by a feature attention module. More specifically, input features extracted from the far-end reference and the echo estimated by the LAF are scaled with time-frequency attention weights, depending on their correlation with the residual interference in LAF's output. Moreover, a scale-independent mean square error and perceptual loss function are further suggested for training the RENC model. Experimental results validate the efficacy of the proposed feature attention module and multi-domain loss function, which achieve an 8.4%, 14.9% and 29.5% improvement in perceptual evaluation of speech quality (PESQ), scale-invariant signal-to-distortion ratio (SI-SDR) and echo return loss enhancement (ERLE), respectively.

MIMO Self-Attentive RNN Beamformer for Multi-Speaker Speech Separation

*Xiyun Li*¹, *Yong Xu*², *Meng Yu*², *Shi-Xiong Zhang*², *Jiaming Xu*¹, *Bo Xu*¹, *Dong Yu*²; ¹*CAS, China*; ²*Tencent, USA*

Tue-E-V-2-4, Time: 19:00

Recently, our proposed recurrent neural network (RNN) based all deep learning minimum variance distortionless response (ADL-MVDR) beamformer method yielded superior performance over the conventional MVDR by replacing the matrix inversion and eigenvalue decomposition with two RNNs. In this work, we present a

NOTES

self-attentive RNN beamformer to further improve our previous RNN-based beamformer by leveraging on the powerful modeling capability of self-attention. Temporal-spatial self-attention module is proposed to better learn the beamforming weights from the speech and noise spatial covariance matrices. The temporal self-attention module could help RNN to learn global statistics of covariance matrices. The spatial self-attention module is designed to attend on the cross-channel correlation in the covariance matrices. Furthermore, a multi-speaker speech separation outputs (MIMO) model is developed to improve the inference efficiency. The evaluations demonstrate that our proposed MIMO self-attentive RNN beamformer improves both the automatic speech recognition (ASR) accuracy and the perceptual estimation of speech quality (PESQ) against prior arts.

Personalized PercepNet: Real-Time, Low-Complexity Target Voice Separation and Enhancement

Ritwik Giri¹, Shrikant Venkataramani¹, Jean-Marc Valin², Umut Isik¹, Arvindh Krishnaswamy¹; ¹Amazon, USA; ²Amazon, Canada Tue-E-V-2-5, Time: 19:00

The presence of multiple talkers in the surrounding environment poses a difficult challenge for real-time speech communication systems considering the constraints on network size and complexity. In this paper, we present Personalized PercepNet, a real-time speech enhancement model that separates a target speaker from a noisy multi-talker mixture without compromising on complexity of the recently proposed PercepNet. To enable speaker-dependent speech enhancement, we first show how we can train a perceptually motivated speaker embedder network to produce a representative embedding vector for the given speaker. Personalized PercepNet uses the target speaker embedding as additional information to pick out and enhance only the target speaker while suppressing all other competing sounds. Our experiments show that the proposed model significantly outperforms PercepNet and other baselines, both in terms of objective speech enhancement metrics and human opinion scores.

Scene-Agnostic Multi-Microphone Speech Dereverberation

*Yochai Yemini*¹, *Ethan Fetaya*¹, *Haggai Maron*², *Sharon Gannot*¹; ¹*Bar-Ilan University, Israel*; ²*NVIDIA, Israel* Tue-E-V-2-6, Time: 19:00

Neural networks (NNs) have been widely applied in speech processing tasks, and, in particular, those employing microphone arrays. Nevertheless, most existing NN architectures can only deal with fixed and position-specific microphone arrays. In this paper, we present an NN architecture that can cope with microphone arrays whose number and positions of the microphones are unknown, and demonstrate its applicability in the speech dereverberation task. To this end, our approach harnesses recent advances in deep learning on set-structured data to design an architecture that enhances the reverberant log-spectrum. We use noisy and noiseless versions of a simulated reverberant dataset to test the proposed architecture. Our experiments on the noisy data show that the proposed scene-agnostic setup outperforms a powerful scene-aware framework, sometimes even with fewer microphones. With the noiseless dataset we show that, in most cases, our method outperforms the position-aware network as well as the state-of-the-art weighted linear prediction error (WPE) algorithm.

Manifold-Aware Deep Clustering: Maximizing Angles Between Embedding Vectors Based on Regular Simplex

*Keitaro Tanaka*¹, *Ryosuke Sawata*², *Shusuke Takahashi*²; ¹*Waseda University, Japan*; ²*Sony, Japan* Tue-E-V-2-7, Time: 19:00

This paper presents a new deep clustering (DC) method called manifold-aware DC (M-DC) that can enhance hyperspace utilization more effectively than the original DC. The original DC has a limitation in that a pair of two speakers has to be embedded having an orthogonal relationship due to its use of the one-hot vector-based loss function, while our method derives a unique loss function aimed at maximizing the target angle in the hyperspace based on the nature of a regular simplex. Our proposed loss imposes a higher penalty than the original DC when the speaker is assigned incorrectly. The change from DC to M-DC can be easily achieved by rewriting just one term in the loss function of DC, without any other modifications to the network architecture or model parameters. As such, our method has high practicability because it does not affect the original inference part. The experimental results show that the proposed method improves the performances of the original DC and its expansion method.

A Deep Learning Approach to Multi-Channel and Multi-Microphone Acoustic Echo Cancellation

Hao Zhang, DeLiang Wang; Ohio State University, USA Tue-E-V-2-8, Time: 19:00

Building on deep learning based acoustic echo cancellation (AEC) in the single-loudspeaker (single-channel) and single-microphone setup, this paper investigates multi-channel (multi-loudspeaker) AEC (MCAEC) and multi-microphone AEC (MMAEC). A convolutional recurrent network (CRN) is trained to predict the near-end speech from microphone signals with far-end signals used as additional information. We find that the deep learning based MCAEC approach avoids the non-uniqueness problem in traditional MCAEC algorithms. For the AEC setup with multiple microphones, rather than employing AEC for each microphone, we propose to train a single network to achieve echo removal for all microphones. Combining deep learning based AEC with supervised beamforming further improves the system performance. Experimental results show the effectiveness of deep learning approach to MCAEC and MMAEC. Furthermore, deep learning based methods are capable of removing echo and noise simultaneously and work well in the presence of nonlinear distortions.

Joint Online Multichannel Acoustic Echo Cancellation, Speech Dereverberation and Source Separation

Yueyue Na, Ziteng Wang, Zhang Liu, Biao Tian, Qiang Fu; Alibaba, China Tue-E-V-2-9, Time: 19:00

This paper presents a joint source separation algorithm that simultaneously reduces acoustic echo, reverberation and interfering sources. Target speeches are separated from the mixture by maximizing independence with respect to the other sources. It is shown that the separation process can be decomposed into cascading subprocesses that separately relate to acoustic echo cancellation, speech dereverberation and source separation, all of which are solved using the auxiliary function based independent component/vector analysis techniques, and their solving orders are exchangeable. The cascaded solution not only leads to lower computational complexity but also better separation performance than the vanilla joint algorithm.

Should We Always Separate?: Switching Between Enhanced and Observed Signals for Overlapping Speech Recognition

Hiroshi Sato, Tsubasa Ochiai, Marc Delcroix, Keisuke Kinoshita, Takafumi Moriya, Naoyuki Kamo; NTT, Japan Tue-E-V-2-10, Time: 19:00

Although recent advances in deep learning technology improved automatic speech recognition (ASR), it remains difficult to recognize speech when it overlaps other people's voices. Speech separation or extraction is often used as a front-end to ASR to handle such overlapping speech. However, deep neural network-based speech enhancement can generate 'processing artifacts' as a side effect of the enhancement, which degrades ASR performance. For example, it is well known that single-channel noise reduction for non-speech noise (non-overlapping speech) often does not improve ASR. Likewise, the processing artifacts may also be detrimental to ASR in some conditions when processing overlapping speech with a separation/extraction method, although it is usually believed that separation/extraction improves ASR. In order to answer the question 'Do we always have to separate/extract speech from mixtures?', we analyze ASR performance on observed and enhanced speech at various noise and interference conditions, and show that speech enhancement degrades ASR under some conditions even for overlapping speech. Based on these findings, we propose a simple switching algorithm between observed and enhanced speech based on the estimated signal-to-interference ratio and signal-to-noise ratio. We demonstrated experimentally that such a simple switching mechanism can improve recognition performance when processing artifacts are detrimental to ASR.

Tue-E-V-3: Speech Signal Analysis and Representation I

19:00-21:00, Tuesday 31 August 2021 Chairs: Bhuvana Ramabhadran and Jinyu Li

Estimating Articulatory Movements in Speech Production with Transformer Networks

Sathvik Udupa¹, Anwesha Roy¹, Abhayjeet Singh¹, Aravind Illa², Prasanta Kumar Ghosh¹; ¹Indian Institute of Science, India; ²Amazon, India Tue-E-V-3-1, Time: 19:00

We estimate articulatory movements in speech production from different modalities - acoustics and phonemes. Acoustic-to-articulatory inversion (AAI) is a sequence-to-sequence task. On the other hand, phoneme to articulatory (PTA) motion estimation faces a key challenge in reliably aligning the text and the articulatory movements. To address this challenge, we explore the use of a transformer architecture - FastSpeech, with explicit duration modelling to learn hard alignments between the phonemes and articulatory movements. We also train a transformer model on AAI. We use correlation coefficient (CC) and root mean squared error (rMSE) to assess the estimation performance in comparison to existing methods on both tasks. We observe 154%, 11.8% & 4.8% relative improvement in CC with subject-dependent, pooled and fine-tuning strategies, respectively, for PTA estimation. Additionally, on the AAI task, we obtain 1.5%, 3% and 3.1% relative gain in CC on the same setups compared to the state-of-the-art baseline. We further present the computational benefits of having transformer architecture as representation blocks.

Unsupervised Multi-Target Domain Adaptation for Acoustic Scene Classification

Dongchao Yang, Helin Wang, Yuexian Zou; Peking University, China

Tue-E-V-3-2, Time: 19:00

It is well known that the mismatch between training (source) and test (target) data distribution will significantly decrease the performance of acoustic scene classification (ASC) systems. To address this issue, domain adaptation (DA) is one solution and many unsupervised DA methods have been proposed. These methods focus on a scenario of single source domain to single target domain. However, we will face such problem that test data comes from multiple target domains. This problem can be addressed by producing one model per target domain, but this solution is too costly. In this paper, we propose a novel unsupervised multi-target domain adaption (MTDA) method for ASC, which can adapt to multiple target domains simultaneously and make use of the underlying relation among multiple domains. Specifically, our approach combines traditional adversarial adaptation with two novel discriminator tasks that learns a common subspace shared by all domains. Furthermore, we propose to divide the target domain into the easy-to-adapt and hard-to-adapt domain, which enables the system to pay more attention to hard-to-adapt domain in training. The experimental results on the DCASE 2020 Task 1-A dataset and the DCASE 2019 Task 1-B dataset show that our proposed method significantly outperforms the previous unsupervised DA methods.

Speech Decomposition Based on a Hybrid Speech Model and Optimal Segmentation

Alfredo Esquivel Jaramillo, Jesper Kjær Nielsen, Mads Græsbøll Christensen; Aalborg University, Denmark

Tue-E-V-3-3, Time: 19:00

In a hybrid speech model, both voiced and unvoiced components can coexist in a segment. Often, the voiced speech is regarded as the deterministic component, and the unvoiced speech and additive noise are the stochastic components. Typically, the speech signal is considered stationary within fixed segments of 20-40 ms, but the degree of stationarity varies over time. For decomposing noisy speech into its voiced and unvoiced components, a fixed segmentation may be too crude, and we here propose to adapt the segment length according to the signal local characteristics. The segmentation relies on parameter estimates of a hybrid speech model and the maximum a posteriori (MAP) and log-likelihood criteria as rules for model selection among the possible segment lengths, for voiced and unvoiced speech, respectively. Given the optimal segmentation markers and the estimated statistics, both components are estimated using linear filtering. A codebook-based approach differentiates between unvoiced speech and noise. A better extraction of the components is possible by taking into account the adaptive segmentation, compared to a fixed one. Also, a lower distortion for voiced speech and higher segSNR for both components is possible, as compared to other decomposition methods.

Dropout Regularization for Self-Supervised Learning of Transformer Encoder Speech Representation

Jian Luo, Jianzong Wang, Ning Cheng, Jing Xiao; Ping An Technology, China

Tue-E-V-3-4, Time: 19:00

Predicting the altered acoustic frames is an effective way of selfsupervised learning for speech representation. However, it is challenging to prevent the pretrained model from overfitting. In this paper, we proposed to introduce two dropout regularization methods into the pretraining of transformer encoder: (1) attention dropout, (2) layer dropout. Both of the two dropout methods encourage the model to utilize global speech information, and avoid just copying local spectrum features when reconstructing the masked frames. We evaluated the proposed methods on phoneme classification and speaker recognition tasks. The experiments demonstrate that our dropout approaches achieve competitive results, and improve the performance of classification accuracy on downstream tasks.

Noise Robust Pitch Stylization Using Minimum Mean Absolute Error Criterion

*Chiranjeevi Yarra*¹, *Prasanta Kumar Ghosh*²; ¹*IIIT Hyderabad, India*; ²*Indian Institute of Science, India* Tue-E-V-3-5, Time: 19:00

We propose a pitch stylization technique in the presence of pitch halving and doubling errors. The technique uses an optimization criterion based on a minimum mean absolute error to make the stylization robust to such pitch estimation errors, particularly under noisy conditions. We obtain segments for the stylization automatically using dynamic programming. Experiments are performed at the frame level and the syllable level. At the frame level, the closeness of stylized pitch is analyzed with the ground truth pitch, which is obtained using a laryngograph signal, considering root mean square error (RMSE) measure. At the syllable level, the effectiveness of perceptual relevant embeddings in the stylized pitch is analyzed by estimating syllabic tones and comparing those with manual tone markings using the Levenshtein distance measure. The proposed approach performs better than a minimum mean squared error criterion based pitch stylization scheme at the frame level and a knowledge-based tone estimation scheme at the syllable level under clean and 20dB, 10dB and 0dB SNR conditions with five noises and four pitch estimation techniques. Among all the combinations of SNR, noise and pitch estimation techniques, the highest absolute RMSE and mean distance improvements are found to be 6.49Hz and 0.23. respectively.

An Attribute-Aligned Strategy for Learning Speech Representation

Yu-Lin Huang, Bo-Hao Su, Y.-W. Peter Hong, Chi-Chun Lee; National Tsing Hua University, Taiwan Tue-E-V-3-6, Time: 19:00

Advancement in speech technology has brought convenience to our life. However, the concern is on the rise as speech signal contains multiple personal attributes, which would lead to either sensitive information leakage or bias toward decision. In this work, we propose an attribute-aligned learning strategy to derive speech representation that can flexibly address these issues by attribute-selection mechanism. Specifically, we propose a layered-representation variational autoencoder (LR-VAE), which factorizes speech representation into attribute-sensitive nodes, to derive an identity-free representation for speech emotion recognition (SER), and an emotionless representation for speaker verification (SV). Our proposed method achieves competitive performances on identity-free SER and a better performance on emotionless SV, comparing to the current state-of-the-art method of using adversarial learning applied on a large emotion corpora, the MSP-Podcast. Also, our proposed learning strategy reduces the model and training process needed to achieve multiple privacy-preserving tasks.
Raw Speech-to-Articulatory Inversion by Temporal Filtering and Decimation

Abdolreza Sabzi Shahrebabaki, Sabato Marco Siniscalchi, Torbjørn Svendsen; NTNU, Norway Tue-E-V-3-7, Time: 19:00

We propose a novel sequence-to-sequence acoustic-to-articulatory inversion (AAI) neural architecture in the temporal waveform domain. In contrast to traditional AAI approaches that leverage hand-crafted short-time spectral features obtained from the windowed signal, such as LSFs, or MFCCs, our solution directly process the input speech signal in the time domain, avoiding any intermediate signal transformation, using a cascade of 1D convolutional filters in a deep model. The time-rate synchronization between raw speech signal and the articulatory signal is obtained through a decimation process that acts upon each convolution step. Decimation in time thus avoids degradation phenomena observed in the conventional AAI procedure, caused by the need of framing the speech signal to produce a feature sequence that perfectly matches the articulatory data rate. Experimental evidence on the "Haskins Production Rate Comparison" corpus demonstrates the effectiveness of the proposed solution, which outperforms a conventional state-of-the-art AAI system leveraging MFCCs with an 20% relative improvement in terms of Pearson correlation coefficient (PCC) in mismatched speaking rate conditions. Finally, the proposed approach attains the same accuracy as the conventional AAI solution in the typical matched speaking rate condition.

Unsupervised Training of a DNN-Based Formant Tracker

Jason Lilley, H. Timothy Bunnell; Nemours, USA Tue-E-V-3-8, Time: 19:00

Phonetic analysis often requires reliable estimation of formants, but estimates provided by popular programs can be unreliable. Recently, Dissen et al. [1] described DNN-based formant trackers that produced more accurate frequency estimates than several others, but require manually-corrected formant data for training. Here we describe a novel unsupervised training method for corpus-based DNN formant parameter estimation and tracking with accuracy similar to [1]. Frame-wise spectral envelopes serve as the input. The output is estimates of the frequencies and bandwidths plus amplitude adjustments for a prespecified number of poles and zeros, hereafter referred to as "formant parameters." A custom loss measure based on the difference between the input envelope and one generated from the estimated formant parameters is calculated and back-propagated through the network to establish the gradients with respect to the formant parameters. The approach is similar to that of autoencoders, in that the model is trained to reproduce its input in order to discover latent features, in this case, the formant parameters. Our results demonstrate that a reliable formant tracker can be constructed for a speech corpus without the need for handcorrected training data.

SUPERB: Speech Processing Universal PERformance Benchmark

Shu-wen Yang¹, Po-Han Chi¹, Yung-Sung Chuang¹, Cheng-I Jeff Lai², Kushal Lakhotia³, Yist Y. Lin¹, Andy T. Liu¹, Jiatong Shi⁴, Xuankai Chang⁵, Guan-Ting Lin¹, Tzu-Hsien Huang¹, Wei-Cheng Tseng¹, Ko-tik Lee¹, Da-Rong Liu¹, Zili Huang⁴, Shuyan Dong⁶, Shang-Wen Li⁶, Shinji Watanabe⁵, Abdelrahman Mohamed³, Hung-yi Lee¹; ¹National Taiwan University, Taiwan; ²MIT, USA; ³Facebook, USA; ⁴Johns Hopkins University, USA; ⁵Carnegie Mellon University, USA; ⁶Amazon, USA Tue-E-V-3-9, Time: 19:00

Self-supervised learning (SSL) has proven vital for advancing research in natural language processing (NLP) and computer vision (CV). The paradigm pretrains a shared model on large volumes of unlabeled data and achieves state-of-the-art (SOTA) for various tasks with minimal adaptation. However, the speech processing community lacks a similar setup to systematically explore the paradigm. To bridge this gap, we introduce Speech processing Universal PERformance Benchmark (SUPERB). SUPERB is a leaderboard to benchmark the performance of a shared model across a wide range of speech processing tasks with minimal architecture changes and labeled data. Among multiple usages of the shared model, we especially focus on extracting the representation learned from SSL for its preferable re-usability. We present a simple framework to solve SUPERB tasks by learning task-specialized lightweight prediction heads on top of the frozen shared model. Our results demonstrate that the framework is promising as SSL representations show competitive generalizability and accessibility across SUPERB tasks. We release SUPERB as a challenge with a leaderboard and a benchmark toolkit to fuel the research in representation learning and general speech processing.

Synchronising Speech Segments with Musical Beats in Mandarin and English Singing

*Cong Zhang*¹, *Jian Zhu*²; ¹*Radboud Universiteit, The Netherlands*; ²*University of Michigan, USA* Tue-E-V-3-10, Time: 19:00

Generating synthesised singing voice with models trained on speech data has many advantages due to the models' flexibility and controllability. However, since the information about the temporal relationship between segments and beats are lacking in speech training data, the synthesised singing may sound off-beat at times. Therefore, the availability of the information on the temporal relationship between speech segments and music beats is crucial. The current study investigated the segment-beat synchronisation in singing data, with hypotheses formed based on the linguistics theories of P-centre and sonority hierarchy. A Mandarin corpus and an English corpus of professional singing data were manually annotated and analysed. The results showed that the presence of musical beats was more dependent on segment duration than sonority. However, the sonority hierarchy and the P-centre theory were highly related to the location of beats. Mandarin and English demonstrated cross-linguistic variations despite exhibiting common patterns.

FRILL: A Non-Semantic Speech Embedding for Mobile Devices

Jacob Peplinski¹, Joel Shor², Sachin Joglekar³, Jake Garrison³, Shwetak Patel¹; ¹University of Washington, USA; ²Google, Japan; ³Google, USA Tue-E-V-3-11, Time: 19:00

Learned speech representations can drastically improve performance on tasks with limited labeled data. However, due to their size and complexity, learned representations have limited utility in mobile settings where run-time performance can be a significant bottleneck. In this work, we propose a class of lightweight non-semantic speech embedding models that run efficiently on mobile devices based on the recently proposed TRILL speech embedding. We combine novel architectural modifications with existing speed-up techniques to create embedding models that are fast enough to run in real-time on a mobile device and exhibit minimal performance degradation on a benchmark of non-semantic speech tasks. One such model (FRILL) is $32 \times$ faster on a Pixel 1 smartphone and 40% the size of TRILL, with an average decrease in accuracy of only 2%. To our knowledge, FRILL is the highest-quality non-semantic embedding designed for use on mobile devices. Furthermore, we demonstrate that these representations are useful for mobile health tasks such as non-speech human sounds detection and face-masked speech detection. Our models and code are publicly available.

Pitch Contour Separation from Overlapping Speech

Hiroki Mori; Utsunomiya University, Japan Tue-E-V-3-12, Time: 19:00

In everyday conversation, speakers' utterances often overlap. For conversation corpora that are recorded in diverse environments, results of pitch extraction in the overlapping parts may be incorrect. The goal of this study is to establish the technique of separating each speaker's pitch contour from an overlapping speech in conversation. The proposed method estimates statistically most plausible f_0 contour from the spectrogram of overlapping speech, along with the information of the speaker to extract. Visual inspection of the separation results showed that the proposed model was able to extract accurate f_0 contours from overlapping speeches of specified speakers. By applying this method, voicing decision errors and gross pitch errors were reduced by 63% compared to simple pitch extraction for overlapping speech.

Do Sound Event Representations Generalize to Other Audio Tasks? A Case Study in Audio Transfer Learning

Anurag Kumar, Yun Wang, Vamsi Krishna Ithapu, Christian Fuegen; Facebook, USA Tue-E-V-3-13, Time: 19:00

Transfer learning is critical for efficient information transfer across multiple related learning problems. A simple, yet effective transfer learning approach utilizes deep neural networks trained on a largescale task for feature extraction. Such representations are then used to learn related downstream tasks. In this paper, we investigate transfer learning capacity of audio representations obtained from neural networks trained on a large-scale sound event detection dataset. We build and evaluate these representations across a wide range of other audio tasks, via a simple linear classifier transfer mechanism. We show that such simple linear transfer is already powerful enough to achieve high performance on the downstream tasks. We also provide insights into the attributes of sound event representations that enable such efficient information transfer.

Tue-E-V-4: Spoken Language Understanding I

19:00-21:00, Tuesday 31 August 2021 Chairs: Gina-Anne Levow and Ian Lane

Data Augmentation for Spoken Language Understanding via Pretrained Language Models

Baolin Peng, Chenguang Zhu, Michael Zeng, Jianfeng Gao; Microsoft, USA Tue-E-V-4-1, Time: 19:00

The training of spoken language understanding (SLU) models often faces the problem of data scarcity. In this paper, we put forward a data augmentation method using pretrained language models to boost the variability and accuracy of generated utterances. Furthermore, we investigate and propose solutions to two previously overlooked semi-supervised learning scenarios of data scarcity in SLU: i) *Rich-in-Ontology*: ontology information with numerous valid dialogue acts is given; ii) *Rich-in-Utterance*: a large number of unlabelled utterances are available. Empirical results show that our method can produce synthetic training data that boosts the performance of language understanding models in various scenarios.

FANS: Fusing ASR and NLU for On-Device SLU

Martin Radfar, Athanasios Mouchtaris, Siegfried Kunzmann, Ariya Rastrow; Amazon, USA Tue-E-V-4-2, Time: 19:00

Spoken language understanding (SLU) systems translate voice input commands to semantics which are encoded as an intent and pairs of slot tags and values. Most current SLU systems deploy a cascade of two neural models where the first one maps the input audio to a transcript (ASR) and the second predicts the intent and slots from the transcript (NLU). In this paper, we introduce FANS, a new end-to-end SLU model that fuses an ASR audio encoder to a multi-task NLU decoder to infer the intent, slot tags, and slot values directly from a given input audio, obviating the need for transcription. FANS consists of a shared audio encoder and three decoders, two of which are seq-to-seq decoders that predict non null slot tags and slot values in parallel and in an auto-regressive manner. FANS neural encoder and decoders architectures are flexible which allows us to leverage different combinations of LSTM, self-attention, and attenders. Our experiments show compared to the state-of-the-art end-to-end SLU models, FANS reduces ICER and IRER errors relatively by 30% and 7%, respectively, when tested on an in-house SLU dataset and by 0.86% and 2% absolute when tested on a public SLU dataset.

Sequential End-to-End Intent and Slot Label Classification and Localization

*Yiran Cao*¹, *Nihal Potdar*¹, *Anderson R. Avila*²; ¹*University of Waterloo, Canada;* ²*Huawei Technologies, Canada*

<u>Tue-E</u>-V-4-3, Time: 19:00

Human-computer interaction (HCI) is significantly impacted by delayed responses from a spoken dialogue system. Hence, end-to-end (e2e) spoken language understanding (SLU) solutions have recently been proposed to decrease latency. Such approaches allow for the extraction of semantic information directly from the speech signal, thus bypassing the need for a transcript from an automatic speech recognition (ASR) system. In this paper, we propose a compact e2e SLU architecture for streaming scenarios, where chunks of the speech signal are processed continuously to predict intent and slot values. Our model is based on a 3D convolutional neural network (3D-CNN) and a unidirectional long short-term memory (LSTM).

We compare the performance of two alignment-free losses: the connectionist temporal classification (CTC) method and its adapted version, namely connectionist temporal localization (CTL). The latter performs not only the classification but also localization of sequential audio events. The proposed solution is evaluated on the Fluent Speech Command dataset and results show our model ability to process incoming speech signal, reaching accuracy as high as 98.97% for CTC and 98.78% for CTL on single-label classification, and as high as 95.69% for CTC and 95.28% for CTL on two-label prediction.

DEXTER: Deep Encoding of External Knowledge for Named Entity Recognition in Virtual Assistants

Deepak Muralidharan¹, Joel Ruben Antony Moniz¹, Weicheng Zhang¹, Stephen Pulman², Lin Li¹, Megan Barnes³, Jingjing Pan¹, Jason Williams¹, Alex Acero¹; ¹Apple, USA; ²Apple, UK; ³University of Washington, USA Tue-E-V-4-4, Time: 19:00

Named entity recognition (NER) is usually developed and tested on text from well-written sources. However, in intelligent voice assistants, where NER is an important component, input to NER may be noisy because of user or speech recognition error. In applications, entity labels may change frequently, and non-textual properties like topicality or popularity may be needed to choose among alternatives.

We describe a NER system intended to address these problems. We test and train this system on a proprietary user-derived dataset. We compare with a baseline text-only NER system; the baseline enhanced with external gazetteers; and the baseline enhanced with the search and indirect labelling techniques we describe below. The final configuration gives around 6% reduction in NER error rate. We also show that this technique improves related tasks, such as semantic parsing, with an improvement of up to 5% in error rate.

A Context-Aware Hierarchical BERT Fusion Network for Multi-Turn Dialog Act Detection

Ting-Wei Wu, Ruolin Su, Biing-Hwang Juang; Georgia Tech, USA

Tue-E-V-4-5, Time: 19:00

The success of interactive dialog systems is usually associated with the quality of the spoken language understanding (SLU) task, which mainly identifies the corresponding dialog acts and slot values in each turn. By treating utterances in isolation, most SLU systems often overlook the semantic context in which a dialog act is expected. The act dependency between turns is nontrivial and yet critical to the identification of the correct semantic representations. Previous works with limited context awareness have exposed the inadequacy of dealing with complexity in multiproned user intents, which are subject to spontaneous change during turn transitions. In this work, we propose to enhance SLU in multi-turn dialogs, employing a context-aware hierarchical BERT fusion Network (CaBERT-SLU) to not only discern context information within a dialog but also jointly identify multiple dialog acts and slots in each utterance. Experimental results show that our approach reaches new state-of-the-art (SOTA) performances in two complicated multi-turn dialogue datasets with considerable improvements compared with previous methods, which only consider single utterances for multiple intents and slot filling.

Pre-Training for Spoken Language Understanding with Joint Textual and Phonetic Representation Learning

Qian Chen, Wen Wang, Qinglin Zhang; Alibaba, China Tue-E-V-4-6, Time: 19:00

In the traditional cascading architecture for spoken language understanding (SLU), it has been observed that automatic speech recognition errors could be detrimental to the performance of natural language understanding. End-to-end (E2E) SLU models have been proposed to directly map speech input to desired semantic frame with a single model, hence mitigating ASR error propagation. Recently, pre-training technologies have been explored for these E2E models. In this paper, we propose a novel joint textual-phonetic pre-training approach for learning spoken language representations, aiming at exploring the full potentials of phonetic information to improve SLU robustness to ASR errors. We explore phoneme labels as high-level speech features, and design and compare pre-training tasks based on conditional masked language model objectives and inter-sentence relation objectives. We also investigate the efficacy of combining textual and phonetic information during fine-tuning. Experimental results on spoken language understanding benchmarks, Fluent Speech Commands and SNIPS, show that the proposed approach significantly outperforms strong baseline models and improves robustness of spoken language understanding to ASR errors.

Predicting Temporal Performance Drop of Deployed Production Spoken Language Understanding Models

Quynh Do, Judith Gaspers, Daniil Sorokin, Patrick Lehnen; Amazon, Germany Tue-E-V-4-7, Time: 19:00

In deployed real-world spoken language understanding (SLU) applications, data continuously flows into the system. This leads to distributional differences between training and application data that can deteriorate model performance. While regularly retraining the deployed model with new data helps mitigating this problem, it implies significant computational and human costs. In this paper, we develop a method, which can help guiding decisions on whether a model is safe to keep in production without notable performance loss or needs to be retrained. Towards this goal, we build a performance drop regression model for an SLU model that was trained offline to detect a potential model drift in the production phase. We present a wide range of experiments on multiple real-world datasets, indicating that our method is useful for guiding decisions in the SLU model development cycle and to reduce costs for model retraining.

Integrating Dialog History into End-to-End Spoken Language Understanding Systems

Jatin Ganhotra, Samuel Thomas, Hong-Kwang J. Kuo, Sachindra Joshi, George Saon, Zoltán Tüske, Brian Kingsbury; IBM, USA Tue-E-V-4-8, Time: 19:00

End-to-end spoken language understanding (SLU) systems that process human-human or human-computer interactions are often context independent and process each turn of a conversation independently. Spoken conversations on the other hand, are very much context dependent, and dialog history contains useful information that can improve the processing of each conversational turn. In this paper, we investigate the importance of dialog history and how it can be effectively integrated into end-to-end SLU systems. While processing a spoken utterance, our proposed RNN transducer (RNN-T) based SLU model has access to its dialog history in the form of decoded transcripts and SLU labels of previous turns. We encode the dialog history as BERT embeddings, and use them as an additional input to the SLU model along with the speech features for the current utterance. We evaluate our approach on a recently released spoken dialog data set, the HARPERVALLEYBANK corpus. We observe significant improvements: 8% for dialog action and 30% for caller intent recognition tasks, in comparison to a competitive context independent end-to-end baseline system.

Coreference Augmentation for Multi-Domain Task-Oriented Dialogue State Tracking

*Ting Han*¹, *Chongxuan Huang*², *Wei Peng*²; ¹*University of Illinois at Chicago, USA*; ²*Huawei Technologies, China* Tue-E-V-4-9, Time: 19:00

Dialogue State Tracking (DST), which is the process of inferring user goals by estimating belief states given the dialogue history, plays a critical role in task-oriented dialogue systems. A coreference phenomenon observed in multi-turn conversations is not addressed by existing DST models, leading to suboptimal performances. In this paper, we propose Coreference Dialogue State Tracker (CDST) that explicitly models the coreference feature. In particular, at each turn, the proposed model jointly predicts the coreferred domain-slot pair and extracts the coreference values from the dialogue context. Experimental results on MultiWOZ 2.1 dataset show that the proposed model achieves the state-of-the-art joint goal accuracy of 56.47%.

Rethinking End-to-End Evaluation of Decomposable Tasks: A Case Study on Spoken Language Understanding

Siddhant Arora, Alissa Ostapenko, Vijay Viswanathan, Siddharth Dalmia, Florian Metze, Shinji Watanabe, Alan W. Black; Carnegie Mellon University, USA Tue-E-V-4-10, Time: 19:00

Decomposable tasks are complex and comprise of a hierarchy of subtasks. Spoken intent prediction, for example, combines automatic speech recognition and natural language understanding. Existing benchmarks, however, typically hold out examples for only the surface-level sub-task. As a result, models with similar performance on these benchmarks may have unobserved performance differences on the other sub-tasks. To allow insightful comparisons between competitive end-to-end architectures, we propose a framework to construct robust test sets using coordinate ascent over sub-task specific utility functions. Given a dataset for a decomposable task, our method optimally creates a test set for each sub-task to individually assess sub-components of the end-to-end model. Using spoken language understanding as a case study, we generate new splits for the Fluent Speech Commands and Snips SmartLights datasets. Each split has two test sets: one with held-out utterances assessing natural language understanding abilities, and one with held-out speakers to test speech processing skills. Our splits identify performance gaps up to 10% between end-to-end systems that were within 1% of each other on the original test sets. These performance gaps allow more realistic and actionable comparisons between different architectures, driving future model development. We release our splits and tools for the community.

Tue-E-V-5: Topics in ASR: Adaptation, Transfer Learning, Children's Speech, and Low-Resource Settings

19:00–21:00, Tuesday 31 August 2021 Chairs: Abeer Alwan and Ewan Dunbar

Semantic Data Augmentation for End-to-End Mandarin Speech Recognition

Jianwei Sun, Zhiyuan Tang, Hengxin Yin, Wei Wang, Xi Zhao, Shuaijiang Zhao, Xiaoning Lei, Wei Zou, Xiangang Li; KE, China

Tue-E-V-5-1, Time: 19:00

End-to-end models have gradually become the preferred option for automatic speech recognition (ASR) applications. During the

NOTES

training of end-to-end ASR, data augmentation is a quite effective technique for regularizing the neural networks. This paper proposes a novel data augmentation technique based on semantic transposition of the transcriptions via syntax rules for end-to-end Mandarin ASR. Specifically, we first segment the transcriptions based on part-of-speech tags. Then transposition strategies, such as placing the object in front of the subject or swapping the subject and the object, are applied on the segmented sentences. Finally, the acoustic features corresponding to the transposed transcription are reassembled based on the audio-to-text forced-alignment produced by a pre-trained ASR system. The combination of original data and augmented one is used for training a new ASR system. The experiments are conducted on the Transformer[2] and Conformer[3] based ASR. The results show that the proposed method can give consistent performance gain to the system. Augmentation related issues, such as comparison of different strategies and ratios for data combination are also investigated.

Layer-Wise Fast Adaptation for End-to-End Multi-Accent Speech Recognition

Xun Gong, Yizhou Lu, Zhikai Zhou, Yanmin Qian; SJTU, China

Tue-E-V-5-2, Time: 19:00

Accent variability has posed a huge challenge to automatic speech recognition (ASR) modeling. Although one-hot accent vector based adaptation systems are commonly used, they require prior knowledge about the target accent and cannot handle unseen accents. Furthermore, simply concatenating accent embeddings does not make good use of accent knowledge, which has limited improvements. In this work, we aim to tackle these problems with a novel layer-wise adaptation structure injected into the E2E ASR model The adapter layer encodes an arbitrary accent in the encoder. accent space and assists the ASR model in recognizing accented speech. Given an utterance, the adaptation structure extracts the corresponding accent information and transforms the input acoustic feature into an accent-related feature through the linear combination of all accent bases. We further explore the injection position of the adaptation layer, the number of accent bases, and different types of accent bases to achieve better accent adaptation. Experimental results show that the proposed adaptation structure brings 12% and 10% relative word error rate (WER) reduction on the AESRC2020 accent dataset and the Librispeech dataset, respectively, compared to the baseline.

Low Resource German ASR with Untranscribed Data Spoken by Non-Native Children — INTERSPEECH 2021 Shared Task SPAPL System

Jinhan Wang¹, Yunzheng Zhu¹, Ruchao Fan¹, Wei Chu², Abeer Alwan¹; ¹University of California at Los Angeles, USA; ²PAII, USA Tue-E-V-5-3, Time: 19:00

This paper describes the SPAPL system for the INTERSPEECH 2021 Challenge: Shared Task on Automatic Speech Recognition for Non-Native Children's Speech in German. ~5 hours of transcribed data and ~60 hours of untranscribed data are provided to develop a German ASR system for children. For the training of the transcribed data, we propose a non-speech state discriminative loss (NSDL) to mitigate the influence of long-duration non-speech segments within speech utterances. In order to explore the use of the untranscribed data, various approaches are implemented and combined together to incrementally improve the system performance. First, bidirectional autoregressive predictive coding (Bi-APC) is used to learn initial parameters for acoustic modelling using the provided untranscribed data. Second, incremental semi-supervised learning is further used to iteratively generate pseudo-transcribed data. Third, different data augmentation schemes are used at different training stages to increase the variability and size of the training data. Finally, a recurrent neural network language model (RNNLM) is used for rescoring. Our system achieves a word error rate (WER) of 39.68% on the evaluation data, an approximately 12% relative improvement over the official baseline (45.21%).

Robust Continuous On-Device Personalization for Automatic Speech Recognition

Khe Chai Sim, Angad Chandorkar, Fan Gao, Mason Chua, Tsendsuren Munkhdalai, Françoise Beaufays; Google, USA_____

Tue-E-V-5-4, Time: 19:00

On-device personalization of an all-neural automatic speech recognition (ASR) model can be achieved efficiently by fine-tuning the last few layers of the model. This approach has been shown to be effective for adapting the model to recognize rare named entities using only a small amount of data. To reliably perform continuous on-device learning, it is important for the training process to be completely autonomous without manual intervention. Our simulation studies show that training over many rounds may eventually lead to a significant model drift if the personalized model is indiscriminately accepted at the end of each training round. It is important to have appropriate acceptance criteria in place to guard the model against drifting. Moreover, for storage efficiency, it is desirable to persist the model weights in quantized form. We found that quantizing and dequantizing the model weights in between training rounds can prevent the model from learning effectively. This issue can be circumvented by adding noise to the quantized weights at the start of each training round.

Speaker Normalization Using Joint Variational Autoencoder

Shashi Kumar¹, Shakti P. Rath², Abhishek Pandey¹; ¹Samsung, India; ²Reverie Language Technologies, India Tue-E-V-5-5, Time: 19:00

Speaker adaptation is known to provide significant improvement in speech recognition accuracy. However, in practical scenario, only a few seconds of audio is available due to which it may be infeasible to apply speaker adaptation methods such as i-vector and fMLLR robustly. Also, decoding with fMLLR transformation happens in two-passes which is impractical for real-time applications. In recent past, mapping speech features from speaker independent (SI) space to fMLLR normalized space using denoising autoencoder (DA) has been explored. To the best of our knowledge, such mapping generally does not yield consistent improvement. In this paper, we show that our proposed joint VAE based mapping achieves a large improvements over ASR models trained using filterbank SI features. We also show that joint VAE outperforms DA by a large margin. We observe a relative improvement of 17% in word error rate (WER) compared to ASR model trained using filterbank features with i-vectors and 23% without i-vectors.

The TAL System for the INTERSPEECH2021 Shared Task on Automatic Speech Recognition for Non-Native Childrens Speech

Gaopeng Xu, Song Yang, Lu Ma, Chengfei Li, Zhongqin Wu; TAL, China Tue-E-V-5-6, Time: 19:00

speech. In this work, we attempt to apply the self-supervised ap-

This paper describes TAL's system for the INTERSPEECH 2021 shared task on Automatic Speech Recognition (ASR) for non-native children's

NOTES

proach to non-native German children's ASR. First, we conduct some baseline experiments to indicate that self-supervised learning can capture more acoustic information on non-native children's speech. Then, we apply the 11-fold data augmentation and combine it with data clean-up to supplement to the limited training data. Moreover, an in-domain semi-supervised VAD model is utilized to segment untranscribed audio. These strategies can significantly improve the system performance. Furthermore, we use two types of language models to further improve performance, i.e., a 4-gram LM with CTC beam-search and a Transformer LM for 2-pass rescoring. Our ASR system reduces the Word Error Rate (WER) by about 48% relatively in comparison with the baseline, achieving 1st in the evaluation period with the WER of 23.5%.

On-the-Fly Aligned Data Augmentation for Sequence-to-Sequence ASR

Tsz Kin Lam, Mayumi Ohta, Shigehiko Schamoni, Stefan Riezler; Universität Heidelberg, Germany Tue-E-V-5-7, Time: 19:00

We propose an on-the-fly data augmentation method for automatic speech recognition (ASR) that uses alignment information to generate effective training samples. Our method, called Aligned Data Augmentation (ADA) for ASR, replaces transcribed tokens and the speech representations in an aligned manner to generate previously unseen training pairs. The speech representations are sampled from an audio dictionary that has been extracted from the training corpus and inject speaker variations into the training examples. The transcribed tokens are either predicted by a language model such that the augmented data pairs are semantically close to the original data, or randomly sampled. Both strategies result in training pairs that improve robustness in ASR training. Our experiments on a Seq-to-Seq architecture show that ADA can be applied on top of SpecAugment, and achieves about 9-23% and 4-15% relative improvements in WER over SpecAugment alone on LibriSpeech 100h and LibriSpeech 960h test datasets, respectively.

Zero-Shot Cross-Lingual Phonetic Recognition with External Language Embedding

Heting Gao¹, Junrui Ni¹, Yang Zhang², Kaizhi Qian², Shiyu Chang², Mark Hasegawa-Johnson¹; ¹University of Illinois at Urbana-Champaign, USA; ²MIT-IBM Watson AI Lab, USA

Tue-E-V-5-8, Time: 19:00

Many existing languages are too sparsely resourced for monolingual deep learning networks to achieve high accuracy. Multilingual phonetic recognition systems mitigate data sparsity issues by training models on data from multiple languages and learning a speech-tophone or speech-to-text model universal to all languages. However, despite their good performance on the seen training languages, multilingual systems have poor performance on unseen languages. This paper argues that in the real world, even an unseen language has metadata: linguists can tell us the language name, its language family and, usually, its phoneme inventory. Even with no transcribed speech, it is possible to train a language embedding using only data from language typologies (phylogenetic node and phoneme inventory) that reduces ASR error rates. Experiments on a 20-language corpus show that our methods achieve phonetic token error rate (PTER) reduction on all the unseen test languages. An ablation study shows that using the wrong language embedding usually harms PTER if the two languages are from different language families. However, even the wrong language embedding often improves PTER if the language embedding belongs to another member of the same language family.

Rapid Speaker Adaptation for Conformer Transducer: Attention and Bias Are All You Need

Yan Huang, Guoli Ye, Jinyu Li, Yifan Gong; Microsoft, USA

Tue-E-V-5-9, Time: 19:00

Conformer transducer achieves new state-of-the-art end-to-end (E2E) system performance and has become increasingly appealing for production. In this paper, we study how to effectively perform rapid speaker adaptation in a conformer transducer and how it compares with the RNN transducer. We hierarchically decompose the conformer transducer and compare adapting each component through fine-tuning. Among various interesting observations, there are three distinct findings: First, adapting the self-attention can achieve more than 80% gain of the full network adaptation. When the adaptation data is extremely scarce, attention is all you need to adapt. Second, within the self-attention, adapting the value projection outperforms adapting the key or the query projection. Lastly, bias adaptation, despite of its compact parameter space, is surprisingly effective. We conduct experiments on a state-of-the-art conformer transducer for an email dictation task. With 3 to 5 min source speech and 200 minute personalized TTS speech, the best performing encoder and joint network adaptation yields 38.37% and 19.90% relative word error rate (WER) reduction. Combining the attention and bias adaptation can achieve 90% of the gain with significantly smaller footprint. Further comparison with the RNN-T suggests the new state-of-the-art conformer transducer can benefit as much as if not more from personalization.

Best of Both Worlds: Robust Accented Speech Recognition with Adversarial Transfer Learning

Nilaksh Das¹, Sravan Bodapati², Monica Sunkara², Sundararajan Srinivasan², Duen Horng Chau¹; ¹Georgia Tech, USA; ²Amazon, USA Tue-E-V-5-10, Time: 19:00

Training deep neural networks for automatic speech recognition (ASR) requires large amounts of transcribed speech. This becomes a bottleneck for training robust models for accented speech which typically contains high variability in pronunciation and other semantics, since obtaining large amounts of annotated accented data is both tedious and costly. Often, we only have access to large amounts of unannotated speech from different accents. In this work, we leverage this unannotated data to provide semantic regularization to an ASR model that has been trained only on one accent, to improve its performance for multiple accents. We propose Accent Pre-Training (Acc-PT), a semi-supervised training strategy that combines transfer learning and adversarial training. Our approach improves the performance of a state-of-the-art ASR model by 33% on average over the baseline across multiple accents, training only on annotated samples from one standard accent, and as little as 105 minutes of unannotated speech from a target accent.

Extending Pronunciation Dictionary with Automatically Detected Word Mispronunciations to Improve PAII's System for Interspeech 2021 Non-Native Child English Close Track ASR Challenge

Wei Chu, Peng Chang, Jing Xiao; PAII, USA Tue-E-V-5-11, Time: 19:00

This paper proposed to automatically detect mispronounced words over the regions that have low Goodness-of-Pronunciation scores through a constrained phone decoder, then add these word mispronunciations into the orthodox lexicon without colliding with existing pronunciations, finally use the expanded lexicon for decoding nonnative speech. The constrained phone decoder is compiled by using

NOTES

a phone-level automatically generated one-edit-distance network to eliminate the need of extended recognition networks designed by phonologists. Results and analysis have shown that the pronunciation dictionary extension is effective in improving WER performance for non-native speech recognition. This paper also described the details of PAII's single-pass fusion-free hybrid system for this Interspeech 2021 non-native children English close track ASR challenge, especially showed the effective use of non-speech segments in the training set as noise sources to perform noise augmentation on the training data, and also conducted a comparison of acoustic models with different neural network architectures with analysis. Final WERs of 12.10%/28.25% are obtained compared to a well-optimized baseline with WERs of 13.37%/33.51% on development/evaluation set, respectively.

Tue-E-V-6: Voice Conversion and Adaptation I 19:00-21:00, Tuesday 31 August 2021

Chairs: Tomi Kinnunen and Yannis Stylianou

CVC: Contrastive Learning for Non-Parallel Voice Conversion

Tingle Li, Yichen Liu, Chenxu Hu, Hang Zhao; Tsinghua University, China

Tue-E-V-6-1, Time: 19:00

Cycle consistent generative adversarial network (CycleGAN) and variational autoencoder (VAE) based models have gained popularity in non-parallel voice conversion recently. However, they often suffer from difficult training process and unsatisfactory results. In this paper, we propose a contrastive learning-based adversarial approach for voice conversion, namely contrastive voice conversion (CVC). Compared to previous CycleGAN-based methods, CVC only requires an efficient one-way GAN training by taking the advantage of contrastive learning. When it comes to non-parallel one-to-one voice conversion, CVC is on par or better than CycleGAN and VAE while effectively reducing training time. CVC further demonstrates superior performance in many-to-one voice conversion, enabling the conversion from unseen speakers.

A Preliminary Study of a Two-Stage Paradigm for Preserving Speaker Identity in Dysarthric Voice Conversion

*Wen-Chin Huang*¹, *Kazuhiro Kobayashi*¹, *Yu-Huai Peng*², *Ching-Feng Liu*³, *Yu Tsao*², *Hsin-Min Wang*², *Tomoki Toda*¹; ¹*Nagoya University, Japan;* ²*Academia Sinica, Taiwan;* ³*Chi Mei Hospital, Taiwan Tue-E-V-6-2, Time:* 19:00

We propose a new paradigm for maintaining speaker identity in dysarthric voice conversion (DVC). The poor quality of dysarthric speech can be greatly improved by statistical VC, but as the normal speech utterances of a dysarthria patient are nearly impossible to collect, previous work failed to recover the individuality of the patient. In light of this, we suggest a novel, two-stage approach for DVC, which is highly flexible in that no normal speech of the patient is required. First, a powerful parallel sequence-to-sequence model converts the input dysarthric speech into a normal speech of a reference speaker as an intermediate product, and a nonparallel, frame-wise VC model realized with a variational autoencoder then converts the speaker identity of the reference speech back to that of the patient while assumed to be capable of preserving the enhanced quality. We investigate several design options. Experimental evaluation results demonstrate the potential of our approach to improving the quality of the dysarthric speech while maintaining the speaker identity.

One-Shot Voice Conversion with Speaker-Agnostic StarGAN

Sefik Emre Eskimez, Dimitrios Dimitriadis, Kenichi Kumatani, Robert Gmyr; Microsoft, USA Tue-E-V-6-3, Time: 19:00

In this work, we propose a variant of STARGAN for many-to-many voice conversion (VC) conditioned on the d-vectors for short-duration (2-15 seconds) speech. We make several modifications to the STAR-GAN training and employ new network architectures. We employ a transformer encoder in the discriminator network, and we apply the discriminator loss to the cycle consistency and identity samples in addition to the generated (fake) samples. Instead of classifying the samples as either real or fake, our discriminator tries to predict the categorical speaker class, where a fake class is added for the generated samples. Furthermore, we employ a reverse gradient layer after the generator's encoder and use an auxiliary classifier to remove the speaker's information from the encoded representation. We show that our method yields better results than the baseline method in objective and subjective evaluations in terms of voice conversion quality. Moreover, we provide an ablation study and show each component's influence on speaker similarity.

Fine-Tuning Pre-Trained Voice Conversion Model for Adding New Target Speakers with Limited Data

Takeshi Koshizuka, Hidefumi Ohmura, Kouichi Katsurada; Tokyo University of Science, Japan Tue-E-V-6-4, Time: 19:00

Voice conversion (VC) is a technique that converts speaker-dependent non-linguistic information into that of another speaker, while retaining the linguistic information of the input speech. A typical VC system comprises two modules: an encoder module that removes speaker individuality from the input speech and a decoder module that incorporates another speaker's individuality in synthesized speech. This paper proposes a training method for a vocoder-free any-to-many encoder-decoder VC model with limited data. Various pre-training techniques have been proposed to solve problems training to limited training data; some of these techniques employ the text-to-speech (TTS) task for pre-training. We pre-train the decoder module in the voice conversion task for growing our pretraining technique into continuously adding target speakers to the VC system. The experimental results show that good conversion performance can be achieved by conducting VC-based pre-training. We also confirmed that the rehearsal and pseudo-rehearsal methods can effectively fine-tune the model without degrading the conversion performance of the pre-trained target speakers.

VQMIVC: Vector Quantization and Mutual Information-Based Unsupervised Speech Representation Disentanglement for One-Shot Voice Conversion

Disong Wang¹, Liqun Deng², Yu Ting Yeung², Xiao Chen², Xunying Liu¹, Helen Meng¹; ¹CUHK, China; ²Huawei Technologies, China Tue-E-V-6-5, Time: 19:00

One-shot voice conversion (VC), which performs conversion across arbitrary speakers with only a single target-speaker utterance for reference, can be effectively achieved by speech representation disentanglement. Existing work generally ignores the correlation between different speech representations during training, which causes leakage of content information into the speaker representation and thus degrades VC performance. To alleviate this issue, we employ vector quantization (VQ) for content encoding and introduce mutual information (MI) as the correlation metric during training, to achieve proper disentanglement of content, speaker and pitch representations, by reducing their inter-dependencies in an unsupervised manner. Experimental results reflect the superiority of the proposed method in learning effective disentangled speech representations for retaining source linguistic content and intonation variations, while capturing target speaker characteristics. In doing so, the proposed approach achieves higher speech naturalness and speaker similarity than current state-of-the-art one-shot VC systems. Our code, pre-trained models and demo are publicly available.

StarGANv2-VC: A Diverse, Unsupervised, Non-Parallel Framework for Natural-Sounding Voice Conversion

Yinghao Aaron Li, Ali Zare, Nima Mesgarani; Columbia University, USA

Tue-E-V-6-6, Time: 19:00

We present an unsupervised non-parallel many-to-many voice conversion (VC) method using a generative adversarial network (GAN) called StarGAN v2. Using a combination of adversarial source classifier loss and perceptual loss, our model significantly outperforms previous VC models. Although our model is trained only with 20 English speakers, it generalizes to a variety of voice conversion tasks, such as any-to-many, cross-lingual, and singing conversion. Using a style encoder, our framework can also convert plain reading speech into stylistic speech, such as emotional and falsetto speech. Subjective and objective evaluation experiments on a non-parallel many-to-many voice conversion task revealed that our model produces natural sounding voices, close to the sound quality of state-of-the-art text-to-speech (TTS) based voice conversion methods without the need for text labels. Moreover, our model is completely convolutional and with a faster-than-real-time vocoder such as Parallel WaveGAN can perform real-time voice conversion.

Normalization Driven Zero-Shot Multi-Speaker Speech Synthesis

*Neeraj Kumar*¹, *Srishti Goel*¹, *Ankur Narang*¹, *Brejesh Lall*²; ¹*Hike, India*; ²*IIT Delhi, India* Tue-E-V-6-7, Time: 19:00

In this paper, we present a novel zero-shot multi-speaker speech synthesis approach (ZSM-SS) that leverages the normalization architecture and speaker encoder with non-autoregressive multi-head attention driven encoder-decoder architecture. Given an input text and a reference speech sample of an unseen person, ZSM-SS can generate speech in that person's style in a zero-shot manner. Additionally, we demonstrate how the affine parameters of normalization help in capturing the prosodic features such as energy and fundamental frequency in a disentangled fashion and can be used to generate morphed speech output. We demonstrate the efficacy of our proposed architecture on multi-speaker VCTK[1] and LibriTTS [2] datasets, using multiple quantitative metrics that measure generated speech distortion and MOS, along with speaker embedding analysis of the proposed speaker encoder model.

StarGAN-VC+ASR: StarGAN-Based Non-Parallel Voice Conversion Regularized by Automatic Speech Recognition

Shoki Sakamoto¹, Akira Taniguchi¹, Tadahiro Taniguchi¹, Hirokazu Kameoka²; ¹Ritsumeikan University, Japan; ²NTT, Japan Tue-E-V-6-8, Time: 19:00

Preserving the linguistic content of input speech is essential during voice conversion (VC). The star generative adversarial network-based

VC method (StarGAN-VC) is a recently developed method that allows non-parallel many-to-many VC. Although this method is powerful, it can fail to preserve the linguistic content of input speech when the number of available training samples is extremely small. To overcome this problem, we propose the use of automatic speech recognition to assist model training, to improve StarGAN-VC, especially in low-resource scenarios. Experimental results show that using our proposed method, StarGAN-VC can retain more linguistic information than vanilla StarGAN-VC.

Two-Pathway Style Embedding for Arbitrary Voice Conversion

Xuexin Xu¹, Liang Shi¹, Jinhui Chen², Xunquan Chen³, Jie Lian¹, Pingyuan Lin¹, Zhihong Zhang¹, Edwin R. Hancock⁴; ¹Xiamen University, China; ²Prefectural University of Hiroshima, Japan; ³Kobe University, Japan; ⁴University of York, UK Tue-E-V-6-9, Time: 19:00

Arbitrary voice conversion, also referred to as zero-shot voice conversion, has recently attracted increased attention in the literature. Although disentangling the linguistic and style representations

Although disentangling the linguistic and style representations for acoustic features is an effective way to achieve zero-shot voice conversion, the problem of how to convert to a natural speaker style is challenging because of the intrinsic variabilities of speech and the difficulties of completely decoupling them. For this reason, in this paper, we propose a Two-Pathway Style Embedding Voice Conversion framework (TPSE-VC) for realistic and natural speech conversion. The novel feature of this method is to simultaneously embed sentence-level and phoneme-level style information. A novel attention mechanism is proposed to implement the implicit alignment for timbre style and phoneme content, further embedding a phoneme-level style representation. In addition, we consider embedding the complete set of time steps of audio style into a fixed-length vector to obtain the sentence-level style representation. Moreover, TPSEVC does not require any pre-trained models, and is only trained with non-parallel speech data. Experimental results demonstrate that the proposed TPSE-VC outperforms the state-of-the-art results on zero-shot voice conversion.

Non-Parallel Any-to-Many Voice Conversion by Replacing Speaker Statistics

*Yufei Liu*¹, *Chengzhu Yu*¹, *Wang Shuai*¹, *Zhenchuan Yang*¹, *Yang Chao*¹, *Weibin Zhang*²; ¹*Tencent, China*; ²*SCUT, China*

Tue-E-V-6-10, Time: 19:00

This paper proposes a non-parallel any-to-many voice conversion (VC) approach with a novel statistics replacement layer. Non-parallel VC is usually achieved by firstly disentangling linguistic and speaker representations, and then concatenating the linguistic content with the learned target speaker's embedding at the conversion stage. While such a concatenation-based approach could introduce speaker-specific characteristics into the network, it is not very effective as it entirely relies on the network to learn to combine the linguistic content and the speaker characteristics. Inspired by X-vectors, where the statistics of hidden representation such as means and standard deviations are used for speaker differentiation, we propose a statistics replacement layer in VC systems to directly modify the hidden states to have the target speaker's statistics. The speaker-specific statistics of hidden states are learned for each target speaker during training and are used as guidance for the statistics replacement layer during inference. Moreover, to better concentrate the speaker information into the statistics of hidden representation, a multitask training with X-vector based speaker classification is

NOTES

also performed. Experimental results with Librispeech and VCTK datasets show that the proposed method can effectively improve the converted speech's naturalness and similarity.

Cross-Lingual Voice Conversion with a Cycle Consistency Loss on Linguistic Representation

*Yi Zhou*¹, *Xiaohai Tian*¹, *Zhizheng Wu*², *Haizhou Li*¹; ¹*NUS, Singapore;* ²*Facebook, USA* Tue-E-V-6-11, Time: 19:00

Cross-Lingual Voice Conversion (XVC) aims to modify a source speaker identity towards a target while preserving the source linguistic content. This paper introduces a cycle consistency loss on linguistic representation to ensure the speech content unchanged after conversion. The proposed XVC model consists of two loss functions during optimization: a spectral reconstruction loss and a linguistic cycle consistency loss. The cycle consistency loss seeks to maintain the source speech's linguistic content. Specifically, we utilize Phonetic PosteriorGram (PPG) to represent the linguistic content. XVC experiments were conducted between English and Mandarin. Both objective and subjective evaluations demonstrated that with the proposed cycle consistency loss, converted speech is more intelligible.

Improving Robustness of One-Shot Voice Conversion with Deep Discriminative Speaker Encoder

Hongqiang Du, Lei Xie; Northwestern Polytechnical University, China

Tue-E-V-6-12, Time: 19:00

One-shot voice conversion has received significant attention since only one utterance from source speaker and target speaker respectively is required. Moreover, source speaker and target speaker do not need to be seen during training. However, available one-shot voice conversion approaches are not stable for unseen speakers as the speaker embedding extracted from one utterance of an unseen speaker is not reliable. In this paper, we propose a deep discriminative speaker encoder to extract speaker embedding from one utterance more effectively. Specifically, the speaker encoder first integrates residual network and squeeze-and-excitation network to extract discriminative speaker information in frame level by modeling frame-wise and channel-wise interdependence in features. Then attention mechanism is introduced to further emphasize speaker related information via assigning different weights to frame level speaker information. Finally a statistic pooling layer is used to aggregate weighted frame level speaker information to form utterance level speaker embedding. The experimental results demonstrate that our proposed speaker encoder can improve the robustness of one-shot voice conversion for unseen speakers and outperforms baseline systems in terms of speech quality and speaker similarity.

Tue-E-SS-1: Voice Quality Characterization for Clinical Voice Assessment: Voice Production, Acoustics, and Auditory Perception

Room Lacina, 19:00–21:00, Tuesday 31 August 2021 Chairs: Philipp Aichinger and Carlo Drioli

Introduction

Time: 19:00

Optimizing an Automatic Creaky Voice Detection Method for Australian English Speaking Females

Hannah White, Joshua Penney, Andy Gibson, Anita Szakay, Felicity Cox; Macquarie University, Australia Tue-E-SS-1-1, Time: 19:05

Creaky voice is a nonmodal phonation type that has various linguistic and sociolinguistic functions. Manually annotating creaky voice for phonetic analysis is time-consuming and labor-intensive. In recent years, automatic tools for detecting creaky voice have been proposed, which present the possibility for easier, faster and more consistent creak identification. One of these proposed tools is a Creak Detector algorithm that uses an automatic neural network taking its input from several acoustic cues to identify creaky voice. Previous work has suggested that the creak probability threshold at which this tool determines an instance to be creaky may vary depending on the speaker population. The present study investigates the optimal creak detection threshold for female Australian English speakers.

Results show further support for the practice of first finding the optimal threshold when using the Creak Detection algorithm on new data sets. Additionally, results show that accuracy of creaky voice detection using the Creak Detection algorithm can be significantly improved by excluding non-sonorant data.

A Comparison of Acoustic Correlates of Voice Quality Across Different Recording Devices: A Cautionary Tale

Joshua Penney, Andy Gibson, Felicity Cox, Michael Proctor, Anita Szakay; Macquarie University, Australia Tue-E-SS-1-2, Time: 19:20

There has been a recent increase in speech research utilizing data recorded with participants' personal devices, particularly in light of the COVID-19 pandemic and restrictions on face-to-face interactions. This raises important questions about whether these recordings are comparable to those made in traditional lab-based settings. Some previous studies have compared the viability of recordings made with personal devices for the clinical evaluation of voice quality. However, these studies rely on simple statistical analyses and do not examine acoustic correlates of voice quality typically examined in the (socio-) phonetic literature (e.g. H1-H2). In this study, we compare recordings from a set of smartphones/laptops and a solid-state recorder to assess the reliability of a range of acoustic correlates of voice quality. The results show significant differences for many acoustic measures of voice quality across devices. Further exploratory analyses demonstrate that these differences are not simple offsets, but rather that their magnitude depends on the value of the measurement of interest. We therefore urge researchers to exercise caution when examining voice quality based on recordings made with participants' devices, particularly when interested in small effect sizes. We also call on the speech research community to investigate these issues more thoroughly.

Investigating Voice Function Characteristics of Greek Speakers with Hearing Loss Using Automatic Glottal Source Feature Extraction

Anna Sfakianaki, George P. Kafentzis; University of

Crete, Greece

Tue-E-SS-1-3, Time: 19:35

The current study investigates voice quality characteristics of Greek adults with normal hearing and hearing loss, automatically obtained from glottal inverse filtering analysis using the Aalto Aparat toolkit. Aalto Aparat has been employed in glottal flow analysis of disordered speech, but to the best of the authors' knowledge, not as yet in hearing impaired voice analysis and assessment. Five speakers, three women and two men, with normal hearing (NH) and five speakers with prelingual profound hearing impairment (HI), matched for age and sex, produced symmetrical /'pVpV/ disyllables, where V=/i, a, u/. A state-of-the-art method named quasi-closed phase analysis (QCP) is offered in Aparat and it is used to estimate the glottal source signal. Glottal source features were obtained using time- and frequency-domain parametrization methods and analysed statistically. The interpretation of the results attempts to shed light on potential differences between HI and NH phonation strategies, while advantages and limitations of inverse filtering methods in HI voice assessment are discussed.

Automated Detection of Voice Disorder in the Saarbrücken Voice Database: Effects of Pathology Subset and Audio Materials

Mark Huckvale, Catinca Buciuleac; University College London, UK

Tue-E-SS-1-4, Time: 19:50

The Saarbrücken Voice Database contains speech and simultaneous electroglottography recordings of 1002 speakers exhibiting a wide range of voice disorders, together with recordings of 851 controls. Previous studies have used this database to build systems for automated detection of voice disorders and for differential diagnosis. These studies have varied considerably in the subset of pathologies tested, the audio materials analyzed, the cross-validation method used and the performance metric reported. This variation has made it hard to determine the most promising approaches to the problem of detecting voice disorders. In this study we re-implement three recently published systems that have been trained to detect pathology using the SVD and compare their performance on the same pathologies with the same audio materials using a common cross-validation protocol and performance metric. We show that under this approach, there is much less difference in performance across systems than in their original publication. We also show that voice disorder detection on the basis of a short phrase gives similar performance to that based on a sequence of vowels of different pitch. Our evaluation protocol may be useful for future studies on voice disorder detection with the SVD.

Accelerometer-Based Measurements of Voice Quality in Children During Semi-Occluded Vocal Tract Exercise with a Narrow Straw in Air

Steven M. Lulich, Rita R. Patel; Indiana University, USA Tue-E-SS-1-5, Time: 20:05

Non-invasive measures of voice quality, such as H1-H2, rely on oral flow signals, inverse filtered speech signals, or corrections for the effects of formants. Voice quality measures play especially important roles in the assessment of voice disorders and the evaluation of treatment efficacy. One type of treatment that is increasingly common in voice therapy, as well as in voice training for singers and actors, is

semi-occluded vocal tract exercises (SOVTEs). The goal of SOVTEs is to change patterns of vocal fold vibration and thereby improve voice quality and vocal efficiency. Accelerometers applied to the skin of the neck have been used to investigate subglottal acoustics, to inverse-filter speech signals, and to obtain voice quality metrics. This paper explores the application of neck-skin accelerometers to measure voice quality without oral flow, inverse filtering, or formant correction. Accelerometer-based measures (uncorrected K1-K2 and corrected K1*-K2*, analogous to microphone-based H1-H2 and H1*-H2*) were obtained from typically developing children with healthy voice, before and during SOVTEs. Traditional microphone-based H1-H2 measures (corrected and uncorrected) were also obtained. Results showed that K1-K2 and K1*-K2* were not substantially affected by vocal tract acoustic changes in formant frequencies.

Articulatory Coordination for Speech Motor Tracking in Huntington Disease

Matthew Perez¹, Amrit Romana¹, Angela Roberts², Noelle Carlozzi¹, Jennifer Ann Miner¹, Praveen Dayalu¹, Emily Mower Provost¹; ¹University of Michigan, USA; ²Northwestern University, USA Tue-E-SS-1-6, Time: 20:20

Huntington Disease (HD) is a progressive disorder which often manifests in motor impairment. Motor severity (captured via motor score) is a key component in assessing overall HD severity. However, motor score evaluation involves in-clinic visits with a trained medical professional, which are expensive and not always accessible. Speech analysis provides an attractive avenue for tracking HD severity because speech is easy to collect remotely and provides insight into motor changes. HD speech is typically characterized as having irregular articulation. With this in mind, acoustic features that can capture vocal tract movement and articulatory coordination are particularly promising for characterizing motor symptom progression in HD. In this paper, we present an experiment that uses Vocal Tract Coordination (VTC) features extracted from read speech to estimate a motor score. When using an elastic-net regression model, we find that VTC features significantly outperform other acoustic features across varied-length audio segments, which highlights the effectiveness of these features for both short- and long-form reading tasks. Lastly, we analyze the F-value scores of VTC features to visualize which channels are most related to motor score. This work enables future research efforts to consider VTC features for acoustic analyses which target HD motor symptomatology tracking.

Modeling Dysphonia Severity as a Function of Roughness and Breathiness Ratings in the GRBAS Scale

Carlos A. Ferrer¹, Efren Aragón¹, María E. Hdez-Díaz², Marc S. de Bodt², Roman Cmejla³, Marina Englert⁴, Mara Behlau⁴, Elmar Nöth⁵; ¹Universidad Central de Las Villas, Cuba; ²Universiteit Antwerpen, Belgium; ³Czech Technical University in Prague, Czechia; ⁴Universidade Federal de São Paulo, Brazil; ⁵FAU Erlangen-Nürnberg, Germany Tue-E-SS-1-7, Time: 20:35

Dysphonia comprises many perceptually deviating aspects of voice, and its overall severity perception is made by the listener according to methods of aggregating the single dimensions which are personally conceived and not well studied. Roughness and breathiness are constituent dimensions in most devised rating scales in clinical use. In this paper, we evaluate several ways to model the mapping of the overall severity as a function of the particular ratings of roughness

NOTES

and breathiness. The models include the simple linear averaging as well as several non-linear variants suggested elsewhere, and some minor adjustments. The models are evaluated on four datasets from different countries, allowing a more global evaluation of how the mapping is conceived.

Results show the limitations of the most widely assumed linear approach, while also hinting at a need for a more uniform coverage of the sample space in voice pathology datasets. The models explored in this paper can be expanded to higher-dimensional scales.

Panel Discussion Time: 20:50

Wed-M-O-1: Miscellaneous Topics in ASR

Room A+B, 11:00–13:00, Wednesday 1 September 2021 Chairs: Martin Karafiat and Mathew Magimai-Doss

Golos: Russian Dataset for Speech Research

Nikolay Karpov, Alexander Denisenko, Fedor Minkin; SberBank, Russia

Wed-M-O-1-1, Time: 11:00

This paper introduces a novel Russian speech dataset called Golos, a large corpus suitable for speech research. The dataset mainly consists of recorded audio files manually annotated on the crowd-sourcing platform. The total duration of the audio is about 1240 hours. We have made the corpus freely available to download, along with the acoustic model with CTC loss prepared on this corpus. Additionally, transfer learning was applied to improve the performance of the acoustic model. In order to evaluate the quality of the dataset with the beam-search algorithm, we have built a 3-gram language model on the open Common Crawl dataset. The total word error rate (WER) metrics turned out to be about 3.3% and 11.5%.

Radically Old Way of Computing Spectra: Applications in End-to-End ASR

Samik Sadhu, Hynek Hermansky; Johns Hopkins University, USA

Wed-M-O-1-2, Time: 11:20

We propose a technique to compute spectrograms using Frequency Domain Linear Prediction (FDLP) that uses all-pole models to fit the squared Hilbert envelope of speech in different frequency sub-bands. The spectrogram of a complete speech utterance is computed by overlap-add of contiguous all-pole model responses. A long context window of 1.5 seconds allows us to capture the low frequency temporal modulations of speech in the spectrogram. For an end-to-end automatic speech recognition task, the FDLP spectrogram performs on par with the standard mel spectrogram features for clean read speech training and test data. For more realistic speech data with train-test domain mismatches or reverberations, FDLP spectrogram shows up to 25% and 22% relative WER improvements over mel spectrogram respectively.

Self-Supervised End-to-End ASR for Low Resource L2 Swedish

*Ragheb Al-Ghezi*¹, *Yaroslav Getman*¹, *Aku Rouhe*¹, *Raili Hildén*², *Mikko Kurimo*¹; ¹*Aalto University, Finland*; ²*University of Helsinki, Finland* Wed-M-O-1-3, Time: 11:40

Unlike traditional (hybrid) Automatic Speech Recognition (ASR), end-to-end ASR systems simplify the training procedure by directly mapping acoustic features to sequences of graphemes or characters, thereby eliminating the need for specialized acoustic, language, or pronunciation models. However, one drawback of end-to-end ASR systems is that they require more training data than conventional ASR systems to achieve similar word error rate (WER). This makes it difficult to develop ASR systems for tasks where transcribed target data is limited such as developing ASR for Second Language (L2) speakers of Swedish. Nonetheless, recent advancements in self-supervised acoustic learning, manifested in wav2vec models [1, 2, 3], leverage the available untranscribed speech data to provide compact acoustic representation that can achieve low WER when incorporated in end-to-end systems. To this end, we experiment with several monolingual and cross-lingual self-supervised acoustic models to develop end-to-end ASR system for L2 Swedish. Even

NOTES

though our test is very small, it indicates that these systems are competitive in performance with traditional ASR pipeline. Our best model seems to reduce the WER by 7% relative to our traditional ASR baseline trained on the same target data.

SPGISpeech: 5,000 Hours of Transcribed Financial Audio for Fully Formatted End-to-End Speech Recognition

Patrick K. O'Neill¹, Vitaly Lavrukhin², Somshubra Majumdar², Vahid Noroozi², Yuekai Zhang³, Oleksii Kuchaiev², Jagadeesh Balam², Yuliya Dovzhenko¹, Keenan Freyberg¹, Michael D. Shulman¹, Boris Ginsburg², Shinji Watanabe³, Georg Kucsko¹; ¹Kensho Technologies, USA; ²NVIDIA, USA; ³Johns Hopkins University, USA

Wed-M-O-1-4, Time: 12:00

In the English speech-to-text (STT) machine learning task, acoustic models are conventionally trained on uncased Latin characters, and any necessary orthography (such as capitalization, punctuation, and denormalization of non-standard words) is imputed by separate post-processing models. This adds complexity and limits performance, as many formatting tasks benefit from semantic information present in the acoustic signal but absent in transcription. Here we propose a new STT task: end-to-end neural transcription with fully formatted text for target labels. We present baseline Conformer-based models trained on a corpus of 5,000 hours of professionally transcribed earnings calls, achieving a CER of 1.7. As a contribution to the STT research community, we release the corpus free for non-commercial use.

LeBenchmark: A Reproducible Framework for Assessing Self-Supervised Representation Learning from Speech

Solène Evain¹, Ha Nguyen¹, Hang Le¹, Marcely Zanon Boito¹, Salima Mdhaffar², Sina Alisamir¹, Ziyi Tong¹, Natalia Tomashenko², Marco Dinarelli¹, Titouan Parcollet², Alexandre Allauzen³, Yannick Estève², Benjamin Lecouteux¹, François Portet¹, Solange Rossato¹, Fabien Ringeval¹, Didier Schwab¹, Laurent Besacier¹; ¹LIG (UMR 5217), France; ²LIA (EA 4128), France; ³LAMSADE (UMR 7243), France Wed-M-O-1-5, Time: 12:20

Self-Supervised Learning (SSL) using huge unlabeled data has been successfully explored for image and natural language processing. Recent works also investigated SSL from speech. They were notably successful to improve performance on downstream tasks such as automatic speech recognition (ASR). While these works suggest it is possible to reduce dependence on labeled data for building efficient speech systems, their evaluation was mostly made on ASR and using multiple and heterogeneous experimental settings (most of them for English). This questions the objective comparison of SSL approaches and the evaluation of their impact on building speech systems. In this paper, we propose LeBenchmark: a reproducible framework for assessing SSL from speech. It not only includes ASR (high and low resource) tasks but also spoken language understanding, speech translation and emotion recognition. We also focus on speech technologies in a language different than English: French. SSL. models of different sizes are trained from carefully sourced and documented datasets. Experiments show that SSL is beneficial for most but not all tasks which confirms the need for exhaustive and reliable benchmarks to evaluate its real impact. LeBenchmark is shared with the scientific community for reproducible research in SSL from speech.

Wed-M-O-2: Phonetics I

Room C, 11:00–13:00, Wednesday 1 September 2021 Chairs: Radek Skarnitzl and Barbara Schuppler

Prosodic Accommodation in Face-to-Face and Telephone Dialogues

Pavel Šturm, Radek Skarnitzl, Tomáš Nechanský; Charles University, Czechia Wed-M-O-2-1, Time: 11:00

The study of phonetic accommodation in various communicative situations is still relatively limited. This paper examines accommodation in spontaneous conversations of eight pairs of Czech young male speakers in two communicative conditions: unconstrained face-to-face conversation and goal-oriented interaction via mobile telephone. Articulation rate and measures of f0 level, range and variability were measured in 40 prosodic phrases per speaker in each condition. Analyses of LME models did not reveal a significant global effect of time throughout the interaction on the distance between speakers (convergence) in any of the examined parameters, or that of preceding phrase value on the subsequent turn-initial value (synchrony). However, more consistent patterns were observed when speaker pairs were examined separately, revealing substantial individual variation on the one hand and non-linear effects on the other. This shows that aggregate analyses can be misleading in the study of phonetic accommodation and that speakers dynamically employ different strategies throughout natural conversations.

Dialect Features in Heterogeneous and Homogeneous Gheg Speaking Communities

Josiane Riverin-Coutlée, Conceição Cunha, Enkeleida Kapia, Jonathan Harrington; LMU München, Germany Wed-M-O-2-2, Time: 11:20

This apparent and real time study analyses how dialect features in the speech of children and adults are differently affected depending on whether they live in homogeneous or heterogeneous speech communities. The general hypotheses are that speakers in such high contact settings as heterogeneous urban centers are more prone to innovation than speakers in homogeneous tightly-knit communities, and that children accelerate leveling, especially through schooling and socialization. This study is of Gheg Albanian, a dialect spoken in and around the capital Tirana. Two features were investigated: rounding of /a/ and vowel length contrasts. Two groups of adults and children were compared: one from Tirana and one from a nearby village. Additionally, the children were recorded twice over a period of 12 months and were compared longitudinally. The results showed that length contrasts were still present in both communities and age groups. Rounding of /a/ was lost in the city, but undergoing change in the village, with differences measured in apparent time, but also in child speech within the 12-month span. Our study further raises the issue of combining both apparent and real time data within the same design.

An Exploration of the Acoustic Space of Rhotics and Laterals in Ruruuli

Margaret Zellers¹, Alena Witzlack-Makarevich², Lilja Saeboe³, Saudah Namyalo⁴; ¹CAU, Germany; ²Hebrew University of Jerusalem, Israel; ³University of Oxford, UK; ⁴Makerere University, Uganda

Wed-M-O-2-3, Time: 11:40

Liquid consonants - rhotics and laterals - have been shown to demonstrate unique distributional patterns cross-linguistically. It is also claimed that rhotics are more difficult to distinguish from one another phonetically than laterals, and that rhotics are less flexible than laterals when it comes to participation in consonant clusters and coarticulatory patterns. We investigate the phonetic realization of the rhotic and lateral phonemes in a Bantu language, Ruruuli. The acoustic space used for rhotics and laterals in this language is extremely similar, although the density peaks in terms of formant values are different. Formant values as well as formant ratios can be reliably used to distinguish between rhotics and laterals. In common with many other languages, an asymmetry between laterals and rhotics is found in Ruruuli, with laterals being more positionally constrained than rhotics. The overlap in acoustic space between rhotics and laterals may cast doubt on the status or stability of the phonological contrast between rhotics and laterals in this language.

Domain-Initial Strengthening in Turkish: Acoustic Cues to Prosodic Hierarchy in Stop Consonants

Kubra Bodur, Sweeney Branje, Morgane Peirolo, Ingrid Tiscareno, James S. German; LPL (UMR 7309), France Wed-M-O-2-4, Time: 12:00

Studies have shown that cross-linguistically, consonants at the left edge of higher-level prosodic boundaries tend to be more forcefully articulated than those at lower-level boundaries, a phenomenon known as domain-initial strengthening. This study tests whether similar effects occur in Turkish, using the Autosegmental-Metrical model proposed by Ipek & Jun [1, 2] as the basis for assessing boundary strength. Productions of /t/ and /d/ were elicited in four domain-initial prosodic positions corresponding to progressively higher-level boundaries: syllable, word, intermediate phrase, and Intonational Phrase. A fifth position, nuclear word, was included in order to better situate it within the prosodic hierarchy. Acoustic correlates of articulatory strength were measured, including closure duration for /d/ and /t/, as well as voice onset time and burst energy for /t/. Our results show that closure duration increases cumulatively from syllable to intermediate phrase, while voice onset time and burst energy are not influenced by boundary strength. These findings provide corroborating evidence for Ipek & Jun's model, particularly for the distinction between word and intermediate phrase boundaries. Additionally, articulatory strength at the left edge of the nuclear word patterned closely with word-initial position, supporting the view that the nuclear word is not associated with a distinct phrasing domain.

Wed-M-O-3: Target Speaker Detection, Localization and Separation

Room D, 11:00–13:00, Wednesday 1 September 2021 Chairs: Reinhold Haeb-Umbach and Sebastian Braun

Auxiliary Loss Function for Target Speech Extraction and Recognition with Weak Supervision Based on Speaker Characteristics

Katerina Zmolikova¹, Marc Delcroix², Desh Raj³, Shinji Watanabe³, Jan Černocký¹; ¹Brno University of Technology, Czechia; ²NTT, Japan; ³Johns Hopkins University, USA Wed-M-O-3-1, Time: 11:00

Automatic speech recognition systems deteriorate in presence of overlapped speech. A popular approach to alleviate this is target speech extraction. The extraction system is usually trained with a loss function measuring the discrepancy between the estimated and the reference target speech. This often leads to distortions to the target signal which is detrimental to the recognition accuracy. Additionally, it is necessary to have the strong supervision provided by parallel data consisting of speech mixtures and single-speaker signals. We propose an auxiliary loss function for retraining the target speech extraction. It is composed of two parts: first, a speaker identity loss, forcing the estimated speech to have correct speaker characteristics, and second, a mixture consistency loss, making the extracted sources sum back to the original mixture. The only supervision required for the proposed loss is speaker characteristics obtained from several segments spoken by the target speaker. Such weak supervision makes the loss suitable for adapting the system directly on real recordings. We show that the proposed loss yields signals more suitable for speech recognition and further, we can gain additional improvements by adaptation to target data. Overall, we can reduce the word error rate on LibriCSS dataset from 27.4% to 24.0%

Universal Speaker Extraction in the Presence and Absence of Target Speakers for Speech of One and Two Talkers

Marvin Borsdorf¹, Chenglin Xu², Haizhou Li², Tanja Schultz¹; ¹Universität Bremen, Germany; ²NUS, Singapore

Wed-M-O-3-2, Time: 11:20

Speaker extraction has been studied mostly for the scenarios where a target speaker is present in a two or more talkers mixture. Such scenarios do not adequately reflect everyday conversations. For example, a target speaker can be the only active talker, be quiet for a while, or leave the conversation, that means the target speaker is absent from the mixture. Traditional speaker extraction models fail in these scenarios. We propose a novel speaker extraction approach to handle speech mixtures with one or two talkers in which the target speaker can either be present or absent. First, we formulate four speaker extraction conditions to cover the typical scenarios of everyday conversations with one and two talkers. Second, we introduce a joint training scheme with one unified loss function that works for all four conditions. We show that only a small amount of data is required to adapt the model to work well in the four conditions.

Using X-Vectors for Speech Activity Detection in Broadcast Streams

Lukas Mateju, Frantisek Kynych, Petr Cerva, Jindrich Zdansky, Jiri Malek; Technical University of Liberec, Czechia

Wed-M-O-3-3, Time: 11:40

A new approach to speech activity detection (SAD) is presented in this work. It allows us to reduce the complexity and computation demands, namely in services that process streaming speech, where a SAD module usually forms the first block of the data pipeline (e.g., in a platform for 24/7 broadcast transcription). Our approach utilizes x-vectors as input features so that, within the subsequent pipeline stages, these embedding instances can also directly be employed for speaker diarization and recognition. The x-vectors are extracted by feed-forward sequential memory network (FSMN), allowing for modeling long-time dependencies; they thus form an input into a computationally undemanding binary classifier, whose output is smoothed by a decoder. Evaluation is performed on the standardized QUT-NOISE-TIMIT dataset as well as on broadcast data with large portions of music and background noise. The former data allows for comparison with other existing approaches. The latter shows the performance in terms of word error rate (WER) and reduction in real-time factor (RTF) of the transcription process.

Time Delay Estimation for Speaker Localization Using CNN-Based Parametrized GCC-PHAT Features

Daniele Salvati, Carlo Drioli, Gian Luca Foresti; Università di Udine, Italy

Wed-M-O-3-4, Time: 12:00

We propose a time delay estimation (TDE) method for speaker localization based on parametrized generalized cross-correlation phase transform (PGCC-PHAT) functions and convolutional neural networks (CNNs). The PGCC-PHAT is used to build a feature matrix, which gives TDE information of two microphone signals with different normalization levels in the cross-correlation functions. The feature matrix is processed by a CNN, composed by several convolutional layers and fully connected layers and by a regression output for the directly estimation of the time difference of arrival (TDOA). Simulations in noisy and reverberant adverse conditions show that the proposed method improves the TDOA estimation performance if compared to the GCC-PHAT.

Real-Time Speaker Counting in a Cocktail Party Scenario Using Attention-Guided Convolutional Neural Network

Midia Yousefi, John H.L. Hansen; University of Texas at Dallas, USA

Wed-M-O-3-5, Time: 12:20

Most current speech technology systems are designed to operate well even in the presence of multiple active speakers. However, most solutions assume that the number of co-current speakers is known. Unfortunately, this information might not always be available in real-world applications. In this study, we propose a real-time, single-channel attention-guided Convolutional Neural Network (CNN) to estimate the number of active speakers in overlapping speech. The proposed system extracts higher-level information from the speech spectral content using a CNN model. Next, the attention mechanism summarizes the extracted information into a compact feature vector without losing critical information. Finally, the active speakers are classified using a fully connected network. Experiments on simulated overlapping speech using WSJ corpus show that the attention solution is shown to improve the performance by almost 3%

absolute over conventional temporal average pooling. The proposed Attention-guided CNN achieves 76.15% for both Weighted Accuracy and average Recall, and 75.80% Precision on speech segments as short as 20 frames (i.e., 200 ms). All the classification metrics exceed 92% for the attention-guided model in offline scenarios where the input signal is more than 100 frames long (i.e., 1s).

Wed-M-V-1: Language and Accent Recognition

11:00-13:00, Wednesday 1 September 2021 Chairs: Alberto Abad and Quan Wang

End-to-End Language Diarization for Bilingual Code-Switching Speech

Hexin Liu¹, Leibny Paola García Perera², Xinyi Zhang¹, Justin Dauwels³, Andy W.H. Khong¹, Sanjeev Khudanpur², Suzy J. Styles¹; ¹NTU, Singapore; ²Johns Hopkins University, USA; ³Technische Universiteit Delft, The Netherlands

Wed-M-V-1-1, Time: 11:00

We propose two end-to-end neural configurations for language diarization on bilingual code-switching speech. The first, a BLSTM-E2E architecture, includes a set of stacked bidirectional LSTMs to compute embeddings and incorporates the deep clustering loss to enforce grouping of languages belonging to the same class. The second, an XSA-E2E architecture, is based on an x-vector model followed by a self-attention encoder. The former encodes frame-level features into segment-level embeddings while the latter considers all those embeddings to generate a sequence of segment-level language labels. We evaluated the proposed methods on the dataset obtained from the shared task B in WSTCSMC 2020 and our handcrafted simulated data from the SEAME dataset. Experimental results show that our proposed XSA-E2E architecture achieved a relative improvement of 12.1% in equal error rate and a 7.4% relative improvement on accuracy compared with the baseline algorithm in the WSTCSMC 2020 dataset. Our proposed XSA-E2E architecture achieved an accuracy of 89.84% with a baseline of 85.60% on the simulated data derived from the SEAME dataset.

Modeling and Training Strategies for Language Recognition Systems

Raphaël Duroselle, Md. Sahidullah, Denis Jouvet, Irina Illina; Loria (UMR 7503), France Wed-M-V-1-2, Time: 11:00

Automatic speech recognition is complementary to language recognition. The language recognition systems exploit this complementarity by using frame-level bottleneck features extracted from neural networks trained with a phone recognition task. Recent methods apply frame-level bottleneck features extracted from an end-to-end sequence-to-sequence speech recognition model. In this work, we study an integrated approach of the training of the speech recognition feature extractor and language recognition modules. We show that for both classical phone recognition and end-to-end sequenceto-sequence features, sequential training of the two modules is not the optimal strategy. The feature extractor can be improved by supervision with the language identification loss, either in a fine-tuning step or in a multi-task training framework. Besides, we notice that end-to-end sequence-to-sequence bottleneck features are on par with classical phone recognition bottleneck features without requiring a forced alignment of the signal with target tokens. However, for sequence-to-sequence, the architecture of the model seems to play an important role; the Conformer architectures leads to much better results than the conventional stacked DNNs approach; and can even be trained directly with the LID module in an end-to-end approach.

A Weight Moving Average Based Alternate Decoupled Learning Algorithm for Long-Tailed Language Identification

Hui Wang¹, Lin Liu², Yan Song¹, Lei Fang², Ian McLoughlin³, Li-Rong Dai¹; ¹USTC, China; ²iFLYTEK, China; ³SIT, Singapore Wed-M-V-1-3, Time: 11:00

Language identification (LID) research has made tremendous progress in recent years, especially with the introduction of deep learning techniques. However, for real-world applications where the distribution of different language data is highly imbalanced, the performance of existing LID systems is still far from satisfactory. This raises the challenge of *long-tailed LID*. In this paper, we propose an effective weight moving average (WMA) based alternate decoupled learning algorithm, termed WADCL, for long-tailed LID. The system is divided into two components, a frontend feature extractor and a backend classifier. These are then alternately learned in an end-to-end manner using different sampling schemes to alleviate the distribution mismatch between training and test datasets. Furthermore, our WMA method aims to mitigate the side-effects of re-sampling schemes, by fusing the model parameters learned along the trajectory of stochastic gradient descent (SGD) optimization. To validate the effectiveness of the proposed WADCL algorithm, we evaluate and compare several systems over a language dataset constructed to match a long-tailed distribution based on real world application [1]. The experimental results from the long-tailed language dataset demonstrate that the proposed algorithm is able to achieve significant performance gains over existing state-of-the-art x-vector based LID methods.

Improving Accent Identification and Accented Speech Recognition Under a Framework of Self-Supervised Learning

Keqi Deng, Songjun Cao, Long Ma; Tencent, China Wed-M-V-1-4, Time: 11:00

Recently, self-supervised pre-training has gained success in automatic speech recognition (ASR). However, considering the difference between speech accents in real scenarios, how to identify accents and use accent features to improve ASR is still challenging. In this paper, we employ the self-supervised pre-training method for both accent identification and accented speech recognition tasks. For the former task, a standard deviation constraint loss (SDC-loss) based end-to-end (E2E) architecture is proposed to identify accents under the same language. As for accented speech recognition task, we design an accent-dependent ASR system, which can utilize additional accent input features. Furthermore, we propose a frame-level accent feature, which is extracted based on the proposed accent identification model and can be dynamically adjusted. We pre-train our models using 960 hours unlabeled LibriSpeech dataset and fine-tune them on AESRC2020 speech dataset. The experimental results show that our proposed accent-dependent ASR system is significantly ahead of the AESRC2020 baseline and achieves 6.5% relative word error rate (WER) reduction compared with our accent-independent ASR system.

Exploring wav2vec 2.0 on Speaker Verification and Language Identification

Zhiyun Fan, Meng Li, Shiyu Zhou, Bo Xu; CAS, China Wed-M-V-1-5, Time: 11:00

wav2vec 2.0 is a recently proposed self-supervised framework for speech representation learning. It follows a two-stage training

process of pre-training and fine-tuning, and performs well in speech recognition tasks especially ultra-low resource cases. In this work, we attempt to extend the self-supervised framework to speaker verification and language identification. First, we use some preliminary experiments to indicate that wav2vec 2.0 can capture the information about the speaker and language. Then we demonstrate the effectiveness of wav2vec 2.0 on the two tasks respectively. For speaker verification, we obtain a competitive result with the Equal Error Rate (EER) of 3.61% on the VoxCeleb1 dataset. For language identification, we obtain an EER of 12.02% on the 1 second condition and an EER of 3.47% on the full-length condition of the AP17-OLR dataset. Finally, we utilize one model to achieve the unified modeling by the multi-task learning for the two tasks.

Self-Supervised Phonotactic Representations for Language Identification

G. Ramesh, C. Shiva Kumar, K. Sri Rama Murty; IIT Hyderabad, India Wed-M-V-1-6, Time: 11:00

Phonotactic constraints characterize the sequence of permissible phoneme structures in a language and hence form an important cue for language identification (LID) task. As phonotactic constraints span across multiple phonemes, the short-term spectral analysis (20-30 ms) alone is not sufficient to capture them. The speech signal has to be analyzed over longer contexts (100s of milliseconds) in order to extract features representing the phonotactic constraints. The supervised senone classifiers, aimed at modeling triphone context, have been used for extracting language-specific features for the LID task. However, it is difficult to get large amounts of manually labeled data to train the supervised models. In this work, we explore a self-supervised approach to extract long-term contextual features for the LID task. We have used wav2vec architecture to extract contextualized representations from multiple frames of the speech signal. The contextualized representations extracted from the pretrained wav2vec model are used for the LID task. The performance of the proposed features is evaluated on a dataset containing 7 Indian languages. The proposed self-supervised embeddings achieved 23% absolute improvement over the acoustic features and 3% absolute improvement over their supervised counterparts.

E2E-Based Multi-Task Learning Approach to Joint Speech and Accent Recognition

Jicheng Zhang¹, Yizhou Peng¹, Van Tung Pham², Haihua Xu², Hao Huang¹, Eng Siong Chng²; ¹Xinjiang University, China; ²NTU, Singapore Wed-M-V-1-7, Time: 11:00

In this paper, we propose a single multi-task learning framework to perform End-to-End (E2E) speech recognition (ASR) and accent recognition (AR) simultaneously. The proposed framework is not only more compact but can also yield comparable or even better results than standalone systems. Specifically, we found that the overall performance is predominantly determined by the ASR task, and the E2E-based ASR pretraining is essential to achieve improved performance, particularly for the AR task. Additionally, we conduct several analyses of the proposed method. First, though the objective loss for the AR task is much smaller compared with its counterpart of ASR task, a smaller weighting factor with the AR task in the joint objective function is necessary to yield better results for each task. Second, we found that sharing only a few layers of the encoder yields better AR results than sharing the overall encoder. Experimentally, the proposed method produces WER results close to the best standalone E2E ASR ones, while it achieves 7.7% and 4.2% relative improvement over standalone and single-task-based joint recognition methods on test set for accent recognition respectively.

NOTES

Excitation Source Feature Based Dialect Identification in Ao — A Low Resource Language

Moakala Tzudir¹, Shikha Baghel¹, Priyankoo Sarmah¹, S.R. Mahadeva Prasanna²; ¹IIT Guwahati, India; ²IIT Dharwad, India

Wed-M-V-1-8, Time: 11:00

Ao is an under-resourced Tibeto-Burman tonal language spoken in Nagaland, India. There are three distinct dialects of the language, namely, Chungli, Mongsen and Changki. The objective of dialect identification is to identify one dialect from the other within the same language family. The goal of this study is to ascertain the potential of excitation source features for automatic dialect identification in Ao. In this direction, Integrated Linear Prediction Residual (ILPR), an approximate representation of source signal, is explored. The log Mel spectrogram of ILPR (S_{Ext}) signal is used to exploit the time-frequency characteristics of the excitation source. This work proposes attention based CNN-BiGRU architecture for automatic dialect identification tasks. Additionally, log Mel spectrogram (S_{VT}) , extracted from the pre-emphasized speech signal, is used as a baseline method. The (S_{VT}) contains the vocal-tract characteristics of the speech signal. A significant performance improvement of (nearly) 6% accuracy is observed when the excitation source feature (S_{Ext}) is combined with the vocal tract representation (S_{VT}) . To analyse the effect of segment duration, dialect identification performance is reported for three different durations, viz., 1 sec, 3 sec and 6 sec. The effect of gender in dialect identification task for Ao is also studied in this work.

Wed-M-V-2: Low-Resource Speech Recognition

11:00-13:00, Wednesday 1 September 2021 Chairs: Sabato Marco Siniscalchi and Xavier Anguera

Low Resource ASR: The Surprising Effectiveness of High Resource Transliteration

Shreya Khare¹, Ashish Mittal¹, Anuj Diwan², Sunita Sarawagi², Preethi Jyothi², Samarth Bharadwaj¹; ¹IBM, India; ²IIT Bombay, India Wed-M-V-2-1, Time: 11:00

Cross-lingual transfer of knowledge from high-resource languages to low-resource languages is an important research problem in automatic speech recognition (ASR). We propose a new strategy of transfer learning by pretraining using large amounts of speech in the high-resource language but with its text transliterated to the target low-resource language. This simple mapping of scripts explicitly encourages increased sharing between the output spaces of both languages and is surprisingly effective even when the high-resource and low-resource languages are from unrelated language families. The utility of our proposed technique is more evident in very lowresource scenarios, where better initializations are more beneficial. We evaluate our technique on a transformer ASR architecture and the state-of-the-art wav2vec2.0 ASR architecture, with English as the high-resource language and six languages as low-resource targets. With access to 1 hour of target speech, we obtain relative WER reductions of up to 8.2% compared to existing transfer-learning approaches.

Unsupervised Acoustic Unit Discovery by Leveraging a Language-Independent Subword Discriminative Feature Representation

Siyuan Feng¹, Piotr Żelasko², Laureano Moro-Velázquez², Odette Scharenborg¹; ¹Technische Universiteit Delft, The Netherlands; ²Johns Hopkins University, USA Wed-M-V-2-2, Time: 11:00

This paper tackles automatically discovering phone-like acoustic units (AUD) from unlabeled speech data. Past studies usually proposed single-step approaches. We propose a two-stage approach: the first stage learns a subword-discriminative feature representation, and the second stage applies clustering to the learned representation and obtains phone-like clusters as the discovered acoustic units. In the first stage, a recently proposed method in the task of unsupervised subword modeling is improved by replacing a monolingual out-of-domain (OOD) ASR system with a multilingual one to create a subword-discriminative representation that is more language-independent. In the second stage, segment-level k-means is adopted, and two methods to represent the variable-length speech segments as fixed-dimension feature vectors are compared. Experiments on a very low-resource Mboshi language corpus show that our approach outperforms state-of-the-art AUD in both normalized mutual information (NMI) and F-score. The multilingual ASR improved upon the monolingual ASR in providing OOD phone labels and in estimating the phone boundaries. A comparison of our systems with and without knowing the ground-truth phone boundaries showed a 16% NMI performance gap, suggesting that the current approach can significantly benefit from improved phone boundary estimation.

Towards Unsupervised Phone and Word Segmentation Using Self-Supervised Vector-Quantized Neural Networks

Herman Kamper, Benjamin van Niekerk; Stellenbosch University, South Africa Wed-M-V-2-3, Time: 11:00

We investigate segmenting and clustering speech into low-bitrate phone-like sequences without supervision. We specifically constrain pretrained self-supervised vector-quantized (VQ) neural networks so that blocks of contiguous feature vectors are assigned to the same code, thereby giving a variable-rate segmentation of the speech into discrete units. Two segmentation methods are considered. In the first, features are greedily merged until a prespecified number of segments are reached. The second uses dynamic programming to optimize a squared error with a penalty term to encourage fewer but longer segments. We show that these VQ segmentation methods can be used without alteration across a wide range of tasks: unsupervised phone segmentation, ABX phone discrimination, same-different word discrimination, and as inputs to a symbolic word segmentation algorithm. The penalized dynamic programming method generally performs best. While performance on individual tasks is only comparable to the state-of-the-art in some cases, in all tasks a reasonable competing approach is outperformed at a substantially lower bitrate.

Speech SimCLR: Combining Contrastive and Reconstruction Objective for Self-Supervised Speech Representation Learning

Dongwei Jiang¹, Wubo Li², Miao Cao², Wei Zou², Xiangang Li²; ¹YuanFuDao, China; ²DiDi Chuxing, China

Wed-M-V-2-4, Time: 11:00

Self-supervised visual pretraining has shown significant progress recently. Among those methods, SimCLR greatly advanced the state of the art in self-supervised and semi-supervised learning on ImageNet. The input feature representations for speech and visual tasks are both continuous, so it is natural to consider applying similar objective on speech representation learning. In this paper, we propose Speech SimCLR, a new self-supervised objective for speech representation learning. During training, Speech SimCLR applies augmentation on raw speech and its spectrogram. Its objective is the combination of contrastive loss that maximizes agreement between differently augmented samples in the latent space and reconstruction loss of input representation. The proposed method achieved competitive results on speech emotion recognition and speech recognition.

Multilingual Transfer of Acoustic Word Embeddings Improves When Training on Languages Related to the Target Zero-Resource Language

Christiaan Jacobs, Herman Kamper; Stellenbosch University, South Africa Wed-M-V-2-5, Time: 11:00

Acoustic word embedding models map variable duration speech segments to fixed dimensional vectors, enabling efficient speech search and discovery. Previous work explored how embeddings can be obtained in zero-resource settings where no labelled data is available in the target language. The current best approach uses transfer learning: a single supervised multilingual model is trained using labelled data from multiple well-resourced languages and then applied to a target zero-resource language (without fine-tuning). However, it is still unclear how the specific choice of training languages affect downstream performance. Concretely, here we ask whether it is beneficial to use training languages related to the target. Using data from eleven languages spoken in Southern Africa, we experiment with adding data from different language families while controlling for the amount of data per language. In word discrimination and query-by-example search evaluations, we show that training on languages from the same family gives large improvements. Through finer-grained analysis, we show that training on even just a single related language gives the largest gain. We also find that adding data from unrelated languages generally doesn't hurt performance.

Analyzing Speaker Information in Self-Supervised Models to Improve Zero-Resource Speech Processing

Benjamin van Niekerk, Leanne Nortje, Matthew Baas, Herman Kamper; Stellenbosch University, South Africa Wed-M-V-2-6, Time: 11:00

Contrastive predictive coding (CPC) aims to learn representations of speech by distinguishing future observations from a set of negative examples. Previous work has shown that linear classifiers trained on CPC features can accurately predict speaker and phone labels. However, it is unclear how the features actually capture speaker and phonetic information, and whether it is possible to normalize out the irrelevant details (depending on the downstream task). In this paper, we first show that the per-utterance mean of CPC features captures speaker information to a large extent. Concretely, we find

that comparing means performs well on a speaker verification task. Next, probing experiments show that standardizing the features effectively removes speaker information. Based on this observation, we propose a speaker normalization step to improve acoustic unit discovery using *K*-means clustering of CPC features. Finally, we show that a language model trained on the resulting units achieves some of the best results in the ZeroSpeech2021 Challenge.

Unsupervised Neural-Based Graph Clustering for Variable-Length Speech Representation Discovery of Zero-Resource Languages

Shun Takahashi, Sakriani Sakti, Satoshi Nakamura; NAIST, Japan

Wed-M-V-2-7, Time: 11:00

Discovering symbolic units from unannotated speech data is fundamental in zero resource speech technology. Previous studies focused on learning fixed-length frame units based on acoustic features. Although they achieve high quality, they also suffer from a high bit-rate due to time-frame encoding. In this work, to discover variable-length, low bit-rate speech representation from a limited amount of unannotated speech data, we propose an approach based on graph neural networks (GNNs), and we study the temporal closeness of salient speech features. Our approach is built upon vector-quantized neural networks (VQNNs), which learn discrete encoding by contrastive predictive coding (CPC). We exploit the predetermined finite set of embeddings (a codebook) used by VQNNs to encode input data. We consider a codebook a set of nodes in a directed graph, where each arc represents the transition from one feature to another. Subsequently, we extract and encode the topological features of nodes in the graph to cluster them using graph convolution. By this process, we can obtain coarsened speech representation. We evaluated our model on the English data set of the ZeroSpeech 2020 challenge on Track 2019. Our model successfully drops the bit rate while achieving high unit quality.

Speech Representation Learning Combining Conformer CPC with Deep Cluster for the ZeroSpeech Challenge 2021

Takashi Maekaku¹, Xuankai Chang², Yuya Fujita¹, Li-Wei Chen², Shinji Watanabe², Alexander Rudnicky²; ¹Yahoo, Japan; ²Carnegie Mellon University, USA Wed-M-V-2-8, Time: 11:00

We present a system for the Zero Resource Speech Challenge 2021, which combines a Contrastive Predictive Coding (CPC) with deep cluster. In deep cluster, we first prepare pseudo-labels obtained by clustering the outputs of a CPC network with k-means. Then, we train an additional autoregressive model to classify the previously obtained pseudo-labels in a supervised manner. Phoneme discriminative representation is achieved by executing the second-round clustering with the outputs of the final layer of the autoregressive model. We show that replacing a Transformer layer with a Conformer layer leads to a further gain in a lexical metric. Experimental results show that a relative improvement of 35% in a phonetic metric, 1.5% in the lexical metric, and 2.3% in a syntactic metric are achieved on LibriSpeech 460h data. We achieve top results in this challenge with the syntactic metric.

Identifying Indicators of Vulnerability from Short Speech Segments Using Acoustic and Textual Features

*Xia Cui*¹, *Amila Gamage*², *Terry Hanley*¹, *Tingting Mu*¹; ¹*University of Manchester, UK*; ²*VoiceIQ, UK* Wed-M-V-2-9, Time: 11:00

In order to protect vulnerable people in telemarketing, organisations have to investigate the speech recordings to identify them first. Typically, the investigation is manually conducted. As such, the procedure is costly and time-consuming. With an automatic vulnerability detection system, more vulnerable people can be identified and protected. A standard telephone conversation lasts around 5 minutes, the detection system is expected to be able to identify such a potential vulnerable speaker from speech segments. Due to the complexity of the vulnerability definition and the unavailable annotated vulnerability examples, this paper attempts to address the detection problem as three classification tasks: age classification, accent classification and patient/non-patient classification utilising publicly available datasets. In the proposed system, we trained three sub models using acoustic and textual features for each sub task. Each trained model was evaluated on multiple datasets and achieved competitive results compared to a strong baseline (i.e. in-dataset accuracy).

The Zero Resource Speech Challenge 2021: Spoken Language Modelling

Ewan Dunbar¹, Mathieu Bernard², Nicolas Hamilakis², Tu Anh Nguyen², Maureen de Seyssel², Patricia Rozé², Morgane Rivière³, Eugene Kharitonov³, Emmanuel Dupoux²; ¹University of Toronto, Canada; ²LSCP (UMR 8554), France; ³Facebook, France Wed-M-V-2-10, Time: 11:00

We present the Zero Resource Speech Challenge 2021, which asks participants to learn a language model directly from audio, without any text or labels. The challenge is based on the Libri-light dataset, which provides up to 60k hours of audio from English audio books without any associated text. We provide a pipeline baseline system consisting on an encoder based on contrastive predictive coding (CPC), a quantizer (*k*-means) and a standard language model (BERT or LSTM). The metrics evaluate the learned representations at the acoustic (ABX discrimination), lexical (spot-the-word), syntactic (acceptability judgment) and semantic levels (similarity judgment). We present an overview of the eight submitted systems from four groups and discuss the main results.

Zero-Shot Federated Learning with New Classes for Audio Classification

Gautham Krishna Gudur, Satheesh Kumar Perepu; Ericsson, India

Wed-M-V-2-11, Time: 11:00

Federated learning is an effective way of extracting insights from different user devices while preserving the privacy of users. However, new classes with completely unseen data distributions can stream across any device in a federated learning setting, whose data cannot be accessed by the global server or other users. To this end, we propose a unified zero-shot framework to handle these aforementioned challenges during federated learning. We simulate two scenarios here -1) when the new class labels are not reported by the user, the traditional FL setting is used; 2) when new class labels are reported by the user, we synthesize *Anonymized Data Impressions* by calculating class similarity matrices corresponding to each device's new classes followed by unsupervised clustering to distinguish between new classes across different users. Moreover, our proposed

framework can also handle statistical heterogeneities in both labels and models across the participating users. We empirically evaluate our framework on-device across different communication rounds (FL iterations) with new classes in both local and global updates, along with heterogeneous labels and models, on two widely used audio classification applications — keyword spotting and urban sound classification, and observe an average deterministic accuracy increase of ~4.041% and ~4.258% respectively.

AVLnet: Learning Audio-Visual Language Representations from Instructional Videos

Andrew Rouditchenko¹, Angie Boggust¹, David Harwath², Brian Chen³, Dhiraj Joshi⁴, Samuel Thomas⁴, Kartik Audhkhasi⁵, Hilde Kuehne⁴, Rameswar Panda⁴, Rogerio Feris⁴, Brian Kingsbury⁴, Michael Picheny⁶, Antonio Torralba¹, James Glass¹; ¹MIT, USA; ²University of Texas at Austin, USA; ³Columbia University, USA; ⁴IBM, USA; ⁵Google, USA; ⁶NYU, USA

Wed-M-V-2-12, Time: 11:00

Current methods for learning visually grounded language from videos often rely on text annotation, such as human generated captions or machine generated automatic speech recognition (ASR) transcripts. In this work, we introduce the Audio-Video Language Network (AVLnet), a self-supervised network that learns a shared audio-visual embedding space directly from raw video inputs. To circumvent the need for text annotation, we learn audio-visual representations from randomly segmented video clips and their raw audio waveforms. We train AVLnet on HowTo100M, a large corpus of publicly available instructional videos, and evaluate on image retrieval and video retrieval tasks, achieving state-of-the-art performance. Finally, we perform analysis of AVLnet's learned representations, showing our model utilizes speech and natural sounds to learn audio-visual concepts.

Wed-M-V-3: Speech Synthesis: Singing, Multimodal, Crosslingual Synthesis

11:00-13:00, Wednesday 1 September 2021 Chairs: Christophe d'Alessandro and Thomas Hueber

N-Singer: A Non-Autoregressive Korean Singing Voice Synthesis System for Pronunciation Enhancement

Gyeong-Hoon Lee, Tae-Woo Kim, Hanbin Bae, Min-Ji Lee, Young-Ik Kim, Hoon-Young Cho; NCSOFT, Korea Wed-M-V-3-1, Time: 11:00

Recently, end-to-end Korean singing voice systems have been designed to generate realistic singing voices. However, these systems still suffer from a lack of robustness in terms of pronunciation accuracy. In this paper, we propose N-Singer, a non-autoregressive Korean singing voice system, to synthesize accurate and pronounced Korean singing voices in parallel. N-Singer consists of a Transformerbased mel-generator, a convolutional network-based postnet, and voicing-aware discriminators. It can contribute in the following ways. First, for accurate pronunciation, N-Singer separately models linguistic and pitch information without other acoustic features. Second, to achieve improved mel-spectrograms, N-Singer uses a combination of Transformer-based modules and convolutional network-based modules. Third, in adversarial training, voicing-aware conditional discriminators are used to capture the harmonic features of voiced segments and noise components of unvoiced segments.

NOTES

The experimental results prove that N-Singer can synthesize a natural singing voice in parallel with a more accurate pronunciation than the baseline model.

Cross-Lingual Low Resource Speaker Adaptation Using Phonological Features

*Georgia Maniati*¹, *Nikolaos Ellinas*¹, *Konstantinos Markopoulos*¹, *Georgios Vamvoukakis*¹, *June Sig Sung*², *Hyoungmin Park*², *Aimilios Chalamandaris*¹, *Pirros Tsiakoulis*¹; ¹*Samsung, Greece;* ²*Samsung, Korea* Wed-M-V-3-2, Time: 11:00

The idea of using phonological features instead of phonemes as input to sequence-to-sequence TTS has been recently proposed for zero-shot multilingual speech synthesis. This approach is useful for code-switching, as it facilitates the seamless uttering of foreign text embedded in a stream of native text. In our work, we train a language-agnostic multispeaker model conditioned on a set of phonologically derived features common across different languages, with the goal of achieving cross-lingual speaker adaptation. We first experiment with the effect of language phonological similarity on cross-lingual TTS of several source-target language combinations. Subsequently, we fine-tune the model with very limited data of a new speaker's voice in either a seen or an unseen language, and achieve synthetic speech of equal quality, while preserving the target speaker's identity. With as few as 32 and 8 utterances of target speaker data, we obtain high speaker similarity scores and naturalness comparable to the corresponding literature. In the extreme case of only 2 available adaptation utterances, we find that our model behaves as a few-shot learner, as the performance is similar in both the seen and unseen adaptation language scenarios.

Improve Cross-Lingual Text-To-Speech Synthesis on Monolingual Corpora with Pitch Contour Information

Haoyue Zhan, Haitong Zhang, Wenjie Ou, Yue Lin; NetEase, China

Wed-M-V-3-3, Time: 11:00

Cross-lingual text-to-speech (TTS) synthesis on monolingual corpora is still a challenging task, especially when many kinds of languages are involved. In this paper, we improve the cross-lingual TTS model on monolingual corpora with pitch contour information. We propose a method to obtain pitch contour sequences for different languages without manual annotation, and extend the Tacotron-based TTS model with the proposed Pitch Contour Extraction (PCE) module. Our experimental results show that the proposed approach can effectively improve the naturalness and consistency of synthesized mixed-lingual utterances.

Cross-Lingual Voice Conversion with Disentangled Universal Linguistic Representations

*Zhenchuan Yang*¹, *Weibin Zhang*², *Yufei Liu*¹, *Xiaofen Xing*¹; ¹*SCUT, China*; ²*VoiceAI Technologies, China* Wed-M-V-3-4, Time: 11:00

Intra-lingual voice conversion has achieved great progress recently in terms of naturalness and similarity. However, in cross-lingual voice conversion, there is still an urgent need to improve the quality of the converted speech, especially with nonparallel training data. Previous works usually use Phonetic Posteriorgrams (PPGs) as the linguistic representations. In the case of cross-lingual voice conversion, the linguistic information is therefore represented as PPGs. It is well-known that PPGs may suffer from word dropping and mispronunciation, especially when the input speech is noisy. In addition, systems using PPGs can only convert the input into a known target language that is seen during training. This paper proposes an any-to-many voice conversion system based on disentangled universal linguistic representations (ULRs), which are extracted from a mix-lingual phoneme recognition system. Two methods are proposed to remove speaker information from ULRs. Experimental results show that the proposed method can effectively improve the converted speech objectively and subjectively. The system can also convert speech utterances naturally even if the language is not seen during training.

EfficientSing: A Chinese Singing Voice Synthesis System Using Duration-Free Acoustic Model and HiFi-GAN Vocoder

Zhengchen Liu, Chenfeng Miao, Qingying Zhu, Minchuan Chen, Jun Ma, Shaojun Wang, Jing Xiao; Ping An Technology, China Wed-M-V-3-5, Time: 11:00

In this paper, we present EfficientSing, a Chinese singing voice synthesis (SVS) system based on a non-autoregressive duration-free acoustic model and HiFi-GAN neural vocoder. Different from many existing SVS methods, no auxiliary duration prediction module is needed in this work, since a newly proposed monotonic alignment modeling mechanism is adopted. Moreover, we follow the non-autoregressive architecture of EfficientTTS with some singing-specific adaption, making training and inference fully parallel and efficient. HiFi-GAN vocoder is adopted to improve the voice quality of synthesized songs and inference efficiency. Both objective and subjective experimental results show that the proposed system can produce quite natural and high-fidelity songs and outperform the Tacotron-based baseline in terms of pronunciation, pitch and rhythm.

Cross-Lingual Speaker Adaptation Using Domain Adaptation and Speaker Consistency Loss for Text-To-Speech Synthesis

Detai Xin, Yuki Saito, Shinnosuke Takamichi, Tomoki Koriyama, Hiroshi Saruwatari; University of Tokyo, Japan

Wed-M-V-3-6, Time: 11:00

We present a cross-lingual speaker adaptation method based on domain adaptation and a speaker consistency loss for text-to-speech (TTS) synthesis. Existing monolingual speaker adaptation methods based on direct fine-tuning are not applicable for cross-lingual data. The proposed method first trains a language-independent speaker encoder by speaker verification using domain adaption on multilingual data, including the source and the target languages. Then the proposed method trains a monolingual multi-speaker TTS model on the source language's data using the speaker embeddings generated by the speaker encoder. To adapt the TTS model of the source language to new speakers the proposed method uses a speaker consistency loss to maximize the cosine similarity between speaker embeddings generated from the natural speech and the same speaker's synthesized speech. This makes fine-tuning the TTS model of source language on speech data of target language become possible. We conduct experiments on multi-speaker English and Japanese datasets with 207 speakers in total. Results of comprehensive experiments demonstrate that the proposed method can significantly improve speech naturalness compared to the baseline method.

Incorporating Cross-Speaker Style Transfer for Multi-Language Text-to-Speech

Zengqiang Shang¹, Zhihua Huang², Haozhe Zhang¹, Pengyuan Zhang¹, Yonghong Yan¹; ¹CAS, China; ²UCAS, China

Wed-M-V-3-7, Time: 11:00

Recently multilingual TTS systems using only monolingual datasets have obtained significant improvement. However, the quality of cross-language speech synthesis is not comparable to the speaker's own language and often comes with a heavy foreign accent. This paper proposed a multi-speaker multi-style multi-language speech synthesis system (M3), which improves the speech quality by introducing a fine-grained style encoder and overcomes the non-authentic accent problem through cross-speaker style transfer. To avoid leaking timbre information into style encoder, we utilized a speaker conditional variational encoder and conducted adversarial speaker training using the gradient reversal layer. Then, we built a Mixture Density Network (MDN) for mapping text to extracted style vectors for each speaker. At the inference stage, cross-language style transfer could be achieved by assigning any speaker's style type in the target language. Our system uses existing speaker style and genuinely avoids foreign accents. In the MOS-speech-naturalness, the proposed method generally achieves 4.0 and significantly outperform the baseline system.

Investigating Contributions of Speech and Facial Landmarks for Talking Head Generation

Ege Kesim, Engin Erzin; Koç University, Turkey Wed-M-V-3-8, Time: 11:00

Talking head generation is an active research problem. It has been widely studied as a direct speech-to-video or two stage speechto-landmarks-to-video mapping problem. In this study, our main motivation is to assess individual and joint contributions of the speech and facial landmarks to the talking head generation quality through a state-of-the-art generative adversarial network (GAN) architecture. Incorporating frame and sequence discriminators and a feature matching loss, we investigate performances of speech only, landmark only and joint speech and landmark driven talking head generation on the CREMA-D dataset. Objective evaluations using the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and landmark distance (LMD) indicate that while landmarks bring PSNR and SSIM improvements to the speech driven system, speech brings LMD improvement to the landmark driven system. Furthermore, feature matching is observed to improve the speech driven talking head generation models significantly.

Speech2Video: Cross-Modal Distillation for Speech to Video Generation

Shijing Si, Jianzong Wang, Xiaoyang Qu, Ning Cheng, Wenqi Wei, Xinghua Zhu, Jing Xiao; Ping An Technology, China

Wed-M-V-3-9, Time: 11:00

This paper investigates a novel task of talking face video generation solely from speeches. The speech-to-video generation technique can spark interesting applications in entertainment, customer service, and human-computer-interaction industries. Indeed, the timbre, accent and speed in speeches could contain rich information relevant to speakers' appearance. The challenge mainly lies in disentangling the distinct visual attributes from audio signals. In this article, we propose a light-weight, cross-modal distillation method to extract disentangled emotional and identity information from unlabelled video inputs. The extracted features are then integrated by a generative adversarial network into talking face video clips. With

carefully crafted discriminators, the proposed framework achieves realistic generation results. Experiments with observed individuals demonstrated that the proposed framework captures the emotional expressions solely from speeches, and produces spontaneous facial motion in the video output. Compared to the baseline method where speeches are combined with a static image of the speaker, the results of the proposed framework is almost indistinguishable. User studies also show that the proposed method outperforms the existing algorithms in terms of emotion expression in the generated videos.

Wed-M-V-4: Speech Coding and Privacy

11:00-13:00, Wednesday 1 September 2021 Chairs: Md. Sahidullah and Tom Bäckström

NU-Wave: A Diffusion Probabilistic Model for Neural Audio Upsampling

Junhyeok Lee, Seungu Han; MINDs Lab, Korea Wed-M-V-4-1, Time: 11:00

In this work, we introduce *NU-Wave*, the first neural audio upsampling model to produce waveforms of sampling rate 48kHz from coarse 16kHz or 24kHz inputs, while prior works could generate only up to 16kHz. NU-Wave is the first diffusion probabilistic model for audio super-resolution which is engineered based on neural vocoders. NU-Wave generates high-quality audio that achieves high performance in terms of signal-to-noise ratio (SNR), log-spectral distance (LSD), and accuracy of the ABX test. In all cases, NU-Wave outperforms the baseline models despite the substantially smaller model capacity (3.0M parameters) than baselines (5.4–21%). The audio samples of our model are publicly available, and the code will be made available soon.

QISTA-Net-Audio: Audio Super-Resolution via Non-Convex ℓ_a -Norm Minimization

Gang-Xuan Lin, Shih-Wei Hu, Yen-Ju Lu, Yu Tsao, Chun-Shien Lu; Academia Sinica, Taiwan Wed-M-V-4-2, Time: 11:00

Audio super-resolution (ASR) aims to reconstruct the high-resolution signal from its corresponding low-resolution one, which is hard while the correlation between them is low.

In this paper, we propose a learning model, QISTA-Net-Audio, to solve ASR in a paradigm of linear inverse problem. QISTA-Net-Audio is composed of two components. First, an audio waveform can be presented as a complex-valued spectrum, which is composed of a real and an imaginary part, in the frequency domain. We treat the real and imaginary parts as an image, and predict a high-resolution spectrum but only keep the phase information from the viewpoint of image reconstruction. Second, we predict the magnitude information by solving the sparse signal reconstruction problem. By combining the predicted magnitude and the phase together, we can recover the high-resolution waveform. Comparison with the state-of-the-art method MfNet [1], in terms of measure metrics SNR, PESQ, and STOI, demonstrates the superior performance of our method.

X-net: A Joint Scale Down and Scale Up Method for Voice Call

Liang Wen¹, Lizhong Wang¹, Xue Wen¹, Yuxing Zheng¹, Youngo Park², Kwang Pyo Choi²; ¹Samsung, China; ²Samsung, Korea Wed-M-V-4-3, Time: 11:00

This paper proposes X-net, a jointly learned scale-down and scaleup architecture for data pre- and post-processing in voice calls, as a means to bandwidth extension over band-limited channels. Scale-down and scale-up are deployed separately on transmitter and receiver to perform down- and upsampling. Separate supervisions are used on the submodules so that X-net can work properly even if one submodule is missing. A two-stage training method is used to learn X-net for improved perceptual quality. Results show that jointly learned X-net achieves promising improvement over blind audio super-resolution by both objective and subjective metrics, even in a lightweight implementation with only 1k parameters.

WSRGlow: A Glow-Based Waveform Generative Model for Audio Super-Resolution

*Kexun Zhang*¹, Yi Ren¹, *Changliang Xu*², *Zhou Zhao*¹; ¹*Zhejiang University, China*; ²*Xinhua News Agency, China*

Wed-M-V-4-4, Time: 11:00

Audio super-resolution is the task of constructing a high-resolution (HR) audio from a low-resolution (LR) audio by adding the missing band. Previous methods based on convolutional neural networks and mean squared error training objective have relatively low performance, while adversarial generative models are difficult to train and tune. Recently, normalizing flow has attracted a lot of attention for its high performance, simple training and fast inference. In this paper, we propose WSRGlow, a Glow-based waveform generative model to perform audio super-resolution. Specifically, 1) we integrate WaveNet and Glow to directly maximize the exact likelihood of the target HR audio conditioned on LR information; and 2) to exploit the audio information from low-resolution audio, we propose an LR audio encoder and an STFT encoder, which encode the LR information from the time domain and frequency domain respectively. The experimental results show that the proposed model is easier to train and outperforms the previous works in terms of both objective and perceptual quality. WSRGlow is also the first model to produce 48kHz waveforms from 12kHz LR audio. Audio samples are publicly available.

Half-Truth: A Partially Fake Audio Detection Dataset

Jiangyan Yi, Ye Bai, Jianhua Tao, Haoxin Ma, Zhengkun Tian, Chenglong Wang, Tao Wang, Ruibo Fu; CAS, China Wed-M-V-4-5, Time: 11:00

Diverse promising datasets have been designed to further the development of fake audio detection, such as ASVspoof databases. However, previous datasets ignore an attacking situation, in which the hacker hides some small fake clips in real speech audio. This poses a serious threat since that it is difficult to distinguish the small fake clip from the whole speech utterance. Therefore, this paper develops such a dataset for half-truth audio detection (HAD). Partially fake audio in the HAD dataset involves only changing a few words in an utterance. The audio of the words is generated with the very latest state-of-the-art speech synthesis technology. We can not only detect fake utterances but also localize manipulated regions in a speech using this dataset. Some benchmark results are presented on this dataset. The results show that partially fake audio presents much more challenging than fully fake audio for fake audio detection.

Data Quality as Predictor of Voice Anti-Spoofing Generalization

Bhusan Chettri¹, Rosa González Hautamäki¹, Md. Sahidullah², Tomi Kinnunen¹; ¹University of Eastern Finland, Finland; ²Loria (UMR 7503), France Wed-M-V-4-6, Time: 11:00

Voice anti-spoofing aims at classifying a given utterance either as a bonafide human sample, or a spoofing attack (e.g. synthetic or

replayed sample). Many anti-spoofing methods have been proposed but most of them fail to generalize across domains (corpora) — and we do not know *why*. We outline a novel interpretative framework for gauging the impact of data quality upon anti-spoofing performance. Our within- and between-domain experiments pool data from seven public corpora and three anti-spoofing methods based on Gaussian mixture and convolutive neural network models. We assess the impacts of long-term spectral information, speaker population (through x-vector speaker embeddings), signal-to-noise ratio, and selected voice quality features.

Coded Speech Enhancement Using Neural Network-Based Vector-Quantized Residual Features

Youngju Cheon¹, Soojoong Hwang¹, Sangwook Han¹, Inseon Jang², Jong Won Shin¹; ¹GIST, Korea; ²ETRI, Korea

Wed-M-V-4-7, Time: 11:00

Various approaches have been proposed to improve the quality of the speech coded at low bitrates. Recently, deep neural networks have also been used for speech coding, providing a high quality of speech with low bitrates. Although designing an entire codec with neural networks may be more effective, backward compatibility with the existing codecs can be desirable so that the systems with the legacy codec can still decode the coded bitstream. In this paper, we propose to generate side information based on neural networks for an existing codec and enhance the decoded speech with another neural networks using the side information. The vector-quantization variational autoencoder (VQ-VAE) is applied to generate vectorquantized side information and reconstruct the residual features, which are the difference between the features extracted from the original and decoded signals. The post-processor in the decoder side, which is another neural network, takes the decoded signal of the main codec and the reconstructed residual features to estimate the features for the original signal. Experimental results show that the proposed method can significantly improve the quality of the enhanced signals with additional bitrate of 0.6 kbps for two of the implementations of the high-efficiency advanced audio coding (HE-AAC) v1.

Multi-Channel Opus Compression for Far-Field Automatic Speech Recognition with a Fixed Bitrate Budget

Lukas Drude¹, Jahn Heymann¹, Andreas Schwarz¹, Jean-Marc Valin²; ¹Amazon, Germany; ²Amazon, USA Wed-M-V-4-8, Time: 11:00

Automatic speech recognition (ASR) in the cloud allows the use of larger models and more powerful multi-channel signal processing front-ends compared to on-device processing. However, it also adds an inherent latency due to the transmission of the audio signal, especially when transmitting multiple channels of a microphone array. One way to reduce the network bandwidth requirements is client-side compression with a lossy codec such as Opus. However, this compression can have a detrimental effect especially on multichannel ASR front-ends, due to the distortion and loss of spatial information introduced by the codec. In this publication, we propose an improved approach for the compression of microphone array signals based on Opus, using a modified joint channel coding approach and additionally introducing a multi-channel spatial decorrelating transform to reduce redundancy in the transmission. We illustrate the effect of the proposed approach on the spatial information retained in multi-channel signals after compression, and evaluate the performance on far-field ASR with a multi-channel beamforming front-end. We demonstrate that our approach can lead to a 37.5%

NOTES

bitrate reduction or a 5.1% relative word error rate (WER) reduction for a fixed bitrate budget in a seven channel setup.

Effects of Prosodic Variations on Accidental Triggers of a Commercial Voice Assistant

Ingo Siegert; OvG Universität Magdeburg, Germany Wed-M-V-4-9, Time: 11:00

The use of modern voice assistants has rapidly grown and they can be found in more and more households. By design, these systems have to scan every sound in their surroundings waiting for their respective wake-word before being able to react to the users' commands. The drawback of this method is that phonetic similar expressions can activate the voice assistant and thus speech utterances or whole private conversations will be recorded and streamed to the cloud back-end for further processing. Many news articles and scientific work reported on inaccurate wake-word detection. Resulting in at least a user's confusion or at worst security breaches. The current paper is based on a broader analysis of phonetic similar accidental triggers conducted by Schönherr et al., they presented a systematic analysis to detect accidental triggers, using a pronouncing dictionary and a weighted, phone-based Levenshtein distance. In this work, the previously identified accidental triggers are recorded by several speakers under various conditions to investigate the influence of phonetic variances (i.e. intonation and speaking/articulation rate) on the robustness of accidental triggers in a real-world environment.

Improving the Expressiveness of Neural Vocoding with Non-Affine Normalizing Flows

Adam Gabrys¹, Yunlong Jiao², Viacheslav Klimkov³, Daniel Korzekwa¹, Roberto Barra-Chicote²; ¹Amazon, Poland; ²Amazon, UK; ³Amazon, Germany Wed-M-V-4-10, Time: 11:00

This paper proposes a general enhancement to the Normalizing Flows (NF) used in neural vocoding. As a case study, we improve expressive speech vocoding with a revamped Parallel Wavenet (PW). Specifically, we propose to extend the affine transformation of PW to the more expressive invertible non-affine function. The greater expressiveness of the improved PW leads to better-perceived signal quality and naturalness in the waveform reconstruction and text-tospeech (TTS) tasks. We evaluate the model across different speaking styles on a multi-speaker, multi-lingual dataset. In the waveform reconstruction task, the proposed model closes the naturalness and signal quality gap from the original PW to recordings by 10%, and from other state-of-the-art neural vocoding systems by more than 60%. We also demonstrate improvements in objective metrics on the evaluation test set with L2 Spectral Distance and Cross-Entropy reduced by 3% and 6‰ comparing to the affine PW. Furthermore, we extend the probability density distillation procedure proposed by the original PW paper, so that it works with any non-affine invertible and differentiable function.

Voice Privacy Through x-Vector and CycleGAN-Based Anonymization

Gauri P. Prajapati, Dipesh K. Singh, Preet P. Amin, Hemant A. Patil; DA-IICT, India Wed-M-V-4-11, Time: 11:00

With the rise in usage of voice assistants and spoken language interfaces, important concerns regarding voice data privacy have been prompted. In an attempt to reduce the threat of attacks on voice data, in this paper, we propose a speaker anonymization system based on CycleGAN. This method modifies the speaker's gender and accent information from the original speech signal. The proposed method gives a more natural-sounding anonymized voice in addition to a deidentified speaker. We have chosen baseline-1 of The Voice Privacy Challenge-2020 as our baseline system. Training of CycleGAN, ASR, and ASV experiments are performed on the subset of Librispeech corpus. In this paper, the double anonymization technique is also explored in which the CycleGAN-based anonymization technique is adopted on top of the baseline system. Experimental results show that combining the proposed method with the x-vector and neural source-filter (NSF) model-based method (baseline system) gives up to 5.61% relative improvement in EER of original-anonymized, enrolltrial pairs. However, it gives up to 19.30% relative improvement in EER for anonymized-anonymized enroll-trial pairs. We observed that along with the good speaker de-identification, the anonymized utterances have adequate speech intelligibility and naturalness.

A Two-Stage Approach to Speech Bandwidth Extension

Ju Lin¹, Yun Wang², Kaustubh Kalgaonkar², Gil Keren², Didi Zhang², Christian Fuegen²; ¹Clemson University, USA; ²Facebook, USA Wed-M-V-4-12, Time: 11:00

Algorithms for speech bandwidth extension (BWE) may work in either the time domain or the frequency domain. Time-domain methods often do not sufficiently recover the high-frequency content of speech signals; frequency-domain methods are better at recovering the spectral envelope, but have difficulty reconstructing the details of the waveform. In this paper, we propose a two-stage approach for BWE, which enjoys the advantages of both time- and frequencydomain methods. The first stage is a frequency-domain neural network, which predicts the high-frequency part of the wide-band spectrogram from the narrow-band input spectrogram. The wideband spectrogram is then converted into a time-domain waveform, and passed through the second stage to refine the temporal details. For the first stage, we compare a convolutional recurrent network (CRN) with a temporal convolutional network (TCN), and find that the latter is able to capture long-span dependencies equally well as the former while using a lot fewer parameters. For the second stage, we enhance the Wave-U-Net architecture with a multi-resolution short-time Fourier transform (MSTFT) loss function. A series of comprehensive experiments show that the proposed system achieves superior performance in speech enhancement (measured by both time- and frequency-domain metrics) as well as speech recognition.

Development of a Psychoacoustic Loss Function for the Deep Neural Network (DNN)-Based Speech Coder

Joon Byun¹, Seungmin Shin¹, Youngcheol Park¹, Jongmo Sung², Seungkwon Beack²; ¹Yonsei University, Korea; ²ETRI, Korea Wed-M-V-4-13, Time: 11:00

This paper presents a loss function to compensate for the perceptual loss of the deep neural network (DNN)-based speech coder. By utilizing the psychoacoustic model (PAM), we design a loss function to maximize the mask-to-noise ratio (MNR) in multi-resolution Melfrequency scales. Also, a perceptual entropy (PE)-based weighting scheme is incorporated onto the MNR loss so that the DNN model focuses more on perceptually important Mel-frequency bands. The proposed loss function was tested on a CNN-based autoencoder implementing the softmax quantization and entropy-based bitrate control. Objective and subjective tests conducted with speech signals showed that the proposed loss function produced higher perceptual quality than the previous perceptual loss functions.

Protecting Gender and Identity with Disentangled Speech Representations

Dimitrios Stoidis, Andrea Cavallaro; Queen Mary University of London, UK Wed-M-V-4-14, Time: 11:00

Besides its linguistic content, our speech is rich in biometric information that can be inferred by classifiers. Learning privacy-preserving representations for speech signals enables downstream tasks without sharing unnecessary, private information about an individual. In this paper, we show that protecting gender information in speech is more effective than modelling speaker-identity information only when generating a non-sensitive representation of speech. Our method relies on reconstructing speech by decoding linguistic content along with gender information using a variational autoencoder. Specifically, we exploit disentangled representation learning to encode information about different attributes into separate subspaces that can be factorised independently. We present a novel way to encode gender information and disentangle two sensitive biometric identifiers, namely gender and identity, in a privacy-protecting setting. Experiments on the LibriSpeech dataset show that gender recognition and speaker verification can be reduced to a random guess, protecting against classification-based attacks.

Wed-M-V-5: Speech Perception II

11:00-13:00, Wednesday 1 September 2021 Chairs: Bernd Möbius and Anna Sfakianaki

Perception of Standard Arabic Synthetic Speech Rate

Yahya Aldholmi, Rawan Aldhafyan, Asma Alqahtani; King Saud University, Saudi Arabia Wed-M-V-5-1, Time: 11:00

This experiment investigated how Arabic speakers perceive synthetic Standard Arabic speech rate produced by Google TTS, at normal vs. accelerated rates. Twenty syntactically identical Standard Arabic sentences with a similar length (M=22 syllables per sentence, SD=1) were auditorily presented in a female voice to thirty female participants who were instructed to rate the tempo of the normal ($M \approx 4.5$ syllable per second) and accelerated (by 10%, 20%, and 30%) stimuli on a 1-7 Likert scale (1= extremely slow, 4= normal, 7= extremely fast). The results show that differences in the four-condition synthetic speech rates were reflected in the ratings provided by the participants: the more the speech was accelerated, the higher rating it received. More importantly, the findings support the observation that the current normal speech rate of Google TTS synthetic speech is not perceived as normal by Arabic speakers, but rather is perceived as slow. This may negatively affect the likelihood that users are comfortable using this technology. Hence, the outcome of this study does not only call for further investigation into Standard Arabic synthetic speech rates, but also reveals the need to define a baseline for a natural speech rate in Arabic.

The Influence of Parallel Processing on Illusory Vowels

Takeshi Kishiyama; University of Tokyo, Japan Wed-M-V-5-2, Time: 11:00

Research has shown that listeners perceive illusory vowels inside consonant clusters that are not allowed in their L1. This phenomenon has been examined using several psycholinguistic and computational models, including hidden Markov models (HMMs), applied to human phoneme perception. However, the inference algorithm of HMMs assumes that parallel processing, which has not been proven to have psychological reality, is a valid cognitive process. This study tested

the psychological reality of parallel processing by attempting to duplicate two results from previous studies: First, listeners perceive an illusory vowel in consonant clusters that are not permissible in their L1. Second, the illusory vowel is based on the characteristics of the preceding consonant, indicating that listeners integrate phonotactics and acoustic information. The experiment manipulated the number of candidates that the model can refer to, and the algorithm can be considered parallel when it allows models to use more than two candidates that are stored in memory. In addition, the transition probabilities between consonants were manipulated to represent the different phonotactics. The results showed that only the parallel processing condition reproduced the two observations above, supporting the psychological reality of parallel processing.

Exploring the Potential of Lexical Paraphrases for Mitigating Noise-Induced Comprehension Errors

Anupama Chingacham, Vera Demberg, Dietrich Klakow; Universität des Saarlandes, Germany Wed-M-V-5-3, Time: 11:00

Listening in noisy environments can be difficult even for individuals with a normal hearing thresholds. The speech signal can be masked by noise, which may lead to word misperceptions on the side of the listener, and overall difficulty to understand the message. To mitigate hearing difficulties on listeners, a co-operative speaker utilizes voice modulation strategies like Lombard speech to generate noise-robust utterances, and similar solutions have been developed for speech synthesis systems. In this work, we propose an alternate solution of choosing noise-robust lexical paraphrases to represent an intended meaning. Our results show that lexical paraphrases differ in their intelligibility in noise. We evaluate the intelligibility of synonyms in context and find that choosing a lexical unit that is less risky to be misheard than its synonym introduced an average gain in comprehension of 37% at SNR -5 dB and 21% at SNR 0 dB for babble noise.

SPEECHADJUSTER: A Tool for Investigating Listener Preferences and Speech Intelligibility

*Olympia Simantiraki*¹, *Martin Cooke*²; ¹*Universidad del País Vasco, Spain*; ²*Ikerbasque, Spain* Wed-M-V-5-4, Time: 11:00

Most of what we know about speech perception has been gleaned from tests in which listeners respond to stimuli chosen by an experimenter. This paper presents SPEECHADJUSTER, an open source tool that reverses the roles of listener and experimenter by allowing listeners direct control of speech characteristics in real-time. This change of paradigm enables listener preferences — reflecting factors such as cognitive effort, naturalness or distortion — to be measured directly, without recourse to rating scales. Incorporation of a test phase in which listener preferences are frozen also enables intelligibility to be estimated within the same trial. Offline computation and smooth online interpolation within the tool permits the impact of changes in practically any target speech feature (e.g. fundamental frequency or spectral slope) or background characteristic (e.g. noise spectrum), regardless of complexity, to be measured. The paper describes the tool's capabilities, presents a range of visualisations, and notes some potential applications and limitations.

VocalTurk: Exploring Feasibility of Crowdsourced Speaker Identification

Susumu Saito, Yuta Ide, Teppei Nakano, Tetsuji Ogawa; Waseda University, Japan Wed-M-V-5-5, Time: 11:00

This paper presents VocalTurk, a feasibility study of crowdsourced speaker identification based on our worker dataset collected in Amazon Mechanical Turk. Crowdsourced data labeling has already been acknowledged in speech data processing nowadays, but empirical analysis that answer to common questions such as "how accurate are workers capable of labeling speech data?" and "what does a good speech-labeling microtask interface look like?" still remain underexplored, which would limit the quality and scale of the dataset collection. Focusing on the speaker identification task in particular, we thus conducted two studies in Amazon Mechanical Turk: i) hired 3,800+ unique workers to test their performances and confidences in giving answers to voice pair comparison tasks, and ii) additionally assigned more-difficult tasks of 1-vs-N voice set comparisons to 350+ top-scoring workers to test their accuracy-speed performances across patterns of N = 1, 3, 5. The results revealed some positive findings that would motivate speech researchers toward crowdsourced data labeling, such as that the top-scoring workers were capable of giving labels to our voice comparison pairs with 99% accuracy after majority voting, as well as they were even capable of batch-labeling which significantly shortened up to 34% of their completion time but still with no statistically-significant degradation in accuracy.

Effects of Aging and Age-Related Hearing Loss on Talker Discrimination

*Min Xu*¹, *Jing Shao*², *Lan Wang*¹; ¹*CAS*, *China*; ²*HKBU*, *China*

Wed-M-V-5-6, Time: 11:00

Paralinguistic information is as important as linguistic information. Being familiar with talker's voice may facilitate speech perception, especially in challenging conditions. Previous studies have suggested that aging and age-related hearing loss lead to the deterioration of the phonetic and phonological processing ability. The current study aims to explore whether these two factors exert effects on the talker' voice discrimination. Three groups of participants, including young adults (YA) and older adults (OA) with and without hearing loss, were tested on talker discrimination in four types of stimuli varying in language familiarity: Mandarin real words, pseudowords, Arabic words and reversed Mandarin words. The results showed that OA with and without hearing loss performed worse than YA in both nonnative and native conditions. OA with hearing loss further performed worse than OA with normal hearing in Mandarin real word condition. These findings indicated that aging and hearing loss affected both low-level phonetic and high-level phonological processing, but hearing loss had extra effect on phonological processing. Altogether, these results implied that OA could not utilize phonetic and phonological cues as effectively as YA, and OA with hearing loss encountered more difficulties in utilizing phonological cues in talker discrimination.

Relationships Between Perceptual Distinctiveness, Articulatory Complexity and Functional Load in Speech Communication

Yuqing Zhang¹, Zhu Li¹, Bin Wu², Yanlu Xie¹, Binghuai Lin³, Jinsong Zhang¹; ¹BLCU, China; ²NAIST, Japan; ³Tencent, China

Wed-M-V-5-7, Time: 11:00

Work on communicative efficiency has hypothesized that phonological contrasts signaling more meaning distinctions (i.e., of high functional load (FL)) tend to have the least articulatory complexity and the highest perceptual salience. However, only a few studies have examined the preference for perceptual distinctiveness based on the traditional measures of FL (e.g., the number of minimal pairs, the change in entropy of the lexicon), which are weak in modeling contexts of individual words. And little attention has been devoted to investigating the need to minimize effort. This study explores

whether and how the communicative pressures to minimize the likelihood of confusion and minimize articulatory effort influence phonemic contrasts' functional contributions to speech communication. We used a revised definition of FL capable of modeling contextual information (i.e., the change in mutual information between phoneme sequences and spoken texts after the contrast in question is neutralized) and quantified information contributions of phonemic contrasts in English. The results indicated that FL of each phoneme pair increased significantly with its perceptual distinctiveness, and decreased significantly with articulatory complexity of the phoneme requiring less articulatory effort in the contrast. Altogether, these findings suggest that communicative pressures modulate the work a phonemic contrast does in distinguishing words.

Human Spoofing Detection Performance on Degraded Speech

*Camryn Terblanche*¹, *Philip Harrison*², *Amelia J. Gully*²; ¹*University of Cape Town, South Africa*; ²*University of York, UK* Wed-M-V-5-8, Time: 11:00

Over the past few years attention has been focused on the automatic detection of spoofing in the context of automatic speaker verification (ASV) systems. However, little is known about how well humans perform at detecting spoofed speech, particularly under degraded conditions. Using the latest synthesis technologies from ASVspoof 2019, this paper explores human judgements of speech authenticity by considering three common channel degradations a GSM network, a VoIP network, and background noise -– in conjunction with varying synthesis quality. The results reveal that channel degradation reduces the size of the perceptual difference between genuine and spoofed speech, and overall participants correctly identified human and spoofed speech only 56% of the time. In background noise and GSM transmission, lower-quality synthetic speech was judged as more human, and in VoIP transmission all speech, including genuine recordings, was judged as less human. Under all conditions, state-of-the-art synthetic speech was judged as human, or more human than, genuine recorded speech. The paper also considers the listener factors which may contribute to an individual's spoofing detection performance, and finds that a listener's familiarity with the accents involved, their age, and the audio equipment used for playback, have an effect on their spoofing detection performance.

Reliable Estimates of Interpretable Cue Effects with Active Learning in Psycholinguistic Research

Marieke Einfeldt¹, Rita Sevastjanova¹, Katharina Zahner-Ritter², Ekaterina Kazak³, Bettina Braun¹; ¹Universität Konstanz, Germany; ²Universität Trier, Germany; ³University of Manchester, UK Wed-M-V-5-9, Time: 11:00

Studying the relative weighting of different cues for the interpretation of a linguistic phenomenon is a core element in psycholinguistic research. This research needs to strike a balance between two things: generalisability to diverse lexical settings, which requires a high number of different lexicalisations and the investigation of a large number of different cues, which requires a high number of different test conditions. Optimizing both is impossible with classical psycholinguistic designs as this would leave the participants with too many experimental trials. Previously we showed that Active Learning (AL) systems allow to test numerous conditions (eight) and items (32) within the same experiment. As stimulus selection was informed by the system's learning mechanism, AL sped-up the labelling process. In the present study, we extend the use case to an experiment with 16 conditions, manipulated through four binary factors (the experimental setting and three prosodic cues; two levels each). Our findings show that the AL system correctly predicted the intended result pattern after twelve trials only. Hence, AL further confirmed previous findings and proved to be an efficient tool, which offers a promising solution to complex study designs in psycholinguistic research.

Towards the Explainability of Multimodal Speech Emotion Recognition

*Puneet Kumar*¹, *Vishesh Kaushik*², *Balasubramanian Raman*¹; ¹*IIT Roorkee, India*; ²*IIT Kanpur, India* Wed-M-V-5-10, Time: 11:00

In this paper, a multimodal speech emotion recognition system has been developed, and a novel technique to explain its predictions has been proposed. The audio and textual features are extracted separately using attention-based Gated Recurrent Unit (GRU) and pre-trained Bidirectional Encoder Representations from Transformers (BERT), respectively. Then they are concatenated and used to predict the final emotion class. The weighted and unweighted emotion recognition accuracy of 71.7% and 75.0% has been achieved on Emotional Dyadic Motion Capture (IEMOCAP) dataset containing speech utterances and corresponding text transcripts. The training and predictions of network layers have been analyzed qualitatively through emotion embedding plots and quantitatively by analyzing the intersection matrices for various emotion classes' embeddings.

Primacy of Mouth over Eyes: Eye Movement Evidence from Audiovisual Mandarin Lexical Tones and Vowels

Biao Zeng¹, Rui Wang², Guoxing Yu³, Christian Dobel⁴; ¹University of South Wales, UK; ²Guangdong Pharmaceutical University, China; ³University of Bristol, UK; ⁴FSU Jena, Germany Wed-M-V-5-11, Time: 11:00

This study investigated Chinese speakers' eye movements when they were asked to identify audiovisual Mandarin lexical tones and vowels. In the lexical tone identification task, Chinese speakers were presented with an audiovisual clip of Mandarin monosyllables ($/\check{a}$ /, $/\check{a}$ /, $/\check{i}$ /, $/\check{i}$ /) and asked to identify whether the syllables were presented in a dipping ($/\check{a}$ /, $/\check{i}$ /) or falling tone ($/\check{a}$ /, $/\check{i}$ /). In the vowel identification task, they were asked to identify whether the vowels were /a/ or /i/ regardless of lexical tone. These audiovisual syllables were presented in clear, noisy, and silent conditions. An eye-tracker recorded the participants' eye movements.

Results showed participants gazed more at the mouth than the eyes in both lexical tones and vowels. Additionally, when acoustic conditions degraded from clear to noisy and eventually silent, Chinese speakers increased their gaze towards the mouth rather than the eyes. These findings suggest the mouth to be the primary area that is utilised during audiovisual speech perception. The similar patterns of eye movements between vowels and lexical tones indicate that the mouth acts as a perceptual cue that provides articulatory information.

Investigating the Impact of Spectral and Temporal Degradation on End-to-End Automatic Speech Recognition Performance

Takanori Ashihara, Takafumi Moriya, Makio Kashino;

NTT, Japan

Wed-M-V-5-12, Time: 11:00

Humans have a sophisticated capability to robustly handle incomplete sensory input, as often happens in real environments. In earlier studies, the robustness of human speech perception was observed

qualitatively by spectrally and temporally degraded stimuli. The current study investigates how machine speech recognition, especially end-to-end automatic speech recognition (E2E-ASR), can yield similar robustness against distorted acoustic cues. To evaluate the performance of E2E-ASR, we employ four types of distorted speech based on previous studies: locally time-reversed speech, noise-vocoded speech, phonemic restoration, and modulation-filtered speech. Those stimuli are synthesized by spectral and/or temporal manipulation from original speech samples whose human speech intelligibility scores have been well-reported. An experiment was conducted on the TED-LIUM2 for English and the Corpus of Spontaneous Japanese (CSJ) for Japanese. We found that while there is a tendency to exhibit similar robustness in some experiments, full recovery from the harmful effect of the severe spectral degradation is not achieved.

Wed-M-V-6: Streaming for ASR/RNN Transducers

11:00-13:00, Wednesday 1 September 2021 Chairs: Mark Gales and Ralf Schlueter

Super-Human Performance in Online Low-Latency Recognition of Conversational Speech

Thai-Son Nguyen, Sebastian Stüker, Alex Waibel; KIT, Germany

Wed-M-V-6-1, Time: 11:00

Achieving super-human performance in recognizing human speech has been a goal for several decades as researchers have worked on increasingly challenging tasks. In the 1990's it was discovered, that conversational speech between two humans turns out to be considerably more difficult than read speech as hesitations, disfluencies, false starts and sloppy articulation complicate acoustic processing and require robust joint handling of acoustic, lexical and language context. Early attempts with statistical models could only reach word error rates (WER) of over 50% which is far from human performance with shows a WER of around 5.5%. Neural hybrid models and recent attention-based encoder-decoder models have considerably improved performance as such contexts can now be learned in an integral fashion. However, processing such contexts requires an entire utterance presentation and thus introduces unwanted delays before a recognition result can be output. In this paper, we address performance as well as latency. We present results for a system that can achieve super-human performance, i.e. a WER of 5.0% on the Switchboard conversational benchmark, at a word based latency of only 1 second behind a speaker's speech. The system uses multiple attention-based encoder-decoder networks integrated within a novel low latency incremental inference approach.

Multiple Softmax Architecture for Streaming Multilingual End-to-End ASR Systems

Vikas Joshi¹, Amit Das², Eric Sun², Rupesh R. Mehta¹, Jinyu Li², Yifan Gong²; ¹Microsoft, India; ²Microsoft, USA

Wed-M-V-6-2, Time: 11:00

Improving multilingual end-to-end (E2E) automatic speech recognition (ASR) systems have manifold advantages. They simplify the training strategy, are easier to scale and exhibit better performance over monolingual models. However, it is still challenging to use a single multilingual model to recognize multiple languages without knowing the input language, as most multilingual models assume the availability of the input language. In this paper, we introduce multisoftmax model to improve the multilingual recurrent neural network

NOTES

transducer (RNN-T) models, by having language specific softmax, joint and embedding layers, while sharing rest of the parameters. We extend the multi-softmax model to work without knowing the input language, by integrating a language identification (LID) model, that estimates the LID on-the-fly and also does the recognition at the same time. The multi-softmax model outperforms monolingual models with an average word error rate relative (WERR) reduction of 4.65% on Indian languages. Finetuning further improves the WERR reduction to 12.2%. The multi-softmax model with on-the-fly LID estimation, shows WERR reduction of 13.86% compared to the multilingual baseline.

Contextualized Streaming End-to-End Speech Recognition with Trie-Based Deep Biasing and Shallow Fusion

Duc Le, Mahaveer Jain, Gil Keren, Suyoun Kim, Yangyang Shi, Jay Mahadeokar, Julian Chan, Yuan Shangguan, Christian Fuegen, Ozlem Kalinli, Yatharth Saraf, Michael L. Seltzer; Facebook, USA Wed-M-V-6-3, Time: 11:00

How to leverage dynamic contextual information in end-to-end speech recognition has remained an active research area. Previous solutions to this problem were either designed for specialized use cases that did not generalize well to open-domain scenarios, did not scale to large biasing lists, or underperformed on rare long-tail words. We address these limitations by proposing a novel solution that combines shallow fusion, trie-based deep biasing, and neural network language model contextualization. These techniques result in significant 19.5% relative Word Error Rate improvement over existing contextual biasing approaches and 5.4%-9.3% improvement compared to a strong hybrid baseline on both open-domain and constrained contextualization tasks, where the targets consist of mostly rare long-tail words. Our final system remains lightweight and modular, allowing for quick modification without model re-training.

An Efficient Streaming Non-Recurrent On-Device End-to-End Model with Improvements to Rare-Word Modeling

Tara N. Sainath, Yanzhang He, Arun Narayanan, Rami Botros, Ruoming Pang, David Rybach, Cyril Allauzen, Ehsan Variani, James Qin, Quoc-Nam Le-The, Shuo-Yiin Chang, Bo Li, Anmol Gulati, Jiahui Yu, Chung-Cheng Chiu, Diamantino Caseiro, Wei Li, Qiao Liang, Pat Rondon; Google, USA Wed-M-V-6-4, Time: 11:00

On-device end-to-end (E2E) models have shown improvements over a conventional model on Search test sets in both quality, as measured by Word Error Rate (WER) [1], and latency [2], measured by the time the result is finalized after the user stops speaking. However, the E2E model is trained on a small fraction of audio-text pairs compared to the 100 billion text utterances that a conventional language model (LM) is trained with. Thus E2E models perform poorly on rare words and phrases. In this paper, building upon the two-pass streaming Cascaded Encoder E2E model [3], we explore using a Hybrid Autoregressive Transducer (HAT) [4] factorization to better integrate an on-device neural LM trained on text-only data. Furthermore, to further improve decoder latency we introduce a non-recurrent embedding decoder, in place of the typical LSTM decoder, into the Cascaded Encoder model. Overall, we present a streaming on-device model that incorporates an external neural LM and outperforms the conventional model in both search and rare-word quality, as well as latency, and is $318 \times$ smaller.

Streaming Multi-Talker Speech Recognition with Joint Speaker Identification

Liang Lu, Naoyuki Kanda, Jinyu Li, Yifan Gong; Microsoft, USA_____

Wed-M-V-6-5, Time: 11:00

In multi-talker scenarios such as meetings and conversations, speech processing systems are usually required to transcribe the audio as well as identify the speakers for downstream applications. Since overlapped speech is common in this case, conventional approaches usually address this problem in a cascaded fashion that involves speech separation, speech recognition and speaker identification that are trained independently. In this paper, we propose Streaming Unmixing, Recognition and Identification Transducer (SURIT) — a new framework that deals with this problem in an end-to-end streaming fashion. SURIT employs the recurrent neural network transducer (RNN-T) as the backbone for both speech recognition and speaker identification. We validate our idea on the LibrispeechMix dataset — a multi-talker dataset derived from Librispeech, and present encouraging results.

Streaming End-to-End Speech Recognition for Hybrid RNN-T/Attention Architecture

Takafumi Moriya, Tomohiro Tanaka, Takanori Ashihara, Tsubasa Ochiai, Hiroshi Sato, Atsushi Ando, Ryo Masumura, Marc Delcroix, Taichi Asami; NTT, Japan

Wed-M-V-6-6, Time: 11:00

We present a novel architecture with its decoding approach for improving recurrent neural network-transducer (RNN-T) performance. RNN-T is promising for building time-synchronous automatic speech recognition (ASR) systems and thus enhancing streaming ASR applications. We note that encoder-decoder-based sequence-to-sequence models (S2S) have been also used successfully by the ASR community. In this paper, we integrate these popular models in the RNN-T+S2S approach; higher recognition performance than either is achieved due to their integration. However, it is generally deemed to be complicated to use S2S in streaming systems, because the attention mechanism can use arbitrarily long past and future contexts during decoding. Our RNN-T+S2S is composed of the shared encoder, an RNN-T decoder and a triggered attention-based decoder which uses time restricted encoder outputs for attention weight computation. By using the trigger points generated from RNN-T outputs, the S2S branch of RNN-T+S2S activates only when the triggers are detected, which makes streaming ASR practical. Experiments on public and private datasets created to research various tasks demonstrate that our proposal can yield superior recognition performance.

Improving RNN-T ASR Accuracy Using Context Audio

Andreas Schwarz, Ilya Sklyar, Simon Wiesler; Amazon,

Germany

Wed-M-V-6-7, Time: 11:00

We present a training scheme for streaming automatic speech recognition (ASR) based on recurrent neural network transducers (RNN-T) which allows the encoder network to learn to exploit context audio from a stream, using segmented or partially labeled sequences of the stream during training. We show that the use of context audio during training and inference can lead to word error rate reductions of more than 6% in a realistic production setting for a voice assistant ASR system. We investigate the effect of the proposed training approach on acoustically challenging data containing background speech and present data points which indicate that this approach helps the network learn both speaker and environment adaptation. To gain further insight into the ability of a long short-term memory (LSTM)

NOTES

based ASR encoder to exploit long-term context, we also visualize RNN-T loss gradients with respect to the input.

HMM-Free Encoder Pre-Training for Streaming RNN Transducer

Lu Huang, Jingyu Sun, Yufeng Tang, Junfeng Hou, Jinkun Chen, Jun Zhang, Zejun Ma; ByteDance, China Wed-M-V-6-8, Time: 11:00

This work describes an encoder pre-training procedure using frame-wise label to improve the training of streaming recurrent neural network transducer (RNN-T) model. Streaming RNN-T trained from scratch usually performs worse than non-streaming RNN-T. Although it is common to address this issue through pre-training components of RNN-T with other criteria or frame-wise alignment guidance, the alignment is not easily available in end-to-end manner. In this work, frame-wise alignment, used to pre-train streaming RNN-T's encoder, is generated without using a HMM-based system. Therefore an all-neural framework equipping HMM-free encoder pre-training is constructed. This is achieved by expanding the spikes of CTC model to their left/right blank frames, and two expanding strategies are proposed. To our best knowledge, this is the first work to simulate HMM-based frame-wise label using CTC model for pre-training. Experiments conducted on LibriSpeech and MLS English tasks show the proposed pre-training procedure, compared with random initialization, reduces the WER by relatively 5%~11% and the emission latency by 60 ms. Besides, the method is lexicon-free, so it is friendly to new languages without manually designed lexicon.

Reducing Exposure Bias in Training Recurrent Neural Network Transducers

Xiaodong Cui, Brian Kingsbury, George Saon, David Haws, Zoltán Tüske; IBM, USA Wed-M-V-6-9, Time: 11:00

When recurrent neural network transducers (RNNTs) are trained using the typical maximum likelihood criterion, the prediction network is trained only on ground truth label sequences. This leads to a mismatch during inference, known as exposure bias, when the model must deal with label sequences containing errors. In this paper we investigate approaches to reducing exposure bias in training to improve the generalization of RNNT models for automatic speech recognition (ASR). A label-preserving input perturbation to the prediction network is introduced. The input token sequences are perturbed using SwitchOut and scheduled sampling based on an additional token language model. Experiments conducted on the 300-hour SwitchDoard dataset demonstrate their effectiveness. By reducing the exposure bias, we show that we can further improve the accuracy of a high-performance RNNT ASR model and obtain state-of-the-art results on the 300-hour SwitchDoard dataset.

Bridging the Gap Between Streaming and Non-Streaming ASR Systems by Distilling Ensembles of CTC and RNN-T Models

Thibault Doutre, Wei Han, Chung-Cheng Chiu, Ruoming Pang, Olivier Siohan, Liangliang Cao; Google, USA Wed-M-V-6-10, Time: 11:00

Streaming end-to-end automatic speech recognition (ASR) systems are widely used in everyday applications that require transcribing speech to text in real-time. Their minimal latency makes them suitable for such tasks. Unlike their non-streaming counterparts, streaming models are constrained to be causal with no future context and suffer from higher word error rates (WER). To improve streaming models, a recent study [1] proposed to distill a non-streaming teacher model on unsupervised utterances, and then train a streaming student using the teachers' predictions. However, the performance gap between teacher and student WERs remains high. In this paper, we aim to close this gap by using a diversified set of non-streaming teacher models and combining them using Recognizer Output Voting Error Reduction (ROVER). In particular, we show that, despite being weaker than RNN-T models, CTC models are remarkable teachers. Further, by fusing RNN-T and CTC models together, we build the strongest teachers. The resulting student models drastically improve upon streaming models of previous work [1]: the WER decreases by 41% on Spanish, 27% on Portuguese, and 13% on French.

Mixture Model Attention: Flexible Streaming and Non-Streaming Automatic Speech Recognition

Kartik Audhkhasi, Tongzhou Chen, Bhuvana Ramabhadran, Pedro J. Moreno; Google, USA Wed-M-V-6-11, Time: 11:00

Streaming automatic speech recognition (ASR) hypothesizes words as soon as the input audio arrives, whereas non-streaming ASR can potentially wait for the completion of the entire utterance to hypothesize words. Streaming and non-streaming ASR systems have typically used different acoustic encoders. Recent work has attempted to unify them by either jointly training a fixed stack of streaming and non-streaming layers or using knowledge distillation during training to ensure consistency between the streaming and non-streaming predictions. We propose mixture model (MiMo) attention as a simpler and theoretically-motivated alternative that replaces only the attention mechanism, requires no change to the training loss, and allows greater flexibility of switching between streaming and non-streaming mode during inference. Our experiments on the public Librispeech data set and a few Indic language data sets show that MiMo attention endows a single ASR model with the ability to operate in both streaming and non-streaming modes without any overhead and without significant loss in accuracy compared to separately-trained streaming and non-streaming models. We also illustrate this benefit of MiMo attention in a second-pass rescoring setting.

StableEmit: Selection Probability Discount for Reducing Emission Latency of Streaming Monotonic Attention ASR

Hirofumi Inaguma, Tatsuya Kawahara; Kyoto University, Japan Wed-M-V-6-12, Time: 11:00

While attention-based encoder-decoder (AED) models have been successfully extended to the online variants for streaming automatic speech recognition (ASR), such as monotonic chunkwise attention (MoChA), the models still have a large label emission latency because of the unconstrained end-to-end training objective. Previous works tackled this problem by leveraging alignment information to control the timing to emit tokens during training. In this work, we propose a simple alignment-free regularization method, StableEmit, to encourage MoChA to emit tokens earlier. StableEmit discounts the selection probabilities in hard monotonic attention for token boundary detection by a constant factor and regularizes them to recover the total attention mass during training. As a result, the scale of the selection probabilities is increased, and the values can reach a threshold for token emission earlier, leading to a reduction of emission latency and deletion errors. Moreover, StableEmit can be combined with methods that constraint alignments to further improve the accuracy and latency. Experimental evaluations with LSTM and Conformer encoders demonstrate that StableEmit significantly reduces the recognition errors and the emission latency simultaneously. We also show that the use of alignment information is complementary in both metrics.

NOTES

Dual Causal/Non-Causal Self-Attention for Streaming End-to-End Speech Recognition

Niko Moritz, Takaaki Hori, Jonathan Le Roux; MERL, USA Wed-M-V-6-13, Time: 11:00

Attention-based end-to-end automatic speech recognition (ASR) systems have recently demonstrated state-of-the-art results for numerous tasks. However, the application of self-attention and attention-based encoder-decoder models remains challenging for streaming ASR, where each word must be recognized shortly after it was spoken. In this work, we present the dual causal/non-causal self-attention (DCN) architecture, which in contrast to restricted self-attention prevents the overall context to grow beyond the look-ahead of a single layer when used in a deep architecture. DCN is compared to chunk-based and restricted self-attention using streaming transformer and conformer architectures, showing improved ASR performance over restricted self-attention and competitive ASR results compared to chunk-based self-attention, while providing the advantage of frame-synchronous processing. Combined with triggered attention, the proposed streaming end-to-end ASR systems obtained state-of-the-art results on the LibriSpeech, HKUST, and Switchboard ASR tasks.

Multi-Mode Transformer Transducer with Stochastic Future Context

Kwangyoun Kim¹, Felix Wu¹, Prashant Sridhar¹, Kyu J. Han¹, Shinji Watanabe²; ¹ASAPP, USA; ²Carnegie Mellon University, USA Wed-M-V-6-14, Time: 11:00

Automatic speech recognition (ASR) models make fewer errors when more surrounding speech information is presented as context. Unfortunately, acquiring a larger future context leads to higher latency. There exists an inevitable trade-off between speed and accuracy. Naïvely, to fit different latency requirements, people have to store multiple models and pick the best one under the constraints. Instead, a more desirable approach is to have a single model that can dynamically adjust its latency based on different constraints, which we refer to as Multi-mode ASR. A Multi-mode ASR model can fulfill various latency requirements during inference — when a larger latency becomes acceptable, the model can process longer future context to achieve higher accuracy and when a latency budget is not flexible, the model can be less dependent on future context but still achieve reliable accuracy. In pursuit of Multi-mode ASR, we propose Stochastic Future Context, a simple training procedure that samples one streaming configuration in each iteration. Through extensive experiments on AISHELL-1 and LibriSpeech datasets, we show that a Multi-mode ASR model rivals, if not surpasses, a set of competitive streaming baselines trained with different latency budgets.

Wed-M-SS-1: ConferencingSpeech 2021 Challenge: Far-Field Multi-Channel Speech Enhancement for Video Conferencing

Room Lacina, 11:00–13:00, Wednesday 1 September 2021 Chairs: Shinji Watanabe and Wei Rao

Introduction of Challenge

A Causal U-Net Based Neural Beamforming Network for Real-Time Multi-Channel Speech Enhancement

Xinlei Ren, Xu Zhang, Lianwu Chen, Xiguang Zheng, Chen Zhang, Liang Guo, Bing Yu; Kuaishou Technology, China

Wed-M-SS-1-1, Time: 11:20

People are meeting through video conferencing more often. While single channel speech enhancement techniques are useful for the individual participants, the speech quality will be significantly degraded in large meeting rooms where the far-field and reverberate conditions are introduced. Approaches based on microphone array signal processing are proposed to explore the inter-channel correlation among the individual microphone channels. In this work, a new causal U-net based multiple-in-multiple-out structure is proposed for real-time multi-channel speech enhancement. The proposed method incorporates the traditional beamforming structure with the multi-channel causal U-net by explicitly adding a beamforming operation at the end of the neural beamformer. The proposed method has entered the INTERSPEECH Far-field Multi-Channel Speech Enhancement Challenge for Video Conferencing. With 1.97M model parameters and 0.25 real-time factor on Intel Core i7 (2.6GHz) CPU, the proposed method has outperforms the baseline system of this challenge on PESQ, Si-SNR and STOI metrics.

Short Presentations 1

Time: 11:40

Short Presentations 2 Time: 12:00

Panel Discussion Time: 12:20

A Partitioned-Block Frequency-Domain Adaptive Kalman Filter for Stereophonic Acoustic Echo Cancellation

*Rui Zhu*¹, *Feiran Yang*², *Yuepeng Li*¹, *Shidong Shang*¹; ¹*Tencent, China*; ²*CAS, China* Wed-M-SS-1-2, Time: 12:30

The rapid development of online video conferencing systems has caused renewed attention to the multi-channel recording and playback systems. Stereophonic acoustic echo cancellation (SAEC) is the key issue of this systems. This paper proposes an optimally designed partitioned-block frequency-domain Kalman filter (PBFDKF) algorithm for SAEC. We establish the frequency-domain observation equation using the overlap-and-save method and we use the first-order Markov model to describe the state equation. The exact PBFDKF algorithm is derived under the umbrella of Kalman filter theory and two fast implementations are then presented to reduce the complexity. The proposed algorithm is equivalent to the

NOTES

dual-channel partitioned-block frequency-domain gradient-based algorithm with optimum step-size control, and hence it exhibits very good convergence performance and is found to be robust to near-end interference without a double-talk detector. Extensive experiments in different SAEC conditions confirm the effectiveness of the proposed algorithm.

Real-Time Independent Vector Analysis Using Semi-Supervised Nonnegative Matrix Factorization as a Source Model

*Taihui Wang*¹, *Feiran Yang*¹, *Rui Zhu*², *Jun Yang*¹; ¹*CAS, China*; ²*Tencent, China* Wed-M-SS-1-3, Time: 12:30

Online independent vector analysis (IVA) based on auxiliary technology is effective to separate audio source in real time. However, the separated signal may contain residual interference noise because the source model of IVA lacks flexibility and cannot treat the specific harmonic structures of sources. This paper presents a real-time IVA method where the amplitude spectrum of separated signal is modeled by semi-supervised nonnegative matrix factorization (SSNMF). Using the pre-trained basis matrix which contains source structures, we can extract the target source from the separated signal in real time. The advantage of the proposed method is that the extracted source can provide a more accurate variance than the separated signal and hence the proposed method can obtain a better separation performance than the oracle IVA. Experimental results in speech denoising task show the effectiveness and the robustness of the proposed method with different types of noise.

Improving Channel Decorrelation for Multi-Channel Target Speech Extraction

Jiangyu Han¹, Wei Rao², Yannan Wang², Yanhua Long¹; ¹Shanghai Normal University, China; ²Tencent, China

Wed-M-SS-1-4, Time: 12:30

Target speech extraction has attracted widespread attention. When microphone arrays are available, the additional spatial information can be helpful in extracting the target speech. We have recently proposed a channel decorrelation (CD) mechanism to extract the inter-channel differential information to enhance the reference channel encoder representation. Although the proposed mechanism has shown promising results for extracting the target speech from mixtures, the extraction performance is still limited by the nature of the original decorrelation theory. In this paper, we propose two methods to broaden the horizon of the original channel decorrelation, by replacing the original softmax-based inter-channel similarity between encoder representations, using an unrolled probability and a normalized cosine-based similarity at the dimensional-level. Moreover, new combination strategies of the CD-based spatial information and target speaker adaptation of parallel encoder outputs are also investigated. Experiments on the reverberant WSJ0 2-mix show that the improved CD can result in more discriminative differential information and the new adaptation strategy is also very effective to improve the target speech extraction.

Inplace Gated Convolutional Recurrent Neural Network for Dual-Channel Speech Enhancement

Jinjiang Liu, Xueliang Zhang; Inner Mongolia University, China

Wed-M-SS-1-5, Time: 12:30

For dual-channel speech enhancement, it is a promising idea to design an end-to-end model based on the traditional array signal pro-

cessing guideline and the manifold space of multi-channel signals. We found that the idea above can be effectively implemented by the classical convolutional recurrent neural networks (CRN) architecture. We propose a very compact inplace gated convolutional recurrent neural network (inplace GCRN) for end-to-end multi-channel speech enhancement, which utilizes inplace-convolution for frequency pattern extraction and reconstruction. The inplace characteristics efficiently preserve spatial cues in each frequency bin for channel-wise long short-term memory neural networks (LSTM) tracing the spatial source. In addition, we come up with a new spectrum recovery method by predict amplitude mask, mapping, and phase, which effectively improves the speech quality.

SRIB-LEAP Submission to Far-Field Multi-Channel Speech Enhancement Challenge for Video Conferencing

*R.G. Prithvi Raj*¹, *Rohit Kumar*², *M.K. Jayesh*¹, *Anurenjan Purushothaman*², *Sriram Ganapathy*², *M.A. Basha Shaik*¹; ¹*Samsung, India*; ²*Indian Institute of Science, India* Wed-M-SS-1-6, Time: 12:30

This paper presents the details of the SRIB-LEAP submission to the ConferencingSpeech challenge 2021. The challenge involved the task of multi-channel speech enhancement to improve the quality of far field speech from microphone arrays in a video conferencing room. We propose a two stage method involving a beamformer followed by single channel enhancement. For the beamformer, we incorporated self-attention mechanism as inter-channel processing layer in the filter-and-sum network (FaSNet), an end-to-end time-domain beamforming system. The single channel speech enhancement is done in log spectral domain using convolution neural network (CNN)-long short term memory (LSTM) based architecture. We achieved improvements in objective quality metrics — perceptual evaluation of speech quality (PESQ) of 0.5 on the noisy data. On subjective quality evaluation, the proposed approach improved the mean opinion score (MOS) by an absolute measure of 0.9 over the noisy audio.

Real-Time Multi-Channel Speech Enhancement Based on Neural Network Masking with Attention Model

*Cheng Xue*¹, *Weilong Huang*¹, *Weiguang Chen*¹, *Jinwei Feng*²; ¹*Alibaba, China*; ²*Alibaba, USA* Wed-M-SS-1-7, Time: 12:30

In this paper, we propose a real-time multi-channel speech enhancement method for noise reduction and dereverberation in far-field environments. The proposed method consists of two components: differential beamforming and mask estimation network. The differential beamforming is employed to suppress the interference signals from non-target directions such that a relatively clean speech can be obtained. The mask estimation network with an attention model is developed to capture the signal correlation among different channels in the feature extraction stage and enhance the feature representation that needs to be reconstructed into the target speech in the estimation mask stage. In the inference phase, the spectrum after differential beamforming is filtered by the estimated mask to obtain the final output. The spectrum after differential beamforming can provide a higher signal-to-noise ratio (SNR) than the original spectrum, so the estimated mask can more easily filter out the noise. We conducted experiments on the ConferencingSpeech2021 challenge (INTERSPEECH 2021) dataset to evaluate the proposed method. With only 2.9M parameters, the proposed method achieved competitive performance.

Wed-Survey: Survey Talk 2: Sriram Ganapathy Room A+B, 13:00-14:00, Wednesday 1 September 2021

Chairs: TBD

Uncovering the Acoustic Cues of COVID-19 Infection

Sriram Ganapathy; Indian Institute of Science, India Wed-Survey, Time: 13:00

The investigation of acoustic biomarkers of respiratory diseases has societal and public health impact following the onset of COVID-19 pandemic. The efforts in the pre-pandemic period focused on developing smartphone friendly diagnostic tools for the detection of chronic pulmonary diseases, Tuberculosis and asthmatic conditions using cough sounds. During the past two years, several research works of varying scales have been undertaken by the speech and signal processing community for analyzing the acoustic symptoms of COVID. The motivation for the development of acoustic-based tools for COVID diagnostics arises from the key limitations of cost, time, and safety of the current gold standard in COVID testing, namely the reverse transcription polymerase chain reaction (RT-PCR) testing.

In this talk, I will survey the major efforts undertaken by groups across the world in i) developing data resources of acoustic signals for COVID-19 diagnostics, and ii) designing models and learning algorithms for tool development. The landscape of data resources ranges from controlled hospital recordings to crowdsourced smartphonebased data. While the primary signal modality recorded is the cough data, the impact of COVID on other modalities like breathing, speech and symptom data are also studied. In the talk, I will also discuss the considerations in designing data representations and machine learning models for COVID detection from acoustic data. The pointers to open-source data resources and tools will be highlighted with the aim of encouraging budding researchers to pursue this important direction.

The talk will conclude by remarking about the progress made by our group, Coswara, where a multi-modal combination of information from several modalities shows the potential to surpass regulatory requirements needed for a rapid acoustic-based point of care testing (POCT) tool.

Wed-Keynote: Keynote 2: Pascale Fung

Room A+B, 15:00–16:00, Wednesday 1 September 2021 Chairs: TBD

Ethical and Technological Challenges of Conversational AI

Pascale Fung; HKUST, China

Wed-Keynote, Time: 15:00

Conversational AI (ConvAI) systems have applications ranging from personal assistance, health assistance to customer services. They have been in place since the first call centre agent went live in the late 1990s. More recently, smart speakers and smartphones are powered with conversational AI with similar architecture as those from the 90s. On the other hand, research on ConvAI systems has made leaps and bounds in recent years with sequence-to-sequence, generation-Thanks to the advent of large scale pre-trained based models. language models, state-of-the-art ConvAI systems can generate surprisingly human-like responses to user queries in open domain conversations, known as chit-chat. However, these generation based ConvAI systems are difficult to control and can lead to inappropriate, biased and sometimes even toxic responses. In addition, unlike previous modular conversational AI systems, it is also challenging to incorporate external knowledge into these models for task-oriented

dialog scenarios such as personal assistance and customer services, and to maintain consistency.

With great power comes great responsibility. We must address the many ethical and technical challenges of generation based conversational AI systems to control for bias and safety, consistency, style, knowledge incorporation, etc. In this talk, I will introduce state-of-the-art generation based conversational AI approaches, and will point out remaining challenges of conversational AI and possible directions for future research, including how to mitigate inappropriate responses. I will also present some ethical guidelines that conversational AI systems can follow.

Wed-A-O-1: Language Modeling and Text-Based Innovations for ASR

Room A+B, 16:00–18:00, Wednesday 1 September 2021 Chairs: Hermann Ney and Irina Illina

BERT-Based Semantic Model for Rescoring N-Best Speech Recognition List

Dominique Fohr, Irina Illina; Loria (UMR 7503), France Wed-A-O-1-1, Time: 16:00

This work aims to improve automatic speech recognition (ASR) by modeling long-term semantic relations. We propose to perform this through rescoring the ASR N-best hypotheses list. To achieve this, we propose two deep neural network (DNN) models and combine semantic, acoustic, and linguistic information. Our DNN rescoring models are aimed at selecting hypotheses that have better semantic consistency and therefore lower WER. We investigate a powerful representation as part of input features to our DNN model: dynamic contextual embeddings from Transformer-based BERT. Acoustic and linguistic features are also included. We perform experiments on the publicly available dataset TED-LIUM. We evaluate in clean and in noisy conditions, with n-gram and Recurrent Neural Network Language Model (RNNLM), more precisely Long Short-Term Memory (LSTM) model. The proposed rescoring approaches give significant WER improvements over the ASR system without rescoring models. Furthermore, the combination of rescoring methods based on BERT and GPT-2 scores achieves the best results.

Text Augmentation for Language Models in High Error Recognition Scenario

Karel Beneš, Lukáš Burget; Brno University of Technology, Czechia Wed-A-O-1-3, Time: 16:40

In this paper, we explore several data augmentation strategies for training of language models for speech recognition. We compare augmentation based on global error statistics with one based on unigram statistics of ASR errors and with label-smoothing and its sampled variant. Additionally, we investigate the stability and the predictive power of perplexity estimated on augmented data. Despite being trivial, augmentation driven by global substitution, deletion and insertion rates achieves the best rescoring results. On the other hand, even though the associated perplexity measure is stable, it gives no better prediction of the final error rate than the vanilla one. Our best augmentation scheme increases the WER improvement from second-pass rescoring from 1.1% to 1.9% absolute on the CHiMe-6 challenge.

On Sampling-Based Training Criteria for Neural Language Modeling

Yingbo Gao, David Thulke, Alexander Gerstenberger, Khoa Viet Tran, Ralf Schlüter, Hermann Ney; RWTH Aachen University, Germany Wed-A-O-1-4, Time: 17:00

As the vocabulary size of modern word-based language models becomes ever larger, many sampling-based training criteria are proposed and investigated. The essence of these sampling methods is that the softmax-related traversal over the entire vocabulary can be simplified, giving speedups compared to the baseline. A problem we notice about the current landscape of such sampling methods is the lack of a systematic comparison and some myths about preferring one over another. In this work, we consider Monte Carlo sampling, importance sampling, a novel method we call compensated partial summation, and noise contrastive estimation. Linking back to the three traditional criteria, namely mean squared error, binary cross-entropy, and cross-entropy, we derive the theoretical solutions to the training problems. Contrary to some common belief, we show that all these sampling methods can perform equally well, as long as we correct for the intended class posterior probabilities. Experimental results in language modeling and automatic speech recognition on Switchboard and LibriSpeech support our claim, with all sampling-based methods showing similar perplexities and word error rates while giving the expected speedups.

Fast Text-Only Domain Adaptation of RNN-Transducer Prediction Network

Janne Pylkkönen, Antti Ukkonen, Juho Kilpikoski, Samu Tamminen, Hannes Heikinheimo; Speechly, Finland Wed-A-O-1-5, Time: 17:20

Adaption of end-to-end speech recognition systems to new tasks is known to be challenging. A number of solutions have been proposed which apply external language models with various fusion methods, possibly with a combination of two-pass decoding. Also TTS systems have been used to generate adaptation data for the end-to-end models. In this paper we show that RNN-transducer models can be effectively adapted to new domains using only small amounts of textual data. By taking advantage of model's inherent structure, where the prediction network is interpreted as a language model, we can apply fast adaptation to the model. Adapting the model avoids the need for complicated decoding time fusions and external language models. Using appropriate regularization, the prediction network can be adapted to new domains while still retaining good generalization capabilities. We show with multiple ASR evaluation tasks how this method can provide relative gains of 10-45% in target task WER. We also share insights how RNN-transducer prediction network performs as a language model.

Wed-A-O-2: Speaker, Language, and Privacy

Room C, 16:00–18:00, Wednesday 1 September 2021 Chairs: Omid Ghahabi and Ondřej Glembek

Using Games to Augment Corpora for Language Recognition and Confusability

Christopher Cieri, James Fiumara, Jonathan Wright; University of Pennsylvania, USA Wed-A-O-2-2, Time: 16:20

We present a Game with a Purpose to elicit judgements of the language spoken in short audio clips of broadcast and conversational telephone speech, the resulting corpus and their potential use in research on language recognition and confusability.

Fair Voice Biometrics: Impact of Demographic Imbalance on Group Fairness in Speaker Recognition

*Gianni Fenu*¹, *Mirko Marras*², *Giacomo Medda*¹, *Giacomo Meloni*¹; ¹Università di Cagliari, Italy; ²EPFL, Switzerland

Wed-A-O-2-3, Time: 16:40

Speaker recognition systems are playing a key role in modern online applications. Though the susceptibility of these systems to discrimination according to group fairness metrics has been recently studied, their assessment has been mainly focused on the difference in equal error rate across groups, not accounting for other fairness criteria important in anti-discrimination policies, defined for demographic groups characterized by sensitive attributes. In this paper, we therefore study how existing group fairness metrics relate with the balancing settings of the training data set in speaker recognition. We conduct this analysis by operationalizing several definitions of fairness and monitoring them under varied data balancing settings. Experiments performed on three deep neural architectures, evaluated on a data set including gender/age-based groups, show that balancing group representation positively impacts on fairness and that the friction across security, usability, and fairness depends on the fairness metric and the recognition threshold.

Knowledge Distillation from Multi-Modality to Single-Modality for Person Verification

Leying Zhang, Zhengyang Chen, Yanmin Qian; SJTU, China

Wed-A-O-2-4, Time: 17:00

Voice and face are two important biometric characteristics that can be used for person identity verification. Previous works have proved the strong complementarity between audio and visual modalities in person verification tasks that multi-modality system can achieve significant performance improvement compared to single-modality system. However, due to the limitations in the real world, it is hard to access both audio and visual data at the same time. In this paper, we investigate several strategies to distill the knowledge from a multi-modality system and transfer it to the single-modality system in a teacher-student mode. We applied the knowledge distillation at three different levels: label level, embedding level, and distribution level. All the experiments are based on the VoxCeleb dataset. The results show that the visual single-modality system achieves 10% EER (equal error rate) improvement on the VoxCeleb1 evaluation set using our proposed knowledge distillation method. Besides, the improvement on the audio system is only reflected on part of the evaluation trials, and we give a detailed analysis for this phenomenon.

Adversarial Disentanglement of Speaker Representation for Attribute-Driven Privacy Preservation

Paul-Gauthier Noé¹, Mohammad Mohammadamini¹, Driss Matrouf¹, Titouan Parcollet¹, Andreas Nautsch², Jean-François Bonastre¹; ¹LIA (EA 4128), France; ²EURECOM, France Wed-A-Q-2-6, Time: 17:40

In speech technologies, speaker's voice representation is used in many applications such as speech recognition, voice conversion, speech synthesis and, obviously, user authentication. Modern vocal representations of the speaker are based on neural embeddings. In addition to the targeted information, these representations usually contain sensitive information about the speaker, like the age, sex, physical state, education level or ethnicity. In order to allow

NOTES

the user to choose which information to protect, we introduce in this paper the concept of *attribute-driven privacy preservation* in speaker voice representation. It allows a person to hide one or more personal aspects to a potential malicious interceptor and to the application provider. As a first solution to this concept, we propose to use an adversarial autoencoding method that disentangles in the voice representation a given speaker attribute thus allowing its concealment. We focus here on the sex attribute for an Automatic Speaker Verification (ASV) task. Experiments carried out using the VoxCeleb datasets have shown that the proposed method enables the concealment of this attribute while preserving ASV ability.

Wed-A-O-3: Assessment of Pathological Speech and Language I

Room Lacina, 16:00–18:00, Wednesday 1 September 2021 Chairs: Khalid Daoudi and RaviShankar Prasad

Automatically Detecting Errors and Disfluencies in Read Speech to Predict Cognitive Impairment in People with Parkinson's Disease

Amrit Romana¹, John Bandon¹, Matthew Perez¹, Stephanie Gutierrez², Richard Richter², Angela Roberts², Emily Mower Provost¹; ¹University of Michigan, USA; ²Northwestern University, USA Wed-A-O-3-1, Time: 16:00

Parkinson's disease (PD) is a central nervous system disorder that causes motor impairment. Recent studies have found that people with PD also often suffer from cognitive impairment (CI). While a large body of work has shown that speech can be used to predict motor symptom severity in people with PD, much less has focused on cognitive symptom severity. Existing work has investigated if acoustic features, derived from speech, can be used to detect CI in people with PD. However, these acoustic features are general and are not targeted toward capturing CI. Speech errors and disfluencies provide additional insight into CI. In this study, we focus on read speech, which offers a controlled template from which we can detect errors and disfluencies, and we analyze how errors and disfluencies vary with CI. The novelty of this work is an automated pipeline, including transcription and error and disfluency detection, capable of predicting CI in people with PD. This will enable efficient analyses of how cognition modulates speech for people with PD, leading to scalable speech assessments of CI.

Automatic Extraction of Speech Rhythm Descriptors for Speech Intelligibility Assessment in the Context of Head and Neck Cancers

*Robin Vaysse*¹, *Jérôme Farinas*¹, *Corine Astésano*², *Régine André-Obrecht*¹; ¹*IRIT (UMR 5505), France;* ²*URI Octogone-Lordat (EA 4156), France* Wed-A-O-3-2, Time: 16:20

The temporal dimension of speech acoustics is rarely taken into account in automatic models for Speech Intelligibility evaluation, although the rhythmic recurrence of phonemes, syllables and prosodic groups are allegedly good predictors of speech intelligibility. The present study aims at unravelling those automatic parameters that best account for the different levels of the speech signal's rhythmic structure, and to evaluate their correlation with a perceptual intelligibility measure. The parameters are extracted from the Fourier Transform of the amplitude modulation of the signal (Envelope Modulation Spectrum) [1, 2]. A Lasso linear model for feature selection is first implemented to select the most relevant

BRNO | CZECHIA

parameters, and a SVR regression analysis is run to reveal the best parameters' combination. Our analyses of EMS, using data from the French corpora of cancer speech C2SI [3], show strong performances of the automatic prediction, with a correlation of 0.70 between our model and an intelligibility evaluation score by speech-pathologists. In particular, the highest correlation with speech intelligibility lies in the ratio between the energy in the low frequency band (0.5-4 Hz that represents slow rhythmic modulations indicative of prosodic groups) and in the higher one (4-10 Hz that represents fast rhythmic modulations like phonemes).

Speech Disorder Classification Using Extended Factorized Hierarchical Variational Auto-Encoders

Jinzi Qi, Hugo Van hamme; KU Leuven, Belgium Wed-A-O-3-3, Time: 16:40

Objective speech disorder classification for speakers with communication difficulty is desirable for diagnosis and administering therapy. With the current state of speech technology, it is evident to propose neural networks for this application. But neural network model training is hampered by a lack of labeled disordered speech data. In this research, we apply an extended version of Factorized Hierarchical Variational Auto-encoders (FHVAE) for representation learning on disordered speech. The FHVAE model extracts both content-related and sequence-related latent variables from speech data, and we utilize the extracted variables to explore how disorder type information is represented in the latent variables. For better classification performance, the latent variables are aggregated at the word and sentence level. We show that an extension of the FHVAE model succeeds in the better disentanglement of the content-related and sequence-related related representations, but both representations are still required for best results on disorder type classification.

The Impact of Forced-Alignment Errors on Automatic Pronunciation Evaluation

*Vikram C. Mathad*¹, *Tristan J. Mahr*², *Nancy Scherer*¹, *Kathy Chapman*³, *Katherine C. Hustad*², *Julie Liss*¹, *Visar Berisha*¹; ¹*Arizona State University, USA*; ²*UW–Madison, USA*; ³*University of Utah, USA* Wed-A-O-3-4, Time: 17:00

Automatic evaluation of phone-level pronunciation scores typically involves two stages: (1) automatic phonetic segmentation via textconstrained phoneme alignment and (2) quantification of acoustic deviation for each phoneme-level relative to a database of correctlypronounced speech. It's clear that the second stage depends on the first. That is, if there is misalignment, the acoustic deviation will also be impacted. In this paper, we analyzed the impact of alignment error on a measure of goodness of pronunciation. We computed (1) automatic pronunciation scores using force-aligned samples, (2) the forced-alignment error rate, and (3) acoustic deviation using manually-aligned samples. We used a bivariate linear regression model to characterize the contributions of forced alignment errors and acoustic deviation on the automatic pronunciation scores. This was done across two different children speech databases, namely children with cleft lip/palate and typically developing children between the ages of 3-6 years. The analysis shows that, for speech from typically-developing children, most of the variation in the automatic pronunciation scores is explained by acoustic deviation, with the errors in forced alignment playing a relatively minor role. The forced alignment errors have a small but significant downstream impact on pronunciation assessment for children with cleft lip/palate.

Late Fusion of the Available Lexicon and Raw Waveform-Based Acoustic Modeling for Depression and Dementia Recognition

Esaú Villatoro-Tello¹, S. Pavankumar Dubagunta², Julian Fritsch², Gabriela Ramírez-de-la-Rosa¹, Petr Motlicek², Mathew Magimai-Doss²; ¹UAM, Mexico; ²Idiap Research Institute, Switzerland Wed-A-O-3-5, Time: 17:20

Mental disorders, e.g. depression and dementia, are categorized as priority conditions according to the World Health Organization (WHO). When diagnosing, psychologists employ structured questionnaires/interviews, and different cognitive tests. Although accurate, there is an increasing necessity of developing digital mental health support technologies to alleviate the burden faced by professionals. In this paper, we propose a multi-modal approach for modeling the communication process employed by patients being part of a clinical interview or a cognitive test. The language-based modality, inspired by the Lexical Availability (LA) theory from psycho-linguistics, identifies the most accessible vocabulary of the interviewed subject and use it as features in a classification process. The acoustic-based modality is processed by a Convolutional Neural Network (CNN) trained on signals of speech that predominantly contained voice source characteristics. In the end, a late fusion technique, based on majority voting, assigns the final classification. Results show the complementarity of both modalities, reaching an overall Macro-F1 of 84% and 90% for Depression and Alzheimer's dementia respectively.

Neural Speaker Embeddings for Ultrasound-Based Silent Speech Interfaces

Amin Honarmandi Shandiz¹, László Tóth¹, Gábor Gosztolya², Alexandra Markó³, Tamás Gábor Csapó⁴; ¹University of Szeged, Hungary; ²MTA-SZTE RGAI, Hungary; ³ELTE, Hungary; ⁴MTA-ELTE LingArt, Hungary Wed-A-O-3-6, Time: 17:40

Articulatory-to-acoustic mapping seeks to reconstruct speech from a recording of the articulatory movements, for example, an ultrasound video. Just like speech signals, these recordings represent not only the linguistic content, but are also highly specific to the actual speaker. Hence, due to the lack of multi-speaker data sets, researchers have so far concentrated on speaker-dependent modeling. Here, we present multi-speaker experiments using the recently published TaL80 corpus. To model speaker characteristics, we adjusted the x-vector framework popular in speech processing to operate with ultrasound tongue videos. Next, we performed speaker recognition experiments using 50 speakers from the corpus. Then, we created speaker embedding vectors and evaluated them on the remaining speakers. Finally, we examined how the embedding vector influences the accuracy of our ultrasound-to-speech conversion network in a multi-speaker scenario. In the experiments we attained speaker recognition error rates below 3%, and we also found that the embedding vectors generalize nicely to unseen speakers. Our first attempt to apply them in a multi-speaker silent speech framework brought about a marginal reduction in the error rate of the spectral estimation step.

Wed-A-V-1: Communication and Interaction, Multimodality

16:00–18:00, Wednesday 1 September 2021 Chairs: Catharine Oertel and Saturnino Luz

Cross-Modal Learning for Audio-Visual Video Parsing

Jatin Lamba, Abhishek, Jayaprakash Akula, Rishabh Dabral, Preethi Jyothi, Ganesh Ramakrishnan; IIT Bombay, India Wed-A-V-1-1, Time: 16:00

In this paper, we present a novel approach to the audio-visual video parsing (AVVP) task that demarcates events from a video separately for audio and visual modalities. The proposed parsing approach simultaneously detects the temporal boundaries in terms of start and end times of such events. We show how AVVP can benefit from the following techniques geared towards effective cross-modal learning: (i) adversarial training and skip connections (ii) global context aware attention and, (iii) self-supervised pretraining using an audio-video grounding objective to obtain cross-modal audio-video representations. We present extensive experimental evaluations on the Look, Listen, and Parse (LLP) dataset and show that we outperform the state-of-the-art Hybrid Attention Network (HAN) on all five metrics proposed for AVVP. We also present several ablations to validate the effect of pretraining, global attention and adversarial training.

A Psychology-Driven Computational Analysis of Political Interviews

Darren Cook¹, Miri Zilka², Simon Maskell¹, Laurence Alison¹; ¹University of Liverpool, UK; ²University of Sussex, UK

Wed-A-V-1-2, Time: 16:00

Can an interviewer influence the cooperativeness of an interviewee? The role of an interviewer in actualising a successful interview is an active field of social psychological research. A large-scale analysis of interviews, however, typically involves time-exorbitant manual tasks and considerable human effort. Despite recent advances in computational fields, many automated methods continue to rely on manually labelled training data to establish ground-truth. This reliance obscures explainability and hinders the mobility of analysis between applications. In this work, we introduce a crossdisciplinary approach to analysing interviewer efficacy. We suggest computational success measures as a transparent, automated, and reproducible alternative for pre-labelled data. We validate these measures with a small-scale study with human-responders. To study the interviewer's influence on the interviewee we utilise features informed by social psychological theory to predict interview quality based on the interviewer's linguistic behaviour. Our psychologically informed model significantly outperforms a bag-of-words model, demonstrating the strength of a cross-disciplinary approach toward the analysis of conversational data at scale.

Speech Emotion Recognition Based on Attention Weight Correction Using Word-Level Confidence Measure

Jennifer Santoso¹, Takeshi Yamada¹, Shoji Makino¹, Kenkichi Ishizuka², Takekatsu Hiramura²; ¹University of Tsukuba, Japan; ²Revcomm, Japan Wed-A-V-1-3, Time: 16:00

Emotion recognition is essential for human behavior analysis and possible through various inputs such as speech and images. However,

in practical situations, such as in call center analysis, the available information is limited to speech. This leads to the study of speech emotion recognition (SER). Considering the complexity of emotions, SER is a challenging task. Recently, automatic speech recognition (ASR) has played a role in obtaining text information from speech. The combination of speech and ASR results has improved the SER performance. However, ASR results are highly affected by speech recognition errors. Although there is a method to improve ASR performance on emotional speech, it requires the fine-tuning of ASR, which is costly. To mitigate the errors in SER using ASR systems, we propose the use of the combination of a self-attention mechanism and a word-level confidence measure (CM), which indicates the reliability of ASR results, to reduce the importance of words with a high chance of error. Experimental results confirmed that the combination of self-attention mechanism and CM reduced the effects of incorrectly recognized words in ASR results, providing a better focus on words that determine emotion recognition. Our proposed method outperformed the state-of-the-art methods on the IEMOCAP dataset.

Effects of Voice Type and Task on L2 Learners' Awareness of Pronunciation Errors

Alif Silpachai¹, Ivana Rehman¹, Taylor Anne Barriuso¹, John Levis¹, Evgeny Chukharev-Hudilainen¹, Guanlong Zhao², Ricardo Gutierrez-Osuna²; ¹Iowa State University, USA; ²Texas A&M University, USA Wed-A-V-1-4, Time: 16:00

Research suggests learners may improve their second language (L2) pronunciation by imitating voices with similar acoustic profiles. However, previously reported improvements have been in suprasegmentals (prosodic features such as intonation). It remains unclear if voice similarity applies to L2 segmentals (consonants and vowels). To address this issue, this study investigates how voice similarity facilitates awareness of pronunciation errors, a necessary step in pronunciation improvement. In two experiments, advanced L2 learners identified their pronunciation errors by comparing their production to the production of a resynthesized model voice using learners' voices as the base (Golden Speaker voice), or to an unfamiliar resynthesized voice with the same gender as the learner (Silver Speaker voice). In Experiment 1, L2 learners identified all syllables with vowel and consonant errors when comparing their production to the model voice. Their choices were compared to identifications by expert judges. In Experiment 2, learners were told how many errors the expert judges had identified before identifying the same number of errors. Results did not support facilitative effects of Golden Speaker voices in either experiment, but Experiment 2 resulted in higher identification percentages. Discussion of the challenges in self-identification of errors in relation to voice similarity are offered.

Lexical Entrainment and Intra-Speaker Variability in Cooperative Dialogues

Alla Menshikova, Daniil Kocharov, Tatiana Kachkovskaia; Saint Petersburg State University, Russia Wed-A-V-1-5, Time: 16:00

In dialogues, intra-speaker variability is often explained by the relationship between interlocutors. A person may speak differently with a friend and a stranger or depending on the interlocutor's gender or age — in all these cases we expect speech entrainment, but the degree of entrainment may vary. In this research, we measured lexical entrainment in a series of dialogues, where each one of 20 "core" speakers talked to five different interlocutors: a sibling, a close friend, an unfamiliar person of the same gender and similar age, and an unfamiliar person of the same gender, greater age and higher job

position. We hypothesized that the degree of speech entrainment systematically varies according to the type of interlocutor, across all the "core" speakers. The following measures of entrainment were used: parts of speech statistics, verb forms statistics, language style matching, and lexical density. Our data have shown that a person speaks very similarly to his/her sibling; dialogues with a friend or a same-gender stranger of similar age show fewer similarities; the least "common language" is observed in dialogues with a stranger of the opposite gender and with a stranger of greater age and higher job position.

Detecting Alzheimer's Disease Using Interactional and Acoustic Features from Spontaneous Speech

Shamila Nasreen, Julian Hough, Matthew Purver; Queen Mary University of London, UK

Wed-A-V-1-6, Time: 16:00

Alzheimer's Disease (AD) is a form of Dementia that manifests in cognitive decline including memory, language, and changes in behavior. Speech data has proven valuable for inferring cognitive status, used in many health assessment tasks, and can be easily elicited in natural settings. Much work focuses on analysis using linguistic features; here, we focus on non-linguistic features and their use in distinguishing AD patients from similar-age Non-AD patients with other health conditions in the Carolinas Conversation Collection (CCC) dataset. We used two types of features: patterns of interaction including pausing behaviour and floor control, and acoustic features including pitch, amplitude, energy, and cepstral coefficients. Fusion of the two kinds of features, combined with feature selection, obtains very promising classification results: classification accuracy of 90% using standard models such as support vector machines and logistic regression. We also obtain promising results using interactional features alone (87% accuracy), which can be easily extracted from natural conversations in daily life and thus have the potential for future implementation as a non-invasive method for AD diagnosis and monitoring.

Investigating the Interplay Between Affective, Phonatory and Motoric Subsystems in Autism Spectrum Disorder Using a Multimodal Dialogue Agent

Hardik Kothare¹, Vikram Ramanarayanan¹, Oliver Roesler¹, Michael Neumann¹, Jackson Liscombe¹, William Burke¹, Andrew Cornish¹, Doug Habberstad¹, Alaa Sakallah², Sara Markuson², Seemran Kansara², Afik Faerman², Yasmine Bensidi-Slimane², Laura Fry², Saige Portera², David Suendermann-Oeft¹, David Pautler¹, Carly Demopoulos²; ¹Modality.AI, USA; ²University of California at San Francisco, USA Wed-A-V-1-7, Time: 16:00

We explore the utility of an on-demand multimodal conversational platform in extracting speech and facial metrics in children with Autism Spectrum Disorder (ASD). We investigate the extent to which these metrics correlate with objective clinical measures, particularly as they pertain to the interplay between the affective, phonatory and motoric subsystems. 22 participants diagnosed with ASD engaged with a virtual agent in conversational affect production tasks designed to elicit facial and vocal affect. We found significant correlations between vocal pitch and loudness extracted by our platform during these tasks and accuracy in recognition of facial and vocal affect, assessed via the Diagnostic Analysis of Nonverbal Accuracy-2 (DANVA-2) neuropsychological task. We also found significant correlations between jaw kinematic metrics extracted

NOTES

using our platform and motor speed of the dominant hand assessed via a standardised neuropsychological finger tapping task. These findings offer preliminary evidence for the usefulness of these audiovisual analytic metrics and could help us better model the interplay between different physiological subsystems in individuals with ASD.

Analysis of Eye Gaze Reasons and Gaze Aversions During Three-Party Conversations

Carlos Toshinori Ishi, Taiken Shintani; RIKEN, Japan Wed-A-V-1-8, Time: 16:00

The background of this study is the generation of natural gaze behaviors in human-robot multimodal interaction. For that purpose, in this study we analyzed gaze behaviors of multiple speakers in a dataset containing three-party conversations, in terms of the reasons/intentions of their gaze events.

Analyses of the gaze reasons were conducted separately for the gaze behaviors towards a dialogue partner, and for gaze aversions (i.e., gazing away from a person's face). Analysis on the eyeball movements during gaze aversions was also conducted. Different distributions for average durations and gaze direction patterns were observed depending on the gaze reasons (e.g., in listening mode, speaking mode, towards dialogue partner's reactions, in gaze aversions during thinking and remembering, and during the speaker's own behaviors like nodding and laughing).

Wed-A-V-2: Language and Lexical Modeling for ASR

16:00–18:00, Wednesday 1 September 2021 Chairs: Preethi Jyothi and Richard Rose

Semantic Distance: A New Metric for ASR Performance Analysis Towards Spoken Language Understanding

Suyoun Kim, Abhinav Arora, Duc Le, Ching-Feng Yeh, Christian Fuegen, Ozlem Kalinli, Michael L. Seltzer; Facebook, USA

Wed-A-V-2-1, Time: 16:00

Word Error Rate (WER) has been the predominant metric used to evaluate the performance of automatic speech recognition (ASR) systems. However, WER is sometimes not a good indicator for downstream Natural Language Understanding (NLU) tasks, such as intent recognition, slot filling, and semantic parsing in task-oriented dialog systems. This is because WER takes into consideration only literal correctness instead of semantic correctness, the latter of which is typically more important for these downstream tasks. In this study, we propose a novel Semantic Distance (SemDist) measure as an alternative evaluation metric for ASR systems to address this issue. We define SemDist as the distance between a reference and hypothesis pair in a sentence-level embedding space. To represent the reference and hypothesis as a sentence embedding, we exploit RoBERTa, a state-of-the-art pre-trained deep contextualized language model based on the transformer architecture. We demonstrate the effectiveness of our proposed metric on various downstream tasks, including intent recognition, semantic parsing, and named entity recognition.

A Light-Weight Contextual Spelling Correction Model for Customizing Transducer-Based Speech Recognition Systems

*Xiaoqiang Wang*¹, *Yanqing Liu*¹, *Sheng Zhao*¹, *Jinyu Li*²; ¹*Microsoft, China*; ²*Microsoft, USA* Wed-A-V-2-2, Time: 16:00

It's challenging to customize transducer-based automatic speech recognition (ASR) system with context information which is dynamic and unavailable during model training. In this work, we introduce a light-weight contextual spelling correction model to correct contextrelated recognition errors in transducer-based ASR systems. We incorporate the context information into the spelling correction model with a shared context encoder and use a filtering algorithm to handle large-size context lists. Experiments show that the model improves baseline ASR model performance with about 50% relative word error rate reduction, which also significantly outperforms the baseline method such as contextual LM biasing. The model also shows excellent performance for out-of-vocabulary terms not seen during training.

Incorporating External POS Tagger for Punctuation Restoration

*Ning Shi*¹, *Wei Wang*¹, *Boxin Wang*², *Jinfeng Li*¹, *Xiangyu Liu*¹, *Zhouhan Lin*³; ¹*Alibaba, China;* ²*University of Illinois at Urbana-Champaign, USA;* ³*SJTU, China*

Wed-A-V-2-3, Time: 16:00

Punctuation restoration is an important post-processing step in automatic speech recognition. Among other kinds of external information, part-of-speech (POS) taggers provide informative tags, suggesting each input token's syntactic role, which has been shown to be beneficial for the punctuation restoration task. In this work, we incorporate an external POS tagger and fuse its predicted labels into the existing language model to provide syntactic information. Besides, we propose sequence boundary sampling (SBS) to learn punctuation positions more efficiently as a sequence tagging task. Experimental results show that our methods can consistently obtain performance gains and achieve a new state-of-the-art on the common IWSLT benchmark. Further ablation studies illustrate that both large pre-trained language models and the external POS tagger take essential parts to improve the model's performance.

Phonetically Induced Subwords for End-to-End Speech Recognition

Vasileios Papadourakis, Markus Müller, Jing Liu, Athanasios Mouchtaris, Maurizio Omologo; Amazon, USA

Wed-A-V-2-4, Time: 16:00

End-to-end automatic speech recognition systems map a sequence of acoustic features to text. In modern systems, text is encoded to grapheme subwords which are generated by methods designed for text processing tasks and therefore don't model or take advantage of the statistics of the acoustic features. Here, we present a novel method for generating grapheme subwords that are derived from phoneme sequences, therefore capturing phonetical statistics. The phonetically induced subwords can be used for training and inference in any system that benefits from subwords, regardless of architecture and without the need of a pronunciation lexicon. We compare our method to other commonly used methods, which are based on text statistics or on text-phoneme correspondence and present experiments on CTC and RNN-T architectures, evaluating subword sets of different sizes. We find that our phonetically

NOTES

induced subwords can improve performance of RNN-T models with relative improvements of up to 15.21% compared to other subword methods.

Revisiting Parity of Human vs. Machine Conversational Speech Transcription

Courtney Mansfield, Sara Ng, Gina-Anne Levow, Richard A. Wright, Mari Ostendorf; University of Washington, USA Wed-A-V-2-5, Time: 16:00

A number of studies have compared human and machine transcription, showing that automatic speech recognition (ASR) is approaching human performance in some contexts. Most studies look at differences as measured by the standard speech recognition scoring criterion: word error rate (WER). This study looks at more fine-grained analysis of differences for conversational speech data where systems have reached human parity in terms of average WER, specifically insertions vs. deletions, word category, and word context characterized by linguistic surprisal. In contrast to ASR systems, humans are more likely to miss words than to misrecognize them, and they are much more likely to make errors in transcribing words associated primarily with conversational contexts (fillers, backchannels and discourse cue words). The differences are more pronounced for more informal contexts, i.e. conversations between family members. Although human transcribers may miss these words, conversational partners seem to use them in turntaking and processing disfluencies. Thus, ASR systems may need superhuman transcription performance for spoken language technology to achieve human-level conversation skills.

Lookup-Table Recurrent Language Models for Long Tail Speech Recognition

W. Ronny Huang, Tara N. Sainath, Cal Peyser, Shankar Kumar, David Rybach, Trevor Strohman; Google, USA Wed-A-V-2-6, Time: 16:00

We introduce Lookup-Table Language Models (LookupLM), a method for scaling up the size of RNN language models with only a constant increase in the floating point operations, by increasing the expressivity of the embedding table. In particular, we instantiate an (additional) embedding table which embeds the previous n-gram token sequence, rather than a single token. This allows the embedding table to be scaled up arbitrarily - with a commensurate increase in performance — without changing the token vocabulary. Since embeddings are sparsely retrieved from the table via a lookup; increasing the size of the table adds neither extra operations to each forward pass nor extra parameters that need to be stored on limited GPU/TPU memory. We explore scaling n-gram embedding tables up to nearly a billion parameters. When trained on a 3-billion sentence corpus, we find that LookupLM improves long tail log perplexity by 2.44 and long tail WER by 23.4% on a downstream speech recognition task over a standard RNN language model baseline, an improvement comparable to a scaling up the baseline by $6.2 \times$ the number of floating point operations.

Contextual Density Ratio for Language Model Biasing of Sequence to Sequence ASR Systems

Jesús Andrés-Ferrer¹, Dario Albesano², Puming Zhan³, Paul Vozila³; ¹Nuance Communications, Spain; ²Nuance Communications, Italy; ³Nuance Communications, USA Wed-A-V-2-7, Time: 16:00

End-2-end (E2E) models have become increasingly popular in some ASR tasks because of their performance and advantages. These E2E

models directly approximate the posterior distribution of tokens given the acoustic inputs. Consequently, the E2E systems implicitly define a language model (LM) over the output tokens, which makes the exploitation of independently trained language models less straightforward than in conventional ASR systems. This makes it difficult to dynamically adapt E2E ASR system to contextual profiles for better recognizing special words such as named entities. In this work, we propose a contextual density ratio approach for both training a contextual aware E2E model and adapting the language model to named entities. We apply the aforementioned technique to an E2E ASR system, which transcribes doctor and patient conversations, for better adapting the E2E system to the names in the conversations. Our proposed technique achieves a relative improvement of up to 46.5% on the names over an E2E baseline without degrading the overall recognition accuracy of the whole test set. Moreover, it also surpasses a contextual shallow fusion baseline by 22.1% relative.

Token-Level Supervised Contrastive Learning for Punctuation Restoration

*Qiushi Huang*¹, *Tom Ko*¹, *H. Lilian Tang*², *Xubo Liu*², *Bo Wu*³; ¹*SUSTech, China*; ²*University of Surrey, UK*; ³*MIT-IBM Watson AI Lab, USA* Wed-A-V-2-8, Time: 16:00

Punctuation is critical in understanding natural language text. Currently, most automatic speech recognition (ASR) systems do not generate punctuation, which affects the performance of downstream tasks, such as intent detection and slot filling. This gives rise to the need for punctuation restoration. Recent work in punctuation restoration heavily utilizes pre-trained language models without considering data imbalance when predicting punctuation classes. In this work, we address this problem by proposing a token-level supervised contrastive learning method that aims at maximizing the distance of representation of different punctuation marks in the embedding space. The result shows that training with token-level supervised contrastive learning obtains up to 3.2% absolute F_1 improvement on the test set.

BART Based Semantic Correction for Mandarin Automatic Speech Recognition System

Yun Zhao, Xuerui Yang, Jinchao Wang, Yongyu Gao, Chao Yan, Yuanfu Zhou; Cloudwalk Technology, China Wed-A-V-2-9, Time: 16:00

Although automatic speech recognition (ASR) systems achieved significantly improvements in recent years, spoken language recognition error occurs which can be easily spotted by human beings. Various language modeling techniques have been developed on post recognition tasks like semantic correction. In this paper, we propose a Transformer based semantic correction method with pretrained BART initialization, Experiments on 10000 hours Mandarin speech dataset show that character error rate (CER) can be effectively reduced by 21.7% relatively compared to our baseline ASR system. Expert evaluation demonstrates that actual improvement of our model surpasses what CER indicates.

Class-Based Neural Network Language Model for Second-Pass Rescoring in ASR

Lingfeng Dai, Qi Liu, Kai Yu; SJTU, China Wed-A-V-2-10, Time: 16:00

Language model rescoring, especially neural network language model (NNLM) rescoring, is widely used to achieve improved performance in a second-pass automatic speech recognition (ASR) system. The rescoring NNLM is usually trained separately from the ASR sys-

NOTES

tem. Typically, the two's training corpora are different, leading to the vocabulary mismatch problem, consequently degrading ASR performance. Previous research focuses more on the language domain mismatch problem, while the vocabulary mismatch problem, which may also cause significant performance degradation, has not been well studied. This paper proposes a novel class-based NNLM framework to address the vocabulary mismatch problem for language model rescoring. Here, OOV words (unknown words to the rescoring NNLM are called OOV words for short) are assigned to well-trained classes of NNLM and inherit the class probability. Experiments show that class-based NNLM rescoring can significantly reduce performance degradation due to vocabulary mismatch.

Improving Customization of Neural Transducers by Mitigating Acoustic Mismatch of Synthesized Audio

*Gakuto Kurata*¹, *George Saon*², *Brian Kingsbury*², *David Haws*², *Zoltán Tüske*²; ¹*IBM*, *Japan*; ²*IBM*, *USA* Wed-A-V-2-11, Time: 16:00

Customization of automatic speech recognition (ASR) models using text data from a target domain is essential to deploying ASR in various domains. End-to-end (E2E) modeling for ASR has made remarkable progress, but the advantage of E2E modeling, where all neural network parameters are jointly optimized, is offset by the challenge of customizing such models. In conventional hybrid models, it is easy to directly modify a language model or a lexicon using text data, but this is not true for E2E models. One popular approach for customizing E2E models uses audio synthesized from the target domain text, but the acoustic mismatch between the synthesized and real audio can be problematic. We propose a method that avoids the negative effect of synthesized audio by (1) adding a mapping network before the encoder network to map the acoustic features of the synthesized audio to those of the source domain, (2) training the added mapping network using text and synthesized audio from the source domain while freezing all layers in the E2E model, (3) training the E2E model with text and synthesized audio from the target domain, and (4) removing the added mapping network when decoding real audio from the target domain. Experiments on customizing RNN Transducer and Conformer Transducer models demonstrate the advantage of the proposed method over encoder freezing, a popular customization method for E2E models.

A Discriminative Entity-Aware Language Model for Virtual Assistants

Mandana Saebi¹, Ernest Pusateri², Aaksha Meghawat², Christophe Van Gysel²; ¹University of Notre Dame, USA; ²Apple, USA

Wed-A-V-2-12, Time: 16:00

High-quality automatic speech recognition (ASR) is essential for virtual assistants (VAs) to work well. However, ASR often performs poorly on VA requests containing named entities. In this work, we start from the observation that many ASR errors on named entities are inconsistent with real-world knowledge. We extend previous discriminative n-gram language modeling approaches to incorporate real-world knowledge from a Knowledge Graph (KG), using features that capture entity type-entity and entity-entity relationships. We apply our model through an efficient lattice rescoring process, achieving relative sentence error rate reductions of more than 25% on some synthesized test sets covering less popular entities, with minimal degradation on a uniformly sampled VA test set.
Correcting Automated and Manual Speech Transcription Errors Using Warped Language Models

Mahdi Namazifar, John Malik, Li Erran Li, Gokhan Tur, Dilek Hakkani Tür; Amazon, USA Wed-A-V-2-13, Time: 16:00

Masked language models have revolutionized natural language processing systems in the past few years. A recently introduced generalization of masked language models called warped language models are trained to be more robust to the types of errors that appear in automatic or manual transcriptions of spoken language by exposing the language model to the same types of errors during the training of language models. In this work we propose a novel approach that takes advantage of the robustness of warped language models to transcription noise for correcting transcriptions of spoken language. We show that our proposed approach is able to achieve up to 10% reduction in word error rates of both automatic and manual transcriptions of spoken language.

Wed-A-V-3: Novel Neural Network Architectures for ASR

16:00-18:00, Wednesday 1 September 2021 Chairs: Penny Karanasou and Michael Auli

Dynamic Encoder Transducer: A Flexible Solution for Trading Off Accuracy for Latency

Yangyang Shi, Varun Nagaraja, Chunyang Wu, Jay Mahadeokar, Duc Le, Rohit Prabhavalkar, Alex Xiao, Ching-Feng Yeh, Julian Chan, Christian Fuegen, Ozlem Kalinli, Michael L. Seltzer; Facebook, USA Wed-A-V-3-1, Time: 16:00

We propose a dynamic encoder transducer (DET) for on-device speech recognition. One DET model scales to multiple devices with different computation capacities without retraining or finetuning. To trading off accuracy and latency, DET assigns different encoders to decode different parts of an utterance. We apply and compare the layer dropout and the collaborative learning for DET training. The layer dropout method that randomly drops out encoder layers in the training phase, can do on-demand layer dropout in decoding. Collaborative learning jointly trains multiple encoders with different depths in one single model. Experiment results on Librispeech and in-house data show that DET provides a flexible accuracy and latency trade-off. Results on Librispeech show that the full-size encoder in DET relatively reduces the word error rate of the same size baseline by over 8%. The lightweight encoder in DET trained with collaborative learning reduces the model size by 25% but still gets similar WER as the full-size baseline. DET gets similar accuracy as a baseline model with better latency on a large in-house data set by assigning a lightweight encoder for the beginning part of one utterance and a full-size encoder for the rest.

Domain-Aware Self-Attention for Multi-Domain Neural Machine Translation

*Shiqi Zhang*¹, *Yan Liu*², *Deyi Xiong*², *Pei Zhang*¹, *Boxing Chen*¹; ¹*Alibaba, China*; ²*Tianjin University, China*

Wed-A-V-3-2, Time: 16:00

In this paper, we investigate multi-domain neural machine translation (NMT) that translates sentences of different domains in a single model. To this end, we propose a domain-aware self-attention mechanism that jointly learns domain representations with the single NMT

NOTES

model. The learned domain representations are integrated into both the encoder and decoder. We further propose two different domain representation learning approaches: 1) word-level unsupervised learning via a domain attention network and 2) guided learning with an auxiliary loss. The two learning approaches allow our multidomain NMT to work in different settings as to whether the domain information is available or not. Experiments on both Chinese-English and English-French demonstrate that our multi-domain model outperforms a strong baseline built on the Transformer and other previous multi-domain NMT approaches. Further analyses show that our model is able to learn domain clusters even without prior knowledge about the domain structure.

Librispeech Transducer Model with Internal Language Model Prior Correction

Albert Zeyer, André Merboldt, Wilfried Michel, Ralf Schlüter, Hermann Ney; RWTH Aachen University, Germany

Wed-A-V-3-3, Time: 16:00

We present our transducer model on Librispeech. We study variants to include an external language model (LM) with shallow fusion and subtract an estimated internal LM. This is justified by a Bayesian interpretation where the transducer model prior is given by the estimated internal LM. The subtraction of the internal LM gives us over 14% relative improvement over normal shallow fusion. Our transducer has a separate probability distribution for the non-blank labels which allows for easier combination with the external LM, and easier estimation of the internal LM. We additionally take care of including the end-of-sentence (EOS) probability of the external LM in the last blank probability which further improves the performance. All our code and setups are published.

A Deliberation-Based Joint Acoustic and Text Decoder

Sepand Mavandadi, Tara N. Sainath, Ke Hu, Zelin Wu; Google, USA

Wed-A-V-3-4, Time: 16:00

We propose a new two-pass E2E speech recognition model that improves ASR performance by training on a combination of paired data and unpaired text data. Previously, the joint acoustic and text decoder (JATD) has shown promising results through the use of text data during model training and the recently introduced deliberation architecture has reduced recognition errors by leveraging first-pass decoding results. Our method, dubbed Deliberation-JATD, combines the spelling correcting abilities of deliberation with JATD's use of unpaired text data to further improve performance. The proposed model produces substantial gains across multiple test sets, especially those focused on rare words, where it reduces word error rate (WER) by between 12% and 22.5% relative. This is done without increasing model size or requiring multi-stage training, making Deliberation-JATD an efficient candidate for on-device applications.

On the Limit of English Conversational Speech Recognition

Zoltán Tüske, George Saon, Brian Kingsbury; IBM, USA Wed-A-V-3-5, Time: 16:00

In our previous work we demonstrated that a single headed attention encoder-decoder model is able to reach state-of-the-art results in conversational speech recognition. In this paper, we further improve the results for both Switchboard 300 and 2000. Through use of an improved optimizer, speaker vector embeddings, and alternative speech representations we reduce the recognition errors of our LSTM system on Switchboard-300 by 4% relative. Compensation of the decoder model with the probability ratio approach allows more efficient integration of an external language model, and we report 5.9% and 11.5% WER on the SWB and CHM parts of Hub5'00 with very simple LSTM models. Our study also considers the recently proposed conformer, and more advanced self-attention based language models. Overall, the conformer shows similar performance to the LSTM; nevertheless, their combination and decoding with an improved LM reaches a new record on Switchboard-300, 5.0% and 10.0% WER on SWB and CHM. Our findings are also confirmed on Switchboard-2000, and a new state of the art is reported, practically reaching the limit of the benchmark.

Deformable TDNN with Adaptive Receptive Fields for Speech Recognition

Keyu An, Yi Zhang, Zhijian Ou; Tsinghua University, China

Wed-A-V-3-6, Time: 16:00

Time Delay Neural Networks (TDNNs) are widely used in both DNN-HMM based hybrid speech recognition systems and recent end-to-end systems. Nevertheless, the receptive fields of TDNNs are limited and fixed, which is not desirable for tasks like speech recognition, where the temporal dynamics of speech are varied and affected by many factors. In this paper, we propose to use deformable TDNNs for adaptive temporal dynamics modeling in end-to-end speech recognition. Inspired by deformable ConvNets, deformable TDNNs augment the temporal sampling locations with additional offsets and learn the offsets automatically based on the ASR criterion, without additional supervision. Experiments show that deformable TDNNs obtain state-of-the-art results on WSJ benchmarks (1.42%/3.45% WER on WSJ eval92/dev93 respectively), outperforming standard TDNNs significantly. Furthermore, we propose the latency control mechanism for deformable TDNNs, which enables deformable TDNNs to do streaming ASR without accuracy degradation.

Transformer-Based End-to-End Speech Recognition with Residual Gaussian-Based Self-Attention

Chengdong Liang, Menglong Xu, Xiao-Lei Zhang; Northwestern Polytechnical University, China Wed-A-V-3-7, Time: 16:00

Self-attention (SA), which encodes vector sequences according to their pairwise similarity, is widely used in speech recognition due to its strong context modeling ability. However, when applied to long sequence data, its accuracy is reduced. This is caused by the fact that its weighted average operator may lead to the dispersion of the attention distribution, which results in the relationship between adjacent signals ignored. To address this issue, in this paper, we introduce relative-position-awareness self-attention (RPSA). It not only maintains the global-range dependency modeling ability of selfattention, but also improves the localness modeling ability. Because the local window length of the original RPSA is fixed and sensitive to different test data, here we propose Gaussian-based self-attention (GSA) whose window length is learnable and adaptive to the test data automatically. We further generalize GSA to a new residual Gaussian self-attention (resGSA) for the performance improvement. We apply RPSA, GSA, and resGSA to Transformer-based speech recognition respectively. Experimental results on the AISHELL-1 Mandarin speech recognition corpus demonstrate the effectiveness of the proposed methods. For example, the resGSA-Transformer achieves a character error rate (CER) of 5.86% on the test set, which is relative 7.8% lower than that of the SA-Transformer. Although the performance of the proposed resGSA-Transformer is only slightly better than that of the RPSA-Transformer, it does not have to tune the window length manually.

SpeechMoE: Scaling to Large Acoustic Models with Dynamic Routing Mixture of Experts

Zhao You¹, *Shulin* Feng¹, *Dan* Su¹, *Dong* Yu²; ¹*Tencent, China*; ²*Tencent, USA* Wed-A-V-3-8, Time: 16:00

Recently, Mixture of Experts (MoE) based Transformer has shown promising results in many domains. This is largely due to the following advantages of this architecture: firstly, MoE based Transformer can increase model capacity without computational cost increasing both at training and inference time. Besides, MoE based Transformer is a dynamic network which can adapt to the varying complexity of input instances in real-world applications. In this work, we explore the MoE based model for speech recognition, named SpeechMoE. To further control the sparsity of router activation and improve the diversity of gate values, we propose a sparsity L1 loss and a mean importance loss respectively. In addition, a new router architecture is used in SpeechMoE which can simultaneously utilize the information from a shared embedding network and the hierarchical representation of different MoE layers. Experimental results show that SpeechMoE can achieve lower character error rate (CER) with comparable computation cost than traditional static networks, providing 7.0%~23.0% relative CER improvements on four evaluation datasets.

Online Compressive Transformer for End-to-End Speech Recognition

Chi-Hang Leong, Yu-Han Huang, Jen-Tzung Chien; NYCU, Taiwan

Wed-A-V-3-9, Time: 16:00

Traditionally, transformer with connectionist temporal classification (CTC) was developed for offline speech recognition where the transcription was generated after the whole utterance has been spoken. However, it is crucial to carry out online transcription of speech signal for many applications including live broadcasting and meeting. This paper presents an online transformer for real-time speech recognition where online transcription is generated chunk by chuck. In particular, an online compressive transformer (OCT) is proposed for end-to-end speech recognition. This OCT aims to generate immediate transcription for each audio chunk while the comparable performance with offline speech recognition can be still achieved. In the implementation, OCT tightly combines with both CTC and recurrent neural network transducer by minimizing their losses for training. In addition, this OCT systematically merges with compressive memory to reduce potential performance degradation due to online processing. This degradation is caused by online transcription which is generated by the chunks without history information. The experiments on speech recognition show that OCT does not only obtain comparable performance with offline transformer, but also work faster than the baseline model.

End to End Transformer-Based Contextual Speech Recognition Based on Pointer Network

Binghuai Lin, Liyuan Wang; Tencent, China Wed-A-V-3-10, Time: 16:00

Most spoken language assessment systems rely on the text features extracted from the automatic speech recognition (ASR) transcripts and thus depend heavily on the accuracy of the ASR systems. Automatic speech scoring tasks such as reading aloud and spontaneous speech are commonly provided with the prompts in advance to guide test takers' answers, which contain information that should be included in the answers (e.g., listening passage, and sample response). Utilizing these texts to improve ASR performance is of great importance for these tasks. In this paper, we develop an

end-to-end (E2E) ASR system incorporating contextual information provided by prompts. Specifically, we add an extra prompt encoder to a transformer-based E2E ASR system. To fuse the probabilities of the ASR output and the prompts dynamically, we train a soft gate based on the pointer network with carefully constructed prompt training corpus. We experiment the proposed method with data collected from English speaking proficiency tests recorded by Chinese teenagers from 16 to 18 years old. The results show the improved performance of speech recognition with a nearly 50% drop in word error rate (WER) utilizing prompts. Furthermore, the proposed network performs well in rare word recognition such as locations and personal names.

A Comparative Study on Neural Architectures and Training Methods for Japanese Speech Recognition

Shigeki Karita, Yotaro Kubo, Michiel Adriaan Unico Bacchiani, Llion Jones; Google, Japan Wed-A-V-3-11, Time: 16:00

End-to-end (E2E) modeling is advantageous for automatic speech recognition (ASR) especially for Japanese since word-based tokenization of Japanese is not trivial, and E2E modeling is able to model character sequences directly. This paper focuses on the latest E2E modeling techniques, and investigates their performances on character-based Japanese ASR by conducting comparative experiments. The results are analyzed and discussed in order to understand the relative advantages of long short-term memory (LSTM), and Conformer models in combination with connectionist temporal classification, transducer, and attention-based loss functions. Furthermore, the paper investigates on effectivity of the recent training techniques such as data augmentation (SpecAugment), variational noise injection, and exponential moving average. The best configuration found in the paper achieved the state-of-the-art character error rates of 4.1%, 3.2%, and 3.5% for Corpus of Spontaneous Japanese (CSJ) eval1, eval2, and eval3 tasks, respectively. The system is also shown to be computationally efficient thanks to the efficiency of Conformer transducers.

Advanced Long-Context End-to-End Speech Recognition Using Context-Expanded Transformers

Takaaki Hori, Niko Moritz, Chiori Hori, Jonathan Le Roux; MERL, USA

Wed-A-V-3-12, Time: 16:00

This paper addresses end-to-end automatic speech recognition (ASR) for long audio recordings such as lecture and conversational speeches. Most end-to-end ASR models are designed to recognize independent utterances, but contextual information (e.g., speaker or topic) over multiple utterances is known to be useful for ASR. In our prior work, we proposed a context-expanded Transformer that accepts multiple consecutive utterances at the same time and predicts an output sequence for the last utterance, achieving 5-15% relative error reduction from utterance-based baselines in lecture and conversational ASR benchmarks. Although the results have shown remarkable performance gain, there is still potential to further improve the model architecture and the decoding process. In this paper, we extend our prior work by (1) introducing the Conformer architecture to further improve the accuracy, (2) accelerating the decoding process with a novel activation recycling technique, and (3) enabling streaming decoding with triggered attention. We demonstrate that the extended Transformer provides state-of-the-art end-to-end ASR performance, obtaining a 17.3% character error rate for the HKUST dataset and 12.0%/6.3% word error rates for the Switchboard-300 Eval2000 CallHome/Switchboard test sets. The new decoding method reduces decoding time by more than 50% and further enables streaming ASR with limited accuracy degradation.

NOTES

Transformer-Based ASR Incorporating Time-Reduction Layer and Fine-Tuning with Self-Knowledge Distillation

Md. Akmal Haidar, Chao Xing, Mehdi Rezagholizadeh; Huawei Technologies, Canada Wed-A-V-3-13, Time: 16:00

Reducing the input sequence length of speech features to alleviate the complexity of alignment between speech features and text transcript by sub-sampling approaches is an important way to get better results in end-to-end (E2E) automatic speech recognition (ASR) systems. This issue is more important in Transformer-based ASR, because the self-attention mechanism in Transformers has $O(n^2)$ order of complexity in both training and inference. In this paper, we propose a Transformer-based ASR model with the time-reduction layer, in which we incorporate time-reduction layer inside transformer encoder layers in addition to traditional sub-sampling methods to input features that further reduce the frame-rate. This can help in reducing the computational cost of the self-attention process for training and inference with performance improvement. Moreover, we introduce a fine-tuning approach for pre-trained ASR models using self-knowledge distillation (S-KD) which further improves the performance of our ASR model. Experiments on LibriSpeech datasets show that our proposed methods outperform all other Transformer-based ASR systems. Furthermore, with language model (LM) fusion, we achieve new state-of-the-art word error rate (WER) results for Transformer-based ASR models with just 30 million parameters trained without any external data.

Flexi-Transducer: Optimizing Latency, Accuracy and Compute for Multi-Domain On-Device Scenarios

Jay Mahadeokar, Yangyang Shi, Yuan Shangguan, Chunyang Wu, Alex Xiao, Hang Su, Duc Le, Ozlem Kalinli, Christian Fuegen, Michael L. Seltzer; Facebook, USA

Wed-A-V-3-14, Time: 16:00

Often, the storage and computational constraints of embedded devices demand that a single on-device ASR model serve multiple use-cases / domains. In this paper, we propose a Flexible Transducer (FlexiT) for on-device automatic speech recognition to flexibly deal with multiple use-cases / domains with different accuracy and latency requirements. Specifically, using a single compact model, FlexiT provides a fast response for voice commands, and accurate transcription but with more latency for *dictation*. In order to achieve flexible and better accuracy and latency trade-offs, the following techniques are used. Firstly, we propose using domain-specific altering of segment size for Emformer encoder that enables FlexiT to achieve flexible decoding. Secondly, we use Alignment Restricted RNNT loss to achieve flexible fine-grained control on token emission latency for different domains. Finally, we add a domain indicator vector as an additional input to the FlexiT model. Using the combination of techniques, we show that a single model can be used to improve WERs and real time factor for dictation scenarios while maintaining optimal latency for voice commands use-cases.

Wed-A-V-4: Speech Localization, Enhancement, and Quality Assessment

16:00–18:00, Wednesday 1 September 2021 Chairs: Marc Delcroix and Yulan Liu

Difference in Perceived Speech Signal Quality Assessment Among Monolingual and Bilingual Teenage Students

Przemyslaw Falkowski-Gilski; Gdansk University of Technology, Poland Wed-A-V-4-1, Time: 16:00

The user perceived quality is a mixture of factors, including the background of an individual. The process of auditory perception is discussed in a wide variety of fields, ranging from engineering to medicine. Many studies examine the difference between musicians and non-musicians. Since musical training develops musical hearing and other various auditory capabilities, similar enhancements should be observable in case of bilingual people. This paper examines the difference in perceived speech signal quality between students from monolingual and bilingual classes. The subjective study was carried out on a group of 30 people, with 15 individuals in each class, aged 16–18 years old, considering three languages: English, German, and Polish. Results of this study may aid researchers as well as professionals active in the field of auditory perception, hearing loss related with ageing, and of course evaluation of networks and services.

PILOT: Introducing Transformers for Probabilistic Sound Event Localization

*Christopher Schymura*¹, *Benedikt Bönninghoff*¹, *Tsubasa Ochiai*², *Marc Delcroix*², *Keisuke Kinoshita*², *Tomohiro Nakatani*², *Shoko Araki*², *Dorothea Kolossa*¹; ¹*Ruhr-Universität Bochum, Germany*; ²*NTT, Japan* Wed-A-V-4-2, Time: 16:00

Sound event localization aims at estimating the positions of sound sources in the environment with respect to an acoustic receiver (e.g. a microphone array). Recent advances in this domain most prominently focused on utilizing deep recurrent neural networks. Inspired by the success of transformer architectures as a suitable alternative to classical recurrent neural networks, this paper introduces a novel transformer-based sound event localization framework, where temporal dependencies in the received multi-channel audio signals are captured via self-attention mechanisms. Additionally, the estimated sound event positions are represented as multivariate Gaussian variables, yielding an additional notion of uncertainty, which many previously proposed deep learning-based systems designed for this application do not provide. The framework is evaluated on three publicly available multi-source sound event localization datasets and compared against state-of-the-art methods in terms of localization error and event detection accuracy. It outperforms all competing systems on all datasets with statistical significant differences in performance.

Sound Source Localization with Majorization Minimization

Masahito Togami, Robin Scheibler; LINE, Japan Wed-A-V-4-3, Time: 16:00

We propose a sound source localization technique that estimates a speech source location without precise grid searching. The source location is estimated in a parameter optimization manner to mini-

NOTES

mize the steered-response power (SRP) function with the near-field assumption. Because there is no closed-form solution for the SRP function, we introduce an auxiliary function of the SRP function based on the majorization-minimization (MM) algorithm. Parameters are updated iteratively to minimize the auxiliary function with alternate execution of time-difference-of-arrival (TDOA) estimation and range-difference (RD) based localization. When TDOA estimation and RD-based localization are performed in a cascade manner, the estimation accuracy of the source location is strongly affected by the estimation accuracy of the TDOA. On contrary, the proposed method corrects the estimated TDOA by referring to the estimated source location in the previous iteration. Thus, it is expected for the proposed method to be robust against TDOA estimation error which occurs under reverberant environments. Experimental results show that the proposed method outperforms conventional techniques under a reverberant environment.

NISQA: A Deep CNN-Self-Attention Model for Multidimensional Speech Quality Prediction with Crowdsourced Datasets

Gabriel Mittag, Babak Naderi, Assmaa Chehadi, Sebastian Möller; Technische Universität Berlin,

Germany

Wed-A-V-4-4, Time: 16:00

In this paper, we present an update to the NISQA speech quality prediction model that is focused on distortions that occur in communication networks. In contrast to the previous version, the model is trained end-to-end and the time-dependency modelling and time-pooling is achieved through a Self-Attention mechanism. Besides overall speech quality, the model also predicts the four speech quality dimensions *Noisiness, Coloration, Discontinuity*, and *Loudness*, and in this way gives more insight into the cause of a quality degradation. Furthermore, new datasets with over 13,000 speech files were created for training and validation of the model. The model was finally tested on a new, live-talking test dataset that contains recordings of real telephone calls. Overall, NISQA was trained and evaluated on 81 datasets from different sources and showed to provide reliable predictions also for unknown speech samples. The code, model weights, and datasets are open-sourced.

Subjective Evaluation of Noise Suppression Algorithms in Crowdsourcing

Babak Naderi¹, Ross Cutler²; ¹Technische Universität Berlin, Germany; ²Microsoft, USA Wed-A-V-4-5, Time: 16:00

The quality of the speech communication systems, which include noise suppression algorithms, are typically evaluated in laboratory experiments according to the ITU-T Rec. P.835, in which participants rate background noise, speech signal, and overall quality separately. This paper introduces an open-source toolkit for conducting subjective quality evaluation of noise suppressed speech in crowdsourcing. We followed the ITU-T Rec. P.835, and P.808 and highly automate the process to prevent moderator's error. To assess the validity of our evaluation method, we compared the Mean Opinion Scores (MOS), calculated using ratings collected with our implementation and the MOS values from a standard laboratory experiment conducted according to the ITU-T Rec P.835. Results show a high validity in all three scales, namely background noise, speech signal and overall quality (average Pearson Correlation Coefficient (PCC) = 0.961). Results of a round-robin test (N=5) showed that our implementation is also a highly reproducible evaluation method (PCC=0.99). Finally, we used our implementation in the INTERSPEECH 2021 Deep Noise Suppression Challenge [1] as the primary evaluation metric, which demonstrates it is practical to use at scale. The results are analyzed to determine why the overall performance was the best in terms of background noise and speech quality.

Reliable Intensity Vector Selection for Multi-Source Direction-of-Arrival Estimation Using a Single Acoustic Vector Sensor

Jianhua Geng, Sifan Wang, Juan Li, JingWei Li, Xin Lou; ShanghaiTech University, China Wed-A-V-4-6, Time: 16:00

In the context of multi-source direction of arrival (DOA) estimation using a single acoustic vector sensor (AVS), the received signal is usually a mixture of noise, reverberation and source signals. The identification of the time-frequency (TF) bins that are dominated by the source signals can significantly improve the robustness of the DOA estimation. In this paper, a TF bin selection based DOA estimation pipeline is proposed. The proposed pipeline mainly involves three key steps: key frame identification, TF bin selection and DOA extraction. We identify the key frames by frame-wisely examining the effective rank. Subsequently, the geometric medians of the selected key frames are extracted to alleviate the impact of extreme outliers. The simulation results show that the accuracy and the robustness of the proposed pipeline outperform the state-of-the-art (SOTA) techniques.

MetricNet: Towards Improved Modeling For Non-Intrusive Speech Quality Assessment

Meng Yu, Chunlei Zhang, Yong Xu, Shi-Xiong Zhang, Dong Yu; Tencent, USA Wed-A-V-4-7, Time: 16:00

The objective speech quality assessment is usually conducted by comparing received speech signal with its clean reference, while human beings are capable of evaluating the speech quality without any reference, such as in the mean opinion score (MOS) tests. Nonintrusive speech quality assessment has attracted much attention recently due to the lack of access to clean reference signals for objective evaluations in real scenarios. In this paper, we propose a novel non-intrusive speech quality measurement model, MetricNet, which leverages label distribution learning and joint speech reconstruction learning to achieve significantly improved performance compared to the existing non-intrusive speech quality measurement models. We demonstrate that the proposed approach yields promisingly high correlation to the intrusive objective evaluation of speech quality on clean, noisy and processed speech data.

CNN-Based Processing of Acoustic and Radio Frequency Signals for Speaker Localization from MAVs

Andrea Toma, Daniele Salvati, Carlo Drioli, Gian Luca Foresti; Università di Udine, Italy Wed-A-V-4-8, Time: 16:00

A novel speaker localization algorithm from micro aerial vehicles (MAVs) is investigated. It introduces a joint direction of arrival (DOA) and distance prediction method based on processing and fusion of the multi-channel speech data with radio frequency (RF) measurements of the received signal strength. Possible applications include unmanned aerial vehicles (UAVs)-based reconnaissance and surveillance against intrusions and search and rescue in hostile environments. A 3-stages convolutional neural network (CNN) with a fusion layer is proposed to perform this task with the objective of augmenting the source localization from multi-channel speech signals. Two parallel CNNs process the speech and RF data, and the regression network produces predictions of the angle and distance

NOTES

from the source after the fusion layer. To show the performance and effectiveness of this RF-assisted method, the experimental scenario and datasets are presented and experiments are then discussed along with the results that have been obtained.

Assessment of von Mises-Bernoulli Deep Neural **Network in Sound Source Localization**

Katsutoshi Itoyama¹, Yoshiya Morimoto¹, Shungo Masaki¹, Ryosuke Kojima², Kenji Nishida¹, Kazuhiro Nakadai¹: ¹Tokyo Tech, Japan: ²Kyoto University, Japan Wed-A-V-4-9, Time: 16:00

This paper addresses the properties and effectiveness of the von Mises-Bernoulli deep neural network (vM-B DNN), a neural network capable of learning periodic information, in sound source localization. The phase, which is periodic information, is an important cue in sound source localization, but typical neural network cannot handle periodic input values properly. The vM-B DNN has been theoretically revealed to be able to handle periodic input values and its effectiveness has been shown in a simple case study of sound source localization using artificial sinusoids, but it was not in the case of speech signals. We conducted both numerical simulation and actual environment experiments. We compared a sound source localization method using vM-B DNN with those using ordinary neural networks, and showed that the vM-B DNN outperforms other methods under various conditions.

Feature Fusion by Attention Networks for Robust **DOA Estimation**

Rongliang Liu, Nengheng Zheng, Xi Chen; Shenzhen University, China

Wed-A-V-4-10, Time: 16:00

Direction of arrival (DOA) estimation is a key front-end technology for many speech-based intelligent systems. Deep neural networksbased DOA systems have recently demonstrated better performances than conventional ones. However, most of the existing networks use only one specific acoustical feature as input, limiting their noise-robustness. This paper proposes an attention-based feature fusion approach for DOA estimation. Two classical DOA estimation approaches, i.e., the least mean square-based adaptive filtering and the generalized cross-correlation, are adopted, and the respective features are served as input to the networks. Network with attention mechanism is built to learn the optimal weighting scheme, which can take advantage of the two features' complementary contributions in DOA estimation. Simulation and real test results show that the proposed method could use the complementary DOA information in different features and improve estimation accuracy under acoustic conditions with both noise and reverberation.

Far-Field Speaker Localization and Adaptive GLMB Tracking

Shoufeng Lin¹, Zhaojie Luo²; ¹Curtin University, Australia; ²Osaka University, Japan Wed-A-V-4-11, Time: 16:00

In the speech signal processing area, far-field speaker localization using only the audio modality has been a fundamental but challenging problem, especially in presence of reverberation and a varying number of moving speakers. Many existing methods use speech onsets as reliable directional cues against reverberation and interference. However, signal processing can be computationally costly especially in time domain. In this paper, we present a computationally efficient implementation of the recently proposed Onset-Multichannel Cross Correlation Coefficient (MCCC) method. Instead of scanning the entire spatial grid, reverse mapping and linear interpolation are used. The proposed algorithm with better efficiency is referred to as the Onset-MCC in this paper. Performance of the Onset-MCC is studied over various reverberant and noisy scenarios. To further suppress outliers and address miss-detections, as well as for the adaptive tracking of a varying number of moving speakers, we present an adaptive implementation of the generalized labeled multi-Bernoulli (GLMB) filter. As shown in studied cases, the proposed system demonstrates reliable and accurate location estimates in far-field ($T_{60} = 1$ s), and is applicable to tracking an unknown and time-varying number of moving speakers.

On the Design of Deep Priors for Unsupervised Audio Restoration

*Vivek Sivaraman Narayanaswamy*¹, Jayaraman J. *Thiagarajan*², *Andreas Spanias*¹; ¹*Arizona State University, USA*; ²*LLNL, USA* Wed-A-V-4-12, Time: 16:00

Unsupervised deep learning methods for solving audio restoration problems extensively rely on carefully tailored neural architectures that carry strong inductive biases for defining priors in the time or spectral domain. In this context, lot of recent success has been achieved with sophisticated convolutional network constructions that recover audio signals in the spectral domain. However, in practice, audio priors require careful engineering of the convolutional kernels to be effective at solving ill-posed restoration tasks, while also being easy to train. To this end, in this paper, we propose a new U-Net based prior that does not impact either the network complexity or convergence behavior of existing convolutional architectures, yet leads to significantly improved restoration. In particular, we advocate the use of carefully designed dilation schedules and dense connections in the U-Net architecture to obtain powerful audio priors. Using empirical studies on standard benchmarks and a variety of ill-posed restoration tasks, such as audio denoising, in-painting and source separation, we demonstrate that our proposed approach consistently outperforms widely adopted audio prior architectures.

Cramér-Rao Lower Bound for DOA Estimation with an Array of Directional Microphones in Reverberant Environments

Weiguang Chen, Cheng Xue, Xionghu Zhong; Hunan University, China

Wed-A-V-4-13, Time: 16:00

Existing direction-of-arrival (DOA) estimation methods usually assume that signals are received by an array of omnidirectional microphones. The performance can be seriously degraded due to heavy reverberation and noise. In this paper, DOA estimation using an array with directional microphones is considered. As the signal response varies over different DOAs, the magnitude information as well as the phase information can be employed to estimate the DOA. We first introduce the spherically isotropic noise field using directional microphones. The Cramér-Rao Lower Bound (CRLB) for DOA estimation is then derived and compared with that using omnidirectional microphones under different signal-to-reverberation ratio (SRR) environments. In addition, we extend existing steered response power (SRP), minimum variance distortionless response (MVDR) and multiple signal classification (MUSIC) estimators for the DOA estimation using directional microphone arrays. Both CRLB Analysis and DOA estimation show that better DOA estimation performance can be achieved by using a directional microphone arrav.

Wed-A-V-5: Speech Synthesis: Neural Waveform Generation

16:00-18:00, Wednesday 1 September 2021 Chairs: Shinnosuke Takamichi and Xin Wang

GAN Vocoder: Multi-Resolution Discriminator Is All You Need

Jaeseong You, Dalhyun Kim, Gyuhyeon Nam, Geumbyeol Hwang, Gyeongsu Chae; MoneyBrain, Korea Wed-A-V-5-1, Time: 16:00

Several of the latest GAN-based vocoders show remarkable achievements, outperforming autoregressive and flow-based competitors in both qualitative and quantitative measures while synthesizing orders of magnitude faster. In this work, we hypothesize that the common factor underlying their success is the multi-resolution discriminating framework, not the minute details in architecture, loss function, or training strategy. We experimentally test the hypothesis by evaluating six different generators paired with one shared multi-resolution discriminating framework. For all evaluative measures with respect to text-to-speech syntheses and for all perceptual metrics, their performances are not distinguishable from one another, which supports our hypothesis.

Glow-WaveGAN: Learning Speech Representations from GAN-Based Variational Auto-Encoder for High Fidelity Flow-Based Speech Synthesis

*Jian Cong*¹, *Shan Yang*², *Lei Xie*¹, *Dan Su*²; ¹*Northwestern Polytechnical University, China*; ²*Tencent, China* Wed-A-V-5-2, Time: 16:00

Current two-stage TTS framework typically integrates an acoustic model with a vocoder — the acoustic model predicts a low resolution intermediate representation such as Mel-spectrum while the vocoder generates waveform from the intermediate representation. Although the intermediate representation is served as a bridge, there still exists critical mismatch between the acoustic model and the vocoder as they are commonly separately learned and work on different distributions of representation, leading to inevitable artifacts in the synthesized speech. In this work, different from using pre-designed intermediate representation in most previous studies, we propose to use VAE combining with GAN to learn a latent representation directly from speech and then utilize a flow-based acoustic model to model the distribution of the latent representation from text. In this way, the mismatch problem is migrated as the two stages work on the same distribution. Results demonstrate that the flow-based acoustic model can exactly model the distribution of our learned speech representation and the proposed TTS framework, namely Glow-WaveGAN, can produce high fidelity speech outperforming the state-of-the-art GAN-based model.

Unified Source-Filter GAN: Unified Source-Filter Network Based On Factorization of Quasi-Periodic Parallel WaveGAN

Reo Yoneyama, Yi-Chiao Wu, Tomoki Toda; Nagoya University, Japan

Wed-A-V-5-3, Time: 16:00

We propose a unified approach to data-driven source-filter modeling using a single neural network for developing a neural vocoder capable of generating high-quality synthetic speech waveforms while retaining flexibility of the source-filter model to control their voice

characteristics. Our proposed network called unified source-filter generative adversarial networks (uSFGAN) is developed by factorizing quasi-periodic parallel WaveGAN (QPPWG), one of the neural vocoders based on a single neural network, into a source excitation generation network and a vocal tract resonance filtering network by additionally implementing a regularization loss. Moreover, inspired by neural source filter (NSF), only a sinusoidal waveform is additionally used as the simplest clue to generate a periodic source excitation waveform while minimizing the effect of approximations in the source filter model. The experimental results demonstrate that uSFGAN outperforms conventional neural vocoders, such as QPPWG and NSF in both speech quality and pitch controllability.

Harmonic WaveGAN: GAN-Based Speech Waveform Generation Model with Harmonic Structure Discriminator

Kazuki Mizuta, Tomoki Koriyama, Hiroshi Saruwatari; University of Tokyo, Japan Wed-A-V-5-4, Time: 16:00

This paper proposes Harmonic WaveGAN, a GAN-based waveform generation model that focuses on the harmonic structure of a speech waveform. Our proposed model uses two discriminators to capture characteristics of a speech waveform in a time domain and in a frequency domain, respectively. In one of them, a harmonic structure discriminator, a 2-D convolution layer called "harmonic convolution" is inserted to model a harmonic structure of a speech waveform. Although harmonic convolution has been shown to perform well in audio restoration tasks, this convolution layer has not yet been fully explored in the field of speech synthesis. Therefore, we seek to improve the perceptual quality of speech samples synthesized by the waveform generation model and investigate the usefulness of harmonic convolution in the field of speech synthesis. Mean opinion score tests showed that the Harmonic WaveGAN can synthesize more natural speech than conventional Parallel WaveGAN. We also showed that a spectrogram of a speech waveform showed a clearer harmonic structure when synthesized by our model than a speech waveform synthesized by the original Parallel WaveGAN.

Fre-GAN: Adversarial Frequency-Consistent Audio Synthesis

Ji-Hoon Kim, Sang-Hoon Lee, Ji-Hyun Lee, Seong-Whan Lee; Korea University, Korea Wed-A-V-5-5, Time: 16:00

Although recent works on neural vocoder have improved the quality of synthesized audio, there still exists a gap between generated and ground-truth audio in frequency space. This difference leads to spectral artifacts such as hissing noise or reverberation, and thus degrades the sample quality. In this paper, we propose Fre-GAN which achieves frequency-consistent audio synthesis with highly improved generation quality. Specifically, we first present resolutionconnected generator and resolution-wise discriminators, which help learn various scales of spectral distributions over multiple frequency bands. Additionally, to reproduce high-frequency components accurately, we leverage discrete wavelet transform in the discriminators. From our experiments, Fre-GAN achieves high-fidelity waveform generation with a gap of only 0.03 MOS compared to ground-truth audio while outperforming standard models in quality.

GANSpeech: Adversarial Training for High-Fidelity Multi-Speaker Speech Synthesis

Jinhyeok Yang, Jae-Sung Bae, Taejun Bak, Young-Ik Kim, Hoon-Young Cho; NCSOFT, Korea Wed-A-V-5-6, Time: 16:00

Recent advances in neural multi-speaker text-to-speech (TTS) models

have enabled the generation of reasonably good speech quality with a single model and made it possible to synthesize the speech of a speaker with limited training data. Fine-tuning to the target speaker data with the multi-speaker model can achieve better quality, however, there still exists a gap compared to the real speech sample and the model depends on the speaker. In this work, we propose GANSpeech, which is a high-fidelity multi-speaker TTS model that adopts the adversarial training method to a non-autoregressive multi-speaker TTS model. In addition, we propose simple but efficient automatic scaling methods for feature matching loss used in adversarial training. In the subjective listening tests, GANSpeech significantly outperformed the baseline multi-speaker FastSpeech and FastSpeech2 models, and showed a better MOS score than the speaker-specific fine-tuned FastSpeech2.

UnivNet: A Neural Vocoder with Multi-Resolution Spectrogram Discriminators for High-Fidelity Waveform Generation

Won Jang, Dan Lim, Jaesam Yoon, Bongwan Kim, Juntae Kim; Kakao, Korea

Wed-A-V-5-7, Time: 16:00

Most neural vocoders employ band-limited mel-spectrograms to generate waveforms. If full-band spectral features are used as the input, the vocoder can be provided with as much acoustic information as possible. However, in some models employing full-band melspectrograms, an over-smoothing problem occurs as part of which non-sharp spectrograms are generated. To address this problem, we propose UnivNet, a neural vocoder that synthesizes high-fidelity waveforms in real time. Inspired by works in the field of voice activity detection, we added a multi-resolution spectrogram discriminator that employs multiple linear spectrogram magnitudes computed using various parameter sets. Using full-band mel-spectrograms as input, we expect to generate high-resolution signals by adding a discriminator that employs spectrograms of multiple resolutions as the input. In an evaluation on a dataset containing information on hundreds of speakers, UnivNet obtained the best objective and subjective results among competing models for both seen and unseen speakers. These results, including the best subjective score for text-to-speech, demonstrate the potential for fast adaptation to new speakers without a need for training from scratch.

Continuous Wavelet Vocoder-Based Decomposition of Parametric Speech Waveform Synthesis

Mohammed Salah Al-Radhi, Tamás Gábor Csapó, Csaba Zainkó, Géza Németh; BME, Hungary Wed-A-V-5-8, Time: 16:00

To date, various speech technology systems have adopted the vocoder approach, a method for synthesizing speech waveform that shows a major role in the performance of statistical parametric speech synthesis. However, conventional source-filter systems (i.e., STRAIGHT) and sinusoidal models (i.e., MagPhase) tend to produce over-smoothed spectra, which often result in muffled and buzzy synthesized text-to-speech (TTS). WaveNet, one of the best models that nearly resembles the human voice, has to generate a waveform in a time-consuming sequential manner with an extremely complex structure of its neural networks. WaveNet needs large quantities of voice data before accurate predictions can be obtained. In order to motivate a new, alternative approach to these issues, we present an updated synthesizer, which is a simple signal model to train and easy to generate waveforms, using Continuous Wavelet Transform (CWT) to characterize and decompose speech features. CWT provides time and frequency resolutions different from those of the short-time Fourier transform. It can also retain the fine spectral envelope and achieve high controllability of the structure closer to human auditory

scales. We confirmed through experiments that our speech synthesis system was able to provide natural-sounding synthetic speech and outperformed the state-of-the-art WaveNet vocoder.

High-Fidelity and Low-Latency Universal Neural Vocoder Based on Multiband WaveRNN with Data-Driven Linear Prediction for Discrete Waveform Modeling

Patrick Lumban Tobing, Tomoki Toda; Nagoya University, Japan Wed-A-V-5-9, Time: 16:00

This paper presents a novel high-fidelity and low-latency universal neural vocoder framework based on multiband WaveRNN with datadriven linear prediction for discrete waveform modeling (MWDLP). MWDLP employs a coarse-fine bit WaveRNN architecture for 10-bit mu-law waveform modeling. A sparse gated recurrent unit with a relatively large size of hidden units is utilized, while the multiband modeling is deployed to achieve real-time low-latency usage. A novel technique for data-driven linear prediction (LP) with discrete waveform modeling is proposed, where the LP coefficients are estimated in a data-driven manner. Moreover, a novel loss function using short-time Fourier transform (STFT) for discrete waveform modeling with Gumbel approximation is also proposed. The experimental results demonstrate that the proposed MWDLP framework generates high-fidelity synthetic speech for seen and unseen speakers and/or language on 300 speakers training data including clean and noisy/reverberant conditions, where the number of training utterances is limited to 60 per speaker, while allowing for real-time low-latency processing using a single core of ~2.1-2.7 GHz CPU with $\sim 0.57-0.64$ real-time factor including input/output and feature extraction.

Basis-MelGAN: Efficient Neural Vocoder Based on Audio Decomposition

Zhengxi Liu¹, Yanmin Qian²; ¹Sun Yat-sen University, China; ²SJTU, China

Wed-A-V-5-10, Time: 16:00

Recent studies have shown that neural vocoders based on generative adversarial network (GAN) can generate audios with high quality. While GAN based neural vocoders have shown to be computationally much more efficient than those based on autoregressive predictions, the real-time generation of the highest quality audio on CPU is still a very challenging task. One major computation of all GAN-based neural vocoders comes from the stacked upsampling layers, which were designed to match the length of the waveform's length of output and temporal resolution. Meanwhile, the computational complexity of upsampling networks is closely correlated with the numbers of samples generated for each window. To reduce the computation of upsampling layers, we propose a new GAN based neural vocoder called Basis-MelGAN where the raw audio samples are decomposed with a learned basis and their associated weights. As the prediction targets of Basis-MelGAN are the weight values associated with each learned basis instead of the raw audio samples, the upsampling layers in Basis-MelGAN can be designed with much simpler networks. Compared with other GAN based neural vocoders, the proposed Basis-MelGAN could produce comparable high-quality audio but significantly reduced computational complexity from HiFi-GAN V1's 17.74 GFLOPs to 7.95 GFLOPs.

High-Fidelity Parallel WaveGAN with Multi-Band Harmonic-Plus-Noise Model

*Min-Jae Hwang*¹, *Ryuichi Yamamoto*², *Eunwoo Song*³, *Jae-Min Kim*³; ¹*Search Solutions, Korea*; ²*LINE, Japan*; ³*Naver, Korea*

Wed-A-V-5-11, Time: 16:00

This paper proposes a multi-band harmonic-plus-noise (HN) Parallel WaveGAN (PWG) vocoder. To generate a high-fidelity speech signal, it is important to well-reflect the harmonic-noise characteristics of the speech waveform in the time-frequency domain. However, it is difficult for the conventional PWG model to accurately match this condition, as its single generator inefficiently represents the complicated nature of harmonic-noise structures. In the proposed method, the HN WaveNet models are employed to overcome this limitation, which enable the separate generation of the harmonic and noise components of speech signals from the pitch-dependent sine wave and Gaussian noise sources, respectively. Then, the energy ratios between harmonic and noise components in multiple frequency bands (i.e., subband harmonicities) are predicted by an additional harmonicity estimator. Weighted by the estimated harmonicities, the gain of harmonic and noise components in each subband is adjusted. and finally mixed together to compose the full-band speech signal. Subjective evaluation results showed that the proposed method significantly improved the perceptual quality of the synthesized speech.

Wed-A-V-6: Spoken Machine Translation 16:00-18:00, Wednesday 1 September 2021

Chairs: Chiori Hori and Helena Moniz

SpecRec: An Alternative Solution for Improving End-to-End Speech-to-Text Translation via Spectrogram Reconstruction

Junkun Chen¹, Mingbo Ma², Renjie Zheng², Liang Huang²; ¹Oregon State University, USA; ²Baidu, USA Wed-A-V-6-1, Time: 16:00

End-to-end Speech-to-text Translation (E2E-ST), which directly translates source language speech to target language text, is widely useful in practice, but traditional cascaded approaches (ASR+MT) often suffer from error propagation in the pipeline. On the other hand, existing end-to-end solutions heavily depend on the source language transcriptions for pre-training or multi-task training with Automatic Speech Recognition (ASR). We instead propose a simple technique to learn a robust speech encoder in a self-supervised fashion only on the speech side, which can utilize speech data without transcription. This technique termed Spectrogram Reconstruction (SpecRec), learns better speech representation via recovering the missing speech frames and provides an alternative solution to improving E2E-ST. We conduct our experiments over 8 different translation directions. In the setting without using any transcriptions, our technique achieves an average improvement of +1.1 BLEU. SpecRec also improves the translation accuracy with +0.7 BLEU over the baseline in speech translation with ASR multitask training setting.

Subtitle Translation as Markup Translation

Colin Cherry, Naveen Arivazhagan, Dirk Padfield, Maxim Krikun; Google, USA Wed-A-V-6-2, Time: 16:00

Automatic subtitle translation is an important technology to make video content available across language barriers. Subtitle translation complicates the normal translation problem by adding the challenge

of how to format the system output into subtitles. We propose a simple technique that treats subtitle translation as standard sentence translation plus alignment driven markup transfer, which enables us to reliably maintain timing and formatting information from the source subtitles. We also introduce two metrics to measure the quality of subtitle boundaries: a Timed BLEU that penalizes mistimed tokens with respect to a reference subtitle sequence, and a measure of how much Timed BLEU is lost due to suboptimal subtitle boundary placement. In experiments on TED and YouTube subtitles, we show that we are able to achieve much better translation quality than a baseline that translates each subtitle independently, while coming very close to optimal subtitle boundary placement.

Large-Scale Self- and Semi-Supervised Learning for Speech Translation

Changhan Wang, Anne Wu, Juan Pino, Alexei Baevski, Michael Auli, Alexis Conneau; Facebook, USA Wed-A-V-6-3, Time: 16:00

In this paper, we improve speech translation (ST) through effectively leveraging large quantities of unlabeled speech and text data in different and complementary ways. We explore both pretraining and self-training by using the large Libri-Light speech audio corpus and language modeling with CommonCrawl. Our experiments improve over the previous state of the art by 2.8 BLEU on average on all four considered CoVoST 2 language pairs via a simple recipe of combining wav2vec 2.0 pretraining, a single iteration of self-training and decoding with a language model. Different from existing work, our approach does not leverage any other supervision than ST data. Code and models are publicly released.

CoVoST 2 and Massively Multilingual Speech Translation

Changhan Wang, Anne Wu, Jiatao Gu, Juan Pino; Facebook, USA Wed-A-V-6-4, Time: 16:00

Speech translation (ST) is an increasingly popular topic of research, partly due to the development of benchmark datasets. Nevertheless, current datasets cover a limited number of languages. With the aim to foster research into massive multilingual ST and ST for low resource languages, we release CoVoST 2, a large-scale multilingual ST corpus covering translations from 21 languages into English and from English into 15 languages. This represents the largest open dataset available to date for volume and language coverage. Data checks provide evidence about the data quality. We provide extensive speech recognition (ASR), machine translation (MT) and ST baselines. We demonstrate the value of CoVoST 2 for multilingual ST research by leveraging it in 4 investigations: simplify multilingual model scaling properties and investigate zero-shot and transfer learning capabilities of models trained on CoVoST 2.

AlloST: Low-Resource Speech Translation Without Source Transcription

Yao-Fei Cheng, Hung-Shin Lee, Hsin-Min Wang; Academia Sinica, Taiwan Wed-A-V-6-5, Time: 16:00

The end-to-end architecture has made promising progress in speech translation (ST). However, the ST task is still challenging under low-resource conditions. Most ST models have shown unsatisfactory results, especially in the absence of word information from the source speech utterance. In this study, we survey methods to improve ST performance without using source transcription, and

propose a learning framework that utilizes a language-independent universal phone recognizer. The framework is based on an attentionbased sequence-to-sequence model, where the encoder generates the phonetic embeddings and phone-aware acoustic representations, and the decoder controls the fusion of the two embedding streams to produce the target token sequence. In addition to investigating different fusion strategies, we explore the specific usage of byte pair encoding (BPE), which compresses a phone sequence into a syllable-like segmented sequence. Due to the conversion of symbols, a segmented sequence represents not only pronunciation but also language-dependent information lacking in phones. Experiments conducted on the Fisher Spanish-English and Taigi-Mandarin drama corpora show that our method outperforms the conformer-based baseline, and the performance is close to that of the existing best method using source transcription.

Weakly-Supervised Speech-to-Text Mapping with Visually Connected Non-Parallel Speech-Text Data Using Cyclic Partially-Aligned Transformer

Johanes Effendi, Sakriani Sakti, Satoshi Nakamura; NAIST, Japan Wed-A-V-6-6, Time: 16:00

Despite the successful development of automatic speech recognition (ASR) systems for several of the world's major languages, they require a tremendous amount of parallel speech-text data. Unfortunately, for many other languages, such resources are usually unavailable. This study addresses the speech-to-text mapping problem given only a collection of visually connected non-parallel speech-text data. We call this "mapping" since the system attempts to learn the semantic association between speech and text instead of recognizing the speech with the exact word-by-word transcription. Here, we propose utilizing our novel cyclic partially-aligned Transformer with two-fold mechanisms. First, we train a Transformer-based vectorquantized variational autoencoder (VQ-VAE) to produce a discrete speech representation in a self-supervised manner. Then, we use a Transformer-based sequence-to-sequence model inside a chain mechanism to map from unknown untranscribed speech utterances into a semantically equivalent text. Because this is not strictly recognizing speech, we focus on evaluating the semantic equivalence of the generated text hypothesis. Our evaluation shows that our proposed method is also effective for a multispeaker natural speech dataset and can also be applied for a cross-lingual application.

Transcribing Paralinguistic Acoustic Cues to Target Language Text in Transformer-Based Speech-to-Text Translation

Hirotaka Tokuyama, Sakriani Sakti, Katsuhito Sudoh, Satoshi Nakamura; NAIST, Japan Wed-A-V-6-7, Time: 16:00

In spoken communication, a speaker may convey their message in words (linguistic cues) with supplemental information (paralinguistic cues) such as emotion and emphasis. Transforming all spoken information into a written or verbal form is not trivial, especially if the transformation has to be done across languages. Most existing speech-to-text translation systems focus only on translating linguistic information while ignoring paralinguistic translation. A few recent studies that proposed paralinguistic translation used a machine translation with hidden Markov model (HMM)-based automatic speech recognition (ASR) and text-to-speech (TTS) that were complicated and suboptimal. Furthermore, paralinguistic information was kept in the acoustic form. Here, we focused on transcribing paralinguistic acoustic cues of emphasis in the target language text. Specifically, we constructed cascade and direct neural

Transformer-based speech-to-text translation, and we investigated various methods of expressing emphasis information in the written form of the target language. We performed our experiments on a Japanese-to-English linguistic and paralinguistic speech-to-text translation framework. The results revealed that our proposed method can translate both linguistic and paralinguistic information while keeping the performance as in standard linguistic translation.

End-to-End Speech Translation via Cross-Modal Progressive Training

Rong Ye, Mingxuan Wang, Lei Li; ByteDance, China Wed-A-V-6-8, Time: 16:00

End-to-end speech translation models have become a new trend in research due to their potential of reducing error propagation. However, these models still suffer from the challenge of data scarcity. How to effectively use unlabeled or other parallel corpora from machine translation is promising but still an open problem. In this paper, we propose Cross Speech-Text Network (XSTNet), an end-to-end model for speech-to-text translation. XSTNet takes both speech and text as input and outputs both transcription and translation text. The model benefits from its three key design aspects: a self-supervised pre-trained sub-network as the audio encoder, a multi-task training objective to exploit additional parallel bilingual text, and a progressive training procedure. We evaluate the performance of XSTNet and baselines on the MuST-C En-X and LibriSpeech En-Fr datasets. In particular, XSTNet achieves state-of-the-art results on all language directions with an average BLEU of 28.8, outperforming the previous best method by 3.2 BLEU. Code, models, cases, and more detailed analysis are publicly available.

ASR Posterior-Based Loss for Multi-Task End-to-End Speech Translation

Yuka Ko, Katsuhito Sudoh, Sakriani Sakti, Satoshi Nakamura; NAIST, Japan Wed-A-V-6-9, Time: 16:00

End-to-end speech translation (ST) translates source language speech directly into target language without an intermediate automatic speech recognition (ASR) output, as in a cascading approach. Endto-end ST has the advantage of avoiding error propagation from the intermediate ASR results, but its performance still lags behind the cascading approach. A recent effort to increase performance is multi-task learning using an auxiliary task of ASR. However, previous multi-task learning for end-to-end ST using cross entropy (CE) loss in ASR-task targets one-hot references and does not consider ASR confusion. In this study, we propose a novel end-to-end ST training method using ASR loss against ASR posterior distributions given by a pre-trained model, which we call ASR posterior-based loss. The proposed method is expected to consider possible ASR confusion due to competing hypotheses with similar pronunciations. The proposed method demonstrated better BLEU results in our Fisher Spanish-to-English translation experiments than the baseline with standard CE loss with label smoothing.

Towards Simultaneous Machine Interpretation

Alejandro Pérez-González-de-Martos, Javier Iranzo-Sánchez, Adrià Giménez Pastor, Javier Jorge, Joan-Albert Silvestre-Cerdà, Jorge Civera, Albert Sanchis, Alfons Juan; Universitat Politècnica de València, Spain Wed-A-V-6-10, Time: 16:00

Automatic speech-to-speech translation (S2S) is one of the most challenging speech and language processing tasks, especially when considering its application to real-time settings. Recent advances on streaming Automatic Speech Recognition (ASR), simultaneous Machine Translation (MT) and incremental neural Text-To-Speech (TTS) make it possible to develop real-time cascade S2S systems with greatly improved accuracy. On the way to simultaneous machine interpretation, a state-of-the-art cascade streaming S2S system is described and empirically assessed in the simultaneous interpretation of European Parliament debates. We pay particular attention to the TTS component, particularly in terms of speech naturalness under a variety of response-time settings, as well as in terms of speaker similarity for its cross-lingual voice cloning capabilities.

Lexical Modeling of ASR Errors for Robust Speech Translation

*Giuseppe Martucci*¹, *Mauro Cettolo*², *Matteo Negri*², *Marco Turchi*²; ¹*Università di Trento, Italy*; ²*FBK, Italy* Wed-A-V-6-11, Time: 16:00

Error propagation from automatic speech recognition (ASR) to machine translation (MT) is a critical issue for the (still) dominant cascade approach to speech translation. To robustify MT to illformed inputs, we propose a technique to artificially corrupt clean transcripts so as to emulate noisy automatic transcripts. Our Lexical Noise model relies on estimating from ASR data: i) the probability distribution of the possible edit operations applicable to each word, and ii) the probability distribution of possible lexical substitutes for that word. Corrupted data generated from these probabilities are paired with their original clean counterpart for MT adaptation via fine-tuning. Contrastive experiments on three language pairs led to three main findings. First, on noisy transcripts, the adapted models outperform MT systems fine-tuned on synthetic data corrupted with previous noising techniques, approaching the upper bound performance obtained by fine-tuning on real ASR data. Second, the increased robustness does not come at the cost of performance drops on clean test data. Third, and crucial from the application standpoint, our approach is domain/ASR-independent: noising patterns learned from a given ASR system in a certain domain can be successfully applied to robustify MT to errors made by other ASR systems in a different domain.

Optimally Encoding Inductive Biases into the Transformer Improves End-to-End Speech Translation

Piyush Vyas, Anastasia Kuznetsova, Donald S. Williamson; Indiana University, USA Wed-A-V-6-12, Time: 16:00

Transformer-based encoder-decoder architectures have recently shown promising results in end-to-end speech translation. However, the content-based attention mechanism employed by the Transformer was designed for text sequences and can only encode global inductive bias, that alone is not sufficient for learning good representations from speech signals. In this work, we address this by putting architectural constraints on the Transformer to allow encoding of both local and global inductive biases. This is accomplished by replacing the Transformer encoder with a Conformer encoder that, in contrast to the Transformer encoder, employs convolution in addition to self-attention and feed-forward. As a result, the new model named Conformer-Transformer has an encoder that captures both local feature correlations and long-range dependencies from speech signals. Experiments on seven non-English to English language directions show that the Conformer-Transformer, compared to strong Transformer-based baselines, achieves up to 3.54 BLEU score improvements with a pre-trained encoder and up to 10.53 BLEU score improvements when trained from scratch.

Effects of Feature Scaling and Fusion on Sign Language Translation

Tejaswini Ananthanarayana, Lipisha Chaudhary, Ifeoma Nwogu; Rochester Institute of Technology, USA Wed-A-V-6-13, Time: 16:00

Sign language translation without transcription has only recently started to gain attention. In our work, we focus on improving the state-of-the-art translation by introducing a multi-feature fusion architecture with enhanced input features. As sign language is challenging to segment, we obtain the input features by extracting overlapping scaled segments across the video and obtaining their 3D CNN representations. We exploit the attention mechanism in the fusion architecture by initially learning dependencies between different frames of the same video and later fusing them to learn the relations between different features from the same video. In addition to 3D CNN features, we also analyze pose-based features.

Our robust methodology outperforms the state-of-the-art sign language translation model by achieving higher BLEU 3 – BLEU 4 scores and also outperforms the state-of-the-art sequence attention models by achieving a 43.54% increase in BLEU 4 score. We conclude that the combined effects of feature scaling and feature fusion make our model more robust in predicting longer n-grams which are crucial in continuous sign language translation.

Wed-A-SS-1: SdSV Challenge 2021: Analysis and Exploration of New Ideas on Short-Duration Speaker Verification

16:00–18:00, Wednesday 1 September 2021 Chairs: Jahangir Alam and Kong Aik Lee

The ID R&D System Description for Short-Duration Speaker Verification Challenge 2021

Alexander Alenin, Anton Okhotnikov, Rostislav Makarov, Nikita Torgashov, Ilya Shigabeev, Konstantin Simonchik; ID R&D, USA

Wed-A-SS-1-1, Time: 16:00

This paper describes ID R&D team submission to the textindependent task of the Short-duration Speaker Verification (SdSV) Challenge 2021. The top performed system is a fusion of 9 Convolutional Neural Networks based on the ResNet architecture. Experiments' results of optimal NN architecture search are shown. We also present and investigate the subnetwork approach to solve the auxiliary tasks such as gender or language detection. Verification scores refinement step using quality measurements of a trial pair allowed to further minimize the target metrics. A comparative analysis of all systems used in the fusion has been provided on the VoxCeleb-1 test set, SdSV-2021 development and evaluation sets. The final submission achieves **0.69**% EER and **0.0319** minDCF on the challenge evaluation set.

Integrating Frequency Translational Invariance in TDNNs and Frequency Positional Information in 2D ResNets to Enhance Speaker Verification

Jenthe Thienpondt, Brecht Desplanques, Kris Demuynck; Ghent University, Belgium

Wed-A-SS-1-2, Time: 16:00

This paper describes the IDLab submission for the text-independent task of the Short-duration Speaker Verification Challenge 2021 (SdSVC-21). This speaker verification competition focuses on short duration test recordings and cross-lingual trials, along with the constraint of limited availability of in-domain DeepMine Farsi training data. Currently, both Time Delay Neural Networks (TDNNs) and ResNets achieve state-of-the-art results in speaker verification. These architectures are structurally very different and the construction of hybrid networks looks a promising way forward. We introduce a 2D convolutional stem in a strong ECAPA-TDNN baseline to transfer some of the strong characteristics of a ResNet based model to this hybrid CNN-TDNN architecture. Similarly, we incorporate absolute frequency positional encodings in an SE-ResNet34 architecture. These learnable feature map biases along the frequency axis offer this architecture a straightforward way to exploit frequency positional information. We also propose a frequency-wise variant of Squeeze-Excitation (SE) which better preserves frequency-specific information when rescaling the feature maps. Both modified architectures significantly outperform their corresponding baseline on the SdSVC-21 evaluation data and the original VoxCeleb1 test set. A four system fusion containing the two improved architectures achieved a third place in the final SdSVC-21 Task 2 ranking.

SdSVC Challenge 2021: Tips and Tricks to Boost the Short-Duration Speaker Verification System Performance

Aleksei Gusev, Alisa Vinogradova, Sergey Novoselov, Sergei Astapov; ITMO University, Russia Wed-A-SS-1-3, Time: 16:00

This paper presents speaker recognition (SR) systems for the textindependent speaker verification under the cross-lingual (English vs Persian) task (task 2) of the Short-duration Speaker Verification Challenge (SdSVC) 2021.

We present the description of applied ResNet-like and ECAPA-TDNNlike topology design solutions as well as an analysis of multi-session scoring techniques benchmarked on the SdSVC challenge datasets. We overview various modifications of the basic ResNet-like architecture and training strategies, allowing us to obtain the improved quality of speaker verification. Also, we introduce the alpha query expansion-based technique (α QE) to the enrollment embeddings aggregation at test time, which results in a 0.042 minDCF improvement from 0.12 to 0.078 for the ECAPA-TDNN system compared to the embeddings mean. We also propose a trial-level distance-based non-parametric imposter/target detector (KrTC) used to filter out the worst enrollment samples at test time to further improve the performance of the system.

Team02 Text-Independent Speaker Verification System for SdSV Challenge 2021

*Woo Hyun Kang*¹, *Nam Soo Kim*²; ¹*CRIM, Canada;* ²*Seoul National University, Korea* Wed-A-SS-1-4, Time: 16:00

In this paper, we provide description of our submitted systems to the Short Duration Speaker Verification (SdSV) Challenge 2021 Task The challenge provides a difficult set of cross-language textindependent speaker verification trials. Our submissions employ ResNet-based embedding networks which are trained using various strategies exploiting both in-domain and out-of-domain datasets. The results show that using the recently proposed joint factor embedding (JFE) scheme can enhance the performance by disentangling the language-dependent information from the speaker embedding. However, upon analyzing the speaker embeddings, it was found that there exists a clear discrepancy between the in-domain and out-of-domain datasets. Therefore, among our submitted systems, the best performance was achieved by pre-training the embedding system using out-of-domain dataset and fine-tuning it with only the in-domain data, which resulted in a MinDCF of 0.142716 on the SdSV2021 evaluation set.

Our Learned Lessons from Cross-Lingual Speaker Verification: The CRMI-DKU System Description for the Short-Duration Speaker Verification Challenge 2021

*Xiaoyi Qin*¹, *Chao Wang*², *Yong Ma*², *Min Liu*², *Shilei Zhang*², *Ming Li*¹; ¹*Wuhan University, China*; ²*China Mobile, China*

Wed-A-SS-1-5, Time: 16:00

In this paper, we present our CRMI-DKU system description for the Short-duration Speaker Verification Challenge (SdSVC) 2021. We introduce the whole pipeline of our cross-lingual speaker verification system, including data preprocessing, training strategy, utterance-level speaker embedding extractor, domain-adaptation, and score calibration. We also propose methods to learn language-invariant features and perform domain adaptation to reduce the cross-lingual mismatch. In addition, we explore a semi-supervised method to utilize the unlabeled training data. The final submitted score level fusion system achieves 0.0476 minDCF and 0.98% EER on the evaluation set.

Investigation of IMU&Elevoc Submission for the Short-Duration Speaker Verification Challenge 2021

*Peng Zhang*¹, *Peng Hu*², *Xueliang Zhang*¹; ¹*Inner Mongolia University, China;* ²*Elevoc Technology, China* Wed-A-SS-1-6, Time: 16:00

In this paper, we present the IMU&Elevoc systems submitted to the Short-duration Verification Challenge (SdSVC) 2021. Our submissions focus on both text-dependent speaker verification (Task 1) and text-independent speaker verification (Task 2). First, we investigate several frame-level feature extractor architectures based on ResNet, Res2Net and TDNN. Then, we integrate Squeeze-Excitation block and dimension cardinality to further improve the Res2Net-based backbone network. In particular, we probe an effective transfer learning strategy that overcomes the lack of Task 1 datasets and improves in-domain performance. A knowledge distillation method fusing multiple models is proposed to obtain a stronger single model. Experimental results on the SdSVC 2021 show that our primary system yields 0.0500MinDCF in Task 1 (ranked as 4th) and 0.0448 MinDCF in Task 2 (ranked as 6th).

The Sogou System for Short-Duration Speaker Verification Challenge 2021

Jie Yan, Shengyu Yao, Yiqian Pan, Wei Chen; Sogou, China

Wed-A-SS-1-7, Time: 16:00

In this paper we present our system for the task 2 of the Shortduration Speaker Verification (SdSV) Challenge 2021. This task focuses on benchmarking and varying degrees of phonetic variability analysis of short-duration speaker recognition system. The main difficulty exists in the variance between cross-lingual trials, along with the limited in-domain Farsi training data. Based on the state-of-the-art ResNetSE speaker embedding network, we propose a novel network architecture with in-domain data finetuning and novel scoring methods, and achieve significant improvement over the ResNetSE baselines. Furthermore, score calibration on duration efficiently improve the robustness. Finally, our system with fusion of 10 subsystems achieve satisfying results in MinDCF and EER of 0.0394 and 0.84% respectively on the SdSVC evaluation set.

NOTES

The SJTU System for Short-Duration Speaker Verification Challenge 2021

Bing Han, Zhengyang Chen, Zhikai Zhou, Yanmin Qian; SJTU, China

Wed-A-SS-1-8, Time: 16:00

This paper presents the SJTU system for both text-dependent and text-independent tasks in short-duration speaker verification (SdSV) challenge 2021. In this challenge, we explored different strong embedding extractors to extract robust speaker embedding. For text-independent task, language-dependent adaptive snorm is explored to improve the system performance under the cross-lingual verification condition. For text-dependent task, we mainly focus on the in-domain fine-tuning strategies based on the model pre-trained on large-scale out-of-domain data. In order to improve the distinction between different speakers uttering the same phrase, we proposed several novel phrase-aware fine-tuning strategies and phrase-aware neural PLDA. With such strategies, the system performance is further improved. Finally, we fused the scores of different systems, and our fusion systems achieved 0.0473 in Task1 (rank 3) and 0.0581 in Task2 (rank 8) on the primary evaluation metric.

Wed-A-S&T-1 : Show and Tell 2

16:00-18:00, Wednesday 1 September 2021

Multi-Speaker Emotional Text-to-Speech Synthesizer

*Sungjae Cho*¹, *Soo-Young Lee*²; ¹*KIST, Korea*; ²*KAIST, Korea*

Wed-A-S&T-1-1, Time: 16:00

We present a methodology to train our multi-speaker emotional text-to-speech synthesizer that can express speech for 10 speakers' 7 different emotions. All silences from audio samples are removed prior to learning. This results in fast learning by our model. Curriculum learning is applied to train our model efficiently. Our model is first trained with a large single-speaker neutral dataset, and then trained with neutral speech from all speakers. Finally, our model is trained using datasets of emotional speech from pair have equal probability to appear in mini-batches. Through this procedure, our model can synthesize audio sets are available on our web page.

Live TV Subtitling Through Respeaking

Aleš Pražák¹, Zdeněk Loose², Josef V. Psutka¹, Vlasta Radová¹, Josef Psutka¹, Jan Švec¹; ¹University of West Bohemia, Czechia; ²SpeechTech, Czechia Wed-A-S&T-1-2, Time: 16:00

In this paper, we describe our solution for live TV subtitling. The subtitling system uses the respeaking concept with respeakers closely tied with the automatic speech recognition system. The ASR is specially tailored to the live subtitling task by using respeaker-specific acoustic models and TV-show-dependent language models. The output stream of ASR could be online modified by keyboard shortcuts controlled by the respeaker. The whole subtitling service is used by Czech Television to provide high-quality subtitles of live shows for people with hearing impairments.

Autonomous Robot for Measuring Room Impulse Responses

*Stefan Fragner*¹, *Tobias Topar*¹, *Maximilian Giller*¹, *Lukas Pfeifenberger*², *Franz Pernkopf*¹; ¹*Technische Universität Graz, Austria*; ²*Evolve, Austria* Wed-A-S&T-1-3, Time: 16:00

Far-field speech recognition for e.g. home automation or smart assistants has to cope with moving speakers in reverberant environments. Simulating stationary or even moving speakers in realistic environments enables to make speech processing technology more robust. This paper introduces an autonomous robot for recording a database of Room Impulse Responses (RIRs) at a high spatial resolution. This supports the creation of realistic simulation environments. These RIRs can be exploited to generate multi-channel speech mixtures of static or moving speakers for various applications.

Expressive Robot Performance Based on Facial Motion Capture

Jonas Beskow, Charlie Caper, Johan Ehrenfors, Nils Hagberg, Anne Jansen, Chris Wood; Furhat Robotics, Sweden

Wed-A-S&T-1-4, Time: 16:00

The Furhat robot is a social robot that uses facial projection technology to achieve a high degree of expressivity and flexibility. In this demonstration, we will present new features that takes this facial expressiveness further. A new face engine for the robot is presented which not only drastically improves the visual fidelity of the face and the eyes, it also adds increased flexibility when it comes to designing new robotic characters as well as modifying existing ones. Most importantly, we will present a new toolset and a workflow that allows users to record their own face motion and incorporate them into skills (i.e. custom robot applications) as gestures, prompts or entire canned performances.

ThemePro 2.0: Showcasing the Role of Thematic Progression in Engaging Human-Computer Interaction

Mónica Domínguez, Juan Soler-Company, Leo Wanner; Universitat Pompeu Fabra, Spain Wed-A-S&T-1-5, Time: 16:00

Structuring speech into informative units is certainly a desirable feature in efficient human-machine communication. This paper introduces ThemePro 2.0, a toolkit that pre-processes long monologues into smaller cohesive units to be consumed by the text-to-speech module within a conversational agent. The methodology used is based upon the text's discourse structure modelled as thematic progression patterns. As shown in the demonstration, thematic progression modelling captures the underlying information structure at the discourse level and is, therefore, instrumental for cohesive speech output in the TTS component.

Addressing Compliance in Call Centers with Entity Extraction

Sai Guruju, Jithendra Vepa; Observe.AI, India Wed-A-S&T-1-6, Time: 16:00

Call centers record and store customer-agent conversations for the purpose of coaching, quality assurance and to comply with *Industry Regulations*. Good amount of these audio recordings contain sensitive information pertaining to their customers' financial or personal details. To ensure data security, compliance and to reduce the risk of abuse/theft, it becomes important to identify such instances in audio recordings and mask these segments. To automate this

NOTES

process, we propose a cascaded system; first, Automatic Speech Recognition (ASR) generates transcript and text-to-audio alignment information for an audio recording. Then, Entity Extraction is performed on generated transcripts to identify and locate sensitive information, and the corresponding sensitive segments are masked in audio recordings using alignment information. We introduce a novel system for selective masking of sensitive information in both audio and transcript.

Audio Segmentation Based Conversational Silence Detection for Contact Center Calls

Krishnachaitanya Gogineni, Tarun Reddy Yadama, Jithendra Vepa; Observe.AI, India Wed-A-S&T-1-7, Time: 16:00

In a typical contact-center call, more than 35% of the call has neither the contact-center agent nor the customer speaking, we usually refer to such areas in the call as *Conversational Silences*. *Conversational silences* comprise mostly of hold-music, automatic-recorded-messages, or just silences when the agent or customer is engaged in some offcall work. Most of these conversational silences negatively affect important KPIs for call-centers, like dead-airs affect customer satisfaction, long-holds affect average call handling time and so on. In this paper we showcase how Observe.AI helps contact-centers identify agents who are breaching accepted levels of conversational silences by using an in-house Audio Segmenter system paired with an NLP system to classify the contexts around these *Conversational Silences*. This solution is provided by Observe.AI to hundreds of contact centers who use it to improve their average call handling time and customer satisfaction scores.

Wed-E-O-1: Graph and End-to-End Learning for Speaker Recognition

Room A+B, 19:00-21:00, Wednesday 1 September 2021 Chairs: Johan Rohdin and Madhu Kamble

Reformulating DOVER-Lap Label Mapping as a Graph Partitioning Problem

Desh Raj, Sanjeev Khudanpur; Johns Hopkins University, USA

Wed-E-O-1-1, Time: 19:00

We recently proposed DOVER-Lap, a method for combining overlapaware speaker diarization system outputs. DOVER-Lap improved upon its predecessor DOVER by using a label mapping method based on globally-informed greedy search. In this paper, we analyze this label mapping in the framework of a maximum orthogonal graph partitioning problem, and present three inferences. First, we show that DOVER-Lap label mapping is exponential in the input size, which poses a challenge when combining a large number of hypotheses. We then revisit the DOVER label mapping algorithm and propose a modification which performs similar to DOVER-Lap while being computationally tractable. We also derive an approximation bound for the algorithm in terms of the maximum number of hypotheses speakers. Finally, we describe a randomized local search algorithm which provides a near-optimal $(1-\varepsilon)$ -approximate solution to the problem with high probability. We empirically demonstrate the effectiveness of our methods on the AMI meeting corpus. Our code is publicly available.

Graph Attention Networks for Anti-Spoofing

Hemlata Tak¹, Jee-weon Jung², Jose Patino¹, Massimiliano Todisco¹, Nicholas Evans¹; ¹EURECOM, France; ²Naver, Korea Wed-E-O-1-2, Time: 19:20

The cues needed to detect spoofing attacks against automatic speaker verification are often located in specific spectral sub-bands or temporal segments. Previous works show the potential to learn these using either spectral or temporal self-attention mechanisms but not the relationships between neighbouring sub-bands or segments. This paper reports our use of graph attention networks (GATs) to model these relationships and to improve spoofing detection performance. GATs leverage a self-attention mechanism over graph structured data to model the data manifold and the relationships between nodes. Our graph is constructed from representations produced by a ResNet. Nodes in the graph represent information either in specific sub-bands or temporal segments. Experiments performed on the ASVspoof 2019 logical access database show that our GAT-based model with temporal attention outperforms all of our baseline single systems. Furthermore, GAT-based systems are complementary to a set of existing systems. The fusion of GAT-based models with more conventional countermeasures delivers a 47% relative improvement in performance compared to the best performing single GAT system.

Log-Likelihood-Ratio Cost Function as Objective Loss for Speaker Verification Systems

Victoria Mingote, Antonio Miguel, Alfonso Ortega, Eduardo Lleida; Universidad de Zaragoza, Spain Wed-E-O-1-3, Time: 19:40

Many recent studies in Speaker Verification (SV) have been focused on the design of the most appropriate training loss function, which plays an important role to improve the recognition ability of the systems. However, the verification loss functions created often do not take into account the performance measures which are used for the final system evaluation. For this reason, this paper presents an alternative approach to optimize the parameters of a neural network using a loss function based on the log-likelihood-ratio cost function (CLLR). This function is an application-independent metric that measures the cost of soft detection decisions over all the operating points. Thus, prior or relevance cost parameters assumptions are not employed to obtain it. Moreover, this metric has a differentiable expression, so no approximation is needed to use it as the objective loss to train a neural network. CLLR function as optimization loss was tested on the RSR2015-Part II database for text-dependent speaker verification, providing competitive results without using score normalization and outperforming other similar loss functions as Cross-Entropy combined with Ring Loss, as well as our previous loss function based on an approximation of the Detection Cost Function (DCF).

Effective Phase Encoding for End-To-End Speaker Verification

Junyi Peng¹, Xiaoyang Qu¹, Rongzhi Gu², Jianzong Wang¹, Jing Xiao¹, Lukáš Burget³, Jan Černocký³; ¹Ping An Technology, China; ²Peking University, China; ³Brno University of Technology, Czechia Wed-E-O-1-4, Time: 20:00

The widely used magnitude spectrum based features have shown their superiority in the field of speech processing. In contrast, the importance of phase spectrum is always ignored. This is because the patterns hidden in phase cannot be intuitively modelled and interpreted, due to phase wrapping phenomenon. In this paper,

we explore novel phase spectrum based features, named Learnable Group Delay (LearnGD), to capture useful information in speech signals. Specifically, firstly, the negative of the spectral derivative of the phase spectrum, called group delay (GD), is used to unwrap the phase. Then, to suppress the spiky nature of GD, which is caused by its roots close to the unit circle in the Z domain, a carefully designed light convolutional smoothing layer is employed to reconstruct the GD. Finally, an exponential hyper-parameter is introduced to reconstruct GD features to restore the spectrum range and generate LearnGD features. For performance evaluation, speaker verification experiments are conducted on the VoxCeleb2 corpus. Compared to the traditional acoustic feature derived from the magnitude spectrum, the proposed phase-based features reach a 27.8% relative improvement in terms of EER. Furthermore, experimental results on TIMIT phoneme recognition task also demonstrate the effectiveness of our proposed phase-based features.

Wed-E-O-2: Spoken Language Processing II

Room C, 19:00–21:00, Wednesday 1 September 2021 Chairs: Shantipriya Parida and Sebastian Möller

Impact of Encoding and Segmentation Strategies on End-to-End Simultaneous Speech Translation

Ha Nguyen¹, Yannick Estève², Laurent Besacier¹; ¹LIG (UMR 5217), France; ²LIA (EA 4128), France Wed-E-O-2-1, Time: 19:00

Boosted by the simultaneous translation shared task at IWSLT 2020, promising end-to-end online speech translation approaches were recently proposed. They consist in incrementally encoding a speech input (in a source language) and decoding the corresponding text (in a target language) with the best possible trade-off between latency and translation quality. This paper investigates two key aspects of end-to-end simultaneous speech translation: (a) how to encode efficiently the continuous speech flow, and (b) how to segment the speech flow in order to alternate optimally between reading (R: encoding input) and writing (W: decoding output) operations. We extend our previously proposed end-to-end online decoding strategy and show that while replacing BLSTM by ULSTM encoding degrades performance in offline mode, it actually improves both efficiency and performance in online mode. We also measure the impact of different methods to segment the speech signal (using fixed interval boundaries, oracle word boundaries or randomly set boundaries) and show that our best end-to-end online decoding strategy is surprisingly the one that alternates R/W operations on fixed size blocks on our English-German speech translation setup.

Lost in Interpreting: Speech Translation from Source or Interpreter?

Dominik Macháček, Matúš Žilinec, Ondřej Bojar; Charles University, Czechia

Wed-E-O-2-2, Time: 19:20

Interpreters facilitate multi-lingual meetings but the affordable set of languages is often smaller than what is needed. Automatic simultaneous speech translation can extend the set of provided languages. We investigate if such an automatic system should rather follow the original speaker, or an interpreter to achieve better translation quality at the cost of increased delay.

To answer the question, we release Europarl Simultaneous Interpreting Corpus (ESIC), 10 hours of recordings and transcripts of European Parliament speeches in English, with simultaneous interpreting into Czech and German. We evaluate quality and latency of speaker-based and interpreter-based spoken translation systems from English to

Czech. We study the differences in implicit simplification and summarization of the human interpreter compared to a machine translation system trained to shorten the output to some extent. Finally, we perform human evaluation to measure information loss of each of these approaches.

Active Speaker Detection as a Multi-Objective Optimization with Uncertainty-Based Multimodal Fusion

Baptiste Pouthier¹, Laurent Pilati¹, Leela K. Gudupudi¹, Charles Bouveyron², Frederic Precioso²; ¹NXP Semiconductors, France; ²I3S (UMR 7271), France Wed-E-O-2-3, Time: 19:40

It is now well established from a variety of studies that there is a significant benefit from combining video and audio data in detecting active speakers. However, either of the modalities can potentially mislead audiovisual fusion by inducing unreliable or deceptive information. This paper outlines active speaker detection as a multi-objective learning problem to leverage best of each modalities using a novel self-attention, uncertainty-based multimodal fusion scheme. Results obtained show that the proposed multi-objective learning architecture outperforms traditional approaches in improving both mAP and AUC scores. We further demonstrate that our fusion strategy surpasses, in active speaker detection, other modality fusion methods reported in various disciplines. We finally show that the proposed method significantly improves the state-of-the-art on the AVA-ActiveSpeaker dataset.

It's Not What You Said, it's How You Said it: Discriminative Perception of Speech as a Multichannel Communication System

Sarenne Wallbridge, Peter Bell, Catherine Lai; University of Edinburgh, UK

Wed-E-O-2-4, Time: 20:00

People convey information extremely effectively through spoken interaction using multiple channels of information transmission: the lexical channel of what is said, and the non-lexical channel of how it is said. We propose studying human perception of spoken communication as a means to better understand how information is encoded across these channels, focusing on the question What characteristics of communicative context affect listener's expectations of speech?. To investigate this, we present a novel behavioural task testing whether listeners can discriminate between the true utterance in a dialogue and utterances sampled from other contexts with the same lexical content. We characterize how perception and subsequent discriminative capability - is affected by different degrees of additional contextual information across both the lexical and non-lexical channel of speech. Results demonstrate that people can effectively discriminate between different prosodic realisations, that non-lexical context is informative, and that this channel provides more salient information than the lexical channel, highlighting the importance of the non-lexical channel in spoken interaction.

Wed-E-O-3: Speech and Audio Analysis

Room D, 19:00-21:00, Wednesday 1 September 2021 Chairs: Douglas O'Shaughnessy and Hugo Van hamme

Extending the Fullband E-Model Towards Background Noise, Bursty Packet Loss, and Conversational Degradations

Thilo Michael, Gabriel Mittag, Andreas Bütow, Sebastian Möller; Technische Universität Berlin, Germany Wed-E-O-3-1, Time: 19:00

Quality engineering of speech communication services in the full speech transmission band (0-20,000 Hz) is facilitated by the fullband E-model, a planning tool that predicts overall quality on the basis of parameters describing the setting of the service. We presented a first version of this model at Interspeech 2019, which has since then been standardized by the International Telecommunication Union in ITU-T Rec. G.107.2. Whereas that model was limited to predict the effects of speech codecs, random packet loss, and transmission delay, more realistic settings such as ambient background noise, bursty packet loss, as well as interactive conversational degradations could not be predicted. Based on the results of two new listening-only and conversational tests, we present an approach to extend the E-model to better predict these effects in the present paper. The results show that background noise effects at both sending and receiving side can be predicted well, whereas bursty packet loss predictions still have some limitations which result from the available database. Finally, approaches from conversational analysis help to better predict the effects of delay on conversational quality.

ORCA-SLANG: An Automatic Multi-Stage Semi-Supervised Deep Learning Framework for Large-Scale Killer Whale Call Type Identification

*Christian Bergler*¹, *Manuel Schmitt*¹, *Andreas Maier*¹, *Helena Symonds*², *Paul Spong*², *Steven R. Ness*³, *George Tzanetakis*³, *Elmar Nöth*¹; ¹*FAU Erlangen-Nürnberg, Germany*; ²*OrcaLab, Canada*; ³*University of Victoria, Canada*

Wed-E-O-3-2, Time: 19:20

Identification of animal-specific vocalization patterns is an imperative requirement to decode animal communication. In bioacoustics, passive acoustic recording setups are increasingly deployed to acquire large-scale datasets. Previous knowledge about established animal-specific call types is usually present due to historically conducted research. However, time- and human-resource constraints, combined with a lack of available machine-based approaches, only allow manual analysis of comparatively small data corpora and strongly distort the actual data representation and information value. Such data limitations cause restrictions in terms of identifying existing population-, group-, and individual-specific call types, subcategories, as well as unseen vocalization patterns. Thus, machine learning forms the basis for animal-specific call type recognition, to facilitate more profound insights into communication. The current study is the first fusing task-specific neural networks to develop a fully automated, multi-stage, deep-learning-based framework, entitled ORCA-SLANG, performing semi-supervised call type identification in one of the largest animal-specific bioacoustic archives the Orchive. Orca/noise segmentation, denoising, and subsequent feature learning provide robust representations for semi-supervised clustering/classification. This results in a machine-annotated call type data repository containing 235,369 unique calls.

Audiovisual Transfer Learning for Audio Tagging and Sound Event Detection

Wim Boes, Hugo Van hamme; KU Leuven, Belgium Wed-E-O-3-3, Time: 19:40

We study the merit of transfer learning for two sound recognition problems, i.e., audio tagging and sound event detection. Employing feature fusion, we adapt a baseline system utilizing only spectral acoustic inputs to also make use of pretrained auditory and visual features, extracted from networks built for different tasks and trained with external data.

We perform experiments with these modified models on an audiovisual multi-label data set, of which the training partition contains a large number of unlabeled samples and a smaller amount of clips with weak annotations, indicating the clip-level presence of 10 sound categories without specifying the temporal boundaries of the active auditory events.

For clip-based audio tagging, this transfer learning method grants marked improvements. Addition of the visual modality on top of audio also proves to be advantageous in this context.

When it comes to generating transcriptions of audio recordings, the benefit of pretrained features depends on the requested temporal resolution: for coarse-grained sound event detection, their utility remains notable. But when more fine-grained predictions are required, performance gains are strongly reduced due to a mismatch between the problem at hand and the goals of the models from which the pretrained vectors were obtained.

Non-Intrusive Speech Quality Assessment with Transfer Learning and Subject-Specific Scaling

Natalia Nessler¹, Milos Cernak², Paolo Prandoni¹, Pablo Mainar²; ¹EPFL, Switzerland; ²Logitech, Switzerland Wed-E-O-3-4, Time: 20:00

In communication systems, it is crucial to estimate the perceived quality of audio and speech. The industrial standards for many vears have been PESO, 3QUEST, and POLOA, which are intrusive methods. This restricts the possibilities of using these metrics in real-world conditions, where we might not have access to the clean reference signal. In this work, we develop a new non-intrusive metric based on crowd-sourced data. We build a new speech dataset by combining publicly available speech, noises, and reverberations. Then we follow the ITU P.808 recommendation to label the dataset with mean opinion scores (MOS). Finally, we train a deep neural network to estimate the MOS from the speech data in a non-intrusive way. We propose two novelties in our work. First, we explore transfer learning by pre-training a model using a larger set of POLQA scores and finetuning with the smaller (and thus cheaper) human-labeled Secondly, we perform a subject-specific scaling in the MOS scores to adjust for their different subjective scales. Our model yields better accuracy than PESQ, POLQA, and other non-intrusive methods when evaluated on the independent VCTK test set. We also report misleading POLQA scores for reverberant speech.

Audio Retrieval with Natural Language Queries

Andreea-Maria Oncescu¹, A. Sophia Koepke², João F. Henriques¹, Zeynep Akata², Samuel Albanie¹; ¹University of Oxford, UK; ²Universität Tübingen, Germany Wed-E-O-3-5, Time: 20:20

We consider the task of retrieving audio using free-form natural language queries. To study this problem, which has received limited attention in the existing literature, we introduce challenging new benchmarks for text-based audio retrieval using text annotations

NOTES

sourced from the AUDIOCAPS and CLOTHO datasets. We then employ these benchmarks to establish baselines for cross-modal audio retrieval, where we demonstrate the benefits of pre-training on diverse audio tasks. We hope that our benchmarks will inspire further research into cross-modal text-based audio retrieval with free-form text queries.

Wed-E-V-1: Cross/Multi-Lingual and Code-Switched ASR

19:00–21:00, Wednesday 1 September 2021 Chairs: Brian Kingsbury and Mikko Kurimo

Bootstrap an End-to-End ASR System by Multilingual Training, Transfer Learning, Text-to-Text Mapping and Synthetic Audio

Manuel Giollo¹, Deniz Gunceler², Yulan Liu³, Daniel Willett²; ¹Amazon, Italy; ²Amazon, Germany; ³Amazon, UK

Wed-E-V-1-1, Time: 19:00

Bootstrapping speech recognition on limited data resources has been an area of active research for long. The recent transition to all-neural models and end-to-end (E2E) training brought along particular challenges as these models are known to be data hungry, but also came with opportunities around language-agnostic representations derived from multilingual data as well as shared word-piece output representations across languages that share script and roots. We investigate here the effectiveness of different strategies to bootstrap an RNN-Transducer (RNN-T) based automatic speech recognition (ASR) system in the low resource regime, while exploiting the abundant resources available in other languages as well as the synthetic audio from a text-to-speech (TTS) engine. Our experiments demonstrate that transfer learning from a multilingual model, using a post-ASR text-to-text mapping and synthetic audio deliver additive improvements, allowing us to bootstrap a model for a new language with a fraction of the data that would otherwise be needed. The best system achieved a 46% relative word error rate (WER) reduction compared to the monolingual baseline, among which 25% relative WER improvement is attributed to the post-ASR text-to-text mappings and the TTS synthetic data.

Efficient Weight Factorization for Multilingual Speech Recognition

Ngoc-Quan Pham, Tuan-Nam Nguyen, Sebastian Stüker, Alex Waibel; KIT, Germany

Wed-E-V-1-2, Time: 19:00

End-to-end multilingual speech recognition involves using a single model training on a compositional speech corpus including many languages, resulting in a single neural network to handle transcribing different languages. Due to the fact that each language in the training data has different characteristics, the shared network may struggle to optimize for all various languages simultaneously. In this paper we propose a novel multilingual architecture that targets the core operation in neural networks: linear transformation functions. The key idea of the method is to assign fast weight matrices for each language by decomposing each weight matrix into a shared component and a language dependent component. The latter is then factorized into vectors using rank-1 assumptions to reduce the number of parameters per language. This efficient factorization scheme is proved to be effective in two multilingual settings with 7 and 27 languages, reducing the word error rates by 26% and 27% rel. for two popular architectures LSTM and Transformer, respectively.

Unsupervised Cross-Lingual Representation Learning for Speech Recognition

Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, Michael Auli; Facebook, USA Wed-E-V-1-3, Time: 19:00

This paper presents XLSR which learns cross-lingual speech representations by pretraining a single model from the raw waveform of speech in multiple languages. We build on wav2vec 2.0 which is trained by solving a contrastive task over masked latent speech representations and jointly learns a quantization of the latents shared across languages. The resulting model is fine-tuned on labeled data and experiments show that cross-lingual pretraining significantly outperforms monolingual pretraining. On the CommonVoice benchmark, XLSR shows a relative phoneme error rate reduction of 72% compared to the best known results. On BABEL. our approach improves word error rate by 16% relative compared to a comparable system. Our approach enables a single multilingual speech recognition model which is competitive to strong individual We hope to catalyze research in low-resource speech models. understanding by releasing XLSR-53, a large model pretrained in 53 languages.

Language and Speaker-Independent Feature Transformation for End-to-End Multilingual Speech Recognition

Tomoaki Hayakawa¹, Chee Siang Leow¹, Akio Kobayashi², Takehito Utsuro³, Hiromitsu Nishizaki¹; ¹University of Yamanashi, Japan; ²NTUT, Japan; ³University of Tsukuba, Japan Wed-E-V-1-4, Time: 19:00

This paper proposes a method to improve the performance of multilingual automatic speech recognition (ASR) systems through language- and speaker-independent feature transformation in a framework of end-to-end (E2E) ASR. Specifically, we propose a multi-task training method that combines a language recognizer and a speaker recognizer with an E2E ASR system based on connectionist temporal classification (CTC) loss functions. We introduce the language and speaker recognition sub-tasks into the E2E ASR network and introduce a gradient reversal layer (GRL) for each sub-task to achieve language and speaker-independent feature transformation. The evaluation results of the proposed method in the multilingual ASR system in six sorts of languages show that the proposed method achieves higher accuracy than the ASR models for each language by introducing multi-tasking and GRL.

Using Large Self-Supervised Models for Low-Resource Speech Recognition

Krishna D. N., Pinyi Wang, Bruno Bozza; Freshworks, India

Wed-E-V-1-5, Time: 19:00

Recently, self-supervised pre-training has shown significant improvements in many areas of machine learning, including speech and NLP. The self-supervised models are trained on a large amount of unlabelled data to learn higher-level representations for downstream tasks. In this work, we investigate the effectiveness of many self-supervised pre-trained models for the low-resource speech recognition task. We adopt pre-trained wav2vec2.0 [1] models for the speech recognition task for three Indian languages Telugu, Tamil, and Gujarati. We examine both English and multilingual pre-trained models. Our experiments show that fine-tuning the multilingual pre-trained model obtains an average relative reduction in WER of 2.88% compared to the previous state-of-the-art supervised method.

NOTES

We carefully analyze the generalization capability of multilingual pre-trained models for both seen and unseen languages. We also show that fine-tuning with only 25% of the training data gives competitive WER to the previous best methods.

Dual Script E2E Framework for Multilingual and Code-Switching ASR

Mari Ganesh Kumar, Jom Kuriakose, Anand Thyagachandran, Arun Kumar A., Ashish Seth, Lodagala V.S.V. Durga Prasad, Saish Jaiswal, Anusha Prakash, Hema A. Murthy; IIT Madras, India Wed-E-V-1-6, Time: 19:00

India is home to multiple languages, and training automatic speech recognition (ASR) systems is challenging. Over time, each language has adopted words from other languages, such as English, leading to code-mixing. Most Indian languages also have their own unique scripts, which poses a major limitation in training multilingual and code-switching ASR systems.

Inspired by results in text-to-speech synthesis, in this paper, we use an in-house rule-based phoneme-level common label set (CLS) representation to train multilingual and code-switching ASR for Indian languages. We propose two end-to-end (E2E) ASR systems. In the first system, the E2E model is trained on the CLS representation, and we use a novel data-driven backend to recover the native language script. In the second system, we propose a modification to the E2E model, wherein the CLS representation and the native language characters are used simultaneously for training. We show our results on the multilingual and code-switching (MUCS) ASR challenge 2021. Our best results achieve $\approx 6\%$ and 5% improvement in word error rate over the baseline system for the multilingual and code-switching tasks, respectively, on the challenge development data.

MUCS 2021: Multilingual and Code-Switching ASR Challenges for Low Resource Indian Languages

Anuj Diwan¹, Rakesh Vaideeswaran², Sanket Shah³, Ankita Singh¹, Srinivasa Raghavan⁴, Shreya Khare⁵, Vinit Unni¹, Saurabh Vyas⁴, Akash Rajpuria⁴, Chiranjeevi Yarra⁶, Ashish Mittal⁵, Prasanta Kumar Ghosh², Preethi Jyothi¹, Kalika Bali³, Vivek Seshadri³, Sunayana Sitaram³, Samarth Bharadwaj⁵, Jai Nanavati⁴, Raoul Nanavati⁴, Karthik Sankaranarayanan⁵; ¹IIT Bombay, India; ²Indian Institute of Science, India; ³Microsoft, India; ⁴Navana Tech, India; ⁵IBM, India; ⁶IIIT Hyderabad, India Wed-E-V-1-7, Time: 19:00

Recently, there is an increasing interest in multilingual automatic speech recognition (ASR) where a speech recognition system caters to multiple low resource languages by taking advantage of low amounts of labelled corpora in multiple languages. With multilingualism becoming common in today's world, there has been increasing interest in code-switching ASR as well. In code-switching, multiple languages are freely interchanged within a single sentence or between sentences. The success of low-resource multilingual and code-switching (MUCS) ASR often depends on the variety of languages in terms of their acoustics, linguistic characteristics as well as the amount of data available and how these are carefully considered in building the ASR system. In this MUCS 2021 challenge, we would like to focus on building MUCS ASR systems through two different subtasks related to a total of seven Indian languages, namely Hindi, Marathi, Odia, Tamil, Telugu, Gujarati and Bengali. For this purpose, we provide a total of ~600 hours of transcribed speech data, comprising train and test sets, in these languages, including two code-switched language pairs, Hindi-English and Bengali-English. We also provide baseline recipes for both the subtasks with 30.73% and 32.45% word error rate on the MUCS test sets, respectively.

Adapt-and-Adjust: Overcoming the Long-Tail Problem of Multilingual Speech Recognition

*Genta Indra Winata*¹, *Guangsen Wang*², *Caiming Xiong*³, *Steven Hoi*²; ¹*HKUST, China*; ²*Salesforce, Singapore*; ³*Salesforce, USA* Wed-E-V-1-8, Time: 19:00

One crucial challenge of real-world multilingual speech recognition is the long-tailed distribution problem, where some resource-rich languages like English have abundant training data, but a long tail of low-resource languages have varying amounts of limited training data. To overcome the long-tail problem, in this paper, we propose Adapt-and-Adjust (A2), a transformer-based multi-task learning framework for end-to-end multilingual speech recognition. The A2 framework overcomes the long-tail problem via three techniques: (1) exploiting a pretrained multilingual language model to improve the performance of low-resource languages; (2) proposing dual adapters consisting of both language-specific and language-agnostic adaptation with minimal additional parameters; and (3) overcoming the class imbalance, either by imposing class priors in the loss during training or adjusting the logits of the softmax output during inference. Extensive experiments on the CommonVoice corpus show that A2 significantly outperforms conventional approaches.

SRI-B End-to-End System for Multilingual and Code-Switching ASR Challenges for Low Resource Indian Languages

Hardik Sailor, Kiran Praveen T., Vikas Agrawal, Abhinav Jain, Abhishek Pandey; Samsung, India Wed-E-V-1-9, Time: 19:00

This paper describes SRI-B's end-to-end Automated Speech Recognition (ASR) system proposed for the subtask-1 on multilingual ASR challenges for Indian languages. Our end-to-end (E2E) ASR model is based on the transformer architecture trained by jointly minimizing Connectionist Temporal Classification (CTC) & Cross-Entropy (CE) losses. A conventional multilingual model which is trained by pooling data from multiple languages helps in terms of generalization, but it comes at the expense of performance degradation compared to their monolingual counterparts. In our experiments, a multilingual model is trained by conditioning the input features using a language-specific embedding vector. These language-specific embedding vectors are obtained by training a language classifier using an attention-based transformer architecture, and then considering its bottleneck features as language identification (LID) embeddings. We further adapt the multilingual system with language specific data to reduce the degradation on specific languages. We propose a novel hypothesis elimination strategy based on LID scores and length-normalized probabilities that optimally select the model from the pool of available models. The experimental results show that the proposed multilingual training and hypothesis elimination strategy gives an average 3.02% of relative word error recognition (WER) improvement for the blind set over the challenge hybrid ASR baseline system.

Hierarchical Phone Recognition with Compositional Phonetics

Xinjian Li, Juncheng Li, Florian Metze, Alan W. Black; Carnegie Mellon University, USA Wed-E-V-1-10, Time: 19:00

There is growing interest in building phone recognition systems for low-resource languages as the majority of languages do not have any writing systems. Phone recognition systems proposed so far typically derive their phone inventory from the training languages, therefore the derived inventory could only cover a limited number of phones existing in the world. It fails to recognize unseen phones in low-resource or zero-resource languages. In this work, we tackle this problem with a hierarchical model, in which we explicitly model three different entities in a hierarchical manner: phoneme, phone, and phonological articulatory attributes. In particular, we decompose phones into articulatory attributes and compute the phone embedding from the attribute embedding. The model would first predict the distribution over the phones using their embeddings, next, the language-independent phones are aggregated to the language-dependent phonemes and then optimized by the CTC loss. This compositional approach enables us to recognize phones even they do not appear in the training set. We evaluate our model on 47 unseen languages and find the proposed model outperforms baselines by 13.1% PER.

Towards One Model to Rule All: Multilingual Strategy for Dialectal Code-Switching Arabic ASR

Shammur Absar Chowdhury, Amir Hussein, Ahmed Abdelali, Ahmed Ali; HBKU, Qatar Wed-E-V-1-11, Time: 19:00

With the advent of globalization, there is an increasing demand for multilingual automatic speech recognition (ASR), handling language and dialectal variation of spoken content. Recent studies show its efficacy over monolingual systems. In this study, we design a large multilingual end-to-end ASR using self-attention based conformer architecture. We trained the system using Arabic (Ar), English (En) and French (Fr) languages. We evaluate the system performance handling: (i) monolingual (Ar, En and Fr); (ii) multi-dialectal (Modern Standard Arabic, along with dialectal variation such as Egyptian and Moroccan); (iii) code-switching - cross-lingual (Ar-En/Fr) and dialectal (MSA-Egyptian dialect) test cases, and compare with current state-of-the-art systems. Furthermore, we investigate the influence of different embedding/character representations including character vs word-piece; shared vs distinct input symbol per language. Our findings demonstrate the strength of such a model by outperforming state-of-the-art monolingual dialectal Arabic and code-switching Arabic ASR.

Differentiable Allophone Graphs for Language-Universal Speech Recognition

Brian Yan, Siddharth Dalmia, David R. Mortensen, Florian Metze, Shinji Watanabe; Carnegie Mellon University, USA Wed-E-V-1-12, Time: 19:00 Building language-universal speech recognition systems

Building language-universal speech recognition systems entails producing phonological units of spoken sound that can be shared across languages. While speech annotations at the language-specific phoneme or surface levels are readily available, annotations at a universal phone level are relatively rare and difficult to produce. In this work, we present a general framework to derive phone-level supervision from only phonemic transcriptions and phone-to-phoneme mappings with *learnable* weights represented using weighted finite-state transducers, which we call *differentiable allophone*

graphs. By training multilingually, we build a universal phone-based speech recognition model with interpretable probabilistic phone-to-phoneme mappings for each language. These phone-based systems with learned allophone graphs can be used by linguists to document new languages, build phone-based lexicons that capture rich pronunciation variations, and re-evaluate the allophone mappings of seen language. We demonstrate the aforementioned benefits of our proposed framework with a system trained on 7 diverse languages.

Wed-E-V-2: Health and Affect II

19:00–21:00, Wednesday 1 September 2021 Chairs: Khiet Truong and Nicholas Cummins

Automatic Speech Recognition Systems Errors for Objective Sleepiness Detection Through Voice

*Vincent P. Martin*¹, *Jean-Luc Rouas*¹, *Florian Boyer*¹, *Pierre Philip*²; ¹*LaBRI (UMR 5800), France;* ²*SANPSY (USR 3413), France* Wed-E-V-2-1, Time: 19:00

Chronic sleepiness, and specifically Excessive Daytime Sleepiness (EDS), impacts everyday life and increases the risks of accidents. Compared with traditional measures (EEG), the detection of objective EDS through voice benefits from its ease to be implemented in ecological conditions and to be sober in terms of data processing and costs. Contrary to previous works focusing on short-term sleepiness estimation, this study focuses on long-term sleepiness detection through voice. Using the Multiple Sleep Latency Test corpus, this study introduces new features based on Automatic Speech Recognition systems errors, in an attempt to replace hand-labeled reading mistakes features. We also introduce a selection feature pipeline inspired by clinical validation practices allowing ASR features to perform on par with the state-of-the-art systems on short-term sleepiness detection through voice (73.2% of UAR). Moreover, we give insights on the decision process during classification and the specificity of the system regarding the threshold delimiting the two sleepiness classes, Sleepy and Non-Sleepy.

Robust Laughter Detection in Noisy Environments

Jon Gillick, Wesley Deng, Kimiko Ryokai, David Bamman; University of California at Berkeley, USA Wed-E-V-2-2, Time: 19:00

We investigate the problem of automatically identifying and extracting laughter from audio files in noisy environments. We conduct an empirical evaluation of several machine learning models using audio data of varying sound quality, finding that while previously published methods work relatively well in controlled environments, performance drops precipitously in real-world settings with background noise. In the process, we contribute a new dataset of laughter annotations on top of the existing AudioSet corpus, with precise segmentations for the start and end points of each laugh, and we present a new approach to laughter detection that performs comparatively well in uncontrolled environments. We discuss the utility of our approach as well as the importance of understanding the variability of model performance in a range of real-world testing environments.

Impact of Emotional State on Estimation of Willingness to Buy from Advertising Speech

Mizuki Nagano, Yusuke Ijima, Sadao Hiroya; NTT, Japan Wed-E-V-2-3, Time: 19:00

The characteristics of a speaker's voice can affect the perceived impression or behavior of the listener. Previous studies of consumer behavior have shown that this can be well explained by the emotion-mediated behavior model. However, few studies of the emotion-mediated behavior model have used advertising speech. In this paper, we examine whether the stimulus-organism-response theory using emotional state can explain willingness to buy from advertising speech stimulus. The subjects listened to speech with modified speech features (mean F0, speech rate, spectral tilt, or standard deviation of F0) and rated their willingness to buy the products advertised in the speech and their own perceived emotions (pleasure, arousal, dominance). We found that the emotions partially mediate the influence of speech features on the willingness to buy. These results will be useful for developing a method of speech synthesis to increase people's willingness to buy.

Stacked Recurrent Neural Networks for Speech-Based Inference of Attachment Condition in School Age Children

Huda Alsofyani, Alessandro Vinciarelli; University of Glasgow, UK

Wed-E-V-2-4, Time: 19:00

In Attachment Theory, children that have a positive perception of their parents are said to be secure, while the others are said to be insecure. Once adult, unless identified and supported early enough, insecure children have higher chances to experience major issues (e.g., suicidal tendencies and antisocial behavior). For this reason, this article proposes a speech-based automatic approach for the recognition of attachment in school-age children. The experiments are based on stacked RNNs and have involved 104 children of age between 5 and 9. The accuracy is up to 68.9% (F1 59.6%), meaning that the approach makes the right decision two times out of three, on average. To the best of our knowledge, this is the first work aimed at inferring attachment from speech in school-age children.

Language or Paralanguage, This is the Problem: Comparing Depressed and Non-Depressed Speakers Through the Analysis of Gated Multimodal Units

Nujud Aloshban¹, Anna Esposito², Alessandro Vinciarelli¹; ¹University of Glasgow, UK; ²Università della Campania "Luigi Vanvitelli", Italy Wed-E-V-2-5, Time: 19:00

Speech-based depression detection has attracted significant attention over the last years. A debated problem is whether it is better to use language (what people say), paralanguage (how they say it) or a combination of the two. This article addresses the question through the analysis of a Gated Multimodal Unit trained to weight modalities according to how effectively they account for the condition of a speaker (depressed or non-depressed). The experiments involved 29 individuals diagnosed with depression and 30 non-depressed participants. Besides an accuracy of 83.0% (F1 score 80.0%), the results show that the Gated Multimodal Unit tends to give more weight to paralanguage. However, the relative contribution of language tends to be higher, to a statistically significant extent, in the case of non-depressed speakers.

Emotion Carrier Recognition from Personal Narratives

Aniruddha Tammewar, Alessandra Cervone, Giuseppe Riccardi; Università di Trento, Italy Wed-E-V-2-6. Time: 19:00

Personal Narratives (PN) — recollections of facts, events, and thoughts from one's own experience — are often used in everyday conversations. So far, PNs have mainly been explored for tasks such as valence prediction or emotion classification (e.g. *happy, sad*).

However, these tasks might overlook more fine-grained information that could prove to be relevant for understanding PNs. In this work, we propose a novel task for Narrative Understanding: Emotion Carrier Recognition (ECR). Emotion carriers, the text fragments that carry the emotions of the narrator (e.g. *loss of a grandpa, high school reunion*), provide a fine-grained description of the emotion state. We explore the task of ECR in a corpus of PNs manually annotated with emotion carriers and investigate different machine learning models for the task. We propose evaluation strategies for ECR including metrics that can be appropriate for different tasks.

Non-Verbal Vocalisation and Laughter Detection Using Sequence-to-Sequence Models and Multi-Label Training

Scott Condron, Georgia Clarke, Anita Klementiev, Daniela Morse-Kopp, Jack Parry, Dimitri Palaz; Speech Graphics, UK

Wed-E-V-2-7, Time: 19:00

Non-verbal vocalisations (NVVs) such as laughter are an important part of communication in social interactions and carry important information about a speaker's state or intention. There remains no clear definition of NVVs and there is no clearly defined protocol for transcribing or detecting NVVs. As such, the standard approach has been to focus on detecting a single NVV such as laughter and map all other NVVs to an "other" class. In this paper we hypothesise that for this task such an approach hurts performance, and that giving more information by using more classes is beneficial. To address this, we present studies using sequence-to-sequence deep neural networks where we include multiple NVV classes rather than mapping them to "other" and allow more than one label per sample. We show that this approach yields better performance than the standard approach on NVV detection. We also evaluate the same model on laughter detection using frame-based and utterance-based metrics and show that the proposed approach yields state-of-the-art performance on the ICSI corpus.

TDCA-Net: Time-Domain Channel Attention Network for Depression Detection

Cong Cai, Mingyue Niu, Bin Liu, Jianhua Tao, Xuefei Liu; CAS, China

Wed-E-V-2-8, Time: 19:00

Depression is a psychiatric disorder and has many adverse effects on our society. Some studies have shown that speech signals are closely related to emotion and stress, and many speech-based automatic depression detection methods have been proposed. However, previous work is based on spectrogram or hand-crafted features, which may lose some useful information related to depression patterns. And there is no evidence that the filter bank designed from perceptual evidence is optimal for depression detection. In order to learn the more discriminative feature representation related to depression, we propose an end-to-end time-domain channel attention network (TDCA-Net) for depression detection. The TDCA-Net directly models time-domain speech signals based on dilated convolution block, which can increase the receptive field exponentially and aggregate multiscale contextual information associated with depression. Besides, we employ the efficient channel attention (ECA) module to model dependencies of channels and improve the sensitivity of the model to information related to depression. Experimental results on the AVEC2013 and the AVEC2014 datasets illustrate the effectiveness of our method.

Visual Speech for Obstructive Sleep Apnea Detection

Catarina Botelho¹, Alberto Abad¹, Tanja Schultz², Isabel Trancoso¹; ¹INESC-ID Lisboa, Portugal; ²Universität Bremen, Germany Wed-E-V-2-9, Time: 19:00

Obstructive sleep apnea (OSA) affects almost one billion people worldwide and limits peoples' quality of life substantially. Furthermore, it is responsible for significant morbidity and mortality associated with hypertension, cardiovascular diseases, work and traffic accidents. Thus, the early detection of OSA can save lives. In our previous work we used speech as biomarker for automatic OSA detection. More recently, we leveraged the fact that OSA patients have anatomical and functional abnormalities of the upper airway and an altered craniofacial morphology, and therefore explore information from facial images for OSA detection. In this work, we propose to combine speech and facial image information to detect OSA from YouTube vlogs. This in-the-wild data poses an inexpensive alternative to standard data collected for medical applications, which is often scarce, imbalanced and costly to acquire. Besides speech and facial images, we propose to include visual speech as a third modality, inspired by the emerging field of silent computational paralinguistics. We hypothesize that embeddings trained from lip reading integrate information on the craniofacial structure, on speech articulation and breathing patterns, thus containing relevant cues for OSA detection. Fusion of the three modalities achieves an accuracy of 82.5% at the speaker level.

Analysis of Contextual Voice Changes in Remote Meetings

Hector A. Cordourier Maruri¹, Sinem Aslan², Georg Stemmer³, Nese Alyuz², Lama Nachman²; ¹Intel, Mexico; ²Intel, USA; ³Intel, Germany Wed-E-V-2-10, Time: 19:00

People participating in remote meetings in open spaces might choose to speak with a restrained voice due to concerns around privacy or disturbing others. These contextual voice changes might impact the quality of communications. To investigate how people adjust their voices in certain situations, we performed an exploratory data collection study with 41 participants in 18 simulated remote meetings. A scenario was provided to the participants to naturally trigger contextual voice changes. We collected multi-modal data from the participants including in-situ labels for the voice quality. We implemented content analysis, t-test, and linear regression to analyze the multi-modal data. Results showed that the participants primarily preferred to use soft voice over whispered voice to avoid being overheard during the meetings. Speaking softly was often sufficient to successfully conceal private conversations, while using whispered voice had only a negative impact on the intelligibility. Overall, we found that participants perceived soft voice as less pleasant to listen to than normal voice during meetings and discovered factors related to speaker demographics and meeting context that impacted the concealing behavior (soft or whispered). For our future research, we will expand to different scenarios and consider the impact of audio feedback on voice concealing.

Speech Based Depression Severity Level Classification Using a Multi-Stage Dilated CNN-LSTM Model

Nadee Seneviratne, Carol Espy-Wilson; University of Maryland at College Park, USA Wed-E-V-2-11, Time: 19:00

Speech based depression classification has gained immense popularity over the recent years. However, most of the classification studies

have focused on binary classification to distinguish depressed subjects from non-depressed subjects. In this paper, we formulate the depression classification task as a severity level classification problem to provide more granularity to the classification outcomes. We use articulatory coordination features (ACFs) developed to capture the changes of neuromotor coordination that happens as a result of psychomotor slowing, a necessary feature of Major Depressive Disorder. The ACFs derived from the vocal tract variables (TVs) are used to train a dilated Convolutional Neural Network based depression classification model to obtain segment-level predictions. Then, we propose a Recurrent Neural Network based approach to obtain session-level predictions from segment-level predictions. We show that strengths of the segment-wise classifier are amplified when a session-wise classifier is trained on embeddings obtained from it. The model trained on ACFs derived from TVs show relative improvement of 27.47% in Unweighted Average Recall (UAR) at the session-level classification task, compared to the ACFs derived from Mel Frequency Cepstral Coefficients (MFCCs).

Wed-E-V-3: Neural Network Training Methods for ASR

19:00-21:00, Wednesday 1 September 2021 Chairs: Peter Bell and Suyoun Kim

Multi-Domain Knowledge Distillation via Uncertainty-Matching for End-to-End ASR Models

*Ho-Gyeong Kim*¹, *Min-Joong Lee*¹, *Hoshik Lee*¹, *Tae Gyoon Kang*¹, *Jihyun Lee*¹, *Eunho Yang*², *Sung Ju Hwang*²; ¹*Samsung, Korea*; ²*KAIST, Korea* Wed-E-V-3-1, Time: 19:00

Knowledge Distillation basically matches predictive distributions of student and teacher networks to improve performance in an environment with model capacity and/or data constraints. However, it is well known that predictive distribution of neural networks not only tends to be overly confident, but also cannot directly model various factors properly that contribute to uncertainty. Recently, deep learning studies based on uncertainty have been successful in various fields, especially in several computer vision tasks. The prediction probability can implicitly show the information about how confident the network is, however, we can explicitly utilize confidence of the output by modeling the uncertainty of the network. In this paper, we propose a novel knowledge distillation method for automatic speech recognition that directly models and transfers the uncertainty inherent in data observation such as speaker variations or confusing pronunciations. Moreover, we investigate an effect of transferring knowledge more effectively using multiple teachers learned from various domains. Evaluated on WSJ which is the standard benchmark dataset with limited instances, the proposed knowledge distillation method achieves significant improvements over student baseline models.

Learning a Neural Diff for Speech Models

Jonathan Macoskey, Grant P. Strimel, Ariya Rastrow; Amazon, USA

Wed-E-V-3-2, Time: 19:00

As more speech processing applications execute locally on edge devices, a set of resource constraints must be considered. In this work we address one of these constraints, namely over-thenetwork data budgets for transferring models from server to device. We present neural update approaches for release of subsequent speech model generations abiding by a data budget. We detail two architecture-agnostic methods which learn compact representa-

NOTES

tions for transmission to devices. We experimentally validate our techniques with results on two tasks (automatic speech recognition and spoken language understanding) on open source data sets by demonstrating when applied in succession, our budgeted updates outperform comparable model compression baselines by significant margins.

Stochastic Attention Head Removal: A Simple and Effective Method for Improving Transformer Based ASR Models

Shucong Zhang, Erfan Loweimi, Peter Bell, Steve Renals; University of Edinburgh, UK Wed-E-V-3-3, Time: 19:00

Recently, Transformer based models have shown competitive automatic speech recognition (ASR) performance. One key factor in the success of these models is the multi-head attention mechanism. However, for trained models, we have previously observed that many attention matrices are close to diagonal, indicating the redundancy of the corresponding attention heads. We have also found that some architectures with reduced numbers of attention heads have better performance. Since the search for the best structure is time prohibitive, we propose to randomly remove attention heads during training and keep all attention heads at test time, thus the final model is an ensemble of models with different architectures. The proposed method also forces each head independently learn the most useful patterns. We apply the proposed method to train Transformer based and Convolution-augmented Transformer (Conformer) based ASR models. Our method gives consistent performance gains over strong baselines on the Wall Street Journal, AISHELL, Switchboard and AMI datasets. To the best of our knowledge, we have achieved state-of-the-art end-to-end Transformer based model performance on Switchboard and AMI.

Model-Agnostic Fast Adaptive Multi-Objective Balancing Algorithm for Multilingual Automatic Speech Recognition Model Training

Jiabin Xue, Tieran Zheng, Jiqing Han; Harbin Institute of Technology, China

Wed-E-V-3-4, Time: 19:00

This paper regards multilingual automatic speech recognition model training as a multi-objective problem because learning different languages may conflict, necessitating a trade-off. Most previous works on multilingual ASR model training mainly used data sampling to balance the performance of multiple languages but ignore the conflicts between different languages, resulting in an imbalance in multiple languages. The language-specific parameters of the multilingual ASR model are updated by the single language gradients while the update of the shared parameter is jointly determined by the gradient of every language on its shared parameter, namely shared gradient. Therefore, we propose a model-agnostic fast adaptive (MAFA) multi-objective balancing algorithm to balance multiple languages by avoiding the mutual interferences between their shared gradients. In the algorithm, based on the decrease in the training loss, we dynamically normalize the shared gradient magnitudes representing the speed of learning to balance the learning speed. To evenly learn multiple languages, the language with the worst performance is selected, and a balancing gradient nearest to the normalized gradient of the selected language and positively correlated with other normalized ones is obtained to eliminate the mutual interferences. The model trained by MAFA outperforms the baseline model on the Common Voice corpus.

Towards Lifelong Learning of End-to-End ASR

Heng-Jui Chang, Hung-yi Lee, Lin-shan Lee; National Taiwan University, Taiwan Wed-E-V-3-5, Time: 19:00

Automatic speech recognition (ASR) technologies today are primarily optimized for given datasets; thus, any changes in the application environment (e.g., acoustic conditions or topic domains) may inevitably degrade the performance. We can collect new data describing the new environment and fine-tune the system, but this naturally leads to higher error rates for the earlier datasets, referred to as catastrophic forgetting. The concept of lifelong learning (LLL) aiming to enable a machine to sequentially learn new tasks from new datasets describing the changing real world without forgetting the previously learned knowledge is thus brought to attention. This paper reports, to our knowledge, the first effort to extensively consider and analyze the use of various approaches of LLL in end-to-end (E2E) ASR, including proposing novel methods in saving data for past domains to mitigate the catastrophic forgetting problem. An overall relative reduction of 28.7% in WER was achieved compared to the fine-tuning baseline when sequentially learning on three very different benchmark corpora. This can be the first step toward the highly desired ASR technologies capable of synchronizing with the continuously changing real world.

Self-Adaptive Distillation for Multilingual Speech Recognition: Leveraging Student Independence

Isabel Leal, Neeraj Gaur, Parisa Haghani, Brian Farris, Pedro J. Moreno, Manasa Prasad, Bhuvana Ramabhadran, Yun Zhu; Google, USA Wed-E-V-3-6, Time: 19:00

With a large population of the world speaking more than one language, multilingual automatic speech recognition (ASR) has gained popularity in the recent years. While lower resource languages can benefit from quality improvements in a multilingual ASR system, including unrelated or higher resource languages in the mix often results in performance degradation. In this paper, we propose distilling from multiple teachers, with each language using its best teacher during training, to tackle this problem. We introduce self-adaptive distillation, a novel technique for automatic weighting of the distillation loss that uses the student/ teachers confidences. We analyze the effectiveness of the proposed techniques on two real world use-cases and show that the performance of the multilingual ASR models can be improved by up to 11.5% without any increase in model capacity. Furthermore, we show that when our methods are combined with increase in model capacity, we can achieve quality gains of up to 20.7%.

Regularizing Word Segmentation by Creating Misspellings

Hainan Xu, Kartik Audhkhasi, Yinghui Huang, Jesse Emond, Bhuvana Ramabhadran; Google, USA Wed-E-V-3-7, Time: 19:00

This work focuses on improving subword segmentation algorithms for end-to-end speech recognition models, and makes two major contributions. Firstly, we propose a novel word segmentation algorithm. The algorithm uses the same vocabulary generated by a regular wordpiece model, is easily extensible and supports a variety of regularization techniques in the segmentation space, and outperforms the regular wordpiece model. Secondly, we propose a number of novel regularization methods that introduce randomness into the tokenization algorithm, which bring further improvements in speech recognition accuracy, with relative gains up to 8.4% compared to the original wordpiece model. We analyze the methods and show

NOTES

that our proposed methods are equivalent to a sophisticated form of label smoothing, which performs smoothing based on the prefix structures of subword units. A noteworthy discovery from this work is that creating artificial misspellings in words results in the best performance among all the methods, which could inspire future research for strategies in this area.

Multitask Training with Text Data for End-to-End Speech Recognition

Peidong Wang, Tara N. Sainath, Ron J. Weiss; Google, USA

Wed-E-V-3-8, Time: 19:00

We propose a multitask training method for attention-based end-toend speech recognition models. We regularize the decoder in a listen, attend, and spell model by multitask training it on both audio-text and text-only data. Trained on the 100-hour subset of LibriSpeech, the proposed method, without requiring an additional language model, leads to an 11% relative performance improvement over the baseline and approaches the performance of language model shallow fusion on the test-clean evaluation set. We observe a similar trend on the whole 960-hour LibriSpeech training set. Analyses of different types of errors and sample output sentences demonstrate that the proposed method can incorporate language level information, suggesting its effectiveness in real-world applications.

Emitting Word Timings with HMM-Free End-to-End System in Automatic Speech Recognition

*Xianzhao Chen*¹, *Hao Ni*², *Yi He*², *Kang Wang*², *Zejun Ma*², *Zongxia Xie*¹; ¹*Tianjin University, China*; ²*BvteDance, China*

Wed-E-V-3-9, Time: 19:00

Word timings, which mark the start and end times of each word in ASR results, play an important part in many applications, such as computer assisted language learning. To date, end-to-end (E2E) systems outperform conventional DNN-HMM hybrid systems in ASR accuracy but have challenges to obtain accurate word timings. In this paper, we propose a two-pass method to estimate word timings under an E2E-based LAS modeling framework, which is completely free of using the DNN-HMM ASR system. Specifically, we first employ the LAS system to obtain word-piece transcripts of the input audio, we then compute forced-alignments with a frame-level-based word-piece classifier. In order to make the classifier yield accurate word-piece timing results, we propose a novel objective function to learn the classifier, utilizing the spike timings of the connectionist temporal classification (CTC) model. On Librispeech data, our E2E-based LAS system achieves 2.8%/7.0% WERs, while its word timing (start/end) accuracy are 99.0%/95.3% and 98.6%/93.7% on test-clean and test-other two test sets respectively. Compared with a DNN-HMM hybrid ASR system (here, TDNN), the LAS system is better in ASR performance, and the generated word timings are close to what the TDNN ASR system presents.

Scaling Laws for Acoustic Models

Jasha Droppo, Oguz Elibol; Amazon, USA Wed-E-V-3-10, Time: 19:00

There is a recent trend in machine learning to increase model quality by growing models to sizes previously thought to be unreasonable. Recent work has shown that autoregressive generative models with cross-entropy objective functions exhibit smooth power-law relationships, or scaling laws, that predict model quality from model size, training set size, and the available compute budget. These scaling laws allow one to choose nearly optimal hyper-parameters given constraints on available training data, model parameter count, or training computation budget. In this paper, we demonstrate that acoustic models trained with an auto-predictive coding loss behave as if they are subject to similar scaling laws. We extend previous work to jointly predict loss due to model size, to training set size, and to the inherent "irreducible loss" of the task. We find that the scaling laws accurately match model performance over two orders of magnitude in both model size and training set size, and make predictions about the limits of model performance.

Leveraging Non-Target Language Resources to Improve ASR Performance in a Target Language

Jayadev Billa; University of Southern California, USA Wed-E-V-3-11, Time: 19:00

This paper investigates approaches to improving automatic speech recognition (ASR) performance in a target language using resources in other languages. In particular, we assume that we have untranscribed speech in a different language and a well trained ASR system in yet another language. Concretely, we structure this as a multi-task problem, where the primary task is acoustic model training in the target language, and the secondary task is also acoustic model training but using a synthetic data set. The synthetic data set consists of pseudo transcripts generated by decoding the untranscribed speech using a well trained ASR model. We compare and contrast this with using labeled data sets, i.e. matched audio and human-generated transcripts, and show that our approach compares favorably. In most cases, we see performance improvements, and in some cases, depending on the selection of languages and nature of speech data, performance exceeds that of systems using labeled data sets as the secondary task. When extended to larger sets of data, we show that the mismatched data approach performs similarly to in-language semi-supervised training (SST) when the secondary task pseudo transcripts are generated by ASR models trained on large diverse data sets.

4-Bit Quantization of LSTM-Based Speech Recognition Models

Andrea Fasoli, Chia-Yu Chen, Mauricio Serrano, Xiao Sun, Naigang Wang, Swagath Venkataramani, George Saon, Xiaodong Cui, Brian Kingsbury, Wei Zhang, Zoltán Tüske, Kailash Gopalakrishnan; IBM, USA Wed-E-V-3-12, Time: 19:00

We investigate the impact of aggressive low-precision representations of weights and activations in two families of large LSTM-based architectures for Automatic Speech Recognition (ASR): hybrid Deep Bidirectional LSTM - Hidden Markov Models (DBLSTM-HMMs) and Recurrent Neural Network - Transducers (RNN-Ts). Using a 4-bit integer representation, a naïve quantization approach applied to the LSTM portion of these models results in significant Word Error Rate (WER) degradation. On the other hand, we show that minimal accuracy loss is achievable with an appropriate choice of quantizers and initializations. In particular, we customize quantization schemes depending on the local properties of the network, improving recognition performance while limiting computational time. We demonstrate our solution on the Switchboard (SWB) and CallHome (CH) test sets of the NIST Hub5-2000 evaluation. DBLSTM-HMMs trained with 300 or 2000 hours of SWB data achieves <0.5% and <1% average WER degradation, respectively. On the more challenging RNN-T models, our quantization strategy limits degradation in 4-bit inference to 1.3%.

Unified Autoregressive Modeling for Joint End-to-End Multi-Talker Overlapped Speech Recognition and Speaker Attribute Estimation

Ryo Masumura, Daiki Okamura, Naoki Makishima, Mana Ihori, Akihiko Takashima, Tomohiro Tanaka, Shota Orihashi; NTT, Japan Wed-E-V-3-13, Time: 19:00

In this paper, we present a novel modeling method for single-channel multi-talker overlapped automatic speech recognition (ASR) systems. Fully neural network based end-to-end models have dramatically improved the performance of multi-taker overlapped ASR tasks. One promising approach for end-to-end modeling is autoregressive modeling with serialized output training in which transcriptions of multiple speakers are recursively generated one after another. This enables us to naturally capture relationships between speakers. However, the conventional modeling method cannot explicitly take into account the speaker attributes of individual utterances such as gender and age information. In fact, the performance deteriorates when each speaker is the same gender or is close in age. To address this problem, we propose unified autoregressive modeling for joint end-to-end multi-talker overlapped ASR and speaker attribute estimation. Our key idea is to handle gender and age estimation tasks within the unified autoregressive modeling. In the proposed method, transformer-based autoregressive model recursively generates not only textual tokens but also attribute tokens of each speaker. This enables us to effectively utilize speaker attributes for improving multi-talker overlapped ASR. Experiments on Japanese multi-talker overlapped ASR tasks demonstrate the effectiveness of the proposed method.

Minimum Word Error Rate Training with Language Model Fusion for End-to-End Speech Recognition

Zhong Meng¹, Yu Wu², Naoyuki Kanda¹, Liang Lu¹, Xie Chen¹, Guoli Ye¹, Eric Sun¹, Jinyu Li¹, Yifan Gong¹; ¹Microsoft, USA; ²Microsoft, China Wed-E-V-3-14, Time: 19:00

Integrating external language models (LMs) into end-to-end (E2E) models remains a challenging task for domain-adaptive speech recognition. Recently, internal language model estimation (ILME)-based LM fusion has shown significant word error rate (WER) reduction from Shallow Fusion by subtracting a weighted internal LM score from an interpolation of E2E model and external LM scores during beam search. However, on different test sets, the optimal LM interpolation weights vary over a wide range and have to be tuned extensively on well-matched validation sets. In this work, we perform LM fusion in the minimum WER (MWER) training of an E2E model to obviate the need for LM weights tuning during inference. Besides MWER training with Shallow Fusion (MWER-SF), we propose a novel MWER training with ILME (MWER-ILME) where the ILME-based fusion is conducted to generate N-best hypotheses and their posteriors. Additional gradient is induced when internal LM is engaged in MWER-ILME loss computation. During inference, LM weights pre-determined in MWER training enable robust LM integrations on test sets from different domains. Experimented with 30K-hour trained transformer transducers, MWER-ILME achieves on average 8.8% and 5.8% relative WER reductions from MWER and MWER-SF training, respectively, on 6 different test sets.

Variable Frame Rate Acoustic Models Using Minimum Error Reinforcement Learning

Dongcheng Jiang, Chao Zhang, Philip C. Woodland; University of Cambridge, UK Wed-E-V-3-15, Time: 19:00

Frame selection in automatic speech recognition (ASR) systems can potentially improve the trade-off between speed and accuracy relative to fixed low frame rate methods. In this paper, a sequence training approach based on minimum error and reinforcement learning is proposed for a hybrid ASR system to operate at a variable frame rate. and uses a frame selection controller to predict the number of frames to skip before taking the next inference action. The controller is integrated into the acoustic model in a multi-task training framework as an additional regression task and the controller output can be used for distribution characterisation during reinforcement learning exploration. The reinforcement learning objective minimises a combined measure of the phone error and average frame rate. ASR experiments using British English multi-genre broadcast (MGB3) data show that the proposed approach achieved a smaller frame rate than using a fixed 1/3 low frame rate method and was able to reduce the word error rate relative to both fixed low frame rate and full frame rate systems.

Wed-E-V-4: Prosodic Features and Structure

19:00–21:00, Wednesday 1 September 2021 Chairs: Mariapaola D'Imperio and Maria Paola Bissiri

How f0 and Phrase Position Affect Papuan Malay Word Identification

Constantijn Kaland¹, Matthew Gordon²; ¹Universität zu Köln, Germany; ²University of California at Santa Barbara, USA Wed-E-V-4-1, Time: 19:00

This paper reports a perception experiment on Papuan Malay, an Eastern Indonesian language for which phrase prosody is largely underresearched. While phrase-final f0 movements are the most prominent ones in this language, it remains to be seen to what extent they signal phrase boundaries (demarcating) or whether they contribute to the prosodic prominence of words in that position (highlighting). Crucially, it is unclear whether these functions can actually be teased apart. In an attempt to investigate this issue, a word identification experiment was carried out using manipulated and original f0 word contours in phrase-medial and phrase-final positions. Results indicate that Papuan Malay listeners recognize words faster in phrase-final position, although the shape of the f0 movement did not significantly affect response latencies. The outcomes are discussed in a typological perspective, with particular attention to Trade Malay languages.

On the Feasibility of the Danish Model of Intonational Transcription: Phonetic Evidence from Jutlandic Danish

Anna Bothe Jespersen¹, Pavel Šturm², Míša Hejná¹; ¹Aarhus University, Denmark; ²Charles University,

Czechia

Wed-E-V-4-2, Time: 19:00

Most of our knowledge of Danish f0 variation and intonation is based on the work of Grønnum and colleagues, who developed an a-phonological model in which a series of repeated "default" contours are superpositioned onto an overarching f0 slope. The current paper tests a range of predictions stemming from this model, most importantly the adequacy of analysing f0 modulations as a string of repeated contours differing in range but not in shape. To facilitate comparison with earlier work in the area, our material is based on read speech, 45 speakers of Jutland Danish participated in the experiment. Analyses of f0 in sentences of differing complexity supplied little evidence in favour of the existence of default contours. Instead, our acoustic data revealed an array of f0 shapes associated with various prosodic anchor points, which are influenced in both range and shape by positional context and the presence or absence of focus.

An Experiment in Paratone Detection in a Prosodically Annotated EAP Spoken Corpus

Adrien Méli¹, Nicolas Ballier¹, Achille Falaise², Alice Henderson³; ¹CLILLAC-ARP (EA 3967), France; ²LLF (UMR 7110), France; ³Lidilem (EA 609), France Wed-E-V-4-3, Time: 19:00

This article describes an experiment in paratone detection based on a spoken corpus of English for Academic Purposes (EAP) recently automatically re-annotated with prosodic information. The Momel and INTSINT annotations were carried out using SPPAS. The EIIDA corpus was chosen as it offered long uninterrupted stretches of speech of academic presentations. We describe the clustering method adopted for automatic detection, contrasting a supervised and an unsupervised method of paratone boundary detection. We showcase the relevance of the annotation scheme followed for this corpus and contribute to the investigation of the phonostyle of lecture delivery. We discuss the relevance of clustering methods applied to the labels of the pitch targets for the analysis of paratones.

ProsoBeast Prosody Annotation Tool

Branislav Gerazov¹, Michael Wagner²; ¹UKiM, Macedonia; ²McGill University, Canada Wed-E-V-4-4, Time: 19:00

The labelling of speech corpora is a laborious and time-consuming process. The ProsoBeast Annotation Tool seeks to ease and accelerate this process by providing an interactive 2D representation of the prosodic landscape of the data, in which contours are distributed based on their similarity. This interactive map allows the user to inspect and label the utterances. The tool integrates several state-of-the-art methods for dimensionality reduction and feature embedding, including variational autoencoders. The user can use these to find a good representation for their data. In addition, as most of these methods are stochastic, each can be used to generate an unlimited number of different prosodic maps. The web app then allows the user to seamlessly switch between these alternative representations in the annotation process. Experiments with a sample prosodically rich dataset have shown that the tool manages to find good representations of varied data and is helpful both for annotation and label correction. The tool is released as free software for use by the community.

Assessing the Use of Prosody in Constituency Parsing of Imperfect Transcripts

*Trang Tran*¹, *Mari Ostendorf*²; ¹*University of Southern California, USA*; ²*University of Washington, USA* Wed-E-V-4-5, Time: 19:00

This work explores constituency parsing on automatically recognized transcripts of conversational speech. The neural parser is based on a sentence encoder that leverages word vectors contextualized with prosodic features, jointly learning prosodic feature extraction with parsing. We assess the utility of the prosody in parsing on imperfect

transcripts, i.e. transcripts with automatic speech recognition (ASR) errors, by applying the parser in an N-best reranking framework. In experiments on Switchboard, we obtain 13–15% of the oracle N-best gain relative to parsing the 1-best ASR output, with insignificant impact on word recognition error rate. Prosody provides a significant part of the gain, and analyses suggest that it leads to more grammatical utterances via recovering function words.

Targeted and Targetless Neutral Tones in Taiwanese Southern Min

Roger Cheng-yen Liu, Feng-fan Hsieh, Yueh-chin Chang; National Tsing Hua University, Taiwan Wed-E-V-4-6, Time: 19:00

This article is an acoustic study on the two types of neutral tone in Taiwanese Southern Min (TSM). Recording materials included a set of verb-clitic constructions with different preceding tones and clitics. Pitch contours in different conditions were compared using Smoothing Spline ANOVA. Our results confirmed that Type 1 neutral tone (NT1) has a low pitch target and that Type 2 neutral tone (NT2) is contextually dependent. Whether NT1 or NT2 is chosen has been treated as the lexical idiosyncrasy of the clitics in question, with idiolectal and dialectal variations. However, we found in this study that the onsets have a bearing on determining the type of neutral tone: the more sonorous the onset, the more possible it is for the clitic to be in NT2. In sum, the two distinct types of neutral tones in TSM not only are unusual among the neutral tones in Sinitic languages, but they also offer novel data for the consonant-tone interaction.

The Interaction of Word Complexity and Word Duration in an Agglutinative Language

*Mária Gósy*¹, *Kálmán Abari*²; ¹*ELKH*, *Hungary*; ²*University of Debrecen*, *Hungary* Wed-E-V-4-7, Time: 19:00

The mental lexicon comprises the representations of various words either in a morphologically decomposed form, or in a conceptually non-decomposed form. The durations of mono-morphemic and multimorphemic words are assumed to contain information on the routes of their lexical access.

The durations of Hungarian nouns with various lengths produced spontaneously by 10 young and 10 elderly speakers (with 55 years of difference between them) were measured. Findings showed significant differences depending on the words' complexity and on age. The nouns both with and without suffixes were significantly longer in old than in young speakers. The durational differences depending on age were more pronounced in monomorphemic nouns as opposed to multimorphemic nouns. Along with the increasing number of syllables of the nouns, old speakers produced increasingly longer simple nouns (stems) than young ones did.

We suggest that multimorphemic nouns are accessed decompositionally in spontaneous utterances when the stem activation is followed by the activation of the suffixes. The specific storage and the corresponding lexical access of the morphemes explain the longer durations of the inflected nouns.

Taiwan Min Nan (Taiwanese) Checked Tones Sound Change

Ho-hsien Pan, Shao-ren Lyu; NYCU, Taiwan Wed-E-V-4-8, Time: 19:00

The multifaced changes of Taiwan Min Nan (TMN) checked sandhi tones, S3 and S5 were investigated as well as the checked base tones, B3 and B5. Simultaneous EGG data, CQ_H and acoustic data,

including duration, f0 offset at 80% vowel interval, and spectral tilt H1^{*}-A3^{*} from forty male and female speakers above 40 and under 30 years of age were analyzed. Though different measures progress at different paces, in general, as the coda stops [p, t, k, ?] from full stop closure, to energy damping and finally to complete deletion, vowel duration lengthening, f0 offset lowering, and more modal phonation were observed. Gender effects were found on f0 offset and CQ_H offset. The pace of progress is more advanced for base tone B5 with glottal coda stops. After coda deletion, the contexts conditioning the anticipatory co-articulation were removed and vowel and tone characteristics were modified to be similar to those found in open syllables.

In-Group Advantage in the Perception of Emotions: Evidence from Three Varieties of German

Moritz Jakob¹, Bettina Braun¹, Katharina Zahner-Ritter²; ¹Universität Konstanz, Germany; ²Universität Trier, Germany Wed-E-V-4-9, Time: 19:00

Various studies on the perception of vocally expressed emotions have shown that recognition rates are higher if speaker and listener belong to the same cultural or linguistic group. This so-called in-group advantage is commonly attributed to prosodic differences in the expression of emotion across groups. Evidence comes mostly from using cross-linguistic and/or cross-cultural study designs. Previous research suggests that varieties of German differ in their use of prosody and can be discriminated based on prosodic features alone. In this paper, we tested whether emotion recognition rates differ across varieties of German: Listeners from three dialectal areas (Hamburg, Vienna, Zurich) identified emotions on semantically neutral sentences (choosing between anger, happiness, relief, surprise or "other"), spoken by actors from the three regions. Correctness rates show that emotions are recognized better if speakers and listeners are native speakers of the same variety. However, further analyses suggest that the in-group advantage does not surface consistently across individual emotions. To explain these results, the prosodic realization of the sentences was tested for interactions between emotion and variety. Here, intensity seemed to differ most across varieties and emotions. Importantly, we show that the in-group advantage extends from cultural groups to dialectal groups of a language.

The LF Model in the Frequency Domain for Glottal Airflow Modelling Without Aliasing Distortion

Christer Gobl; Trinity College Dublin, Ireland Wed-E-V-4-10, Time: 19:00

Many of the commonly used voice source models are based on piecewise elementary functions defined in the time domain. The discrete-time implementation of such models generally causes aliasing distortion, which make them less useful for certain applications. This paper presents a method which eliminates this distortion. The key component of the proposed method is the frequency domain description of the source model. By deploying the Laplace transform and phasor arithmetic, closed-form expressions of the source model spectrum can be derived. This facilitates the calculation of the spectrum directly from the model parameters, which in turn makes it possible to obtain the ideal discrete spectrum of the model given the sampling frequency used. This discrete spectrum is entirely free of aliasing distortion, and the inverse discrete Fourier transform is used to compute the sampled glottal flow pulse. The proposed method was applied to the widely used LF model, and the complete Laplace transform of the model is presented. Also included are closed-form expressions of the amplitude spectrum and the phase spectrum for the calculation of the LF model spectrum.

Parsing Speech for Grouping and Prominence, and the Leveraging the Uniformity Framework to Examine **Typology of Rhythm**

Michael Wagner, Alvaro Iturralde Zurita, Sijia Zhang; McGill University, Canada Wed-E-V-4-11, Time: 19:00

Humans appear to be wired to perceive acoustic events rhythmically. English speakers, for example, tend to perceive alternating short and long sounds as a series of binary groups with a final beat (iambs), and alternating soft and loud sounds as a series of trochees. This generalization, often called the 'Iambic-trochaic Law' (ITL), although viewed as an auditory universal by some, has been argued to be shaped by language experience. Earlier work on the ITL had a crucial limitation, in that it did not tease apart the percepts of grouping and prominence, which the notions of iamb and trochee inherently confound. We explore how intensity and duration relate to percepts of prominence and grouping in six languages (English, French, German, Japanese, Mandarin, and Spanish). The results show that the ITL is not universal, and that cue interpretation is shaped by language experience. However, there are also invariances: Duration appears relatively robust across languages as a cue to prominence (longer syllables are perceived as stressed), and intensity for grouping (louder syllables are perceived as initial). The results show the beginnings of a rhythmic typology based on how the dimensions of grouping and prominence are cued.

Prosody of Case Markers in Urdu

Benazir Mumtaz, Massimiliano Canzi, Miriam Butt; Universität Konstanz, Germany Wed-E-V-4-12, Time: 19:00

This paper studies the prosody of case clitics in Urdu, for which various different claims exist in the literature. We conducted a production experiment and controlled for effects potentially arising from the phonetics of the case clitics, the syntactic function they express and clausal position. We find that case clitics are incorporated into the prosodic phrase of the noun and that they become part of the overall LH contour found on accentual phrases in Urdu/Hindi. We also find some differences across case type and position which we tie to information structural effects.

Articulatory Characteristics of Icelandic Voiced Fricative Lenition: Gradience, Categoricity, and Speaker/Gesture-Specific Effects

Brynhildur Stefansdottir, Francesco Burroni, Sam Tilsen; Cornell University, USA

Wed-E-V-4-13, Time: 19:00

Icelandic voiced fricatives frequently reduce in connected speech. However, systematic investigations of the phenomenon from acoustic and articulatory perspectives are lacking. To further the understanding of this lenition process, we present electromagnetic articulography and acoustic data from four speakers concerning the intervocalic realization of the dental and velar fricatives. The results show that lenition is mostly gradient, but some speakers and places of articulation exhibit two distinct modes suggesting a categorical distinction. Moreover, in some tokens, the fricative constriction is absent from the articulatory trajectories. Finally, the relation between lenition and speech rate, style, and stress is also subject to speaker- and gesture-specific effects. We conclude by evaluating how our findings challenge the common assumptions, made in the literature, that lenition is a change in gestural target or a perceptually driven phenomenon.

Crosslinguistic Similarity for Long-Lag Stops in Spontaneous Cantonese-English Bilingual Speech

Khia A. Johnson; University of British Columbia, Canada Wed-E-V-4-14, Time: 19:00

While crosslinguistic influence is widespread in bilingual speech production, it is less clear which aspects of representation are shared across languages, if any. Most prior work examines phonetically distinct yet phonologically similar sounds, for which phonetic convergence suggests a cross-language link within individuals [1]. Convergence is harder to assess when sounds are already similar, as with English and Cantonese initial long-lag stops. Here, the articulatory uniformity framework [2, 3, 4] is leveraged to assess whether bilinguals share an underlying laryngeal feature across languages, and describe the nature of cross-language links. Using the SpiCE corpus of spontaneous Cantonese-English bilingual speech [5], this paper asks whether Cantonese-English bilinguals exhibit uniform voice-onset time for long-lag stops within and across languages. Results indicate moderate patterns of uniformity within-language replicating prior work [2, 6] — and weaker patterns across languages. The analysis, however, raises many questions, as correlations were generally lower compared to prior work, and talkers did not adhere to expected ordinal relationships by place of articulation. Talkers also retained clear differences for /t/ and /k/, despite expectations of similarity. Yet at the same time, more of the overall variation seems to derive from individual-specific differences. While many questions remain, the uniformity framework shows promise.

Wed-E-V-5: Single-Channel Speech Enhancement

19:00-21:00, Wednesday 1 September 2021 Chairs: Dorothea Kolossa and Pejman Mowlaee

Personalized Speech Enhancement Through Self-Supervised Data Augmentation and Purification

Aswin Sivaraman, Sunwoo Kim, Minje Kim; Indiana University, USA

Wed-E-V-5-1, Time: 19:00

Training personalized speech enhancement models is innately a noshot learning problem due to privacy constraints and limited access to noise-free speech from the target user. If there is an abundance of unlabeled noisy speech from the test-time user, one may train a personalized speech enhancement model using self-supervised learning. One straightforward approach to model personalization is to use the target speaker's noisy recordings as pseudo-sources. Then, a pseudo denoising model learns to remove injected training noises and recover the pseudo-sources. However, this approach is volatile as it depends on the quality of the pseudo-sources, which may be too noisy. To remedy this, we propose a data purification step that refines the self-supervised approach. We first train an SNR predictor model to estimate the frame-by-frame SNR of the pseudo-sources. Then, we convert the predictor's estimates into weights that adjust the pseudo-sources' frame-by-frame contribution towards training the personalized model. We empirically show that the proposed data purification step improves the usability of the speaker-specific noisy data in the context of personalized speech enhancement. Our approach may be seen as privacy-preserving as it does not rely on any clean speech recordings or speaker embeddings.

Speech Denoising with Auditory Models

Mark R. Saddler¹, Andrew Francl¹, Jenelle Feather¹, Kaizhi Qian², Yang Zhang², Josh H. McDermott¹; ¹MIT, USA; ²MIT-IBM Watson AI Lab, USA Wed-E-V-5-2, Time: 19:00

Contemporary speech enhancement predominantly relies on audio transforms that are trained to reconstruct a clean speech waveform. The development of high-performing neural network sound recognition systems has raised the possibility of using deep feature representations as 'perceptual' losses with which to train denoising systems. We explored their utility by first training deep neural networks to classify either spoken words or environmental sounds from audio. We then trained an audio transform to map noisy speech to an audio waveform that minimized the difference in the deep feature representations between the output audio and the corresponding clean audio. The resulting transforms removed noise substantially better than baseline methods trained to reconstruct clean waveforms, and also outperformed previous methods using deep feature losses. However, a similar benefit was obtained simply by using losses derived from the filter bank inputs to the deep networks. The results show that deep features can guide speech enhancement, but suggest that they do not yet outperform simple alternatives that do not involve learned features.

Human Listening and Live Captioning: Multi-Task Training for Speech Enhancement

Sefik Emre Eskimez, Xiaofei Wang, Min Tang, Hemin Yang, Zirun Zhu, Zhuo Chen, Huaming Wang, Takuya Yoshioka; Microsoft, USA Wed-E-V-5-3, Time: 19:00

With the surge of online meetings, it has become more critical than ever to provide high-quality speech audio and live captioning under various noise conditions. However, most monaural speech enhancement (SE) models introduce processing artifacts and thus degrade the performance of downstream tasks, including automatic speech recognition (ASR). This paper proposes a multi-task training framework to make the SE models unharmful to ASR. Because most ASR training samples do not have corresponding clean signal references, we alternately perform two model update steps called SE-step and ASR-step. The SE-step uses clean and noisy signal pairs and a signalbased loss function. The ASR-step applies a pre-trained ASR model to training signals enhanced with the SE model. A cross-entropy loss between the ASR output and reference transcriptions is calculated to update the SE model parameters. Experimental results with realistic large-scale settings using ASR models trained on 75,000-hour data show that the proposed framework improves the word error rate for the SE output by 11.82% with little compromise in the SE quality. Performance analysis is also carried out by changing the ASR model, the data used for the ASR-step, and the schedule of the two update steps.

Multi-Stage Progressive Speech Enhancement Network

Xinmeng Xu, Yang Wang, Dongxiang Xu, Yiyuan Peng, Cong Zhang, Jie Jia, Binbin Chen; vivo, China Wed-E-V-5-4, Time: 19:00

Speech enhancement is a fundamental way to separate and generate clean speech from adverse environment where the received speech is seriously corrupted by noise. This paper applies a novel progressive network for speech enhancement by using multi-stage structure, where each stage contains a channel attention block followed by dilated encoder-decoder convolutional network with gated linear units. In addition, each stage generates a prediction that is refined

NOTES

by a supervised attention block. What is more, a fusion block is inserted between original inputs and outputs of previous stage. Multi-stage architecture is introduced to sequentially invoke multiple deep-learning networks, and its key ingredient is the information exchange between different stages. Thus, a more flexible and robust outputs can be generated. Experimental results show that the proposed architecture obtains consistently better performance than recent state-of-the-art models in terms of both PESQ and STOI scores.

Single-Channel Speech Enhancement Using Learnable Loss Mixup

*Oscar Chang*¹, *Dung N. Tran*², *Kazuhito Koishida*²; ¹*Columbia University, USA*; ²*Microsoft, USA* Wed-E-V-5-5, Time: 19:00

Generalization remains a major problem in supervised learning of single-channel speech enhancement. In this work, we propose *learnable loss mixup (LLM)*, a simple and effortless training diagram, to improve the generalization of deep learning-based speech enhancement models. *Loss mixup*, of which *learnable loss mixup* is a special variant, optimizes a mixture of the loss functions of random sample pairs to train a model on virtual training data constructed from these pairs of samples. In *learnable loss mixup*, by conditioning on the mixed data, the loss functions are mixed using a non-linear mixing function automatically learned via neural parameterization. Our experimental results on the VCTK benchmark show that *learnable loss mixup* achieves 3.26 PESQ, outperforming the state-of-the-art.

A Maximum Likelihood Approach to SNR-Progressive Learning Using Generalized Gaussian Distribution for LSTM-Based Speech Enhancement

*Xiao-Qi Zhang*¹, *Jun Du*¹, *Li Chai*¹, *Chin-Hui Lee*²; ¹*USTC, China*; ²*Georgia Tech, USA* Wed-E-V-5-6, Time: 19:00

A maximum likelihood (ML) approach to characterizing regression errors in each target layer of SNR progressive learning (PL) using long short-term memory (LSTM) networks is proposed to improve performances of speech enhancement at low SNR levels. Each LSTM layer is guided to learn an intermediate target with a specific SNR gain. In contrast to using previously proposed minimum squared error criterion (MMSE-PL-LSTM) which leads to an un-even distribution and a broad dynamic range of the prediction errors, we model the errors with a generalized Gaussian distribution (GGD) at all intermediate layers in the newly proposed ML-PL-LSTM framework. The shape factors in GGD can be automatically updated when training the LSTM networks in a layer-wise manner to estimate the network parameters progressively. Tested on the CHiME-4 simulation set for speech enhancement in unseen noise conditions, the proposed ML-PL-LSTM approach outperforms MMSE-PL-LSTM in terms of both PESQ and STOI measures. Furthermore, when evaluated on the CHiME-4 real test set for speech recognition, using ML-enhanced speech also results in less word error rates than those obtained with MMSE-enhanced speech.

Whisper Speech Enhancement Using Joint Variational Autoencoder for Improved Speech Recognition

*Vikas Agrawal*¹, *Shashi Kumar*¹, *Shakti P. Rath*²; ¹*Samsung, India;* ²*Reverie Language Technologies, India* Wed-E-V-5-7, Time: 19:00

Whispering is the natural choice of communication when one wants to interact quietly and privately. Due to vast differences in acoustic characteristics of whisper and natural speech, there is drastic degradation in the performance of whisper speech when decoded by the Automatic Speech Recognition (ASR) system trained on neutral speech. Recently, to handle this mismatched train and test scenario Denoising Autoencoders (DA) are used which gives some improvement. To improve over DA performance we propose another method to map speech from whisper domain to neutral speech domain via Joint Variational Auto-Encoder (JVAE). The proposed method requires time-aligned parallel data which is not available, so we developed an algorithm to convert parallel data to time-aligned parallel data. JVAE jointly learns the characteristics of whisper and neutral speech in a common latent space which significantly improves whisper recognition accuracy and outperforms traditional autoencoder based techniques. We benchmarked our method against two baselines, first being ASR trained on neutral speech and tested on whisper dataset and second being whisper test set mapped using DA and tested on same neutral ASR. We achieved an absolute improvement of 22.31% in Word Error Rate (WER) over the first baseline and an absolute 5.52% improvement over DA.

DEMUCS-Mobile : On-Device Lightweight Speech Enhancement

Lukas Lee, Youna Ji, Minjae Lee, Min-Seok Choi; Naver, Korea

Wed-E-V-5-8, Time: 19:00

As the importance of speech enhancement for real-world application increases, the compactness of the model is also becoming a crucial study. In this paper, we present compression techniques to reduce the model size and applied them to the state-of-the-art real-time speech enhancement system. We successfully reduce the model size by actively applying channel pruning while maintaining performance. In particular, we propose a method to prune more channels of convolutional neural networks (CNN) by utilizing gated linear unit (GLU) activation. In addition, lower-bit-quantization is applied to reduce model size, while minimizing performance degradation caused by quantization. We show the performance of our proposed model on a mobile device where computing resources are limited. In particular, it is implemented to enable streaming, and speech enhancement works in real-time.

Speech Denoising Without Clean Training Data: A Noise2Noise Approach

Madhav Mahesh Kashyap, Anuj Tambwekar, Krishnamoorthy Manohara, S. Natarajan; PES University, India Wed-E-V-5-9, Time: 19:00

This paper tackles the problem of the heavy dependence of clean speech data required by deep learning based audio-denoising methods by showing that it is possible to train deep speech denoising networks using only noisy speech samples. Conventional wisdom dictates that in order to achieve good speech denoising performance, there is a requirement for a large quantity of both noisy speech samples and perfectly clean speech samples, resulting in a need for expensive audio recording equipment and extremely controlled soundproof recording studios. These requirements pose significant challenges in data collection, especially in economically disadvantaged regions and for low resource languages. This work shows that speech denoising deep neural networks can be successfully trained utilizing only noisy training audio. Furthermore it is revealed that such training regimes achieve superior denoising performance over conventional training regimes utilizing clean training audio targets, in cases involving complex noise distributions and low Signal-to-Noise ratios (high noise environments). This is demonstrated through experiments studying the efficacy of our proposed approach over both real-world noises and synthetic noises using the 20 layered Deep Complex U-Net architecture.

NOTES

Improved Speech Enhancement Using a Complex-Domain GAN with Fused Time-Domain and Time-Frequency Domain Constraints

Feng Dang, Pengyuan Zhang, Hangting Chen; CAS, China

Wed-E-V-5-10, Time: 19:00

Complex-domain models have achieved promising results for speech enhancement (SE) tasks. Some complex-domain models consider only time-frequency (T-F) domain constraints and do not take advantage of the information at the time-domain waveform level. Some complex-domain models consider only time-domain constraints and do not take into account T-F domain constraints that have rich harmonic structure information. Indeed some complex-domain models consider both time-domain and T-F domain constraints but only use the simple mean square loss as time-frequency-domain constraints. This paper proposes a complex-domain-based speech enhancement method that integrates time-domain constraints and T-F domain constraints into a unified framework using a Generative Adversarial Network (GAN). The proposed framework captures information at the time-domain waveform level features while paying attention to the harmonic structure by time-domain and T-F domain constraints. We conducted experiments on the Voice Bank + DEMAND dataset to evaluate the proposed method. Experimental results show that the proposed method improves the PESQ score by 0.09 and the STOI score by 1% over the strong baseline deep complex convolution recurrent network (DCCRN) and outperforms the state-of-the-art GAN-based SE systems.

Speech Enhancement with Topology-Enhanced Generative Adversarial Networks (GANs)

*Xudong Zhang*¹, *Liang Zhao*², *Feng Gu*³; ¹*CUNY Graduate Center, USA*; ²*CUNY Lehman College, USA*; ³*CUNY CSI, USA* Wed-E-V-5-11, Time: 19:00

Speech enhancement is one of the effective approaches in improving speech quality. Neural network models have been widely used in speech enhancement, such as recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and generative adversarial networks (GANs). However, some of them either handle the speech noise removal tasks in the spectral domain or lack the waveform recovery capability. As a result, the enhanced speeches still include noisy signals. In this study, we propose a topology-enhanced GAN model to tackle noisy speeches in an end-to-end structure. We use the topology features of speech waves as additional constraints and modify the objective function of the GAN by adding a penalty term. The penalty term is a Wasserstein distance of topology features measuring the difference between the generated speech and the corresponding clean speech. We evaluate the proposed speechenhanced model on the public speech data set with 56 speakers and 20 different types of noisy conditions. The experimental results indicate that the topology features improve the performance of GANs on speech enhancement in metrics of PESQ, CBAK, COVL, and SSNR.

Learning Speech Structure to Improve Time-Frequency Masks

*Suliang Bu*¹, *Yunxin Zhao*¹, *Shaojun Wang*², *Mei Han*²; ¹*Mizzou, USA*; ²*PAII, USA*

Wed-E-V-5-12, Time: 19:00

Time-frequency (TF) masks are widely used in speech enhancement (SE). However, accurately estimating TF masks from noisy speech remains a challenge to both statistical or neural network approaches.

Statistical model-based mask estimation usually depends on a good parameter initialization, while NN-based mask estimation relies on setting proper and stable learning targets. To address these issues, we propose a novel approach to extracting TF speech structures from clean speech data, and partition a noisy speech spectrogram into mutually exclusive regions of core speech, core noise, and transition. Using such region targets derived from clean speech, we train bidirectional LSTM to learn region prediction from noisy speech, which is easier to do than mask prediction. The predicted regions can further be used in place of masks in beamforming, or integrated with statistical and NN based mask estimation to constrain mask values and model parameter updates. Our experimental results on ASR (CHiME-3) and SE (CHiME-3 and LibriSpeech) have demonstrated the effectiveness of our approach of learning speech region structure to improve TF masks.

SE-Conformer: Time-Domain Speech Enhancement Using Conformer

Eesung Kim, Hyeji Seo; Kakao, Korea Wed-E-V-5-13, Time: 19:00

Convolution-augmented transformer (conformer) has recently shown competitive results in speech-domain applications, such as automatic speech recognition, continuous speech separation, and sound event detection. Conformer can capture both the short and long-term temporal sequence information by attending to the whole sequence at once with multi-head self-attention and convolutional neural network. However, the effectiveness of conformer in speech enhancement has not been demonstrated. In this paper, we propose an end-to-end speech enhancement architecture (SE-Conformer), incorporating a convolutional encoder-decoder and conformer, designed to be directly applied to the time-domain signal. We performed evaluations on both the VoiceBank-DEMAND Corpus (VCTK) and Librispeech datasets in terms of objective speech quality metrics. The experimental results show that the proposed model outperforms other competitive baselines in speech enhancement performance.

Wed-E-V-6: Speech Synthesis: Tools, Data, Evaluation

19:00-21:00, Wednesday 1 September 2021 Chairs: Sébastien Le Maguer and Tim Bunnell

Spectral and Latent Speech Representation Distortion for TTS Evaluation

Thananchai Kongthaworn, Burin Naowarat, Ekapol Chuangsuwanich; Chulalongkorn University, Thailand Wed-E-V-6-1, Time: 19:00

One of the main problems in the development of text-to-speech (TTS) systems is its reliance on subjective measures, typically the Mean Opinion Score (MOS). MOS requires a large number of people to reliably rate each utterance, making the development process slow and expensive. Recent research on speech quality assessment tends to focus on training models to estimate MOS, which requires a large number of training data, something that might not be available in low-resource languages. We propose an objective assessment metric based on the DTW distance using the spectrogram and the high-level features from an Automatic Speech Recognition (ASR) model to cover both acoustic and linguistic information. Experiments on Thai TTS and the Blizzard Challenge datasets show that our method outperformed other baselines in both utterance- and system-level by a large margin in terms of correlation coefficients. Our metric also outperformed the best baseline by 9.58% when used in head-to-head

NOTES

utterance-level comparisons. Ablation studies suggest that the middle layers of the ASR model are most suitable for TTS evaluation when used in conjunction with spectral features.

Detection and Analysis of Attention Errors in Sequence-to-Sequence Text-to-Speech

Cassia Valentini-Botinhao, Simon King; University of Edinburgh, UK Wed-E-V-6-2, Time: 19:00

Sequence-to-sequence speech synthesis models are notorious for gross errors such as skipping and repetition, commonly associated with failures in the attention mechanism. While a lot has been done to improve attention and decrease errors, this paper focuses instead on automatic error detection and analysis. We evaluated three objective metrics against error detection scores collected by human listening. All metrics were derived from the synthesised attention matrix alone and do not require a reference signal, relying on the expectation that errors occur when attention is dispersed or insufficient. Using one of this metrics as an analysis tool, we observed that gross errors are more likely to occur in longer sentences and in sentences with punctuation marks that indicate pause or break. We also found that mechanisms such as forcibly incremented attention have the potential for decreasing gross errors but to the detriment of naturalness. The results of the error detection evaluation revealed that two of the evaluated metrics were able to detect errors with a relatively high success rate, obtaining F-scores of up to 0.89 and 0.96

RyanSpeech: A Corpus for Conversational Text-to-Speech Synthesis

Rohola Zandie¹, Mohammad H. Mahoor¹, Julia Madsen², Eshrat S. Emamian²; ¹University of Denver, USA; ²DreamFace Technologies, USA Wed-E-V-6-3, Time: 19:00

This paper introduces *RyanSpeech*, a new speech corpus for research on automated text-to-speech (TTS) systems. Publicly available TTS corpora are often noisy, recorded with multiple speakers, or lack quality male speech data. In order to meet the need for a high quality, publicly available male speech corpus within the field of speech recognition, we have designed and created *RyanSpeech* which contains textual materials from real-world conversational settings. These materials contain over 10 hours of a professional male voice actor's speech recorded at 44.1 kHz. This corpus's design and pipeline make *RyanSpeech* ideal for developing TTS systems in real-world applications. To provide a baseline for future research, protocols, and benchmarks, we trained 4 state-of-the-art speech models and a vocoder on *RyanSpeech*. The results show 3.36 in mean opinion scores (MOS) in our best model. We have made both the corpus and trained models for public use.

AISHELL-3: A Multi-Speaker Mandarin TTS Corpus

Yao Shi¹, Hui Bu², Xin Xu², Shaoji Zhang², Ming Li¹; ¹Wuhan University, China; ²Beijing Shell Shell Technology, China Wed-E-V-6-4, Time: 19:00

In this paper, we present AISHELL-3, a large-scale multi-speaker Mandarin speech corpus which could be used to train multi-speaker Text-To-Speech (TTS) systems. The corpus contains roughly 85 hours of emotion-neutral recordings spanning across 218 native Chinese mandarin speakers. Their auxiliary attributes such as gender, age group and native accents are explicitly marked and provided in the corpus. Moreover, transcripts in Chinese character-level and pinyin-level are provided along with the recordings. We also present some data processing strategies and techniques which match with the characteristics of the presented corpus and conduct experiments on multiple speech-synthesis systems to assess the quality of the generated speech samples, showing promising results. The corpus is available online under Apache v2.0 license.

Comparing Speech Enhancement Techniques for Voice Adaptation-Based Speech Synthesis

Nicholas Eng, C.T. Justine Hui, Yusuke Hioka, Catherine I. Watson; University of Auckland, New Zealand

Wed-E-V-6-5, Time: 19:00

This study investigates the use of speech enhancement techniques in creating text-to-speech voices with degraded or noisy speech. A number of synthetic voices were created using speech that was first degraded by different noise types at various signal-to-noise ratios (SNRs), then enhanced through four speech enhancement algorithms: Subspace, Wiener filter, SEGAN and a DNN-based method. Subjective listening tests show that the quality of the synthetic voices produced by subspace and the DNN-based method enhanced speech outperforms the quality of the voices created using Wiener filter or SEGAN enhanced speech at low SNRs, and speech enhanced by the subspace method results in higher quality synthetic speech at higher SNRs.

EMOVIE: A Mandarin Emotion Speech Dataset with a Simple Emotional Text-to-Speech Model

*Chenye Cui*¹, *Yi Ren*¹, *Jinglin Liu*¹, *Feiyang Chen*¹, *Rongjie Huang*¹, *Ming Lei*², *Zhou Zhao*¹; ¹*Zhejiang University, China*; ²*Alibaba, China* Wed-E-V-6-6, Time: 19:00

Recently, there has been an increasing interest in neural speech synthesis. While the deep neural network achieves the state-ofthe-art result in text-to-speech (TTS) tasks, how to generate a more emotional and more expressive speech is becoming a new challenge to researchers due to the scarcity of high-quality emotion speech dataset and the lack of advanced emotional TTS model. In this paper, we first briefly introduce and publicly release a Mandarin emotion speech dataset including 9,724 samples with audio files and its emotion human-labeled annotation. After that, we propose a simple but efficient architecture for emotional speech synthesis called EMSpeech. Unlike those models which need additional reference audio as input, our model could predict emotion labels just from the input text and generate more expressive speech conditioned on the emotion embedding. In the experiment phase, we first validate the effectiveness of our dataset by an emotion classification task. Then we train our model on the proposed dataset and conduct a series of subjective evaluations. Finally, by showing a comparable performance in the emotional speech synthesis task, we successfully demonstrate the ability of the proposed model.

Perception of Social Speaker Characteristics in Synthetic Speech

Sai Sirisha Rallabandi, Abhinav Bharadwaj, Babak Naderi, Sebastian Möller; Technische Universität Berlin, Germany

Wed-E-V-6-7, Time: 19:00

With the improved computational abilities, the usage of chatbots and conversational agents has become more prevalent. Therefore, it is essential that these agents exhibit certain social speaker characteristics in the generated speech. In this paper, we study the perception of such speaker characteristics in two commercial Textto-Speech (TTS) systems, Amazon Polly and Google TTS. We carried out a 15-item semantic differential scaling test. The factor analysis provided us with three underlying dimensions that can be perceived from synthetic speech, warmth, competence, and extraversion. Our results show that we can perceive both interpersonal relationships and also personality traits from synthetic voices. Additionally, we observed that the female participants perceived male voices to be more responsible, energetic, relaxed, and enthusiastic. In comparison, male participants found female voices to be more reliable, accessible, and confident. A discussion on the comparison of our results with that of the studies on natural speech is also provided.

Hi-Fi Multi-Speaker English TTS Dataset

Evelina Bakhturina, Vitaly Lavrukhin, Boris Ginsburg, Yang Zhang; NVIDIA, USA

Wed-E-V-6-8, Time: 19:00

This paper introduces a new multi-speaker English dataset for training text-to-speech models. The dataset is based on LibriVox audiobooks and Project Gutenberg texts, both in the public domain. The new dataset contains about 292 hours of speech from 10 speakers with at least 17 hours per speaker sampled at 44.1 kHz. To select speech samples with high quality, we considered audio recordings with a signal bandwidth of at least 13 kHz and a signal-to-noise ratio (SNR) of at least 32 dB. The dataset is publicly released.

Utilizing Self-Supervised Representations for MOS Prediction

Wei-Cheng Tseng, Chien-yu Huang, Wei-Tsung Kao, Yist Y. Lin, Hung-yi Lee; National Taiwan University, Taiwan

Wed-E-V-6-9, Time: 19:00

Speech quality assessment has been a critical issue in speech processing for decades. Existing automatic evaluations usually require clean references or parallel ground truth data, which is infeasible when the amount of data soars. Subjective tests, on the other hand, do not need any additional clean or parallel data and correlates better to human perception. However, such a test is expensive and time-consuming because crowd work is necessary. It thus becomes highly desired to develop an automatic evaluation approach that correlates well with human perception while not requiring ground truth data. In this paper, we use self-supervised pre-trained models for MOS prediction. We show their representations can distinguish between clean and noisy audios. Then, we fine-tune these pre-trained models followed by simple linear layers in an end-to-end manner. The experiment results showed that our framework outperforms the two previous state-of-the-art models by a significant improvement on Voice Conversion Challenge 2018 and achieves comparable or superior performance on Voice Conversion Challenge 2016. We also conducted an ablation study to further investigate how each module benefits the task. The experiment results are implemented and reproducible with publicly available toolkits.

KazakhTTS: An Open-Source Kazakh Text-to-Speech Synthesis Dataset

Saida Mussakhojayeva, Aigerim Janaliyeva, Almas Mirzakhmetov, Yerbolat Khassanov, Huseyin Atakan Varol; Nazarbayev University, Kazakhstan Wed-E-V-6-10, Time: 19:00

This paper introduces a high-quality open-source speech synthesis dataset for Kazakh, a low-resource language spoken by over 13 million people worldwide. The dataset consists of about 93 hours of

transcribed audio recordings spoken by two professional speakers (female and male). It is the first publicly available large-scale dataset developed to promote Kazakh text-to-speech (TTS) applications in both academia and industry. In this paper, we share our experience by describing the dataset development procedures and faced challenges, and discuss important future directions. To demonstrate the reliability of our dataset, we built baseline end-to-end TTS models and evaluated them using the subjective mean opinion score (MOS) measure. Evaluation results show that the best TTS models trained on our dataset achieve MOS above 4 for both speakers, which makes them applicable for practical use. The dataset, training recipe, and pretrained TTS models are freely available.

Confidence Intervals for ASR-Based TTS Evaluation

Jason Taylor, Korin Richmond; University of Edinburgh, UK

Wed-E-V-6-11, Time: 19:00

Automatic speech recognition (ASR) is increasingly used to evaluate the intelligibility of text-to-speech synthesis (TTS). ASR is less costly than traditional listening tests, but questions remain about its reliability. We re-evaluate the Blizzard Challenge's intelligibility tasks in English since 2011 using ASR. Re-analysing transcriptions collected by paid in-lab participants, online volunteers and Amazon Mechanical Turkers (the latter used only in 2011), we compare their word error rates (WERs) and statistically-significant system-groupings with those generated by an open-source, Transformer-based ASR model. This ASR model consistently decodes test stimuli with more reliable WERs than the Blizzard Challenge's (mostly non-native) speech experts and online volunteers. The model also groups systems according to statistical significance similarly to the paid in-lab participants. Using surplus semantically unpredictable sentences (SUS) submitted every year to the challenge, we investigate how confidence intervals in ASR WERs change as the number of transcribed stimuli increases. We plot the Frobenius norm of pairwise significance matrices with increasing stimuli. We find that finer groupings of systems are detected as confidence intervals narrow. The number of stimuli where p-values start to converge ranges from 400-800 stimuli. We conclude that, with enough stimuli, ASR can be more reliable than humans.

Wed-E-SS-1: INTERSPEECH 2021 Deep Noise Suppression Challenge

Room Lacina, 19:00–21:00, Wednesday 1 September 2021 Chairs: Harishchandra Dubey and Ross Cutler

INTERSPEECH 2021 Deep Noise Suppression Challenge

*Chandan K.A. Reddy*¹, *Harishchandra Dubey*¹, *Kazuhito Koishida*¹, *Arun Nair*², *Vishak Gopal*¹, *Ross Cutler*¹, *Sebastian Braun*¹, *Hannes Gamper*¹, *Robert Aichner*¹, *Sriram Srinivasan*¹; ¹*Microsoft, USA*; ²*Johns Hopkins University, USA* Wed-E-SS-1-1, Time: 19:00

The Deep Noise Suppression (DNS) challenge was designed to unify the research efforts in the area of noise suppression targeted for human perception. We recently organized a DNS challenge special session at INTERSPEECH 2020 and ICASSP 2021. We open-sourced training and test datasets for the wideband scenario along with a subjective evaluation framework based on ITU-T standard P.808, which was used to evaluate participants of the challenge. Many researchers from academia and industry made significant contributions to push the field forward, yet even the best noise suppressor was far

NOTES

from achieving superior speech quality in challenging scenarios. In this version of the challenge organized at INTERSPEECH 2021, we expanded our training and test datasets to accommodate fullband scenarios and challenging test conditions. We used ITU-T P.835 to evaluate the challenge winners as it gives additional information about the quality of processed speech and residual noise. The two tracks in this challenge focused on real-time denoising for (i) wideband, and (ii) fullband scenarios. We also made available a reliable non-intrusive objective speech quality metric for wideband called DNSMOS for the participants to use during their development phase.

A Simultaneous Denoising and Dereverberation Framework with Target Decoupling

Andong Li, Wenzhe Liu, Xiaoxue Luo, Guochen Yu, Chengshi Zheng, Xiaodong Li; CAS, China Wed-E-SS-1-2. Time: 19:00

Background noise and room reverberation are regarded as two major factors to degrade the subjective speech quality. In this paper, we propose an integrated framework to address simultaneous denoising and dereverberation under complicated scenario environments. It adopts a chain optimization strategy and designs four sub-stages accordingly. In the first two stages, we decouple the multi-task learning w.r.t. complex spectrum into magnitude and phase, and only implement noise and reverberation removal in the magnitude domain. Based on the estimated priors above, we further polish the spectrum in the third stage, where both magnitude and phase information are explicitly repaired with the residual learning. Due to the data mismatch and nonlinear effect of DNNs, the residual noise often exists in the DNN-processed spectrum. To resolve the problem, we adopt a light-weight algorithm as the post-processing module to capture and suppress the residual noise in the non-active regions. In the Interspeech 2021 Deep Noise Suppression (DNS) Challenge, our submitted system ranked top-1 for the real-time track in terms of Mean Opinion Score (MOS) with ITU-T P.835 framework.

Deep Noise Suppression with Non-Intrusive PESQNet Supervision Enabling the Use of Real Training Data

Ziyi Xu, Maximilian Strake, Tim Fingscheidt; Technische Universität Braunschweig, Germany Wed-E-SS-1-3, Time: 19:00

Data-driven speech enhancement employing deep neural networks (DNNs) can provide state-of-the-art performance even in the presence of non-stationary noise. During the training process, most of the speech enhancement neural networks are trained in a fully supervised way with losses requiring noisy speech to be synthesized by clean speech and additive noise. However, in a real implementation, only the noisy speech mixture is available, which leads to the question, how such data could be advantageously employed in training. In this work, we propose an end-to-end non-intrusive PESQNet DNN which estimates perceptual evaluation of speech quality (PESQ) scores, allowing a reference-free loss for real data. As a further novelty. we combine the PESQNet loss with denoising and dereverberation loss terms, and train a complex mask-based fully convolutional recurrent neural network (FCRN) in a "weakly" supervised way, each training cycle employing some synthetic data, some real data, and again synthetic data to keep the PESQNet up-to-date. In a subjective listening test, our proposed framework outperforms the Interspeech 2021 Deep Noise Suppression (DNS) Challenge baseline overall by 0.09 MOS points and in particular by 0.45 background noise MOS points.

DPCRN: Dual-Path Convolution Recurrent Network for Single Channel Speech Enhancement

Xiaohuai Le, Hongsheng Chen, Kai Chen, Jing Lu; Nanjing University, China Wed-E-SS-1-4, Time: 19:00

The dual-path RNN (DPRNN) was proposed to more effectively model extremely long sequences for speech separation in the time domain. By splitting long sequences to smaller chunks and applying intra-chunk and inter-chunk RNNs, the DPRNN reached promising performance in speech separation with a limited model size. In this paper, we combine the DPRNN module with Convolution Recurrent Network (CRN) and design a model called Dual-Path Convolution Recurrent Network (DPCRN) for speech enhancement in the timefrequency domain. We replace the RNNs in the CRN with DPRNN modules, where the intra-chunk RNNs are used to model the spectrum pattern in a single frame and the inter-chunk RNNs are used to model the dependence between consecutive frames. With only 0.8M parameters, the submitted DPCRN model achieves an overall mean opinion score (MOS) of 3.57 in the wide band scenario track of the Interspeech 2021 Deep Noise Suppression (DNS) challenge. Evaluations on some other test sets also show the efficacy of our model.

DCCRN+: Channel-Wise Subband DCCRN with SNR Estimation for Speech Enhancement

Shubo Lv, Yanxin Hu, Shimin Zhang, Lei Xie; Northwestern Polytechnical University, China Wed-E-SS-1-5, Time: 20:00

Deep complex convolution recurrent network (DCCRN), which extends CRN with complex structure, has achieved superior performance in MOS evaluation in Interspeech 2020 deep noise suppression challenge (DNS2020). This paper further extends DCCRN with the following significant revisions. We first extend the model to sub-band processing where the bands are split and merged by learnable neural network filters instead of engineered FIR filters, leading to a faster noise suppressor trained in an end-to-end manner. Then the LSTM is further substituted with a complex TF-LSTM to better model temporal dependencies along both time and frequency axes. Moreover, instead of simply concatenating the output of each encoder layer to the input of the corresponding decoder layer, we use convolution blocks to first aggregate essential information from the encoder output before feeding it to the decoder layers. We specifically formulate the decoder with an extra a priori SNR estimation module to maintain good speech quality while removing noise. Finally a post-processing module is adopted to further suppress the unnatural residual noise. The new model, named DCCRN+, has surpassed the original DCCRN as well as several competitive models in terms of PESQ and DNSMOS, and has achieved superior performance in the new Interspeech 2021 DNS challenge.

DBNet: A Dual-Branch Network Architecture Processing on Spectrum and Waveform for Single-Channel Speech Enhancement

Kanghao Zhang, Shulin He, Hao Li, Xueliang Zhang; Inner Mongolia University, China Wed-E-SS-1-6, Time: 20:00

In real acoustic environment, speech enhancement is an arduous task to improve the quality and intelligibility of speech interfered by background noise and reverberation. Over the past years, deep learning has shown great potential on speech enhancement. In this paper, we propose a novel real-time framework called DBNet which is a dual-branch structure with alternate interconnection. Each branch incorporates an encoder-decoder architecture with

NOTES

skip connections. The two branches are responsible for spectrum and waveform modeling, respectively. A bridge layer is adopted to exchange information between the two branches. Systematic evaluation and comparison show that the proposed system substantially outperforms related algorithms under very challenging environments. And in INTERSPEECH 2021 Deep Noise Suppression (DNS) challenge, the proposed system ranks the top 8 in real-time track 1 in terms of the Mean Opinion Score (MOS) of the ITU-T P.835 framework.

Low-Delay Speech Enhancement Using Perceptually Motivated Target and Loss

Xu Zhang, Xinlei Ren, Xiguang Zheng, Lianwu Chen, Chen Zhang, Liang Guo, Bing Yu; Kuaishou Technology, China

Wed-E-SS-1-7, Time: 20:00

Speech enhancement approaches based on deep neural network have outperformed the traditional signal processing methods. This paper presents a low-delay speech enhancement method that employs a new perceptually motivated training target and loss function. The proposed approach can achieve similar speech enhancement performance compared to the state-of-the-art approaches, but with significantly less latency and computational complexities. Judged by the MOS tests conducted by the INTERSPEECH 2021 Deep Noise Suppression Challenge organizer, the proposed method is ranked the 2nd place for Background Noise MOS, and the 6th place for overall MOS.

Lightweight Causal Transformer with Local Self-Attention for Real-Time Speech Enhancement

Koen Oostermeijer, Qing Wang, Jun Du; USTC, China Wed-E-SS-1-8, Time: 20:00

In this paper, we describe a novel speech enhancement transformer architecture. The model uses local causal self-attention, which makes it lightweight and therefore particularly well-suited for real-time speech enhancement in computation resource-limited environments. In addition, we provide several ablation studies that focus on different parts of the model and the loss function to figure out which modifications yield best improvements. Using this knowledge, we propose a final version of our architecture, that we sent in to the INTERSPEECH 2021 DNS Challenge, where it achieved competitive results, despite using only 2% of the maximally allowed computation. Furthermore, we performed experiments to compare it with with LSTM and CNN models, that had 127% and 257% more parameters, respectively. Despite this difference in model size, we achieved significant improvements on the considered speech quality and intelligibility measures.

Thu-M-O-1: Neural Network Training Methods and Architectures for ASR

Room A+B, 11:00–13:00, Thursday 2 September 2021 Chairs: Sanjeev Khudanpur and Philip Garner

Self-Paced Ensemble Learning for Speech and Audio Classification

*Nicolae-Cătălin Ristea*¹, *Radu Tudor Ionescu*²; ¹*UPB*, *Romania*; ²*University of Bucharest, Romania* Thu-M-O-1-1, Time: 11:00

Combining multiple machine learning models into an ensemble is known to provide superior performance levels compared to the individual components forming the ensemble. This is because models can complement each other in taking better decisions. Instead of just combining the models, we propose a self-paced ensemble learning scheme in which models learn from each other over several iterations. During the self-paced learning process based on pseudo-labeling, in addition to improving the individual models, our ensemble also gains knowledge about the target domain. To demonstrate the generality of our self-paced ensemble learning (SPEL) scheme, we conduct experiments on three audio tasks. Our empirical results indicate that SPEL significantly outperforms the baseline ensemble models. We also show that applying self-paced learning on individual models is less effective, illustrating the idea that models in the ensemble actually learn from each other.

Knowledge Distillation for Streaming Transformer-Transducer

Atsushi Kojima; Advanced Media, Japan Thu-M-O-1-2, Time: 11:20

We explore knowledge distillation methods from nonstreaming to streaming Transformer-Transducer (T-T) models. Streaming T-T truncates future context. It leads to recognition quality degradation compared with the original T-T. In this work, we explore knowledge distillation, which minimizes internal representations in all Transformer layers between nonstreaming and streaming T-T models. In the experiment, we compared two different methods: the minimization of the L2 distance of hidden vectors and the minimization of the L2 distance of heads. All experiments were conducted using the public LibriSpeech corpus. Results of the experiment showed that hidden vector similarity-based knowledge distillation is better than multi-head similarity-based knowledge distillation. We observed 3.5% and 2.1% relative reductions in word error rate compared with the original streaming T-T in test-clean set and test-other set, respectively.

Multi-Encoder Learning and Stream Fusion for Transformer-Based End-to-End Automatic Speech Recognition

Timo Lohrenz, Zhengyang Li, Tim Fingscheidt; Technische Universität Braunschweig, Germany Thu-M-O-1-3, Time: 11:40

Stream fusion, also known as system combination, is a common technique in automatic speech recognition for traditional hybrid hidden Markov model approaches, yet mostly unexplored for modern deep neural network end-to-end model architectures. Here, we investigate various fusion techniques for the all-attention-based encoder-decoder architecture known as the transformer, striving to achieve optimal fusion by investigating different fusion levels in an example single-microphone setting with fusion of standard magnitude and phase features. We introduce a novel multi-encoder learning method that performs a weighted combination of two encoder-decoder multi-head attention outputs *only* during training. Employing then only the magnitude feature encoder in inference, we are able to show consistent improvement on Wall Street Journal (WSJ) with language model and on Librispeech, without increase in runtime or parameters. Combining two such multi-encoder trained models by a simple late fusion in inference, we achieve state-of-the-art performance for transformer-based models on WSJ with a significant WER reduction of 19% relative compared to the current benchmark approach.

Conditional Independence for Pretext Task Selection in Self-Supervised Speech Representation Learning

Salah Zaiem¹, Titouan Parcollet², Slim Essid¹; ¹LTCI (UMR 5141), France; ²LIA (EA 4128), France Thu-M-O-1-4, Time: 12:00

Through solving pretext tasks, self-supervised learning (SSL) leverages unlabeled data to extract useful latent representations replacing traditional input features in the downstream task. A common pretext task consists in pretraining a SSL model on pseudo-labels derived from the original signal. This technique is particularly relevant for speech data where various meaningful signal processing features may serve as pseudo-labels. However, the process of selecting pseudo-labels, for speech or other types of data, remains mostly unexplored and currently relies on observing the results on the final downstream task. Nevertheless, this methodology is not sustainable at scale due to substantial computational (hence carbon) costs. Thus, this paper introduces a practical and theoretical framework to select relevant pseudo-labels with respect to a given downstream task. More precisely, we propose a functional estimator of the pseudo-label utility grounded in the conditional independence theory, which does not require any training. The experiments conducted on speaker recognition and automatic speech recognition validate our estimator, showing a significant correlation between the performance observed on the downstream task and the utility estimates obtained with our approach, facilitating the prospection of relevant pseudo-labels for self-supervised speech representation learning.

Investigating Methods to Improve Language Model Integration for Attention-Based Encoder-Decoder ASR Models

Mohammad Zeineldeen, Aleksandr Glushko, Wilfried Michel, Albert Zeyer, Ralf Schlüter, Hermann Ney; RWTH Aachen University, Germany Thu-M-O-1-5, Time: 12:20

Attention-based encoder-decoder (AED) models learn an implicit internal language model (ILM) from the training transcriptions. The integration with an external LM trained on much more unpaired text usually leads to better performance. A Bayesian interpretation as in the hybrid autoregressive transducer (HAT) suggests dividing by the prior of the discriminative acoustic model, which corresponds to this implicit LM, similarly as in the hybrid hidden Markov model approach. The implicit LM cannot be calculated efficiently in general and it is yet unclear what are the best methods to estimate it. In this work, we compare different approaches from the literature and propose several novel methods to estimate the ILM directly from the AED model. Our proposed methods outperform all previous approaches. We also investigate other methods to suppress the ILM mainly by decreasing the capacity of the AED model, limiting the label context, and also by training the AED model together with a pre-existing LM.

Comparing CTC and LFMMI for Out-of-Domain Adaptation of wav2vec 2.0 Acoustic Model

Apoorv Vyas, Srikanth Madikeri, Hervé Bourlard; Idiap Research Institute, Switzerland Thu-M-O-1-6, Time: 12:40

In this work, we investigate if the wav2vec 2.0 self-supervised pretraining helps mitigate the overfitting issues with connectionist temporal classification (CTC) training to reduce its performance gap with flat-start lattice-free MMI (E2E-LFMMI) for automatic speech recognition with limited training data. Towards that objective, we use the pretrained wav2vec 2.0 BASE model and fine-tune it on three different datasets including out-of-domain (Switchboard) and cross-lingual (Babel) scenarios. Our results show that for supervised adaptation of the wav2vec 2.0 model, both E2E-LFMMI and CTC achieve similar results; significantly outperforming the baselines trained only with supervised data. Fine-tuning the wav2vec 2.0 model with E2E-LFMMI and CTC we obtain the following relative WER improvements over the supervised baseline trained with E2E-LFMMI. We get relative improvements of 40% and 44% on the clean-set and 64% and 58% on the test set of Librispeech (100h) respectively. On Switchboard (300h) we obtain relative improvements of 33% and 35% respectively. Finally, for Babel languages, we obtain relative improvements of 26% and 23% on Swahili (38h) and 18% and 17% on Tagalog (84h) respectively.

Thu-M-O-2: Emotion and Sentiment Analysis I

Room C, 11:00–13:00, Thursday 2 September 2021 Chairs: Ioana Vasilescu and Nicole Holliday

Speaker Attentive Speech Emotion Recognition

Clément Le Moine, Nicolas Obin, Axel Roebel; STMS (UMR 9912), France

Thu-M-O-2-1, Time: 11:00

Speech Emotion Recognition (SER) task has known significant improvements over the last years with the advent of Deep Neural Networks (DNNs). However, even the most successful methods are still rather failing when adaptation to specific speakers and scenarios is needed, inevitably leading to poorer performances when compared to humans. In this paper, we present novel work based on the idea of teaching the emotion recognition network about speaker identity. Our system is a combination of two ACRNN classifiers respectively dedicated to speaker and emotion recognition. The first informs the latter through a Self Speaker Attention (SSA) mechanism that is shown to considerably help to focus on emotional information of the speech signal. Speaker-dependant experiments on social attitudes database Att-HACK and IEMOCAP corpus demonstrate the effectiveness of the proposed method and achieve the state-of-the-art performance in terms of unweighted average recall.

Separation of Emotional and Reconstruction Embeddings on Ladder Network to Improve Speech Emotion Recognition Robustness in Noisy Conditions

Seong-Gyun Leem¹, Daniel Fulford², Jukka-Pekka Onnela³, David Gard⁴, Carlos Busso¹; ¹University of Texas at Dallas, USA; ²Boston University, USA; ³Harvard University, USA; ⁴San Francisco State University, USA Thu-M-O-2-2, Time: 11:20

When *speech emotion recognition* (SER) is applied in an actual application, the system should be able to cope with audio acquired in a noisy, unconstrained environment. Most studies on noise-robust SER require a parallel dataset with emotion labels, which is impractical to collect, or use speech with artificially added noise, which does not resemble practical conditions. This study builds upon the ladder network formulation, which can effectively compensate the environmental differences between a clean speech corpus and real-life recordings. This study proposes a decoupled ladder network, which increases the robustness of the SER system against the influences of non-stationary background noise by decoupling the last hidden layer embedding into emotion and reconstruction embeddings. This novel implementation allows the emotion embedding to focus exclusively on building a discriminative representation, without worrying about the reconstruction task. We introduce a noisy version of the MSP-Podcast database, which contains audio segments collected with a smartphone that simultaneously records sentences from the corpus and non-stationary noise at different signal-to-noise ratios (SNRs). We test the effectiveness of our proposed model with this corpus, showing that the decoupled ladder network can increase the performance of the regular ladder network when dealing with noisy recordings.

M³: MultiModal Masking Applied to Sentiment Analysis

Efthymios Georgiou, Georgios Paraskevopoulos, Alexandros Potamianos; NTUA, Greece Thu-M-O-2-3, Time: 11:40

A common issue when training multimodal architectures is that not all modalities contribute equally to the model's prediction and the network tends to over-rely on the strongest modality. In this work, we present M^3 , a training procedure based on modality masking for deep multimodal architectures. During network training, we randomly select one modality and mask its features, forcing the model to make its prediction in the absence of this modality. This structured regularization allows the network to better exploit complementary information in input modalities. We implement M^3 as a generic layer that can be integrated with any multimodal architecture. Our experiments show that M^3 outperforms other masking schemes and improves performance for our strong baseline. We evaluate M^3 for multimodal sentiment analysis on CMU-MOSEI, achieving results comparable to the state-of-the-art.

Thu-M-O-3: Linguistic Components in End-to-End ASR

Room D, 11:00–13:00, Thursday 2 September 2021 Chairs: Petr Schwarz and Thomas Schaaf

The CSTR System for Multilingual and Code-Switching ASR Challenges for Low Resource Indian Languages

Ondřej Klejch, Electra Wallington, Peter Bell; University of Edinburgh, UK

Thu-M-O-3-1, Time: 11:00

This paper describes the CSTR submission to the Multilingual and Code-Switching ASR Challenges at Interspeech 2021. For the multilingual track of the challenge, we trained a multilingual CNN-TDNN acoustic model for Gujarati, Hindi, Marathi, Odia, Tamil and Telugu and subsequently fine-tuned the model on monolingual training data. A language model built on a mixture of training and CommonCrawl data was used for decoding. We also demonstrate that crawled data from YouTube can be successfully used to improve the performance of the acoustic model with semi-supervised training. These models together with confidence based language identification achieve the average WER of 18.1%, a 41% relative improvement compared to the provided multilingual baseline model. For the code-switching track

of the challenge we again train a multilingual model on Bengali and Hindi technical lectures and we employ a language model trained on CommonCrawl Bengali and Hindi data mixed with in-domain English data, using a novel transliteration method to generate pronunciations for the English terms. The final model improves by 18% and 34% relative compared to our multilingual baseline. Both our systems were among the top-ranked entries to the challenge.

Acoustic Data-Driven Subword Modeling for End-to-End Speech Recognition

Wei Zhou, Mohammad Zeineldeen, Zuoyun Zheng, Ralf Schlüter, Hermann Ney; RWTH Aachen University, Germany

Thu-M-O-3-2, Time: 11:20

Subword units are commonly used for end-to-end automatic speech recognition (ASR), while a fully acoustic-oriented subword modeling approach is somewhat missing. We propose an acoustic data-driven subword modeling (ADSM) approach that adapts the advantages of several text-based and acoustic-based subword methods into one pipeline. With a fully acoustic-oriented label design and learning process, ADSM produces acoustic-structured subword units and acoustic-matched target sequence for further ASR training. The obtained ADSM labels are evaluated with different end-to-end ASR approaches including CTC, RNN-Transducer and attention models. Experiments on the LibriSpeech corpus show that ADSM clearly outperforms both byte pair encoding (BPE) and pronunciation-assisted subword modeling (PASM) in all cases. Detailed analysis shows that ADSM achieves acoustically more logical word segmentation and more balanced sequence length, and thus, is suitable for both time-synchronous and label-synchronous models. We also briefly describe how to apply acoustic-based subword regularization and unseen text segmentation using ADSM.

Equivalence of Segmental and Neural Transducer Modeling: A Proof of Concept

Wei Zhou, Albert Zeyer, André Merboldt, Ralf Schlüter, Hermann Ney; RWTH Aachen University, Germany Thu-M-O-3-3, Time: 11:40

With the advent of direct models in automatic speech recognition (ASR), the formerly prevalent frame-wise acoustic modeling based on hidden Markov models (HMM) diversified into a number of modeling architectures like encoder-decoder attention models, transducer models and segmental models (direct HMM). While transducer models stay with a frame-level model definition, segmental models are defined on the level of label segments directly. While (soft-)attention-based models avoid explicit alignment, transducer and segmental approach internally do model alignment, either by segment hypotheses or, more implicitly, by emitting so-called blank symbols. In this work, we prove that the widely used class of RNN-Transducer models and segmental models (direct HMM) are equivalent and therefore show equal modeling power. It is shown that blank probabilities translate into segment length probabilities and vice versa. In addition, we provide initial experiments investigating decoding and beam-pruning, comparing time-synchronous and label-/segment-synchronous search strategies and their properties using the same underlying model.

Modeling Dialectal Variation for Swiss German Automatic Speech Recognition

Abbas Khosravani¹, Philip N. Garner¹, Alexandros Lazaridis²; ¹Idiap Research Institute, Switzerland; ²Swisscom, Switzerland Thu-M-O-3-4, Time: 12:00

We describe a speech recognition system for Swiss German, a dialectal spoken language in German-speaking Switzerland. Swiss German has no standard orthography, with a significant variation in its written form. To alleviate the uncertainty associated with this variability, we automatically generate a lexicon from which multiple written forms of a given word in any dialect can be generated. The lexicon is built from a small (incomplete) handcrafted lexicon designed by linguistic experts and contains forms of common words in various Swiss German dialects. We exploit the powerful speech representation of self-supervised acoustic pre-training (wav2vec) to address the low-resource nature of the spoken dialects. The proposed approach results in an overall relative improvement of 9% word error rate compared to one based on an expert-generated lexicon for our TV Box voice assistant application.

Out-of-Vocabulary Words Detection with Attention and CTC Alignments in an End-to-End ASR System

Ekaterina Egorova, Hari Krishna Vydana, Lukáš Burget, Jan Černocký; Brno University of Technology, Czechia Thu-M-O-3-5, Time: 12:20

This work explores the effectiveness of detecting positions of outof-vocabulary words (OOVs) in a decoded utterance using attention weights and CTC per-frame outputs of an end-to-end system predicting word sequences. We show that the end-to-end approach can be effective for the task of OOV detection. CTC alignments are shown to provide better temporal information about the positions of OOV words than attention, and therefore are more suitable for the task. The detected positions of OOV occurrences are utilized for the recurrent OOV recovery task in which probabilistic representations of the pronunciations of the detected OOVs are clustered in order to find repeating words. Improved detection results are shown to correlate with better performance of the recovery of recurrent OOVs.

Training Hybrid Models on Noisy Transliterated Transcripts for Code-Switched Speech Recognition

Matthew Wiesner¹, Mousmita Sarma², Ashish Arora¹, Desh Raj¹, Dongji Gao¹, Ruizhe Huang¹, Supreet Preet², Moris Johnson², Zikra Iqbal², Nagendra Goel², Jan Trmal¹, Leibny Paola García Perera¹, Sanjeev Khudanpur¹; ¹Johns Hopkins University, USA; ²GoVivace, USA Thu-M-O-3-6, Time: 12:40

Thu-M-O-3-6, Time: 12:40

In this paper, we describe the JHU-GoVivace submission for subtask 2 (code-switching task) of the Multilingual and Code-switching ASR challenges for low resource Indian languages. We built a hybrid HMM-DNN system with several improvements over the provided baseline in terms of lexical, language, and acoustic modeling. For lexical modeling, we investigate using unified pronunciations and phonesets derived from the baseline lexicon and publicly available Wikipron lexicons in Bengali and Hindi to expand the pronunciation lexicons. We explore several neural network architectures, along with supervised pretraining and multilingual training for acoustic modeling. We also describe how we used large externally crawled web text for language modeling. Since the challenge data contain artefacts such as misalignments, various data cleanup methods are

explored, including acoustic-driven pronunciation learning to help discover Indian-accented pronunciations for English words as well as transcribed punctuation. As a result of these efforts, our best systems achieve transliterated WERs of 19.5% and 23.2% on the non-duplicated development sets for Hindi-English and Bengali-English, respectively.

Thu-M-V-1: Assessment of Pathological Speech and Language II

11:00-13:00, Thursday 2 September 2021 Chairs: Marzieh Razavi and Fei Chen

Speech Intelligibility of Dysarthric Speech: Human Scores and Acoustic-Phonetic Features

Wei Xue, Roeland van Hout, Fleur Boogmans, Mario Ganzeboom, Catia Cucchiarini, Helmer Strik; Radboud Universiteit, The Netherlands Thu-M-V-1-1, Time: 11:00

We investigated speech intelligibility in dysarthric and non-dysarthric speakers as measured by two commonly used metrics, ratings through the Visual Analogue Scale (VAS) and word accuracy (AcW) through orthographic transcriptions. To gain a better understanding of how acoustic-phonetic correlates could be employed to obtain more objective measures of speech intelligibility and a better classification of dysarthric and non-dysarthric speakers, we studied the relation between these measures of intelligibility and some important acoustic-phonetic correlates. We found that the two intelligibility measures are related, but distinct, and that they might refer to different components of the intelligibility construct. The acoustic-phonetic features showed no difference in the mean values between the two speaker types at the utterance level, but more than half of them played a role in classifying the two speaker types. We computed an acoustic-phonetic probability index (API) at the speaker level. API is moderately correlated to VAS ratings but not correlated to AcW. In addition, API and VAS complement each other in classifying dysarthric and non-dysarthric speakers. This suggests that the intelligibility measures assigned by human raters and acoustic-phonetic features relate to different constructs of intelligibility.

Analyzing Short Term Dynamic Speech Features for Understanding Behavioral Traits of Children with Autism Spectrum Disorder

Young-Kyung Kim¹, Rimita Lahiri¹, Md. Nasir², So Hyun Kim³, Somer Bishop⁴, Catherine Lord⁵, Shrikanth S. Narayanan¹; ¹University of Southern California, USA; ²Microsoft, USA; ³Cornell University, USA; ⁴University of California at San Francisco, USA; ⁵University of California at Los Angeles, USA Thu-M-V-1-2, Time: 11:00

Computational methodologies have shown promise in advancing diagnostic and intervention research in the domain of *Autism Spectrum Disorder (ASD)*. Prior works have investigated speech features to assess disorder severity and also to differentiate between children with and without an ASD diagnosis. In this work, we explore short term dynamic functionals of speech features both within and across speakers to understand if local changes in speech provide information toward phenotyping of ASD.We compare the contributions of static and dynamic functionals representing conversational speech toward the clinical diagnosis state. Our results show that predictions obtained from a combination of dynamic

and static functionals have comparable or superior performance to the predictions obtained from just static speech functionals. We also analyze the relationship between speech production and ASD diagnosis through correlation analyses between speech functionals and manually-derived behavioral codes related to autism severity. The experimental results support the notion that dynamic speech functionals capture complementary information which can facilitate enriched analysis of clinically-meaningful behavioral inference tasks.

Vocalization Recognition of People with Profound Intellectual and Multiple Disabilities (PIMD) Using Machine Learning Algorithms

Waldemar Jęśko; PSNC, Poland Thu-M-V-1-3, Time: 11:00

We investigate vocalization recognition for people with Profound Intellectual and Multiple Disabilities using various machine learning algorithms. The amount of training data available for people with PIMD is typically significantly limited. Due to this fact, data augmentation process was used. Various types of Machine Learning algorithms were tested: k-NN, NB, DT, RDF, MLP and LSTM. During research we also tested various regularization techniques to improve recognition performance. The best results were obtained in case of MLP network with dropout and batch normalization: 90%.

Phonetic Complexity, Speech Accuracy and Intelligibility Assessment of Italian Dysarthric Speech

Barbara Gili Fivela¹, Vincenzo Sallustio², Silvia Pede², Danilo Patrocinio³; ¹Università del Salento, Italy; ²ASL Lecce, Italy; ³Università Cattolica del Sacro Cuore, Italy Thu-M-V-1-4, Time: 11:00

Intelligibility is the degree to which the speech of a person may be understood by a listener, and is related to functional limitation and disability. In protocols for the clinical assessment of dysarthria, intelligibility checks are included, as well as evaluations of speech accuracy, which is more directly related to the disease severity. However, both evaluations are usually based on subjective ratings.

Aim of this work is checking the correlation between intelligibility judgements, subjectively assigned as it may be the case in clinical procedures, and acoustic measures related to linguistically contrasting units. Two novelties characterize this work: a) acoustic measurements considered in the paper relate to both segments (vowel and consonants) and prosodic-intonational phonological events (e.g., pitch accents), that is linguistically relevant speech units; b) contexts of increasing phonetic-phonological complexity are considered, in order for the phonetic characteristics to challenge production accuracy, possibly affecting the realization of phonological features and intelligibility. Increasing complexity is expected to challenge intelligibility indeed and to have an impact on the correlation between intelligibility rates and acoustic measures. Results are preliminary, but confirm both 1) the correlation between acoustic measures of linguistically relevant events and speech intelligibility, as for both the segmental and the prosodic-intonational level, and 2) the role of increasing phonetic-phonological complexity in enhancing the above mentioned correlation.

Detection of Consonant Errors in Disordered Speech Based on Consonant-Vowel Segment Embedding

Si-Ioi Ng, Cymie Wing-Yee Ng, Jingyu Li, Tan Lee; CUHK, China

Thu-M-V-1-5, Time: 11:00

Speech sound disorder (SSD) refers to a type of developmental disorder in young children who encounter persistent difficulties in producing certain speech sounds at the expected age. Consonant
errors are the major indicator of SSD in clinical assessment. Previous studies on automatic assessment of SSD revealed that detection of speech errors concerning short and transitory consonants is less satisfactory. This paper investigates a neural network based approach to detecting consonant errors in disordered speech using consonant-vowel (CV) diphone segment in comparison to using consonant monophone segment. The underlying assumption is that the vowel part of a CV segment carries important information of co-articulation from the consonant. Speech embeddings are extracted from CV segments by a recurrent neural network model. The similarity scores between the embeddings of the test segment and the reference segments are computed to determine if the test segment is the expected consonant or not. Experimental results show that using CV segments achieves improved performance on detecting speech errors concerning those "difficult" consonants reported in the previous studies.

Assessing Posterior-Based Mispronunciation Detection on Field-Collected Recordings from Child Speech Therapy Sessions

Adam Hair¹, Guanlong Zhao¹, Beena Ahmed², Kirrie J. Ballard³, Ricardo Gutierrez-Osuna¹; ¹Texas A&M University, USA; ²UNSW Sydney, Australia; ³University of Sydney, Australia Thu-M-V-1-6, Time: 11:00

A critical component of child speech therapy is home practice with a caregiver, who can provide feedback. However, caregivers oftentimes struggle with accurately rating speech and with perceiving pronunciation errors. One potential solution for this issue is to embed automatic mispronunciation-detection (MPD) algorithms within digital speech therapy applications. To address the need for MPD within child speech therapy, we investigated posterior-based mispronunciation detection using a custom corpus of disordered speech from children that had been manually annotated by an expert clinician. Namely, we trained a family of phoneme-specific logistic regression classifiers (LRC) and support vector machines (SVM) on log posterior probability and log posterior ratio features. Our results show that these classifiers outperformed baseline Goodness of Pronunciation scoring by 11% and 10%, respectively. Even more importantly, in an offline test, the LRC and SVM classifiers outperformed student clinicians at identifying mispronunciations by 18% and 16%, respectively. These results suggest that posterior-based mispronunciation detection may be suitable to provide at-home therapy feedback for children.

Identifying Cognitive Impairment Using Sentence Representation Vectors

Bahman Mirheidari, Yilin Pan, Daniel Blackburn, Ronan O'Malley, Heidi Christensen; University of Sheffield, UK Thu-M-V-1-7, Time: 11:00

The widely used word vectors can be extended at the sentence level to perform a wide range of natural language processing (NLP) tasks. Recently the Bidirectional Encoder Representations from Transformers (BERT) language representation achieved state-of-the-art performance for these applications. The model is trained with punctuated and well-formed (writ-ten) text, however, the performance of the model drops significantly when the input text is the — erroneous and unpunctuated — output of automatic speech recognition (ASR). We use a sliding window and averaging approach for pre-processing text for BERT to extract features for classifying three diagnostic categories relating to cognitive impairment: neurodegenerative dis-order (ND), mild cognitive impairment (MCI), and healthy controls (HC). The in-house dataset contains the audio recordings of an intelligent virtual agent (IVA) who asks the participants several

NOTES

conversational questions prompts in addition to giving a picture description prompt. For the three-way classification, we achieve a 73.88% F-score (accuracy: 76.53%) using the pre-trained, uncased base BERT and for the two-way classifier (HC vs. ND) we achieve 89.80% (accuracy: 90%). We further improve these by using a prompt selection technique, reaching the F-scores of 79.98% (accuracy: 81.63%) and 93.56% (accuracy: 93.75%) respectively.

Parental Spoken Scaffolding and Narrative Skills in Crowd-Sourced Storytelling Samples of Young Children

Zhengjun Yue¹, Jon Barker¹, Heidi Christensen¹, Cristina McKean², Elaine Ashton², Yvonne Wren³, Swapnil Gadgil⁴, Rebecca Bright⁴; ¹University of Sheffield, UK; ²Newcastle University, UK; ³North Bristol NHS Trust, UK; ⁴Therapy Box, UK Thu-M-V-1-8, Time: 11:00

A novel crowdsourcing project to gather children's storytelling based language samples using a mobile app was undertaken across the United Kingdom. Parents' scaffolding of children's narratives was observed in many of the samples. This study was designed to examine the relationship of scaffolding and young children's narrative language ability in a story retell context which is analysed at the macro-structural (total macro-structure score), the micro-structural (mean length of utterances in morphemes) and verbal productivity (total number of utterances) levels. Young children with and without scaffolding were statistically compared. The interaction between the level of scaffolding support, the grammar complexity and the narrative structure was explored. A bidirectional relationship was observed between scaffolding and young children's narrative language ability. Young children with better performance were observed to receive less scaffolding from parents. Scaffolding was shown to support early narrative development of young children and was more able to benefit those with low-level grammatical complexity skills. It is crucial to encourage parental scaffolding to be well-attuned to the child's narrative ability.

Uncertainty-Aware COVID-19 Detection from Imbalanced Sound Data

Tong Xia, Jing Han, Lorena Qendro, Ting Dang, Cecilia Mascolo; University of Cambridge, UK

Thu-M-V-1-9, Time: 11:00

Recently, sound-based COVID-19 detection studies have shown great promise to achieve scalable and prompt digital pre-screening. However, there are still two unsolved issues hindering the practice. First, collected datasets for model training are often imbalanced, with a considerably smaller proportion of users tested positive, making it harder to learn representative and robust features. Second, deep learning models are generally overconfident in their predictions. Clinically, false predictions aggravate healthcare costs. Estimation of the uncertainty of screening would aid this. To handle these issues, we propose an ensemble framework where multiple deep learning models for sound-based COVID-19 detection are developed from different but balanced subsets from original data. As such, data are utilized more effectively compared to traditional up-sampling and down-sampling approaches: an AUC of 0.74 with a sensitivity of 0.68 and a specificity of 0.69 is achieved. Simultaneously, we estimate uncertainty from the disagreement across multiple models. It is shown that false predictions often yield higher uncertainty, enabling us to suggest the users with certainty higher than a threshold to repeat the audio test on their phones or to take clinical tests if digital diagnosis still fails. This study paves the way for a more robust sound-based COVID-19 automated screening system.

Unsupervised Domain Adaptation for Dysarthric Speech Detection via Domain Adversarial Training and Mutual Information Minimization

Disong Wang¹, Liqun Deng², Yu Ting Yeung², Xiao Chen², Xunying Liu¹, Helen Meng¹; ¹CUHK, China; ²Huawei Technologies, China

Thu-M-V-1-10, Time: 11:00

Dysarthric speech detection (DSD) systems aim to detect characteristics of the neuromotor disorder from speech. Such systems are particularly susceptible to domain mismatch where the training and testing data come from the source and target domains respectively, but the two domains may differ in terms of speech stimuli, disease etiology, etc. It is hard to acquire labelled data in the target domain, due to high costs of annotating sizeable datasets. This paper makes a first attempt to formulate cross-domain DSD as an unsupervised domain adaptation (UDA) problem. We use labelled source-domain data and unlabelled target-domain data, and propose a multi-task learning strategy, including dysarthria presence classification (DPC), domain adversarial training (DAT) and mutual information minimization (MIM), which aim to learn dysarthria-discriminative and domain-invariant biomarker embeddings. Specifically, DPC helps biomarker embeddings capture critical indicators of dysarthria; DAT forces biomarker embeddings to be indistinguishable in source and target domains; and MIM further reduces the correlation between biomarker embeddings and domain-related cues. By treating the UASPEECH and TORGO corpora respectively as the source and target domains, experiments show that the incorporation of UDA attains absolute increases of 22.2% and 20.0% respectively in utterance-level weighted average recall and speaker-level accuracy.

Source and Vocal Tract Cues for Speech-Based Classification of Patients with Parkinson's Disease and Healthy Subjects

Tanuka Bhattacharjee¹, Jhansi Mallela¹, Yamini Belur², Nalini Atchayaram², Ravi Yadav², Pradeep Reddy², Dipanjan Gope¹, Prasanta Kumar Ghosh¹; ¹Indian Institute of Science, India; ²NIMHANS, India Thu-M-V-1-11, Time: 11:00

Parkinson's disease (PD) affects both source and vocal tract components of speech. Various speech cues explored in literature for automatic classification of individuals with PD and healthy controls (HC) implicitly carry information about both these components. This work explicitly analyzes the contribution of source and vocal tract attributes toward automatic PD vs. HC classification, which has not been done earlier to the best of our knowledge. Here fundamental frequency (f_0) is used to capture source information. For quantifying vocal tract information, speech waveforms are converted to unvoiced forms and mel-frequency cepstral coefficients (MFCC), denoted by voicing-removed MFCC, are obtained from them. Experimental results suggest that (1) the relative merit of source and vocal tract cues in classifying PD vs. HC largely depends on the speech task being considered, (2) both cues complement each other across all tasks, (3) while MFCC encodes both source and vocal tract features, source information captured by fo is different and further complements MFCC when the classifiers are trained and tested under clean or matched noise conditions, thereby enabling the feature-level fusion of f_0 and MFCC to achieve the best classification accuracy, (4) under unseen noise conditions, fo alone proves to be a highly noise-robust feature

CLAC: A Speech Corpus of Healthy English Speakers

R'mani Haulcy, James Glass; MIT, USA Thu-M-V-1-12, Time: 11:00

This paper introduces the Crowdsourced Language Assessment Corpus (CLAC), a speech corpus consisting of audio recordings and automatically-generated transcripts for several speech and language tasks, as well as metadata for each of the speakers. The CLAC was created to provide the community with a collection of audio samples from various speakers that could be used to learn a general representation for speech from healthy subjects, as well as complement other health-related speech datasets, which tend to be limited. In this paper, we describe the data collection protocol and summarize the contents of the dataset. We also extract timing metrics from the recordings of each task to explore what those metrics look like for a large, English-speaking population. Lastly, we provide an example of how the dataset can be used by comparing the metrics to those extracted from a small sample of Frontotemporal Dementia subjects. We hope that this dataset will help advance the state of the art in the health and speech domain.

Thu-M-V-2: Multimodal Systems

11:00-13:00, Thursday 2 September 2021 Chairs: Jonas Beskow and Helen Meng

Direct Multimodal Few-Shot Learning of Speech and Images

Leanne Nortje, Herman Kamper; Stellenbosch University, South Africa Thu-M-V-2-1, Time: 11:00

We propose direct multimodal few-shot models that learn a shared embedding space of spoken words and images from only a few paired examples. Imagine an agent is shown an image along with a spoken word describing the object in the picture, e.g. pen, book and eraser. After observing a few paired examples of each class, the model is asked to identify the "book" in a set of unseen pictures. Previous work used a two-step indirect approach relying on speech-speech and image-image comparisons across the support set of given speech-image pairs. Instead, we propose two direct models which learn a single multimodal space where inputs from different modalities are directly comparable: a multimodal triplet network (MTriplet) and a multimodal correspondence autoencoder (MCAE). To train these direct models, we *mine* speech-image pairs by using the support set to pair up unlabelled in-domain speech and images. In a speech-to-image digit matching task, direct models outperform indirect models, with the MTriplet achieving the best multimodal five-shot accuracy. We show that the improvements are due to the combination of unsupervised and transfer learning in the direct models, and the absence of two-step compounding errors.

Talk, Don't Write: A Study of Direct Speech-Based Image Retrieval

*Ramon Sanabria*¹, *Austin Waters*², *Jason Baldridge*²; ¹*University of Edinburgh, UK*; ²*Google, USA* Thu-M-V-2-2, Time: 11:00

Speech-based image retrieval has been studied as a proxy for joint representation learning, usually without emphasis on retrieval itself. As such, it is unclear how well speech-based retrieval can work in practice — both in an absolute sense and versus alternative strategies that combine automatic speech recognition (ASR) with strong text encoders. In this work, we extensively study and expand choices of encoder architectures, training methodology (including unimodal and multimodal pretraining), and other factors. Our experiments

cover different types of speech in three datasets: Flickr Audio, Places Audio, and Localized Narratives. Our best model configuration achieves large gains over state of the art, e.g., pushing recall-at-one from 21.8% to 33.2% for Flickr Audio and 27.6% to 53.4% for Places Audio. We also show our best speech-based models can match or exceed cascaded ASR-to-text encoding when speech is spontaneous, accented, or otherwise hard to automatically transcribe.

A Fast Discrete Two-Step Learning Hashing for Scalable Cross-Modal Retrieval

Huan Zhao, Kaili Ma; Hunan University, China Thu-M-V-2-3, Time: 11:00

Recently, some cross-modal hashing methods are proposed to search data for different modality effectively. Hashing has received wide attention because of its low storage and high efficiency. Hashing-based methods project the data instances from different modalities into a Hamming space to learn hash codes for retrieval between different modality. Although obtaining promising performance, hashingbased methods have still several common limitations. First, they learn the hash codes by constructing semantic similarity matrices, resulting in the loss of information. Second, most existing methods simultaneously learn the hash codes and the hash functions, which bring a high computational complexity. Third, they utilize the relaxation-based optimization strategy to generate the hash codes which leads to the large quantization error of the hash codes. To solve the above problems, we propose a novel fast supervised hashing method, termed Fast Discrete Two-Step Learning Hashing (FDTLH) for scalable cross-modal retrieval, which learns the discriminative hash codes by adopting a effective two-step learning scheme. Extensive experiments show that the FDTLH outperforms several state-of-the-art hashing methods in terms of retrieval performance and learning efficiency.

Cross-Modal Knowledge Distillation Method for Automatic Cued Speech Recognition

*Jianrong Wang*¹, *Ziyue Tang*¹, *Xuewei Li*¹, *Mei Yu*¹, *Qiang Fang*², *Li Liu*³; ¹*Tianjin University, China*; ²*CASS, China*; ³*CUHK, China Thu-M-V-2-4, Time: 11:00*

Cued Speech (CS) is a visual communication system for the deaf or hearing impaired people. It combines lip movements with hand cues to obtain a complete phonetic repertoire. Current deep learning based methods on automatic CS recognition suffer from a common problem, which is the data scarcity. Until now, there are only two public single speaker datasets for French (238 sentences) and British English (97 sentences). In this work, we propose a cross-modal knowledge distillation method with teacher-student structure, which transfers audio speech information to CS to overcome the limited data problem. Firstly, we pretrain a teacher model for CS recognition with a large amount of open source audio speech data, and simultaneously pretrain the feature extractors for lips and hands using CS data. Then, we distill the knowledge from teacher model to the student model with frame-level and sequence-level distillation strategies. Importantly, for frame-level, we exploit multi-task learning to weigh losses automatically, to obtain the balance coefficient. Besides, we establish a five-speaker British English CS dataset for the first time. The proposed method is evaluated on French and British English CS datasets, showing superior CS recognition performance to the state-of-the-art (SOTA) by a large margin.

Attention-Based Keyword Localisation in Speech Using Visual Grounding

Kayode Olaleye, Herman Kamper; Stellenbosch University, South Africa Thu-M-V-2-5, Time: 11:00

Visually grounded speech models learn from images paired with spoken captions. By tagging images with soft text labels using a trained visual classifier with a fixed vocabulary, previous work has shown that it is possible to train a model that can *detect* whether a particular text keyword occurs in speech utterances or not. Here we investigate whether visually grounded speech models can also do keyword *localisation*: predicting where, within an utterance, a given textual keyword occurs without any explicit text-based or alignment supervision. We specifically consider whether incorporating attention into a convolutional model is beneficial for localisation. Although absolute localisation performance with visually supervised models is still modest (compared to using unordered bag-of-word text labels for supervision), we show that attention provides a large gain in performance over previous visually grounded models. As in many other speech-image studies, we find that many of the incorrect localisations are due to semantic confusions, e.g. locating the word 'backstroke' for the query keyword 'swimming'.

Evaluation of Audio-Visual Alignments in Visually Grounded Speech Models

Khazar Khorrami, Okko Räsänen; Tampere University, Finland

Thu-M-V-2-6, Time: 11:00

Systems that can find correspondences between multiple modalities, such as between speech and images, have great potential to solve different recognition and data analysis tasks in an unsupervised manner. This work studies multimodal learning in the context of visually grounded speech (VGS) models, and focuses on their recently demonstrated capability to extract spatiotemporal alignments between spoken words and the corresponding visual objects without ever been explicitly trained for object localization or word recognition. As the main contributions, we formalize the alignment problem in terms of an audio-visual alignment tensor that is based on earlier VGS work, introduce systematic metrics for evaluating model performance in aligning visual objects and spoken words, and propose a new VGS model variant for the alignment task utilizing cross-modal attention layer. We test our model and a previously proposed model in the alignment task using SPEECH-COCO captions coupled with MSCOCO images. We compare the alignment performance using our proposed evaluation metrics to the semantic retrieval task commonly used to evaluate VGS models. We show that cross-modal attention layer not only helps the model to achieve higher semantic cross-modal retrieval performance, but also leads to substantial improvements in the alignment performance between image object and spoken words.

Automatic Lip-Reading with Hierarchical Pyramidal Convolution and Self-Attention for Image Sequences with No Word Boundaries

Hang Chen¹, Jun Du¹, Yu Hu¹, Li-Rong Dai¹, Bao-Cai Yin², Chin-Hui Lee³; ¹USTC, China; ²iFLYTEK, China; ³Georgia Tech, USA Thu-M-V-2-7, Time: 11:00

In this paper, we propose a novel deep learning architecture for improving word-level lip-reading. We first incorporate multi-scale processing into spatial feature extraction for lip-reading using hierarchical pyramidal convolution (HPConv) and self-attention.

Specifically, HPConv is proposed to replace the conventional convolution features, leading to an improvement over the model's ability to discover fine-grained lip movements. Next to deal with fixedlength image sequences representing words in a given database, a self-attention mechanism is proposed to integrate local information in all lip frames without assuming known word boundaries, so that our deep models automatically utilize key feature in relevant frames of a given word. Experiments on the Lip Reading in the Wild corpus show that our proposed architecture achieves an accuracy of 86.83%, yielding a relative error rate reduction of about 10% from that obtained with a state-of-the-art scheme of averaging frame scores for information fusion. A detailed analysis of the experimental results also confirms that weights learned from self-attention tend to be zero at both sides of an image sequence and focus non-zero weights in the middle part of a given word.

Cascaded Multilingual Audio-Visual Learning from Videos

Andrew Rouditchenko¹, Angie Boggust¹, David Harwath², Samuel Thomas³, Hilde Kuehne³, Brian Chen⁴, Rameswar Panda³, Rogerio Feris³, Brian Kingsbury³, Michael Picheny⁵, James Glass¹; ¹MIT, USA; ²University of Texas at Austin, USA; ³IBM, USA; ⁴Columbia University, USA; ⁵NYU, USA Thu-M-V-2-8, Time: 11:00

In this paper, we explore self-supervised audio-visual models that learn from instructional videos. Prior work has shown that these models can relate spoken words and sounds to visual content after training on a large-scale dataset of videos, but they were only trained and evaluated on videos in English. To learn multilingual audio-visual representations, we propose a cascaded approach that leverages a model trained on English videos and applies it to audio-visual data in other languages, such as Japanese videos. With our cascaded approach, we show an improvement in retrieval performance of nearly $10 \times$ compared to training on the Japanese videos solely. We also apply the model trained on English videos to Japanese and Hindi spoken captions of images, achieving state-of-the-art performance.

LiRA: Learning Visual Speech Representations from Audio Through Self-Supervision

Pingchuan Ma¹, Rodrigo Mira¹, Stavros Petridis², Björn W. Schuller¹, Maja Pantic¹; ¹Imperial College London, UK; ²Facebook, UK Thu-M-V-2-9, Time: 11:00

The large amount of audiovisual content being shared online today has drawn substantial attention to the prospect of audio-visual self-supervised learning. Recent works have focused on each of these modalities separately, while others have attempted to model both simultaneously in a cross-modal fashion. However, comparatively little attention has been given to leveraging one modality as a training objective to learn from the other. In this work, we propose Learning visual speech Representations from Audio via selfsupervision (LiRA). Specifically, we train a ResNet+Conformer model to predict acoustic features from unlabelled visual speech. We find that this pre-trained model can be leveraged towards word-level and sentence-level lip-reading through feature extraction and fine-tuning experiments. We show that our approach significantly outperforms other self-supervised methods on the Lip Reading in the Wild (LRW) dataset and achieves state-of-the-art performance on Lip Reading Sentences 2 (LRS2) using only a fraction of the total labelled data.

NOTES

End-to-End Audio-Visual Speech Recognition for Overlapping Speech

Richard Rose, Olivier Siohan, Anshuman Tripathi, Otavio Braga; Google, USA Thu-M-V-2-10, Time: 11:00

This paper investigates an end-to-end audio-visual (A/V) modeling approach for transcribing utterances in scenarios where there are overlapping speech utterances from multiple talkers. It assumes that overlapping audio signals and video signals in the form of mouthtracks aligned with speech are available for overlapping talkers. The approach builds on previous work in audio-only multi-talker ASR. In that work, a conventional recurrent neural network transducer (RNN-T) architecture was extended to include a masking model for separation of encoded audio features and multiple label encoders to encode transcripts from overlapping speakers. It is shown here that incorporating an attention weighted combination of visual features in A/V multi-talker RNN-T models significantly improves speaker disambiguation in ASR on overlapping speech relative to audio-only performance. The A/V multi-talker ASR systems described here are trained and evaluated on a two speaker A/V overlapping speech dataset created from YouTube videos. A 17% reduction in WER was observed for A/V multi-talker models relative to audio-only multi-talker models.

Audio-Visual Multi-Talker Speech Recognition in a Cocktail Party

*Yifei Wu*¹, *Chenda Li*¹, *Song Yang*², *Zhongqin Wu*², *Yanmin Qian*¹; ¹*SJTU*, *China*; ²*TAL*, *China* Thu-M-V-2-11, Time: 11:00

Speech from microphones is vulnerable in a complex acoustic environment due to noise and reverberation, while the cameras are not. Thus, utilizing the visual modality in the "cocktail party" scenario with multi-talkers has become a promising and popular approach. In this paper, we have explored the incorporating of visual modality into the end-to-end multi-talker speech recognition task. We propose two methods based on the modality fusion position, which are encoder-based fusion and decoder-based fusion. And for each method, advanced audio-visual fusion techniques including attention mechanism and dual decoder have been explored to find the best usage of the visual modality. With the proposed methods, our best audio-visual multi-talker automatic speech recognition (ASR) model gets almost $\sim 50.0\%$ word error rate (WER) reduction compared to the audio-only multi-talker ASR system.

Thu-M-V-3: Source Separation I

11:00–13:00, Thursday 2 September 2021 Chairs: Keisuke Kinoshita and Min-Jae Hwang

Ultra Fast Speech Separation Model with Teacher Student Learning

Sanyuan Chen¹, Yu Wu², Zhuo Chen³, Jian Wu², Takuya Yoshioka³, Shujie Liu², Jinyu Li³, Xiangzhan Yu¹; ¹Harbin Institute of Technology, China; ²Microsoft, China; ³Microsoft, USA Thu-M-V-3-1, Time: 11:00

Transformer has been successfully applied to speech separation recently with its strong long-dependency modeling capacity using a self-attention mechanism. However, Transformer tends to have heavy run-time costs due to the deep encoder layers, which hinders its deployment on edge devices. A small Transformer model with fewer encoder layers is preferred for computational efficiency, but it is prone to performance degradation. In this paper, an ultra fast speech separation Transformer model is proposed to achieve both better performance and efficiency with teacher student learning (T-S learning). We introduce layer-wise T-S learning and objective shifting mechanisms to guide the small student model to learn intermediate representations from the large teacher model. Compared with the small Transformer model trained from scratch, the proposed T-S learning method reduces the word error rate (WER) by more than 5% for both multi-channel and single-channel speech separation on LibriCSS dataset. Utilizing more unlabeled speech data, our ultra fast speech separation models achieve more than 10% relative WER reduction.

Group Delay Based Re-Weighted Sparse Recovery Algorithms for Robust and High-Resolution Source Separation in DOA Framework

Murtiza Ali, Ashwani Koul, Karan Nathwani; IIT Jammu, India

Thu-M-V-3-2, Time: 11:00

Sparse Recovery (SR) algorithms have been used widely for directionof-arrival (DOA) estimation in spatially contiguous plane wave for their robust performance. But these algorithms have proven to be computationally costly. With a few sensors and at low SNRs, the noise dominates the data singular vectors and the sparse estimation of contiguous sources is incorrect. The magnitude spectrum-based re-weighted sparse recovery (RWSR) algorithms improve the robustness by re-weighting the sparse estimates. However, their efficiency degrades with decreasing the number of sensors at low SNRs. Therefore, this paper exhibits the significance of the phase spectrum, in the form of group-delay, for sparse and robust source estimation using RWSR algorithms for spatially contiguous sources. Further, an optimal re-weighted methodology based on simultaneously minimizing average-root-mean-square-error and maximizing the probability of separation is also proposed. The simulation results are carried out for Gaussian noise to demonstrate the excellent performance of the proposed algorithms.

Continuous Speech Separation Using Speaker Inventory for Long Recording

Cong Han¹, Yi Luo¹, Chenda Li², Tianyan Zhou³, Keisuke Kinoshita⁴, Shinji Watanabe⁵, Marc Delcroix⁴, Hakan Erdogan⁶, John R. Hershey⁶, Nima Mesgarani¹, Zhuo Chen³; ¹Columbia University, USA; ²SJTU, China; ³Microsoft, USA; ⁴NTT, Japan; ⁵Johns Hopkins University, USA; ⁶Google, USA Thu-M-V-3-3, Time: 11:00

Leveraging additional speaker information to facilitate speech separation has received increasing attention in recent years. Recent research includes extracting target speech by using the target speaker's voice snippet and jointly separating all participating speakers by using a pool of additional speaker signals, which is known as speech separation using speaker inventory (SSUSI). However, all these systems ideally assume that the pre-enrolled speaker signals are available and are only evaluated on simple data configurations. In realistic multi-talker conversations, the speech signal contains a large proportion of non-overlapped regions, where we can derive robust speaker embedding of individual talkers. In this work, we adopt the SSUSI model in long recordings and propose a self-informed, clustering-based inventory forming scheme for long recording, where the speaker inventory is fully built from the input signal without the need for external speaker signals. Experiment results on simulated noisy reverberant long recording datasets show that the proposed method can significantly improve the separation performance across various conditions.

Crossfire Conditional Generative Adversarial Networks for Singing Voice Extraction

Weitao Yuan¹, Shengbei Wang¹, Xiangrui Li¹, Masashi Unoki², Wenwu Wang³; ¹Tiangong University, China; ²JAIST, Japan; ³University of Surrey, UK Thu-M-V-3-4, Time: 11:00

Generative adversarial networks (GANs) and Conditional GANs (cGANs) have recently been applied for singing voice extraction (SVE), since they can accurately model the vocal distributions and effectively utilize a large amount of unlabelled datasets. However, current GANs/cGANs based SVE frameworks have no explicit mechanism to eliminate the mutual interferences between different sources. In this work, we introduce a novel 'crossfire' criterion into GANs to complement its standard adversarial training, which forms a dual-objective GANs, namely Crossfire GANs (Cr-GANs). In addition, we design a Generalized Projection Method (GPM) for cGANs based frameworks to extract more effective conditional information for SVE. Using the proposed GPM, we extend our Cr-GANs to conditional version, i.e., Crossfire Conditional GANs (Cr-cGANs). The proposed methods were evaluated on the DSD100 and CCMixter datasets. The numerical results have shown that the 'crossfire' criterion and GPM are beneficial to each other and considerably improve the separation performance of existing GANs/cGANs based SVE methods.

End-to-End Speech Separation Using Orthogonal Representation in Complex and Real Time-Frequency Domain

*Kai Wang*¹, *Hao Huang*¹, *Ying Hu*¹, *Zhihua Huang*¹, *Sheng Li*²; ¹*Xinjiang University, China*; ²*NICT, Japan* Thu-M-V-3-5, Time: 11:00

Traditional single channel speech separation in the time-frequency (T-F) domain often faces the problem of phase reconstruction. Due to the fact that the real-valued network is not suitable for dealing with complex-valued representation, the performance of the T-F domain speech separation method is often constrained from reaching the state-of-the-art. In this paper, we propose improved speech separation methods in both complex and real T-F domain using orthogonal representation. For the complex-valued case, we combine the deep complex network (DCN) and Conv-TasNet to design an end-to-end complex-valued model. Specifically, we incorporate short-time Fourier transform (STFT) and learnable complex layers to build a hybrid encoder-decoder structure, and use a DCN based separator. Then we present the importance of weights orthogonality in the T-F domain transformation and propose a multi-segment orthogonality (MSO) architecture for further improvements. For the real-valued case, we performed separation in real T-F domain by introducing the short-time DCT (STDCT) with orthogonal representation as well. Experimental results show that the proposed complex model outperforms the baseline Conv-TasNet with a comparable parameter size by 1.8 dB, and the STDCT-based real-valued T-F model by 1.2 dB, showing the advantages of speech separation in the T-F domain.

Efficient and Stable Adversarial Learning Using Unpaired Data for Unsupervised Multichannel Speech Separation

Yu Nakagome¹, Masahito Togami², Tetsuji Ogawa¹, Tetsunori Kobayashi¹; ¹Waseda University, Japan; ²LINE, Japan

Thu-M-V-3-6, Time: 11:00

This study presents a framework to enable efficient and stable adversarial learning of unsupervised multichannel source separation models. When the paired data, i.e., the mixture and the correspond-

ing clean speech, are not available for training, it is promising to exploit generative adversarial networks (GANs), where a source separation system is treated as a generator and trained to bring the distribution of the separated (fake) speech closer to that of the clean (real) speech. The separated speech, however, contains many errors, especially when the system is trained unsupervised and can be easily distinguished from the clean speech. A real/fake binary discriminator therefore will stop the adversarial learning process unreasonably early. This study aims to balance the convergence of the generator and discriminator to achieve efficient and stable learning. For that purpose, the autoencoder-based discriminator and more stable adversarial loss, which are designed in boundary equilibrium GAN (BEGAN), are introduced. In addition, generatorspecific distortions are added to real examples so that the models can be trained to focus only on source separation. Experimental comparisons demonstrated that the present stabilizing learning techniques improved the performance of multiple unsupervised source separation systems.

Stabilizing Label Assignment for Speech Separation by Self-Supervised Pre-Training

Sung-Feng Huang¹, Shun-Po Chuang¹, Da-Rong Liu¹, Yi-Chen Chen¹, Gene-Ping Yang², Hung-yi Lee¹; ¹National Taiwan University, Taiwan; ²University of Edinburgh, UK Thu-M-V-3-7, Time: 11:00

Speech separation has been well developed, with the very successful permutation invariant training (PIT) approach, although the frequent label assignment switching happening during PIT training remains to be a problem when better convergence speed and achievable performance are desired. In this paper, we propose to perform self-supervised pre-training to stabilize the label assignment in training the speech separation model. Experiments over several types of self-supervised approaches, several typical speech separation models and two different datasets showed that very good improvements are achievable if a proper self-supervised approach is chosen.

Dual-Path Filter Network: Speaker-Aware Modeling for Speech Separation

Fan-Lin Wang, Yu-Huai Peng, Hung-Shin Lee, Hsin-Min Wang; Academia Sinica, Taiwan Thu-M-V-3-8, Time: 11:00

Speech separation has been extensively studied to deal with the cocktail party problem in recent years. All related approaches can be divided into two categories: time-frequency domain methods and time domain methods. In addition, some methods try to generate speaker vectors to support source separation. In this study, we propose a new model called dual-path filter network (DPFN). Our model focuses on the post-processing of speech separation to improve speech separation performance. DPFN is composed of two parts: the speaker module and the separation module. First, the speaker module infers the identities of the speakers. Then, the separation module uses the speakers' information to extract the voices of individual speakers from the mixture. DPFN constructed based on DPRNN-TasNet is not only superior to DPRNN-TasNet, but also avoids the problem of permutation-invariant training (PIT).

Investigation of Practical Aspects of Single Channel Speech Separation for ASR

Jian Wu¹, Zhuo Chen², Sanyuan Chen¹, Yu Wu¹, Takuya Yoshioka², Naoyuki Kanda², Shujie Liu¹, Jinyu Li²; ¹Microsoft, China; ²Microsoft, USA Thu-M-V-3-9, Time: 11:00

Speech separation has been successfully applied as a front-end processing module of conversation transcription systems thanks to its ability to handle overlapped speech and its flexibility to combine with downstream tasks such as automatic speech recognition (ASR). However, a speech separation model often introduces target speech distortion, resulting in a sub-optimum word error rate (WER). In this paper, we describe our efforts to improve the performance of a single channel speech separation system. Specifically, we investigate a two-stage training scheme that firstly applies a feature level optimization criterion for pre-training, followed by an ASR-oriented optimization criterion using an end-to-end (E2E) speech recognition model. Meanwhile, to keep the model light-weight, we introduce a modified teacher-student learning technique for model compression. By combining those approaches, we achieve a absolute average WER improvement of 2.70% and 0.77% using models with less than 10M parameters compared with the previous state-of-the-art results on the LibriCSS dataset for utterance-wise evaluation and continuous evaluation, respectively.

Implicit Filter-and-Sum Network for End-to-End Multi-Channel Speech Separation

Yi Luo, Nima Mesgarani; Columbia University, USA Thu-M-V-3-10, Time: 11:00

Various neural network architectures have been proposed in recent years for the task of multi-channel speech separation. Among them, the filter-and-sum network (FaSNet) performs end-to-end time-domain filter-and-sum beamforming and has shown effective in both ad-hoc and fixed microphone array geometries. However, whether such explicit beamforming operation is a necessary and valid formulation remains unclear. In this paper, we investigate the beamforming operation and show that it is not necessary. To further improve the performance, we change the explicit waveform-level filter-and-sum operation into an implicit feature-level filter-and-sum operation around a context of features. A feature-level normalized cross correlation (fNCC) feature is also proposed to better match the implicit operation for an improved performance. Experiment results on a simulated ad-hoc microphone array dataset show that the proposed modification to the FaSNet, which we refer to as the implicit filter-and-sum network (iFaSNet), achieve better performance than the explicit FaSNet with a similar model size and a faster training and inference speed.

Generalized Spatio-Temporal RNN Beamformer for Target Speech Separation

Yong Xu¹, Zhuohuang Zhang², Meng Yu¹, Shi-Xiong Zhang¹, Dong Yu¹; ¹Tencent, USA; ²Indiana University, USA

Thu-M-V-3-11, Time: 11:00

Although the conventional mask-based minimum variance distortionless response (MVDR) could reduce the non-linear distortion, the residual noise level of the MVDR separated speech is still high. In this paper, we propose a spatio-temporal recurrent neural network based beamformer (RNN-BF) for target speech separation. This new beamforming framework directly learns the beamforming weights from the estimated speech and noise spatial covariance matrices. Leveraging on the temporal modeling capability of RNNs, the RNN-BF

could automatically accumulate the statistics of the speech and noise covariance matrices to learn the frame-level beamforming weights in a recursive way. An RNN-based generalized eigenvalue (RNN-GEV) beamformer and a more generalized RNN beamformer (GRNN-BF) are proposed. We further improve the RNN-GEV and the GRNN-BF by using layer normalization to replace the commonly used mask normalization on the covariance matrices. The proposed GRNN-BF obtains better performance against prior arts in terms of speech quality (PESQ), speech-to-noise ratio (SNR) and word error rate (WER).

Thu-M-V-4: Speaker Diarization I

11:00–13:00, Thursday 2 September 2021 Chairs: Alicia Lozano-Diez and Jose Patino

End-to-End Neural Diarization: From Transformer to Conformer

*Yi Chieh Liu*¹, *Eunjung Han*², *Chul Lee*², *Andreas Stolcke*²; ¹*Georgia Tech, USA*; ²*Amazon, USA* Thu-M-V-4-1, Time: 11:00

We propose a new end-to-end neural diarization (EEND) system that is based on Conformer, a recently proposed neural architecture that combines convolutional mappings and Transformer to model both local and global dependencies in speech. We first show that data augmentation and convolutional subsampling layers enhance the original self-attentive EEND in the Transformer-based EEND, and then Conformer gives an additional gain over the Transformer-based EEND. However, we notice that the Conformer-based EEND does not generalize as well from simulated to real conversation data as the Transformer-based model. This leads us to quantify the mismatch between simulated data and real speaker behavior in terms of temporal statistics reflecting turn-taking between speakers, and investigate its correlation with diarization error. By mixing simulated and real data in EEND training, we mitigate the mismatch further, with Conformer-based EEND achieving 24% error reduction over the baseline SA-EEND system, and 10% improvement over the best augmented Transformer-based system, on two-speaker CALLHOME data.

Three-Class Overlapped Speech Detection Using a Convolutional Recurrent Neural Network

Jee-weon Jung, Hee-Soo Heo, Youngki Kwon, Joon Son Chung, Bong-Jin Lee; Naver, Korea Thu-M-V-4-2, Time: 11:00

In this work, we propose an overlapped speech detection system trained as a three-class classifier. Unlike conventional systems that perform binary classification as to whether or not a frame contains overlapped speech, the proposed approach classifies into three classes: non-speech, single speaker speech, and overlapped speech. By training a network with the more detailed label definition, the model can learn a better notion on deciding the number of speakers included in a given frame. A convolutional recurrent neural network architecture is explored to benefit from both convolutional layer's capability to model local patterns and recurrent layer's ability to model sequential information. The proposed overlapped speech detection model establishes a state-of-the-art performance with a precision of 0.6648 and a recall of 0.3222 on the DIHARD II evaluation set, showing a 20% increase in recall along with higher precision. In addition, we also introduce a simple approach to utilize the proposed overlapped speech detection model for speaker diarization which ranked third place in the Track 1 of the DIHARD III challenge.

Online Speaker Diarization Equipped with Discriminative Modeling and Guided Inference

Xucheng Wan, Kai Liu, Huan Zhou; Huawei Technologies, China Thu-M-V-4-3, Time: 11:00

Despite considerable efforts, online speaker diarization remains an ongoing challenge. In this study, we propose to tackle the challenge from two perspectives, to endow diarization model with discriminability and to rectify less-reliable online inference with guidance. Specifically, based on the current prior art, UIS-RNN, two enhancement approaches are proposed to concretize our motivations. The effectiveness of our proposals is experimentally validated by results on the AMI evaluation set. With substantial relative improvement of 48.7%, our online speaker diarization system significantly outperformed its baseline. More impressively, its performance in terms of diarization error rate is better than most state-of-the-art offline systems.

Semi-Supervised Training with Pseudo-Labeling for End-To-End Neural Diarization

Yuki Takashima¹, Yusuke Fujita¹, Shota Horiguchi¹, Shinji Watanabe², Leibny Paola García Perera³, Kenji Nagamatsu¹; ¹Hitachi, Japan; ²Carnegie Mellon University, USA; ³Johns Hopkins University, USA Thu-M-V-4-4, Time: 11:00

In this paper, we present a semi-supervised training technique using pseudo-labeling for end-to-end neural diarization (EEND). The EEND system has shown promising performance compared with traditional clustering-based methods, especially in the case of overlapping speech. However, to get a well-tuned model, EEND requires labeled data for all the joint speech activities of every speaker at each time frame in a recording. In this paper, we explore a pseudo-labeling approach that employs unlabeled data. First, we propose an iterative pseudo-label method for EEND, which trains the model using unlabeled data of a target condition. Then, we also propose a committee-based training method to improve the performance of EEND. To evaluate our proposed method, we conduct the experiments of model adaptation using labeled and unlabeled data. Experimental results on the CALLHOME dataset show that our proposed pseudo-label achieved a 37.4% relative diarization error rate reduction compared to a seed model. Moreover, we analyzed the results of semi-supervised adaptation with pseudo-labeling. We also show the effectiveness of our approach on the third DIHARD dataset.

Adapting Speaker Embeddings for Speaker Diarisation

Youngki Kwon, Jee-weon Jung, Hee-Soo Heo, You Jin Kim, Bong-Jin Lee, Joon Son Chung; Naver, Korea Thu-M-V-4-5, Time: 11:00

The goal of this paper is to adapt speaker embeddings for solving the problem of speaker diarisation. The quality of speaker embeddings is paramount to the performance of speaker diarisation systems. Despite this, prior works in the field have directly used embeddings designed only to be effective on the speaker verification task. In this paper, we propose three techniques that can be used to better adapt the speaker embedding aggregation, and non-speech clustering. A wide range of experiments is performed on various challenging datasets. The results demonstrate that all three techniques contribute positively to the performance of the diarisation system achieving an average relative improvement of 25.07% in terms of diarisation error rate over the baseline.

Scenario-Dependent Speaker Diarization for DIHARD-III Challenge

Yu-Xuan Wang¹, Jun Du¹, Maokui He¹, Shu-Tong Niu¹, Lei Sun², Chin-Hui Lee³; ¹USTC, China; ²iFLYTEK, China; ³Georgia Tech, USA Thu-M-V-4-6, Time: 11:00

In this study, we propose a scenario-dependent speaker diarization approach to handling the diversified scenarios of 11 domains encountered in DIHARD-III challenge with a divide-and-conquer strategy. First, using a ResNet-based audio domain classifier, all domains in DIHARD-III challenge could be divided into several scenarios by different impact factors, such as background noise level, speaker number, and speaker overlap ratio. In each scenario, different combinations of techniques are designed, aiming at achieving the best performance in terms of both diarization error rate (DER) and run-time efficiency. For low signal-to-noise-ration (SNR) scenarios, speech enhancement based on a progressive learning network with multiple intermediate SNR targets is adopted for pre-processing. Conventional clustering-based speaker diarization is utilized to mainly handle speech segments with non-overlapping speakers, while separation-based or neural speaker diarization is used to cope with the overlapping speech regions, which is combined with an iterative fine-tuning strategy to boost the generalization ability. We also explore post-processing to perform system fusion and selection. For DIHARD-III challenge, our scenario-dependent system won the first place among all submitted systems, and significantly outperforms the state-of-the-art clustering-based speaker diarization system, yielding relative DER reductions of 32.17% and 28.34% on development set and evaluation set on Track 1, respectively.

End-To-End Speaker Segmentation for Overlap-Aware Resegmentation

*Hervé Bredin*¹, *Antoine Laurent*²; ¹*IRIT (UMR 5505), France;* ²*LIUM (EA 4023), France* Thu-M-V-4-7, Time: 11:00

Speaker segmentation consists in partitioning a conversation between one or more speakers into speaker turns. Usually addressed as the late combination of three sub-tasks (voice activity detection, speaker change detection, and overlapped speech detection), we propose to train an end-to-end segmentation model that does it directly. Inspired by the original end-to-end neural speaker diarization approach (EEND), the task is modeled as a multi-label classification problem using permutation-invariant training. The main difference is that our model operates on short audio chunks (5 seconds) but at a much higher temporal resolution (every 16ms). Experiments on multiple speaker diarization datasets conclude that our model can be used with great success on both voice activity detection and overlapped speech detection. Our proposed model can also be used as a post-processing step, to detect and correctly assign overlapped speech regions. Relative diarization error rate improvement over the best considered baseline (VBx) reaches 17% on AMI, 13% on DIHARD 3 and 13% on VoxConverse

Online Streaming End-to-End Neural Diarization Handling Overlapping Speech and Flexible Numbers of Speakers

Yawen Xue¹, Shota Horiguchi¹, Yusuke Fujita¹, Yuki Takashima¹, Shinji Watanabe², Leibny Paola García Perera³, Kenji Nagamatsu¹; ¹Hitachi, Japan; ²Carnegie Mellon University, USA; ³Johns Hopkins University, USA Thu-M-V-4-8, Time: 11:00

We propose a streaming diarization method based on an end-to-end

neural diarization (EEND) model, which handles flexible numbers of speakers and overlapping speech. In our previous study, the speaker-tracing buffer (STB) mechanism was proposed to achieve a chunk-wise streaming diarization using a pre-trained EEND model. STB traces the speaker information in previous chunks to map the speakers in a new chunk. However, it only worked with two-speaker recordings. In this paper, we propose an extended STB for flexible numbers of speakers, FLEX-STB. The proposed method uses a zeropadding followed by speaker-tracing, which alleviates the difference in the number of speakers between a buffer and a current chunk. We also examine buffer update strategies to select important frames for tracing multiple speakers. Experiments on CALLHOME and DIHARD II datasets show that the proposed method achieves comparable performance to the offline EEND method with 1-second latency. The results also show that our proposed method outperforms recently proposed chunk-wise diarization methods based on EEND (BW-EDA-EEND).

A Thousand Words are Worth More Than One Recording: *Word-Embedding* Based Speaker Change Detection

*Or Haim Anidjar*¹, Itshak Lapidot², Chen Hajaj¹, Amit *Dvir*¹; ¹*Ariel University, Israel*; ²*Afeka College, Israel* Thu-M-V-4-9, Time: 11:00

Speaker Change Detection (SCD) is the task of segmenting an input audio-recording according to speaker interchanges. This task is essential for many applications, such as automatic voice transcription or Speaker Diarization (SD). This paper focuses on the essential task of audio segmentation and suggests a word-embedding-based solution for the SCD problem. Moreover, we show how to use our approach in order to outperform voice-based solutions for the SD problem. We empirically show that our method can accurately identify the speaker-turns in an audio-recording with 82.12% and 89.02% success in the Recall and F1-score measures.

Thu-M-V-5: Speech Synthesis: Prosody Modeling I

11:00–13:00, Thursday 2 September 2021 Chairs: Branislav Gerazov and Mahsa Elyasi

Phrase Break Prediction with Bidirectional Encoder Representations in Japanese Text-to-Speech Synthesis

Kosuke Futamata, Byeongseon Park, Ryuichi Yamamoto, Kentaro Tachibana; LINE, Japan

Thu-M-V-5-1, Time: 11:00

We propose a novel phrase break prediction method that combines implicit features extracted from a pre-trained large language model, a.k.a BERT, and explicit features extracted from BiLSTM with linguistic features. In conventional BiLSTM-based methods, word representations and/or sentence representations are used as independent components. The proposed method takes account of both representations to extract the latent semantics, which cannot be captured by previous methods. The objective evaluation results show that the proposed method obtains an absolute improvement of 3.2 points for the F1 score compared with BiLSTM-based conventional methods using linguistic features. Moreover, the perceptual listening test results verify that a TTS system that applied our proposed method achieved a mean opinion score of 4.39 in prosody naturalness, which is highly competitive with the score of 4.37 for synthesized speech with ground-truth phrase breaks.

Improving Multi-Speaker TTS Prosody Variance with a Residual Encoder and Normalizing Flows

Iván Vallés-Pérez¹, Julian Roth¹, Grzegorz Beringer², Roberto Barra-Chicote¹, Jasha Droppo³; ¹Amazon, UK; ²Amazon, Poland; ³Amazon, USA Thu-M-V-5-2, Time: 11:00

Text-to-speech systems recently achieved almost indistinguishable quality from human speech. However, the prosody of those systems is generally flatter than natural speech, producing samples with low expressiveness. Disentanglement of speaker id and prosody is crucial in text-to-speech systems to improve on naturalness and produce more variable syntheses. This paper proposes a new neural text-to-speech model that approaches the disentanglement problem by conditioning a Tacotron2-like architecture on flow-normalized speaker embeddings, and by substituting the reference encoder with a new learned latent distribution responsible for modeling the intra-sentence variability due to the prosody. By removing the reference encoder dependency, the speaker-leakage problem typically happening in this kind of systems disappears, producing more distinctive syntheses at inference time. The new model achieves significantly higher prosody variance than the baseline in a set of quantitative prosody features, as well as higher speaker distinctiveness, without decreasing the speaker intelligibility. Finally, we observe that the normalized speaker embeddings enable much richer speaker interpolations, substantially improving the distinctiveness of the new interpolated speakers.

Rich Prosody Diversity Modelling with Phone-Level Mixture Density Network

Chenpeng Du, Kai Yu; SJTU, China Thu-M-V-5-3, Time: 11:00

Generating natural speech with a diverse and smooth prosody pattern is a challenging task. Although random sampling with phone-level prosody distribution has been investigated to generate different prosody patterns, the diversity of the generated speech is still very limited and far from what can be achieved by humans. This is largely due to the use of uni-modal distribution, such as single Gaussian, in the prior works of phone-level prosody modelling. In this work, we propose a novel approach that models phone-level prosodies with GMM based mixture density network (GMM-MDN). Experiments on the LJSpeech dataset demonstrate that phone-level prosodies can precisely control the synthetic speech and GMM-MDN can generate a more natural and smooth prosody pattern than a single Gaussian. Subjective evaluations further show that the proposed approach not only achieves better naturalness, but also significantly improves the prosody diversity in synthetic speech without the need of manual control.

Phoneme Duration Modeling Using Speech Rhythm-Based Speaker Embeddings for Multi-Speaker Speech Synthesis

Kenichi Fujita, Atsushi Ando, Yusuke Ijima; NTT, Japan Thu-M-V-5-4, Time: 11:00

This paper proposes a novel speech-rhythm-based method for speaker embeddings. Conventionally spectral feature-based speaker embedding vectors such as the x-vector are used as auxiliary information for multi-speaker speech synthesis. However, speech synthesis with conventional embeddings has difficulty reproducing the target speaker's speech rhythm, one of the important factors among speaker characteristics, because spectral features do not explicitly include speech rhythm. In this paper, speaker embeddings that take speech rhythm information into account are introduced to achieve phoneme duration modeling using a few utterances

NOTES

by the target speaker. A novel point of the proposed method is that rhythm-based embeddings are extracted with phonemes and their durations. They are extracted with a speaker identification model similar to the conventional spectral feature-based one. We conducted two experiments: speaker embeddings generation and speech synthesis with generated embeddings. We show that the proposed model has an EER of 10.3% in speaker identification even with only speech rhythm. Visualizing the embeddings shows that utterances with similar rhythms are also similar in their speaker embeddings. The results of an objective and subjective evaluation on speech synthesis demonstrate that the proposed method can synthesize speech with speech rhythm closer to the target speaker.

Fine-Grained Prosody Modeling in Neural Speech Synthesis Using ToBI Representation

Yuxiang Zou, Shichao Liu, Xiang Yin, Haopeng Lin, Chunfeng Wang, Haoyu Zhang, Zejun Ma; ByteDance, China

Thu-M-V-5-5, Time: 11:00

Benefiting from the great development of deep learning, modern neural text-to-speech (TTS) models can generate speech indistinguishable from natural speech. However, The generated utterances often keep an average prosodic style of the database instead of having rich prosodic variation. For pitch-stressed languages, such as English, accurate intonation and stress are important for conveying semantic information. In this work, we propose a fine-grained prosody modeling method in neural speech synthesis with ToBI (Tones and Break Indices) representation. The proposed system consists of a text frontend for ToBI prediction and a Tacotron-based TTS module for prosody modeling. By introducing the ToBI representation, we can control the system to synthesize speech with accurate intonation and stress at syllable level. Compared with the two baselines (Tacotron and unsupervised method), experiments show that our model can generate more natural speech with more accurate prosody, as well as effectively control the stress, intonation, and pause of the speech.

Intra-Sentential Speaking Rate Control in Neural Text-To-Speech for Automatic Dubbing

Mayank Sharma¹, Yogesh Virkar², Marcello Federico², Roberto Barra-Chicote³, Robert Enyedi²; ¹Amazon, India; ²Amazon, USA; ³Amazon, UK Thu-M-V-5-6, Time: 11:00

Automatically dubbed speech of a video involves: (i) segmenting the target sentences into phrases to reflect the speech-pause arrangement used by the original speaker, and (ii) adjusting the speaking rate of the synthetic voice at the phrase-level to match the exact timing of each corresponding source phrase. In this work, we investigate a post-segmentation approach to control the speaking rate of neural Text-to-Speech (TTS) at the phrase-level after generating the entire sentence. Our post-segmentation method relies on the attention matrix generated by the context generation step to perform a force-alignment over pause markers inserted in the input text. We show that: (i) our approach can be more accurate than applying an off-the-shelf forced aligner, and (ii) post-segmentation approach described in [1].

Applying the Information Bottleneck Principle to Prosodic Representation Learning

*Guangyan Zhang*¹, *Ying Qin*², *Daxin Tan*¹, *Tan Lee*¹; ¹*CUHK*, *China*; ²*Beijing Jiaotong University, China* Thu-M-V-5-7, Time: 11:00

This paper describes a novel design of a neural network-based speech generation model for learning prosodic representation. The problem of representation learning is formulated according to the information bottleneck (IB) principle. A modified VQ-VAE quantized layer is incorporated in the speech generation model to control the IB capacity and adjust the balance between reconstruction power and disentangle capability of the learned representation. The proposed model is able to learn word-level prosodic representations from speech data. With an optimized IB capacity, the learned representations not only are adequate to reconstruct the original speech but also can be used to transfer the prosody onto different textual content. Extensive results of the objective and subjective evaluation are presented to demonstrate the effect of IB capacity control, the effectiveness, and potential usage of the learned prosodic representation in controllable neural speech generation.

A Prototypical Network Approach for Evaluating Generated Emotional Speech

Alice Baird, Silvan Mertes, Manuel Milling, Lukas Stappen, Thomas Wiest, Elisabeth André, Björn W. Schuller; Universität Augsburg, Germany Thu-M-V-5-8, Time: 11:00

The collection of emotional speech data is a time-consuming and costly endeavour. Generative networks can be applied to augment the limited audio data artificially. However, it is challenging to evaluate generated audio for its similarity to source data, as current quantitative metrics are not necessarily suited to the audio domain. We explore the use of a prototypical network to evaluate four classes of generated emotional audio with this in mind. We first extract spectrogram images from WAVEGAN generated audio and other audio augmentation approaches, comparing similarity to the class prototype and diversity within the embedding space. Furthermore, we augment the source training set with each augmentation type and perform a classification to explore the generated audio plausibility. Results suggest that quality and diversity can be quantitatively observed with this approach. In the chosen context, we see that WAVEGAN generated data is recognisable as a source data class (F_1 score 43.6%), and the samples add similar diversity as unseen source data. This result leads to more plausible data for augmentation of the source training set — achieving up to $63.9\% F_1$ which is a 3.5%improvement over the source data baseline.

Thu-M-V-6: Speech Production II

11:00–13:00, Thursday 2 September 2021 Chairs: Amelia Gully and Rita Patel

A Simplified Model for the Vocal Tract of [s] with Inclined Incisors

Tsukasa Yoshinaga¹, Kohei Tada¹, Kazunori Nozaki², Akiyoshi Iida¹; ¹Toyohashi Tech, Japan; ²Osaka University Dental Hospital, Japan Thu-M-V-6-1, Time: 11:00

To examine the effects of inclined incisors on the phonation of [s], a simplified vocal tract model is proposed, and the acoustic characteristics with different maxillary incisor angles are predicted

NOTES

by the model. As a control model, a realistic vocal tract replica of [s] was constructed from medical images, and the angle of the maxillary incisor was changed from the original position up to 30°. The simplified model was constructed with a rectangular flow channel using the average dimensions of the vocal tracts for five Japanese subjects. Both geometries were set in an anechoic chamber, and sounds generated from the geometries were recorded with a microphone. The results showed that amplitudes of the sound generated by the realistic geometry were decreased by increasing the incisor angle, and this tendency agreed well with the simplified model. Moreover, the slope value of the decrease in overall pressure levels estimated by the model was consistent with that of the realistic geometry, indicating the capability of estimating the effects of inclined incisors with dental prostheses on the production of [s] by using the simplified model.

Vocal-Tract Models to Visualize the Airstream of Human Breath and Droplets While Producing Speech

Takayuki Arai; Sophia University, Japan Thu-M-V-6-2, Time: 11:00

Due to the COVID-19 pandemic, visualizing the airstream of human breath during speech production has become extremely important from the viewpoint of preventing infection. In addition, visualizing droplets and the larger drops expelled when we speak consonantal sounds may help for the same reason. One visualization technique is to pass a laser sheet through the droplet cloud produced by a human speaker. However, the laser poses certain health risks for human beings. Therefore, we developed an alternative method to passing a laser against a human body in which we utilize physical models of the human vocal tract. First, we tested a head-shaped model with a lung model from our previous study to visualize the exhaled breath during vowel production (with and without a mask). Then, we implemented an extended version of the anatomical-type vocal-tract model introduced in our previous study. With this newly developed model, lips are made of the same flexible material that was used to form the tongue part in the previous model. We also attached these lips to another previous model for producing sounds including /b/. Finally, the lip models were tested to visualize the droplet cloud including expelled drops present while producing a bilabial plosive sound.

Using Transposed Convolution for Articulatory-to-Acoustic Conversion from Real-Time MRI Data

Ryo Tanji, Hidefumi Ohmura, Kouichi Katsurada; Tokyo University of Science, Japan Thu-M-V-6-3, Time: 11:00

We herein propose a deep neural network-based model for articulatory-to-acoustic conversion from real-time MRI data. Although rtMRI, which can record entire articulatory organs with a high resolution, has an advantage in articulatory-to-acoustic conversion, it has a relatively low sampling rate. To address this, we incorporated the super-resolution technique in the temporal dimension with a transposed convolution. With the use of transposed convolution, the resolution can be increased by applying the inversion process of resolution reduction of a standard CNN. To evaluate the performance on the datasets with different temporal resolutions, we conducted experiments using two datasets: USC-TIMIT and Japanese rtMRI dataset. Results of the experiments performed using mel-cepstrum distortion and PESQ showed that transposed convolution is effective for generating accurate acoustic features. We also confirmed that increasing the magnification of the super-resolution leads to an improvement in the PESQ score.

Comparison Between Lumped-Mass Modeling and Flow Simulation of the Reed-Type Artificial Vocal Fold

Rafia Inaam¹, Tsukasa Yoshinaga¹, Takayuki Arai², Hiroshi Yokoyama¹, Akiyoshi Iida¹; ¹Toyohashi Tech, Japan; ²Sophia University, Japan Thu-M-V-6-4, Time: 11:00

The sound generated by a reed-type artificial vocal fold was predicted by a one-mass modeling and numerical flow simulation to examine the sound generation mechanisms of the artificial vocal fold. For the one-mass modeling, the reed oscillation was modeled with an equivalent spring constant, and the flow rate was estimated by Bernoulli's equation. For the flow simulation, the flow and acoustic fields were predicted with compressible Navier-Stokes Equations. while the reed oscillation was calculated by a one-dimensional beam equation. The experimentation was conducted by measuring the sound of an artificial vocal fold in an anechoic chamber. The results of the acoustic measurement showed that the sound amplitudes in the flow simulation agreed well with the experiment, while the onemass model underestimated the amplitudes in a higher frequency range. Reed displacement and flow rate comparisons indicated that the flow retention in the reed retainer caused the asymmetry in the flow rate waveform, hence producing larger amplitudes for the flow simulation in the higher frequency range. The flow simulation enabled to predict this flow retention which cannot be modeled in the one-dimensional one-mass model, and it is anticipated to apply the flow simulation to develop a better artificial vocal fold.

Inhalations in Speech: Acoustic and Physiological Characteristics

*Raphael Werner*¹, *Susanne Fuchs*², *Jürgen Trouvain*¹, *Bernd Möbius*¹; ¹*Universität des Saarlandes, Germany*; ²*ZAS, Germany*

Thu-M-V-6-5, Time: 11:00

This paper examines the acoustic properties of breath noises in speech pauses in relation to similar speech segments and with regard to their inhalation speed. We measured intensity, center of gravity, and formants, as well as kinematic data (via Respiratory Inductance Plethysmography) for inhalations, aspirations of stops, glottal fricatives, and schwa vowels. We find that inhalations within speech are louder than those initiating speech, share spectral properties (center of gravity) with the aspiration phase of /k/-realizations, and generally involve a more open vocal tract (higher F1) than schwa-realizations. Intensity, center of gravity, and F1 are found to be positively correlated to inhalation speed. Overall, we conclude that jaw openness and inhalation speed are major contributors to inhalation noises in speech pauses.

Model-Based Exploration of Linking Between Vowel Articulatory Space and Acoustic Space

Anqi Xu¹, Daniel van Niekerk¹, Branislav Gerazov², Paul Konstantin Krug³, Santitham Prom-on⁴, Peter Birkholz³, Yi Xu¹; ¹University College London, UK; ²UKiM, Macedonia; ³Technische Universität Dresden, Germany; ⁴KMUTT, Thailand Thu-M-V-6-6, Time: 11:00

While the acoustic vowel space has been extensively studied in previous research, little is known about the high-dimensional articulatory space of vowels. The articulatory imaging techniques are limited to tracking only a few key articulators, leaving the rest of the articulators unmonitored. In the present study, we attempted to develop a detailed articulatory space obtained by training a 3D articulatory synthesizer to learn eleven British English vowels. An analysis-by-synthesis strategy was used to acoustically optimize vocal tract parameters that represent twenty articulatory dimensions. The results show that tongue height and retraction, larynx location and lip roundness are the most perceptually distinctive articulatory dimensions. Yet, even for these dimensions, there is a fair amount of articulatory overlap between vowels, unlike the fine-grained acoustic space. This method opens up the possibility of using modelling to investigate the link between speech production and perception.

Take a Breath: Respiratory Sounds Improve Recollection in Synthetic Speech

Mikey Elmers, Raphael Werner, Beeke Muhlack, Bernd Möbius, Jürgen Trouvain; Universität des Saarlandes, Germany

Thu-M-V-6-7, Time: 11:00

This study revisits Whalen et al. (1995, JASA) by evaluating English speaking participants in a perception experiment to determine if their recollection is affected by including breath noises in sentences generated by a speech synthesis system. Whalen found an improvement in recollection for sentences that were preceded by a breath noise compared to sentences without one. While Whalen and colleagues used formant synthesis to render the English sentences, we use a modern concatenative synthesis system. The present study uses inhalations of three different lengths: 0 ms (no breath noise), 300 ms (short breath noise), and 600 ms (long breath noise). Our results are consistent with Whalen and colleagues for the 600 ms condition, but not for the 300 ms condition, indicating that not all inhalations improved recollection. The present study also found a significant effect for sentence length, illustrating that shorter sentences have higher accuracy for recollection than longer sentences. Overall, the present study indicates that respiratory sounds are important to the recollection of synthesized speech and that researchers should focus on longer and more complex types of speech, such as paragraphs or dialogues, for future studies.

Modeling Sensorimotor Adaptation in Speech Through Alterations to Forward and Inverse Models

*Taijing Chen*¹, *Adam Lammert*², *Benjamin Parrell*¹; ¹*UW–Madison, USA*; ²*Worcester Polytechnic Institute,*

USA

Thu-M-V-6-8, Time: 11:00

When speakers are exposed to auditory feedback perturbations of a particular vowel, they not only adapt their productions of that vowel but also transfer this change to other, untrained, vowels. However, current models of speech sensorimotor adaptation, which rely on changes in the feedforward control of specific speech units, are unable to account for this type of generalization. Here, we developed a neural-network based model to simulate speech sensorimotor adaptation, and assess whether updates to internal control models can account for observed patterns of generalization. Based on a dataset generated from the Maeda plant, we trained two independent neural networks: 1) an inverse model, which generates motor commands for desired acoustic outcomes and 2) a forward model, which maps motor commands to acoustic outcomes (prediction). When vowel formant perturbations were given, both forward and inverse models were updated when there was a mismatch between predicted and perceived output. Our results replicate behavioral experiments: the model altered its production to counteract the perturbation, and showed gradient transfer of this learning dependent on acoustic distance between training and test vowels. These results suggest that updating paired forward and inverse models provides a plausible account for sensorimotor adaptation in speech.

Mixture of Orthogonal Sequences Made from Extended Time-Stretched Pulses Enables Measurement of Involuntary Voice Fundamental Frequency Response to Pitch Perturbation

Hideki Kawahara¹, Toshie Matsui², Kohei Yatabe³, Ken-Ichi Sakakibara⁴, Minoru Tsuzaki⁵, Masanori Morise⁶, Toshio Irino¹; ¹Wakayama University, Japan; ²Toyohashi Tech, Japan; ³Waseda University, Japan; ⁴HSUH, Japan; ⁵KCUA, Japan; ⁶Meiji University, Japan Thu-M-V-6-9, Time: 11:00

Auditory feedback plays an essential role in the regulation of the fundamental frequency of voiced sounds. The fundamental frequency also responds to auditory stimulation other than the speaker's voice. We propose to use this response of the fundamental frequency of sustained vowels to frequency-modulated test signals for investigating involuntary control of voice pitch. This involuntary response is difficult to identify and isolate by the conventional paradigm, which uses step-shaped pitch perturbation. We recently developed a versatile measurement method using a mixture of orthogonal sequences made from a set of extended time-stretched pulses (TSP). In this article, we extended our approach and designed a set of test signals using the mixture to modulate the fundamental frequency of artificial signals. For testing the response, the experimenter presents the modulated signal aurally while the subject is voicing sustained We developed a tool for conducting this test quickly vowels and interactively. We make the tool available as an open-source and also provide executable GUI-based applications. Preliminary tests revealed that the proposed method consistently provides compensatory responses with about 100 ms latency, representing involuntary control. Finally, we discuss future applications of the proposed method for objective and non-invasive auditory response measurements.

Thu-M-V-7: Spoken Dialogue Systems II

11:00–13:00, Thursday 2 September 2021 Chairs: Ilya Oparin and Dilek Hakkani-Tür

Contextualized Attention-Based Knowledge Transfer for Spoken Conversational Question Answering

*Chenyu You*¹, *Nuo Chen*², *Yuexian Zou*²; ¹*Yale University, USA*; ²*Peking University, China* Thu-M-V-7-1, Time: 11:00

Spoken conversational question answering (SCQA) requires machines to model the flow of multi-turn conversation given the speech utterances and text corpora. Different from traditional text question answering (QA) tasks, SCQA involves audio signal processing, passage comprehension, and contextual understanding. However, ASR systems introduce unexpected noisy signals to the transcriptions, which result in performance degradation on SCQA. To overcome the problem, we propose CADNet, a novel contextualized attentionbased distillation approach, which applies both cross-attention and self-attention to obtain ASR-robust contextualized embedding representations of the passage and dialogue history for performance improvements. We also introduce the spoken conventional knowledge distillation framework to distill the ASR-robust knowledge from the estimated probabilities of the *teacher* model to the *student*. We conduct extensive experiments on the Spoken-CoQA dataset and demonstrate that our approach achieves remarkable performance in this task.

Injecting Descriptive Meta-Information into Pre-Trained Language Models with Hypernetworks

Wenying Duan¹, Xiaoxi He², Zimu Zhou³, Hong Rao¹, Lothar Thiele²; ¹Nanchang University, China; ²ETH Zürich, Switzerland; ³Singapore Management University, Singapore Thu-M-V-7-2, Time: 11:00

Pre-trained language models have been widely adopted as backbones in various natural language processing tasks. However, existing pretrained language models ignore the descriptive meta-information in the text such as the distinction between the title and the mainbody, leading to over-weighted attention to insignificant text. In this paper, we propose a hypernetwork-based architecture to model the descriptive meta-information and integrate it into pre-trained language models. Evaluations on three natural language processing tasks show that our method notably improves the performance of pre-trained language models and achieves the state-of-the-art results on keyphrase extraction.

Causal Confusion Reduction for Robust Multi-Domain Dialogue Policy

Mahdin Rohmatillah, Jen-Tzung Chien; NYCU, Taiwan Thu-M-V-7-3, Time: 11:00

In the multi-domain dialogue system, dialog policy plays an important role since it determines the suitable actions based on the user's goals. However, in many recent works, most of the dialogue optimizations, especially that use reinforcement learning (RL) methods, do not perform well. The main problem is that the initial step of optimization that involves the behavior cloning (BC) methods suffer from the causal confusion problem, which means that the agent misidentifies true cause of an expert action in current state. This paper proposes a novel method to improve the performance of BC method in dialogue system. Instead of only predicting correct action given a state from dataset, we introduce the auxiliary tasks to predict both of current belief state and recent user utterance in order to reduce causal confusion of the expert action in the dataset since those features are important in every dialog turn. Experiments on ConvLab-2 shows that, by using this method, all of RL based optimizations are improved. Furthermore, the agent based on the proximal policy optimization shows very significant improvement with the help of the proposed BC agent weights both in policy evaluation as well as in end-to-end system evaluation.

Timing Generating Networks: Neural Network Based Precise Turn-Taking Timing Prediction in Multiparty Conversation

Shinya Fujie¹, Hayato Katayama², Jin Sakuma², Tetsunori Kobayashi²; ¹Chiba Institute of Technology, Japan; ²Waseda University, Japan Thu-M-V-7-4, Time: 11:00

A brand new neural network based precise timing generation framework, named the Timing Generating Network (TGN), is proposed and applied to turn-taking timing decision problems. Although turn-taking problems have conventionally been formalized as users' end-of-turn detection, this approach cannot estimate the precise timing at which a spoken dialogue system should take a turn to start its utterance. Since several conventional approaches estimate precise timings but the estimation executed only at/after the end of preceding user's utterance, they highly depend on the accuracy of intermediate decision modules, such as voice activity detection, etc. The advantages of the TGN are that its parameters are tunable via error backpropagation as it is described in a differentiable form

as a whole, and it is free from inter-module error propagation as it has no deterministic intermediate modules. The experimental results show that the proposed system is superior to a conventional turn-taking system that adopts the hard decisions on user's voice activity detection and response time estimation.

Human-to-Human Conversation Dataset for Learning Fine-Grained Turn-Taking Action

Kehan Chen, Zezhong Li, Suyang Dai, Wei Zhou, Haiqing Chen; Alibaba, China

Thu-M-V-7-5, Time: 11:00

Conducting natural turn-taking behavior takes a crucial part in the user experience of modern spoken dialogue systems. One way to build such system is to learn those behaviors from real-world humanto-human dialogues, which have the most diverse and fine-grained turn-taking actions than any manual constructed sessions.

In this paper, we propose a Dataset — FTAD which could be used to learn turn-taking policies directly from human. First, we design an annotation mechanism to transform existing human-to-human dialogue session into structural data with most fine-grained turn-taking actions reserved. Then we explored a set of supervised learning tasks on it, showing the challenge and potential of learning complete fine-grained turn-taking policies based on such data.

PhonemeBERT: Joint Language Modelling of Phoneme Sequence and ASR Transcript

Mukuntha Narayanan Sundararaman, Ayush Kumar, Jithendra Vepa; Observe.AI, India Thu-M-V-7-6, Time: 11:00

Recent years have witnessed significant improvement in ASR systems to recognize spoken utterances. However, it is still a challenging task for noisy and out-of-domain data, where ASR errors are prevalent in the transcribed text. These errors significantly degrade the performance of downstream tasks such as intent and sentiment detection. In this work, we propose a BERT-style language model, referred to as PhonemeBERT that learns a joint language model with phoneme sequence and ASR transcript to learn phonetic-aware representations that are robust to ASR errors. We show that PhonemeBERT leverages phoneme sequences as additional features that outperform word-only models on downstream tasks. We evaluate our approach extensively by generating noisy data for three benchmark datasets Stanford Sentiment Treebank, TREC and ATIS for sentiment, question and intent classification tasks respectively in addition to a real-life sentiment dataset. The results of the proposed approach beats the state-of-the-art baselines comprehensively on each dataset. Additionally, we show that PhonemeBERT can also be utilized as a pre-trained encoder in a low-resource setup where we only have ASR-transcripts for the downstream tasks.

Joint Retrieval-Extraction Training for Evidence-Aware Dialog Response Selection

Hongyin Luo¹, James Glass¹, Garima Lalwani², Yi Zhang², Shang-Wen Li²; ¹MIT, USA; ²Amazon, USA Thu-M-V-7-7, Time: 11:00

Neural dialog response selection models infer by scoring each candidate response given the dialog context, and the cross-encoder method yields state-of-the-art (SOTA) results for the task. In the method, the candidate scores are computed by feeding the output embedding of the first token in the input sequence, which is a concatenation of response and context, to a linear layer for making prediction. However, the embeddings of the other tokens in the sequence are not modeled explicitly, and inferring the candidate

NOTES

scores only with the first token makes the result not interpretable. To address the challenge, we propose a Retrieval-EXtraction encoder (REX) for dialog response selection. We augment the existing first-token- or sequence- based retrieval approach with an extraction loss. The loss provides gradient signal from each token during training and allows the model to learn token-level evidence and to select response based on important keywords. We show that REX achieves the new SOTA in the dialog response selection task. Also, our qualitative analysis suggests that REX highlights evidence it infers selections from and makes the inference result interpretable.

Adapting Long Context NLM for ASR Rescoring in Conversational Agents

Ashish Shenoy, Sravan Bodapati, Monica Sunkara, Srikanth Ronanki, Katrin Kirchhoff; Amazon, USA Thu-M-V-7-8, Time: 11:00

Neural Language Models (NLM), when trained and evaluated with context spanning multiple utterances, have been shown to consistently outperform both conventional n-gram language models and NLMs that use limited context. In this paper, we investigate various techniques to incorporate turn based context history into both recurrent (LSTM) and Transformer-XL based NLMs. For recurrent based NLMs, we explore context carry over mechanism and feature based augmentation, where we incorporate other forms of contextual information such as bot response and system dialogue acts as classified by a Natural Language Understanding (NLU) model. To mitigate the sharp nearby, fuzzy far away problem with contextual NLM, we propose the use of attention layer over lexical metadata to improve feature based augmentation. Additionally, we adapt our contextual NLM towards user provided on-the-fly speech patterns by leveraging encodings from a large pre-trained masked language model and performing fusion with a Transformer-XL based NLM. We test our proposed models using N-best rescoring of ASR hypotheses of task-oriented dialogues and also evaluate on downstream NLU tasks such as intent classification and slot labeling. The best performing model shows a relative WER between 1.6% and 9.1% and a slot labeling F1 score improvement of 4% over non-contextual baselines.

Thu-M-SS-1: Oriental Language Recognition

Room Virtual, 11:00–13:00, Thursday 2 September 2021 Chairs: Dong Wang and Qingyang Hong

Oriental Language Recognition (OLR) 2020: Summary and Analysis

*Jing Li*¹, *Binling Wang*¹, *Yiming Zhi*¹, *Zheng Li*¹, *Lin Li*¹, *Qingyang Hong*¹, *Dong Wang*²; ¹*Xiamen University, China*; ²*Tsinghua University, China* Thu-M-SS-1-1, Time: 11:00

The fifth Oriental Language Recognition (OLR) Challenge focuses on language recognition in a variety of complex environments to promote its development. The OLR 2020 Challenge includes three tasks: (1) cross-channel language identification, (2) dialect identification, and (3) noisy language identification. We choose C_{avg} as the principle evaluation metric, and the Equal Error Rate (EER) as the secondary metric. There were 58 teams participating in this challenge and one third of the teams submitted valid results. Compared with the best baseline, the C_{avg} values of Top 1 system for the three tasks were relatively reduced by 82%, 62% and 48%, respectively. This paper describes the three tasks, the database profile, and the final results. We also outline the novel approaches that improve the performance of language recognition systems most significantly, such as the utilization of auxiliary information.

Language Recognition on Unknown Conditions: The LORIA-Inria-MULTISPEECH System for AP20-OLR Challenge

Raphaël Duroselle, Md. Sahidullah, Denis Jouvet, Irina Illina; Loria (UMR 7503), France

Thu-M-SS-1-2, Time: 11:20

We describe the LORIA-Inria-MULTISPEECH system submitted to the Oriental Language Recognition AP20-OLR Challenge. This system has been specifically designed to be robust to unknown conditions: channel mismatch (task 1) and noisy conditions (task 3). Three sets of studies have been carried out for elaborating the system: design of multilingual bottleneck features, selection of robust features by evaluating language recognition performance on an unobserved channel, and design of the final models with different loss functions which exploit channel diversity within the training set. Key factors for robustness to unknown conditions are data augmentation techniques, stochastic weight averaging, and regularization of TDNNs with domain robustness loss functions. The final system is the combination of four TDNNs using bottleneck features and one GMM using SDC-MFCC features. Within the AP20-OLR Challenge, it achieves the top performance for tasks 1 and 3 with a C_{avg} of respectively 0.0239 and 0.0374. This validates the approach for generalization to unknown conditions.

Dynamic Multi-Scale Convolution for Dialect Identification

Tianlong Kong¹, Shouyi Yin¹, Dawei Zhang², Wang Geng², Xin Wang², Dandan Song¹, Jinwen Huang², Huiyu Shi¹, Xiaorui Wang²; ¹Tsinghua University, China; ²Kwai, China Thu-M-SS-1-3, Time: 11:40

Time Delay Neural Networks (TDNN)-based methods are widely used in dialect identification. However, in previous work with TDNN application, subtle variant is being neglected in different feature scales. To address this issue, we propose a new architecture, named dynamic multi-scale convolution, which consists of dynamic kernel convolution, local multi-scale learning, and global multi-scale pooling. Dynamic kernel convolution captures features between short-term and long-term context adaptively. Local multi-scale learning, which represents multi-scale features at a granular level, is able to increase the range of receptive fields for convolution operation. Besides, global multi-scale pooling is applied to aggregate features from different bottleneck layers in order to collect information from multiple aspects. The proposed architecture significantly outperforms state-of-the-art system on the AP20-OLR-dialect-task of oriental language recognition (OLR) challenge 2020, with the best average cost performance (C_{avq}) of 0.067 and the best equal error rate (EER) of 6.52%. Compared with the known best results, our method achieves 9% of C_{avg} and 45% of EER relative improvement, respectively. Furthermore, the parameters of proposed model are 91% fewer than the best known model.

An End-to-End Dialect Identification System with Transfer Learning from a Multilingual Automatic Speech Recognition Model

*Ding Wang*¹, *Shuaishuai* Ye¹, *Xinhui* Hu¹, *Sheng* Li², *Xinkang Xu*¹; ¹*HiThink RoyalFlush, China*; ²*NICT, Japan* Thu-M-SS-1-4, Time: 12:00

In this paper, we propose an end-to-end (E2E) dialect identification system trained using transfer learning from a multilingual automatic speech recognition (ASR) model. This is also an extension of our submitted system to the Oriental Language Recognition Challenge 2020 (AP20-OLR). We verified its applicability using the dialect identification (DID) task of the AP20-OLR. First, we trained a robust conformer-based joint connectionist temporal classification (CTC) /attention multilingual E2E ASR model using the training corpora of eight languages, independent of the target dialects. Second, we initialized the E2E-based classifier with the ASR model's shared encoder using a transfer learning approach. Finally, we trained the classifier on the target dialect corpus. We obtained the final classifier by selecting the best model from the following: (1) the averaged model in term of the loss values; and (2) the averaged model in term of classification accuracy.

Our experiments on the DID test-set of the AP20-OLR demonstrated that significant identification improvements were achieved for three Chinese dialects. The performances of our system outperforms the winning team of the AP20-OLR, with the largest relative reductions of 19.5% in C_{avg} and 25.2% in EER.

Language Recognition Based on Unsupervised Pretrained Models

Haibin Yu¹, Jing Zhao¹, Song Yang², Zhongqin Wu², Yuting Nie¹, Wei-Qiang Zhang¹; ¹Tsinghua University, China; ²TAL, China Thu-M-SS-1-5, Time: 12:20

Unsupervised pretrained models have been proven to rival or even outperform supervised systems in various speech recognition tasks. However, their performance for language recognition is still left to be explored. In this paper, we construct several language recognition systems based on existing unsupervised pretraining approaches, and explore their credibility and performance to learn high-level generalization of language. We discover that unsupervised pretrained models capture expressive and highly linear-separable features. With these representations, language recognition can perform well even when the classifiers are relatively simple or only a small amount of labeled data is available. Although linear classifiers are usable, neural nets with RNN structures improve the results. Meanwhile, unsupervised pretrained models are able to gain refined representations on audio frame level that are strongly coupled with the acoustic features of the input sequence. Therefore these features contain redundant information of speakers and channels with few relations to the identity of the language. This nature of unsupervised pretrained models causes a performance degradation in language recognition tasks on cross-channel tests.

Additive Phoneme-Aware Margin Softmax Loss for Language Recognition

Zheng Li, Yan Liu, Lin Li, Qingyang Hong; Xiamen University, China Thu-M-SS-1-6, Time: 12:40

This paper proposes an additive phoneme-aware margin softmax (APM-Softmax) loss to train the multi-task learning network with phonetic information for language recognition. In additive margin softmax (AM-Softmax) loss, the margin is set as a constant during the entire training for all training samples, and that is a suboptimal method since the recognition difficulty varies in training samples. In additive angular margin softmax (AAM-Softmax) loss, the additional angular margin is set as a constant as well. In this paper, we propose an APM-Softmax loss for language recognition with phoneitc multi-task learning, in which the additive phoneme-aware margin is automatically tuned for different training samples. More specifically, the margin of language recognition is adjusted according to the results of phoneme recognition. Experiments are reported on Oriental Language Recognition (OLR) datasets, and the proposed method improves AM-Softmax loss and AAM-Softmax loss in different language recognition testing conditions.

Thu-M-SS-2: Automatic Speech Recognition in Air Traffic Management

Room Lacina, 11:00–13:00, Thursday 2 September 2021 Chairs: Hartmut Helmke and Pavel Kolčárek

Introduction

Time: 11:00

Towards an Accent-Robust Approach for ATC Communications Transcription

Nataly Jahchan¹, Florentin Barbier², Ariyanidevi Dharma Gita¹, Khaled Khelif², Estelle Delpech²; ¹APSYS, France; ²Airbus, France Thu-M-SS-2-1, Time: 11:10

Air Traffic Control (ATC) communications are a typical example where Automatic Speech Recognition could face various challenges: audio data are quite noisy due to the characteristics of capturing mechanisms. All speakers involved use a specific English-based phraseology and a significant number of pilots and controllers are non-native English speakers. The aim of this work is to enhance pilot-ATC communications by adding a Speech to Text (STT) capability that will transcribe ATC speech into text on the cockpit interfaces to help the pilot understand ATC speech in a more optimal manner (be able to verify what he/she heard on the radio by looking at the text transcription, be able to decipher non-native English accents from controllers, not lose time asking the ATC to repeat the message several times). In this paper, we first describe an accent analysis study which was carried out both on a theoretical level but also with the help of feedback from several hundred airline pilots. Then, we present the dataset that was set up for this work. Finally, we describe the experiments we have implemented and the impact of the speaker accent on the performance of a speech to text engine.

Detecting English Speech in the Air Traffic Control Voice Communication

Igor Szöke, Santosh Kesiraju, Ondřej Novotný, Martin Kocour, Karel Veselý, Jan Černocký; Brno University of Technology, Czechia

Thu-M-SS-2-2, Time: 11:25

Developing in-cockpit voice enabled applications require a real-world dataset with labels and annotations. We launched a community platform for collecting the Air-Traffic Control (ATC) speech, world-wide in the ATCO² project. Filtering out non-English speech is one of the main components in the data processing pipeline. The proposed English Language Detection (ELD) system is based on the embeddings from Bayesian subspace multinomial model. It is trained on the word confusion network from an ASR system. It is robust, easy to train, and light weighted. We achieved 0.0439 equal-error-rate (EER), a 50% relative reduction as compared to the state-of-the-art acoustic ELD system based on x-vectors, in the in-domain scenario. Further, we achieved an EER of 0.1352, a 33% relative reduction as compared to the acoustic ELD, in the unseen language (out-of-domain) condition. We plan to publish the evaluation dataset from the ATCO² project.

Robust Command Recognition for Lithuanian Air Traffic Control Tower Utterances

Oliver Ohneiser¹, Seyyed Saeed Sarfjoo², Hartmut Helmke¹, Shruthi Shetty¹, Petr Motlicek², Matthias Kleinert¹, Heiko Ehr¹, Šarūnas Murauskas³; ¹DLR, Germany; ²Idiap Research Institute, Switzerland; ³Oro navigacija, Lithuania Thu-M-SS-2-3, Time: 11:40

The maturity of automatic speech recognition (ASR) systems at controller working positions is currently a highly relevant technological topic in air traffic control (ATC). However, ATC service providers are less interested in pure word error rate (WER). They want to see benefits of ASR applications for ATC. Such applications transform recognized word sequences into semantic meanings, i.e., a number of related concepts such as callsign, type, value, unit, etc., which are combined to form commands. Digitized concepts or recognized commands can enter ATC systems based on an ontology for utterance annotation agreed between European ATC stakeholders. Command recognition (CR) has already been performed in approach control. However, spoken utterances of tower controllers are longer, include more free speech, and contain other command types than in approach. An automatic CR rate of 95.8% is achievable on perfect word recognition, i.e., manually transcribed audio recordings (gold transcriptions), taken from Lithuanian controllers in a multiple remote tower environment. This paper presents CR results for various speech-to-text models with different WERs on tower utterances. Although WERs were around 9%, we achieve CR rates of 85%. CR rates only slightly decrease with higher WERs, which enables to bring ASR applications closer to operational ATC environment.

Contextual Semi-Supervised Learning: An Approach to Leverage Air-Surveillance and Untranscribed ATC Data in ASR Systems

Juan Zuluaga-Gomez¹, Iuliia Nigmatulina¹, Amrutha Prasad¹, Petr Motlicek¹, Karel Veselý², Martin Kocour², Igor Szöke³; ¹Idiap Research Institute, Switzerland; ²Brno University of Technology, Czechia; ³ReplayWell, Czechia

Thu-M-SS-2-4, Time: 11:55

Air traffic management and specifically air-traffic control (ATC) rely mostly on voice communications between Air Traffic Controllers (ATCos) and pilots. In most cases, these voice communications follow a well-defined grammar that could be leveraged in Automatic Speech Recognition (ASR) technologies. The callsign used to address an airplane is an essential part of all ATCo-pilot communications. We propose a two-step approach to add contextual knowledge during semi-supervised training to reduce the ASR system error rates at recognizing the part of the utterance that contains the callsign. Initially, we represent in a WFST the contextual knowledge (i.e. air-surveillance data) of an ATCo-pilot communication. Then, during Semi-Supervised Learning (SSL) the contextual knowledge is added by second-pass decoding (i.e. lattice re-scoring). Results show that 'unseen domains' (e.g. data from airports not present in the supervised training data) are further aided by contextual SSL when compared to standalone SSL. For this task, we introduce the Callsign Word Error Rate (CA-WER) as an evaluation metric, which only assesses ASR performance of the spoken callsign in an utterance. We obtained a 32.1% CA-WER relative improvement applying SSL with an additional 17.5% CA-WER improvement by adding contextual knowledge during SSL on a challenging ATC-based test set gathered from LiveATC.

Boosting of Contextual Information in ASR for Air-Traffic Call-Sign Recognition

Martin Kocour¹, Karel Veselý¹, Alexander Blatt², Juan Zuluaga Gomez³, Igor Szöke¹, Jan Černocký¹, Dietrich Klakow², Petr Motlicek³; ¹Brno University of Technology, Czechia; ²Universität des Saarlandes, Germany; ³Idiap Research Institute, Switzerland Thu-M-SS-2-5, Time: 12:10

Contextual adaptation of ASR can be very beneficial for multi-accent and often noisy Air-Traffic Control (ATC) speech. Our focus is call-sign recognition, which can be used to track conversations of ATC operators with individual airplanes. We developed a two-stage boosting strategy, consisting of HCLG boosting and Lattice boosting. Both are implemented as WFST compositions and the contextual information is specific to each utterance. In HCLG boosting we give score discounts to individual words, while in Lattice boosting the score discounts are given to word sequences. The context data have origin in surveillance database of OpenSky Network. From this, we obtain lists of call-signs that are made more likely to appear in the best hypothesis of ASR. This also improves the accuracy of the NLU module that recognizes the call-signs from the best hypothesis of ASR.

As part of $ATCO^2$ project, we collected liveatc test set2. The boosting of call-signs leads to 4.7% absolute WER improvement and 27.1% absolute increase of Call-Sign recognition Accuracy (CSA). Our best result of 82.9% CSA is quite good, given that the data is noisy, and WER 28.4% is relatively high. We believe there is still room for improvement.

Modeling the Effect of Military Oxygen Masks on Speech Characteristics

Benjamin Elie¹, Jodie Gauvain², Jean-Luc Gauvain¹, Lori Lamel¹; ¹LISN (UMR 9015), France; ²Vocapia Research, France Thu-M-SS-2-6, Time: 12:25

Wearing an oxygen mask changes the speech production of speakers. It indeed modifies the vocal apparatus and perturbs the articulatory movements of the speaker. This paper studies the impact of the oxygen mask of military aircraft pilots on formant trajectories, both dynamically (variations of the formants at a utterance level) and globally (mean value at the utterance level) for 12 speakers.

A comparative analysis of speech collected with and without an oxygen mask shows that the mask has a significant impact on the formant trajectories, both on the mean values and on the formant variations at the utterance level. This impact is strongly dependent on the speaker and also on the mask model. These observations suggest that the articulatory movements of the speaker are modified by the presence of the mask.

These observations are validated via a preliminary ASR experiment that uses a data augmentation technique based on articulatory perturbations that are driven by our experimental observations.

Panel Discussion

Time: 12:40

Thu-M-S&T-1 : Show and Tell 3

11:00-13:00, Thursday 2 September 2021

MoM: Minutes of Meeting Bot

Benjamin Milde, Tim Fischer, Steffen Remus, Chris Biemann; Universität Hamburg, Germany Thu-M-S&T-1-1, Time: 11:00

We present MoM (Minutes of Meeting) bot, an automatic meeting transcription system with real-time recognition, summarization and visualization capabilities. MoM works without any cloud processing and does not require a network connection. Every processing step is local, even its speech recognition component, to address privacy concerns of meetings. MoM can be used to assisted writing a (summarized) protocol of a meeting, but may also help the hearing-impaired to follow a discussion. We address meeting-related issues, e.g. local vocabulary of an organization or company with active learning of G2P models and custom vocabulary extensions.

Articulatory Data Recorder: A Framework for Real-Time Articulatory Data Recording

Alexander Wilbrandt, Simon Stone, Peter Birkholz; Technische Universität Dresden, Germany Thu-M-S&T-1-2, Time: 11:00

Articulatory data can be collected using numerous modalities, such as video, ultrasound, electromagnetic articulography, or palatographic techniques. Every measurement technique requires software to visualize the incoming data and export the data for further analysis. This has led to an increase of available recording software over the past decades, including properly maintained software in regular use but also many abandoned and dead projects. In this paper, we present a new framework for real-time, simultaneous recording of acoustic and articulatory data. With the release of the Articulatory Data Recorder, our aim is to provide the experimental phonetics and articulatory research community with a common framework that is simple to use and easy to extend. It is specifically designed to cover the most common use cases in experimental phonetics: Elicit speech utterances using text prompts and record simultaneous audio and articulatory data. By following the FURPS+-system, we offer a combination of high performance and a low barrier of entrance for enrollment of any new articulatory measurement technique. The current version already supports various palatographic measurement techniques in use at our institute and future work will incorporate feedback and feature requests from the community.

The INGENIOUS Multilingual Operations App

Joan Codina-Filbà¹, Guillermo Cámbara¹, Alex Peiró-Lilja¹, Jens Grivolla¹, Roberto Carlini¹, Mireia Farrús²; ¹Universitat Pompeu Fabra, Spain; ²Universitat de Barcelona, Spain

Thu-M-S&T-1-3, Time: 11:00

This paper presents the integration of a speech-to-speech translation service into a Telegram bot as a part of the EU funded INGENIOUS project. The bot is thought as a multilingual communication channel where First Responders talk in their own language and receive other's messages in English. The Speech-to-Speech translation system is currently being adapted to the emergency domains, so it will correctly deal with emergency codes and geographical data.

Digital Einstein Experience: Fast Text-to-Speech for Conversational AI

Joanna Rownicka, Kilian Sprenkamp, Antonio Tripiana, Volodymyr Gromoglasov, Timo P. Kunz; Aflorithmic Labs, UK

Thu-M-S&T-1-4, Time: 11:00

We describe our approach to create and deliver a custom voice for a conversational AI use-case. More specifically, we provide a voice for a Digital Einstein character, to enable human-computer interaction within the digital conversation experience. To create the voice which fits the context well, we first design a voice character and we produce the recordings which correspond to the desired speech attributes. We then model the voice. Our solution utilizes Fastspeech 2 for log-scaled mel-spectrogram prediction from phonemes and Parallel WaveGAN to generate the waveforms. The system supports a character input and gives a speech waveform at the output. We use a custom dictionary for selected words to ensure their proper pronunciation. Our proposed cloud architecture enables for fast voice delivery, making it possible to talk to the digital version of Albert Einstein in real-time.

Live Subtitling for BigBlueButton with Open-Source Software

Robert Geislinger¹, Benjamin Milde¹, Timo Baumann², Chris Biemann²; ¹HITeC, Germany; ²Universität Hamburg, Germany Thu-M-S&T-1-5, Time: 11:00

We present an open source plugin for live subtitling in the popular open source video conferencing software BigBlueButton. Our plugin decodes each speaker's audio stream separately and in parallel, thereby obliviating the need for speaker diarization and seamlessly handling overlapped talk. Any Kaldi-compatible nnet3 model can be used with our plugin and we demonstrate it using freely available TDNN-HMM-based ASR models for English and German. Our subtitles can be used as they are (e.g., in loud environments) or can form the basis for further NLP processes. Our tool can also simplify the collection of remotely recorded multi-party dialogue corpora.

Expressive Latvian Speech Synthesis for Dialog Systems

Dāvis Nicmanis, Askars Salimbajevs; Tilde, Latvia Thu-M-S&T-1-6, Time: 11:00

To fully enable spoken human-computer interaction, the text-tospeech (TTS) component of such a system must produce natural human-like speech and adjust the prosody according to the dialog context.

While the current publicly available TTS services can produce naturalsounding speech, they usually lack emotional expressiveness.

In this paper, we present an expressive speech synthesis prototype for the Latvian language. The prototype is integrated into our chatbot management system and enables bot designers to specify the stylistic information for each bot response, thus making the interaction with the chatbot more natural.

ViSTAFAE: A Visual Speech-Training Aid with Feedback of Articulatory Efforts

Pramod H. Kachare¹, Prem C. Pandey¹, Vishal Mane², Hirak Dasgupta¹, K.S. Nataraj¹, Akshada Rathod², Sheetal K. Pathak²; ¹IIT Bombay, India; ²Digital India, India

Thu-M-S&T-1-7, Time: 11:00

An app is presented as a speech-training aid for providing visual feedback of articulatory efforts using information obtained from the utterances' audiovisual recording. It has two panels to enable comparison between the articulatory efforts of the learner and the teacher or a pre-recorded reference speaker. The visual feedback consists of a slow-motion animation of lateral vocal tract shape, level, and pitch, and time-aligned display of the frontal view of the speaker's face along with playback of the time-scaled speech signal. The app comprises a graphical user interface and modules for signal acquisition, analysis, and animation. It is developed using Python as a Windows-based app and may be accessed remotely through a web browser.

Thu-Survey: Survey Talk 3: Karen Livescu

Room A+B, 13:00–14:00, Thursday 2 September 2021 Chairs: TBD

Learning Speech Models from Multi-Modal Data

Karen Livescu; TTIC, USA Thu-Survey, Time: 13:00

Speech is usually recorded as an acoustic signal, but it often appears in context with other signals. In addition to the acoustic signal, we may have available a corresponding visual scene, the video of the speaker, physiological signals such as the speaker's movements or neural recordings, or other related signals. It is often possible to learn a better speech model or representation by considering the context provided by these additional signals, or to learn with less training data. Typical approaches to training from multi-modal data are based on the idea that models or representations of each modality should be in some sense predictive of the other modalities. Multi-modal approaches can also take advantage of the fact that the sources of noise or nuisance variables are different in different measurement modalities, so an additional (non-acoustic) modality can help learn a speech representation that suppresses such noise. This talk will survey several lines of work in this area, both older and newer. It will cover some basic techniques from machine learning and statistics, as well as specific models and applications for speech.

Thu-Keynote: Keynote 3: Mounya Elhilali

Room A+B, 15:00–16:00, Thursday 2 September 2021 Chairs: TBD

Adaptive Listening to Everyday Soundscapes

Mounya Elhilali; Johns Hopkins University, USA Thu-Keynote, Time: 15:00

As we navigate our everyday life, we are continuously parsing through a cacophony of sounds that are constantly impinging on our senses. This ability to sieve through everyday sounds and pick-out signals of interest may seem intuitive and effortless, but it is a real feat that involves complex brain networks that balance the sensory signal with our goals, expectations, attentional state and prior knowledge (what we hear, what we want to hear, what we expect to hear, what we know). A similar challenge faces computer systems that need to adapt to dynamic inputs, evolving objectives and novel surrounds. A growing body of work in neuroscience has been amending our views of processing in the brain; replacing the conventional view of 'static' processing with a more 'active' and malleable mapping that rapidly adapts to the task at hand and listening conditions. After all, humans and most animals are not specialists, but generalists whose perception is shaped by experience, context and changing behavioral demands. The talk will discuss theoretical formulations of these adaptive processes and lessons to leverage attentional feedback in algorithms for detecting and separating sounds of interest (e.g. speech, music) amidst competing distractors.

Thu-A-O-1: Speech Production I

Room C, 16:00–18:00, Thursday 2 September 2021 Chairs: Pavel Šturm and Petra Wagner

Towards the Prediction of the Vocal Tract Shape from the Sequence of Phonemes to be Articulated

Vinicius Ribeiro¹, Karyna Isaieva², Justine Leclere², Pierre-André Vuissoz², Yves Laprie¹; ¹Loria (UMR 7503), France; ²IADI (Inserm U1254), France Thu-A-0-1-1, Time: 16:00

In this work, we address the prediction of speech articulators' temporal geometric position from the sequence of phonemes to be articulated. We start from a set of real-time MRI sequences uttered by a female French speaker. The contours of five articulators were tracked automatically in each of the frames in the MRI video. Then, we explore the capacity of a bidirectional GRU to correctly predict each articulator's shape and position given the sequence of phonemes and their duration. We propose a 5-fold cross-validation experiment to evaluate the generalization capacity of the model. In a second experiment, we evaluate our model's data efficiency by reducing training data. We evaluate the point-to-point Euclidean distance and the Pearson's correlations along time between the predicted and the target shapes. We also evaluate produced shapes of the critical articulators of specific phonemes. We show that our model can achieve good results with minimal data, producing very realistic vocal tract shapes.

Comparison of the Finite Element Method, the Multimodal Method and the Transmission-Line Model for the Computation of Vocal Tract Transfer Functions

Rémi Blandin¹, Marc Arnela², Simon Félix³, Jean-Baptiste Doc⁴, Peter Birkholz¹; ¹Technische Universität Dresden, Germany; ²Universitat Ramon Llull, Spain; ³LAUM (UMR 6613), France; ⁴LMSSC (EA 3196), France

Thu-A-O-1-2, Time: 16:20

The acoustic properties of vocal tract are usually characterized by its transfer function from the input acoustic volume flow at the glottis to the radiated acoustic pressure. These transfer functions can be computed with acoustic models. Three-dimensional acoustic simulation are used to take into account accurately the three-dimensional vocal tract shape and to generate valid results even at high frequency. Finite element models, finite difference methods, three-dimensional waveguide meshes, or the multimodal method have been used for this purpose. However, these methods require much more computation time than simple one-dimensional models. Among these methods, the multimodal method can achieve the shortest computation times. However, all the previous implementations had limitations regarding the geometrical shapes and the losses. In this work, we evaluate a new implementation that intends to overcome these limitations. Vowel transfer functions obtained with this new implementation are compared with a transmission-line model and a proven, robust and highly accurate method: the finite element method. While the finite element method remains the most reliable, the multimodal method generates similar transfer functions in much less time. The transmission line model gives valid results for the four first resonances.

Effects of Time Pressure and Spontaneity on Phonotactic Innovations in German Dialogues

Petra Wagner, Sina Zarrieß, Joana Cholin; Universität Bielefeld, Germany Thu-A-O-1-3, Time: 16:40

Speech variation is often explained by speakers' balancing of production constraints (favoring phonetic reduction of high frequency, expected items) and listener orientation (favoring more canonical productions for low frequency, unexpected items). Less well understood are processes involving a structural reorganization of articulatory plans due to re-syllabification, e.g., resulting from processes involving massive reduction, epenthesis or metathesis. In this paper, we want to focus on two kinds of re-syllabifications: (1) within-system innovations, in which non-canonical forms occur, and (2) beyond-system inventions, which do not follow the phonotactic constraints of the language under consideration. We examine these processes in a corpus of spontaneous and read dyadic interactions of German, in which time pressure was controlled as an additional factor. Results show that spontaneity and time pressure will mostly lead to within-system innovations, favoring highly trained, unmarked articulatory routines, while minimizing information loss. However, occasionally speakers leave the beaten paths of highly trained articulatory routines, and invent novel phonotactic sequences which are at odds with the phonotactic grammar of German. Our results are discussed in the light of their implications for contemporary models of speech production.

Importance of Parasagittal Sensor Information in Tongue Motion Capture Through a Diphonic Analysis

Salvador Medina¹, Sarah Taylor², Mark Tiede³, Alexander Hauptmann¹, Iain Matthews⁴; ¹Carnegie Mellon University, USA; ²University of East Anglia, UK; ³Haskins Laboratories, USA; ⁴Epic Games, USA Thu-A-O-1-4, Time: 17:00

Our study examines the information obtained by adding two parasagittal sensors to the standard midsagittal configuration of an Electromagnetic Articulography (EMA) observation of lingual articulation. In this work, we present a large and phonetically balanced corpus obtained from an EMA recording session of a single English native speaker reading 1899 sentences from the Harvard and TIMIT corpora. According to a statistical analysis of the diphones produced during the recording session, the motion captured by the parasagittal sensors has a low correlation to the midsagittal sensors in the mediolateral direction. We perform a geometric analysis of the lateral tongue by the measure of its width and using a proxy of the tongue's curvature that is computed using the Menger curvature. To provide a better understanding of the tongue sensor motion we present dynamic visualizations of all diphones. Finally, we present a summary of the velocity information computed from the tongue sensor information.

Learning Robust Speech Representation with an Articulatory-Regularized Variational Autoencoder

Marc-Antoine Georges, Laurent Girin, Jean-Luc Schwartz, Thomas Hueber; GIPSA-lab (UMR 5216), France

Thu-A-O-1-5, Time: 17:20

It is increasingly considered that human speech perception and production both rely on articulatory representations. In this paper, we investigate whether this type of representation could improve the performances of a deep generative model (here a variational autoencoder) trained to encode and decode acoustic speech features. First we develop an articulatory model able to associate articulatory parameters describing the jaw, tongue, lips and velum configurations with vocal tract shapes and spectral features. Then we incorporate these articulatory parameters into a variational autoencoder applied on spectral features by using a regularization technique that constrains part of the latent space to represent articulatory trajectories. We show that this articulatory constraint improves model training by decreasing time to convergence and reconstruction loss at convergence, and yields better performance in a speech denoising task.

Changes in Glottal Source Parameter Values with Light to Moderate Physical Load

*Heather Weston*¹, *Laura L. Koenig*², *Susanne Fuchs*¹; ¹*ZAS, Germany*; ²*Adelphi University, USA* Thu-A-O-1-6, Time: 17:40

Engaging in everyday physical activities, like walking, initiates physiological processes that also affect parts of the body used for speech. However, it is currently unclear to what extent such activities affect phonatory processes, and in turn, the voice. The present exploratory study investigates how selected glottal source parameters are affected by light and moderate physical activity. Recordings of sustained vowel /a/ were obtained from 39 female speakers of German at rest, and during low-intensity and moderate-intensity cycling. Ten glottal source parameters thought to reflect different physiological states were investigated using VoiceSauce. Even during light activity, significant increases were found in f0, strength of excitation and H1, and a decrease in harmonics-to-noise ratio at

NOTES

higher frequencies. During moderate-intensity activity, significant effects were stronger and found for most parameters. However, considerable intra- and interspeaker variability was observed. These findings may be relevant for applications in automatic speaker-state recognition. They also underscore the importance of investigating individual-level responses to better understand stress-voice interactions.

Thu-A-O-2: Speech Enhancement and Coding

Room A+B, 16:00–18:00, Thursday 2 September 2021 Chairs: Vladimir Malenovsky and Lukas Pfeifenberger

End-to-End Optimized Multi-Stage Vector Quantization of Spectral Envelopes for Speech and Audio Coding

Mohammad Hassan Vali, Tom Bäckström; Aalto University, Finland Thu-A-O-2-1, Time: 16:00

Spectral envelope modeling is an instrumental part of speech and audio codecs, which can be used to enable efficient entropy coding of spectral components. Overall optimization of codecs, including envelope models, has however been difficult due to the complicated interactions between different modules of the codec. In this paper, we study an end-to-end optimization methodology to optimize all modules in a codec integrally with respect to each other while capturing all these complex interactions with a global loss function. For the quantization of the spectral envelope parameters with a fixed bitrate, we use multi-stage vector quantization which gives high quality, but yet has a computational complexity which can be realistically applied in embedded devices. The obtained results demonstrate benefits in terms of PESQ and PSNR in comparison to the 3GPP EVS, as well as our recently proposed PyAWNeS codecs.

Fusion-Net: Time-Frequency Information Fusion Y-Network for Speech Enhancement

Santhan Kumar Reddy Nareddula, Subrahmanyam Gorthi, Rama Krishna Sai S. Gorthi; IIT Tirupati, India Thu-A-O-2-2, Time: 16:20

This paper proposes a deep learning-based densely connected Y-Net as an effective network architecture for the fusion of time and frequency domain loss functions for speech enhancement. The proposed architecture performs speech enhancement in the time domain while fusing information from the frequency domain. Y-network consists of an encoder branch followed by two decoder branches, where the first and second decoder loss functions enforce speech enhancement in time and frequency domains respectively. Each layer of the proposed network is formed with densely connected blocks comprising dilated and causal convolutions for significant feature collection and error backpropagation. The proposed model is trained on a publicly available data set of 28 speakers with 40 different noise conditions. The evaluations are performed on an independent, unseen test set of 2 speakers and 20 different noise conditions. The results from the proposed method are compared with five state-of-the-art methods using various metrics. The proposed method has resulted in an overall perceptual evaluation of speech quality of 3.4. It has outperformed the existing methods by a significant margin in terms of all the evaluation metrics.

N-MTTL SI Model: Non-Intrusive Multi-Task Transfer Learning-Based Speech Intelligibility Prediction Model with Scenery Classification

Luboš Marcinek, Michael Stone, Rebecca Millman, Patrick Gaydecki; University of Manchester, UK Thu-A-O-2-3, Time: 16:40

The application of speech enhancement algorithms for hearing aids may not always be beneficial to increasing speech intelligibility. Therefore, a prior environment classification could be important. However, previous speech intelligibility models do not provide any additional information regarding the reason for a decrease in speech intelligibility. We propose a unique non-intrusive multi-task transfer learning-based speech intelligibility prediction model with scenery classification (N-MTTL SI model). The solution combines a Mel-spectrogram analysis of the degraded speech signal with transfer learning and multi-task learning to provide simultaneous speech intelligibility prediction (task 1) and scenery classification of ten real-world noise conditions (task 2). The model utilises a pre-trained ResNet architecture as an encoder for feature extraction. The prediction accuracy of the N-MTTL SI model for both tasks is high. Specifically, RMSE of speech intelligibility predictions for seen and unseen conditions is 3.76% and 4.06%. The classification accuracy is 98%. In addition, the proposed solution demonstrates the potential of using pre-trained deep learning models in the domain of speech intelligibility prediction.

Thu-A-V-1: Emotion and Sentiment Analysis II

16:00–18:00, Thursday 2 September 2021 Chairs: Carlos Busso and Shrikanth Narayanan

Temporal Context in Speech Emotion Recognition

Yangyang Xia, Li-Wei Chen, Alexander Rudnicky, Richard M. Stern; Carnegie Mellon University, USA Thu-A-V-1-1, Time: 16:00

We investigate the importance of temporal context for speech emotion recognition (SER). Two SER systems trained on traditional and learned features, respectively, are developed to predict categorical labels of emotion. For traditional acoustical features, we study the combination of filterbank features and prosodic features and the impact on SER when the temporal context of these features is expanded by learnable spectro-temporal receptive fields (STRFs). Experiments show that the system trained on learnable STRFs outperforms other reported systems evaluated with a similar setup. We also demonstrate that the wav2vec features, pretrained with long temporal context, are superior to traditional features. We then introduce a novel segment-based learning objective to constrain our classifier to extract local emotion features from the large temporal context. Combined with the learning objective and fine-tuning strategy, our top-line system using wav2vec features reaches state-of-the-art performance on the IEMOCAP dataset.

Learning Fine-Grained Cross Modality Excitement for Speech Emotion Recognition

Hang Li, Wenbiao Ding, Zhongqin Wu, Zitao Liu; TAL, China

Thu-A-V-1-2, Time: 16:00

Speech emotion recognition is a challenging task because the emotion expression is complex, multimodal and fine-grained. In this paper, we propose a novel multimodal deep learning approach to perform fine-grained emotion recognition from real-life speeches. We design a temporal alignment mean-max pooling mechanism to capture the subtle and fine-grained emotions implied in every

NOTES

utterance. In addition, we propose a cross modality excitement module to conduct sample-specific adjustment on cross modality embeddings and adaptively recalibrate the corresponding values by its aligned latent features from the other modality. Our proposed model is evaluated on two well-known real-world speech emotion recognition datasets. The results demonstrate that our approach is superior on the prediction tasks for multimodal speech utterances, and it outperforms a wide range of baselines in terms of prediction accuracy. Furthermore, we conduct detailed ablation studies to show that our temporal alignment mean-max pooling mechanism and cross modality excitement significantly contribute to the promising results. In order to encourage the research reproducibility, we make the code publicly available.

Automatic Analysis of the Emotional Content of Speech in Daylong Child-Centered Recordings from a Neonatal Intensive Care Unit

*Einari Vaaras*¹, *Sari Ahlqvist-Björkroth*², *Konstantinos Drossos*¹, *Okko Räsänen*¹; ¹*Tampere University, Finland*; ²*University of Turku, Finland* Thu-A-V-1-3, Time: 16:00

Researchers have recently started to study how the emotional speech heard by young infants can affect their developmental outcomes. As a part of this research, hundreds of hours of daylong recordings from preterm infants' audio environments were collected from two hospitals in Finland and Estonia in the context of so-called APPLE study. In order to analyze the emotional content of speech in such a massive dataset, an automatic speech emotion recognition (SER) system is required. However, there are no emotion labels or existing in-domain SER systems to be used for this purpose. In this paper, we introduce this initially unannotated large-scale real-world audio dataset and describe the development of a functional SER system for the Finnish subset of the data. We explore the effectiveness of alternative state-of-the-art techniques to deploy a SER system to a new domain, comparing cross-corpus generalization, WGAN-based domain adaptation, and active learning in the task. As a result, we show that the best-performing models are able to achieve a classification performance of 73.4% unweighted average recall (UAR) and 73.2% UAR for a binary classification for valence and arousal, respectively. The results also show that active learning achieves the most consistent performance compared to the two alternatives.

Multimodal Sentiment Analysis with Temporal Modality Attention

Fan Qian, Jiqing Han; Harbin Institute of Technology, China

Thu-A-V-1-4, Time: 16:00

Multimodal sentiment analysis is an important research that involves integrating information from multiple modalities to identify a speaker underlying attitude. The core challenge is to model crossmodal interactions which span across both the different modalities and time. Although great progress has been made, the existing methods are still not sufficient for modeling cross-modal interactions. Inspired by previous research in cognitive neuroscience that humans perceive intentions through focusing on different modalities over time, in this paper we propose a novel attention mechanism called Temporal Modality Attention (TMA) to simulate this process. Cross-modal interactions are modeled using this human-like TMA mechanism which focuses on specific modalities dynamically as recurrent modeling proceed. To verify the effectiveness of TMA, we conduct comprehensive experiments on multiple benchmark datasets for multimodal sentiment analysis. The results show a consistently significant improvement compared to the baseline models.

Stochastic Process Regression for Cross-Cultural Speech Emotion Recognition

Mani Kumar T.¹, Enrique Sanchez¹, Georgios Tzimiropoulos², Timo Giesbrecht³, Michel Valstar¹; ¹University of Nottingham, UK; ²Queen Mary University of London, UK; ³Unilever, UK Thu-A-V-1-5, Time: 16:00

In this work, we pose continuous apparent emotion recognition from speech as a problem of learning distributions of functions, and do so using Stochastic Processes Regression. We presume that the relation between speech signals and their corresponding emotion labels is governed by some underlying stochastic process, in contrast to existing speech emotion recognition methods that are mostly based on deterministic regression models (static or recurrent). We treat each training sequence as an instance of the underlying stochastic process which we aim to discover using a neural latent variable model, which approximates the distribution of functions with a stochastic latent variable using an encoder-decoder composition: the encoder infers the distribution over the latent variable, which the decoder uses to predict the distribution of output emotion labels. To this end, we build on the previously proposed Neural Processes theory by using (a). noisy label predictions of a backbone instead of ground truth labels for latent variable inference and (b). recurrent encoder-decoder models to alleviate the effect of commonly encountered temporal misalignment between audio features and emotion labels due to annotator reaction lag. We validated our method on AVEC'19 cross-cultural emotion recognition dataset, achieving state-of-the-art results.

Acted vs. Improvised: Domain Adaptation for Elicitation Approaches in Audio-Visual Emotion Recognition

Haoqi Li¹, Yelin Kim¹, Cheng-Hao Kuo¹, Shrikanth S. Narayanan²; ¹Amazon, USA; ²University of Southern California, USA

Thu-A-V-1-6, Time: 16:00

Key challenges in developing generalized automatic emotion recognition systems include scarcity of labeled data and lack of gold-standard references. Even for the cues that are labeled as the same emotion category, the variability of associated expressions can be high depending on the elicitation context e.g., emotion elicited during improvised conversations vs. acted sessions with predefined scripts. In this work, we regard the emotion elicitation approach as domain knowledge, and explore domain transfer learning techniques on emotional utterances collected under different emotion elicitation approaches, particularly with limited labeled target samples. Our emotion recognition model combines the gradient reversal technique with an entropy loss function as well as the softlabel loss, and the experiment results show that domain transfer learning methods can be employed to alleviate the domain mismatch between different elicitation approaches. Our work provides new insights into emotion data collection, particularly the impact of its elicitation strategies, and the importance of domain adaptation in emotion recognition aiming for generalized systems.

Emotion Recognition from Speech Using wav2vec 2.0 Embeddings

Leonardo Pepino, Pablo Riera, Luciana Ferrer; UBA-CONICET ICC, Argentina Thu-A-V-1-7, Time: 16:00

Emotion recognition datasets are relatively small, making the use of deep learning techniques challenging. In this work, we propose a

NOTES

transfer learning method for speech emotion recognition (SER) where features extracted from pre-trained wav2vec 2.0 models are used as input to shallow neural networks to recognize emotions from speech. We propose a way to combine the output of several layers from the pre-trained model, producing richer speech representations than the model's output alone. We evaluate the proposed approaches on two standard emotion databases, IEMOCAP and RAVDESS, and compare different feature extraction techniques using two wav2vec 2.0 models: a generic one, and one finetuned for speech recognition. We also experiment with different shallow architectures for our speech emotion recognition model, and report baseline results using traditional features. Finally, we show that our best performing models have better average recall than previous approaches that use deep neural networks trained on features extracted from wav2vec 1.0.

Graph Isomorphism Network for Speech Emotion Recognition

Jiawang Liu, Haoxiang Wang; SCUT, China Thu-A-V-1-8, Time: 16:00

Previous deep learning approaches such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) have been broadly used in speech emotion recognition (SER). In these approaches, speech signals are generally modeled in the Euclidean space. In this paper, a novel SER model (LSTM-GIN) is proposed, which applies Graph Isomorphism Network (GIN) on LSTM outputs for global emotion modeling in the non-Euclidean space. In our LSTM-GIN model, speech signals are represented as graph-structured data so that we can better extract global feature representation. The deep frame-level features generated from the bidirectional LSTM are converted into an undirected graph with nodes represented by frame-level features and connections defined according to temporal relations between speech frames. GIN is adopted to classify the graph representations of utterances, as it is proved of excellent discriminative power in comparative experiments. We conduct experiments on the IEMOCAP dataset, and the results show that our proposed LSTM-GIN model surpasses other recent graph-based models and deep learning models by achieving 64.65% of weighted accuracy (WA) and 65.53% of unweighted accuracy (UA).

Applying TDNN Architectures for Analyzing Duration Dependencies on Speech Emotion Recognition

Pooja Kumawat, Aurobinda Routray; IIT Kharagpur, India

Thu-A-V-1-9, Time: 16:00

We have analyzed the Time Delay Neural Network (TDNN) based architectures for speech emotion classification. TDNN models efficiently capture the temporal information and provide an utterance level prediction. Emotions are dynamic in nature and require temporal context for reliable prediction. In our work, we have applied the TDNN based x-vector and emphasized channel attention, propagation & aggregation based TDNN (ECAPA-TDNN) architectures for speech emotion identification with RAVDESS, Emo-DB, and IEMOCAP databases. The results show that the TDNN architectures are very efficient for predicting emotion classes and ECAPA-TDNN outperforms the TDNN based x-vector architecture. Next, we investigated the performance of ECAPA-TDNN with various training chunk durations and test utterance durations. We have identified that in spite of very promising emotion recognition performance the TDNN models have a strong training chunk duration-based bias. Earlier research work revealed that individual emotion class accuracy depends largely on the test utterance duration. Most of these studies were based on frame level emotions predictions. However, utterance level based emotion recognition is relatively less explored. The results show that even with the TDNN models, the accuracy of the different emotion classes is dependent on the utterance duration.

Acoustic Features and Neural Representations for Categorical Emotion Recognition from Speech

Aaron Keesing, Yun Sing Koh, Michael Witbrock; University of Auckland, New Zealand Thu-A-V-1-10, Time: 16:00

Many features have been proposed for use in speech emotion recognition, from signal processing features to bag-of-audio-words (BoAW) models to abstract neural representations. Some of these feature types have not been directly compared across a large number of speech corpora to determine performance differences. We propose a full factorial design and to compare speech processing features, BoAW and neural representations on 17 emotional speech datasets. We measure the performance of features in a categorical emotion classification problem for each dataset, using speaker-independent cross-validation with diverse classifiers. Results show statistically significant differences between features and between classifiers, with large effect sizes between features. In particular, standard acoustic feature sets still perform competitively to neural representations, while neural representations have a larger range of performance, and BoAW features lie in the middle. The best and worst neural representations were wav2veq and VGGish, respectively, with wav2vec performing best out of all tested features. These results indicate that standard acoustic feature sets are still very useful baselines for emotional classification, but high quality neural speech representations can be better.

Leveraging Pre-Trained Language Model for Speech Sentiment Analysis

Suwon Shon¹, Pablo Brusco¹, Jing Pan¹, Kyu J. Han¹, Shinji Watanabe²; ¹ASAPP, USA; ²Carnegie Mellon University, USA Thu-A-V-1-11, Time: 16:00

In this paper, we explore the use of pre-trained language models to learn sentiment information of written texts for speech sentiment analysis. First, we investigate how useful a pre-trained language model would be in a 2-step pipeline approach employing Automatic Speech Recognition (ASR) and transcripts-based sentiment analysis separately. Second, we propose a pseudo label-based semi-supervised training strategy using a language model on an end-to-end speech sentiment approach to take advantage of a large, but unlabeled speech dataset for training. Although spoken and written texts have different linguistic characteristics, they can complement each other in understanding sentiment. Therefore, the proposed system can not only model acoustic characteristics to bear sentiment-specific information in speech signals, but learn latent information to carry sentiments in the text representation. In these experiments, we demonstrate the proposed approaches improve F1 scores consistently compared to systems without a language model. Moreover, we also show that the proposed framework can reduce 65% of human supervision by leveraging a large amount of data without human sentiment annotation and boost performance in a low-resource condition where the human sentiment annotation is not available enough.

Thu-A-V-2: Multi- and Cross-Lingual ASR, Other Topics in ASR

16:00–18:00, Thursday 2 September 2021 Chairs: Khe Sim and Alexei Baevski

Cross-Domain Speech Recognition with Unsupervised Character-Level Distribution Matching

*Wenxin Hou*¹, *Jindong Wang*², *Xu Tan*², *Tao Qin*², *Takahiro Shinozaki*¹; ¹*Tokyo Tech, Japan*; ²*Microsoft, China*

Thu-A-V-2-1, Time: 16:00

End-to-end automatic speech recognition (ASR) can achieve promising performance with large-scale training data. However, it is known that domain mismatch between training and testing data often leads to a degradation of recognition accuracy. In this work, we focus on the unsupervised domain adaptation for ASR and propose CMatch, a Character-level distribution matching method to perform fine-grained adaptation between each character in two domains. First, to obtain labels for the features belonging to each character, we achieve frame-level label assignment using the Connectionist Temporal Classification (CTC) pseudo labels. Then, we match the character-level distributions using Maximum Mean Discrepancy. We train our algorithm using the self-training technique. Experiments on the Libri-Adapt dataset show that our proposed approach achieves 14.39% and 16.50% relative Word Error Rate (WER) reduction on both cross-device and cross-environment ASR. We also comprehensively analyze the different strategies for frame-level label assignment and Transformer adaptations.

Large-Scale Pre-Training of End-to-End Multi-Talker ASR for Meeting Transcription with Single Distant Microphone

Naoyuki Kanda¹, Guoli Ye¹, Yu Wu², Yashesh Gaur¹, Xiaofei Wang¹, Zhong Meng¹, Zhuo Chen¹, Takuya Yoshioka¹; ¹Microsoft, USA; ²Microsoft, China Thu-A-V-2-2, Time: 16:00

Transcribing meetings containing overlapped speech with only a single distant microphone (SDM) has been one of the most challenging problems for automatic speech recognition (ASR). While various approaches have been proposed, all previous studies on the monaural overlapped speech recognition problem were based on either simulation data or small-scale real data. In this paper, we extensively investigate a two-step approach where we first pre-train a serialized output training (SOT)-based multi-talker ASR by using large-scale simulation data and then fine-tune the model with a small amount of real meeting data. Experiments are conducted by utilizing 75 thousand (K) hours of our internal single-talker recording to simulate a total of 900K hours of multi-talker audio segments for supervised pre-training. With fine-tuning on the 70 hours of the AMI-SDM training data, our SOT ASR model achieves a word error rate (WER) of 21.2% for the AMI-SDM evaluation set while automatically counting speakers in each test segment. This result is not only significantly better than the previous state-of-the-art WER of 36.4% with oracle utterance boundary information but also better than a result by a similarly fine-tuned single-talker ASR model applied to beamformed audio.

On Minimum Word Error Rate Training of the Hybrid Autoregressive Transducer

Liang Lu, Zhong Meng, Naoyuki Kanda, Jinyu Li, Yifan Gong; Microsoft, USA Thu-A-V-2-3, Time: 16:00

Hybrid Autoregressive Transducer (HAT) is a recently proposed end-to-end acoustic model that extends the standard Recurrent Neural Network Transducer (RNN-T) for the purpose of the external language model (LM) fusion. In HAT, the blank probability and the label probability are estimated using two separate probability distributions, which provides a more accurate solution for internal LM score estimation, and thus works better when combining with an external LM. Previous work mainly focuses on HAT model training with the negative log-likelihood loss, while in this paper, we study the minimum word error rate (MWER) training of HAT - a criterion that is closer to the evaluation metric for speech recognition, and has been successfully applied to other types of end-to-end models such as sequence-to-sequence (S2S) and RNN-T models. From experiments with around 30,000 hours of training data, we show that MWER training can improve the accuracy of HAT models, while at the same time, improving the robustness of the model against the decoding hyper-parameters such as length normalization and decoding beam during inference.

Reducing Streaming ASR Model Delay with Self Alignment

Jaeyoung Kim, Han Lu, Anshuman Tripathi, Qian Zhang, Hasim Sak; Google, USA Thu-A-V-2-4, Time: 16:00

Reducing prediction delay for streaming end-to-end ASR models with minimal performance regression is a challenging problem. Constrained alignment is a well-known existing approach that penalizes predicted word boundaries using external low-latency acoustic models. On the contrary, recently proposed FastEmit is a sequence-level delay regularization scheme encouraging vocabulary tokens over blanks without any reference alignments. Although all these schemes are successful in reducing delay, ASR word error rate (WER) often severely degrades after applying these delay constraining schemes. In this paper, we propose a novel delay constraining method, named self alignment. Self alignment does not require external alignment models. Instead, it utilizes Viterbi forced-alignments from the trained model to find the lower latency alignment direction. From LibriSpeech evaluation, self alignment outperformed existing schemes: 25% and 56% less delay compared to FastEmit and constrained alignment at the similar word error rate. For Voice Search evaluation, 12% and 25% delay reductions were achieved compared to FastEmit and constrained alignment with more than 2% WER improvements.

Reduce and Reconstruct: ASR for Low-Resource Phonetic Languages

Anuj Diwan, Preethi Jyothi; IIT Bombay, India Thu-A-V-2-5, Time: 16:00

This work presents a seemingly simple but effective technique to improve low-resource ASR systems for phonetic languages. By identifying sets of acoustically similar graphemes in these languages, we first reduce the output alphabet of the ASR system using linguistically meaningful reductions and then reconstruct the original alphabet using a standalone module. We demonstrate that this lessens the burden and improves the performance of low-resource end-to-end ASR systems (because only reduced-alphabet predictions are needed) and that it is possible to design a very simple but effective reconstruction module that recovers sequences in the original

NOTES

alphabet from sequences in the reduced alphabet. We present a finite state transducer-based reconstruction module that operates on the 1-best ASR hypothesis in the reduced alphabet. We demonstrate the efficacy of our proposed technique using ASR systems for two Indian languages, Gujarati and Telugu. With access to only 10 hrs of speech data, we obtain relative WER reductions of up to 7% compared to systems that do not use any reduction.

Knowledge Distillation Based Training of Universal ASR Source Models for Cross-Lingual Transfer

*Takashi Fukuda*¹, *Samuel Thomas*²; ¹*IBM*, *Japan*; ²*IBM*, *USA*

Thu-A-V-2-6, Time: 16:00

In this paper we introduce a novel knowledge distillation based framework for training universal source models. In our proposed approach for automatic speech recognition (ASR), multilingual source models are first trained using multiple language-dependent resources before being used to initialize language specific target models in low resource settings. For the proposed source models to be effective in cross-lingual transfer to novel target languages, the training framework encourages the models to perform accurate universal phone classification while ignoring any language-dependent characteristics present in the training data set. These two goals are achieved by applying knowledge distillation to improve the models' universal phone classification performance along with a shuffling mechanism that alleviates any language specific dependencies that might be learned. The benefits of this proposed technique are demonstrated in several practical settings, where either large amounts or only limited quantities of unbalanced multilingual data resources are available for source model creation. Compared to a conventional knowledge transfer learning method, the proposed approaches achieve a relative WER reduction of 8-10% in streaming ASR settings for various low resource target languages.

Listen with Intent: Improving Speech Recognition with Audio-to-Intent Front-End

Swayambhu Nath Ray¹, Minhua Wu², Anirudh Raju², Pegah Ghahremani², Raghavendra Bilgi¹, Milind Rao², Harish Arsikere¹, Ariya Rastrow², Andreas Stolcke², Jasha Droppo²; ¹Amazon, India; ²Amazon, USA Thu-A-V-2-7, Time: 16:00

Comprehending the overall intent of an utterance helps a listener recognize the individual words spoken. Inspired by this fact, we perform a novel study of the impact of explicitly incorporating intent representations as additional information to improve a recurrent neural network-transducer (RNN-T) based automatic speech recognition (ASR) system. An audio-to-intent (A2I) model encodes the intent of the utterance in the form of embeddings or posteriors, and these are used as auxiliary inputs for RNN-T training and inference. Experimenting with a 50k-hour far-field English speech corpus, this study shows that when running the system in *non-streaming* mode, where intent representation is extracted from the entire utterance and then used to bias streaming RNN-T search from the start, it provides a 5.56% relative word error rate reduction (WERR). On the other hand, a *streaming* system using per-frame intent posteriors as extra inputs for the RNN-T ASR system yields a 3.33% relative WERR. A further detailed analysis of the streaming system indicates that our proposed method brings especially good gain on media-playing related intents (e.g. 9.12% relative WERR on PlayMusicIntent).

Exploring Targeted Universal Adversarial Perturbations to End-to-End ASR Models

Zhiyun Lu, Wei Han, Yu Zhang, Liangliang Cao; Google, USA

Thu-A-V-2-8, Time: 16:00

Although end-to-end automatic speech recognition (e2e ASR) models are widely deployed in many applications, there have been very few studies to understand models' robustness against adversarial perturbations. In this paper, we explore whether a targeted universal perturbation vector exists for e2e ASR models. Our goal is to find perturbations that can mislead the models to predict the given targeted transcript such as "thank you" or empty string on any input utterance. We study two different attacks, namely additive and prepending perturbations, and their performances on the state-of-the-art LAS, CTC and RNN-T models. We find that LAS is the most vulnerable to perturbations among the three models. RNN-T is more robust against additive perturbations, especially on long utterances. And CTC is robust against both additive and prepending perturbations. To attack RNN-T, we find prepending perturbation is more effective than the additive perturbation, and can mislead the models to predict the same short target on utterances of arbitrary length.

Earnings-21: A Practical Benchmark for ASR in the Wild

*Miguel Del Rio*¹, *Natalie Delworth*¹, *Ryan Westerman*¹, *Michelle Huang*¹, *Nishchal Bhandari*¹, *Joseph Palakapilly*¹, *Quinten McNamara*¹, *Joshua Dong*¹, *Piotr Żelasko*², *Miguel Jetté*¹; ¹*Rev.com, USA*; ²*Johns Hopkins University, USA*

Thu-A-V-2-9, Time: 16:00

Commonly used speech corpora inadequately challenge academic and commercial ASR systems. In particular, speech corpora lack metadata needed for detailed analysis and WER measurement. In response, we present Earnings-21, a 39-hour corpus of earnings calls containing entity-dense speech from nine different financial sectors. This corpus is intended to benchmark ASR systems in the wild with special attention towards named entity recognition. We benchmark four commercial ASR models, two internal models built with open-source tools, and an open-source LibriSpeech model and discuss their differences in performance on Earnings-21. Using our recently released *fstalign* tool, we provide a candid analysis of each model's recognition capabilities under different partitions. Our analysis finds that ASR accuracy for certain NER categories is poor, presenting a significant impediment to transcript comprehension Earnings-21 bridges academic and commercial ASR and usage. system evaluation and enables further research on entity modeling and WER on real world audio.

Improving Multilingual Transformer Transducer Models by Reducing Language Confusions

*Eric Sun*¹, *Jinyu Li*¹, *Zhong Meng*¹, *Yu Wu*², *Jian Xue*¹, *Shujie Liu*², *Yifan Gong*¹; ¹*Microsoft, USA*; ²*Microsoft, China*

Thu-A-V-2-10, Time: 16:00

In end-to-end multilingual speech recognition, the hypotheses in one language could include word tokens from other languages. Language confusions happen even more frequently when language identifier (LID) is not present during inference. In this paper, we explore to reduce language confusions without using LID in model inference by creating models with multiple output heads and use sequence probability to select the correct head for output hypotheses. We

NOTES

propose head grouping to merge several language outputs into one head to save runtime cost. Head groups are decided by the distances among language clusters learned through language embedding vectors to separate confusable languages apart. We further propose prediction network sharing for languages from the same family. By jointly applying head grouping and prediction network sharing, training data from the same family languages is better shared while confusable languages are divided into different heads to reduce language confusions. Our experiments demonstrate that our multilingual transformer transducer models based on multi-head outputs achieve on average 7.8% and 10.9% relative word error rate reductions without LID being used in inference from one-head baseline model with affordably increased runtime cost on 10 European languages.

Arabic Code-Switching Speech Recognition Using Monolingual Data

Ahmed Ali¹, Shammur Absar Chowdhury¹, Amir Hussein¹, Yasser Hifny²; ¹HBKU, Qatar; ²Helwan University, Egypt Thu-A-V-2-11, Time: 16:00

Code-switching in automatic speech recognition (ASR) is an important challenge due to globalization. Recent research in multilingual ASR shows potential improvement over monolingual systems. We study key issues related to multilingual modeling for ASR through a series of large-scale ASR experiments. Our innovative framework deploys a multi-graph approach in the weighted finite state transducers (WFST) framework. We compare our WFST decoding strategies with a transformer sequence to sequence system trained on the same data. Given a code-switching scenario between Arabic and English languages, our results show that the WFST decoding datasets. In addition, the transformer system performed better for intrasentential code-switching task. With this study, we release an artificially generated development and test sets, along with ecological code-switching test set, to benchmark the ASR performance.

Thu-A-V-3: Source Separation II

16:00–18:00, Thursday 2 September 2021 Chairs: Mariem Bouafif and Lukas Drude

Online Blind Audio Source Separation Using Recursive Expectation-Maximization

Aviad Eisenberg, Boaz Schwartz, Sharon Gannot; Bar-Ilan University, Israel Thu-A-V-3-1, Time: 16:00

The challenging problem of online multi-microphone blind audio source separation (BASS) in noisy environment is addressed in this paper. We present a sequential, non-iterative, algorithm based on the recursive EM (REM) framework. In the proposed algorithm, the compete-data, which constitutes the separated sources and residual noise, is estimated in the E-step by applying a multichannel Wiener filter (MCWF); and the corresponding parameters, comprised of acoustic transfer functions (ATFs) relating the sources and the microphones and power spectral densities (PSDs) of the desired sources, are sequentially estimated in the M-step. The separated speech signals are further enhanced using matched-filter beamformers. The performance of the algorithm is demonstrated in terms of the separation capabilities, the resulting speech intelligibility and the ability to track the direction of arrival (DOA) of the moving sources.

Empirical Analysis of Generalized Iterative Speech Separation Networks

Yi Luo, Cong Han, Nima Mesgarani; Columbia University, USA

Thu-A-V-3-2, Time: 16:00

Although most existing speech separation networks are designed as a one-pass pipeline where the sources are directly estimated from the mixture, multi-pass or iterative pipelines have been shown to be effective by designing multiple rounds of separation and utilizing separation outputs from a previous iteration as additional inputs for the next iteration. Moreover, such iterative separation pipeline can also be extended to a more general framework where a training objective designed to minimize the discrepancy between the estimated and target sources is applied to different parts of the network. In this paper, we empirically investigate the effect of such generalized iterative separation pipeline by adjusting its configuration in multiple aspects in both training and inference phases. For the training phase, we compare the separation performance of both time-domain and frequency-domain networks with different numbers of iterations following the recent discussions on the model architecture organizations. We also evaluate the effect of parameter sharing across iterations and the necessity of additional training objectives. For the inference phase, we measure the separation performance of various numbers of iterations. Our results show that iterative speech separation is a promising direction and deserves more in-depth analysis and exploration.

Graph-PIT: Generalized Permutation Invariant Training for Continuous Separation of Arbitrary Numbers of Speakers

Thilo von Neumann¹, Keisuke Kinoshita², Christoph Boeddeker¹, Marc Delcroix², Reinhold Haeb-Umbach¹; ¹Universität Paderborn, Germany; ²NTT, Japan Thu-A-V-3-3, Time: 16:00

Automatic transcription of meetings requires handling of overlapped speech, which calls for continuous speech separation (CSS) systems. The uPIT criterion was proposed for utterance-level separation with neural networks and introduces the constraint that the total number of speakers must not exceed the number of output channels. When processing meeting-like data in a segment-wise manner, i.e., by separating overlapping segments independently and stitching adjacent segments to continuous output streams, this constraint has to be fulfilled for any segment. In this contribution, we show that this constraint can be significantly relaxed. We propose a novel graph-based PIT criterion, which casts the assignment of utterances to output channels in a graph coloring problem. It only requires that the number of concurrently active speakers must not exceed the number of output channels. As a consequence, the system can process an arbitrary number of speakers and arbitrarily long segments and thus can handle more diverse scenarios. Further, the stitching algorithm for obtaining a consistent output order in neighboring segments is of less importance and can even be eliminated completely, not the least reducing the computational effort. Experiments on meeting-style WSJ data show improvements in recognition performance over using the uPIT criterion.

Teacher-Student MixIT for Unsupervised and Semi-Supervised Speech Separation

*Jisi Zhang*¹, *Cătălin Zorilă*², *Rama Doddipatla*², *Jon Barker*¹; ¹*University of Sheffield, UK*; ²*Toshiba, UK* Thu-A-V-3-4, Time: 16:00

In this paper, we introduce a novel semi-supervised learning framework for end-to-end speech separation. The proposed method first uses mixtures of unseparated sources and the mixture invariant training (MixIT) criterion to train a teacher model. The teacher model then estimates separated sources that are used to train a student model with standard permutation invariant training (PIT). The student model can be fine-tuned with supervised data, i.e., paired artificial mixtures and clean speech sources, and further improved via model distillation. Experiments with single and multi channel mixtures show that the teacher-student training resolves the overseparation problem observed in the original MixIT method. Further, the semi-supervised performance is comparable to a fully-supervised separation system trained using ten times the amount of supervised data.

Few-Shot Learning of New Sound Classes for Target Sound Extraction

Marc Delcroix, Jorge Bennasar Vázquez, Tsubasa Ochiai, Keisuke Kinoshita, Shoko Araki; NTT, Japan Thu-A-V-3-5, Time: 16:00

Target sound extraction consists of extracting the sound of a target acoustic event (AE) class from a mixture of AE sounds. It can be realized using a neural network that extracts the target sound conditioned on a 1-hot vector that represents the desired AE class. With this approach, embedding vectors associated with the AE classes are directly optimized for the extraction of sound classes seen during training. However, it is not easy to extend this framework to new AE classes, i.e. unseen during training. Recently, speech, music, or AE sound extraction based on enrollment audio of the desired sound offers the potential of extracting any target sound in a mixture given only a short audio signal of a similar sound. In this work, we propose combining 1-hot- and enrollment-based target sound extraction, allowing optimal performance for seen AE classes and simple extension to new classes. In experiments with synthesized sound mixtures generated with the Freesound Dataset (FSD) datasets, we demonstrate the benefit of the combined framework for both seen and new AE classes. Besides, we also propose adapting the embedding vectors obtained from a few enrollment audio samples (few-shot) to further improve performance on new classes.

Binaural Speech Separation of Moving Speakers With Preserved Spatial Cues

Cong Han, Yi Luo, Nima Mesgarani; Columbia University, USA_____

Thu-A-V-3-6, Time: 16:00

Binaural speech separation algorithms designed for augmented hearing technologies need to both improve the signal-to-noise ratio of individual speakers and preserve their perceived location in space. The majority of binaural speech separation methods assume nonmoving speakers. As a result, their application to real-world scenarios with freely moving speakers requires block-wise adaptation which relies on short-term contextual information and limits their performance. In this study, we propose an alternative approach for utterance-level source separation with moving speakers and in reverberant conditions. Our model makes use of spectral and spatial features of speakers in a larger context compared to the block-wise adaption methods. The model can implicitly track speakers within the utterance without the need for explicit tracking modules. Experimental results on simulated moving multitalker speech show that the proposed method can significantly outperform block-wise adaptation methods in both separation performance and preserving the interaural cues across multiple conditions, which makes it suitable for real-world augmented hearing applications.

AvaTr: One-Shot Speaker Extraction with Transformers

Shell Xu Hu, Md. Rifat Arefin, Viet-Nhat Nguyen, Alish Dipani, Xaq Pitkow, Andreas Savas Tolias; Upload AI, USA

Thu-A-V-3-7, Time: 16:00

To extract the voice of a target speaker when mixed with a variety of other sounds, such as white and ambient noises or the voices of interfering speakers, we extend the Transformer network [1] to attend the most relevant information with respect to the target speaker given the characteristics of his or her voices as a form of contextual information. The idea has a natural interpretation in terms of the *selective attention theory* [2]. Specifically, we propose two models to incorporate the voice characteristics in Transformer based on different insights of where the feature selection should take place. Both models yield excellent performance, on par or better than published state-of-the-art models on the *speaker extraction* task, including separating speech of novel speakers not seen during training.

Vocal Harmony Separation Using Time-Domain Neural Networks

Saurjya Sarkar, Emmanouil Benetos, Mark Sandler; Queen Mary University of London, UK Thu-A-V-3-8, Time: 16:00

Polyphonic vocal recordings are an inherently challenging source separation task due to the melodic structure of the vocal parts and unique timbre of its constituents. In this work we utilise a time-domain neural network architecture re-purposed from speech separation research and modify it to separate a capella mixtures at a high sampling rate. We use four-part (soprano, alto, tenor and bass) a capella recordings of Bach Chorales and Barbershop Quartets for our experiments. Unlike current deep learning based choral separation models where the training objective is to separate constituent sources based on their class, we train our model using a permutation invariant objective. Using this we achieve state-of-the-art results for choral music separation. We introduce a novel method to estimate harmonic overlap between sung musical notes as a measure of task complexity. We also present an analysis of the impact of randomised mixing, input lengths and filterbank lengths for our task. Our results show a moderate negative correlation between the harmonic overlap of the target sources and source separation performance. We report that training our models with randomly mixed musically-incoherent mixtures drastically reduces the performance of vocal harmony separation as it decreases the average harmonic overlap presented during training.

Speaker Verification-Based Evaluation of Single-Channel Speech Separation

Matthew Maciejewski, Shinji Watanabe, Sanjeev Khudanpur; Johns Hopkins University, USA Thu-A-V-3-9, Time: 16:00

Speech enhancement techniques typically focus on intrinsic metrics of signal quality. The overwhelming majority of deep learning-based single-channel speech separation studies, for instance, have relied on a single class of metrics to evaluate the systems by. These metrics, usually variants of Signal-to-Distortion Ratio (SDR), measure fidelity to the "ground truth" waveform. This can be problematic, not only for lack of diversity in evaluation metrics, but also in cases where a perfect ground truth waveform may be unavailable. In this work, we explore the value of speaker verification as an extrinsic metric of separation quality, with additional utility as evidence of the benefits of separation as pre-processing for downstream tasks.

NOTES

Improved Speech Separation with Time-and-Frequency Cross-Domain Feature Selection

Tian Lan, Yuxin Qian, Yilan Lyu, Refuoe Mokhosi, Wenxin Tai, Qiao Liu; UESTC, China Thu-A-V-3-10, Time: 16:00

Most deep learning-based monaural speech separation models only use either spectrograms or time domain speech signal as the input feature. The recently proposed cross-domain network (CDNet) demonstrates that concatenated frequency domain and time domain features helps to reach better performance. Although concatenation is a widely used feature fusion method, it has been proved that using frequency domain and time domain features to reconstruct signal makes minor difference compared with only using time domain feature in CDNet. To make better use of frequency domain feature in decoder, we propose using selection weights to select and fuse features from different domains and unify the features used in separator and decoder. In this paper, we propose using trainable weights or the global information calculated from the different domain features to generate selection weights. Given that our proposed models use element-wise fusing in the encoder, only one deconvolution layer in the decoder is needed to reconstruct signals. Experiments show that proposed methods achieve encouraging results on the large and challenging Libri2Mix dataset with a small increasing in parameters, which proves the frequency domain information is beneficial for signal reconstruction. Furthermore, proposed method has shown good generalizability on the unmatched VCTK2Mix dataset.

Robust Speaker Extraction Network Based on Iterative Refined Adaptation

Chengyun Deng¹, Shiqian Ma¹, Yongtao Sha¹, Yi Zhang¹, Hui Zhang², Hui Song¹, Fei Wang¹; ¹DiDi Chuxing, China; ²Baidu, China Thu-A-V-3-11, Time: 16:00

Speaker extraction aims to extract target speech signal from a multitalker environment with interference speakers and surrounding noise, given a reference speech from target speaker. Most speaker extraction systems achieve satisfactory performance in the closed condition. Such systems suffer from performance degradation given unseen target speakers and/or mismatched reference speech. In this paper we propose a novel strategy named Iterative Refined Adaptation (IRA) to improve the robustness and generalization capability of speaker extraction systems in the aforementioned scenarios. Given an initial speaker embedding encoded by an auxiliary network, the extraction network can obtain a latent representation of the target speaker as the feedback of the auxiliary network to refine the speaker embedding, which provides more accurate guidance for the extraction network. Experiments show that the network with IRA confirm the superior performance over comparison approaches in terms of SI-SDRi and PESQ on WSJ0-2mix-extr and WHAM! dataset.

Neural Speaker Extraction with Speaker-Speech Cross-Attention Network

Wupeng Wang, Chenglin Xu, Meng Ge, Haizhou Li; NUS, Singapore

Thu-A-V-3-12, Time: 16:00

In this paper, we propose a novel time-domain speaker-speech cross-attention network as a variant of SpEx [1] architecture, that features speaker-speech cross-attention. The speaker-speech cross-attention network consists of speech semantic layers that capture the high-level dependency of audio feature, and cross-attention layers that fuse speaker embedding and speech features to estimate the speaker mask. We implement cross-attention layers with both

parallel and sequential concatenation techniques. Experiments show that the proposed models consistently outperform the state-of-theart time-domain speaker extraction baseline on WSJ0-2mix dataset.

Deep Audio-Visual Speech Separation Based on Facial Motion

*Rémi Rigal*¹, *Jacques Chodorowski*¹, *Benoît Zerr*²; ¹*Orange Labs, France;* ²*Lab-STICC (UMR 6285), France* Thu-A-V-3-13, Time: 16:00

We present a deep neural network that relies on facial motion and time-domain audio for isolating speech signals from a mixture of speeches and background noises. Recent studies in deep learningbased audio-visual speech separation and speech enhancement have proven that leveraging visual information in addition to audio can yield substantial improvement to the prediction quality and robustness. We propose to use facial motion, inferred from optical flow techniques, as a visual feature input for our model. Combined with state-of-the-art audio-only speech separation approaches, we demonstrate that facial motion significantly improves the speech quality as well as the versatility of the model. Our proposed method offers a signal-to-distortion improvement of up to 4.2 dB on twospeaker mixtures when compared to other audio-visual approaches.

Thu-A-V-4: Speaker Diarization II

16:00-18:00, Thursday 2 September 2021 Chairs: Andreas Stolcke and Leibny Paola García Perera

LEAP Submission for the Third DIHARD Diarization Challenge

Prachi Singh, Rajat Varma, Venkat Krishnamohan, Srikanth Raj Chetupalli, Sriram Ganapathy; Indian Institute of Science, India Thu-A-V-4-1, Time: 16:00

The LEAP submission for DIHARD-III challenge is described in this paper. The proposed system is composed of a speech bandwidth classifier, and diarization systems fine-tuned for narrowband and wideband speech separately. We use an end-to-end speaker diarization system for the narrowband conversational telephone speech recordings. For the wideband multi-speaker recordings, we use a neural embedding based clustering approach, similar to the baseline system. The embeddings are extracted from a time-delay neural network (called x-vectors) followed by the graph based path integral clustering (PIC) approach. The LEAP system showed 24% and 18% relative improvements for Track-1 and Track-2 respectively over the baseline system provided by the organizers. This paper describes the challenge submission, the post-evaluation analysis and improvements observed on the DIHARD-III dataset.

Investigation of Spatial-Acoustic Features for Overlapping Speech Detection in Multiparty Meetings

Shiliang Zhang¹, Siqi Zheng¹, Weilong Huang¹, Ming Lei¹, Hongbin Suo¹, Jinwei Feng², Zhijie Yan¹; ¹Alibaba, China; ²Alibaba, USA Thu-A-V-4-2, Time: 16:00

In this paper, we propose an overlapping speech detection (OSD) system for real multiparty meetings. Different from previous works on single-channel recordings or simulated data, we conduct research on real multi-channel data recorded by an 8-microphone array. We investigate how spatial information provided by multi-channel beamforming can benefit OSD. Specifically, we propose a two-stream

DFSMN to jointly model acoustic and spatial features. Instead of performing frame-level OSD, we try to perform segment-level OSD. We come up with an attention pooling layer to model speech segments with variable length. Experimental results show that two-stream DFSMN with attention pooling can effectively model acoustic-spatial feature and significantly boost the performance of OSD, result in 3.5% (from 85.57% to 89.12%) absolute detection accuracy improvement compared to the baseline system.

Target-Speaker Voice Activity Detection with Improved i-Vector Estimation for Unknown Number of Speaker

Maokui He¹, Desh Raj², Zili Huang², Jun Du¹, Zhuo Chen³, Shinji Watanabe²; ¹USTC, China; ²Johns Hopkins University, USA; ³Microsoft, USA Thu-A-V-4-3, Time: 16:00

Target-speaker voice activity detection (TS-VAD) has recently shown promising results for speaker diarization on highly overlapped speech. However, the original model requires a fixed (and known) number of speakers, which limits its application to real conversations. In this paper, we extend TS-VAD to speaker diarization with unknown numbers of speakers. This is achieved by two steps: first, an initial diarization system is applied for speaker number estimation, followed by TS-VAD network output masking according to this estimate. We further investigate different diarization methods, including clustering-based and region proposal networks, for estimating the initial i-vectors. Since these systems have complementary strengths, we propose a fusion-based method to combine framelevel decisions from the systems for an improved initialization. We demonstrate through experiments on variants of the LibriCSS meeting corpus that our proposed approach can improve the DER by up to 50% relative across varying numbers of speakers. This improvement also results in better downstream ASR performance approaching that using oracle segments.

ECAPA-TDNN Embeddings for Speaker Diarization

Nauman Dawalatabad¹, Mirco Ravanelli², François Grondin³, Jenthe Thienpondt⁴, Brecht Desplanques⁴, Hwidong Na⁵; ¹IIT Madras, India; ²Mila, Canada; ³Université de Sherbrooke, Canada; ⁴Ghent University, Belgium; ⁵Samsung, Korea

Thu-A-V-4-4, Time: 16:00

Learning robust speaker embeddings is a crucial step in speaker diarization. Deep neural networks can accurately capture speaker discriminative characteristics and popular deep embeddings such as x-vectors are nowadays a fundamental component of modern diarization systems. Recently, some improvements over the standard TDNN architecture used for x-vectors have been proposed. The ECAPA-TDNN model, for instance, has shown impressive performance in the speaker verification domain, thanks to a carefully designed neural model.

In this work, we extend, for the first time, the use of the ECAPA-TDNN model to speaker diarization. Moreover, we improved its robustness with a powerful augmentation scheme that concatenates several contaminated versions of the same signal within the same training batch. The ECAPA-TDNN model turned out to provide robust speaker embeddings under both close-talking and distant-talking conditions. Our results on the popular AMI meeting corpus show that our system significantly outperforms recently proposed approaches.

Advances in Integration of End-to-End Neural and Clustering-Based Diarization for Real Conversational Speech

Keisuke Kinoshita, Marc Delcroix, Naohiro Tawara; NTT, Japan

Thu-A-V-4-5, Time: 16:00

Recently, we proposed a novel speaker diarization method called End-to-End-Neural-Diarization-vector clustering (EEND-vector clustering) that integrates clustering-based and end-to-end neural network-based diarization approaches into one framework. The proposed method combines advantages of both frameworks, i.e. high diarization performance and handling of overlapped speech based on EEND, and robust handling of long recordings with an arbitrary number of speakers based on clustering-based approaches. However, the method was only evaluated so far on simulated 2speaker meeting-like data. This paper is to (1) report recent advances we made to this framework, including newly introduced robust constrained clustering algorithms, and (2) experimentally show that the method can now outperform competitive diarization methods such as Encoder-Decoder Attractor (EDA)-EEND, on CALLHOME data which comprises real conversational speech data including overlapped speech and an arbitrary number of speakers. By further analyzing the experimental results, this paper also discusses pros and cons of the proposed method and reveals potential for further improvement.

The Third DIHARD Diarization Challenge

*Neville Ryant*¹, *Prachi Singh*², *Venkat Krishnamohan*², *Rajat Varma*², *Kenneth Church*³, *Christopher Cieri*¹, *Jun Du*⁴, *Sriram Ganapathy*², *Mark Liberman*¹; ¹*University of Pennsylvania, USA*; ²*Indian Institute of Science, India*; ³*Baidu, USA*; ⁴*USTC, China* Thu-A-V-4-6. Time: 16:00

DIHARD III was the third in a series of speaker diarization challenges intended to improve the robustness of diarization systems to variability in recording equipment, noise conditions, and conversational domain. Speaker diarization was evaluated under two speech activity conditions (diarization from a reference speech activity vs. diarization from scratch) and 11 diverse domains. The domains span a range of recording conditions and interaction types, including read audio-books, meeting speech, clinical interviews, web videos, and, for the first time, conversational telephone speech. A total of 30 organizations (forming 21 teams) from industry and academia submitted 499 valid system outputs. The evaluation results indicate that speaker diarization has improved markedly since DIHARD I, particularly for two-party interactions, but that for many domains (e.g., web video) the problem remains far from solved.

Robust End-to-End Speaker Diarization with Conformer and Additive Margin Penalty

Tsun-Yat Leung, Lahiru Samarakoon; Fano Labs, China Thu-A-V-4-7, Time: 16:00

Traditionally, a speaker diarization system has multiple components to extract and cluster speaker embeddings. However, end-to-end diarization is more desirable as it facilitates optimizing one model in contrast to multiple components in a traditional set up. Moreover, end-to-end diarization systems are capable of handling overlapped speech. Recently proposed self-attentive end-to-end diarization model with encoder-decoder based attractors (EEND-EDA) is capable of processing speech from an unknown number of speakers, and has reported comparable performances to traditional systems. In this work, we aim to improve the EEND-EDA model. First, we increase

NOTES

the robustness of the model by incorporating an additive margin penalty for minimizing the intra-class variance. Second, we propose to replace the Transformer encoders with Conformer encoders to capture local information. Third, we propose to use convolutional subsampling and upsampling instead of manual subsampling only. Our proposed improvements report 21.6% relative reduction in DER on the evaluation full set of the track 2 of the DIHARD III challenge.

Anonymous Speaker Clusters: Making Distinctions Between Anonymised Speech Recordings with Clustering Interface

Benjamin O'Brien¹, Natalia Tomashenko², Anaïs Chanclu², Jean-François Bonastre²; ¹LPL (UMR 7309), France; ²LIA (EA 4128), France Thu-A-V-4-8, Time: 16:00

Our study examined the performance of evaluators tasked to group natural and anonymised speech recordings into clusters based on their perceived similarities. Speech stimuli were selected from the VCTK corpus; two systems developed for the VoicePrivacy 2020 Challenge were used for anonymisation. The Baseline-1 (B1) system was developed by using x-vectors and neural waveform models, while the Baseline-2 (B2) system relied on digital-signal-processing techniques. 74 evaluators completed three trials composed of 16 recordings with either natural or anonymised speech generated from a single system. F-measure and cluster purity metrics were used to assess evaluator accuracy. Probabilistic linear discriminant analysis (PLDA) scores from an automatic speaker verification system were generated to quantify similarity between recordings and used to correlate subjective results. Our findings showed that non-native English speaking evaluators significantly lowered their F-measure means when presented anonymised recordings. We observed no significance for cluster purity. Pearson correlation procedures revealed that PLDA scores generated from natural and B2-anonymised speech recordings correlated positively to F-measure and cluster purity metrics. These findings show evaluators were able to use the interface to cluster natural and anonymised speech recordings and suggest anonymisation systems modelled like B1 are more effective at suppressing identifiable speech characteristics.

Speaker Diarization Using Two-Pass Leave-One-Out Gaussian PLDA Clustering of DNN Embeddings

Kiran Karra, Alan McCree; Johns Hopkins University, USA

Thu-A-V-4-9, Time: 16:00

Many modern systems for speaker diarization, such as the recentlydeveloped VBx approach, rely on clustering of DNN speaker embeddings followed by resegmentation. Two problems with this approach are that the DNN is not directly optimized for this task, and the parameters need significant retuning for different applications. We have recently presented progress in this direction with a Leave-One-Out Gaussian PLDA (LGP) clustering algorithm and an approach to training the DNN such that embeddings directly optimize performance of this scoring method. This paper presents a new two-pass version of this system, where the second pass uses finer time resolution to significantly improve overall performance. For the Callhome corpus, we achieve the first published error rate below 4% without any task-dependent parameter tuning. We also show significant progress towards a robust single solution for multiple diarization tasks.

Thu-A-V-5: Speech Synthesis: Toward End-to-End Synthesis I

16:00–18:00, Thursday 2 September 2021 Chairs: Simon King and Wei-Ning Hsu

Federated Learning with Dynamic Transformer for Text to Speech

Zhenhou Hong, Jianzong Wang, Xiaoyang Qu, Jie Liu, Chendong Zhao, Jing Xiao; Ping An Technology, China Thu-A-V-5-1, Time: 16:00

Text to speech (TTS) is a crucial task for user interaction, but TTS model training relies on a sizable set of high-quality original datasets. Due to privacy and security issues, the original datasets are usually unavailable directly. Recently, federated learning proposes a popular distributed machine learning paradigm with an enhanced privacy protection mechanism. It offers a practical and secure framework for data owners to collaborate with others, thus obtaining a better global model trained on the larger dataset. However, due to the high complexity of transformer models, the convergence process becomes slow and unstable in the federated learning setting. Besides, the transformer model trained in federated learning is costly communication and limited computational speed on clients, impeding its popularity. To deal with these challenges, we propose the federated dynamic transformer. On the one hand, the performance is greatly improved comparing with the federated transformer, approaching centralize-trained Transformer-TTS when increasing clients number. On the other hand, it achieves faster and more stable convergence in the training phase and significantly reduces communication time. Experiments on the LJSpeech dataset also strongly prove our method's advantage.

LiteTTS: A Lightweight Mel-Spectrogram-Free Text-to-Wave Synthesizer Based on Generative Adversarial Networks

Huu-Kim Nguyen¹, Kihyuk Jeong¹, Seyun Um¹, Min-Jae Hwang², Eunwoo Song³, Hong-Goo Kang¹; ¹Yonsei University, Korea; ²Search Solutions, Korea; ³Naver, Korea

Thu-A-V-5-2, Time: 16:00

In this paper, we propose a lightweight end-to-end text-to-speech model that can generate high-quality speech at breakneck speed. In our proposed model, a feature prediction module and a waveform generation module are combined within a single framework. The feature prediction module, which consists of two independent sub-modules, estimates latent space embeddings for input text and prosodic information, and the waveform generation module generates speech waveforms by conditioning on the estimated latent space embeddings. Unlike conventional approaches that estimate prosodic information using a pre-trained model, our model jointly trains the prosodic embedding network with the speech waveform generation task using an effective domain transfer technique. Experimental results show that our proposed model can generate samples 7 times faster than real-time, and about 1.6 times faster than FastSpeech 2, as we use only 13.4 million parameters. We confirm that the generated speech quality is still of a high standard as evaluated by mean opinion scores.

Zero-Shot Text-to-Speech for Text-Based Insertion in Audio Narration

*Chuanxin Tang*¹, *Chong Luo*¹, *Zhiyuan Zhao*¹, *Dacheng Yin*², *Yucheng Zhao*², *Wenjun Zeng*¹; ¹*Microsoft, China*; ²*USTC, China* Thu-A-V-5-3, Time: 16:00

Given a piece of speech and its transcript text, text-based speech editing aims to generate speech that can be seamlessly inserted into the given speech by editing the transcript. Existing methods adopt a two-stage approach: synthesize the input text using a generic text-to-speech (TTS) engine and then transform the voice to the desired voice using voice conversion (VC). A major problem of this framework is that VC is a challenging problem which usually needs a moderate amount of parallel training data to work satisfactorily. In this paper, we propose a one-stage context-aware framework to generate natural and coherent target speech without any training data of the target speaker. In particular, we manage to perform accurate zero-shot duration prediction for the inserted text. The predicted duration is used to regulate both text embedding and speech embedding. Then, based on the aligned cross-modality input, we directly generate the mel-spectrogram of the edited speech with a transformer-based decoder. Subjective listening tests show that despite the lack of training data for the speaker, our method has achieved satisfactory results. It outperforms a recent zero-shot TTS engine by a large margin.

Diff-TTS: A Denoising Diffusion Model for Text-to-Speech

*Myeonghun Jeong*¹, *Hyeongju Kim*², *Sung Jun Cheon*¹, *Byoung Jin Choi*¹, *Nam Soo Kim*¹; ¹*Seoul National University, Korea*; ²*Neosapience, Korea* Thu-A-V-5-4, Time: 16:00

Although neural text-to-speech (TTS) models have attracted a lot of attention and succeeded in generating human-like speech, there is still room for improvements to its naturalness and architectural efficiency. In this work, we propose a novel non-autoregressive TTS model, namely Diff-TTS, which achieves highly natural and efficient speech synthesis. Given the text, Diff-TTS exploits a denoising diffusion framework to transform the noise signal into a mel-spectrogram via diffusion time steps. In order to learn the mel-spectrogram distribution conditioned on the text, we present a likelihood-based optimization method for TTS. Furthermore, to boost up the inference speed, we leverage the accelerated sampling method that allows Diff-TTS to generate raw waveforms much faster without significantly degrading perceptual quality. Through experiments, we verified that Diff-TTS generates 28 times faster than the real-time with a single NVIDIA 2080Ti GPU.

Hierarchical Context-Aware Transformers for Non-Autoregressive Text to Speech

Jae-Sung Bae, Taejun Bak, Young-Sun Joo, Hoon-Young Cho; NCSOFT, Korea Thu-A-V-5-5, Time: 16:00

In this paper, we propose methods for improving the modeling performance of a Transformer-based non-autoregressive text-to-speech (TNA-TTS) model. Although the text encoder and audio decoder handle different types and lengths of data (i.e., text and audio), the TNA-TTS models are not designed considering these variations. Therefore, to improve the modeling performance of the TNA-TTS model we propose a hierarchical Transformer structure-based text encoder and audio decoder that are designed to accommodate the characteristics of each module. For the text encoder, we constrain

each self-attention layer so the encoder focuses on a text sequence from the local to the global scope. Conversely, the audio decoder constrains its self-attention layers to focus in the reverse direction, i.e., from global to local scope. Additionally, we further improve the pitch modeling accuracy of the audio decoder by providing sentence and word-level pitch as conditions. Various objective and subjective evaluations verified that the proposed method outperformed the baseline TNA-TTS.

Speech Resynthesis from Discrete Disentangled Self-Supervised Representations

Adam Polyak¹, Yossi Adi¹, Jade Copet², Eugene Kharitonov², Kushal Lakhotia³, Wei-Ning Hsu³, Abdelrahman Mohamed³, Emmanuel Dupoux²; ¹Facebook, Israel; ²Facebook, France; ³Facebook, USA Thu-A-V-5-6, Time: 16:00

We propose using self-supervised discrete representations for the task of speech resynthesis. To generate disentangled representation, we separately extract low-bitrate representations for speech content, prosodic information, and speaker identity. This allows to synthesize speech in a controllable manner. We analyze various state-of-the-art, self-supervised representation learning methods and shed light on the advantages of each method while considering reconstruction quality and disentanglement properties. Specifically, we evaluate the F0 reconstruction, speaker identification performance (for both resynthesis and voice conversion), recordings' intelligibility, and overall quality using subjective human evaluation. Lastly, we demonstrate how these representations can be used for an ultra-lightweight speech codec. Using the obtained representations, we can get to a rate of 365 bits per second while providing better speech quality than the baseline methods. Audio samples are publicly available.

A Learned Conditional Prior for the VAE Acoustic Space of a TTS System

Penny Karanasou, Sri Karlapati, Alexis Moinet, Arnaud Joly, Ammar Abbas, Simon Slangen, Jaime Lorenzo-Trueba, Thomas Drugman; Amazon, UK Thu-A-V-5-7, Time: 16:00

Many factors influence speech yielding different renditions of a given sentence. Generative models, such as variational autoencoders (VAEs), capture this variability and allow multiple renditions of the same sentence via sampling. The degree of prosodic variability depends heavily on the prior that is used when sampling. In this paper, we propose a novel method to compute an informative prior for the VAE latent space of a neural text-to-speech (TTS) system. By doing so, we aim to sample with more prosodic variability, while gaining controllability over the latent space's structure.

By using as prior the posterior distribution of a secondary VAE, which we condition on a speaker vector, we can sample from the primary VAE taking explicitly the conditioning into account and resulting in samples from a specific region of the latent space for each condition (i.e. speaker). A formal preference test demonstrates significant preference of the proposed approach over standard Conditional VAE. We also provide visualisations of the latent space where well-separated condition-specific clusters appear, as well as ablation studies to better understand the behaviour of the system.

A Universal Multi-Speaker Multi-Style Text-to-Speech via Disentangled Representation Learning Based on Rényi Divergence Minimization

Dipjyoti Paul¹, Sankar Mukherjee², Yannis Pantazis³, Yannis Stylianou¹; ¹University of Crete, Greece; ²IIT, Italy; ³FORTH, Greece Thu-A-V-5-8, Time: 16:00

In this paper, we present a universal multi-speaker, multi-style Text-to-Speech (TTS) synthesis system which is able to generate speech from text with speaker characteristics and speaking style similar to a given reference signal. Training is conducted on nonparallel data and generates voices in an unsupervised manner, i.e., neither style annotation nor speaker label are required. To avoid leaking content information into the style embeddings (referred to as "content leakage") and leaking speaker information into style embeddings (referred to as "style leakage") we suggest a novel **R**ényi Divergence based Disentangled Representation framework through adversarial learning. Similar to mutual information minimization, the proposed approach explicitly estimates via a variational formula and then minimizes the Rényi divergence between the joint distribution and the product of marginals for the content-style and style-speaker pairs. By doing so, content, style and speaker spaces become representative and (ideally) independent of each other. Our proposed system greatly reduces content leakage by improving the word error rate by approximately 17-19% relative to the baseline system. In MOS-speech-quality, the proposed algorithm achieves an improvement of about 16-20% whereas MOS-style-similarly boost up 15% relative performance.

Relational Data Selection for Data Augmentation of Speaker-Dependent Multi-Band MelGAN Vocoder

Yi-Chiao Wu¹, Cheng-Hung Hu², Hung-Shin Lee², Yu-Huai Peng², Wen-Chin Huang¹, Yu Tsao², Hsin-Min Wang², Tomoki Toda¹; ¹Nagoya University, Japan; ²Academia Sinica, Taiwan Thu-A-V-5-9, Time: 16:00

Nowadays, neural vocoders can generate very high-fidelity speech when a bunch of training data is available. Although a speakerdependent (SD) vocoder usually outperforms a speaker-independent (SI) vocoder, it is impractical to collect a large amount of data of a specific target speaker for most real-world applications. To tackle the problem of limited target data, a data augmentation method based on speaker representation and similarity measurement of speaker verification is proposed in this paper. The proposed method selects utterances that have similar speaker identity to the target speaker from an external corpus, and then combines the selected utterances with the limited target data for SD vocoder adaptation. The evaluation results show that, compared with the vocoder adapted using only limited target data, the vocoder adapted using augmented data improves both the quality and similarity of synthesized speech.

Reinforce-Aligner: Reinforcement Alignment Search for Robust End-to-End Text-to-Speech

Hyunseung Chung, Sang-Hoon Lee, Seong-Whan Lee; Korea University, Korea Thu-A-V-5-10, Time: 16:00

Text-to-speech (TTS) synthesis is the process of producing synthesized speech from text or phoneme input. Traditional TTS models contain multiple processing steps and require external aligners, which provide attention alignments of phoneme-to-frame sequences. As the complexity increases and efficiency decreases with every additional step, there is expanding demand in modern synthesis

pipelines for end-to-end TTS with efficient internal aligners. In this work, we propose an end-to-end text-to-waveform network with a novel reinforcement learning based duration search method. Our proposed generator is feed-forward and the aligner trains the agent to make optimal duration predictions by receiving active feedback from actions taken to maximize cumulative reward. We demonstrate accurate alignments of phoneme-to-frame sequence generated from trained agents enhance fidelity and naturalness of synthesized audio. Experimental results also show the superiority of our proposed model compared to other state-of-the-art TTS models with internal and external aligners.

Triple M: A Practical Text-to-Speech Synthesis System with Multi-Guidance Attention and Multi-Band Multi-Time LPCNet

Shilun Lin, Fenglong Xie, Li Meng, Xinhui Li, Li Lu; Tencent, China

Thu-A-V-5-11, Time: 16:00

In this work, a robust and efficient text-to-speech (TTS) synthesis system named Triple M is proposed for large-scale online application. The key components of Triple M are: 1) A sequence-to-sequence model adopts a novel multi-guidance attention to transfer complementary advantages from guiding attention mechanisms to the basic attention mechanism without in-domain performance loss and online service modification. Compared with single attention mechanism, multi-guidance attention not only brings better naturalness to long sentence synthesis, but also reduces the word error rate by 26.8%. 2) A new efficient multi-band multi-time vocoder framework, which reduces the computational complexity from 2.8 to 1.0 GFLOP and speeds up LPCNet by $2.75 \times$ on a single CPU.

SC-GlowTTS: An Efficient Zero-Shot Multi-Speaker Text-To-Speech Model

Edresson Casanova¹, Christopher Shulby², Eren Gölge³, Nicolas Michael Müller⁴, Frederico Santos de Oliveira⁵, Arnaldo Candido Jr.⁶, Anderson da Silva Soares⁵, Sandra Maria Aluisio¹, Moacir Antonelli Ponti¹; ¹Universidade de São Paulo, Brazil; ²DefinedCrowd, USA; ³Coqui, Germany; ⁴Fraunhofer AISEC, Germany; ⁵Universidade Federal de Goiás, Brazil; ⁶Universidade Tecnológica Federal do Paraná, Brazil Thu-A-V-5-12, Time: 16:00

In this paper, we propose SC-GlowTTS: an efficient zero-shot multi-speaker text-to-speech model that improves similarity for speakers unseen during training. We propose a speaker-conditional architecture that explores a flow-based decoder that works in a zero-shot scenario. As text encoders, we explore a dilated residual convolutional-based encoder, gated convolutional-based encoder, and transformer-based encoder. Additionally, we have shown that adjusting a GAN-based vocoder for the spectrograms predicted by the TTS model on the training dataset can significantly improve the similarity and speech quality for new speakers. Our model converges using only 11 speakers, reaching state-of-the-art results for similarity with new speakers, as well as high speech quality.

Thu-A-V-6: Tools, Corpora and Resources

16:00–18:00, Thursday 2 September 2021 Chairs: Neville Ryant and Audrey Tong

Spoken ObjectNet: A Bias-Controlled Spoken Caption Dataset

Ian Palmer, Andrew Rouditchenko, Andrei Barbu, Boris Katz, James Glass; MIT, USA Thu-A-V-6-1, Time: 16:00

Visually-grounded spoken language datasets can enable models to learn cross-modal correspondences with very weak supervision. However, modern audio-visual datasets contain biases that undermine the real-world performance of models trained on that data. We introduce Spoken ObjectNet, which is designed to remove some of these biases and provide a way to better evaluate how effectively models will perform in real-world scenarios. This dataset expands upon ObjectNet, which is a bias-controlled image dataset that features similar image classes to those present in ImageNet.

We detail our data collection pipeline, which features several methods to improve caption quality, including automated language model checks. Lastly, we show baseline results on image retrieval and audio retrieval tasks. These results show that models trained on other datasets and then evaluated on Spoken ObjectNet tend to perform poorly due to biases in other datasets that the models have learned. We also show evidence that the performance decrease is due to the dataset controls, and not the transfer setting.

The Multilingual TEDx Corpus for Speech Recognition and Translation

*Elizabeth Salesky*¹, *Matthew Wiesner*¹, *Jacob Bremerman*², *Roldano Cattoni*³, *Matteo Negri*³, *Marco Turchi*³, *Douglas W. Oard*², *Matt Post*¹; ¹*Johns Hopkins University, USA*; ²*University of Maryland, USA*; ³*FBK, Italy*

Thu-A-V-6-2, Time: 16:00

We present the Multilingual TEDx corpus, built to support speech recognition (ASR) and speech translation (ST) research across many non-English source languages. The corpus is a collection of audio recordings from TEDx talks in 8 source languages. We segment transcripts into sentences and align them to the source-language audio and target-language translations. The corpus is released along with open-sourced code enabling extension to new talks and languages as they become available. Our corpus creation methodology can be applied to more languages than previous work, and creates multi-way parallel evaluation sets. We provide baselines in multiple ASR and ST settings, including multilingual models to improve translation performance for low-resource language pairs.

Tusom2021: A Phonetically Transcribed Speech Dataset from an Endangered Language for Universal Phone Recognition Experiments

*David R. Mortensen*¹, *Jordan Picone*², *Xinjian Li*¹, *Kathleen Siminyu*³; ¹*Carnegie Mellon University, USA*; ²*University of Pittsburgh, USA*; ³*Georgia Tech, USA* Thu-A-V-6-3, Time: 16:00

There is growing interest in ASR systems that can recognize phones in a language-independent fashion [1, 2, 3]. There is additionally interest in building language technologies for low-resource and endangered languages. However, there is a paucity of realistic data that can be used to test such systems and technologies. This paper

presents a publicly available, phonetically transcribed corpus of 2255 utterances (words and short phrases) in the endangered Tangkhulic language East Tusom (no ISO 639-3 code), a Tibeto-Burman language variety spoken mostly in India. Because the dataset is transcribed in terms of phones, rather than phonemes, it is a better match for universal phone recognition systems than many larger (phonemically transcribed) datasets. This paper describes the dataset and the methodology used to produce it. It further presents basic benchmarks of state-of-the-art universal phone recognition systems on the dataset as baselines for future experiments.

AISHELL-4: An Open Source Dataset for Speech Enhancement, Separation, Recognition and Speaker Diarization in Conference Scenario

Yihui Fu¹, Luyao Cheng¹, Shubo Lv¹, Yukai Jv¹, Yuxiang Kong¹, Zhuo Chen², Yanxin Hu¹, Lei Xie¹, Jian Wu³, Hui Bu⁴, Xin Xu⁴, Jun Du⁵, Jingdong Chen¹; ¹Northwestern Polytechnical University, China; ²Microsoft, USA; ³Microsoft, China; ⁴Beijing Shell Shell Technology, China; ⁵USTC, China Thu-A-V-6-4, Time: 16:00

In this paper, we present AISHELL-4, a sizable real-recorded Mandarin speech dataset collected by 8-channel circular microphone array for speech processing in conference scenario. The dataset consists of 211 recorded meeting sessions, each containing 4 to 8 speakers, with a total length of 120 hours. This dataset aims to bridge the advanced research on multi-speaker processing and the practical application scenario in three aspects. With real recorded meetings, AISHELL-4 provides realistic acoustics and rich natural speech characteristics in conversation such as short pause, speech overlap, quick speaker turn, noise, etc. Meanwhile, accurate transcription and speaker voice activity are provided for each meeting in AISHELL-4. This allows the researchers to explore different aspects in meeting processing, ranging from individual tasks such as speech front-end processing, speech recognition and speaker diarization, to multi-modality modeling and joint optimization of relevant tasks. Given most open source dataset for multi-speaker tasks are in English, AISHELL-4 is the only Mandarin dataset for conversation speech, providing additional value for data diversity in speech community. We also release a PyTorch-based training and evaluation framework as baseline system to promote reproducible research in this field.

GigaSpeech: An Evolving, Multi-Domain ASR Corpus with 10,000 Hours of Transcribed Audio

Guoguo Chen¹, Shuzhou Chai¹, Guan-Bo Wang¹, Jiayu Du¹, Wei-Qiang Zhang¹, Chao Weng², Dan Su², Daniel Povey³, Jan Trmal⁴, Junbo Zhang³, Mingjie Jin², Sanjeev Khudanpur⁴, Shinji Watanabe⁴, Shuaijiang Zhao⁵, Wei Zou⁵, Xiangang Li⁵, Xuchen Yao⁶, Yongqing Wang³, Zhao You², Zhiyong Yan³; ¹SpeechColab, China; ²Tencent, China; ³Xiaomi, China; ⁴Johns Hopkins University, USA; ⁵KE, China; ⁶Seasalt AI, USA Thu-A-V-6-5, Time: 16:00

This paper introduces GigaSpeech, an evolving, multi-domain English speech recognition corpus with 10,000 hours of high quality labeled audio suitable for supervised training, and 33,000 hours of total audio suitable for semi-supervised and unsupervised training. Around 33,000 hours of transcribed audio is first collected from audiobooks, podcasts and YouTube, covering both read and spontaneous speaking styles, and a variety of topics, such as arts, science, sports, etc. A new forced alignment and segmentation pipeline is proposed to create sentence segments suitable for speech

NOTES

recognition training, and to filter out segments with low-quality transcription. For system training, GigaSpeech provides five subsets of different sizes, 10h, 250h, 1000h, 2500h, and 10000h. For our 10,000-hour *XL* training subset, we cap the word error rate at 4% during the filtering/ validation stage, and for all our other smaller training subsets, we cap it at 0%. The *DEV* and *TEST* evaluation sets, on the other hand, are re-processed by professional human transcribers to ensure high transcription quality. Baseline systems are provided for popular speech recognition toolkits, namely Athena, ESPnet, Kaldi and Pika.

Look Who's Talking: Active Speaker Detection in the Wild

You Jin Kim, Hee-Soo Heo, Soyeon Choe, Soo-Whan Chung, Yoohwan Kwon, Bong-Jin Lee, Youngki Kwon, Joon Son Chung; Naver, Korea Thu-A-V-6-6, Time: 16:00

In this work, we present a novel audio-visual dataset for active speaker detection in the wild. A speaker is considered active when his or her face is visible and the voice is audible simultaneously. Although active speaker detection is a crucial pre-processing step for many audio-visual tasks, there is no existing active speaker detection dataset to evaluate the performance using natural human speech. We therefore curate the Active Speakers in the Wild (ASW) dataset which contains videos and co-occurring speech segments with dense speech activity labels. Videos and timestamps of audible segments are parsed and adopted from VoxConverse, an existing speaker diarisation dataset that consists of videos in the wild. Face tracks are extracted from the videos and active segments are annotated based on the timestamps of VoxConverse in a semi-automatic way. Two reference systems, one is self-supervised and the other is supervised system, are evaluated on the dataset to provide the baseline performances of ASW. Cross-domain evaluation and case study are conducted, in order to show the negative effect of the dubbed videos that are excluded in ASW.

AusKidTalk: An Auditory-Visual Corpus of 3- to 12-Year-Old Australian Children's Speech

Beena Ahmed¹, Kirrie J. Ballard², Denis Burnham³, Tharmakulasingam Sirojan¹, Hadi Mehmood¹, Dominique Estival³, Elise Baker³, Felicity Cox⁴, Joanne Arciuli⁵, Titia Benders⁴, Katherine Demuth⁴, Barbara Kelly⁶, Chloé Diskin-Holdaway⁷, Mostafa Shahin¹, Vidhyasaharan Sethu¹, Julien Epps¹, Chwee Beng Lee³, Eliathamby Ambikairajah¹; ¹UNSW Sydney, Australia; ²University of Sydney, Australia; ³Western Sydney University, Australia; ⁴Macquarie University, Australia; ⁵Flinders University, Australia; ⁶University of Melbourne, Australia; ⁷University of Mel bourne, Australia Thu-A-V-6-7, Time: 16:00

Here we present AusKidTalk [1], an audio-visual (AV) corpus of Australian children's speech collected to facilitate the development of speech based technological solutions for children. It builds upon the technology and expertise developed through the collection of an earlier corpus of Australian adult speech, AusTalk [2,3]. This multi-site initiative was established to remedy the dire shortage of children's speech corpora in Australia and around the world that are sufficiently sized to train accurate automated speech processing tools for children. We are collecting ~600 hours of speech from children aged 3-12 years that includes single word and sentence productions as well as narrative and emotional speech. In this paper, we discuss the key requirements for AusKidTalk and how we designed the recording setup and protocol to meet them. We also discuss key findings from our feasibility study of the recording protocol, recording tools, and user interface.

Human-in-the-Loop Efficiency Analysis for Binary Classification in Edyson

Per Fallgren, Jens Edlund; KTH, Sweden Thu-A-V-6-8, Time: 16:00

Edyson is a human-in-the-loop (HITL) tool for browsing and annotating large amounts of audio data quickly. It builds on temporally disassembled audio and massively multi-component audio environments to overcome the cumbersome time constraints that come with linear exploration of large audio data. This study adds the following contributions to Edyson: 1) We add the new use case of HITL binary classification by sample; 2) We explore the new domain oceanic hydrophone recordings with whale song, along with speech activity detection in noisy audio; 3) We propose a repeatable method of analysing the efficiency of HITL in Edyson for binary classification, specifically designed to measure the return on human time spent in a given domain. We exemplify this method on two domains, and show that for a manageable initial cost in terms of HITL, it does differentiate between suitable and unsuitable domains for our new use case — a valuable insight when working with large collections of audio.

Annotation Confidence vs. Training Sample Size: Trade-Off Solution for Partially-Continuous Categorical Emotion Recognition

Elena Ryumina, Oxana Verkholyak, Alexey Karpov; RAS, Russia

Thu-A-V-6-9, Time: 16:00

Commonly adapted design of emotional corpora includes multiple annotations for the same instance from several annotators. Most of the previous studies assume the ground truth to be an average between all labels or the most frequently used label. Current study shows that this approach may not be optimal for training. By filtering training data according to the level of annotation agreement, it is possible to increase the performance of the system even on unreliable test samples. However, increasing the annotation confidence inevitably leads to a loss of data. Therefore, balancing the trade-off between annotation quality and sample size requires careful investigation. This study presents experimental findings of audio-visual emotion classification on a recently introduced RAMAS dataset, which contains rich categorical partially-continuous annotation for 6 basic emotions, and reveals important conclusions about optimal formulation of ground truth. By applying the proposed approach, it is possible to achieve classification accuracy of UAR=70.51% on the speech utterances with more than 60% agreement, which surpasses previously reported values on this corpus in the literature.

Europarl-ASR: A Large Corpus of Parliamentary Debates for Streaming ASR Benchmarking and Speech Data Filtering/Verbatimization

Gonçal V. Garcés Díaz-Munío, Joan-Albert Silvestre-Cerdà, Javier Jorge, Adrià Giménez Pastor, Javier Iranzo-Sánchez, Pau Baquero-Arnal, Nahuel Roselló, Alejandro Pérez-González-de-Martos, Jorge Civera, Albert Sanchis, Alfons Juan; Universitat Politècnica de València, Spain Thu-A-V-6-10, Time: 16:00

We introduce Europarl-ASR, a large speech and text corpus of parliamentary debates including 1 300 hours of transcribed speeches and 70 million tokens of text in English extracted from European Parliament sessions. The training set is labelled with the Parliament's non-fully-verbatim official transcripts, time-aligned. As verbatimness is critical for acoustic model training, we also provide automatically noise-filtered and automatically verbatimized transcripts of all speeches based on speech data filtering and verbatimization techniques. Additionally, 18 hours of transcribed speeches were manually verbatimized to build reliable speaker-dependent and speaker-independent development/test sets for streaming ASR benchmarking. The availability of manual non-verbatim and verbatim transcripts for dev/test speeches makes this corpus useful for the assessment of automatic filtering and verbatimization techniques. This paper describes the corpus and its creation, and provides off-line and streaming ASR baselines for both the speaker-dependent and speaker-independent tasks using the three training transcription sets. The corpus is publicly released under an open licence.

Towards Automatic Speech to Sign Language Generation

Parul Kapoor¹, Rudrabha Mukhopadhyay², Sindhu B. Hegde², Vinay Namboodiri¹, C.V. Jawahar²; ¹IIT Kanpur, India; ²IIIT Hyderabad, India Thu-A-V-6-11, Time: 16:00

We aim to solve the highly challenging task of generating continuous sign language videos solely from speech segments for the first time. Recent efforts in this space have focused on generating such videos from human-annotated text transcripts without considering other modalities. However, replacing speech with sign language proves to be a practical solution while communicating with people suffering from hearing loss. Therefore, we eliminate the need of using text as input and design techniques that work for more natural, continuous, freely uttered speech covering an extensive vocabulary. Since the current datasets are inadequate for generating sign language directly from speech, we collect and release the first Indian sign language dataset comprising speech-level annotations, text transcripts, and the corresponding sign-language videos. Next, we propose a multitasking transformer network trained to generate signer's poses from speech segments. With speech-to-text as an auxiliary task and an additional cross-modal discriminator, our model learns to generate continuous sign pose sequences in an end-to-end manner. Extensive experiments and comparisons with other baselines demonstrate the effectiveness of our approach. We also conduct additional ablation studies to analyze the effect of different modules of our network. A demo video containing several results is attached to the supplementary material.

kosp2e: Korean Speech to English Translation Corpus

*Won Ik Cho*¹, *Seok Min Kim*¹, *Hyunchang Cho*², *Nam Soo Kim*¹; ¹*Seoul National University, Korea*; ²*Naver, Korea*

Thu-A-V-6-12, Time: 16:00

Most speech-to-text (S2T) translation studies use English speech as a source, which makes it difficult for non-English speakers to take advantage of the S2T technologies. For some languages, this problem was tackled through corpus construction, but the farther linguistically from English or the more under-resourced, this deficiency and underrepresentedness becomes more significant. In this paper, we introduce *kosp2e* (read as 'kospi'), a corpus that allows Korean speech to be translated into English text in an end-to-end manner. We adopt open license speech recognition corpus, translation corpus, and spoken language corpora to make our dataset freely available to the public, and check the performance through the pipeline and training-based approaches. Using pipeline and various end-to-end schemes, we obtain the highest BLEU of 21.3 and 18.0 for each based

on the English hypothesis, validating the feasibility of our data. We plan to supplement annotations for other target languages through community contributions in the future.

speechocean762: An Open-Source Non-Native English Speech Corpus for Pronunciation Assessment

Junbo Zhang¹, Zhiwen Zhang², Yongqing Wang¹, Zhiyong Yan¹, Qiong Song², Yukai Huang², Ke Li², Daniel Povey¹, Yujun Wang¹; ¹Xiaomi, China; ²SpeechOcean, China Thu-A-V-6-13, Time: 16:00

This paper introduces a new open-source speech corpus named "speechocean762" designed for pronunciation assessment use, consisting of 5000 English utterances from 250 non-native speakers, where half of the speakers are children. Five experts annotated each of the utterances at sentence-level, word-level and phoneme-level. A baseline system is released in open source to illustrate the phoneme-level pronunciation assessment workflow on this corpus. This corpus is allowed to be used freely for commercial and non-commercial purposes. It is available for free download from OpenSLR, and the corresponding baseline system is published in the Kaldi speech recognition toolkit.

Thu-A-SS-1: Non-Autoregressive Sequential Modeling for Speech Processing

Room D, 16:00–18:00, Thursday 2 September 2021 Chairs: Yuya Fujita and Katrin Kirchhoff

Introduction

Time: 16:00

Short Presentations

Time: 16:06

An Improved Single Step Non-Autoregressive Transformer for Automatic Speech Recognition

Ruchao Fan¹, Wei Chu², Peng Chang², Jing Xiao², Abeer Alwan¹; ¹University of California at Los Angeles, USA; ²PAII, USA

Thu-A-SS-1-1, Time: 16:45

Non-autoregressive mechanisms can significantly decrease inference time for speech transformers, especially when the single step variant is applied. Previous work on CTC alignment-based single step nonautoregressive transformer (CASS-NAT) has shown a large real time factor (RTF) improvement over autoregressive transformers (AT). In this work, we propose several methods to improve the accuracy of the end-to-end CASS-NAT, followed by performance analyses. First, convolution augmented self-attention blocks are applied to both the encoder and decoder modules. Second, we propose to expand the trigger mask (acoustic boundary) for each token to increase the robustness of CTC alignments. In addition, iterated loss functions are used to enhance the gradient update of low-layer parameters. Without using an external language model, the WERs of the improved CASS-NAT, when using the three methods, are 3.1%/7.2% on Librispeech test clean/other sets and the CER is 5.4% on the Aishell1 test set, achieving a 7%~21% relative WER/CER improvement. For the analyses, we plot attention weight distributions in the decoders to visualize the relationships between token-level acoustic embeddings. When the acoustic embeddings are visualized, we find that they have a similar behavior to word embeddings, which explains why the improved CASS-NAT performs similarly to AT.

NOTES

Multi-Speaker ASR Combining Non-Autoregressive Conformer CTC and Conditional Speaker Chain

Pengcheng Guo¹, Xuankai Chang², Shinji Watanabe², Lei Xie¹; ¹Northwestern Polytechnical University, China; ²Carnegie Mellon University, USA Thu-A-SS-1-2, Time: 16:45

Non-autoregressive (NAR) models have achieved a large inference computation reduction and comparable results with autoregressive (AR) models on various sequence to sequence tasks. However, there has been limited research aiming to explore the NAR approaches on sequence to multi-sequence problems, like multi-speaker automatic speech recognition (ASR). In this study, we extend our proposed conditional chain model to NAR multi-speaker ASR. Specifically, the output of each speaker is inferred one-by-one using both the input mixture speech and previously-estimated conditional speaker features. In each step, a NAR connectionist temporal classification (CTC) encoder is used to perform parallel computation. With this design, the total inference steps will be restricted to the number of mixed speakers. Besides, we also adopt the Conformer and incorporate an intermediate CTC loss to improve the performance. Experiments on WSJ0-Mix and LibriMix corpora show that our model outperforms other NAR models with only a slight increase of latency, achieving WERs of 22.3% and 24.9%, respectively. Moreover, by including the data of variable numbers of speakers, our model can even better than the PIT-Conformer AR model with only 1/7 latency, obtaining WERs of 19.9% and 34.3% on WSJ0-2mix and WSJ0-3mix sets. All of our codes are publicly available.

Pushing the Limits of Non-Autoregressive Speech Recognition

*Edwin G. Ng*¹, *Chung-Cheng Chiu*¹, *Yu Zhang*¹, *William Chan*²; ¹*Google, USA*; ²*Google, Canada* Thu-A-SS-1-3, Time: 16:45

We combine recent advancements in end-to-end speech recognition to non-autoregressive automatic speech recognition. We push the limits of non-autoregressive state-of-the-art results for multiple datasets: LibriSpeech, Fisher+Switchboard and Wall Street Journal. Key to our recipe, we leverage CTC on giant Conformer neural network architectures with SpecAugment and wav2vec2 pre-training. We achieve 1.8%/3.6% WER on LibriSpeech test/test-other sets, 5.1%/9.8% WER on Switchboard, and 3.4% on the Wall Street Journal, all without a language model.

Non-Autoregressive Predictive Coding for Learning Speech Representations from Local Dependencies

Alexander H. Liu, Yu-An Chung, James Glass; MIT, USA Thu-A-SS-1-4, Time: 16:45

Self-supervised speech representations have been shown to be effective in a variety of speech applications. However, existing representation learning methods generally rely on the autoregressive model and/or observed global dependencies while generating the representation. In this work, we propose Non-Autoregressive Predictive Coding (NPC), a self-supervised method, to learn a speech representation in a non-autoregressive manner by relying only on local dependencies of speech. NPC has a conceptually simple objective and can be implemented easily with the introduced Masked Convolution Blocks. NPC offers a significant speedup for inference since it is parallelizable in time and has a fixed inference time for each time step regardless of the input sequence length. We discuss and verify the effectiveness of NPC by theoretically and empirically comparing it with other methods. We show that the NPC representation is comparable to other methods in our experiments while being more efficient.

Relaxing the Conditional Independence Assumption of CTC-Based ASR by Conditioning on Intermediate Predictions

Jumon Nozaki, Tatsuya Komatsu; LINE, Japan Thu-A-SS-1-5, Time: 16:45

This paper proposes a method to relax the conditional independence assumption of connectionist temporal classification (CTC)-based automatic speech recognition (ASR) models. We train a CTC-based ASR model with auxiliary CTC losses in intermediate layers in addition to the original CTC loss in the last layer. During both training and inference, each generated prediction in the intermediate layers is summed to the input of the next layer to condition the prediction of the last layer on those intermediate predictions. Our method is easy to implement and retains the merits of CTC-based ASR: a simple model architecture and fast decoding speed. We conduct experiments on three different ASR corpora. Our proposed method improves a standard CTC model significantly (e.g., more than 20% relative word error rate reduction on the WSJ corpus) with a little computational overhead. Moreover, for the TEDLIUM2 corpus and the AISHELL-1 corpus, it achieves a comparable performance to a strong autoregressive model with beam search, but the decoding speed is at least 30 times faster.

Toward Streaming ASR with Non-Autoregressive Insertion-Based Model

Yuya Fujita¹, Tianzi Wang², Shinji Watanabe³, Motoi Omachi¹; ¹Yahoo, Japan; ²Johns Hopkins University, USA; ³Carnegie Mellon University, USA Thu-A-SS-1-6, Time: 16:45

Neural end-to-end (E2E) models have become a promising technique to realize practical automatic speech recognition (ASR) systems. When realizing such a system, one important issue is the segmentation of audio to deal with streaming input or long recording. After audio segmentation, the ASR model with a small real-time factor (RTF) is preferable because the latency of the system can be faster. Recently, E2E ASR based on non-autoregressive models becomes a promising approach since it can decode an N-length token sequence with less than N iterations. We propose a system to concatenate audio segmentation and non-autoregressive ASR to realize high accuracy and low RTF ASR. As a non-autoregressive ASR, the insertion-based model is used. In addition, instead of concatenating separated models for segmentation and ASR, we introduce a new architecture that realizes audio segmentation and non-autoregressive ASR by a single neural network. Experimental results on Japanese and English dataset show that the method achieved a reasonable trade-off between accuracy and RTF compared with baseline autoregressive Transformer and connectionist temporal classification.

Layer Pruning on Demand with Intermediate CTC

Jaesong Lee¹, Jingu Kang¹, Shinji Watanabe²; ¹Naver, Korea; ²Carnegie Mellon University, USA Thu-A-SS-1-7, Time: 16:45

Deploying an end-to-end automatic speech recognition (ASR) model on mobile/embedded devices is a challenging task, since the device computational power and energy consumption requirements are dynamically changed in practice. To overcome the issue, we present a training and pruning method for ASR based on the connectionist temporal classification (CTC) which allows reduction of model depth at run-time without any extra fine-tuning. To achieve the goal, we adopt two regularization methods, intermediate CTC and stochastic depth, to train a model whose performance does not degrade much after pruning. We present an in-depth analysis of layer behaviors

NOTES

using singular vector canonical correlation analysis (SVCCA), and efficient strategies for finding layers which are safe to prune. Using the proposed method, we show that a Transformer-CTC model can be pruned in various depth on demand, improving real-time factor from 0.005 to 0.002 on GPU, while each pruned sub-model maintains the accuracy of individually trained model of the same depth.

Real-Time End-to-End Monaural Multi-Speaker Speech Recognition

Song Li, Beibei Ouyang, Fuchuan Tong, Dexin Liao, Lin Li, Qingyang Hong; Xiamen University, China Thu-A-SS-1-8, Time: 16:45

The rising interest in single-channel multi-speaker speech separation has triggered the development of end-to-end multi-speaker automatic speech recognition (ASR). However, until now, most systems have adopted autoregressive mechanisms for decoding, resulting in slow decoding speed, which is not conducive to the application of multi-speaker speech recognition in real-world environments. In this paper, we first comprehensively investigate and compare the mainstream end-to-end multi-speaker speech recognition systems. Secondly, we improve the recently proposed non-autoregressive end-to-end speech recognition to achieve real-time decoding. Our experiments on the LibriMix data set show that under the premise of the same amount of parameters, the non-autoregressive model achieves performance close to that of the autoregressive model while having a faster decoding speed.

Streaming End-to-End ASR Based on Blockwise Non-Autoregressive Models

Tianzi Wang¹, Yuya Fujita², Xuankai Chang¹, Shinji Watanabe¹; ¹Johns Hopkins University, USA; ²Yahoo, Japan

Thu-A-SS-1-9, Time: 16:45

Non-autoregressive (NAR) modeling has gained more and more attention in speech processing. With recent state-of-the-art attentionbased automatic speech recognition (ASR) structure, NAR can realize promising real-time factor (RTF) improvement with only small degradation of accuracy compared to the autoregressive (AR) models. However, the recognition inference needs to wait for the completion of a full speech utterance, which limits their applications on low latency scenarios. To address this issue, we propose a novel end-to-end streaming NAR speech recognition system by combining blockwise-attention and connectionist temporal classification with mask-predict (Mask-CTC) NAR. During inference, the input audio is separated into small blocks and then processed in a blockwise streaming way. To address the insertion and deletion error at the edge of the output of each block, we apply an overlapping decoding strategy with a dynamic mapping trick that can produce more coherent sentences. Experimental results show that the proposed method improves online ASR recognition in low latency conditions compared to vanilla Mask-CTC. Moreover, it can achieve a much faster inference speed compared to the AR attention-based models. All of our codes will be publicly available.

TalkNet: Non-Autoregressive Depth-Wise Separable Convolutional Model for Speech Synthesis

Stanislav Beliaev, Boris Ginsburg; NVIDIA, USA Thu-A-SS-1-10, Time: 16:45

We propose TalkNet, a non-autoregressive convolutional neural model for speech synthesis with explicit pitch and duration prediction. The model consists of three feed-forward convolutional networks. The first network predicts grapheme durations. An input text is then expanded by repeating each symbol according to the predicted duration. The second network predicts pitch value for every mel frame. The third network generates a mel-spectrogram from the expanded text conditioned on predicted pitch. All networks are based on 1D depth-wise separable convolutional architecture. The explicit duration prediction eliminates word skipping and repeating. The quality of the generated speech nearly matches the best auto-regressive models — TalkNet trained on the LJSpeech dataset got a MOS of 4.08. The model has only 13.2M parameters, almost $2\times$ less than the present state-of-the-art text-to-speech models. The non-autoregressive architecture allows for fast training and inference. The small model size and fast inference make TalkNet an attractive candidate for embedded speech synthesis.

WaveGrad 2: Iterative Refinement for Text-to-Speech Synthesis

Nanxin Chen¹, Yu Zhang², Heiga Zen³, Ron J. Weiss², Mohammad Norouzi⁴, Najim Dehak¹, William Chan⁴; ¹Johns Hopkins University, USA; ²Google, USA; ³Google, Japan; ⁴Google, Canada

Thu-A-SS-1-11, Time: 16:45

This paper introduces WaveGrad 2, a non-autoregressive generative model for text-to-speech synthesis. WaveGrad 2 is trained to estimate the gradient of the log conditional density of the waveform given a phoneme sequence. The model takes an input phoneme sequence, and through an iterative refinement process, generates an audio waveform. This contrasts to the original WaveGrad vocoder which conditions on mel-spectrogram features, generated by a separate model. The iterative refinement process starts from Gaussian noise, and through a series of refinement steps (e.g., 50 steps), progressively recovers the audio sequence. WaveGrad 2 offers a natural way to trade-off between inference speed and sample quality, through adjusting the number of refinement steps. Experiments show that the model can generate high fidelity audio, approaching the performance of a state-of-the-art neural TTS system. We also report various ablation studies over different model configurations. Audio samples are publicly available.

Align-Denoise: Single-Pass Non-Autoregressive Speech Recognition

Nanxin Chen, Piotr Żelasko, Laureano Moro-Velázquez, Jesús Villalba, Najim Dehak; Johns Hopkins University, USA

Thu-A-SS-1-12, Time: 16:45

Deep autoregressive models start to become comparable or superior to the conventional systems for automatic speech recognition. However, for the inference computation, they still suffer from inference speed issue due to their token-by-token decoding characteristic. Non-autoregressive models greatly improve decoding speed by supporting decoding within a constant number of iterations. For example, Align-Refine was proposed to improve the performance of the non-autoregressive system by refining the alignment iteratively. In this work, we propose a new perspective to connect Align-Refine and denoising autoencoder. We introduce a novel noisy distribution to sample the alignment directly instead of obtaining it from the decoder output. The experimental results reveal that the proposed Align-Denoise speeds up both training and inference with performance improvement up to 5% relatively using single-pass decoding.

NOTES

VAENAR-TTS: Variational Auto-Encoder Based Non-AutoRegressive Text-to-Speech Synthesis

Hui Lu¹, Zhiyong Wu¹, Xixin Wu², Xu Li¹, Shiyin Kang³, Xunying Liu¹, Helen Meng¹; ¹CUHK, China; ²University of Cambridge, UK; ³Huya, China Thu-A-SS-I-13, Time: 16:45

This paper describes a variational auto-encoder based nonautoregressive text-to-speech (VAENAR-TTS) model. The autoregressive TTS (AR-TTS) models based on the sequence-to-sequence architecture can generate high-quality speech, but their sequential decoding process can be time-consuming. Recently, non-autoregressive TTS (NAR-TTS) models have been shown to be more efficient with the parallel decoding process. However, these NAR-TTS models rely on phoneme-level durations to generate a hard alignment between the text and the spectrogram. Obtaining duration labels, either through forced alignment or knowledge distillation, is cumbersome. Furthermore, hard alignment based on phoneme expansion can degrade the naturalness of the synthesized speech. In contrast, the proposed model of VAENAR-TTS is an end-to-end approach that does not require phoneme-level durations. The VAENAR-TTS model does not contain recurrent structures and is completely non-autoregressive in both the training and inference phases. Based on the VAE architecture, the alignment information is encoded in the latent variable, and attention-based soft alignment between the text and the latent variable is used in the decoder to reconstruct the spectrogram. Experiments show that VAENAR-TTS achieves stateof-the-art synthesis quality, while the synthesis speed is comparable with other NAR-TTS models.

Thu-A-SS-2: The ADReSSo Challenge: Detecting Cognitive Decline Using Speech Only

16:00-18:00, Thursday 2 September 2021 Chairs: Fasih Haider and Davida Fromm

Detecting Cognitive Decline Using Speech Only: The ADReSSo Challenge

Saturnino Luz¹, Fasih Haider¹, Sofia de la Fuente¹, Davida Fromm², Brian MacWhinney²; ¹University of Edinburgh, UK; ²Carnegie Mellon University, USA Thu-A-SS-2-1, Time: 16:00

Building on the success of the ADReSS Challenge at Interspeech 2020, which attracted the participation of 34 teams from across the world, the ADReSSo Challenge targets three difficult automatic prediction problems of societal and medical relevance, namely: detection of Alzheimer's Dementia, inference of cognitive testing scores, and prediction of cognitive decline. This paper presents these prediction tasks in detail, describes the datasets used, and reports the results of the baseline classification and regression models we developed for each task. A combination of acoustic and linguistic features extracted directly from audio recordings, without human intervention, yielded a baseline accuracy of 78.87% for the AD classification task, a root mean squared error (RMSE) of 5.28 for prediction of cognitive scores , and 68.75% accuracy ($F_1 = 66.67$) for the cognitive decline prediction task.
Influence of the Interviewer on the Automatic Assessment of Alzheimer's Disease in the Context of the ADReSSo Challenge

P.A. Pérez-Toro¹, S.P. Bayerl², T. Arias-Vergara¹, J.C. Vásquez-Correa¹, P. Klumpp¹, M. Schuster³, Elmar Nöth¹, J.R. Orozco-Arroyave¹, K. Riedhammer²; ¹FAU Erlangen-Nürnberg, Germany; ²TH Nürnberg, Germany; ³LMU München, Germany Thu-A-SS-2-2, Time: 16:00

Alzheimer's Disease (AD) results from the progressive loss of neurons in the hippocampus, which affects the capability to produce coherent language. It affects lexical, grammatical, and semantic processes as well as speech fluency. This paper considers the analyses of speech and language for the assessment of AD in the context of the Alzheimer's Dementia Recognition through Spontaneous Speech (ADReSSo) 2021 challenge. We propose to extract acoustic features such as X-vectors, prosody, and emotional embeddings as well as linguistic features such as perplexity, and word-embeddings. The data consist of speech recordings from AD patients and healthy controls. The transcriptions are obtained using a commercial automatic speech recognition system. We outperform baseline results on the test set, both for the classification and the Mini-Mental State Examination (MMSE) prediction. We achieved a classification accuracy of 80% and an RMSE of 4.56 in the regression. Additionally, we found strong evidence for the influence of the interviewer on classification results. In cross-validation on the training set, we get classification results of 85% accuracy using the combined speech of the interviewer and the participant. Using interviewer speech only we still get an accuracy of 78%. Thus, we provide strong evidence for interviewer influence on classification results.

WavBERT: Exploiting Semantic and Non-Semantic Speech Using Wav2vec and BERT for Dementia Detection

Youxiang Zhu¹, Abdelrahman Obyat¹, Xiaohui Liang¹, John A. Batsis², Robert M. Roth³; ¹UMass Boston, USA; ²University of North Carolina, USA; ³Geisel School of Medicine at Dartmouth, USA Thu-A-SS-2-3, Time: 16:00

In this paper, we exploit semantic and non-semantic information from patient's speech data using Wav2vec and Bidirectional Encoder Representations from Transformers (BERT) for dementia detection. We first propose a basic WavBERT model by extracting semantic information from speech data using Wav2vec, and analyzing the semantic information using BERT for dementia detection. While the basic model discards the non-semantic information, we propose extended WavBERT models that convert the output of Wav2vec to the input to BERT for preserving the non-semantic information in dementia detection. Specifically, we determine the locations and lengths of inter-word pauses using the number of blank tokens from Wav2vec where the threshold for setting the pauses is automatically generated via BERT. We further design a pre-trained embedding conversion network that converts the output embedding of Wav2vec to the input embedding of BERT, enabling the fine-tuning of WavBERT with non-semantic information. Our evaluation results using the ADReSSo dataset showed that the WavBERT models achieved the highest accuracy of 83.1% in the classification task, the lowest Root-Mean-Square Error (RMSE) score of 4.44 in the regression task, and a mean F1 of 70.91% in the progression task. We confirmed the effectiveness of WavBERT models exploiting both semantic and non-semantic speech.

NOTES

Alzheimer Disease Recognition Using Speech-Based Embeddings From Pre-Trained Models

Lara Gauder, Leonardo Pepino, Luciana Ferrer, Pablo Riera; UBA-CONICET ICC, Argentina Thu-A-SS-2-4, Time: 16:00

This paper describes our submission to the ADreSSo Challenge, which focuses on the problem of automatic recognition of Alzheimer's Disease (AD) from speech. The audio samples contain speech from the subjects describing a picture with the guidance of an experimenter. Our approach to the problem is based on the use of embeddings extracted from different pre-trained models - trill. allosaurus, and wav2vec 2.0 — which were trained to solve different speech tasks. These features are modeled with a neural network that takes short segments of speech as input, generating an AD score per segment. The final score for an audio file is given by the average over all segments in the file. We include ablation results to show the performance of different feature types individually and in combination, a study of the effect of the segment size, and an analysis of statistical significance. Our results on the test data for the challenge reach an accuracy of 78.9%, outperforming both the acoustic and linguistic baselines provided by the organizers.

Comparing Acoustic-Based Approaches for Alzheimer's Disease Detection

Aparna Balagopalan, Jekaterina Novikova; Winterlight Labs, Canada

Thu-A-SS-2-5, Time: 16:00

Robust strategies for Alzheimer's disease (AD) detection is important, given the high prevalence of AD. In this paper, we study the performance and generalizability of three approaches for AD detection from speech on the recent ADReSSo challenge dataset:1) using conventional acoustic features 2) using novel pre-trained acoustic embeddings 3) combining acoustic features and embeddings. We find that while feature-based approaches have a higher precision, classification approaches relying on the combination of embeddings and features prove to have a higher, and more balanced performance across multiple metrics of performance. Our best model, using such a combined approach, outperforms the acoustic baseline in the challenge by 2.8%.

Alzheimer's Disease Detection from Spontaneous Speech Through Combining Linguistic Complexity and (Dis)Fluency Features with Pretrained Language Models

*Yu Qiao*¹, *Xuefeng Yin*¹, *Daniel Wiechmann*², *Elma Kerz*¹; ¹*RWTH Aachen University, Germany;* ²*Universiteit van Amsterdam, The Netherlands* Thu-A-SS-2-6, Time: 16:00

In this paper, we combined linguistic complexity and (dis)fluency features with pretrained language models for the task of Alzheimer's disease detection of the 2021 ADReSSo (Alzheimer's Dementia Recognition through Spontaneous Speech) challenge. An accuracy of 83.1% was achieved on the test set, which amounts to an improvement of 4.23% over the baseline model. Our best-performing model that integrated component models using a stacking ensemble technique performed equally well on cross-validation and test data, indicating that it is robust against overfitting.

Using the Outputs of Different Automatic Speech Recognition Paradigms for Acoustic- and BERT-Based Alzheimer's Dementia Detection Through Spontaneous Speech

Yilin Pan¹, Bahman Mirheidari¹, Jennifer M. Harris², Jennifer C. Thompson², Matthew Jones², Julie S. Snowden², Daniel Blackburn¹, Heidi Christensen¹; ¹University of Sheffield, UK; ²University of Manchester, UK

Thu-A-SS-2-7, Time: 16:00

Exploring acoustic and linguistic information embedded in spontaneous speech recordings has proven to be efficient for automatic Alzheimer's dementia detection. Acoustic features can be extracted directly from the audio recordings, however, linguistic features, in fully automatic systems, need to be extracted from transcripts generated by an automatic speech recognition (ASR) system. We explore two state-of-the-art ASR paradigms, Wav2vec2.0 (for transcription and feature extraction) and time delay neural networks (TDNN) on the ADReSSo dataset containing recordings of people describing the Cookie Theft (CT) picture. As no manual transcripts are provided, we train an ASR system using our in-house CT data. We further investigate the use of confidence scores and multiple ASR hypotheses to guide and augment the input for the BERT-based classification. In total, five models are proposed for exploring how to use the audio recordings only for acoustic and linguistic information extraction. The test results on best acoustic-only and best linguistic-only are 74.65% and 84.51% respectively (representing a 15% and 9% relative increase to published baseline results).

Tackling the ADRESSO Challenge 2021: The MUET-RMIT System for Alzheimer's Dementia Recognition from Spontaneous Speech

Zafi Sherhan Syed¹, Muhammad Shehram Shah Syed², Margaret Lech², Elena Pirogova²; ¹MUET, Pakistan; ²RMIT University, Australia Thu-A-SS-2-8, Time: 16:00

This paper addresses the Interspeech Alzheimer's Dementia Recognition through Spontaneous Speech only (ADReSSo) challenge 2021. The objective of our study is to propose the approach to a three task automated screening that will aid in distinguishing between healthy individuals and subjects with dementia. The first task is to differentiate between speech recordings from individuals with dementia. The second task requires participants to estimate the Mini-Mental State Examination (MMSE) score based on an individual's speech. The third task requires participants to leverage speech recordings to identify whether individuals have suffered from cognitive decline. Here, we propose a system based on functionals of deep textual embeddings with special preprocessing steps integrating the effect of silence segments. We report that the developed system outperforms the challenge baseline for all three tasks. For Task 1, we achieve an accuracy of 84.51% compared to the baseline of 77.46%, for Task 2, we achieve a root-mean-square-error (RMSE) of 4.35 compared to the baseline of 5.28, and for Task 3, we achieve an average-f1score of 73.80% compared to the baseline of 66.67%. These results are a testament of the effectiveness of our proposed system.

Alzheimer's Dementia Recognition Using Acoustic, Lexical, Disfluency and Speech Pause Features Robust to Noisy Inputs

Morteza Rohanian, Julian Hough, Matthew Purver; Queen Mary University of London, UK Thu-A-SS-2-9, Time: 16:00

We present two multimodal fusion-based deep learning models that consume ASR transcribed speech and acoustic data simultaneously to classify whether a speaker in a structured diagnostic task has Alzheimer's Disease and to what degree, evaluating the ADReSSo challenge 2021 data. Our best model, a BiLSTM with highway layers using words, word probabilities, disfluency features, pause information, and a variety of acoustic features, achieves an accuracy of 84% and RSME error prediction of 4.26 on MMSE cognitive scores. While predicting cognitive decline is more challenging, our models show improvement using the multimodal approach and word probabilities, disfluency, and pause information over word-only models. We show considerable gains for AD classification using multimodal fusion and gating, which can effectively deal with noisy inputs from acoustic features and ASR hypotheses.

Automatic Detection and Assessment of Alzheimer Disease Using Speech and Language Technologies in Low-Resource Scenarios

Raghavendra Pappagari, Jaejin Cho, Sonal Joshi, Laureano Moro-Velázquez, Piotr Żelasko, Jesús Villalba, Najim Dehak; Johns Hopkins University, USA Thu-A-SS-2-10, Time: 16:00

In this study, we analyze the use of speech and speaker recognition technologies and natural language processing to detect Alzheimer disease (AD) and estimate mini-mental status evaluation (MMSE) scores. We used speech recordings from Interspeech 2021 ADReSSo challenge dataset. Our work focuses on adapting state-of-the-art speaker recognition and language models individually and later collectively to examine their complementary behavior for the tasks. We used speech embedding techniques such as x-vectors and prosody features to characterize the speech signals. We also employed automatic speech recognition (ASR) with interpolated language models to obtain transcriptions used to fine-tune the BERT models that classify and assess the speakers. Our results indicate that the fusion of scores obtained from the multiple acoustic and linguistic models provides the best detection results, suggesting that they contain complementary information. A separate analysis of the models indicates that linguistic models outperform acoustic models in detection and prediction tasks. However, acoustic models can provide better results than linguistic models under certain circumstances due to the errors in ASR transcriptions, which indicates that the performance of linguistic models relies on the performance of ASRs. Our best models provide 84.51% accuracy in automatic detection of AD and 3.85 RMSE in MMSE prediction.

Automatic Detection of Alzheimer's Disease Using Spontaneous Speech Only

*Jun Chen*¹, *Jieping Ye*¹, *Fengyi Tang*², *Jiayu Zhou*²; ¹*University of Michigan, USA*; ²*Michigan State*

University, USA Thu-A-SS-2-11, Time: 16:00

Alzheimer's disease (AD) is a neurodegenerative syndrome which affects tens of millions of elders worldwide. Although there is no treatment currently available, early recognition can improve the lives of people with AD and their caretakers and families. To find a cost-effective and easy-to-use method for dementia detection and ad-

dress the dementia classification task of InterSpeech 2021 ADReSSo (Alzheimer's Dementia Recognition through Spontaneous Speech only) challenge, we conduct a systematic comparison of approaches to detection of cognitive impairment based on spontaneous speech. We investigated the characteristics of acoustic modality and linguistic modality directly based on the audio recordings of narrative speech, and explored a variety of modality fusion strategies. With an ensemble over top-10 classifiers on the training set, we achieved an accuracy of 81.69% compared to the baseline of 78.87% on the test set. The results suggest that although transcription errors will be introduced through automatic speech recognition, integrating textual information generally improves classification performance. Besides, ensemble methods can boost both the accuracy and the robustness of models.

Modular Multi-Modal Attention Network for Alzheimer's Disease Detection Using Patient Audio and Language Data

*Ning Wang*¹, *Yupeng Cao*¹, *Shuai Hao*¹, *Zongru Shao*², *K.P. Subbalakshmi*¹; ¹*Stevens Institute of Technology, USA*; ²*CASUS, Germany*

Thu-A-SS-2-12, Time: 16:00

In this work, we propose a modular multi-modal architecture to automatically detect Alzheimer's disease using the dataset provided in the ADReSSo challenge. Both acoustic and text-based features are used in this architecture. Since the dataset provides only audio samples of controls and patients, we use Google cloud-based speech-to-text API to automatically transcribe the audio files to extract text-based features. Several kinds of audio features are extracted using standard packages. The proposed approach consists of 4 networks: C-attention-acoustic network (for acoustic features only), C-Attention-FT network (for linguistic features only), C-Attention-Embedding network (for language embeddings and acoustic embeddings), and a unified network (uses all of those features). The architecture combines attention networks and a convolutional neural network (C-Attention network) in order to process these features. Experimental results show that the C-Attention-Unified network with Linguistic features and X-Vector embeddings achieves the best accuracy of 80.28% and F1 score of 0.825 on the test dataset.

Fri-M-O-1: Robust and Far-Field ASR

Room A+B, 11:00-13:00, Friday 3 September 2021 Chairs: Richard Stern and Alessio Brutti

Self-Attention Channel Combinator Frontend for End-to-End Multichannel Far-Field Speech Recognition

Rong Gong¹, Carl Quillen², Dushyant Sharma², Andrew Goderre², José Laínez³, Ljubomir Milanović¹; ¹Nuance Communications, Austria; ²Nuance Communications, USA; ³Nuance Communications, Spain Fri-M-O-1-1, Time: 11:00

When a sufficiently large far-field training data is presented, jointly optimizing a multichannel frontend and an end-to-end (E2E) Automatic Speech Recognition (ASR) backend shows promising results. Recent literature has shown traditional beamformer designs, such as MVDR (Minimum Variance Distortionless Response) or fixed beamformers can be successfully integrated as the frontend into an E2E ASR system with learnable parameters. In this work, we propose the self-attention channel combinator (SACC) ASR frontend, which leverages the self-attention mechanism to combine multichannel audio signals in the magnitude spectral domain. Experiments conducted on a multichannel playback test data shows that the SACC achieved a 9.3% WERR compared to a state-of-the-art fixed beamformer-based frontend, both jointly optimized with a ContextNet-based ASR backend. We also demonstrate the connection between the SACC and the traditional beamformers, and analyze the intermediate outputs of the SACC.

ETLT 2021: Shared Task on Automatic Speech Recognition for Non-Native Children's Speech

R. Gretter¹, Marco Matassoni¹, D. Falavigna¹, A. Misra², C.W. Leong², K. Knill³, L. Wang³; ¹FBK, Italy; ²Educational Testing Service, USA; ³University of Cambridge, UK Fri-M-O-1-2, Time: 11:20

The paper presents the Second ASR Challenge for Non-native Children's Speech proposed as a Special Session at Interspeech 2021, following the successful first challenge at Interspeech 2020. The goal of the challenge is to advance research on non-native children's speech recognition technology, as speech technology still struggles when applied to both children and non-native speakers. The audio data consists of spoken responses provided by L2 students in the context of both English and German speaking proficiency examinations, the latter language added for 2021. Additional training data and a new evaluation set was released for L2 English recorded by speakers of different native languages. Participants could build systems for one or both languages. Each had a closed track where a predetermined set of audio and linguistic resources were selected, and an open track where additional data was allowed. After a description of the released corpora, the paper analyzes the results achieved by the participating systems. Some issues suggested from these results are discussed.

Age-Invariant Training for End-to-End Child Speech Recognition Using Adversarial Multi-Task Learning

Lars Rumberg, Hanna Ehlert, Ulrike Lüdtke, Jörn Ostermann; Leibniz Universität Hannover, Germany Fri-M-O-1-3, Time: 11:40

Automatic speech recognition for children's speech is a challenging task mainly due to scarcity of publicly available child speech corpora

and wide inter- and intra-speaker variability in terms of acoustic and linguistic characteristics of children's speech. We propose a framework for age-invariant training of the acoustic model of end-to-end speech recognition systems based on adversarial multi-task learning. We use age information additionally to just differentiating between the child and adult domains and thus force the acoustic model to learn age invariant features. Our results on publicly available data sets show that this leads to better leveraging of existing data during training. We further show that usage of adversarial multitask learning should not necessarily be regarded as a substitute for traditional feature space adaptation methods, but that both should be used together for best performance.

Learning to Rank Microphones for Distant Speech Recognition

Samuele Cornell¹, Alessio Brutti², Marco Matassoni², Stefano Squartini¹; ¹Università Politecnica delle Marche, Italy; ²FBK, Italy Fri-M-O-1-4, Time: 12:00

Fully exploiting ad-hoc microphone networks for distant speech recognition is still an open issue. Empirical evidence shows that being able to select the best microphone leads to significant improvements in recognition without any additional effort on front-end processing. Current channel selection techniques either rely on signal, decoder or posterior-based features. Signal-based features are inexpensive to compute but do not always correlate with recognition performance. Instead decoder and posterior-based features exhibit better correlation but require substantial computational resources.

In this work, we tackle the channel selection problem by proposing MicRank, a learning to rank framework where a neural network is trained to rank the available channels using directly the recognition performance on the training set. The proposed approach is agnostic with respect to the array geometry and type of recognition back-end. We investigate different learning to rank strategies using a synthetic dataset developed on purpose and the CHiME-6 data. Results show that the proposed approach considerably improves over previous selection techniques, reaching comparable and in some instances better performance than oracle signal-based measures.

Simulating Reading Mistakes for Child Speech Transformer-Based Phone Recognition

*Lucile Gelin*¹, *Thomas Pellegrini*¹, *Julien Pinquier*¹, *Morgane Daniel*²; ¹*IRIT (UMR 5505), France;* ²*Lalilo, France*

Fri-M-O-1-5, Time: 12:20

Current performance of automatic speech recognition (ASR) for children is below that of the latest systems dedicated to adult speech. Child speech is particularly difficult to recognise, and substantial corpora are missing to train acoustic models. Furthermore, in the scope of our reading assistant for 5-8-year-old children learning to read, models need to cope with disfluencies and reading mistakes, which remain considerable challenges even for state-of-the-art ASR systems. In this paper, we adapt an end-to-end Transformer acoustic model to speech from children learning to read. Transfer learning (TL) with a small amount of child speech improves the phone error rate (PER) by 48.7% relative over an adult model and outperforms a TL-adapted DNN-HMM model by 21.0% relative PER. Multi-objective training with a Connectionist Temporal Classification (CTC) function further reduces the PER by 4.8% relative. We propose a method of reading mistakes data augmentation, where we simulate word-level repetitions and substitutions with phonetically or graphically close words. Combining these two types of reading mistakes reaches a 19.9% PER, with a 13.1% relative improvement over the baseline. A

NOTES

detailed analysis shows that both the CTC multi-objective training and the augmentation with synthetic repetitions help the attention mechanisms better detect children's disfluencies.

Fri-M-O-2: Speech Synthesis: Prosody Modeling II

Room C, 11:00–13:00, Friday 3 September 2021 Chairs: Esther Klabbers and Gérard Bailly

Alternate Endings: Improving Prosody for Incremental Neural TTS with Predicted Future Text Input

Brooke Stephenson¹, Thomas Hueber¹, Laurent Girin¹, Laurent Besacier²; ¹GIPSA-lab (UMR 5216), France; ²LIG (UMR 5217), France Fri-M-O-2-1, Time: 11:00

Inferring the prosody of a word in text-to-speech synthesis requires information about its surrounding context. In incremental textto-speech synthesis, where the synthesizer produces an output before it has access to the complete input, the full context is often unknown which can result in a loss of naturalness. In this paper, we investigate whether the use of predicted future text from a transformer language model can attenuate this loss in a neural TTS system. We compare several test conditions of next future word: (a) unknown (zero-word), (b) language model predicted, (c) randomly predicted and (d) ground-truth. We measure the prosodic features (pitch, energy and duration) and find that predicted text provides significant improvements over a zero-word lookahead, but only slight gains over random-word lookahead. We confirm these results with a perceptive test.

Exploring Emotional Prototypes in a High Dimensional TTS Latent Space

Pol van Rijn¹, Silvan Mertes², Dominik Schiller², Peter M.C. Harrison¹, Pauline Larrouy-Maestri¹, Elisabeth André², Nori Jacoby¹; ¹MPI for Empirical Aesthetics, Germany; ²Universität Augsburg, Germany Fri-M-O-2-2, Time: 11:20

Recent TTS systems are able to generate prosodically varied and realistic speech. However, it is unclear how this prosodic variation contributes to the perception of speakers' emotional states. Here we use the recent psychological paradigm 'Gibbs Sampling with People' to search the prosodic latent space in a trained Global Style Token Tacotron model to explore prototypes of emotional prosody. Participants are recruited online and collectively manipulate the latent space of the generative speech model in a sequentially adaptive way so that the stimulus presented to one group of participants is determined by the response of the previous groups. We demonstrate that (1) particular regions of the model's latent space are reliably associated with particular emotions, (2) the resulting emotional prototypes are well-recognized by a separate group of human raters, and (3) these emotional prototypes can be effectively transferred to new sentences. Collectively, these experiments demonstrate a novel approach to the understanding of emotional speech by providing a tool to explore the relation between the latent space of generative models and human semantics.

Ctrl-P: Temporal Control of Prosodic Variation for Speech Synthesis

Devang S. Ram Mohan, Vivian Hu, Tian Huey Teh, Alexandra Torresquintero, Christopher G.R. Wallis, Marlene Staib, Lorenzo Foglianti, Jiameng Gao, Simon King; Papercup Technologies, UK Fri-M-O-2-3, Time: 11:40

Text does not fully specify the spoken form, so text-to-speech models must be able to learn from speech data that vary in ways not explained by the corresponding text. One way to reduce the amount of unexplained variation in training data is to provide acoustic information as an additional learning signal. When generating speech, modifying this acoustic information enables multiple distinct renditions of a text to be produced.

Since much of the unexplained variation is in the prosody, we propose a model that generates speech explicitly conditioned on the three primary acoustic correlates of prosody: F_0 , energy and duration. The model is flexible about how the values of these features are specified: they can be externally provided, or predicted from text, or predicted then subsequently modified.

Compared to a model that employs a variational auto-encoder to learn unsupervised latent features, our model provides more interpretable, temporally-precise, and disentangled control. When automatically predicting the acoustic features from text, it generates speech that is more natural than that from a Tacotron 2 model with reference encoder. Subsequent human-in-the-loop modification of the predicted acoustic features can significantly further increase naturalness.

ADEPT: A Dataset for Evaluating Prosody Transfer

Alexandra Torresquintero, Tian Huey Teh, Christopher G.R. Wallis, Marlene Staib, Devang S. Ram Mohan, Vivian Hu, Lorenzo Foglianti, Jiameng Gao, Simon King; Papercup Technologies, UK Fri-M-O-2-4, Time: 12:00

Text-to-speech is now able to achieve near-human naturalness and research focus has shifted to increasing expressivity. One popular method is to transfer the prosody from a reference speech sample. There have been considerable advances in using prosody transfer to generate more expressive speech, but the field lacks a clear definition of what successful prosody transfer means and a method for measuring it. We introduce a dataset of prosodically-varied reference natural speech samples for evaluating prosody transfer. The samples include global variations reflecting emotion and interpersonal attitude, and local variations reflecting topical emphasis, propositional attitude, syntactic phrasing and marked tonicity. The corpus only includes prosodic variations that listeners are able to distinguish with reasonable accuracy, and we report these figures as a benchmark against which text-to-speech prosody transfer can be compared. We conclude the paper with a demonstration of our proposed evaluation methodology, using the corpus to evaluate two text-to-speech models that perform prosody transfer.

Prosodic Boundary Prediction Model for Vietnamese Text-To-Speech

Nguyen Thi Thu Trang¹, Nguyen Hoang Ky¹, Albert Rilliard², Christophe d'Alessandro³; ¹Hanoi University of Science & Technology, Vietnam; ²LISN (UMR 9015), France; ³∂'Alembert (UMR 7190), France Fri-M-O-2-5, Time: 12:20

This research aims to build a prosodic boundary prediction model for improving the naturalness of Vietnamese speech synthesis. This model can be used directly to predict prosodic boundaries in the synthesis phase of the statistical parametric or end-to-end speech Beside conventional features related to Part-Of-Speech systems. (POS), this paper proposes two efficient features to predict prosodic boundaries: syntactic blocks and syntactic links, based on a thorough analysis of a Vietnamese dataset. Syntactic blocks are syntactic phrases whose sizes are bounded in their constituent syntactic tree. A syntactic link of two adjacent words is calculated based on the distance between them in the syntax tree. The experimental results show that the two proposed predictors improve the quality of the boundary prediction model using a decision tree classification algorithm, about 36.4% (F1 score) higher than the model with only POS features. The final boundary prediction model with POS, syntactic block, and syntactic link features using the LightGBM algorithm gives the best F1-score results at 87.0% in test data. The proposed model helps the TTS systems, developed by either HMM-based, DNN-based, or End-to-end speech synthesis techniques, improve about 0.3 MOS points (i.e. 6 to 10%) compared to the ones without the proposed model.

Fri-M-O-3: Source Separation III

Room D, 11:00–13:00, Friday 3 September 2021 Chairs: Zbynek Koldovský and Kateřina Žmolíková

Many-Speakers Single Channel Speech Separation with Optimal Permutation Training

Shaked Dovrat, Eliya Nachmani, Lior Wolf; Tel Aviv University, Israel

Fri-M-O-3-1, Time: 11:00

Single channel speech separation has experienced great progress in the last few years. However, training neural speech separation for a large number of speakers (e.g., more than 10 speakers) is out of reach for the current methods, which rely on the Permutation Invariant Training (PIT). In this work, we present a permutation invariant training that employs the Hungarian algorithm in order to train with an $O(C^3)$ time complexity, where *C* is the number of speakers, in comparison to O(C!) of PIT based methods. Furthermore, we present a modified architecture that can handle the increased number of speakers. Our approach separates up to 20 speakers and improves the previous results for large *C* by a wide margin.

Combating Reverberation in NTF-Based Speech Separation Using a Sub-Source Weighted Multichannel Wiener Filter and Linear Prediction

Mieszko Fraś, Marcin Witkowski, Konrad Kowalczyk; AGH UST, Poland Fri-M-O-3-2, Time: 11:20

Sound source separation (SS) from the microphone signals capturing speech in reverberant conditions is a formidable task. This paper addresses the problem of joint separation and dereverberation of speech using the multichannel Wiener filter (MWF) that is tailored to the sub-source modeling of each speech source with a full-rank mixing matrix. Specifically, the parameters of the proposed sub-source-weighted (SSW) spatial filter are estimated using the sub-source based expectation maximization (EM) algorithm with multiplicative updates (MU) and the localization prior distribution (LP) on the mixing matrix (SSEM-MU-LP). In addition, we strengthen dereverberation by incorporating a Generalized Weighted Prediction Error (GWPE) algorithm. The proposed method is evaluated using a large dataset of two-channel recordings of clean speech convolved with both real and synthesized impulse responses. The results of the experiments show the superior performance of the

proposed method in reverberant conditions in comparison to using the standard NTF-based separation with the vanilla MWF in terms of signal-to-distortion ratio (improvement of 3–5.6 dB) and other commonly used sound separation metrics.

A Hands-On Comparison of DNNs for Dialog Separation Using Transfer Learning from Music Source Separation

Martin Strauss¹, Jouni Paulus¹, Matteo Torcoli², Bernd Edler¹; ¹AudioLabs, Germany; ²Fraunhofer IIS, Germany

Fri-M-O-3-3, Time: 11:40

This paper describes a hands-on comparison on using state-of-the-art music source separation deep neural networks (DNNs) before and after task-specific fine-tuning for separating speech content from non-speech content in broadcast audio (i.e., dialog separation). The music separation models are selected as they share the number of channels (2) and sampling rate (44.1 kHz or higher) with the considered broadcast content, and vocals separation in music is considered as a parallel for dialog separation in the target application domain. These similarities are assumed to enable transfer learning between the tasks. Three models pre-trained on music (Open-Unmix, Spleeter, and Conv-TasNet) are considered in the experiments, and fine-tuned with real broadcast data. The performance of the models is evaluated before and after fine-tuning with computational evaluation metrics (SI-SIRi, SI-SDRi, 2f-model), as well as with a listening test simulating an application where the non-speech signal is partially attenuated, e.g., for better speech intelligibility. The evaluations include two reference systems specifically developed for dialog separation. The results indicate that pre-trained music source separation models can be used for dialog separation to some degree, and that they benefit from the fine-tuning, reaching a performance close to task-specific solutions.

GlobalPhone Mix-To-Separate Out of 2: A Multilingual 2000 Speakers Mixtures Database for Speech Separation

*Marvin Borsdorf*¹, *Chenglin Xu*², *Haizhou Li*², *Tanja Schultz*¹; ¹*Universität Bremen, Germany*; ²*NUS, Singapore*

Fri-M-O-3-4, Time: 12:00

Monaural speech separation has been well studied on various databases. However, these databases mostly concern English speech. Research in multi-speaker scenarios, such as speech recognition, speaker recognition, speaker diarization, and speech separation calls for speaker mixtures databases comprising multiple languages. In this paper, we propose a new extensive multilingual database for speech separation tasks derived from the GlobalPhone 2000 Speaker Package, called "GlobalPhone Mix-to-Separate out of 2" (GlobalPhoneMS2). We describe the construction of the database and conduct speech separation experiments in monolingual and multilingual as well as seen and unseen languages settings. When trained on a multilingual dataset, the networks improve their performances for unseen languages, and across almost all seen languages. We show that replacing a monolingual dataset with a trilingual one, while keeping the data size roughly the same, helps to improve the performance in most cases. We attribute this to a larger diversity in speech, language, speaker, and recording characteristics. Based on the GlobalPhoneMS2 database, speech separation results for two-speaker mixing scenarios are reported in 22 spoken languages for the first time.

Fri-M-V-1: Non-Native Speech

11:00-13:00, Friday 3 September 2021 Chairs: Joakim Gustafson and Anne Bonneau

Cross-Linguistic Perception of the Japanese Singleton/Geminate Contrast: Korean, Mandarin and Mongolian Compared

*Kimiko Tsukada*¹, *Yurong*², *Joo-Yeon Kim*³, *Jeong-Im Han*³, *John Hajek*⁴; ¹*Macquarie University, Australia;* ²*Inner Mongolia University, China;* ³*Konkuk University, Korea;* ⁴*University of Melbourne, Australia* Fri-M-V-1-1, Time: 11:00

The perception of Japanese consonant length contrasts (i.e. short/singleton vs long/geminate) by native and non-native speakers was compared to examine the extent to which difficult foreign language (FL) sounds are processed accurately. Three groups of participants had Korean, Mandarin or Mongolian as their first language (L1) and had no experience with Japanese. Unlike Japanese, Mandarin and Mongolian do not use consonant length contrastively. The phonemic status of consonant length in Korean is debatable. Further, unlike Japanese and Mandarin which predominantly use open syllables and restrict the occurrence of consonants in coda position, Korean and Mongolian permit a wide range of consonants in that syllable position. Via the AXB task, the participants' discrimination accuracy of Japanese consonant length contrasts was assessed and compared to that of a group of 10 native Japanese speakers who served as controls. The Japanese group was at near ceiling with little individual variation. The Mongolian (but not Korean and Mandarin) group did not significantly differ from the control group when the target token (X) contained a geminate. All non-native groups were significantly less accurate than the control group when X contained a singleton. These results were interpreted as reflecting the participants' L1 quantity system.

Detection of Lexical Stress Errors in Non-Native (L2) English with Data Augmentation and Attention

Daniel Korzekwa¹, Roberto Barra-Chicote², Szymon Zaporowski³, Grzegorz Beringer¹, Jaime Lorenzo-Trueba², Alicja Serafinowicz¹, Jasha Droppo⁴, Thomas Drugman², Bozena Kostek³; ¹Amazon, Poland; ²Amazon, UK; ³Gdansk University of Technology, Poland; ⁴Amazon, USA Fri-M-V-1-2, Time: 11:00

This paper describes two novel complementary techniques that improve the detection of lexical stress errors in non-native (L2) English speech: attention-based feature extraction and data augmentation based on Neural Text-To-Speech (TTS). In a classical approach, audio features are usually extracted from fixed regions of speech such as the syllable nucleus. We propose an attention-based deep learning model that automatically derives optimal syllable-level representation from frame-level and phoneme-level audio features. Training this model is challenging because of the limited amount of incorrect stress patterns. To solve this problem, we propose to augment the training set with incorrectly stressed words generated with Neural TTS. Combining both techniques achieves 94.8% precision and 49.2% recall for the detection of incorrectly stressed words in L2 English speech of Slavic and Baltic speakers.

Testing Acoustic Voice Quality Classification Across Languages and Speech Styles

Bettina Braun¹, Nicole Dehé¹, Marieke Einfeldt¹, Daniela Wochner¹, Katharina Zahner-Ritter²; ¹Universität Konstanz, Germany; ²Universität Trier, Germany

Fri-M-V-1-3, Time: 11:00

Many studies relate acoustic voice quality measures to perceptual classification. We extend this line of research by training a classifier on a balanced set of perceptually annotated voice quality categories with high inter-rater agreement, and test it on speech samples from a different language and on a different speech style. Annotations were done on continuous speech from different laboratory settings. In Experiment 1, we trained a random forest with Standard Chinese and German recordings labelled as modal, breathy, or glottalized. The model had an accuracy of 78.7% on unseen data from the same sample (most important variables were harmonics-to-noise ratio, cepstral-peak prominence, and H1-A2). This model was then used to classify data from a different language (Icelandic, Experiment 2) and to classify a different speech style (German infant-directed speech (IDS), Experiment 3). Cross-linguistic generalizability was high for Icelandic (78.6% accuracy), but lower for German IDS (71.7% accuracy). Accuracy of recordings of adult-directed speech from the same speakers as in Experiment 3 (77%, Experiment 4) suggests that it is the special speech style of IDS, rather than the recording setting that led to lower performance. Results are discussed in terms of efficiency of coding and generalizability across languages and speech styles.

Acquisition of Prosodic Focus Marking by Three- to Six-Year-Old Children Learning Mandarin Chinese

Qianyutong Zhang, Kexin Lyu, Zening Chen, Ping Tang; NJUST, China

Fri-M-V-1-4, Time: 11:00

Prosodic focus plays an important role during speech communication, delivering speakers' pragmatical intention to emphasize key information, especially in contrastive scenarios. Previous studies exploring children's acquisition of prosodic focus have generally focused on Germanic and Romance languages, while it was unclear when children learning Mandarin Chinese were able to correctly interpret the pragmatic meaning of prosodic focus and integrate it into speech comprehension. The current study explored Mandarin-learning 3-6-year-olds' online interpretation of prosodic focus to identify contrastive referents. Twenty 3-4-year-olds, 23 5-6-year-olds, and 22 adult controls were tested. The visual-world paradigm was adopted, where participants were instructed to search for target pictures while listening to contrastive objects in discourse sequences, e.g., Find the red cat. Now, find the PURPLE/purple cat, where the second adjective was produced with or without prosodic focus. Participants' fixation patterns were recorded via eye-trackers. The results showed that while adults and 5-6 years showed faster fixation toward target pictures in the presence of prosodic focus, this was not the case for 3-4 years. These results indicated that Mandarin-learning children at 5–6 years have acquired the pragmatic meaning of prosodic focus and utilize it to guide their identification of contrastive referents.

Adaptive Listening Difficulty Detection for L2 Learners Through Moderating ASR Resources

Maryam Sadat Mirzaei, Kourosh Meshgi; RIKEN, Japan Fri-M-V-1-5, Time: 11:00

Teaching listening skills to those learning a second language (L2) is one of the most challenging tasks mainly because predicting

L2 listening difficulties is not always straightforward. Complex processes are involved in decoding connected speech, constructing meaning, and comprehending the audio material. Many studies have attempted to identify the significant factors leading to listening difficulties, yet, a comprehensive model is to be constructed. We argue that an automatic speech recognition (ASR) system with limited training can be viewed as a rough model for an L2 listener with particular language proficiency. We proposed a method to select the training samples for the ASR system to match the mistakes of L2 listeners when listening to the authentic listening materials. This model can predict the learners' listening difficulties, thus allowing for generating tailored captions to assist them with L2 listening.

$F_{0}\ Patterns$ of L2 English Speech by Mandarin Chinese Learners

*Hongwei Ding*¹, *Binghuai Lin*², *Liyuan Wang*²; ¹*SJTU*, *China*; ²*Tencent, China* Fri-M-V-1-6, Time: 11:00

Prosodic speech characteristics are important in the evaluation of both intelligibility and naturalness of oral proficiency for learners of English as a Second Language (ESL). Different f_0 movement patterns between native and Mandarin Chinese learners have been an important research topic for second-language (L2) English speech learning. However, previous studies have seldom examined f_0 movement patterns between lower-level and higher-level Mandarin ESL learners. The current study compared f_0 change patterns extracted from the same 20 English sentences read by 20 lower- and 20 higherlevel Mandarin ESL learners, and 20 native English speakers from a speech database. Appropriate procedures were applied to ensure a more accurate estimation of f_0 values and to catch characteristic deviation in f_0 movement patterns of ESL learners. The results showed that lower-level Mandarin speakers displayed more frequent f_0 fluctuations and smaller standard deviation of intervals between f_0 peaks than both native speakers and higher-level learners. The special characteristic of many smaller "ripples" on pitch contours of lower-level L2 English speech resembles Mandarin Chinese f_0 movements, which suggests a negative transfer from the first language (L1) Mandarin. The findings can shed light on the assessment and learning of L2 English prosody by Mandarin ESL learners.

A Neural Network-Based Noise Compensation Method for Pronunciation Assessment

Binghuai Lin, Liyuan Wang; Tencent, China Fri-M-V-1-7, Time: 11:00

Automatic pronunciation assessment plays an important role in computer-assisted pronunciation training (CAPT). Goodness of pronunciation (GOP) based on automatic speech recognition (ASR) has been commonly used in pronunciation assessment. It has been found that GOP normally shows deteriorating performance under noisy conditions. Traditional noise compensation methods, which compensate distorted GOP under noisy situations based on the Gaussian mixture model (GMM) or other simple mapping functions, ignore contextual influence and phonemic attributes of the utterance. This usually leads to a lack of robustness with changed conditions. In this paper, we adopt a bidirectional long short-term (BLSTM) network combining phonemic attributes to conduct the compensation for distorted GOP under noisy conditions. We evaluate the model performance based on English words recorded by Chinese learners in clean and noisy situations. Experimental results show the proposed model outperforms the traditional baselines in Pearson correlation coefficient (PCC) and accuracy for pronunciation assessment under various noisy conditions.

Phonetic Distance and Surprisal in Multilingual Priming: Evidence from Slavic

Jacek Kudera, Philip Georgis, Bernd Möbius, Tania Avgustinova, Dietrich Klakow; Universität des Saarlandes, Germany Fri-M-V-1-8, Time: 11:00

This study reveals the relation between surprisal, phonetic distance, and latency based on a multilingual, short-term priming framework. Four Slavic languages (Bulgarian, Czech, Polish, and Russian) are investigated across two priming conditions: associative and phonetic priming, involving true cognates and near-homophones, respectively. This research is grounded in the methodology of information theory and proposes new methods for quantifying differences between meaningful lexical primes and targets for closely related languages. It also outlines the influence of phonetic distance between cognate and noncognate pairs of primes and targets on response times in a cross-lingual lexical decision task. The experimental results show that phonetic distance moderates response times only in Polish and Czech, whereas the surprisal-based correspondence effect is an accurate predictor of latency for all tested languages. The information-theoretic approach of quantifying feature-based alternations between Slavic cognates and near-homophones appears to be a valid method for latency moderation in the auditory modality. The outcomes of this study suggest that the surprisal-based (un)expectedness of spoken stimuli is an accurate predictor of human performance in multilingual lexical decision tasks.

A Preliminary Study on Discourse Prosody Encoding in L1 and L2 English Spontaneous Narratives

*Yuqing Zhang*¹, *Zhu Li*¹, *Binghuai Lin*², *Jinsong Zhang*¹; ¹*BLCU, China*; ²*Tencent, China* Fri-M-V-1-9, Time: 11:00

Relatively little attention has been devoted to the discourse-level prosodic encoding and speech planning in second language (L2) speech. This study reports a preliminary study on learners' discourse prosody encoding pattern and makes a comparison with that of native speakers. Using a corpus of spontaneously produced picture story narratives, we analyzed general characteristics of prosodic units (PUs) and explored relationships between pitch encoding (cross-boundary f0 heights and f0 slopes) of PUs and the semantic completeness of PUs in English spontaneous speech by native speakers, beginning learners, and advanced learners. The results indicated that beginning learners showed neither sensitivity to semantic units in discourse (DUs) in their f0 encoding nor distinct signs of pitch-related preplanning based on DUs, suggesting improper phrasing of the least proficient non-native speakers. Both native speakers and advanced learners were sensitive to the initiation and termination of DUs in their prosodic encoding; however, only native speakers showed clear signs of DU-based preplanning. We argue that the observed between-group differences in L1 and L2 speech might be attributed to differences in the scope of speech planning, i.e., compared with native speakers, who mostly produce complete semantic units, learners' speech is produced step by step with pauses between phrases

Transformer Based End-to-End Mispronunciation Detection and Diagnosis

*Minglin Wu*¹, *Kun Li*², *Wai-Kim Leung*¹, *Helen Meng*¹; ¹*CUHK*, *China*; ²*SpeechX*, *China* Fri-M-V-1-10, Time: 11:00

This paper introduces two Transformer-based architectures for Mispronunciation Detection and Diagnosis (MDD). The first Trans-

former architecture (T-1) is a standard setup with an encoder, a decoder, a projection part and the Cross Entropy (CE) loss. T-1 takes in Mel-Frequency Cepstral Coefficients (MFCC) as input. The second architecture (T-2) is based on wav2vec 2.0, a pretraining framework. T-2 is composed of a CNN feature encoder, several Transformer blocks capturing contextual speech representations, a projection part and the Connectionist Temporal Classification (CTC) loss. Unlike T-1, T-2 takes in raw audio data as input. Both models are trained in an end-to-end manner. Experiments are conducted on the CU-CHLOE corpus, where T-1 achieves a Phone Error Rate (PER) of 8.69% and F-measure of 77.23%; and T-2 achieves a PER of 5.97% and F-measure of 80.98%. Both models significantly outperform the previously proposed AGPM and CNN-RNN-CTC models, with PERs at 11.1% and 12.1% respectively, and F-measures at 72.61% and 74.65% respectively.

L1 Identification from L2 Speech Using Neural Spectrogram Analysis

Calbert Graham; University of Cambridge, UK Fri-M-V-1-11, Time: 11:00

It is well-known that the characteristics of L2 speech are highly influenced by the speakers' L1. The main objective of this study was to uncover discriminative speech features to identify the L1 background of a speaker from their L2 English speech. Traditional phonetic approaches tend to compare speakers based on a preselected set of acoustic features, which may not be sufficient to capture all the unique traces of the L1 in the L2 speech for forensic speaker profiling purposes. Convolutional Neural Network (CNN) has the potential to remedy this issue through the automatic processing of the visual spectrogram.

This paper reports a series of CNN classification experiments modelled on spectrogram images. The classification problem consisted of determining whether English speech samples are spoken by a native speaker of English, Japanese, Dutch, French, or Polish. Both phonetically transcribed and untranscribed speech data were used.

Overall, results showed that the CNN achieved a high level of accuracy in identifying the speakers' L1s based on spectrogram pictures without explicit phonetic segmentation. However, the results also showed that training the classifiers on certain combinations of phonetically modelled spectrogram images, which would make features more transparent, can produce results with comparable accuracy rates.

Fri-M-V-2: Phonetics II

11:00–13:00, Friday 3 September 2021 Chairs: Daniel Brenner and Priyankoo Sarmah

Leveraging Real-Time MRI for Illuminating Linguistic Velum Action

Miran Oh, Dani Byrd, Shrikanth S. Narayanan; University of Southern California, USA Fri-M-V-2-1, Time: 11:00

Velum actions are critical to differentiating oral and nasal sounds in spoken language; specifically in the latter, the velum is lowered to open the nasal port and allow nasal airflow. However, details on how the velum is lowered for nasal production in speech are scarce. State-of-the-art real-time Magnetic Resonance Imaging (rtMRI) can directly image the entirety of the moving vocal tract, providing spatiotemporal kinematic data of articulatory actions. Most instrumental studies of speech production explore oral constriction actions such as lip or tongue movements. RtMRI makes possible a quantitative assessment of non-oral and non-constriction

actions, such as velum (and larynx) dynamics. This paper illustrates articulatory aspects of consonant nasality, which have previously been inferred from acoustic or aerodynamic data. Velum actions are quantified in spatial and temporal domains: i) vertical and horizontal velum positions during nasal consonant production are quantified to measure, respectively, the degree of velum lowering and velic opening, and ii) duration intervals for velum lowering, plateau, and raising are obtained to understand which portion of the velum action is lengthened to generate phonologically long nasality. Findings demonstrate that velum action tracking using rtMRI can illuminate linguistic modulations of nasality strength and length.

Segmental Alignment of English Syllables with Singleton and Cluster Onsets

Zirui Liu, Yi Xu; University College London, UK Fri-M-V-2-2, Time: 11:00

Recent research has shown fresh evidence that consonant and vowel are synchronised at the syllable onset, as predicted by a number of theoretical models. The finding was made by using a minimal contrast paradigm to determine segment onset in Mandarin CV syllables, which differed from the conventional method of detecting gesture onset with a velocity threshold [1]. It has remained unclear, however, if CV co-onset also occurs between the nucleus vowel and a consonant cluster, as predicted by the articulatory syllable model [2]. This study applied the minimal contrast paradigm to British English in both CV and clusterV (CLV) syllables, and analysed the spectral patterns with signal chopping in conjunction with recurrent neural networks (RNN) with long short-term memory (LSTM) [3]. Results show that vowel onset is synchronised with the onset of the first consonant in a cluster, thus supporting the articulatory syllable model.

Exploration of Welsh English Pre-Aspiration: How Wide-Spread is it?

Míša Hejná; Aarhus University, Denmark Fri-M-V-2-3, Time: 11:00

This study investigates how widespread pre-aspiration and local breathiness are in English spoken in Wales, by speakers identifying as Welsh. While the main purpose is to establish whether the phenomenon is generally present in Welsh English, the data also enables us to explore whether pre-aspiration might be conditioned by sex/gender, age, and the ability to speak Welsh. An acoustic corpus of 45 speakers producing word-final plosives and fricatives is analysed.

Pre-aspiration and local breathiness are produced by all speakers, representing 32 towns and 16 areas included in the analyses. Preaspiration and breathiness are more frequent and longer in L1 and L2 Welsh speakers than those who do not speak Welsh at all. In general, no statistically significant sex and age effects emerge.

In addition, a gradient allophony between pre-aspiration and glottalisation is reported for all speakers in the plosive context: the more frequently they glottalise, the less frequent the pre-aspiration. In fricatives, most speakers do not glottalise. Regarding those who do, 1. some display no relationship between pre-aspiration and glottalisation, and 2. a minority display either an indication of gradient allophony between the two, or 3. a positive correlation.

Revisiting Recall Effects of Filler Particles in German and English

Beeke Muhlack, Mikey Elmers, Heiner Drenhaus, Jürgen Trouvain, Marjolein van Os, Raphael Werner, Margarita Ryzhova, Bernd Möbius; Universität des Saarlandes, Germany

Fri-M-V-2-4, Time: 11:00

This paper reports on two experiments that partially replicate an experiment by Fraundorf and Watson (2011, J Mem. Lang.) on the recall effect of filler particles. Their subjects listened to three passages of a story, either with or without filler particles, which they had to retell afterwards. They analysed the subjects' retelling in terms of whether important plot points were remembered or not. For their English data, they found that filler particles facilitate the recall of the plot points significantly compared to stories that did not include filler particles. As this seems to be a convincing experimental design, we aimed at evaluating this method as a web-based experiment which may, if found to be suitable, easily be applied to other languages. Furthermore, we investigated whether their results are found in German as well (Experiment 1), and evaluated whether filler duration has an effect on recall performance (Experiment 2). Our results could not replicate the findings of the original study: in fact, the opposite effect was found for German. In Experiment 1, participants performed better on recall in the fluent condition, while no significant results were found for English in Experiment 2.

How Reliable Are Phonetic Data Collected Remotely? Comparison of Recording Devices and Environments on Acoustic Measurements

Chunyu Ge, Yixuan Xiong, Peggy Mok; CUHK, China Fri-M-V-2-5, Time: 11:00

The COVID-19 pandemic posed an unprecedented challenge to phonetic research. On-site collection of speech data is difficult, if not impossible. The advancement of technology in mobile devices and online conference platforms offers the opportunity to collect data remotely. This paper aims to answer the question of how reliable speech data collected remotely are based on controlled speech. Seven devices, including smartphones and laptops, were used to record speech simultaneously, locally or on the cloud using ZOOM, both in a sound-attenuated lab and a conference room. Common Local acoustic measurements were made on these recordings. recordings proved to be reliable in duration, but not for recordings made on the cloud. Different devices have comparable performances in F0 and F1. The values acquired by different devices differ a lot for F2 and higher formants, spectral moments, and voice quality measures. These differences can lead to erroneous interpretation of segmental and voice quality contrasts. The recordings made remotely by smartphones and locally using ZOOM can be useful in studying prosody, but should be used with care for segments.

A Cross-Dialectal Comparison of Apical Vowels in Beijing Mandarin, Northeastern Mandarin and Southwestern Mandarin: An EMA and Ultrasound Study

Jing Huang, Feng-fan Hsieh, Yueh-chin Chang; National Tsing Hua University, Taiwan Fri-M-V-2-6, Time: 11:00

This paper is a comparative study of the articulation of the "apical vowels" in three Mandarin dialects: Beijing Mandarin (BJM), Northeastern Mandarin (NEM), and Southwestern Mandarin (SWM), using co-registered EMA and ultrasound. Data from 5 BJM speakers, 5 NEM speakers and 4 SWM speakers in their twenties were analyzed and

discussed. Our recording materials include the dental and retroflex apical vowels, and their *er*-suffixed forms. Results suggest that distinct lingual configurations are found among the three dialects of Mandarin, even though these apical vowels are not perceptually distinguishable. Specifically, the dental apical vowel [1] has a grooved tongue shape in BJM, a retracted tongue dorsum in NEM, and a relatively flat tongue shape in SWM. The retroflex apical vowel [1] has a domed tongue shape as well as a bunched tongue body in NEM, while a slightly domed tongue posture is found in SWM. Moreover, the retroflex apical vowel [1] is, articulatorily speaking, very similar to the *er*-suffix in BJM (cf. [10]). In sum, we observed yet another instance of the articulatory-acoustic mismatch.

Dissecting the Aero-Acoustic Parameters of Open Articulatory Transitions

*Mark Gibson*¹, *Oihane Muxika*¹, *Marianne Pouplier*²; ¹*Universidad de Navarra, Spain;* ²*LMU München, Germany*

Fri-M-V-2-7, Time: 11:00

We capitalize on previously recorded kinematic and acoustic data for three languages (Georgian (GE), Spanish (SP) and Moroccan Arabic (MA)) that exhibit open articulatory transitions between the consonants in clusters in order to dissect the aero-acoustic parameters of the transitions in each language. These particular languages are of interest because they show similar patterns of interarticulatory timing in clusters, offering the unique opportunity to examine the acoustics of open transitions cross-linguistically. Our analysis centers on word initial clusters (/kl/ and /gl/), from which we extract relativized temporal values relevant to clusters and spectral parameters related to open articulatory transitions. We report baseline results using linear mixed effects models, then train a Random Forest model in a supervised learning environment on the significant variables. After training, test tokens are introduced in order to test whether the model can categorize the language based on the spectral and temporal parameters, and rank variables in terms of their feature importance. The results show that the model can categorize the data to the correct language with a 95.59% accuracy rate, where normalized zero-crossing (nzcr), modifications of the amplitude envelope (ΔE), and intensity ratio ranked highest in feature importance.

Quantifying Vocal Tract Shape Variation and its Acoustic Impact: A Geometric Morphometric Approach

Amelia J. Gully; University of York, UK Fri-M-V-2-8, Time: 11:00

The shape of the vocal tract varies considerably between individuals. The relationship between detailed variation in vocal tract shape and the acoustics of speech is not yet well understood, despite its potential for increasing understanding in the fields of voice biometrics, forensic speech science, and personalised speech synthesis. One reason that this topic has not yet been extensively explored is that 3D vocal tract shape is difficult to quantify robustly. Geometric morphometrics is a technique developed in evolutionary biology for statistically valid quantification and comparison of anatomical shapes. This study makes use of 3D magnetic resonance imaging data of the vocal tracts of eight individuals, and accompanying audio recordings, combined with geometric morphometric techniques to determine whether the method offers useful information for speech science. The results suggest a linear relationship between the shapes of the vocal tract and output spectra, and there is evidence of possible sexual dimorphism and allometry (a systematic variation of shape with size) in the vocal tract, although due to the limited sample size the results did not reach statistical significance. The

NOTES

results suggest that geometric morphometrics can provide useful information about the vocal tract, and justify further study using this technique.

Speech Perception and Loanword Adaptations: The Case of Copy-Vowel Epenthesis

Adriana Guevara-Rukoz¹, Shi Yu², Sharon Peperkamp¹; ¹LSCP (UMR 8554), France; ²LPP (UMR 7018), France Fri-M-V-2-9, Time: 11:00

Japanese allows for almost no consonants in syllable codas. In loanwords, illegal codas are transformed into onsets by means of vowel epenthesis. The default epenthetic vowel in loanwords is [ω], and previous work has shown that this [ω]-epenthesis reflects Japanese listeners' perception of illegal coda consonants. Here, we focus on one of the non-default cases: following coda [c] and [x] the epenthetic vowel is a copy of the preceding vowel. Using an identification and a discrimination task, we provide evidence for the perceptual origin of this copy vowel phenomenon: After [c] and [x], Japanese listeners perceive more often an epenthetic copy vowel than the default vowel [ω], whereas after [k] it is the reverse.

Speakers Coarticulate Less When Facing Real and Imagined Communicative Difficulties: An Analysis of Read and Spontaneous Speech from the LUCID Corpus

Zhe-chen Guo, Rajka Smiljanic; University of Texas at Austin, USA

Fri-M-V-2-10, Time: 11:00

This study investigated coarticulation of read and spontaneous speech in different communicative contexts from the LUCID corpus. Spontaneous speech samples were from Southern British English speakers who completed an interactive spot-the-differences task with no communicative barrier (NB), with their voice vocoded (VOC), and with a partner who heard their speech in babble (BABBLE) or was a non-native English speaker (L2). The same speakers also read sentences in a casual (READ-CO) and clear (READ-CL) speaking style. Tokens of a pre-defined set of keywords were extracted from the speech samples and consonant-vowel sequences in these tokens were analyzed using a whole-spectrum measure of coarticulation. Results showed that coarticulatory resistance in the six communicative contexts from highest to lowest was: BABBLE > VOC, L2, READ-CL > NB, READ-CO. Thus, in response to communicative barriers, be they real or imaginary, speakers coarticulated less, in line with the models of targeted speaker adaptations (the H&H theory [1] and Adaptive Speaker Framework [2]).

Developmental Changes of Vowel Acoustics in Adolescents

Einar Meister, Lya Meister; Tallinn University of Technology, Estonia Fri-M-V-2-11, Time: 11:00

The paper explores the developmental changes of vowel acoustics in Estonian adolescents as a function of age and gender. Formant frequencies F1-F4 and the duration of vowels were measured from read speech samples of 305 native Estonian subjects (173 girls and 132 boys) aged from 10 to 18 years. GAM framework was applied for the statistical analysis. The results show that both the formant frequencies and the vowel space area decrease gradually from 10 to 15 years in both gender groups and the quality of vowels stabilizes at the age of 15-18 years, whereas gender-specific differences emerge around the age of 12-13. Age-related change in the duration of vowels shows similar patterns with formants, however, with no gender difference. The findings are in line with the results reported for adolescent speech in other languages. The analysis results based on speech samples of the subjects with normal linguistic development can be considered reference data for distinguishing between normal and abnormal speech development.

Context and Co-Text Influence on the Accuracy Production of Italian L2 Non-Native Sounds

Sonia d'Apolito, Barbara Gili Fivela; Università del Salento, Italy Fri-M-V-2-12, Time: 11:00

Accuracy in production of non-native sounds is analyzed by considering the influence of L1, context and co-text on Italian L2 speech. While the L1 influence is often described in the literature, careful investigations on how production accuracy may change in different contexts and co-texts are needed. This paper describes two experiments on how French learners of Italian as L2 (advanced/beginners) realize geminates depending on different contexts (the global contexts, e.g., the tasks) and co-texts (the amount of information available syntagmatically).

Acoustic data acquired by recording 4 advanced and 4 beginner Italian-L2 learners (and 3 Italian natives as control) were analyzed as for the duration of the target consonant and preceding vowel, as well as speech articulation rate, taken as indexes of geminate production accuracy.

Results confirm the strongest influence of L1 in beginners' production, and depict a complex interplay of context and co- text. Adding information in co-text may induce different effects on speech production, depending on the local context, that is on the speakers' communication needs during speech production. Specifically, a "rich" co-text may favor a decrease in production accuracy or, on the contrary, an increase, depending on the need the speaker have to highlight/contrast information.

A New Vowel Normalization for Sociophonetics

Wilbert Heeringa, Hans Van de Velde; Fryske Akademy, The Netherlands

Fri-M-V-2-13, Time: 11:00

Several studies have shown that in sociophonetic research Lobanov's speaker normalization method outperforms other methods for normalizing vowel formants of speakers. An advantage of Lobanov's method compared to the method that was introduced by Watt & Fabricius in 2002 is that it is independent of the shape of the vowel space area, and also normalizes to the dispersion of the vowels. However, it does depend on the distribution of the vowels within the vowel space. When using Lobanov normalization the formant values are converted to *z*-scores. We present a method where the μ in the *z*-score formula is replaced by the center of the convex hull that encloses the vowels, and the σ is obtained on the basis of the points that constitute the convex hull. When normalizing measurements of two real data sets, and of a series of randomly generated data sets, we found that our method improved in matching vowel spaces in size and overlap.

The Pacific Expansion: Optimizing Phonetic Transcription of Archival Corpora

*Rosey Billington*¹, *Hywel Stoakes*², *Nick Thieberger*²; ¹*ANU, Australia*; ²*University of Melbourne, Australia* Fri-M-V-2-14, Time: 11:00

For most of the world's languages, detailed phonetic analyses across different aspects of the sound system do not exist, due in part to limitations in available speech data and tools for efficiently processing such data for low-resource languages. Archival language documentation collections offer opportunities to extend the scope and scale of phonetic research on low-resource languages, and developments in methods for automatic recognition and alignment of speech facilitate the preparation of phonetic corpora based on these collections. We present a case study applying speech modelling and forced alignment methods to narrative data for Nafsan, an Oceanic language of central Vanuatu. We examine the accuracy of the forcedaligned phonetic labelling based on limited speech data used in the modelling process, and compare acoustic and durational measures of 17,851 vowel tokens for 11 speakers with previous experimental phonetic data for Nafsan. Results point to the suitability of archival data for large-scale studies of phonetic variation in low-resource languages, and also suggest that this approach can feasibly be used as a starting point in expanding to phonetic comparisons across closely-related Oceanic languages.

Fri-M-V-3: Search/Decoding Techniques and Confidence Measures for ASR

11:00-13:00, Friday 3 September 2021 Chairs: Takaaki Hori and Tatsuya Kawahara

FSR: Accelerating the Inference Process of Transducer-Based Models by Applying Fast-Skip Regularization

Zhengkun Tian, Jiangyan Yi, Ye Bai, Jianhua Tao, Shuai Zhang, Zhengqi Wen; CAS, China Fri-M-V-3-1, Time: 11:00

Transducer-based models, such as RNN-Transducer and transformertransducer, have achieved great success in speech recognition. A typical transducer model decodes the output sequence conditioned on the current acoustic state and previously predicted tokens step by step. Statistically, The number of blank tokens in the prediction results accounts for nearly 90% of all tokens. It takes a lot of computation and time to predict the blank tokens, but only the non-blank tokens will appear in the final output sequence. Therefore, we propose a method named fast-skip regularization, which tries to align the blank position predicted by a transducer with that predicted by a connectionist temporal classification (CTC) model. During the inference, the transducer model can predict the blank tokens in advance by a simple CTC project layer without many complicated forward calculations of the transducer decoder and then skip them, which will reduce the computation and improve the inference speed greatly. All experiments are conducted on a public Chinese mandarin dataset AISHELL-1. The results show that the fast-skip regularization can indeed help the transducer model learn the blank position alignments. Besides, the inference with fast-skip can be speeded up nearly 4 times with only a little performance degradation.

LT-LM: A Novel Non-Autoregressive Language Model for Single-Shot Lattice Rescoring

Anton Mitrofanov¹, Mariya Korenevskaya², Ivan Podluzhny¹, Yuri Khokhlov², Aleksandr Laptev¹, Andrei Andrusenko¹, Aleksei Ilin², Maxim Korenevsky², Ivan Medennikov¹, Aleksei Romanenko¹; ¹ITMO University, Russia; ²STC-innovations, Russia Fri-M-V-3-2, Time: 11:00

Neural network-based language models are commonly used in rescoring approaches to improve the quality of modern automatic speech recognition (ASR) systems. Most of the existing methods are computationally expensive since they use autoregressive language

models. We propose a novel rescoring approach, which processes the entire lattice in a single call to the model. The key feature of our rescoring policy is a novel non-autoregressive Lattice Transformer Language Model (LT-LM). This model takes the whole lattice as an input and predicts a new language score for each arc. Additionally, we propose the artificial lattices generation approach to incorporate a large amount of text data in the LT-LM training process. Our single-shot rescoring performs orders of magnitude faster than other rescoring methods in our experiments. It is more than 300 times faster than pruned RNNLM lattice rescoring and N-best rescoring while slightly inferior in terms of WER.

A Hybrid Seq-2-Seq ASR Design for On-Device and Server Applications

Cyril Allauzen, Ehsan Variani, Michael Riley, David Rybach, Hao Zhang; Google, USA Fri-M-V-3-3, Time: 11:00

This paper proposes and evaluates alternative speech recognition design strategies using the hybrid autoregressive transducer (HAT) model. The different strategies are designed with special attention to the choice of modeling units and to the integration of different types of external language models during first-pass beam-search or second-pass re-scoring. These approaches are compared on a large-scale voice search task and the recognition quality over the head and tail of speech data is analyzed. Our experiments show decent improvements in WER over common speech phrases and significant gains on uncommon ones compared to the state-of-the-art approaches.

VAD-Free Streaming Hybrid CTC/Attention ASR for Unsegmented Recording

Hirofumi Inaguma, Tatsuya Kawahara; Kyoto University, Japan

Fri-M-V-3-4, Time: 11:00

In this work, we propose novel decoding algorithms to enable streaming automatic speech recognition (ASR) on unsegmented long-form recordings without voice activity detection (VAD), based on monotonic chunkwise attention (MoChA) with an auxiliary connectionist temporal classification (CTC) objective. We propose a blocksynchronous beam search decoding to take advantage of efficient batched output-synchronous and low-latency input-synchronous searches. We also propose a VAD-free inference algorithm that leverages CTC probabilities to determine a suitable timing to reset the model states to tackle the vulnerability to long-form data. Experimental evaluations demonstrate that the block-synchronous decoding achieves comparable accuracy to the label-synchronous Moreover, the VAD-free inference can recognize long-form one. speech robustly for up to a few hours.

WeNet: Production Oriented Streaming and Non-Streaming End-to-End Speech Recognition Toolkit

Zhuoyuan Yao¹, Di Wu², Xiong Wang¹, Binbin Zhang², Fan Yu¹, Chao Yang², Zhendong Peng², Xiaoyu Chen², Lei Xie¹, Xin Lei²; ¹Northwestern Polytechnical University, China; ²Mobvoi, China Fri-M-V-3-5, Time: 11:00

In this paper, we propose an open source speech recognition toolkit called WeNet, in which a new two-pass approach named U2 is implemented to unify streaming and non-streaming end-to-end (E2E) speech recognition in a single model. The main motivation of WeNet is to close the gap between the research and deployment of E2E speech recognition models. WeNet provides an efficient way to

NOTES

ship automatic speech recognition (ASR) applications in real-world scenarios, which is the main difference and advantage to other open source E2E speech recognition toolkits. We develop a hybrid connectionist temporal classification (CTC)/attention architecture with transformer or conformer as encoder and an attention decoder to rescore the CTC hypotheses. To achieve streaming and non-streaming in a unified model, we use a dynamic chunk-based attention strategy which allows the self-attention to focus on the right context with random length. Our experiments on the AISHELL-1 dataset show that our model achieves 5.03% relative character error rate (CER) reduction in non-streaming ASR compared to a standard non-streaming transformer. After model quantification, our model achieves reasonable RTF and latency at runtime. The toolkit is publicly available.

Cross-Modal Transformer-Based Neural Correction Models for Automatic Speech Recognition

Tomohiro Tanaka, Ryo Masumura, Mana Ihori, Akihiko Takashima, Takafumi Moriya, Takanori Ashihara, Shota Orihashi, Naoki Makishima; NTT, Japan Fri-M-V-3-6, Time: 11:00

We propose a cross-modal transformer-based neural correction models that refines the output of an automatic speech recognition (ASR) system so as to exclude ASR errors. Generally, neural correction models are composed of encoder-decoder networks, which can directly model sequence-to-sequence mapping problems. The most successful method is to use both input speech and its ASR output text as the input contexts for the encoder-decoder networks. However, the conventional method cannot take into account the relationships between these two different modal inputs because the input contexts are separately encoded for each modal. To effectively leverage the correlated information between the two different modal inputs, our proposed models encode two different contexts jointly on the basis of cross-modal self-attention using a transformer. We expect that cross-modal self-attention can effectively capture the relationships between two different modals for refining ASR hypotheses. We also introduce a shallow fusion technique to efficiently integrate the first-pass ASR model and our proposed neural correction model. Experiments on Japanese natural language ASR tasks demonstrated that our proposed models achieve better ASR performance than conventional neural correction models.

Deep Neural Network Calibration for E2E Speech Recognition System

Mun-Hak Lee, Joon-Hyuk Chang; Hanyang University, Korea

Fri-M-V-3-7, Time: 11:00

Cross-entropy loss, which is commonly used in deep-neural-networkbased (DNN) classification model training, induces models to assign a high probability value to one class. Networks trained in this fashion tend to be overconfident, which causes a problem in the decoding process of the speech recognition system, as it uses the combined probability distribution of multiple independently trained networks. Overconfidence in neural networks can be quantified as a calibration error, which is the difference between the output probability of a model and the likelihood of obtaining an actual correct answer. We show that the deep-learning-based components of an end-to-end (E2E) speech recognition system with high classification accuracy contain calibration errors and quantify them using various calibration measures. In addition, it was experimentally shown that the calibration function, which was being trained to minimize calibration errors effectively mitigates those of the speech recognition system, and as a result, can improve the performance of beam-search during decoding.

Residual Energy-Based Models for End-to-End Speech Recognition

*Qiujia Li*¹, *Yu Zhang*², *Bo Li*², *Liangliang Cao*², *Philip C. Woodland*¹; ¹*University of Cambridge, UK*; ²*Google, USA* Fri-M-V-3-8, Time: 11:00

End-to-end models with auto-regressive decoders have shown impressive results for automatic speech recognition (ASR). These models formulate the sequence-level probability as a product of the conditional probabilities of all individual tokens given their histories. However, the performance of locally normalised models can be sub-optimal because of factors such as exposure bias. Consequently, the model distribution differs from the underlying data distribution. In this paper, the residual energy-based model (R-EBM) is proposed to complement the auto-regressive ASR model to close the gap between the two distributions. Meanwhile, R-EBMs can also be regarded as utterance-level confidence estimators, which may benefit many downstream tasks. Experiments on a 100hr LibriSpeech dataset show that R-EBMs can reduce the word error rates (WERs) by 8.2%/6.7% while improving areas under precision-recall curves of confidence scores by 12.6%/28.4% on test-clean/test-other sets. Furthermore, on a state-of-the-art model using self-supervised learning (wav2vec 2.0), R-EBMs still significantly improves both the WER and confidence estimation performance.

Multi-Task Learning for End-to-End ASR Word and Utterance Confidence with Deletion Prediction

David Qiu¹, Yanzhang He¹, Qiujia Li², Yu Zhang¹, Liangliang Cao¹, Ian McGraw¹; ¹Google, USA; ²University of Cambridge, UK Fri-M-V-3-9, Time: 11:00

Confidence scores are very useful for downstream applications of automatic speech recognition (ASR) systems. Recent works have proposed using neural networks to learn word or utterance confidence scores for end-to-end ASR. In those studies, word confidence by itself does not model deletions, and utterance confidence does not take advantage of word-level training signals. This paper proposes to jointly learn word confidence, word deletion, and utterance confidence. Empirical results show that multi-task learning with all three objectives improves confidence metrics (NCE, AUC, RMSE) without the need for increasing the model size of the confidence estimation module. Using the utterance-level confidence for rescoring also decreases the word error rates on Google's Voice Search and Long-tail Maps datasets by 3–5% relative, without needing a dedicated neural rescorer.

Insights on Neural Representations for End-to-End Speech Recognition

Anna Ollerenshaw, Md. Asif Jalal, Thomas Hain; University of Sheffield, UK

Fri-M-V-3-10, Time: 11:00

End-to-end automatic speech recognition (ASR) models aim to learn a generalised speech representation. However, there are limited tools available to understand the internal functions and the effect of hierarchical dependencies within the model architecture. It is crucial to understand the correlations between the layer-wise representations, to derive insights on the relationship between neural representations and performance. Previous investigations of network similarities using correlation analysis techniques have not been explored for End-to-End ASR models. This paper analyses and explores the internal dynamics between layers during training with CNN, LSTM and Transformer based approaches using Canonical correlation analysis (CCA) and centered kernel alignment (CKA) for the experiments. It was found that neural representations within CNN layers exhibit

NOTES

hierarchical correlation dependencies as layer depth increases but this is mostly limited to cases where neural representation correlates more closely. This behaviour is not observed in LSTM architecture, however there is a bottom-up pattern observed across the training process, while Transformer encoder layers exhibit irregular coefficiency correlation as neural depth increases. Altogether, these results provide new insights into the role that neural architectures have upon speech recognition performance. More specifically, these techniques can be used as indicators to build better performing speech recognition models.

Sequence-Level Confidence Classifier for ASR Utterance Accuracy and Application to Acoustic Models

*Amber Afshan*¹, *Kshitiz Kumar*², *Jian Wu*²; ¹*University of California at Los Angeles, USA*; ²*Microsoft, USA* Fri-M-V-3-11, Time: 11:00

Scores from traditional confidence classifiers (CCs) in automatic speech recognition (ASR) systems lack universal interpretation and vary with updates to the underlying confidence or acoustic models (AMs). In this work, we build interpretable confidence scores with an objective to closely align with ASR accuracy. We propose a new sequence-level CC with a richer context providing CC scores highly correlated with ASR accuracy and scores stable across CC updates. Hence, expanding CC applications. Recently, AM customization has gained traction with the widespread use of unified models. Conventional adaptation strategies that customize AM expect well-matched data for the target domain with gold-standard transcriptions. We propose a cost-effective method of using CC scores to select an optimal adaptation data set, where we maximize ASR gains from minimal data. We study data in various confidence ranges and optimally choose data for AM adaptation with KL-Divergence regularization. On the Microsoft voice search task, data selection for supervised adaptation using the sequence-level confidence scores achieves word error rate reduction (WERR) of 8.5% for row-convolution LSTM (RC-LSTM) and 5.2% for latency-controlled bidirectional LSTM (LC-BLSTM). In the semi-supervised case, with ASR hypotheses as labels, our method provides WERR of 5.9% and 2.8% for RC-LSTM and LC-BLSTM, respectively.

Fri-M-V-4: Speech Synthesis: Linguistic Processing, Paradigms and Other Topics

11:00–13:00, Friday 3 September 2021 Chairs: Hema Murthy and Géza Németh

Unsupervised Learning of Disentangled Speech Content and Style Representation

Andros Tjandra¹, Ruoming Pang², Yu Zhang², Shigeki Karita³; ¹NAIST, Japan; ²Google, USA; ³Google, Japan Fri-M-V-4-1, Time: 11:00

Speech is influenced by a number of underlying factors, which can be broadly categorized into linguistic contents and speaking styles. However, collecting the labeled data that annotates both content and style is an expensive and time-consuming task. Here, we present an approach for unsupervised learning of speech representation disentangling contents and styles. Our model consists of: (1) a local encoder that captures per-frame information; (2) a global encoder that captures per-utterance information; and (3) a conditional decoder that reconstructs speech given local and global latent variables. Our experiments show that (1) the local latent variables encode speech contents, as reconstructed speech can be recognized by ASR with low word error rates (WER), even with a different global encoding; (2) the global latent variables encode speaker style, as reconstructed speech shares speaker identity with the source utterance of the global encoding. Additionally, we demonstrate a useful application from our pre-trained model, where we can train a speaker recognition model from the global latent variables and achieve high accuracy by fine-tuning with as few data as one label per speaker.

Label Embedding for Chinese Grapheme-to-Phoneme Conversion

*Eunbi Choi*¹, *Hwa-Yeon Kim*², *Jong-Hwan Kim*², *Jae-Min Kim*²; ¹*KAIST, Korea*; ²*Naver, Korea* Fri-M-V-4-2, Time: 11:00

Chinese grapheme-to-phoneme (G2P) conversion plays a significant role in text-to-speech systems by generating pronunciations corresponding to Chinese input characters. The main challenge in Chinese G2P conversion is polyphone disambiguation, which requires selecting the appropriate pronunciation among several candidates. In polyphone disambiguation, calculating probabilities for the entire pronunciations is unnecessary since each Chinese character has only a few (mostly two or three) candidate pronunciations. In this study, we introduce a label embedding approach that matches the character embedding with the closest label embedding among the possible candidates. Specifically, negative sampling and triplet loss were applied to maximize the difference between the correct embedding and the other candidate embeddings. Experimental results show that the label embedding approach improved the polyphone disambiguation accuracy by 4.50% and 1.74% on two datasets compared to the one-hot label classification approach. Moreover, the bidirectional long short-term memory model with the label embedding approach outperformed the previous most advanced model, BERT, demonstrating outstanding performance in polyphone disambiguation. Lastly, we discuss the effect of contextual information in character embeddings on the G2P conversion task.

PDF: Polyphone Disambiguation in Chinese by Using FLAT

Haiteng Zhang; Databaker Technology, China Fri-M-V-4-3, Time: 11:00

Polyphone disambiguation is an essential procedure in the front-end module of the Chinese text-to-speech (TTS) system. It serves to predict the pronunciation of the input polyphonic character. In the Chinese TTS system, a well-designed pronunciation dictionary plays a crucial role in supplying pinyin to words. However, the conventional system is unable to fully utilize the pronunciation dictionary while modelling because of the unavoidable Chinese segment errors and model structure. In this paper, we proposed a system named PDF: Polyphone Disambiguation by using FLAT. The proposed model encodes both the input character sequence and dictionary matched words of the sentence, enabling the model to both avoid segment errors and leverage the well-designed pronunciation dictionary in the model. Additionally, we also use the pre-trained language model (PLM) as an encoder to extract the contextual information of input sequence. The experimental results verified the effectiveness of the proposed PDF model. Our system obtains an improvement in accuracy by 0.98% compared to Bert on an open-source dataset. The experiential results demonstrate that leveraging pronunciation dictionary while modelling helps improve the performance of polyphone disambiguation system.

Improving Polyphone Disambiguation for Mandarin Chinese by Combining Mix-Pooling Strategy and Window-Based Attention

Junjie Li¹, Zhiyu Zhang², Minchuan Chen¹, Jun Ma¹, Shaojun Wang¹, Jing Xiao¹; ¹Ping An Technology, China; ²National Tsing Hua University, Taiwan Fri-M-V-4-4, Time: 11:00

In this paper, we propose a novel system based on word-level features and window-based attention for polyphone disambiguation, which is a fundamental task for Grapheme-to-phoneme (G2P) conversion of Mandarin Chinese. The framework aims to combine a pre-trained language model with explicit word-level information in order to get meaningful context extraction. Particularly, we employ a pre-trained bidirectional encoder from Transformers (BERT) model to extract character-level features, and an external Chinese word segmentation (CWS) tool is used to obtain the word units. We adopt a mixed pooling mechanism to convert character-level features into word-level features based on the segmentation results. A window-based attention module is utilized to incorporate contextual word-level features for the polyphonic characters. Experimental results show that our method achieves an accuracy of 99.06% on an open benchmark dataset for Mandarin Chinese polyphone disambiguation, which outperforms the baseline systems.

Polyphone Disambiguation in Mandarin Chinese with Semi-Supervised Learning

Yi Shi, Congyi Wang, Yu Chen, Bin Wang; Xmov, China Fri-M-V-4-5, Time: 11:00

The majority of Chinese characters are monophonic, while a special group of characters, called polyphonic characters, have multiple pronunciations. As a prerequisite of performing speech-related generative tasks, the correct pronunciation must be identified among several candidates. This process is called Polyphone Disambiguation. Although the problem has been well explored with both knowledgebased and learning-based approaches, it remains challenging due to the lack of publicly available labeled datasets and the irregular nature of polyphone in Mandarin Chinese. In this paper, we propose a novel semi-supervised learning (SSL) framework for Mandarin Chinese polyphone disambiguation that can potentially leverage unlimited unlabeled text data. We explore the effect of various proxy labeling strategies including entropy-thresholding and lexicon-based labeling. Qualitative and quantitative experiments demonstrate that our method achieves state-of-the-art performance. In addition, we publish a novel dataset specifically for the polyphone disambiguation task to promote further researches.

A Neural-Network-Based Approach to Identifying Speakers in Novels

Yue Chen, Zhen-Hua Ling, Qing-Feng Liu; USTC, China Fri-M-V-4-6, Time: 11:00

Identifying speakers in novels aims at determining who says a quote in a given context by text analysis. This task is important for speech synthesis systems to assign appropriate voices to the quotes when producing audiobooks. However, existing approaches stick with manual features and traditional machine learning classifiers, which constrain the accuracy of speaker identification. In this paper, we propose a method to tackle this challenging problem with the help of deep learning. We formulate speaker identification as a scoring task and build a candidate scoring network (CSN) based on BERT. Candidate-specific segments are put forward to eliminate redundant context information. Moreover, a revision algorithm is designed utilizing the speaker alternation pattern in two-party dialogues. Experiments have been conducted using the dataset built

on the Chinese novel *World of Plainness*. The results show that our proposed method reaches a new state-of-the-art performance with an identification accuracy of 82.5%, which outperforms the baseline using manual features by 12%.

UnitNet-Based Hybrid Speech Synthesis

Xiao Zhou, Zhen-Hua Ling, Li-Rong Dai; USTC, China Fri-M-V-4-7, Time: 11:00

This paper presents a hybrid speech synthesis method based on UnitNet, a unified sequence-to-sequence (Seq2Seq) acoustic model for both statistical parametric speech synthesis (SPSS) and concatenative speech synthesis (CSS). This method combines CSS and SPSS approaches to synthesize different segments in an utterance. Comparing with the Tacotron2 model for Seq2Seq speech synthesis, UnitNet utilizes the phone boundaries of training data and its decoder contains autoregressive structures at both phone and frame levels. This hierarchical architecture can not only extract embedding vectors for representing phone-sized units in the corpus but also measure the dependency among consecutive units, which makes UnitNet capable of guiding the selection of phone-sized units for CSS. Furthermore, hybrid synthesis can be achieved by integrating the units generated by SPSS into the framework of CSS for the target phones without appropriate candidates in the corpus. Experimental results show that UnitNet can achieve comparable naturalness with Tacotron2 for SPSS and outperform our previous Tacotron-based method for CSS. Besides, the naturalness and inference efficiency of SPSS can be further improved through hybrid synthesis.

Dynamically Adaptive Machine Speech Chain Inference for TTS in Noisy Environment: Listen and Speak Louder

Sashi Novitasari, Sakriani Sakti, Satoshi Nakamura; NAIST, Japan

Fri-M-V-4-8, Time: 11:00

Although machine speech chains were originally proposed to mimic a closed-loop human speech chain mechanism with auditory feedback, the existing machine speech chains are only utilized as a semi-supervised learning method that allows automatic speech recognition (ASR) and text-to-speech synthesis systems (TTS) to support each other given unpaired data. During inference, however, ASR and TTS are still performed separately. This paper focuses on machine speech chain inferences in a noisy environment. In human communication, speakers tend to talk more loudly in noisy environments, a phenomenon known as the Lombard effect. Simulating the Lombard effect, we implement a machine speech chain that enables TTS to speak louder in a noisy condition given auditory feedback. The auditory feedback includes speech-to-noise ratio prediction and ASR loss as a speech intelligibility measurement. To the best of our knowledge, this is the first deep learning framework that mimics human speech perception and production behaviors in a noisy environment.

LinearSpeech: Parallel Text-to-Speech with Linear Complexity

*Haozhe Zhang*¹, *Zhihua Huang*², *Zengqiang Shang*¹, *Pengyuan Zhang*¹, *Yonghong Yan*¹; ¹*CAS*, *China*; ²*UCAS*, *China* Fri-M-V-4-9, Time: 11:00

Non-autoregressive text to speech models such as FastSpeech can synthesize speech significantly faster than previous autoregressive models with comparable quality. However, the memory and time complexity $O(N^2)$ of self-attention hinders FastSpeech from gener-

ating long sequences, where N is the length of mel-spectrograms. In this work, we propose LinearSpeech, an efficient parallel textto-speech model with memory and computational complexity O(N). Firstly, we replace standard attention modules in decoder of the model with linear attention modules to reduce the time and memory cost. Secondly, we add a novel positional encoding to standard and linear attention modules, which enable the model to learn the order of input sequence and synthesizing long mel-spectrograms. Furthermore, we use reversible residual layers instead of the standard residuals, which reduce the memory consumption in training stage. In our experiments, LinearSpeech can be trained with doubled batch size than FastSpeech with similar number of parameters. At inference, LinearSpeech achieves more than $2.0 \times$ inference speedup on CPU when synthesizing mel-spectrograms longer than 3,500. And our model can synthesize $5.5 \times$ longer mel-spectrograms than FastSpeech when running out of 12GB GPU memory. Our subjective listening test also shows that the speech quality of LinearSpeech is comparable to FastSpeech.

Fri-M-V-5: Speech Type Classification and Diagnosis

11:00–13:00, Friday 3 September 2021 Chairs: Tanja Schultz and Mahadeva Prasanna

An Agent for Competing with Humans in a Deceptive Game Based on Vocal Cues

Noa Mansbach, Evgeny Hershkovitch Neiterman, Amos Azaria; Ariel University, Israel Fri-M-V-5-1, Time: 11:00

In this work we present the development of an autonomous agent capable of competing with humans in a deception-based game. The agent predicts whether a given statement is true or false based on vocal cues. To this end, we develop a game for collecting a large scale and high quality labeled sound data-set in a controlled environment in English and Hebrew. We develop a model that can detect deception based on vocal statements from the participants of the experiment, and show that the model is more accurate than humans.

We develop an agent that uses the developed deception model and interacts with humans within our deceptive environment. We show that our agent significantly outperforms a simple agent that does not use the deception model; that is, it wins significantly more games when played against human players. In addition, we use our model to detect whether a statement will be perceived as a lie or not by human subjects, based on its vocal cues.

A Multi-Branch Deep Learning Network for Automated Detection of COVID-19

Ahmed Fakhry¹, Xinyi Jiang², Jaclyn Xiao³, Gunvant Chaudhari⁴, Asriel Han⁵; ¹University of Alexandria, Egypt; ²Independent Researcher, USA; ³Duke University, USA; ⁴University of California at San Francisco, USA; ⁵Stanford University, USA Fri-M-V-5-2, Time: 11:00

Fast and affordable solutions for COVID-19 testing are necessary to contain the spread of the global pandemic and help relieve the burden on medical facilities. Currently, limited testing locations and expensive equipment pose difficulties for individuals seeking testing, especially in low-resource settings. Researchers have successfully presented models for detecting COVID-19 infection status using audio samples recorded in clinical settings, suggesting that audiobased Artificial Intelligence models can be used to identify COVID-19. Such models have the potential to be deployed on smartphones for

fast, widespread, and low-resource testing. However, while previous studies have trained models on cleaned audio samples collected mainly from clinical settings, audio samples collected from average smartphones may yield suboptimal quality data that is different from the clean data that models were trained on. This discrepancy may add a bias that affects COVID-19 status predictions. To tackle this issue, we propose a multi-branch deep learning network that is trained and tested on crowdsourced data where most of the data has not been manually processed and cleaned. Furthermore, the model achieves state-of-art results for the COUGHVID dataset. After breaking down results for each category, we have shown an AUC of 0.99 for audio samples with COVID-19 positive labels.

RW-Resnet: A Novel Speech Anti-Spoofing Model Using Raw Waveform

Youxuan Ma, Zongze Ren, Shugong Xu; Shanghai University, China

Fri-M-V-5-3, Time: 11:00

In recent years, synthetic speech generated by advanced text-tospeech (TTS) and voice conversion (VC) systems has caused great harms to automatic speaker verification (ASV) systems, urging us to design a synthetic speech detection system to protect ASV systems. In this paper, we propose a new speech anti-spoofing model named ResWavegram-Resnet (RW-Resnet). The model contains two parts, Conv1D Resblocks and backbone Resnet34. The Conv1D Resblock is based on the Conv1D block with a residual connection. For the first part, we use the raw waveform as input and feed it to the stacked Conv1D Resblocks to get the ResWavegram. Compared with traditional methods, ResWavegram keeps all the information from the audio signal and has a stronger ability in extracting features. For the second part, the extracted features are fed to the backbone Resnet34 for the spoofed or bonafide decision. The ASVspoof2019 logical access (LA) corpus is used to evaluate our proposed RW-Resnet. Experimental results show that the RW-Resnet achieves better performance than other state-of-the-art anti-spoofing models, which illustrates its effectiveness in detecting synthetic speech attacks.

Fake Audio Detection in Resource-Constrained **Settings Using Microfeatures**

Hira Dhamyal, Ayesha Ali, Ihsan Ayyub Qazi, Agha Ali Raza; LUMS, Pakistan

Fri-M-V-5-4, Time: 11:00

Fake audio generation has undergone remarkable improvement with the advancement in deep neural network models. This has made it increasingly important to develop lightweight yet robust mechanisms for detecting fake audios, especially for resource-constrained settings such as on edge devices and embedded controllers as well as with low-resource languages. In this paper, we analyze two microfeatures: Voicing Onset Time (VOT) and coarticulation, to classify bonafide and synthesized audios. Using the ASVSpoof2019 LA dataset, we find that on average, VOT is higher in synthesized speech compared to bonafide speech and exhibits higher variance for multiple occurrences of the same stop consonants. Further, we observe that vowels in CVC form in bonafide speech have greater F1/F2 movement compared to similarly constrained vowels in synthesized speech. We also analyse the predictive power of VOT and coarticulation for detecting bonafide and synthesized speech and achieve equal error rates of 25.2% using VOT, 39.3% using coarticulation, and 23.5% using a fusion of both models. This is the first study analysing VOT and coarticulation as features for fake audio detection. We suggest these microfeatures as standalone features for speaker-dependent forensics, voice-biometrics, and for rapid pre-screening of suspicious audios, and as additional features in bigger feature sets for computationally intensive classifiers.

NOTES

Coughing-Based Recognition of Covid-19 with Spatial Attentive ConvLSTM Recurrent Neural Networks

Tianhao Yan¹, Hao Meng¹, Emilia Parada-Cabaleiro², Shuo Liu³, Meishu Song³, Björn W. Schuller³; ¹Harbin Engineering University, China; ² Johannes Kepler Universität Linz, Austria; ³Universität Augsburg, Germany

Fri-M-V-5-5, Time: 11:00

The rapid emergence of COVID-19 has become a major public health threat around the world. Although early detection is crucial to reduce its spread, the existing diagnostic methods are still insufficient in bringing the pandemic under control. Thus, more sophisticated systems, able to easily identify the infection from a larger variety of symptoms, such as cough, are urgently needed. Deep learning models can indeed convey numerous signal features relevant to fight against the disease; yet, the performance of state-of-the-art approaches is still severely restricted by the feature information loss typically due to the high number of layers. To mitigate this phenomenon, identifying the most relevant feature areas by drawing into attention mechanisms becomes essential. In this paper, we introduce Spatial Attentive ConvLSTM-RNN (SACRNN), a novel algorithm that is using Convolutional Long-Short Term Memory Recurrent Neural Networks with embedded attention that has the ability to identify the most valuable features. The promising results achieved by the fusion between the proposed model and a conventional Attentive Convolutional Recurrent Neural Network, on the automatic recognition of COVID-19 coughing (73.2% of Unweighted Average Recall) show the great potential of the presented approach in developing efficient solutions to defeat the pandemic.

Knowledge Distillation for Singing Voice Detection

Soumava Paul, Gurunath Reddy M., K. Sreenivasa Rao, Partha Pratim Das; IIT Kharagpur, India Fri-M-V-5-6, Time: 11:00

Singing Voice Detection (SVD) has been an active area of research in music information retrieval (MIR). Currently, two deep neural network-based methods, one based on CNN and the other on RNN, exist in literature that learn optimized features for the voice detection (VD) task and achieve state-of-the-art performance on common datasets. Both these models have a huge number of parameters (1.4M for CNN and 65.7K for RNN) and hence not suitable for deployment on devices like smartphones or embedded sensors with limited capacity in terms of memory and computation power. The most popular method to address this issue is known as knowledge distillation in deep learning literature (in addition to model compression) where a large pre-trained network known as the teacher is used to train a smaller student network. Given the wide applications of SVD in music information retrieval, to the best of our knowledge, model compression for practical deployment has not yet been explored. In this paper, efforts have been made to investigate this issue using both conventional as well as ensemble knowledge distillation techniques.

Age Estimation with Speech-Age Model for **Heterogeneous Speech Datasets**

Ryu Takeda, Kazunori Komatani; Osaka University, Japan

Fri-M-V-5-7, Time: 11:00

This paper describes an age estimation method from speech signals for heterogeneous datasets. Although previous studies in the speech field evaluate age prediction models with held-out testing data within the same dataset recorded in a consistent setting, such evaluation does not measure real performance. The difficulty of heterogeneous datasets is overfitting caused by the corpus-specific properties: transfer function of the recording environment and distributions of age and speaker. We propose a speech-age model and its integration with sequence neural networks (NNs). The speech-age model represents the ambiguity of age as a probability distribution, which also virtually extends the limited range of age distribution of each corpus. A Bayesian generative model successfully integrates the speech-age model and the NNs. We also applied mean normalization technique to cope with the transfer function problem. Experiments showed that our proposed method outperformed the baseline neural classifier for completely open test sets in the age distribution and recording setting.

Open-Set Audio Classification with Limited Training Resources Based on Augmentation Enhanced Variational Auto-Encoder GAN with Detection-Classification Joint Training

*Kah Kuan Teh, Huy Dat Tran; A*STAR, Singapore* Fri-M-V-5-8, Time: 11:00

In this paper, we propose a novel method to address practical problems when deploying audio classification systems in operations that are the presence of unseen sound classes (open-set) and the limitation of training resources. To solve it, a novel method which embeds variational auto-encoder (VAE), data augmentation and detection-classification joint training into conventional GAN networks is proposed. The VAE input to GAN-generator helps to generate realistic outlier samples which are not too far from in-distribution class and hence improve the open-set discrimination capabilities of classifiers. Next, the augmentation enhanced GAN scheme developed in our previous work [4] for close-set audio classification, will help to address the limited training resources by in cooperating the physical data augmentation to work together with traditional GAN produced samples to prevent overfitting and improve the optimization convergences. The detection-classification joint training further steps on advantages of VAE and Augmentation GAN to further improving the performances of detection and classification tasks. The experiments carried out on Google Speech Command database show great improvements of open-set classification accuracy from 62.41% to 88.29% when using only 10% amount of training data.

Deep Spectral-Cepstral Fusion for Shouted and Normal Speech Classification

Takahiro Fukumori; Ritsumeikan University, Japan Fri-M-V-5-9, Time: 11:00

Discrimination between shouted and normal speech is crucial in audio surveillance and monitoring. Although deep neural networks are used in recent methods, traditional low-level speech features are applied, such as mel-frequency cepstral coefficients and the mel spectrum. This paper presents a deep spectral-cepstral fusion approach that learns descriptive features for target classification from high-dimensional spectrograms and cepstrograms. We compare the following three types of architectures as base networks: convolutional neural networks (CNNs), gated recurrent unit (GRU) networks, and their combination (CNN-GRU). Using a corpus comprising real shouts and speech, we present a comprehensive comparison with conventional methods to verify the effectiveness of the proposed feature learning method. The results of experiments conducted in various noisy environments demonstrate that the CNN-GRU based on our spectral-cepstral features achieves better classification performance than single feature-based networks. This finding suggests the effectiveness of using high-dimensional sources for speech-type recognition in sound event detection.

Automatic Detection of Shouted Speech Segments in Indian News Debates

Shikha Baghel¹, Mrinmoy Bhattacharjee¹, S.R. Mahadeva Prasanna², Prithwijit Guha¹; ¹IIT Guwahati, India; ²IIT Dharwad, India Fri-M-V-5-10, Time: 11:00

Shouted speech detection is an essential pre-processing step in conventional speech processing systems such as speech and speaker recognition, speaker diarization, and others. Excitation source plays an important role in shouted speech production. This work explores feature computed from the Integrated Linear Prediction Residual (ILPR) signal for shouted speech detection in Indian news debates. The log spectrogram of ILPR signal provides time-frequency characteristics of excitation source signal. The proposed shouted speech detection system is deep network with CNN-based autoencoder and attention-based classifier sub-modules. The Autoencoder sub-network aids the classifier in learning discriminative deep embeddings for better classification. The proposed classifier is equipped with attention mechanism and Bidirectional Gated Recurrent Units. Classification results show that the proposed system with excitation feature performs better than baseline log spectrogram computed from the pre-emphasized speech signal. A score-level fusion of the classifiers trained on the source feature and the baseline feature provides the best performance. The performance of the proposed shouted speech detection is also evaluated at various speech segment durations.

Generalized Spoofing Detection Inspired from Audio Generation Artifacts

*Yang Gao*¹, *Tyler Vuong*¹, *Mahsa Elyasi*², *Gaurav Bharaj*², *Rita Singh*¹; ¹*Carnegie Mellon University, USA*; ²*AI Foundation, USA* Fri-M-V-5-11, Time: 11:00

State-of-the-art methods for audio generation suffer from fingerprint artifacts and repeated inconsistencies across temporal and spectral domains. Such artifacts could be well captured by the frequency domain analysis over the spectrogram. Thus, we propose a novel use of long-range spectro-temporal modulation feature - 2D DCT over log-Mel spectrogram for the audio deepfake detection. We show that this feature works better than log-Mel spectrogram, CQCC, MFCC, as a suitable candidate to capture such artifacts. We employ spectrum augmentation and feature normalization to decrease overfitting and bridge the gap between training and test dataset along with this novel feature introduction. We developed a CNN-based baseline that achieved a 0.0849 t-DCF and outperformed the previously top single systems reported in the ASVspoof 2019 challenge. Finally, by combining our baseline with our proposed 2D DCT spectro-temporal feature, we decrease the t-DCF score down by 14% to 0.0737, making it a state-of-the-art system for spoofing detection. Furthermore, we evaluate our model using two external datasets, showing the proposed feature's generalization ability. We also provide analysis and ablation studies for our proposed feature and results.

Overlapped Speech Detection Based on Spectral and Spatial Feature Fusion

*Weiguang Chen*¹, *Van Tung Pham*², *Eng Siong Chng*², *Xionghu Zhong*¹; ¹*Hunan University, China*; ²*NTU*,

Singapore

Fri-M-V-5-12, Time: 11:00

Overlapped speech is widely present in conversations and can cause significant performance degradation on speech processing such as diarization, enhancement, and recognition. Detection of overlapped

speech, in particular when the speakers are in the far-field, is a challenging task as the overlapped part is usually short, and heavy reverberation and noise may present in the conversation scenario. Existing solutions overwhelmingly rely on spectral features extracted from single microphone signal to perform the detection. In this paper, we propose a novel detection approach which is able to use a microphone array and fuse the spatial and spectral features extracted from multi-channel array signal. Two categories of spatial features, directional statistics which are projected to spherical location grids and generalized cross-correlation function based on phase transform (GCC-PHAT), are considered to model the speaker's spatial characteristic. Such spatial features are then fused with the spectral features to detect the overlapped speech by using a Gated Multimodal Unit (GMU). The performance of the proposed approach is studied under AMI and CHiME-6 corpora. Experimental results show that the proposed feature fusion approach achieves better performance than methods using spectral features only.

Fri-M-V-6: Spoken Term Detection & Voice Search

11:00–13:00, Friday 3 September 2021 Chairs: Jan Trmal and Xiaodong Cui

Do Acoustic Word Embeddings Capture Phonological Similarity? An Empirical Study

Badr M. Abdullah, Marius Mosbach, Iuliia Zaitova, Bernd Möbius, Dietrich Klakow; Universität des Saarlandes, Germany

Fri-M-V-6-1, Time: 11:00

Several variants of deep neural networks have been successfully employed for building parametric models that project variable-duration spoken word segments onto fixed-size vector representations, or acoustic word embeddings (AWEs). However, it remains unclear to what degree we can rely on the distance in the emerging AWE space as an estimate of word-form similarity. In this paper, we ask: does the distance in the acoustic embedding space correlate with phonological dissimilarity? To answer this question, we empirically investigate the performance of supervised approaches for AWEs with different neural architectures and learning objectives. We train AWE models in controlled settings for two languages (German and Czech) and evaluate the embeddings on two tasks: word discrimination and phonological similarity. Our experiments show that (1) the distance in the embedding space in the best cases only moderately correlates with phonological distance, and (2) improving the performance on the word discrimination task does not necessarily yield models that better reflect word phonological similarity. Our findings highlight the necessity to rethink the current intrinsic evaluations for AWEs.

Paraphrase Label Alignment for Voice Application Retrieval in Spoken Language Understanding

Zheng Gao, Radhika Arava, Qian Hu, Xibin Gao, Thahir Mohamed, Wei Xiao, Mohamed AbdelHady; Amazon, USA

Fri-M-V-6-2, Time: 11:00

Spoken language understanding (SLU) smart assistants such as Amazon Alexa host hundreds of thousands of voice applications (skills) to delight end-users and fulfill their utterance requests. Sometimes utterances fail to be claimed by smart assistants due to system problems such as model incapability or routing errors. The failure may lead to customer frustration, dialog termination and eventually cause customer churn. To avoid this, we design a skill retrieval system as a downstream service to suggest fallback skills

NOTES

to unclaimed utterances. If the suggested skill satisfies customer intent, the conversation will be recovered with the assistant. For the sake of smooth customer experience, we only present the most relevant skill to customers, resulting in partial observation problem which constrains retrieval model training. To solve this problem, we propose a two-step approach to automatically align claimed utterance labels to unclaimed utterances. Extensive experiments on two real-world datasets demonstrate that our proposed model significantly outperforms a number of strong alternatives.

Personalized Keyphrase Detection Using Speaker and Environment Information

Rajeev Rikhye, Quan Wang, Qiao Liang, Yanzhang He, Ding Zhao, Yiteng Huang, Arun Narayanan, Ian McGraw; Google, USA Fri-M-V-6-3, Time: 11:00

In this paper, we introduce a streaming keyphrase detection system that can be easily customized to accurately detect any phrase composed of words from a large vocabulary. The system is implemented with an end-to-end trained automatic speech recognition (ASR) model and a text-independent speaker verification model. To address the challenge of detecting these keyphrases under various noisy conditions, a speaker separation model is added to the feature frontend of the speaker verification model, and an adaptive noise cancellation (ANC) algorithm is included to exploit cross-microphone noise coherence. Our experiments show that the text-independent speaker verification model largely reduces the false triggering rate of the keyphrase detection, while the speaker separation model and adaptive noise cancellation largely reduce false rejections.

Streaming Transformer for Hardware Efficient Voice Trigger Detection and False Trigger Mitigation

Vineet Garg¹, Wonil Chang¹, Siddharth Sigtia², Saurabh Adya¹, Pramod Simha¹, Pranay Dighe¹, Chandra Dhir¹; ¹Apple, USA; ²Apple, UK Fri-M-V-6-4, Time: 11:00

We present a unified and hardware efficient architecture for two stage voice trigger detection (VTD) and false trigger mitigation (FTM) tasks. Two stage VTD systems of voice assistants can get falsely activated to audio segments acoustically similar to the trigger phrase of interest. FTM systems cancel such activations by using post trigger audio context. Traditional FTM systems rely on automatic speech recognition lattices which are computationally expensive to obtain on device. We propose a streaming transformer (TF) encoder architecture, which progressively processes incoming audio chunks and maintains audio context to perform both VTD and FTM tasks using only acoustic features. The proposed joint model yields an average 18% relative reduction in false reject rate (FRR) for the VTD task at a given false alarm rate. Moreover, our model suppresses 95% of the false triggers with an additional one second of post-trigger audio. Finally, on-device measurements show 32% reduction in runtime memory and 56% reduction in inference time compared to non-streaming version of the model.

Few-Shot Keyword Spotting in Any Language

Mark Mazumder¹, Colby Banbury¹, Josh Meyer², Pete Warden³, Vijay Janapa Reddi¹; ¹Harvard University, USA; ²Coqui, Germany; ³Google, USA Fri-M-V-6-5, Time: 11:00

We introduce a few-shot transfer learning method for keyword spotting in any language. Leveraging open speech corpora in nine languages, we automate the extraction of a large multilingual keyword bank and use it to train an embedding model. With just five training examples, we fine-tune the embedding model for keyword spotting and achieve an average F_1 score of 0.75 on keyword classification for 180 new keywords unseen by the embedding model in these nine languages. This embedding model also generalizes to new languages. We achieve an average F_1 score of 0.65 on 5-shot models for 260 keywords sampled across 13 new languages unseen by the embedding model. We investigate streaming accuracy for our 5-shot models in two contexts: keyword spotting and keyword search. Across 440 keywords in 22 languages, we achieve an average streaming keyword spotting accuracy of 87.4% with a false acceptance rate of 4.3%, and observe promising initial results on keyword search.

Text Anchor Based Metric Learning for Small-Footprint Keyword Spotting

Li Wang, Rongzhi Gu, Nuo Chen, Yuexian Zou; Peking University, China Fri-M-V-6-6, Time: 11:00

Keyword Spotting (KWS) remains challenging to achieve the trade-off between small footprint and high accuracy. Recently proposed metric learning approaches improved the generalizability of models for the KWS task, and 1D-CNN based KWS models have achieved the state-of-the-arts (SOTA) in terms of model size. However, for metric learning, due to data limitations, the speech anchor is highly susceptible to the acoustic environment and speakers. Also, we note that the 1D-CNN models have limited capability to capture long-term temporal acoustic features. To address the above problems, we propose to utilize text anchors to improve the stability of anchors. Furthermore, a new type of model (LG-Net) is exquisitely designed to promote long-short term acoustic feature modeling based on 1D-CNN and self-attention. Experiments are conducted on Google Speech Commands Dataset version 1 (GSCDv1) and 2 (GSCDv2). The results demonstrate that the proposed text anchor based metric learning method shows consistent improvements over speech anchor on representative CNN-based models. Moreover, our LG-Net model achieves SOTA accuracy of 97.67% and 96.79% on two datasets, respectively. It is encouraged to see that our lighter LG-Net with only 74k parameters obtains 96.82% KWS accuracy on the GSCDv1 and 95.77% KWS accuracy on the GSCDv2.

A Meta-Learning Approach for User-Defined Spoken Term Classification with Varying Classes and Examples

*Yangbin Chen*¹, *Tom Ko*², *Jianping Wang*³; ¹*CUHK*, *China*; ²*SUSTech, China*; ³*CityU, China* Fri-M-V-6-7, Time: 11:00

Recently we formulated a user-defined spoken term classification task as a few-shot learning task and tackled the task using Model-Agnostic Meta-Learning (MAML) algorithm. Our results show that the meta-learning approach performs much better than conventional supervised learning and transfer learning in the task, especially with limited training data. In this paper, we extend our work by addressing a more practical problem in the user-defined scenario where users can define any number of spoken terms and provide any number of enrollment audio examples for each spoken term. From the perspective of few-shot learning, this is an N-way, K-shot problem with varying N and K. In our work, we relax the values of N and K of each meta-task during training instead of assigning fixed values to them, which differs from what most meta-learning algorithms do. We adopt a metric-based meta-learning algorithm named Prototypical Networks (ProtoNet) as it avoids exhaustive fine-tuning when N varies. Furthermore, we use the Max-Mahalanobis Center (MMC) loss as an effective regularizer to address the problem of ProtoNet under the condition of varying K. Experiments on the Google Speech Commands dataset demonstrate that our proposed method outperforms the conventional *N*-way, *K*-shot setting in most testing tasks.

Auxiliary Sequence Labeling Tasks for Disfluency Detection

Dongyub Lee¹, Byeongil Ko¹, Myeong Cheol Shin¹, Taesun Whang², Daniel Lee¹, Eunhwa Kim¹, Eunggyun Kim¹, Jaechoon Jo³; ¹Kakao, Korea; ²Wisenut, Korea; ³Hanshin University, Korea Fri-M-V-6-8, Time: 11:00

Detecting disfluencies in spontaneous speech is an important preprocessing step in natural language processing and speech recognition applications. Existing works for disfluency detection have focused on designing a single objective only for disfluency detection, while auxiliary objectives utilizing linguistic information of a word such as named entity or part-of-speech information can be effective. In this paper, we focus on detecting disfluencies on spoken transcripts and propose a method utilizing named entity recognition (NER) and part-of-speech (POS) as auxiliary sequence labeling (SL) tasks for disfluency detection. First, we investigate cases that utilizing linguistic information of a word can prevent mispredicting important words and can be helpful for the correct detection of disfluencies. Second, we show that training a disfluency detection model with auxiliary SL tasks can improve its F-score in disfluency detection. Then, we analyze which auxiliary SL tasks are influential depending on baseline models. Experimental results on the widely used English Switchboard dataset show that our method outperforms the previous state-of-the-art in disfluency detection.

Energy-Friendly Keyword Spotting System Using Add-Based Convolution

Hang Zhou, Wenchao Hu, Yu Ting Yeung, Xiao Chen; Huawei Technologies, China Fri-M-V-6-9, Time: 11:00

Wake-up keyword of a keyword spotting (KWS) system represents brand name of a smart device. Performance of KWS is also crucial for modern speech based human-device interaction. An on-device KWS with both high accuracy and low power consumption is desired. We propose a KWS with add-based convolution layers, namely Add TC-ResNet. Add-based convolution paves a new way to reduce power consumption of KWS system, as addition is more energy efficient than multiplication at hardware level. On Google Speech Commands dataset V2, Add TC-ResNet achieves an accuracy of 97.1%, with 99% of multiplication operations are replaced by addition operations. The result is competitive to a state-of-the-art fully multiplication-based TC-ResNet KWS. We also investigate knowledge distillation and a mixed addition-multiplication design for the proposed KWS, which leads to further performance improvement.

The 2020 Personalized Voice Trigger Challenge: Open Datasets, Evaluation Metrics, Baseline System and Results

Yan Jia¹, Xingming Wang¹, Xiaoyi Qin¹, Yinping Zhang², Xuyang Wang², Junjie Wang², Dong Zhang³, Ming Li¹; ¹Duke Kunshan University, China; ²Lenovo, China; ³Sun Yat-sen University, China Fri-M-V-6-10, Time: 11:00

The 2020 Personalized Voice Trigger Challenge (PVTC2020) addresses two different research problems in a unified setup: joint

wake-up word detection with speaker verification on close-talking single microphone data and far-field multi-channel microphone array data. Specially, the second task poses an additional cross-channel matching challenge on top of the far-field condition. To simulate the real-life application scenario, the enrollment utterances are recorded from close-talking cell-phone only, while the test utterances are recorded from both the close-talking cell-phone and the far-field microphone arrays. This paper introduces our challenge setup and the released database as well as the evaluation metrics. In addition, we present a sequential two stage end-to-end neural network baseline system trained with the proposed database for speaker-dependent wake-up word detection. Results show that state-of-the-art personalized voice trigger methods are still based on the two stage design, however, this benchmark database could also be used to evaluate multi-task joint learning methods. The official website, the open-source baseline system and results of submitted systems have been released.

Auto-KWS 2021 Challenge: Task, Datasets, and Baselines

Jingsong Wang¹, Yuxuan He¹, Chunyu Zhao¹, Qijie Shao², Wei-Wei Tu¹, Tom Ko³, Hung-yi Lee⁴, Lei Xie²; ¹4Paradigm, China; ²Northwestern Polytechnical University, China; ³SUSTech, China; ⁴National Taiwan University, Taiwan Fri-M-V-6-11, Time: 11:00

Auto-KWS 2021 challenge calls for automated machine learning (AutoML) solutions to automate the process of applying machine learning to a customized keyword spotting task. Compared with other keyword spotting tasks, Auto-KWS challenge has the following three characteristics: 1) The challenge focuses on the problem of customized keyword spotting, where the target device can only be awakened by an enrolled speaker with his/her specified keyword. The speaker can use any language and accent to define his keyword. 2) All data of the challenge is recorded in realistic environment to simulate different user scenarios. 3) Auto-KWS is a "code competition", where participants need to submit AutoML solutions, then the platform automatically runs the enrollment and prediction steps with the submitted code. This challenge aims at promoting the development of a more personalized and flexible keyword spotting system. Two baseline systems are provided to all participants as references.

Keyword Transformer: A Self-Attention Model for Keyword Spotting

Axel Berg, Mark O'Connor, Miguel Tairum Cruz; Arm, UK

Fri-M-V-6-12, Time: 11:00

The Transformer architecture has been successful across many domains, including natural language processing, computer vision and speech recognition. In keyword spotting, self-attention has primarily been used on top of convolutional or recurrent encoders. We investigate a range of ways to adapt the Transformer architecture to keyword spotting and introduce the Keyword Transformer (KWT), a fully self-attentional architecture that exceeds state-of-the-art performance across multiple tasks without any pre-training or additional data. Surprisingly, this simple architecture outperforms more complex models that mix convolutional, recurrent and attentive layers. KWT can be used as a drop-in replacement for these models, setting two new benchmark records on the Google Speech Commands dataset with 98.6% and 97.7% accuracy on the 12 and 35-command tasks respectively.

NOTES

Teaching Keyword Spotters to Spot New Keywords with Limited Examples

Abhijeet Awasthi, Kevin Kilgour, Hassan Rom; Google, Switzerland

Fri-M-V-6-13, Time: 11:00

Learning to recognize new keywords with just a few examples is essential for personalizing keyword spotting (KWS) models to a user's choice of keywords. However, modern KWS models are typically trained on large datasets and restricted to a small vocabulary of keywords, limiting their transferability to a broad range of unseen keywords. Towards easily customizable KWS models, we present KeySEM (Keyword Speech EMbedding), a speech embedding model pre-trained on the task of recognizing a large number of keywords. Speech representations offered by KeySEM are highly effective for learning new keywords from a limited number of examples. Comparisons with a diverse range of related work across several datasets show that our method achieves consistently superior performance with fewer training examples. Although KeySEM was pre-trained only on English utterances, the performance gains also extend to datasets from four other languages indicating that KeySEM learns useful representations well aligned with the task of keyword spotting. Finally, we demonstrate KeySEM's ability to learn new keywords sequentially without requiring to re-train on previously learned keywords. Our experimental observations suggest that KeySEM is well suited to on-device environments where post-deployment learning and ease of customization are often desirable.

Fri-M-V-7: Voice Anti-Spoofing and Countermeasure

11:00-13:00, Friday 3 September 2021 Chairs: Anna Silnova and Massimiliano Todisco

A Comparative Study on Recent Neural Spoofing Countermeasures for Synthetic Speech Detection

Xin Wang, Junichi Yamagishi; NII, Japan Fri-M-V-7-1, Time: 11:00

A great deal of recent research effort on speech spoofing countermeasures has been invested into back-end neural networks and training criteria. We contribute to this effort with a comparative perspective in this study. Our comparison of countermeasure models on the ASVspoof 2019 logical access scenario takes into account common strategies to deal with input trials of varied length, recently proposed margin-based training criteria, and widely used We also measured intra-model differences through front ends. multiple training-evaluation rounds with random initialization. Our statistical analysis demonstrates that the performance of the same model may be statistically significantly different when just changing the random initial seed. We thus recommend similar statistical analysis or reporting results of multiple runs for further research on the database. Despite the intra-model differences, we observed a few promising techniques, including average pooling, to efficiently process varied-length inputs and a new hyper-parameter-free loss function. The two techniques led to the best single model in our experiment, which achieved an equal error rate of 1.92% and was significantly different in statistical sense from most of the other experimental models.

An Initial Investigation for Detecting Partially Spoofed Audio

Lin Zhang¹, Xin Wang¹, Erica Cooper¹, Junichi Yamagishi¹, Jose Patino², Nicholas Evans²; ¹NII, Japan; ²EURECOM, France

Fri-M-V-7-2, Time: 11:00

All existing databases of spoofed speech contain attack data that is spoofed in its entirety. In practice, it is entirely plausible that successful attacks can be mounted with utterances that are only partially spoofed. By definition, partially-spoofed utterances contain a mix of both spoofed and bona fide segments, which will likely degrade the performance of countermeasures trained with entirely spoofed utterances. This hypothesis raises the obvious question: 'Can we detect partially-spoofed audio?' This paper introduces a new database of partially-spoofed data, named PartialSpoof, to help address this question. This new database enables us to investigate and compare the performance of countermeasures on both utterance- and segmental- level labels. Experimental results using the utterance-level labels reveal that the reliability of countermeasures trained to detect fully-spoofed data is found to degrade substantially when tested with partially-spoofed data, whereas training on partially-spoofed data performs reliably in the case of both fullyand partially-spoofed utterances. Additional experiments using segmental-level labels show that spotting injected spoofed segments included in an utterance is a much more challenging task even if the latest countermeasure models are used.

Siamese Network with wav2vec Feature for Spoofing Speech Detection

Yang Xie, Zhenchuan Zhang, Yingchun Yang; Zhejiang University, China

Fri-M-V-7-3, Time: 11:00

Automatic speaker verification is vulnerable to spoofing attacks with synthesized or converted speech. Although high-performance anti-spoofing countermeasures can achieve high accuracy when the training and testing spoofing attack examples are similarly distributed, their performance degrades significantly when confronted with out-of-distribution spoofing speech, which is created by increasingly advanced unseen speech synthesis and voice conversion methods. Since it is unrealistic to collect enough labeled data from each new spoofing attack method, we argue that addressing the problem of out-of-distribution generalization for spoofing speech detection is essential. In this work, we propose a two-phase representation learning system based on a Siamese network for spoofing speech detection tasks. During the representation learning phase, an embedding Siamese neural network is trained with the wav2vec features to distinguish whether the speech samples in a pair belong to the same category. The proposed system decreases the equal error rate from the state-of-the-art result of 4.07% to 1.15% on the ASVspoof 2019 evaluation set.

Cross-Database Replay Detection in Terminal-Dependent Speaker Verification

Xingliang Cheng, Mingxing Xu, Thomas Fang Zheng; Tsinghua University, China Fri-M-V-7-4, Time: 11:00

The vulnerability of automatic speaker verification (ASV) systems against replay attacks becomes a severe problem. Although various methods have been proposed for replay detection, the generalization capability is still limited. For instance, a detection model trained on one database may fully fail when tested on another database. In this paper, we adopt the one-class learning technology to address the cross-database problem. Different from conventional two-class

models that discriminate genuine speeches from replay attacks, the one-class model focuses on the within-class variance of genuine speeches, which is naturally robust to unseen attacks. In this study, we choose the Gaussian mixture model (GMM) as the one-class model and design two utterance-level features which reduce the uncertainties of genuine class while still be distinguishable from non-genuine class. Experiments conducted on three public replay datasets show that, compared to the state-of-the-art methods, the proposed method demonstrates promising generalization capability under cross-database scenarios.

The Effect of Silence and Dual-Band Fusion in Anti-Spoofing System

Yuxiang Zhang, Wenchao Wang, Pengyuan Zhang; CAS, China

Fri-M-V-7-5, Time: 11:00

The current neural network based anti-spoofing systems have poor robustness. Their performance degrades further after voice activity detection (VAD) performed, making it difficult to be applied in practice. This work investigated the effect of silence at the beginning and end of speech, finding that silent differences are part of the basis for countermeasures' judgements. The reason for the performance deterioration caused by VAD is also explored. The experimental results demonstrate that the neural network loses the information about silent segments after the VAD operation removes them. This can lead to more serious overfitting. In order to solve the overfitting problem, the work in this paper also analyzes the reasons for system overfitting from different frequency sub-bands. It is found that the high-frequency part of the feature is the main cause of system overfitting, while the low-frequency part is more robust but less accurate against known attacks. Therefore, we propose the dual-band fusion anti-spoofing algorithm, which requires only two sub-systems but outperforms all but one primary system submitted to the logical access condition of the ASVspoof 2019 challenge. Our system has an EER of 3.50% even after VAD operations performed, thus can be put into practical application.

Pairing Weak with Strong: Twin Models for Defending Against Adversarial Attack on Speaker Verification

Zhiyuan Peng, Xu Li, Tan Lee; CUHK, China Fri-M-V-7-6, Time: 11:00

Vulnerability of speaker verification (SV) systems under adversarial attack receives wide attention recently. Simple and effective countermeasures against such attack are yet to be developed. This paper formulates the task of adversarial defense as a problem of attack detection. The detection is made possible with the verification scores from a pair of purposely selected SV models. The twin-model design comprises a fragile model paired up with a relatively robust one. The two models show prominent score inconsistency under adversarial attack. To detect the score inconsistency, a simple one-class classifier is adopted. The classifier is trained with normal speech samples, which not only bypasses the need of crafting adversarial samples but also prevents itself from over-fitting to the crafted samples, and hence makes the detection robust to unseen attacks. Compared to single-model systems, the proposed system shows consistent and significant performance improvement against different attack strategies. The false acceptance rates (FARs) are reduced from over 63.54% to 2.26% under the strongest attack. Our approach has practical benefits, e.g., no need to modify a well-deployed SV model even it is well-known and can be fully accessed by the adversary. Moreover, it can be combined with existing single-model countermeasures for even stronger defenses.

Attention-Based Convolutional Neural Network for ASV Spoofing Detection

Hefei Ling, Leichao Huang, Junrui Huang, Baiyan Zhang, Ping Li; HUST, China Fri-M-V-7-7, Time: 11:00

In recent years, automatic speaker verification (ASV) algorithms have undergone significant progress. They have been widely deployed in different applications, but the ASV systems are vulnerable to spoofing attacks, such as impersonation, replay, text-to-speech, voice conversion and the recently emerged adversarial attacks. To improve the robustness of the ASV system, researchers have designed anti-spoofing systems to resist spoofing attacks. While previously proposed systems have shown to be effective for spoof attacks detection, they are all ensemble methods based on different speech representations and architectures at the cost of increased model complexity, with similar performance not being achieved with single systems. This paper proposes an attention-based single convolutional neural network to learn discriminative feature embedding for spoof detection, achieving performance comparable to ensemble methods. The key idea is to decrease the information redundancy among channels and focus on the most informative sub-bands of speech representations. The experiments show that our proposed single system achieves an equal error rate of 1.87% on the evaluation set of ASVspoof 2019 Challenge, outperforming all single systems and comparable to the second-ranked system (EER 1.86%) among all known systems.

Voting for the Right Answer: Adversarial Defense for Speaker Verification

Haibin Wu¹, Yang Zhang¹, Zhiyong Wu¹, Dong Wang¹, Hung-yi Lee²; ¹Tsinghua University, China; ²National Taiwan University, Taiwan Fri-M-V-7-8, Time: 11:00

Automatic speaker verification (ASV) is a well developed technology for biometric identification, and has been ubiquitous implemented in security-critic applications, such as banking and access control. However, previous works have shown that ASV is under the radar of adversarial attacks, which are very similar to their original counterparts from human's perception, yet will manipulate the ASV render wrong prediction. Due to the very late emergence of adversarial attacks for ASV, effective countermeasures against them are limited. Given that the security of ASV is of high priority, in this work, we propose the idea of "voting for the right answer" to prevent risky decisions of ASV in blind spot areas, by employing random sampling and voting. Experimental results show that our proposed method improves the robustness against both the limited-knowledge attackers by pulling the adversarial samples out of the blind spots, and the sufficient-knowledge attackers by introducing randomness and increasing the attackers' budgets.

Visualizing Classifier Adjacency Relations: A Case Study in Speaker Verification and Voice Anti-Spoofing

Tomi Kinnunen¹, Andreas Nautsch², Md. Sahidullah³, Nicholas Evans², Xin Wang⁴, Massimiliano Todisco², Héctor Delgado⁵, Junichi Yamagishi⁴, Kong Aik Lee⁶; ¹University of Eastern Finland, Finland; ²EURECOM, France; ³Inria, France; ⁴NII, Japan; ⁵Nuance Communications, Spain; ⁶A*STAR, Singapore Fri-M-V-7-9, Time: 11:00

Whether it be for results summarization, or the analysis of classifier fusion, some means to compare different classifiers can often provide illuminating insight into their behaviour, (dis)similarity or complementarity. We propose a simple method to derive 2D representation from detection scores produced by an arbitrary set of binary classifiers in response to a common dataset. Based upon rank correlations, our method facilitates a visual comparison of classifiers with arbitrary scores and with close relation to receiver operating characteristic (ROC) and detection error trade-off (DET) analyses. While the approach is fully versatile and can be applied to any detection task, we demonstrate the method using scores produced by automatic speaker verification and voice anti-spoofing systems. The former are produced by a Gaussian mixture model system trained with VoxCeleb data whereas the latter stem from submissions to the ASVspoof 2019 challenge.

Representation Learning to Classify and Detect Adversarial Attacks Against Speaker and Speech Recognition Systems

Jesús Villalba, Sonal Joshi, Piotr Żelasko, Najim Dehak; Johns Hopkins University, USA Fri-M-V-7-10, Time: 11:00

Adversarial attacks have become a major threat for machine learning applications. There is a growing interest in studying these attacks in the audio domain, e.g, speech and speaker recognition; and find defenses against them. In this work, we focus on using representation learning to classify/detect attacks w.r.t. the attack algorithm, threat model or signal-to-adversarial-noise ratio. We found that common attacks in the literature can be classified with accuracies as high as 90%. Also, representations trained to classify attacks against speaker identification can be used also to classify attacks against speaker verification and speech recognition. We also tested an attack verification task, where we need to decide whether two speech utterances contain the same attack. We observed that our models did not generalize well to attack algorithms not included in the attack representation model training. Motivated by this, we evaluated an unknown attack detection task. We were able to detect unknown attacks with equal error rates of about 19%, which is promising.

An Empirical Study on Channel Effects for Synthetic Voice Spoofing Countermeasure Systems

You Zhang, Ge Zhu, Fei Jiang, Zhiyao Duan; University of Rochester, USA

Fri-M-V-7-11, Time: 11:00

Spoofing countermeasure (CM) systems are critical in speaker verification; they aim to discern spoofing attacks from bona fide speech trials. In practice, however, acoustic condition variability in speech utterances may significantly degrade the performance of CM systems. In this paper, we conduct a cross-dataset study on several state-of-the-art CM systems and observe significant performance degradation compared with their single-dataset performance. Observing differences of average magnitude spectra of bona fide utterances across the datasets, we hypothesize that channel mismatch among these datasets is one important reason. We then verify it by demonstrating a similar degradation of CM systems trained on original but evaluated on channel-shifted data. Finally, we propose several channel robust strategies (data augmentation, multi-task learning, adversarial learning) for CM systems, and observe a significant performance improvement on cross-dataset experiments.

Channel-Wise Gated Res2Net: Towards Robust Detection of Synthetic Speech Attacks

Xu Li¹, *Xixin* Wu², *Hui* Lu¹, *Xunying* Liu¹, *Helen* Meng¹; ¹*CUHK*, *China*; ²*University of Cambridge*, *UK* Fri-M-V-7-12, Time: 11:00

Existing approaches for anti-spoofing in automatic speaker verification (ASV) still lack generalizability to unseen attacks. The Res2Net

approach designs a residual-like connection between feature groups within one block, which increases the possible receptive fields and improves the system's detection generalizability. However, such a residual-like connection is performed by a direct addition between feature groups without channel-wise priority. We argue that the information across channels may not contribute to spoofing cues equally, and the less relevant channels are expected to be suppressed before adding onto the next feature group, so that the system can generalize better to unseen attacks. This argument motivates the current work that presents a novel, channel-wise gated Res2Net (CG-Res2Net), which modifies Res2Net to enable a channel-wise gating mechanism in the connection between feature groups. This gating mechanism dynamically selects channel-wise features based on the input, to suppress the less relevant channels and enhance the detection generalizability. Three gating mechanisms with different structures are proposed and integrated into Res2Net. Experimental results conducted on ASVspoof 2019 logical access (LA) demonstrate that the proposed CG-Res2Net significantly outperforms Res2Net on both the overall LA evaluation set and individual difficult unseen attacks, which also outperforms other state-of-the-art single systems, depicting the effectiveness of our method.

Partially-Connected Differentiable Architecture Search for Deepfake and Spoofing Detection

Wanying Ge, Michele Panariello, Jose Patino, Massimiliano Todisco, Nicholas Evans; EURECOM, France Fri-M-V-7-13, Time: 11:00

This paper reports the first successful application of a differentiable architecture search (DARTS) approach to the deepfake and spoofing detection problems. An example of neural architecture search, DARTS operates upon a continuous, differentiable search space which enables both the architecture and parameters to be optimised via gradient descent. Solutions based on partially-connected DARTS use random channel masking in the search space to reduce GPU time and automatically learn and optimise complex neural architectures composed of convolutional operations and residual blocks. Despite being learned quickly with little human effort, the resulting networks are competitive with the best performing systems reported in the literature. Some are also far less complex, containing 85% fewer parameters than a Res2Net competitor.

Fri-M-SS-1: OpenASR20 and Low Resource ASR Development

Room Lacina, 11:00–13:00, Friday 3 September 2021 Chairs: Srikanth Madikeri and Emily Prud'hommeaux

Introduction by the Session Chairs

OpenASR20: An Open Challenge for Automatic Speech Recognition of Conversational Telephone Speech in Low-Resource Languages

*Kay Peterson*¹, *Audrey Tong*¹, *Yan Yu*²; ¹*NIST, USA*; ²*Dakota Consulting, USA* Fri-M-SS-1-1, Time: 11:10

In 2020, the National Institute of Standards and Technology (NIST), in cooperation with the Intelligence Advanced Research Project Activity (IARPA), conducted an open challenge on automatic speech recognition (ASR) technology for low-resource languages on a challenging data type — conversational telephone speech. The OpenASR20 Challenge was offered for ten low-resource languages — Amharic,

NOTES

Cantonese, Guarani, Javanese, Kurmanji Kurdish, Mongolian, Pashto, Somali, Tamil, and Vietnamese. A total of nine teams from five countries fully participated, and 128 valid submissions were scored. This paper gives an overview of the challenge setup and procedures, as well as a summary of the results. The results show overall high word error rate (WER), with the best results on a severely constrained training data condition ranging from 0.4 to 0.65, depending on the language. ASR with such limited resources remains a challenging problem. Providing a computing platform may be a way to level the playing field and encourage wider participation in challenges like OpenASR.

Buffer / Break

Time: 11:25

Multitask Adaptation with Lattice-Free MMI for Multi-Genre Speech Recognition of Low Resource Languages

Srikanth Madikeri, Petr Motlicek, Hervé Bourlard; Idiap Research Institute, Switzerland Fri-M-SS-1-2, Time: 11:30

In this paper, we develop Automatic Speech Recognition (ASR) systems for multi-genre speech recognition of low-resource languages where training data is predominantly conversational speech but test data can be in one of the following genres: news broadcast, topical broadcast and conversational speech. ASR for low-resource languages is often developed by adapting a pre-trained model to a target language. When training data is predominantly from one genre and limited, the system's performance for other genres suffer. To handle such out-of-domain scenarios, we employ multitask adaptation by using auxiliary conversational speech data from other languages in addition to the target-language data. We aim to (1) improve adaptation through implicit data augmentation by adding other languages as auxiliary tasks, and (2) prevent the acoustic model from overfitting to the dominant genre in the training set. Pre-trained parameters are obtained from a multilingual model trained with data from 18 languages using the Lattice-Free Maximum Mutual Information (LF-MMI) criterion. The adaptation is performed with the LF-MMI criterion. We present results on MATERIAL datasets for three languages: Kazakh and Farsi and Pashto.

An Improved Wav2Vec 2.0 Pre-Training Approach Using Enhanced Local Dependency Modeling for Speech Recognition

*Qiu-shi Zhu*¹, *Jie Zhang*¹, *Ming-hui Wu*², *Xin Fang*¹, *Li-Rong Dai*¹; ¹*USTC*, *China*; ²*iFLYTEK*, *China* Fri-M-SS-1-3, Time: 11:30

wav2vec 2.0 is a recently proposed self-supervised pre-training framework for learning speech representation. It utilizes a transformer to learn global contextual representation, which is effective especially in low-resource scenarios. Besides, it was shown that combining convolution neural network and transformer to model both local and global dependencies is beneficial for e.g., automatic speech recognition (ASR), natural language processing (NLP). However, how to model the local and global dependence in pre-training models is still an open question in the speech domain. In this paper, we therefore propose a new transformer encoder for enhancing the local dependency by combining convolution and self-attention modules. The transformer encoder first parallels the convolution and self-attention modules, and then serialized with another convolution module, sandwiched by a pair of feed forward modules. Experimental results show that the pre-trained model using the proposed method can reduce the word error rate (WER) compared to the reproduced wav2vec 2.0 at the cost of slightly increasing the size of training parameters.

Systems for Low-Resource Speech Recognition Tasks in Open Automatic Speech Recognition and Formosa Speech Recognition Challenges

Hung-Pang Lin, Yu-Jia Zhang, Chia-Ping Chen; National Sun Yat-sen University, Taiwan Fri-M-SS-1-4, Time: 11:30

We, in the team name of NSYSU-MITLab, have participated in low-resource speech recognition of the Open Automatic Speech Recognition Challenge 2020 (OpenASR20) and Formosa Speech Recognition Challenge 2020 (FSR-2020). For the tasks in the challenges, we build and compare end-to-end (E2E) systems and Deep Neural Network Hidden Markov Model (DNN-HMM) systems. In E2E systems, we implement an encoder with Conformer architecture and a decoder with Transformer architecture. In addition, a speaker classifier with a gradient reversal layer is included in the training phase to improve the robustness to speaker variation. In DNN-HMM systems, we implement the Time-Restricted Self-Attention and Factorized Time Delay Neural Networks for the DNN front-end acoustic representation learning. In OpenASR20, the best word error rates we achieved are 61.45% for Cantonese and 74.61% for Vietnamese. In FSR-2020, the best character error rate we achieved is 43.4% for Taiwanese Southern Min Recommended Characters and the best syllable error rate is 25.4% for Taiwan Minnanyu Luomazi Pinyin.

The TNT Team System Descriptions of Cantonese and Mongolian for IARPA OpenASR20

Jing Zhao¹, Zhiqiang Lv², Ambyera Han², Guan-Bo Wang¹, Guixin Shi¹, Jian Kang², Jinghao Yan², Pengfei Hu², Shen Huang², Wei-Qiang Zhang¹; ¹Tsinghua University, China; ²Tencent, China Fri-M-SS-1-5, Time: 12:10

This paper presents our work for OpenASR20 Challenge. We describe our Automatic Speech Recognition (ASR) systems for Cantonese and Mongolian under both constrained and unconstrained conditions. For constrained condition, a hybrid NN-HMM ASR system play the main role, while for unconstrained condition, an end-to-end ASR system outperforms traditional hybrid systems significantly due to adequate training data. Besides, we adapt to the challenging PSTN conditions using publicly available wideband dictated speech with similar accent, respectively for the two languages. Furthermore, data cleanup, language tailored features, multi-band training, data augmentation, pre-training and system fusions are incorporated. Our submitted systems have achieved excellent performances for the two conditions.

Combining Hybrid and End-to-End Approaches for the OpenASR20 Challenge

Tanel Alumäe, Jiaming Kong; Tallinn University of Technology, Estonia Fri-M-SS-1-6, Time: 12:10

This paper describes the TalTech team submission to the OpenASR20 Challenge. OpenASR20 evaluated low-resource speech recognition technologies across 10 languages, using only 10 hours of training data in the constrained condition. Our ASR systems used hybrid CNN-TDNNF-based acoustic models, trained with different data augmentation strategies. We used language model adaptation, recurrent neural network language models and lattice combination for improving first pass results. The scores of our submissions were

NOTES

the best across all teams in six out of ten languages. The paper also describes post-evaluation experiments that focused on the unconstrained condition. We show that optimized N-best list combination of a CNN-TDNNF based system and a finetuned multilingual XLSR-53 model results in large reductions in word error rate. Using BABEL data and the combination of hybrid and end-to-end systems gives 12–22% relative improvement over the constrained condition results.

One Size Does Not Fit All in Resource-Constrained ASR

Ethan Morris¹, Robbie Jimerson¹, Emily Prud'hommeaux²; ¹Rochester Institute of Technology, USA; ²Boston College, USA Fri-M-SS-1-7, Time: 12:10

The application of deep neural networks to the task of acoustic modeling for automatic speech recognition has resulted in dramatic decreases in ASR word error rates, enabling the use of this technology for interacting with smart phones and personal home assistants in high-resource languages. Developing ASR models of this caliber, however, requires hundreds or thousands of hours of transcribed speech recordings, which presents challenges for the vast majority of the world's languages. In this paper, we investigate the utility of three distinct architectures that have previously been used for ASR in languages with limited training resources. We train and test these systems on publicly available ASR datasets for several typologically and orthographically diverse languages, which were produced under a variety of conditions using different speech collection strategies, practices, and equipment. Although these corpora are comparable in size, we find that no single ASR architecture outperforms all others. In addition, word error rates vary significantly, in some cases within the range of those typically reported for high-resource languages. Our results point to the importance of considering language-specific and corpus-specific factors and experimenting with multiple approaches when developing ASR systems for languages with limited training resources.

Buffer / Break

Time: 12:50

Concluding Remarks by the Session Chairs Time: 12:55

Fri-Survey: Survey Talk 4: Alejandrina Cristia Room A+B, 13:00–14:00, Friday 3 September 2021 Chairs: TBD

Child Language Acquisition Studied with Wearables

Alejandrina Cristia; LSCP (UMR 8554), France Fri-Survey, Time: 13:00

In recent years, the ease with which we can collect audio (and to a lesser extent visual information) with wearables has improved dramatically. These allow unprecedented access to the speech that children produce, and that which they year. Although many conclusions drawn from short observations seem to generalize to these naturalistic datasets, others appear questionable based on human annotations of data collected with wearables. Making the best of such recordings also requires unique tool development.

Fri-Keynote: Keynote 4: Tomáš Mikolov

Room A+B, 15:00–16:00, Friday 3 September 2021 Chairs: TBD

Language Modeling and Artificial Intelligence

Tomáš Mikolov; CIIRC CTU, Czechia

Fri-Keynote, Time: 15:00

Statistical language modeling has been labeled as an AI-complete problem by many famous researchers of the past. However, despite all the progress made in the last decade, it remains unclear how much progress towards truly intelligent language models we made.

In this talk, I will present my view on what has been accomplished so far, and what scientific challenges are still in front of us. We need to focus more on developing new mathematical models with certain properties, such as the ability to learn continually and without explicit supervision, generalize to novel tasks from limited amounts of data, and the ability to form non-trivial long-term memory. I will describe some of our attempts to develop such models within the framework of complex systems.

Fri-A-O-1: Voice Activity Detection

Room A+B, 16:00–18:00, Friday 3 September 2021 Chairs: John Hansen and Hynek Bořil

Unsupervised Representation Learning for Speech Activity Detection in the Fearless Steps Challenge 2021

Pablo Gimeno, Alfonso Ortega, Antonio Miguel, Eduardo Lleida; Universidad de Zaragoza, Spain Fri-A-O-1-1, Time: 16:00

In this paper, we describe the ViVoLab speech activity detection (SAD) system submitted to the Fearless Steps Challenge Phase III. This series of challenges have proposed a number of speech processing task dealing with audio from Apollo space missions over the last few years. The focus in this edition is set on the generalisation capabilities of the systems, with new evaluation data from different channels. Our proposed submission is based on the use of the unsupervised representation learning paradigm, seeking to obtain a new and more discriminative audio representation than traditional perceptual features such as log Mel-filterbank energies. These new features are used to train different variations of a convolutional recurrent neural network (CRNN). Experimental results show that features learned via unsupervised learning provide a much more robust representation, significantly reducing the mismatch observed between development and evaluation partition results. Obtained results largely outperform the organisation baseline, achieving a DCF metric of 2.98% on the evaluation set and ranking third among all the participant teams.

The Application of Learnable STRF Kernels to the 2021 Fearless Steps Phase-03 SAD Challenge

Tyler Vuong, Yangyang Xia, Richard M. Stern; Carnegie Mellon University, USA

Fri-A-O-1-2, Time: 16:20

We describe a deep-learning-based system developed for the Fearless Steps Phase-03 Speech Activity Detection (SAD) challenge. The system includes both learnable spectro-temporal receptive fields (STRFs) and unconstrained 2-dimensional convolutional kernels in the first layer. Experiments show that the inclusion of learnable

STRFs in the first layer increases the system's robustness to additive noise. Additionally, we found that utilizing SpecAugment during training improves generalization on unseen data. By incorporating these enhancements and others our system achieved the best score in the official SAD challenge.

Speech Activity Detection Based on Multilingual Speech Recognition System

Seyyed Saeed Sarfjoo, Srikanth Madikeri, Petr Motlicek; Idiap Research Institute, Switzerland Fri-A-O-1-3, Time: 16:40

To better model the contextual information and increase the generalization ability of the Speech Activity Detection (SAD) system, this paper leverages a multilingual Automatic Speech Recognition (ASR) system to perform SAD. Sequence-discriminative training of Acoustic Model (AM) using Lattice-Free Maximum Mutual Information (LF-MMI) loss function, effectively extracts the contextual information of the input acoustic frame. Multilingual AM training causes the robustness to noise and language variabilities. The index of maximum output posterior is considered as a frame-level speech/non-speech decision function. Majority voting and logistic regression are applied to fuse the language-dependent decisions. The multilingual ASR is trained on 18 languages of BABEL datasets and the built SAD is evaluated on 3 different languages. On out-of-domain datasets, the proposed SAD model shows significantly better performance with respect to baseline models. On the Ester2 dataset, without using any in-domain data, this model outperforms the WebRTC, phoneme recognizer based VAD (Phn_Rec), and Pyannote baselines (respectively by 7.1, 1.7, and 2.7% absolute) in Detection Error Rate (DetER) metrics. Similarly, on the LiveATC dataset, this model outperforms the WebRTC, Phn_Rec, and Pyannote baselines (respectively by 6.4, 10.0, and 3.7% absolutely) in DetER metrics.

Voice Activity Detection with Teacher-Student Domain Emulation

Jarrod Luckenbaugh¹, Samuel Abplanalp², Rachel Gonzalez³, Daniel Fulford², David Gard³, Carlos Busso¹; ¹University of Texas at Dallas, USA; ²Boston University, USA; ³San Francisco State University, USA Fri-A-O-1-4, Time: 17:00

Transfer learning is a promising approach to increase performance for many speech-based systems, including voice activity detection (VAD). Domain adaptation, a subfield of transfer learning, often improves model conditioning in the presence of a mismatch between train-test conditions. This study proposes a formulation for VAD based on the teacher-student training, where the teacher model, trained with clean data, transfers knowledge to the student model trained with a noisy, paired version of the corpus resembling the test conditions. The models leverage temporal information using recurrent neural networks (RNN), implemented with either bidirectional long short term memory (BLSTM) or the modern, continuous-state Hopfield network. We provide evidence that in-domain noise emulation for domain adaptation is viable under unconstrained audio channel conditions for VAD "in the wild." Our application domain is in healthcare, where multimodal sensors, including microphones, from portable devices are used to automatically predict social isolation in patients affected by schizophrenia. We empirically show positive results for domain emulation when the training conditions are similar to the target domain. We also show that the Hopfield network outperforms our best BLSTM for VAD on real-world benchmarks.

EML Online Speech Activity Detection for the Fearless Steps Challenge Phase-III

Omid Ghahabi, Volker Fischer; EML Speech Technology, Germany

Fri-A-O-1-5, Time: 17:20

Speech Activity Detection (SAD), locating speech segments within an audio recording, is a main part of most speech technology applications. Robust SAD is usually more difficult in noisy conditions with varying signal-to-noise ratios (SNR). The Fearless Steps challenge has recently provided such data from the NASA Apollo-11 mission for different speech processing tasks including SAD. Most audio recordings are degraded by different kinds and levels of noise varying within and between channels. This paper describes the EML online algorithm for the most recent phase of this challenge. The proposed algorithm can be trained both in a supervised and unsupervised manner and assigns speech and non-speech labels at runtime approximately every 0.1 sec. The experimental results show a competitive accuracy on both development and evaluation datasets with a real-time factor of about 0.002 using a single CPU machine.

Fri-A-O-2: Keyword Search and Spoken Language Processing

Room D, 16:00–18:00, Friday 3 September 2021 Chairs: Jan Chorowski and Jean-Luc Gauvain

Device Playback Augmentation with Echo Cancellation for Keyword Spotting

Kuba Łopatka, Katarzyna Kaszuba-Miotke, Piotr Klinke, Paweł Trella; Intel, Poland Fri-A-O-2-1, Time: 16:00

Keyword spotting (KWS) is required to operate in device playback conditions in which the device itself plays interfering signals. We propose a new method to augment the training set and adapt the acoustic model to the playback environment. It is based on acoustic simulation which models the coupling between the device's loudspeakers and microphones. The employed model involves frequency response of the device, as well as room impulse response and nonlinear distortions introduced in the playback path. Finally, we pass the simulated signals through Acoustic Echo Cancellation (AEC) to model the artifacts introduced by AEC algorithm. The proposed method reduces False Rejection Rate in device playback noise by 25-60% for a Time-Delay Neural Network-based KWS engine. It is shown that the introduction of device characteristics and nonlinear filtration is necessary to achieve improvement in playback conditions. The augmentation scheme is highly independent of the architecture of the KWS system.

End-to-End Open Vocabulary Keyword Search

Bolaji Yusuf, Alican Gok, Batuhan Gundogdu, Murat Saraclar; Boğaziçi Üniversitesi, Turkey Fri-A-O-2-2, Time: 16:20

Recently, neural approaches to spoken content retrieval have become popular. However, they tend to be restricted in their vocabulary or in their ability to deal with imbalanced test settings. These restrictions limit their applicability in keyword search, where the set of queries is not known beforehand, and where the system should return not just whether an utterance contains a query but the exact location of any such occurrences. In this work, we propose a model directly optimized for keyword search. The model takes a query and an utterance as input and returns a sequence of probabilities for each frame of the utterance of the query having occurred in

NOTES

that frame. Experiments show that the proposed model not only outperforms similar end-to-end models on a task where the ratio of positive and negative trials is artificially balanced, but it is also able to deal with the far more challenging task of keyword search with its inherent imbalance. Furthermore, using our system to rescore the outputs an LVCSR-based keyword search system leads to significant improvements on the latter.

Semantic Sentence Similarity: Size does not Always Matter

Danny Merkx, Stefan L. Frank, Mirjam Ernestus; Radboud Universiteit, The Netherlands Fri-A-O-2-3, Time: 16:40

This study addresses the question whether visually grounded speech recognition (VGS) models learn to capture sentence semantics without access to any prior linguistic knowledge. We produce synthetic and natural spoken versions of a well known semantic textual similarity database and show that our VGS model produces embeddings that correlate well with human semantic similarity judgements. Our results show that a model trained on a small image-caption database outperforms two models trained on much larger databases, indicating that database size is not all that matters. We also investigate the importance of having multiple captions per image and find that this is indeed helpful even if the total number of images is lower, suggesting that paraphrasing is a valuable learning signal. While the general trend in the field is to create ever larger datasets to train models on, our findings indicate other characteristics of the database can just as important.

Spoken Term Detection and Relevance Score Estimation Using Dot-Product of Pronunciation Embeddings

Jan Švec, Luboš Šmídl, Josef V. Psutka, Aleš Pražák; University of West Bohemia, Czechia Fri-A-O-2-4, Time: 17:00

The paper describes a novel approach to Spoken Term Detection (STD) in large spoken archives using deep LSTM networks. The work is based on the previous approach of using Siamese neural networks for STD and naturally extends it to directly localize a spoken term and estimate its relevance score. The phoneme confusion network generated by a phoneme recognizer is processed by the deep LSTM network which projects each segment of the confusion network into an embedding space. The searched term is projected into the same embedding space using another deep LSTM network. The relevance score is then computed using a simple dot-product in the embedding space and calibrated using a sigmoid function to predict the probability of occurrence. The location of the searched term is then estimated from the sequence of output probabilities. The deep LSTM networks are trained in a self-supervised manner from paired recognition hypotheses on word and phoneme levels. The method is experimentally evaluated on MALACH data in English and Czech languages.

Toward Genre Adapted Closed Captioning

François Buet, François Yvon; LISN (UMR 9015), France Fri-A-O-2-5, Time: 17:20

This paper studies the generation of intralingual closed captions from automatic speech transcripts, with the aim to assess techniques for multi-genre captioning. Captions and subtitles greatly vary in form and content depending on the programs genres and subtitling styles, resulting for instance in significantly different compression rates and lexical content. Borrowing ideas from the multi-domain machine translation literature, we implement and contrast several adaptation methods on a diverse set of programs broadcast on the French public TV. Our results show that such multi-domain adaption techniques are effective and help to improve our automatic subtiling system.

Fri-A-V-1: Applications in Transcription, Education and Learning

16:00–18:00, Friday 3 September 2021 Chairs: Kate Knill and Michael Picheny

Weakly-Supervised Word-Level Pronunciation Error Detection in Non-Native English Speech

Daniel Korzekwa¹, Jaime Lorenzo-Trueba², Thomas Drugman², Shira Calamaro², Bozena Kostek³; ¹Amazon, Poland; ²Amazon, UK; ³Gdansk University of Technology, Poland Fri-A-V-1-1, Time: 16:00

We propose a weakly-supervised model for word-level mispronunciation detection in non-native (L2) English speech. To train this model, phonetically transcribed L2 speech is not required and we only need to mark mispronounced words. The lack of phonetic transcriptions for L2 speech means that the model has to learn only from a weak signal of word-level mispronunciations. Because of that and due to the limited amount of mispronounced L2 speech, the model is more likely to overfit. To limit this risk, we train it in a multi-task setup. In the first task, we estimate the probabilities of word-level mispronunciation. For the second task, we use a phoneme recognizer trained on phonetically transcribed L1 speech that is easily accessible and can be automatically annotated. Compared to state-of-the-art approaches, we improve the accuracy of detecting word-level pronunciation errors in AUC metric by 30% on the GUT Isle Corpus of L2 Polish speakers, and by 21.5% on the Isle Corpus of L2 German and Italian speakers.

End-to-End Speaker-Attributed ASR with Transformer

Naoyuki Kanda, Guoli Ye, Yashesh Gaur, Xiaofei Wang, Zhong Meng, Zhuo Chen, Takuya Yoshioka; Microsoft, USA

Fri-A-V-1-2, Time: 16:00

This paper presents our recent effort on end-to-end speakerattributed automatic speech recognition, which jointly performs speaker counting, speech recognition and speaker identification for monaural multi-talker audio. Firstly, we thoroughly update the model architecture that was previously designed based on a long short-term memory (LSTM)-based attention encoder decoder by applying transformer architectures. Secondly, we propose a speaker deduplication mechanism to reduce speaker identification errors in highly overlapped regions. Experimental results on the LibriSpeechMix dataset shows that the transformer-based architecture is especially good at counting the speakers and that the proposed model reduces the speaker-attributed word error rate by 47% over the LSTM-based baseline. Furthermore, for the LibriCSS dataset, which consists of real recordings of overlapped speech, the proposed model achieves concatenated minimum-permutation word error rates of 11.9% and 16.3% with and without target speaker profiles, respectively, both of which are the state-of-the-art results for LibriCSS with the monaural setting.

Understanding Medical Conversations: Rich Transcription, Confidence Scores & Information Extraction

Hagen Soltau, Mingqiu Wang, Izhak Shafran, Laurent El Shafey; Google, USA Fri-A-V-1-3, Time: 16:00

In this paper, we describe novel components for extracting clinically relevant information from medical conversations which will be available as Google APIs. We describe a transformer-based Recurrent Neural Network Transducer (RNN-T) model tailored for long-form audio, which can produce rich transcriptions including speaker segmentation, speaker role labeling, punctuation and capitalization. On a representative test set, we compare performance of RNN-T models with different encoders, units and streaming constraints. Our transformer-based streaming model performs at about 20% WER on the ASR task, 6% WDER on the diarization task, 43% SER on periods, 52% SER on commas, 43% SER on question marks and 30% SER on capitalization. Our recognizer is paired with a confidence model that utilizes both acoustic and lexical features from the recognizer. The model performs at about 0.37 NCE. Finally, we describe a RNN-T based tagging model. The performance of the model depends on the ontologies, with F-scores of 0.90 for medications, 0.76 for symptoms, 0.75 for conditions, 0.76 for diagnosis, and 0.61 for treatments. While there is still room for improvement, our results suggest that these models are sufficiently accurate for practical applications.

Phone-Level Pronunciation Scoring for Spanish Speakers Learning English Using a GOP-DNN System

*Jazmín Vidal*¹, *Cyntia Bonomi*¹, *Marcelo Sancinetti*¹, *Luciana Ferrer*²; ¹*UBA*, *Argentina*; ²*UBA-CONICET ICC*, *Argentina*

Fri-A-V-1-4, Time: 16:00

In today's globalized world being able to communicate in English is crucial to many people. Computer assisted pronunciation training (CAPT) systems can help students achieve English proficiency by providing an accessible way to practice, offering personalized feedback. However, phone-level pronunciation scoring is still a very challenging task, with performance far from that of human annotators. In this paper we compare and present results on the Spanish subset of the L2-ARCTIC corpus and the new Epa-DB database, both containing non-native English speech by native Spanish speakers and intended for the development of pronunciation scoring systems. We show the most frequent errors in each database and compare performance of a state-of-the-art goodness of pronunciation (GOP) system. Results show that both databases have similar error patterns and that performance is similar for most phones, despite differences in recording conditions. For the EpaDB database we also present an analysis of the errors per target phone. This study validates the EpaDB collection and annotations, providing initial results and contributing to the advancement of a challenging low-resource task.

Explore wav2vec 2.0 for Mispronunciation Detection

Xiaoshuo Xu, Yueteng Kang, Songjun Cao, Binghuai Lin, Long Ma; Tencent, China Fri-A-V-1-5, Time: 16:00

This paper presents an initial attempt to use self-supervised learning for Mispronunciation Detection. Unlike existing methods that use speech recognition corpus to train models, we exploit unlabeled data and utilize a self-supervised learning technique, Wav2vec 2.0, for pretraining. After the pretraining process, the training process only requires a little pronunciation-labeled data for finetuning. Formulating Mispronunciation Detection as a binary classification task, we

add convolutional and pooling layers on the top of the pretrained model to detect mispronunciations of the given prompted texts within the alignment segmentations. The training process is simple and effective. Several experiments are conducted to validate the effectiveness of the pretrained method. Our approach outperforms existing methods on a public dataset L2-ARCTIC with a F1 value of 0.610.

Lexical Density Analysis of Word Productions in Japanese English Using Acoustic Word Embeddings

Shintaro Ando, Nobuaki Minematsu, Daisuke Saito; University of Tokyo, Japan Fri-A-V-1-6, Time: 16:00

In L2 pronunciation, what kind of phonetic errors are more influential to intelligibility reduction? Teachers say that learners' utterances become unintelligible when words are pronounced with such errors that make the words misidentified as others. In this paper, we focus on Japanese English (JE), where the number of phonemes of the L1 (Japanese) is much smaller than that of the L2 (American English, AE). Since learners often substitute L1 phonemes when speaking in L2, some words are expected to be pronounced not distinctively enough in JE, which may result in word misidentification. This implies that words of JE will exist phonetically closer to each other in a space where words are distributed. In this paper, lexical density analysis of JE and AE is carried out using acoustic word embeddings. Word productions in JE and AE, extracted from the ERJ corpus, are mapped as points in an acoustic word embedding space obtained by network training with the WSJ corpus. Experiments show that significantly higher density is found in JE than in AE and it is also found in poor learners than in good learners.

Deep Feature Transfer Learning for Automatic Pronunciation Assessment

Binghuai Lin, Liyuan Wang; Tencent, China Fri-A-V-1-7, Time: 16:00

Automatic pronunciation assessment is commonly developed to evaluate pronunciation quality of second language (L2) learners. Traditional methods for automatic pronunciation assessment normally utilize speech features such as Goodness of pronunciation (GOP), which may not provide sufficient information for the pronunciation proficiency assessment [1]. In this paper, we propose a transfer learning method for automatic pronunciation assessment. We directly utilize the deep features from the acoustic model instead of traditional features such as GOP, and transfer the acoustic knowledge from ASR to a specific scoring module. The scoring module is designed to consider the relationship among different granularities in an utterance based on an attention mechanism. Only this module is updated for faster transfer and adaptation of various pronunciation assessment tasks. Experimental results based on the dataset recorded by Chinese English-as-second-language (ESL) learners and the Speechocean762 dataset demonstrate that the proposed method outperforms the traditional GOP-based baselines in Pearson correlation coefficient (PCC) and yields parameter-efficient transfer for different pronunciation assessment tasks.

Multilingual Speech Evaluation: Case Studies on English, Malay and Tamil

Huayun Zhang, Ke Shi, Nancy F. Chen; A*STAR, Singapore

Fri-A-V-1-8, Time: 16:00

Speech evaluation is an essential component in computer-assisted language learning (CALL). While speech evaluation on English has been popular, automatic speech scoring on low resource languages

NOTES

remains challenging. Work in this area has focused on monolingual specific designs and handcrafted features stemming from resource-rich languages like English. Such approaches are often difficult to generalize to other languages, especially if we also want to consider suprasegmental qualities such as rhythm. In this work, we examine three different languages that possess distinct rhythm patterns: English (stress-timed), Malay (syllable-timed), and Tamil (mora-timed). We exploit robust feature representations inspired by music processing and vector representation learning. Empirical validations show consistent gains for all three languages when predicting pronunciation, rhythm and intonation performance.

A Study on Fine-Tuning wav2vec2.0 Model for the Task of Mispronunciation Detection and Diagnosis

Linkai Peng¹, Kaiqi Fu¹, Binghuai Lin², Dengfeng Ke¹, Jinsong Zhan¹; ¹BLCU, China; ²Tencent, China Fri-A-V-1-9, Time: 16:00

Mispronunciation detection and diagnosis (MDD) technology is a key component of computer-assisted pronunciation training system (CAPT). The mainstream method is based on deep neural network automatic speech recognition. Unfortunately, the technique requires massive human-annotated speech recordings for training. Due to the huge variations in mother tongue, age, and proficiency level among second language learners, it is difficult to gather a large amount of matching data for acoustic model training, which greatly limits the model performance. In this paper, we explore the use of Self-Supervised Pretraining (SSP) model wav2vec2.0 for MDD tasks. SSP utilizes a large unlabelled dataset to learn general representation and can be applied in downstream tasks. We conduct experiments using two publicly available datasets (TIMIT, L2-arctic) and our best system achieves 60.44% f1-score. Moreover, our method is able to achieve 55.52% f1-score with 3 times less data, which demonstrates the effectiveness of SSP on MDD.

The Impact of ASR on the Automatic Analysis of Linguistic Complexity and Sophistication in Spontaneous L2 Speech

Yu Qiao, Wei Zhou, Elma Kerz, Ralf Schlüter; RWTH Aachen University, Germany Fri-A-V-1-10, Time: 16:00

In recent years, automated approaches to assessing linguistic complexity in second language (L2) writing have made significant progress in gauging learner performance, predicting human ratings of the quality of learner productions, and benchmarking L2 development. In contrast, there is comparatively little work in the area of speaking, particularly with respect to fully automated approaches to assessing L2 spontaneous speech. While the importance of a well-performing ASR system is widely recognized, little research has been conducted to investigate the impact of its performance on subsequent automatic text analysis. In this paper, we focus on this issue and examine the impact of using a state-of-the-art ASR system for subsequent automatic analysis of linguistic complexity in spontaneously produced L2 speech. A set of 30 selected measures were considered, falling into four categories: syntactic, lexical, n-gram frequency, and information-theoretic measures. The agreement between the scores for these measures obtained on the basis of ASR-generated vs. manual transcriptions was determined through correlation analysis. A more differential effect of ASR performance on specific types of complexity measures when controlling for task type effects is also presented.

End-to-End Rich Transcription-Style Automatic Speech Recognition with Semi-Supervised Learning

Tomohiro Tanaka, Rvo Masumura, Mana Ihori, Akihiko Takashima, Shota Orihashi, Naoki Makishima; NTT, Japan

Fri-A-V-1-11, Time: 16:00

We propose a semi-supervised learning method for building end-toend rich transcription-style automatic speech recognition (RT-ASR) systems from small-scale rich transcription-style and large-scale common transcription-style datasets. In spontaneous speech tasks. various speech phenomena such as fillers, word fragments, laughter and coughs, etc. are often included. While common transcriptions do not give special awareness to these phenomena, rich transcriptions explicitly convert them into special phenomenon tokens as well as textual tokens. In previous studies, the textual and phenomenon tokens were simultaneously estimated in an end-to-end manner. However, it is difficult to build accurate RT-ASR systems because large-scale rich transcription-style datasets are often unavailable. To solve this problem, our training method uses a limited rich transcription-style dataset and common transcription-style dataset simultaneously. The Key process in our semi-supervised learning is to convert the common transcription-style dataset into a pseudorich transcription-style dataset. To this end, we introduce style tokens which control phenomenon tokens are generated or not into transformer-based autoregressive modeling. We use this modeling for generating the pseudo-rich transcription-style datasets and for building RT-ASR system from the pseudo and original datasets. Our experiments on spontaneous ASR tasks showed the effectiveness of the proposed method.

"You don't understand me!": Comparing ASR Results for L1 and L2 Speakers of Swedish

Ronald Cumbal¹, Birger Moell¹, José Lopes², Olov Engwall¹; ¹*KTH*, Sweden; ²*Heriot-Watt University*, UK Fri-A-V-1-12. Time: 16:00

The performance of Automatic Speech Recognition (ASR) systems has constantly increased in state-of-the-art development. However, performance tends to decrease considerably in more challenging conditions (e.g., background noise, multiple speaker social conversations) and with more atypical speakers (e.g., children, non-native speakers or people with speech disorders), which signifies that general improvements do not necessarily transfer to applications that rely on ASR, e.g., educational software for younger students or language learners. In this study, we focus on the gap in performance between recognition results for native and non-native, read and spontaneous, Swedish utterances transcribed by different ASR services. We compare the recognition results using Word Error Rate and analyze the linguistic factors that may generate the observed transcription errors.

NeMo Inverse Text Normalization: From **Development to Production**

Yang Zhang¹, Evelina Bakhturina¹, Kyle Gorman², Boris Ginsburg¹; ¹NVIDIA, USA; ²CUNY Graduate Center, USA Fri-A-V-1-13, Time: 16:00

Inverse text normalization (ITN) converts spoken-domain automatic speech recognition (ASR) output into written-domain text to improve the readability of the ASR output. Many state-of-the-art ITN systems use hand-written weighted finite-state transducer (WFST) grammars since this task has extremely low tolerance to unrecoverable errors. We introduce an open-source Python WFST-based library for ITN which enables a seamless path from development to production. We

describe the specification of ITN grammar rules for English, but the library can be adapted for other languages. It can also be used for written-to-spoken text normalization. We evaluate the NeMo ITN library using a modified version of the Google Text normalization dataset.

Improvement of Automatic English Pronunciation Assessment with Small Number of Utterances Using Sentence Speakability

Satsuki Naijo, Akinori Ito, Takashi Nose; Tohoku University, Japan Fri-A-V-1-14, Time: 16:00

The current Computer-Assisted Pronunciation Training (CAPT) system uses DNN-based speech recognition results to evaluate learner's pronunciation with high accuracy when using many utterances for the evaluation. However, when we use only a few utterances. the accuracy of the CAPT system deteriorates. One reason for the deterioration is that the score calculated by a CAPT system is biased depending on the pronunciation difficulty of the sentences when using a small number of utterances. In this study, we developed a CAPT system that takes the sentence speakability (pronunciation difficulty of sentences) into account. As a result, the correlation coefficient between the human evaluation and the machine score was 0.46 in the conventional method, while it improved to 0.57 with the proposed method.

Fri-A-V-2: Emotion and Sentiment Analysis III

16:00-18:00, Friday 3 September 2021 Chairs: Chi-Chun Lee and Carol Espy-Wilson

Affect Recognition Through Scalogram and **Multi-Resolution Cochleagram Features**

Fasih Haider, Saturnino Luz; University of Edinburgh, UK Fri-A-V-2-1, Time: 16:00

An approach to the categorization of voice samples according to emotions expressed by the speaker is proposed which uses Multi-Resolution Cochleagram (MRCG) and scalogram features in a novel way. Audio recordings from the EmoDB, EMOVO and Savee Data-sets are employed in training and testing of predictive models consisting of different sets of speech features. This study systematically evaluates the performance of the feature sets most commonly used in computational paralinguistic tasks (i.e. emobase, eGeMAPS and ComParE) in addition to MRCG- and scalogram-derived features and their fusion, across five different classifiers. The datasets used in this evaluation include speech in three different languages (German, Italian and English). MRCG features outperform the feature sets most commonly used in computational paralinguistic tasks, including emobase, eGeMAPS and ComParE, for the EmoDB (unweighted average recall, UAR = 59.15%) and SAVEE (UAR = 36.12%) datasets, while eGeMAPS provides the best overall UAR (33.84%) for the EMOVO dataset. A support vector machine (SVM) classifier yields the best UAR for EmoDB (80.05%) through fusion of emobase, eGeMAPS, ComParE and MRCG, and for EMOVO (40.31%), through fusion of emobase, eGeMAPS and ComParE. For SAVEE, random forests provide the best result (46.55%) using the *ComParE* feature set.

A Speech Emotion Recognition Framework for Better **Discrimination of Confusions**

Jiawang Liu, Haoxiang Wang; SCUT, China Fri-A-V-2-2, Time: 16:00

Speech emotion recognition (SER) plays an important role in humanmachine interaction (HMI). Various methods have been proposed for

the SER task. However, a common problem in most of the previous studies is some specific emotions are grossly misclassified. In this paper, we propose a novel SER framework aiming at discriminating the confusions by utilizing triplet loss and data augmentation to enforce a CNN-LSTM model to emphasize more on these emotions which are hard to be correctly classified. Ablation experiments demonstrate the effectiveness of the proposed framework. On Interactive Emotional Dyadic Motion Capture (IEMOCAP) dataset, our framework can achieve 79.52% of Weighted Accuracy (WA) and 78.30% of Unweighted Accuracy (UA). Compared to the other state-of-the-art models, our framework obtains more than 3.34% and 1.94% improvement on WA and UA respectively.

Speech Emotion Recognition via Multi-Level Cross-Modal Distillation

Ruichen Li, Jinming Zhao, Qin Jin; Renmin University of China, China

Fri-A-V-2-3, Time: 16:00

Speech emotion recognition faces the problem that most of the existing speech corpora are limited in scale and diversity due to the high annotation cost and label ambiguity. In this work, we explore the task of learning robust speech emotion representations based on large unlabeled speech data. Under a simple assumption that the internal emotional states across different modalities are similar, we propose a method called Multi-level Cross-modal Emotion Distillation (MCED), which trains the speech emotion model Emotion any labeled speech emotion data by transferring emotion knowledge from a pretrained text emotion model. Extensive experiments on two benchmark datasets, IEMOCAP and MELD, show that our proposed MCED can help learn effective speech emotion recognition tasks.

Audio-Visual Speech Emotion Recognition by Disentangling Emotion and Identity Attributes

Koichiro Ito, Takuya Fujioka, Qinghua Sun, Kenji Nagamatsu; Hitachi, Japan

Fri-A-V-2-4, Time: 16:00

In this paper, we propose an audio-visual speech emotion recognition (AV-SER) that can suppress the disturbance from an identity attribute by disentangling an emotion attribute and an identity one. We developed a model that first disentangles both attributes for each modality. In order to achieve the disentanglement, we introduce a coattention module to our model. Our model disentangles the emotion attribute by giving the identity attribute as conditional features to the module. Conversely, the identity attribute is also obtained with the emotion attribute as a condition. Our model then makes a prediction for each attribute from these disentangled features by considering both modalities. In addition, to ensure the disentanglement capacity of our model, we train the model with an identification task as the auxiliary task and an SER task as the primary task alternately, and we update only the part of parameters responsible for each task. The experimental result shows the effectiveness of our method with the wild CMU-MOSEI dataset.

Parametric Distributions to Model Numerical Emotion Labels

Deboshree Bose, Vidhyasaharan Sethu, Eliathamby Ambikairajah; UNSW Sydney, Australia Fri-A-V-2-5, Time: 16:00

It is common to represent emotional states as values on a set of numerical scales corresponding to attributes such as arousal and valence. Often these labels are obtained from multiple annotators who record their perception of emotion in terms of these attributes. Combining these multiple annotations by taking the mean, as is typical in affective computing systems ignores the inherent ambiguity in the labels. Recently it has been recognised that this ambiguity carries useful information and systems that employ distributions over the numerical scales to represent emotional states have been proposed. In this paper we show that the common and widespread assumption that this distribution is Gaussian may not be suitable since the underlying numerical scales are bounded. We then compare a range of well-known distributions defined on bounded domains to ascertain which of them would be the most suitable alternative. Statistical measures are proposed to enable quantifiable comparisons and the results are reported. All comparisons reported in the paper were carried out on the RECOLA dataset.

Metric Learning Based Feature Representation with Gated Fusion Model for Speech Emotion Recognition

Yuan Gao, Jiaxing Liu, Longbiao Wang, Jianwu Dang; Tianjin University, China Fri-A-V-2-6, Time: 16:00

Due to the lack of sufficient speech emotional data, the recognition performance of existing speech emotion recognition (SER) approaches is relatively low and requires further improvement to meet the needs of real-life applications. For the problem of data scarcity, an increasingly popular solution is to transfer emotional information through pre-training models and extract additional features. However, the feature representation needs further compression because the training object of unsupervised learning is to reconstruct input, making the latent representation contain non-affective information. In this paper, we introduce deep metric learning to constrain the feature distribution of the pre-training model. Specifically, we propose a triplet loss to modify the representation extraction model as a pseudo-siamese network and achieve more efficient knowledge transfer for emotion recognition. Furthermore, we propose a gated fusion method to learn the connection of features extracted from the pre-training model and supervised feature extraction model. We conduct experiments on the common benchmarking dataset IEMOCAP to verify the performance of the proposed model. The experimental results demonstrate the advantages of our model, outperforming the unsupervised transfer learning system by 3.7% and 3.88% in weighted accuracy and unweighted accuracy, respectively.

Speech Emotion Recognition with Multi-Task Learning

Xingyu Cai, Jiahong Yuan, Renjie Zheng, Liang Huang, Kenneth Church; Baidu, USA Fri-A-V-2-7, Time: 16:00

Speech emotion recognition (SER) classifies speech into emotion categories such as: *Happy, Angry, Sad* and *Neutral*. Recently, deep learning has been applied to the SER task. This paper proposes a multi-task learning (MTL) framework to simultaneously perform speech-to-text recognition and emotion classification, with an end-to-end deep neural model based on wav2vec-2.0. Experiments on the IEMOCAP benchmark show that the proposed method achieves the state-of-the-art performance on the SER task. In addition, an ablation study establishes the effectiveness of the proposed MTL framework.

Generalized Dilated CNN Models for Depression Detection Using Inverted Vocal Tract Variables

Nadee Seneviratne, Carol Espy-Wilson; University of Maryland at College Park, USA Fri-A-V-2-8, Time: 16:00

Depression detection using vocal biomarkers is a highly researched area. Articulatory coordination features (ACFs) are developed based on the changes in neuromotor coordination due to psychomotor slowing, a key feature of Major Depressive Disorder. However

findings of existing studies are mostly validated on a single database which limits the generalizability of results. Variability across different depression databases adversely affects the results in cross corpus evaluations (CCEs). We propose to develop a generalized classifier for depression detection using a dilated Convolutional Neural Network which is trained on ACFs extracted from two depression databases. We show that ACFs derived from Vocal Tract Variables (TVs) show promise as a robust set of features for depression detection. Our model achieves relative accuracy improvements of $\sim 10\%$ compared to CCEs performed on models trained on a single database. We extend the study to show that fusing TVs and Mel-Frequency Cepstral Coefficients can further improve the performance of this classifier.

Learning Mutual Correlation in Multimodal Transformer for Speech Emotion Recognition

Yuhua Wang, Guang Shen, Yuezhu Xu, Jiahang Li, Zhengdao Zhao; Harbin Engineering University, China Fri-A-V-2-9, Time: 16:00

Various studies have confirmed the necessity and benefits of leveraging multimodal features for SER, and the latest research results show that the temporal information captured by the transformer is very useful for improving multimodal speech emotion recognition. However, the dependency between different modalities and high-level temporal-feature learning using a deeper transformer is yet to be investigated. Thus, we propose a multimodal transformer with sharing weights for speech emotion recognition. The proposed network shares the weights across the modalities in each transformer layer to learn the correlation among multiple modalities. In addition, since the emotion contained in a speech generally include audio and text features, both of which have not only internal dependence but also mutual dependence, we design a deep multimodal attention mechanism to capture these two kinds of emotional dependence. We evaluated our model on the publicly available IEMOCAP dataset. The experimental results demonstrate that the proposed model yielded a promising result.

Time-Frequency Representation Learning with Graph Convolutional Network for Dialogue-Level Speech Emotion Recognition

Jiaxing Liu, Yaodong Song, Longbiao Wang, Jianwu Dang, Ruiguo Yu; Tianjin University, China Fri-A-V-2-10, Time: 16:00

With the development of speech emotion recognition (SER), dialoguelevel SER (DSER) is more aligned with actual scenarios. In this paper, we propose a DSER approach that includes two stages of representation learning: intra-utterance representation learning and inter-utterance representation learning. In the intra-utterance representation learning stage, traditional convolutional neural network (CNN) has demonstrated great success. However, the basic design of a CNN restricts its ability to model the local and global information in the spectrogram. Therefore, we propose a novel local-global representation learning method for the intra-utterance stage. The local information is learned by a time-frequency convolutional neural network (TFCNN), which we published previously. Here, we propose a time-frequency capsule neural network (TFCap) to model global information that can extract more stable global time-frequency information directly from spectrograms. In the inter-utterance stage, a graph convolutional network (GCN) is introduced to explore the relations between utterances in a dialog. Our proposed methods were evaluated on the IEMOCAP database. The proposed time-frequency based method in the intra-utterance stage achieves an absolute increase of 9.35% compared to CNN. By integrating GCN in the inter-utterance stage, the proposed approach achieves an absolute increase of 4.05% compared to the model in the previous stage.

NOTES

Fri-A-V-3: Resource-Constrained ASR

16:00–18:00, Friday 3 September 2021 Chairs: Liang Lu and Yanzhang He

Compressing 1D Time-Channel Separable Convolutions Using Sparse Random Ternary Matrices

*Gonçalo Mordido*¹, *Matthijs Van keirsbilck*², *Alexander Keller*²; ¹*HPI, Germany*; ²*NVIDIA, Germany* Fri-A-V-3-1, Time: 16:00

We demonstrate that 1×1 -convolutions in 1D time-channel separable convolutions may be replaced by constant, sparse random ternary matrices with weights in {-1, 0, +1}. Such layers do not perform any multiplications and do not require training. Moreover, the matrices may be generated on the chip during computation and therefore do not require any memory access. With the same parameter budget, we can afford deeper and more expressive models, improving the Pareto frontiers of existing models on several tasks. For command recognition on Google Speech Commands v1, we improve the stateof-the-art accuracy from 97.21% to 97.41% at the same network size. Alternatively, we can lower the cost of existing models. For speech recognition on Librispeech, we halve the number of weights to be trained while only sacrificing about 1% of the floating-point baseline's word error rate.

Weakly Supervised Construction of ASR Systems from Massive Video Data

Mengli Cheng, Chengyu Wang, Jun Huang, Xiaobo Wang; Alibaba, China

Fri-A-V-3-2, Time: 16:00

Despite the rapid development of deep learning models, for real-world applications, building large-scale Automatic Speech Recognition (ASR) systems from scratch is still significantly challenging, mostly due to the time-consuming and financially-expensive process of annotating a large amount of audio data with transcripts. Although several self-supervised pre-training models have been proposed to learn speech representations, applying such models directly might be sub-optimal if more labeled, training data could be obtained without a large cost.

In this paper, we present VideoASR, a weakly supervised framework for constructing ASR systems from massive video data. As usergenerated videos often contain human-speech audio roughly aligned with subtitles, we consider videos as an important knowledge source, and propose an effective approach to extract high-quality audio aligned with transcripts from videos based on text detection and Optical Character Recognition. The underlying ASR models can be fine-tuned to fit any domain-specific target training datasets after weakly supervised pre-training on automatically generated datasets. Extensive experiments show that VideoASR can easily produce state-of-the-art results on six public datasets for Mandarin speech recognition. In addition, the VideoASR framework has been deployed on the cloud to support various industrial-scale applications.

Broadcasted Residual Learning for Efficient Keyword Spotting

Byeonggeun Kim, Simyung Chang, Jinkyu Lee, Dooyong Sung; Qualcomm, Korea

Fri-A-V-3-3, Time: 16:00

Keyword spotting is an important research field because it plays a key role in device wake-up and user interaction on smart devices. However, it is challenging to minimize errors while operating efficiently in devices with limited resources such as mobile phones. We present a *broadcasted residual learning* method to achieve high accuracy with small model size and computational load. Our method configures most of the residual functions as 1D temporal convolution while still allows 2D convolution together using a broadcasted-residual connection that expands temporal output to frequency-temporal dimension. This residual mapping enables the network to effectively represent useful audio features with much less computation than conventional convolutional neural networks. We also propose a novel network architecture, Broadcasting-residual network (BC-ResNet), based on broadcasted residual learning and describe how to scale up the model according to the target device's resources. BC-ResNets achieve state-of-the-art 98.0% and 98.7% top-1 accuracy on Google speech command datasets v1 and v2, respectively, and consistently outperform previous approaches, using fewer computations and parameters.

CoDERT: Distilling Encoder Representations with Co-Learning for Transducer-Based Speech Recognition

Rupak Vignesh Swaminathan, Brian King, Grant P. Strimel, Jasha Droppo, Athanasios Mouchtaris; Amazon, USA

Fri-A-V-3-4, Time: 16:00

We propose a simple yet effective method to compress an RNN-Transducer (RNN-T) through the well-known knowledge distillation paradigm. We show that the transducer's encoder outputs naturally have a high entropy and contain rich information about acoustically similar word-piece confusions. This rich information is suppressed when combined with the lower entropy decoder outputs to produce the joint network logits. Consequently, we introduce an auxiliary loss to distill the encoder logits from a teacher transducer's encoder, and explore training strategies where this encoder distillation works effectively. We find that tandem training of teacher and student encoders with an inplace encoder distillation outperforms the use of a pre-trained and static teacher transducer. We also report an interesting phenomenon we refer to as implicit distillation, that occurs when the teacher and student encoders share the same decoder. Our experiments show 5.37-8.4% relative word error rate reductions (WERR) on in-house test sets, and 5.05-6.18% relative WERRs on LibriSpeech test sets.

Extremely Low Footprint End-to-End ASR System for Smart Device

*Zhifu Gao*¹, *Yiwu Yao*¹, *Shiliang Zhang*¹, *Jun Yang*¹, *Ming Lei*¹, *Ian McLoughlin*²; ¹*Alibaba*, *China*; ²*SIT*, *Singapore*

Fri-A-V-3-5, Time: 16:00

Recently, end-to-end (E2E) speech recognition has become popular, since it can integrate the acoustic, pronunciation and language models into a single neural network, which outperforms conventional models. Among E2E approaches, attention-based models, e.g. Transformer, have emerged as being superior. Such models have opened the door to deployment of ASR on smart devices, however they still suffer from requiring a large number of model parameters. We propose an extremely low footprint E2E ASR system for smart devices, to achieve the goal of satisfying resource constraints without sacrificing recognition accuracy. We design cross-layer weight sharing to improve parameter efficiency and further exploit model compression methods including sparsification and quantization, to reduce memory storage and boost decoding efficiency. We evaluate our approaches on the public AISHELL-1 and AISHELL-2 benchmarks. On the AISHELL-2 task, the proposed method achieves more than 10× compression (model size reduces from 248 to 24MB), at the cost of only minor performance loss (CER reduces from 6.49% to 6.92%).

NOTES

Dissecting User-Perceived Latency of On-Device E2E Speech Recognition

Yuan Shangguan, Rohit Prabhavalkar, Hang Su, Jay Mahadeokar, Yangyang Shi, Jiatong Zhou, Chunyang Wu, Duc Le, Ozlem Kalinli, Christian Fuegen, Michael L. Seltzer; Facebook, USA

Fri-A-V-3-6, Time: 16:00

As speech-enabled devices such as smartphones and smart speakers become increasingly ubiquitous, there is growing interest in building automatic speech recognition (ASR) systems that can run directly on-device; end-to-end (E2E) speech recognition models such as recurrent neural network transducers and their variants have recently emerged as prime candidates for this task. Apart from being accurate and compact, such systems need to decode speech with low user-perceived latency (UPL), producing words as soon as they are spoken. This work examines the impact of various techniques model architectures, training criteria, decoding hyperparameters, and endpointer parameters - on UPL. Our analyses suggest that measures of model size (parameters, input chunk sizes), or measures of computation (e.g., FLOPS, RTF) that reflect the model's ability to process input frames are not always strongly correlated with observed UPL. Thus, conventional algorithmic latency measurements might be inadequate in accurately capturing latency observed when models are deployed on embedded devices. Instead, we find that factors affecting token emission latency, and endpointing behavior have a larger impact on UPL. We achieve the best trade-off between latency and word error rate when performing ASR jointly with endpointing, while utilizing the recently proposed alignment regularization mechanism.

Amortized Neural Networks for Low-Latency Speech Recognition

Jonathan Macoskey, Grant P. Strimel, Jinru Su, Ariya Rastrow; Amazon, USA Fri-A-V-3-7, Time: 16:00

We introduce Amortized Neural Networks (AmNets), a compute costand latency-aware network architecture particularly well-suited for sequence modeling tasks. We apply AmNets to the Recurrent Neural Network Transducer (RNN-T) to reduce compute cost and latency for an automatic speech recognition (ASR) task. The AmNets RNN-T architecture enables the network to dynamically switch between encoder branches on a frame-by-frame basis. Branches are constructed with variable levels of compute cost and model capacity. Here, we achieve variable compute for two well-known candidate techniques: one using sparse pruning and the other using matrix factorization. Frame-by-frame switching is determined by an arbitrator network that requires negligible compute overhead. We present results using both architectures on LibriSpeech data and show that our proposed architecture can reduce inference cost by up to 45% and latency to nearly real-time without incurring a loss in accuracy.

Tied & Reduced RNN-T Decoder

Rami Botros, Tara N. Sainath, Robert David, Emmanuel Guzman, Wei Li, Yanzhang He; Google, USA Fri-A-V-3-8, Time: 16:00

Previous works on the Recurrent Neural Network-Transducer (RNN-T) models have shown that, under some conditions, it is possible to simplify its prediction network with little or no loss in recognition accuracy [1, 2, 3]. This is done by limiting the context size of previous labels and/or using a simpler architecture for its layers instead of LSTMs. The benefits of such changes include reduction in model

size, faster inference and power savings, which are all useful for on-device applications.

In this work, we study ways to make the RNN-T decoder (prediction network + joint network) smaller and faster without degradation in recognition performance. Our prediction network performs a simple weighted averaging of the input embeddings, and shares its embedding matrix weights with the joint network's output layer (a.k.a. weight tying, commonly used in language modeling [4]). This simple design, when used in conjunction with additional Edit-based Minimum Bayes Risk (EMBR) training, reduces the RNN-T Decoder from 23M parameters to just 2M, without affecting word-error rate (WER).

PQK: Model Compression via Pruning, Quantization, and Knowledge Distillation

Jangho Kim¹, Simyung Chang¹, Nojun Kwak²; ¹Qualcomm, Korea; ²Seoul National University, Korea Fri-A-V-3-9, Time: 16:00

As edge devices become prevalent, deploying Deep Neural Networks (DNN) on edge devices has become a critical issue. However, DNN requires a high computational resource which is rarely available for edge devices. To handle this, we propose a novel model compression method for the devices with limited computational resources, called PQK consisting of pruning, quantization, and knowledge distillation (KD) processes. Unlike traditional pruning and KD, PQK makes use of unimportant weights pruned in the pruning process to make a teacher network for training a better student network without pre-training the teacher model. PQK has two phases. Phase 1 exploits iterative pruning and quantization-aware training to make a lightweight and power-efficient model. In phase 2, we make a teacher network by adding unimportant weights unused in phase 1 to a pruned network. By using this teacher network, we train the pruned network as a student network. In doing so, we do not need a pre-trained teacher network for the KD framework because the teacher and the student networks coexist within the same network (See Fig. 1). We apply our method to the recognition model and verify the effectiveness of PQK on keyword spotting (KWS) and image recognition.

Collaborative Training of Acoustic Encoders for Speech Recognition

Varun Nagaraja, Yangyang Shi, Ganesh Venkatesh, Ozlem Kalinli, Michael L. Seltzer, Vikas Chandra; Facebook, USA Fri-A-V-3-10, Time: 16:00

On-device speech recognition requires training models of different sizes for deploying on devices with various computational budgets. When building such different models, we can benefit from training them jointly to take advantage of the knowledge shared between them. Joint training is also efficient since it reduces the redundancy in the training procedure's data handling operations. We propose a method for collaboratively training acoustic encoders of different sizes for speech recognition. We use a sequence transducer setup where different acoustic encoders share a common predictor and joiner modules. The acoustic encoders are also trained using co-distillation through an auxiliary task for frame level chenone prediction, along with the transducer loss. We perform experiments using the LibriSpeech corpus and demonstrate that the collaboratively trained acoustic encoders can provide up to a 11% relative improvement in the word error rate on both the test partitions.

Efficient Conformer with Prob-Sparse Attention Mechanism for End-to-End Speech Recognition

*Xiong Wang*¹, *Sining Sun*², *Lei Xie*¹, *Long Ma*²; ¹Northwestern Polytechnical University, China; ²Tencent, China

Fri-A-V-3-11, Time: 16:00

End-to-end models are favored in automatic speech recognition (ASR) because of their simplified system structure and superior performance. Among these models, Transformer and Conformer have achieved state-of-the-art recognition accuracy in which self-attention plays a vital role in capturing important global information. However, the time and memory complexity of self-attention increases squarely with the length of the sentence. In this paper, a probsparse self-attention mechanism is introduced into Conformer to sparse the computing process of self-attention in order to accelerate inference speed and reduce space consumption. Specifically, we adopt a Kullback-Leibler divergence based sparsity measurement for each query to decide whether we compute the attention function on this query. By using the prob-sparse attention mechanism, we achieve impressively 8% to 45% inference speed-up and 15% to 45% memory usage reduction of the self-attention module of Conformer Transducer while maintaining the same level of error rate.

The Energy and Carbon Footprint of Training **End-to-End Speech Recognizers**

Titouan Parcollet¹, Mirco Ravanelli²; ¹LIA (EA 4128), France; ²Mila, Canada Fri-A-V-3-12, Time: 16:00

Deep learning contributes to reaching higher levels of artificial intelligence. Due to its pervasive adoption, however, growing concerns on the environmental impact of this technology have been raised. In particular, the energy consumed at training and inference time by modern neural networks is far from being negligible and will increase even further due to the deployment of ever larger models.

This work investigates for the first time the carbon cost of end-to-end automatic speech recognition (ASR). First, it quantifies the amount of CO2 emitted while training state-of-the-art (SOTA) ASR systems on a university-scale cluster. Then, it shows that a tiny performance improvement comes at an extremely high carbon cost. For instance, the conducted experiments reveal that a SOTA Transformer emits 50% of its total training released CO₂ solely to achieve a final decrease of 0.3 of the word error rate. With this study, we hope to raise awareness on this crucial topic and we provide guidelines, insights, and estimates enabling researchers to better assess the environmental impact of training speech technologies.

Fri-A-V-4: Speaker Recognition: Applications

16:00-18:00, Friday 3 September 2021

Chairs: Anthony Larcher and Rohan Kumar Das

Graph-Based Label Propagation for Semi-Supervised Speaker Identification

Long Chen, Venkatesh Ravichandran, Andreas Stolcke; Amazon, USA

Fri-A-V-4-1, Time: 16:00

Speaker identification in the household scenario (e.g., for smart speakers) is typically based on only a few enrollment utterances but a much larger set of unlabeled data, suggesting semi-supervised learning to improve speaker profiles. We propose a graph-based semisupervised learning approach for speaker identification in the household scenario, to leverage the unlabeled speech samples. In con-

trast to most of the works in speaker recognition that focus on speaker-discriminative embeddings, this work focuses on speaker label inference (scoring). Given a pre-trained embedding extractor, graph-based learning allows us to integrate information about both labeled and unlabeled utterances. Considering each utterance as a graph node, we represent pairwise utterance similarity scores as edge weights. Graphs are constructed per household, and speaker identities are propagated to unlabeled nodes to optimize a global consistency criterion. We show in experiments on the VoxCeleb dataset that this approach makes effective use of unlabeled data and improves speaker identification accuracy compared to two state-of-theart scoring methods as well as their semi-supervised variants based on pseudo-labels.

Fusion of Embeddings Networks for Robust Combination of Text Dependent and Independent Speaker Recognition

Ruirui Li, Chelsea J.-T. Ju, Zeya Chen, Hongda Mao, Oguz Elibol, Andreas Stolcke; Amazon, USA Fri-A-V-4-2, Time: 16:00

By implicitly recognizing a user based on his/her speech input, speaker identification enables many downstream applications, such as personalized system behavior and expedited shopping checkouts. Based on whether the speech content is constrained or not, both text-dependent (TD) and text-independent (TI) speaker recognition models may be used. We wish to combine the advantages of both types of models through an ensemble system to make more reliable predictions. However, any such combined approach has to be robust to incomplete inputs, i.e., when either TD or TI input is missing. As a solution we propose a <u>fusion of embeddings network</u> (FOEnet) architecture, combining joint learning with neural attention. We compare FOEnet with four competitive baseline methods on a dataset of voice assistant inputs, and show that it achieves higher accuracy than the baseline and score fusion methods, especially in the presence of incomplete inputs.

A Generative Model for Duration-Dependent Score Calibration

Sandro Cumani, Salvatore Sarni; Politecnico di Torino, Italy

Fri-A-V-4-3, Time: 16:00

In this work we introduce a generative score calibration model for speaker verification systems able to explicitly account for utterancedependent miscalibration sources, with a focus on segment duration. The model is theoretically motivated by an analysis of the effects of distribution mismatch on the scores produced by Probabilistic Linear Discriminant Analysis (PLDA), and extends our previous investigation on the distribution of well-calibrated PLDA log-likelihood ratios. We characterize target and non-target scores by means of Variance-Gamma densities, whose parameters represent effective between and within-class variabilities. Experimental results on SRE 2019 show that the proposed method improves both calibration and verification accuracy with respect to duration-agnostic models and to duration-aware discriminative methods.

Dr-Vectors: Decision Residual Networks and an Improved Loss for Speaker Recognition

Jason Pelecanos, Quan Wang, Ignacio Lopez Moreno; Google, USA

Fri-A-V-4-4, Time: 16:00

Many neural network speaker recognition systems model each speaker using a fixed-dimensional embedding vector. These embeddings are generally compared using either linear or 2nd-order scoring

NOTES

and, until recently, do not handle utterance-specific uncertainty. In this work we propose scoring these representations in a way that can capture uncertainty, enroll/test asymmetry and additional non-linear information. This is achieved by incorporating a 2nd-stage neural network (known as a decision network) as part of an end-to-end training regimen. In particular, we propose the concept of decision residual networks which involves the use of a compact decision network to leverage cosine scores and to model the residual signal that's needed. Additionally, we present a modification to the generalized end-toend softmax loss function to target the separation of same/different speaker scores. We observed significant performance gains for the two techniques.

Multi-Channel Speaker Verification for Single and Multi-Talker Speech

*Saurabh Kataria*¹, *Shi-Xiong Zhang*², *Dong Yu*²; ¹*Johns Hopkins University, USA*; ²*Tencent, USA* Fri-A-V-4-5, Time: 16:00

To improve speaker verification in real scenarios with interference speakers, noise, and reverberation, we propose to bring together advancements made in multi-channel speech features. Specifically, we combine spectral, spatial, and directional features, which includes inter-channel phase difference, multichannel sinc convolutions, directional power ratio features, and angle features. To maximally leverage supervised learning, our framework is also equipped with multi-channel speech enhancement and voice activity detection. On all simulated, replayed, and real recordings, we observe large and consistent improvements at various degradation levels. On real recordings of multi-talker speech, we achieve a 36% relative reduction in equal error rate w.r.t. single-channel baseline. We find the improvements from speaker-dependent directional features more consistent in multi-talker conditions than clean. Lastly, we investigate if the learned multi-channel speaker embedding space can be made more discriminative through a contrastive loss-based fine-tuning. With a simple choice of Triplet loss, we observe a further 8.3% relative reduction in EER.

Chronological Self-Training for Real-Time Speaker Diarization

Dirk Padfield, Daniel J. Liebling; Google, USA Fri-A-V-4-6, Time: 16:00

Diarization partitions an audio stream into segments based on the voices of the speakers. Real-time diarization systems that include an enrollment step should limit enrollment training samples to reduce user interaction time. Although training on a small number of samples yields poor performance, we show that the accuracy can be improved dramatically using a chronological self-training approach. We studied the tradeoff between training time and classification performance and found that 1 second is sufficient to reach over 95% accuracy. We evaluated on 700 audio conversation files of about 10 minutes each from 6 different languages and demonstrated average diarization error rates as low as 10%.

Adaptive Margin Circle Loss for Speaker Verification

Runqiu Xiao¹, Xiaoxiao Miao¹, Wenchao Wang¹, Pengyuan Zhang¹, Bin Cai², Liuping Luo²; ¹CAS, China; ²Guangdong PSD, China Fri-A-V-4-7, Time: 16:00

Deep-Neural-Network (DNN) based speaker verification systems use the angular softmax loss with margin penalties to enhance the intraclass compactness of speaker embeddings, which achieved remarkable performance. In this paper, we propose a novel angular loss function called adaptive margin circle loss for speaker verification. The stage-based margin and chunk-based margin are applied to improve the angular discrimination of circle loss on the training set. The analysis on gradients shows that, compared with the previous angular loss like Additive Margin Softmax(Am-Softmax), circle loss has flexible optimization and definite convergence status. Experiments are carried out on the Voxceleb and SITW. By applying adaptive margin circle loss, our best system achieves 1.31%EER on Voxceleb1 and 2.13% on SITW core-core.

Presentation Matters: Evaluating Speaker Identification Tasks

Benjamin O'Brien, Christine Meunier, Alain Ghio; LPL (UMR 7309), France Fri-A-V-4-8, Time: 16:00

This paper details our evaluations and comparisons of speaker identification (SID) performance by listeners across different tasks. Experiment 1 participants completed traditional target-lineup (1-outof-N speakers or out-of-set speaker) and binary (speaker verification) tasks. Experiment 2 participants completed trials online by using a clustering method by grouping speech recordings into speakerspecific clusters. Both studies employed similar speech recordings from the PTSVOX corpus. Our results showed participants who completed the binary and clustering tasks had higher accuracy than those who completed the target-lineup task. We also observed that independent of the tasks participants found some speakers significantly more difficult to identify relative to their foils. Pearson correlation procedures showed significant negative correlations between accuracy and task-dependent temporal-based metrics across tasks, where an increase in time required to make determinations yielded a decrease in perceptual SID performance. These findings underscored the important role of SID task design and the process of selecting speech recordings. Future work aims to examine the relationship between different perceptual SID task performances and scores generated by automatic speaker verification systems.

Automatic Error Correction for Speaker Embedding Learning with Noisy Labels

Fuchuan Tong, Yan Liu, Song Li, Jie Wang, Lin Li, Qingyang Hong; Xiamen University, China Fri-A-V-4-9, Time: 16:00

Despite the superior performance deep neural networks have achieved in speaker verification tasks, much of their success benefits from the availability of large-scale and carefully labeled datasets. However, noisy labels often occur during data collection. In this paper, we propose an automatic error correction method for deep speaker embedding learning with noisy labels. Specifically, a label noise correction loss is proposed that leverages a model's generalization capability to correct noisy labels during training. In addition, we improve the vanilla AM-Softmax to estimate a more robust speaker posterior by introducing sub-centers. When applied on the VoxCeleb dataset, the proposed method performs gracefully when noisy labels are introduced. Moreover, when combining with the Bayesian estimation of PLDA with noisy training labels at the back-end, the whole system performs better under conditions in which noisy labels are present.

An Integrated Framework for Two-Pass Personalized Voice Trigger

Dexin Liao, Jing Li, Yiming Zhi, Song Li, Qingyang Hong, Lin Li; Xiamen University, China Fri-A-V-4-10, Time: 16:00

In this paper, we present the XMUSPEECH system for Task 1 of 2020 Personalized Voice Trigger Challenge (PVTC2020). Task 1 is a joint wake-up word detection with speaker verification on close talking

data. The whole system consists of a keyword spotting (KWS) subsystem and a speaker verification (SV) sub-system. For the KWS system, we applied a Temporal Depthwise Separable Convolution Residual Network (TDSC-ResNet) to improve the system's performance. For the SV system, we proposed a multi-task learning network, where phonetic branch is trained with the character label of the utterance, and speaker branch is trained with the label of the speaker. Phonetic branch is optimized with connectionist temporal classification (CTC) loss, which is treated as an auxiliary module for speaker branch. Experiments show that our system gets significant improvements compared with baseline system.

Masked Proxy Loss for Text-Independent Speaker Verification

Jiachen Lian, Aiswarya Vinod Kumar, Hira Dhamyal, Bhiksha Rai, Rita Sinah: Carneaie Mellon University, USA Fri-A-V-4-11. Time: 16:00

Open-set speaker recognition can be regarded as a metric learning problem, which is to maximize inter-class variance and minimize intra-class variance. Supervised metric learning can be categorized into pair-based learning and proxy-based learning [1]. Most of the existing metric learning objectives belong to the former division, the performance of which is either highly dependent on sample mining strategy or restricted by insufficient label information in the minibatch. Proxy-based losses mitigate both shortcomings, however, finegrained connections among entities are either not or indirectly leveraged. This paper proposes a Masked Proxy (MP) loss which directly incorporates both proxy-based relationship and pair-based relationship. We further propose Multinomial Masked Proxy (MMP) loss to leverage the hardness of speaker pairs. These methods have been applied to evaluate on VoxCeleb test set and reach state-of-the-art Equal Error Rate (EER).

Fri-A-V-5: Speech Synthesis: Speaking Style and Emotion

16:00-18:00, Friday 3 September 2021 Chairs: Julie Carson-Berndsen and Javier Latorre

STYLER: Style Factor Modeling with Rapidity and **Robustness via Speech Decomposition for Expressive** and Controllable Neural Text to Speech

Keon Lee, Kyumin Park, Daeyoung Kim; KAIST, Korea Fri-A-V-5-1, Time: 16:00

Previous works on neural text-to-speech (TTS) have been addressed on limited speed in training and inference time, robustness for difficult synthesis conditions, expressiveness, and controllability. Although several approaches resolve some limitations, there has been no attempt to solve all weaknesses at once. In this paper, we propose STYLER, an expressive and controllable TTS framework with highspeed and robust synthesis. Our novel audio-text aligning method called Mel Calibrator and excluding autoregressive decoding enable rapid training and inference and robust synthesis on unseen data. Also, disentangled style factor modeling under supervision enlarges the controllability in synthesizing process leading to expressive TTS. On top of it, a novel noise modeling pipeline using domain adversarial training and Residual Decoding empowers noise-robust style transfer, decomposing the noise without any additional label. Various experiments demonstrate that STYLER is more effective in speed and robustness than expressive TTS with autoregressive decoding and more expressive and controllable than reading style non-autoregressive TTS. Synthesis samples and experiment results are provided via our demo page, and code is available publicly.

Reinforcement Learning for Emotional Text-to-Speech Synthesis with Improved Emotion Discriminability

*Rui Liu*¹, *Berrak Sisman*¹, *Haizhou Li*²; ¹*SUTD*, *Singapore*; ²*NUS*, *Singapore* Fri-A-V-5-2, Time: 16:00

Emotional text-to-speech synthesis (ETTS) has seen much progress in recent years. However, the generated voice is often not perceptually identifiable by its intended emotion category. To address this problem, we propose a new interactive training paradigm for ETTS, denoted as *i*-*ETTS*, which seeks to directly improve the emotion discriminability by interacting with a speech emotion recognition (SER) model. Moreover, we formulate an iterative training strategy with reinforcement learning to ensure the quality of *i*-*ETTS* optimization. Experimental results demonstrate that the proposed *i*-*ETTS* outperforms the state-of-the-art baselines by rendering speech with more accurate emotion style. To our best knowledge, this is the first study of reinforcement learning in emotional text-to-speech synthesis.

Emotional Prosody Control for Speech Generation

Sarath Sivaprasad, Saiteja Kosgi, Vineet Gandhi; IIIT Hyderabad, India

Fri-A-V-5-3, Time: 16:00

Machine-generated speech is characterized by its limited or unnatural emotional variation. Current text to speech systems generates speech with either a flat emotion, emotion selected from a predefined set, average variation learned from prosody sequences in training data or transferred from a source style. We propose a text to speech (TTS) system, where a user can choose the emotion of generated speech from a continuous and meaningful emotion space (Arousal-Valence space). The proposed TTS system can generate speech from the text in any speaker's style, with fine control of emotion. We show that the system works on emotion unseen during training and can scale to previously unseen speakers given his/her speech sample. Our work expands the horizon of the state-of-the-art FastSpeech2 backbone to a multi-speaker setting and gives it much-coveted continuous (and interpretable) affective control, without any observable degradation in the quality of the synthesized speech. Audio samples are publicly available.

Controllable Context-Aware Conversational Speech Synthesis

Jian Cong¹, Shan Yang², Na Hu², Guangzhi Li², Lei Xie¹, Dan Su²; ¹Northwestern Polytechnical University, China; ²Tencent, China Fri-A-V-5-4, Time: 16:00

In spoken conversations, spontaneous behaviors like filled pause and prolongations always happen. Conversational partner tends to align features of their speech with their interlocutor which is known as entrainment. To produce human-like conversations, we propose a unified controllable spontaneous conversational speech synthesis frame-

work to model the above two phenomena. Specifically, we use explicit labels to represent two typical spontaneous behaviors *filled-pause* and *prolongation* in the acoustic model and develop a neural network based predictor to predict the occurrences of the two behaviors from text. We subsequently develop an algorithm based on the predictor to control the occurrence frequency of the behaviors, making the synthesized speech vary from less disfluent to more disfluent. To model the speech entrainment at acoustic level, we utilize a context acoustic encoder to extract a global style embedding from the previous speech conditioning on the synthesizing of current speech. Furthermore, since the current and previous utterances belong to the different speakers in a conversation, we add a domain adversarial training module to eliminate the speaker-related information in the acoustic encoder while maintaining the style-related information. Experiments show that our proposed approach can synthesize realistic conversations and control the occurrences of the spontaneous behaviors naturally.

Expressive Text-to-Speech Using Style Tag

*Minchan Kim*¹, *Sung Jun Cheon*¹, *Byoung Jin Choi*¹, *Jong Jin Kim*², *Nam Soo Kim*¹; ¹*Seoul National University, Korea*; ²*SK Telecom, Korea* Fri-A-V-5-5, Time: 16:00

As recent text-to-speech (TTS) systems have been rapidly improved in speech quality and generation speed, many researchers now focus on a more challenging issue: expressive TTS. To control speaking styles, existing expressive TTS models use categorical style index or reference speech as style input. In this work, we propose StyleTagging-TTS (ST-TTS), a novel expressive TTS model that utilizes a style tag written in natural language. Using a style-tagged TTS dataset and a pre-trained language model, we modeled the relationship between linguistic embedding and speaking style domain, which enables our model to work even with style tags unseen during training. As style tag is written in natural language, it can control speaking style in a more intuitive, interpretable, and scalable way compared with style index or reference speech. In addition, in terms of model architecture, we propose an efficient non-autoregressive (NAR) TTS architecture with single-stage training. The experimental result shows that ST-TTS outperforms the existing expressive TTS model, Tacotron2-GST in speech quality and expressiveness.

Adaptive Text to Speech for Spontaneous Style

Yuzi Yan¹, Xu Tan², Bohan Li², Guangyan Zhang³, Tao Qin², Sheng Zhao², Yuan Shen¹, Wei-Qiang Zhang¹, Tie-Yan Liu²; ¹Tsinghua University, China; ²Microsoft, China; ³CUHK, China Fri-A-V-5-6, Time: 16:00

While recent text to speech (TTS) models perform very well in synthesizing reading-style (e.g., audiobook) speech, it is still challenging to synthesize spontaneous-style speech (e.g., podcast or conversation), mainly because of two reasons: 1) the lack of training data for spontaneous speech; 2) the difficulty in modeling the filled pauses (um and *uh*) and diverse rhythms in spontaneous speech. In this paper, we develop AdaSpeech 3, an adaptive TTS system that fine-tunes a well-trained reading-style TTS model for spontaneous-style speech. Specifically, 1) to insert filled pauses (FP) in the text sequence appropriately, we introduce an FP predictor to the TTS model; 2) to model the varying rhythms, we introduce a duration predictor based on mixture of experts (MoE), which contains three experts responsible for the generation of fast, medium and slow speech respectively, and fine-tune it as well as the pitch predictor for rhythm adaptation; 3) to adapt to other speaker timbre, we fine-tune some parameters in the decoder with few speech data. To address the challenge of lack of training data, we mine a spontaneous speech dataset to support our research this work and facilitate future research on spontaneous TTS. Experiments show that AdaSpeech 3 synthesizes speech with natural FP and rhythms in spontaneous styles, and achieves much better MOS and SMOS scores than previous adaptive TTS systems.

Towards Multi-Scale Style Control for Expressive Speech Synthesis

Xiang Li, Changhe Song, Jingbei Li, Zhiyong Wu, Jia Jia, Helen Meng; Tsinghua University, China Fri-A-V-5-7, Time: 16:00

This paper introduces a multi-scale speech style modeling method for end-to-end expressive speech synthesis. The proposed method em-
ploys a multi-scale reference encoder to extract both the global-scale utterance-level and the local-scale quasi-phoneme-level style features of the target speech, which are then fed into the speech synthesis model as an extension to the input phoneme sequence. During training time, the multi-scale style model could be jointly trained with the speech synthesis model in an end-to-end fashion. By applying the proposed method to style transfer task, experimental results indicate that the controllability of the multi-scale speech are greatly improved. Moreover, by assigning different reference speeches to extraction of style on each scale, the flexibility of the proposed method is further revealed.

Cross-Speaker Style Transfer with Prosody Bottleneck in Neural Speech Synthesis

Shifeng Pan, Lei He; Microsoft, China Fri-A-V-5-8, Time: 16:00

Cross-speaker style transfer is crucial to the applications of multistyle and expressive speech synthesis at scale. It does not require the target speakers to be experts in expressing all styles and to collect corresponding recordings for model training. However, the performances of existing style transfer methods are still far behind real application needs. The root causes are mainly twofold. Firstly, the style embedding extracted from single reference speech can hardly provide fine-grained and appropriate prosody information for arbitrary text to synthesize. Secondly, in these models the content/text, prosody, and speaker timbre are usually highly entangled, it's therefore not realistic to expect a satisfied result when freely combining these components, such as to transfer speaking style between speakers. In this paper, we propose a cross-speaker style transfer text-tospeech (TTS) model with explicit prosody bottleneck. The prosody bottleneck builds up the kernels accounting for speaking style robustly, and disentangles the prosody from content and speaker timbre, therefore guarantees high quality cross-speaker style transfer. Evaluation result shows the proposed method even achieves on-par performance with source speaker's speaker-dependent (SD) model in objective measurement of prosody, and significantly outperforms the cycle consistency and GMVAE-based baselines in objective and subjective evaluations.

Fine-Grained Style Modeling, Transfer and Prediction in Text-to-Speech Synthesis via Phone-Level Content-Style Disentanglement

Daxin Tan, Tan Lee; CUHK, China Fri-A-V-5-9, Time: 16:00

This paper presents a novel design of neural network system for fine-grained style modeling, transfer and prediction in expressive text-to-speech (TTS) synthesis. Fine-grained modeling is realized by extracting style embeddings from the mel-spectrograms of phonelevel speech segments. Collaborative learning and adversarial learning strategies are applied in order to achieve effective disentanglement of content and style factors in speech and alleviate the "content leakage" problem in style modeling. The proposed system can be used for varying-content speech style transfer in the single-speaker scenario. The results of objective and subjective evaluation show that our system performs better than other fine-grained speech style transfer models, especially in the aspect of content preservation. By incorporating a style predictor, the proposed system can also be used for text-to-speech synthesis. Audio samples are provided for system demonstration.

Improving Performance of Seen and Unseen Speech Style Transfer in End-to-End Neural TTS

*Xiaochun An*¹, *Frank K. Soong*², *Lei Xie*¹; ¹*Northwestern Polytechnical University, China;* ²*Microsoft, China Fri-A-V-5-10, Time:* 16:00

End-to-end neural TTS training has shown improved performance in speech style transfer. However, the improvement is still limited by the training data in both target styles and speakers. Inadequate style transfer performance occurs when the trained TTS tries to transfer the speech to a target style from a new speaker with an unknown, arbitrary style. In this paper, we propose a new approach to style transfer for both seen and unseen styles, with disjoint, multi-style datasets, i.e., datasets of different styles are recorded, each individual style is by one speaker with multiple utterances. To encode the style information, we adopt an inverse autoregressive flow (IAF) structure to improve the variational inference. The whole system is optimized to minimize a weighed sum of four different loss functions: 1) a reconstruction loss to measure the distortions in both source and target reconstructions; 2) an adversarial loss to "fool" a well-trained discriminator; 3) a style distortion loss to measure the expected style loss after the transfer; 4) a cycle consistency loss to preserve the speaker identity of the source after the transfer. Experiments demonstrate, both objectively and subjectively, the effectiveness of the proposed approach for seen and unseen style transfer tasks. The performance of the new approach is better and more robust than those of four baseline systems of the prior art.

Synthesis of Expressive Speaking Styles with Limited Training Data in a Multi-Speaker, Prosody-Controllable Sequence-to-Sequence Architecture

*Slava Shechtman*¹, *Raul Fernandez*², *Alexander Sorin*¹, *David Haws*²; ¹*IBM, Israel*; ²*IBM, USA* Fri-A-V-5-11, Time: 16:00

Although Sequence-to-Sequence (S2S) architectures have become state-of-the-art in speech synthesis, the best models benefit from access to moderate-to-large amounts of training data, posing a resource bottleneck when we are interested in generating speech in a variety of expressive styles. In this work we explore a S2S architecture variant that is capable of generating a variety of stylistic expressive variations observed in a limited amount of training data, and of transplanting that style to a neutral target speaker for whom no labeled expressive resources exist. The architecture is furthermore controllable, allowing the user to select an operating point that conveys a desired level of expressiveness. We evaluate this proposal against a classically supervised baseline via perceptual listening tests, and demonstrate that i) it is able to outperform the baseline in terms of its generalizability to neutral speakers, ii) it is strongly preferred in terms of its ability to convey expressiveness, and iii) it provides a reasonable trade-off between expressiveness and naturalness, allowing the user to tune it to the particular demands of a given application.

Fri-A-V-6: Spoken Language Understanding II

16:00–18:00, Friday 3 September 2021 Chairs: Martin Radfar and Isabel Trancoso

Intent Detection and Slot Filling for Vietnamese

Mai Hoang Dao, Thinh Hung Truong, Dat Quoc Nguyen; VinAI Research, Vietnam

Fri-A-V-6-1, Time: 16:00

Intent detection and slot filling are important tasks in spoken and natural language understanding. However, Vietnamese is a low-resource language in these research topics. In this paper, we present the *first* public intent detection and slot filling dataset for Vietnamese. In addition, we also propose a joint model for intent detection and slot filling, that extends the recent state-of-the-art JointBERT+CRF model [1] with an intent-slot attention layer to explicitly incorporate intent context information into slot filling via "soft" intent label embedding. Experimental results on our Vietnamese dataset show that our proposed model significantly outperforms JointBERT+CRF. We publicly release our dataset and the implementation of our model.

Augmenting Slot Values and Contexts for Spoken Language Understanding with Pretrained Models

Haitao Lin, Lu Xiang, Yu Zhou, Jiajun Zhang, Chengqing Zonq; CAS, China

Fri-A-V-6-2, Time: 16:00

Spoken Language Understanding (SLU) is one essential step in building a dialogue system. Due to the expensive cost of obtaining the labeled data, SLU suffers from the data scarcity problem. Therefore, in this paper, we focus on data augmentation for slot filling task in SLU. To achieve that, we aim at generating more diverse data based on existing data. Specifically, we try to exploit the latent language knowledge from pretrained language models by finetuning them. We propose two strategies for finetuning process: value-based and contextbased augmentation. Experimental results on two public SLU datasets have shown that compared with existing data augmentation methods, our proposed method can generate more diverse sentences and significantly improve the performance on SLU.

The Impact of Intent Distribution Mismatch on Semi-Supervised Spoken Language Understanding

Judith Gaspers, Quynh Do, Daniil Sorokin, Patrick Lehnen; Amazon, Germany Fri-A-V-6-3, Time: 16:00

With the expanding role of voice-controlled devices, bootstrapping spoken language understanding models from little labeled data becomes essential. Semi-supervised learning is a common technique to improve model performance when labeled data is scarce. In a realworld production system, the labeled data and the online test data often may come from different distributions. In this work, we use semi-supervised learning based on pseudo-labeling with an auxiliary task on incoming unlabeled noisy data, which is closer to the test distribution. We demonstrate empirically that our approach can mitigate negative effects arising from training with non-representative labeled data as well as the negative impacts of noises in the data, which are introduced by pseudo-labeling and automatic speech recognition.

Knowledge Distillation from BERT Transformer to Speech Transformer for Intent Classification

Yidi Jiang, Bidisha Sharma, Maulik Madhavi, Haizhou Li; NUS, Singapore

Fri-A-V-6-4, Time: 16:00

End-to-end intent classification using speech has numerous advantages compared to the conventional pipeline approach using automatic speech recognition (ASR), followed by natural language processing modules. It attempts to predict intent from speech without using an intermediate ASR module. However, such end-to-end framework suffers from the unavailability of large speech resources with higher acoustic variation in spoken language understanding. In this work, we exploit the scope of the transformer distillation method that is specifically designed for knowledge distillation from a transformer based language model to a transformer based speech model. In this regard, we leverage the reliable and widely used bidirectional encoder representations from transformers (BERT) model as a language model and transfer the knowledge to build an acoustic model for intent classification using the speech. In particular, a multilevel transformer based teacher-student model is designed, and knowledge distillation is performed across attention and hidden sub-layers of different transformer layers of the student and teacher models. We achieve an intent classification accuracy of 99.10% and 88.79% for Fluent speech corpus and ATIS database, respectively. Further, the proposed method demonstrates better performance and robustness in acoustically degraded condition compared to the baseline method.

Three-Module Modeling For End-to-End Spoken Language Understanding Using Pre-Trained DNN-HMM-Based Acoustic-Phonetic Model

Nick J.C. Wang, Lu Wang, Yandan Sun, Haimei Kang, Dejun Zhang; Ping An Technology, China Fri-A-V-6-5, Time: 16:00

In spoken language understanding (SLU), what the user says is converted to his/her intent. Recent work on end-to-end SLU has shown that accuracy can be improved via pre-training approaches. We revisit ideas presented by Lugosch et al. using speech pre-training and three-module modeling; however, to ease construction of the endto-end SLU model, we use as our phoneme module an open-source acoustic-phonetic model from a DNN-HMM hybrid automatic speech recognition (ASR) system instead of training one from scratch. Hence we fine-tune on speech only for the word module, and we apply multitarget learning (MTL) on the word and intent modules to jointly optimize SLU performance. MTL yields a relative reduction of 40% in intent-classification error rates (from 1.0% to 0.6%). Note that our three-module model is a streaming method. The final outcome of the proposed three-module modeling approach yields an intent accuracy of 99.4% on FluentSpeech, an intent error rate reduction of 50% compared to that of Lugosch et al. Although we focus on real-time streaming methods, we also list non-streaming methods for comparison.

Speak or Chat with Me: End-to-End Spoken Language Understanding System with Flexible Inputs

Sujeong Cha¹, Wangrui Hou¹, Hyun Jung¹, My Phung¹, Michael Picheny¹, Hong-Kwang J. Kuo², Samuel Thomas², Edmilson Morais³; ¹NYU, USA; ²IBM, USA; ³IBM, Brazil Fri-A-V-6-6, Time: 16:00

A major focus of recent research in spoken language understanding (SLU) has been on the end-to-end approach where a single model can predict intents directly from speech inputs without intermediate transcripts. However, this approach presents some challenges. First,

since speech can be considered as personally identifiable information, in some cases only automatic speech recognition (ASR) transcripts are accessible. Second, intent-labeled speech data is scarce. To address the first challenge, we propose a novel system that can predict intents from flexible types of inputs: speech, ASR transcripts, or both. We demonstrate strong performance for either modality separately, and when both speech and ASR transcripts are available, through system combination, we achieve better results than using a single input modality. To address the second challenge, we leverage a semantically robust pre-trained BERT model and adopt a cross-modal system that co-trains text embeddings and acoustic embeddings in a shared latent space. We further enhance this system by utilizing an acoustic module pre-trained on LibriSpeech and domain-adapting the text module on our target datasets. Our experiments show significant advantages for these pre-training and fine-tuning strategies, resulting in a system that achieves competitive intent-classification performance on Snips SLU and Fluent Speech Commands datasets.

End-to-End Cross-Lingual Spoken Language Understanding Model with Multilingual Pretraining

Xianwei Zhang, Liang He; Tsinghua University, China Fri-A-V-6-7, Time: 16:00

The spoken language understanding (SLU) plays an essential role in the field of human-computer interaction. Most of the current SLU systems are cascade systems of automatic speech recognition (ASR) and natural language understanding (NLU). Error propagation and scarcity of annotated speech data are two common difficulties for resourcepoor languages. To solve them, we propose a simple but effective endto-end cross-lingual spoken language understanding model based on XLSR-53, which is a pretrained model in 53 languages by the Facebook research team. The end-to-end approach avoids error propagation and the multilingual pretraining reduces data annotation requirements. Our proposed method achieves 99.71% on the Fluent Speech Commands (FSC) English database and 79.89% on the CATSLU-MAP Chinese database, in intent classification accuracy. To the best of our knowledge, the former is the reported best result on the FSC database.

Factorization-Aware Training of Transformers for Natural Language Understanding on the Edge

Hamidreza Saghir, Samridhi Choudhary, Sepehr Eghbali, Clement Chung; Amazon, Canada Fri-A-V-6-8, Time: 16:00

Fine-tuning transformer-based models have shown to outperform other methods for many Natural Language Understanding (NLU) tasks. Recent studies to reduce the size of transformer models have achieved reductions of > 80%, making on-device inference on powerful devices possible. However, other resource-constrained devices, like those enabling voice assistants (VAs), require much further reductions. In this work, we propose factorization-aware training (FAT), wherein we factorize the linear mappings of an already compressed transformer model (DistilBERT) and train jointly on NLU tasks. We test this method on three different NLU datasets and show our method outperforms naive application of factorization after training by 10%-440% across various compression rates. Additionally, We introduce a new metric called factorization gap and use it to analyze the need for FAT across various model components. We also present results for training subsets of factorized components to enable faster training, re-usability and maintainability for multiple on-device models. We further demonstrate the trade-off between memory, inference speed and performance at a given compressionrate for a on-device implementation of a factorized model. Our best performing factorized model, achieves a relative size reduction of 84% with $\approx 10\%$ relative degradation in NLU error rate compared to a non-factorized model on our internal dataset.

End-to-End Spoken Language Understanding for Generalized Voice Assistants

Michael Saxon, Samridhi Choudhary, Joseph P. McKenna, Athanasios Mouchtaris; Amazon, USA Fri-A-V-6-9, Time: 16:00

End-to-end (E2E) spoken language understanding (SLU) systems predict utterance semantics directly from speech using a single model. Previous work in this area has focused on targeted tasks in fixed domains, where the output semantic structure is assumed a priori and the input speech is of limited complexity. In this work we present our approach to developing an E2E model for generalized SLU in commercial voice assistants (VAs). We propose a fully differentiable, transformer-based, hierarchical system that can be pretrained at both the ASR and NLU levels. This is then fine-tuned on both transcription and semantic classification losses to handle a diverse set of intent and argument combinations. This leads to an SLU system that achieves significant improvements over baselines on a complex internal generalized VA dataset with a 43% improvement in accuracy, while still meeting the 99% accuracy benchmark on the popular Fluent Speech Commands dataset. We further evaluate our model on a hard test set, exclusively containing slot arguments unseen in training, and demonstrate a nearly 20% improvement, showing the efficacy of our approach in truly demanding VA scenarios.

Bi-Directional Joint Neural Networks for Intent Classification and Slot Filling

Soyeon Caren Han, Siqu Long, Huichun Li, Henry Weld, Josiah Poon; University of Sydney, Australia Fri-A-V-6-10, Time: 16:00

Intent classification and slot filling are two critical tasks for natural language understanding. Traditionally the two tasks proceeded independently. However, more recently joint models for intent classification and slot filling have achieved state-of-the-art performance, and have proved that there exists a strong relationship between the two tasks. In this paper, we propose a bi-directional joint model for intent classification and slot filling, which includes a multi-stage hierarchical process via BERT and bi-directional joint natural language understanding mechanisms, including intent2slot and slot2intent, to obtain mutual performance enhancement between intent classification and slot filling. The evaluations show that our model achieves state-of-the-art results on intent classification accuracy, slot filling F1, and significantly improves sentence-level semantic frame accuracy when applied to publicly available benchmark datasets, ATIS (88.6%) and SNIPS (92.8%).

Fri-A-SS-1: INTERSPEECH 2021 Acoustic Echo Cancellation Challenge

Room Lacina, 16:00-18:00, Friday 3 September 2021 Chairs: Ross Cutler and Ando Saabas

INTERSPEECH 2021 Acoustic Echo Cancellation Challenge

Ross Cutler, Ando Saabas, Tanel Parnamaa, Markus Loide, Sten Sootla, Marju Purin, Hannes Gamper, Sebastian Braun, Karsten Sorensen, Robert Aichner, Sriram Srinivasan; Microsoft, USA Fri-A-SS-1-1, Time: 16:00

The INTERSPEECH 2021 Acoustic Echo Cancellation Challenge is intended to stimulate research in the area of acoustic echo cancellation (AEC), which is an important part of speech enhancement and still a top issue in audio communication. Many recent AEC studies re-

port good performance on synthetic datasets where the training and testing data may come from the same underlying distribution. However, AEC performance often degrades significantly on real recordings. Also, most of the conventional objective metrics such as echo return loss enhancement and perceptual evaluation of speech quality do not correlate well with subjective speech quality tests in the presence of background noise and reverberation found in realistic environments. In this challenge, we open source two large datasets to train AEC models under both single talk and double talk scenarios. These datasets consist of recordings from more than 5,000 real audio devices and human speakers in real environments, as well as a synthetic dataset. We also open source an online subjective test framework and provide an online objective metric service for researchers to auickly test their results. The winners of this challenge are selected based on the average Mean Opinion Score achieved across all different single talk and double talk scenarios.

Acoustic Echo Cancellation with Cross-Domain Learning

Lukas Pfeifenberger¹, Matthias Zoehrer¹, Franz Pernkopf²; ¹Evolve, Austria; ²Technische Universität Graz. Austria Fri-A-SS-1-2, Time: 16:20

This paper proposes the Cross-Domain Echo-Controller (CDEC), submitted to the Interspeech 2021 AEC-Challenge. The algorithm consists of three building blocks: (i) a Time-Delay Compensation (TDC) module, (ii) a frequency-domain block-based Acoustic Echo Canceler (AEC), and (iii) a Time-Domain Neural-Network (TD-NN) used as a postprocessor. Our system achieves an overall MOS score of 3.80, while only using 2.1 million parameters at a system latency of 32ms.

F-T-LSTM Based Complex Network for Joint Acoustic **Echo Cancellation and Speech Enhancement**

Shimin Zhang, Yuxiang Kong, Shubo Lv, Yanxin Hu, Lei Xie; Northwestern Polytechnical University, China Fri-A-SS-1-3, Time: 16:40

With the increasing demand for audio communication and online conference, ensuring the robustness of Acoustic Echo Cancellation (AEC) under the complicated acoustic scenario including noise, reverberation and nonlinear distortion has become a top issue. Although there have been some traditional methods that consider nonlinear distortion, they are still inefficient for echo suppression and the performance will be attenuated when noise is present. In this paper, we present a real-time AEC approach using complex neural network to better modeling the important phase information and frequencytime-LSTMs (F-T-LSTM), which scan both frequency and time axis, for better temporal modeling. Moreover, we utilize modified SI-SNR as cost function to make the model to have better echo cancellation and noise suppression (NS) performance. With only 1.4M parameters, the proposed approach outperforms the AEC-challenge baseline by 0.27 in terms of Mean Opinion Score (MOS).

Y²-Net FCRN for Acoustic Echo and Noise Suppression

Ernst Seidel, Jan Franzen, Maximilian Strake, Tim Fingscheidt; Technische Universität Braunschweig, Germany

Fri-A-SS-1-4, Time: 17:00

In recent years, deep neural networks (DNNs) were studied as an alternative to traditional acoustic echo cancellation (AEC) algorithms. The proposed models achieved remarkable performance for the separate tasks of AEC and residual echo suppression (RES). A promising network topology is a fully convolutional recurrent network (FCRN) structure, which has already proven its performance on both noise

NOTES

suppression and AEC tasks, individually. However, the combination of AEC, postfiltering, and noise suppression to a single network typically leads to a noticeable decline in the quality of the near-end speech component due to the lack of a separate loss for echo estimation. In this paper, we propose a two-stage model (Y²-Net) which consists of two FCRNs, each with two inputs and one output (Y-Net). The first stage (AEC) yields an echo estimate, which - as a novelty for a DNN AEC model — is further used by the second stage to perform RES and noise suppression. While the subjective listening test of the Interspeech 2021 AEC Challenge mostly yielded results close to the baseline, the proposed method scored an average improvement of 0.46 points over the baseline on the blind testset in double-talk on the instrumental metric DECMOS, provided by the challenge organizers.

Acoustic Echo Cancellation Using Deep Complex Neural Network with Nonlinear Magnitude **Compression and Phase Information**

Renhua Peng, Linjuan Cheng, Chengshi Zheng, Xiaodong Li; CAS, China Fri-A-SS-1-5, Time: 17:20

This paper describes a two-stage acoustic echo cancellation (AEC) and suppression framework for the INTERSPEECH2021 AEC Challenge. In the first stage, four parallel partitioned block frequency domain adaptive filters are used to cancel the linear echo components, where the far-end signal is delayed 0ms, 320ms, 640ms and 960ms for these four adaptive filters, respectively, thus a maximum 1280 ms time delay can be well handled in the blind test dataset. The error signal with minimum energy and its corresponding reference signal are chosen as the input for the second stage, where a gate complex convolutional recurrent neural network (GCCRN) is trained to further suppress the residual echo, late reverberation and environmental noise simultaneously. To improve the performance of GCCRN, we compress both the magnitude of the error signal and that of the far-end reference signal, and then the two compressed magnitudes are combined with the phase of the error signal to regenerate the complex spectra as the input features of GCCRN. Numerous experimental results show that the proposed framework is robust to the blind test dataset, and achieves a promising result with the P.808 evaluation.

Nonlinear Acoustic Echo Cancellation with Deep Learning

Amir Ivry, Israel Cohen, Baruch Berdugo; Technion, Israel

Fri-A-SS-1-6, Time: 17:40

We propose a nonlinear acoustic echo cancellation system, which aims to model the echo path from the far-end signal to the near-end microphone in two parts. Inspired by the physical behavior of modern hands-free devices, we first introduce a novel neural network architecture that is specifically designed to model the nonlinear distortions these devices induce between receiving and playing the far-end signal. To account for variations between devices, we construct this network with trainable memory length and nonlinear activation functions that are not parameterized in advance, but are rather optimized during the training stage using the training data. Second, the network is succeeded by a standard adaptive linear filter that constantly tracks the echo path between the loudspeaker output and the microphone. During training, the network and filter are jointly optimized to learn the network parameters. This system requires 17 thousand parameters that consume 500 Million floating-point operations per second and 40 Kilo-bytes of memory. It also satisfies hands-free communication timing requirements on a standard neural processor, which renders it adequate for embedding on hands-free communication devices. Using 280 hours of real and synthetic data, experiments show advantageous performance compared to competing methods.

Fri-A-SS-2: Speech Recognition of Atypical Speech

Room C, 16:00–18:00, Friday 3 September 2021 Chairs: Katrin Tomanek and Jordan Green

Introduction

Short Presentations of Papers Time: 16:05

Automatic Speech Recognition of Disordered Speech: Personalized Models Outperforming Human Listeners on Short Phrases

Jordan R. Green¹, Robert L. MacDonald², Pan-Pan Jiang², Julie Cattiau², Rus Heywood², Richard Cave³, Katie Seaver¹, Marilyn A. Ladewig⁴, Jimmy Tobin², Michael P. Brenner², Philip C. Nelson², Katrin Tomanek²; ¹MGH Institute of Health Professions, USA; ²Google, USA; ³MND Association, UK; ⁴Cerebral Palsy Associations of New York State, USA Fri-A-SS-2-1, Time: 16:25

This study evaluated the accuracy of personalized automatic speech recognition (ASR) for recognizing disordered speech from a large cohort of individuals with a wide range of underlying etiologies using an open vocabulary. The performance of these models was benchmarked relative to that of expert human transcribers and two different speaker-independent ASR models trained on typical speech. 432 individuals with self-reported disordered speech recorded at least 300 short phrases using a web-based application. Word error rates (WERs) were estimated for three different ASR models and for human transcribers. Metadata were collected to evaluate the potential impact of participants, atypical speech characteristics, and technical factors on recognition accuracy. Personalized models outperformed human transcribers with median and max recognition accuracy gains of 9% and 80%, respectively. The accuracies of personalized models were high (median WER: 4.6%) and better than those of speakerindependent models (median WER: 31%). The most significant improvements were for the most severely affected speakers. Low signalto-noise ratio and fewer training utterances were associated with poor word recognition, even for speakers with mild speech impairments. Our results demonstrate the efficacy of personalized ASR models in recognizing a wide range of speech impairments and severities and using an open vocabulary.

Investigating the Utility of Multimodal Conversational Technology and Audiovisual Analytic Measures for the Assessment and Monitoring of Amyotrophic Lateral Sclerosis at Scale

Michael Neumann¹, Oliver Roesler¹, Jackson Liscombe¹, Hardik Kothare¹, David Suendermann-Oeft¹, David Pautler¹, Indu Navar², Aria Anvar², Jochen Kumm³, Raquel Norel⁴, Ernest Fraenkel⁵, Alexander V. Sherman⁶, James D. Berry⁶, Gary L. Pattee⁷, Jun Wang⁸, Jordan R. Green⁶, Vikram Ramanarayanan¹; ¹Modality.AI, USA; ²Peter Cohen Foundation, USA; ³Pr3vent, USA; ⁴IBM, USA; ⁵MIT, USA; ⁶MGH Institute of Health Professions, USA; ⁷University of Nebraska, USA; ⁸University of Texas at Austin, USA Fri-ASS-2-2, Time: 16:25

We propose a cloud-based multimodal dialog platform for the remote assessment and monitoring of Amyotrophic Lateral Sclerosis (ALS) at scale. This paper presents our vision, technology setup, and an initial investigation of the efficacy of the various acoustic and visual speech metrics automatically extracted by the platform. 82 healthy controls and 54 people with ALS (pALS) were instructed to interact with the platform and completed a battery of speaking tasks designed to probe the acoustic, articulatory, phonatory, and respiratory aspects of their speech. We find that multiple acoustic (rate, duration, voicing) and visual (higher order statistics of the jaw and lip) speech metrics show statistically significant differences between controls, bulbar symptomatic and bulbar pre-symptomatic patients. We report on the sensitivity and specificity of these metrics using five-fold cross-validation. We further conducted a LASSO-LARS regression analysis to uncover the relative contributions of various acoustic and visual features in predicting the severity of patients' ALS (as measured by their selfreported ALSFRSR scores). Our results provide encouraging evidence of the utility of automatically extracted audiovisual analytics for scalable remote patient assessment and monitoring in ALS.

Handling Acoustic Variation in Dysarthric Speech Recognition Systems Through Model Combination

Enno Hermann, Mathew Magimai-Doss; Idiap Research Institute, Switzerland Fri-A-SS-2-3, Time: 16:25

Developing automatic speech recognition (ASR) systems that recognise dysarthric speech as well as control speech from unimpaired speakers remains challenging. Including more highly variable dysarthric speech during training can also negatively affect the performance on control speakers, which is not desirable when developing speech recognisers for a wider audience. In this work, we analyse how the acoustic variability of dysarthric speech affects ASR systems and propose the combination of multiple acoustic models trained on different subsets of speakers to mitigate this effect. This approach shows improvements for both dysarthric and control speakers on the Torgo and UA-Speech corpora.

Spectro-Temporal Deep Features for Disordered Speech Assessment and Recognition

Mengzhe Geng¹, Shansong Liu¹, Jianwei Yu¹, Xurong Xie², Shoukang Hu¹, Zi Ye¹, Zengrui Jin¹, Xunying Liu¹, Helen Meng¹; ¹CUHK, China; ²CAS, China Fri-A-SS-2-4, Time: 16:25

Automatic recognition of disordered speech remains a highly challenging task to date. Sources of variability commonly found in normal

speech including accent, age or gender, when further compounded with the underlying causes of speech impairment and varying severity levels, create large diversity among speakers. To this end, speaker adaptation techniques play a vital role in current speech recognition systems. Motivated by the spectro-temporal level differences between disordered and normal speech that systematically manifest in articulatory imprecision, decreased volume and clarity, slower speaking rates and increased dysfluencies, novel spectro-temporal subspace basis embedding deep features derived by SVD decomposition of speech spectrum are proposed to facilitate both accurate speech intelligibility assessment and auxiliary feature based speaker adaptation of state-of-the-art hybrid DNN and end-to-end disordered speech recognition systems. Experiments conducted on the UASpeech corpus suggest the proposed spectro-temporal deep feature adapted systems consistently outperformed baseline i-Vector adaptation by up to 2.63% absolute (8.6% relative) reduction in word error rate (WER) with or without data augmentation. Learning hidden unit contribution (LHUC) based speaker adaptation was further applied. The final speaker adapted system using the proposed spectral basis embedding features gave an overall WER of 25.6% on the UASpeech test set of 16 dysarthric speakers.

Speaking with a KN95 Face Mask: ASR Performance and Speaker Compensation

Sarah E. Gutz¹, Hannah P. Rowe², Jordan R. Green¹; ¹Harvard University, USA; ²MGH Institute of Health Professions, USA

Fri-A-SS-2-5, Time: 16:25

The increasing prevalence of face masks in the United States due to the COVID-19 pandemic necessitates serious consideration of the functional impact of wearing a mask on speech. This study considers how the presence of a KN95 mask affects the performance of a commercial ASR system, Google Cloud Speech. We present evidence that wearing a mask does not impact ASR performance at the sentence level. Moreover, speakers may be naturally adapting to the mask by increasing their vowel space area. However, when speakers intentionally altered their speech by speaking clearly or loudly (though not slowly), ASR performance improved. These findings suggest that ASR users can employ speech strategies to achieve better ASR results when wearing a mask. Beyond healthy speakers, our study has implications for mask-wearing ASR users with otherwise reduced speech intelligibility.

Adversarial Data Augmentation for Disordered Speech Recognition

Zengrui Jin¹, Mengzhe Geng¹, Xurong Xie², Jianwei Yu¹, Shansong Liu¹, Xunying Liu¹, Helen Meng¹; ¹CUHK, China; ²CAS, China Fri-A-SS-2-6, Time: 16:25

Automatic recognition of disordered speech remains a highly challenging task to date. The underlying neuro-motor conditions, often compounded with co-occurring physical disabilities, lead to the difficulty in collecting large quantities of impaired speech required for ASR system development. To this end, data augmentation techniques play a vital role in current disordered speech recognition systems. In contrast to existing data augmentation techniques only modifying the speaking rate or overall shape of spectral contour, fine-grained spectro-temporal differences between disordered and normal speech are modelled using deep convolutional generative adversarial networks (DCGAN) during data augmentation to modify normal speech spectra into those closer to disordered speech. Experiments conducted on the UASpeech corpus suggest the proposed adversarial data augmentation approach consistently outperformed the baseline

NOTES

augmentation methods using tempo or speed perturbation on a stateof-the-art hybrid DNN system. An overall word error rate (WER) reduction up to 3.05% (9.7% relative) was obtained over the baseline system using no data augmentation. The final learning hidden unit contribution (LHUC) speaker adapted system using the best adversarial augmentation approach gives an overall WER of 25.89% on the UASpeech test set of 16 dysarthric speakers.

Variational Auto-Encoder Based Variability Encoding for Dysarthric Speech Recognition

*Xurong Xie*¹, *Rukiye Ruzi*¹, *Xunying Liu*², *Lan Wang*¹; ¹*CAS, China;* ²*CUHK, China* Fri-A-SS-2-7, Time: 16:25

Dysarthric speech recognition is a challenging task due to acoustic variability and limited amount of available data. Diverse conditions of dysarthric speakers account for the acoustic variability, which make the variability difficult to be modeled precisely. This paper presents a variational auto-encoder based variability encoder (VAEVE) to explicitly encode such variability for dysarthric speech. The VAEVE makes use of both phoneme information and low-dimensional latent variable to reconstruct the input acoustic features, thereby the latent variable is forced to encode the phoneme-independent variability. Stochastic gradient variational Bayes algorithm is applied to model the distribution for generating variability encodings, which are further used as auxiliary features for DNN acoustic modeling. Experiment results conducted on the UASpeech corpus show that the VAEVE based variability encodings have complementary effect to the learning hidden unit contributions (LHUC) speaker adaptation. The systems using variability encodings consistently outperform the comparable baseline systems without using them, and obtain absolute word error rate (WER) reduction by up to 2.2% on dysarthric speech with "Very low" intelligibility level, and up to 2% on the "Mixed" type of dysarthric speech with diverse or uncertain conditions.

Learning Explicit Prosody Models and Deep Speaker Embeddings for Atypical Voice Conversion

*Disong Wang*¹, *Songxiang Liu*¹, *Lifa Sun*², *Xixin Wu*³, *Xunying Liu*¹, *Helen Meng*¹; ¹*CUHK*, *China*; ²*SpeechX*, *China*; ³*University of Cambridge*, *UK* Fri-A-SS-2-8, Time: 16:25

Though significant progress has been made for the voice conversion (VC) of typical speech, VC for atypical speech, e.g., dysarthric and second-language (L2) speech, remains a challenge, since it involves correcting for atypical prosody while maintaining speaker identity. To address this issue, we propose a VC system with explicit prosodic modelling and deep speaker embedding (DSE) learning. First, a speech-encoder strives to extract robust phoneme embeddings from atypical speech. Second, a prosody corrector takes in phoneme embeddings to infer typical phoneme duration and pitch values. Third, a conversion model takes phoneme embeddings and typical prosody features as inputs to generate the converted speech, conditioned on the target DSE that is learned via speaker encoder or speaker adaptation. Extensive experiments demonstrate that speaker adaptation can achieve higher speaker similarity, and the speaker encoder based conversion model can greatly reduce dysarthric and non-native pronunciation patterns with improved speech intelligibility. A comparison of speech recognition results between the original dysarthric speech and converted speech show that absolute reduction of 47.6% character error rate (CER) and 29.3% word error rate (WER) can be achieved.

Bayesian Parametric and Architectural Domain Adaptation of LF-MMI Trained TDNNs for Elderly and Dysarthric Speech Recognition

Jiajun Deng¹, Fabian Ritter Gutierrez¹, Shoukang Hu¹, Mengzhe Geng¹, Xurong Xie², Zi Ye¹, Shansong Liu¹, Jianwei Yu¹, Xunying Liu¹, Helen Meng¹; ¹CUHK, China; ²CAS, China Fri-A-SS-2-9, Time: 16:25

Automatic recognition of elderly and disordered speech remains a highly challenging task to date. Such data is not only difficult to collect in large quantities, but also exhibits a significant mismatch against normal speech trained ASR systems. To this end, conventional deep neural network model adaptation approaches only consider parameter fine-tuning on limited target domain data. In this paper, a novel Bayesian parametric and neural architectural domain adaptation approach is proposed. Both the standard model parameters and architectural hyper-parameters (hidden layer L/R context offsets) of two lattice-free MMI (LF-MMI) factored TDNN systems separately trained using large quantities of normal speech from the English LibriSpeech and Cantonese SpeechOcean corpora were domain adapted to two tasks: a) 16-hour DementiaBank elderly speech corpus; and b) 14-hour CUDYS dysarthric speech database. A Bayesian differentiable architectural search (DARTS) super-network was designed to allow both efficient search over up to 7²⁸ different TDNN structures during domain adaptation, and robust modelling of parameter uncertainty given limited target domain data. Absolute recognition error rate reductions of 1.82% and 2.93% (13.2% and 8.3% relative) were obtained over the baseline systems performing model parameter fine-tuning only. Consistent performance improvements were retained after data augmentation and learning hidden unit contribution (LHUC) based speaker adaptation was performed.

A Voice-Activated Switch for Persons with Motor and Speech Impairments: Isolated-Vowel Spotting Using Neural Networks

Shanqing Cai, Lisie Lillianfeld, Katie Seaver, Jordan R. Green, Michael P. Brenner, Philip C. Nelson, D. Sculley; Google, USA

Fri-A-SS-2-10, Time: 16:25

Severe speech impairments limit the precision and range of producible speech sounds. As a result, generic automatic speech recognition (ASR) and keyword spotting (KWS) systems fail to accurately recognize the utterances produced by individuals with severe speech impairments. This paper describes an approach in a simple speech sound, namely isolated open vowel (/a/), is used in lieu of more motorically-demanding utterances. A neural network (NN) is trained to detect the isolated open vowel uttered by impaired speakers. The NN is trained with a two-phase approach. The pre-training phase uses samples from unimpaired speakers along with samples of background noises and unrelated speech; then the fine-tuning phase uses samples of vowel samples collected from individuals with speech impairments. This model can be built into an experimental mobile app to act as a switch that allows users to activate preconfigured actions such as alerting caregivers. Preliminary user testing indicates the vowel spotter has the potential to be a useful and flexible emergency communication channel for motor- and speech-impaired individuals.

Conformer Parrotron: A Faster and Stronger End-to-End Speech Conversion and Recognition Model for Atypical Speech

Zhehuai Chen, Bhuvana Ramabhadran, Fadi Biadsy, Xia Zhang, Youzheng Chen, Liyang Jiang, Fang Chu, Rohan Doshi, Pedro J. Moreno; Google, USA Fri-A-SS-2-11, Time: 16:25

Parrotron is an end-to-end personalizable model that enables manyto-one voice conversion (VC) and automated speech recognition (ASR) simultaneously for atypical speech. In this work, we present the nextgeneration Parrotron model with improvements in overall accuracy, training and inference speeds. The proposed architecture builds on the recent Conformer encoder comprising of convolution and attention layer based blocks used in ASR. We introduce architectural modifications that subsamples encoder activations to achieve speed-ups in training and inference. In order to jointly improve ASR and voice conversion quality, we show that this requires a corresponding upsampling after the Conformer encoder blocks. We provide an in-depth analysis on how the proposed approach can maximize the efficiency of a speech-to-speech conversion model in the context of atypical speech. Experiments on both many-to-one and one-to-one dysarthric speech conversion tasks show that we can achieve up to $7 \times$ speedup and 35% relative reduction in WER over the previous best Transformer Parrotron.

Disordered Speech Data Collection: Lessons Learned at 1 Million Utterances from Project Euphonia

Robert L. MacDonald¹, Pan-Pan Jiang¹, Julie Cattiau¹, Rus Heywood¹, Richard Cave², Katie Seaver³, Marilyn A. Ladewig⁴, Jimmy Tobin¹, Michael P. Brenner¹, Philip C. Nelson¹, Jordan R. Green³, Katrin Tomanek¹; ¹Google, USA; ²MND Association, UK; ³MGH Institute of Health Professions, USA; ⁴Cerebral Palsy Associations of New York State, USA Fri-A-SS-2-12, Time: 16:25

Speech samples from over 1000 individuals with impaired speech have been submitted for Project Euphonia, aimed at improving automated speech recognition systems for disordered speech. We provide an overview of the corpus, which recently passed 1 million utterances (>1300 hours), and review key lessons learned from this project. The reasoning behind decisions such as phrase set composition, prompted vs extemporaneous speech, metadata and data quality efforts are explained based on findings from both technical and user-facing research.

Automatic Severity Classification of Korean Dysarthric Speech Using Phoneme-Level Pronunciation Features

Eun Jung Yeo, Sunhee Kim, Minhwa Chung; Seoul National University, Korea Fri-A-SS-2-13, Time: 16:25

This paper proposes an automatic severity classification method for Korean dysarthric speech by using two types of phoneme-level pronunciation features. The first type is the percentage of correct phonemes, which consists of percentage of correct consonants, percentage of correct vowels, and percentage of total correct phonemes. The second type is related to the degree of vowel distortion, such as vowel space area, formant centralized ratio, vowel articulatory index, and F2-ratio. The baseline experiments use features from our previous study, consisting of MFCCs, voice quality features, and prosody features. Compared to the baseline, experiments including phoneme

level pronunciation features achieve a relative percentage increase of 32.55% and 33.84% in F1-score for support vector machine and feed-forward neural network classifiers, respectively. Our best performance reaches an F1-score of 77.38%, which is a relative percentage increase of 10.39% compared to the best previous results conducted on the Korean dysarthric QoLT corpus. Furthermore, with feature selection applied, all seven phoneme-level pronunciation features are chosen, accounting for the highest percentage of the selected feature set by both recursive feature elimination and extra trees classifier feature selection algorithms. Results indicate that phoneme-level pronunciation features are useful in enhancing the performance for automatic severity classification of dysarthric speech.

Comparing Supervised Models and Learned Speech Representations for Classifying Intelligibility of Disordered Speech on Selected Phrases

Subhashini Venugopalan¹, Joel Shor², Manoj Plakal¹, Jimmy Tobin¹, Katrin Tomanek¹, Jordan R. Green³, Michael P. Brenner¹; ¹Google, USA; ²Google, Japan; ³MGH Institute of Health Professions, USA Fri-A SS-2-14, Time: 16:25

Automatic classification of disordered speech can provide an objective tool for identifying the presence and severity of a speech impairment. Classification approaches can also help identify hard-torecognize speech samples to teach ASR systems about the variable manifestations of impaired speech. Here, we develop and compare different deep learning techniques to classify the intelligibility of disordered speech on selected phrases. We collected samples from a diverse set of 661 speakers with a variety of self-reported disorders speaking 29 words or phrases, which were rated by speech-language pathologists for their overall intelligibility using a five-point Likert scale. We then evaluated classifiers developed using 3 approaches: (1) a convolutional neural network (CNN) trained for the task, (2) classifiers trained on non-semantic speech representations from CNNs that used an unsupervised objective [1], and (3) classifiers trained on the acoustic (encoder) embeddings from an ASR system trained on typical speech [2]. We found that the ASR encoder's embeddings considerably outperform the other two on detecting and classifying disordered speech. Further analysis shows that the ASR embeddings cluster speech by the spoken phrase, while the non-semantic embeddings cluster speech by speaker. Also, longer phrases are more indicative of intelligibility deficits than single words.

Analysis and Tuning of a Voice Assistant System for Dysfluent Speech

Vikramjit Mitra, Zifang Huang, Colin Lea, Lauren Tooley, Sarah Wu, Darren Botten, Ashwini Palekar, Shrinath Thelapurath, Panayiotis Georgiou, Sachin Kajarekar, Jefferey Bigham; Apple, USA Fri-A-SS-2-15, Time: 16:25

Dysfluencies and variations in speech pronunciation can severely degrade speech recognition performance, and for many individuals with moderate-to-severe speech disorders, voice operated systems do not work. Current speech recognition systems are trained primarily with data from fluent speakers and as a consequence do not generalize well to speech with dysfluencies such as sound or word repetitions, sound prolongations, or audible blocks. The focus of this work is on quantitative analysis of a consumer speech recognition system on individuals who stutter and production-oriented approaches for improving performance for common voice assistant tasks (i.e., "what is the weather?"). At baseline, this system introduces a significant number of insertion and substitution errors resulting in intended speech Word Error Rates (isWER) that are 13.64% worse (absolute) for indi-

NOTES

viduals with fluency disorders. We show that by simply tuning the decoding parameters in an existing hybrid speech recognition system one can improve isWER by 24% (relative) for individuals with fluency disorders. Tuning these parameters translates to 3.6% better domain recognition and 1.7% better intent recognition relative to the default setup for the 18 study participants across all stuttering severities.

Group Discussion

Time: 17:35

Fri-A-S&T-1 : Show and Tell 4 16:00-18:00, Friday 3 September 2021

Interactive and Real-Time Acoustic Measurement Tools for Speech Data Acquisition and Presentation: Application of an Extended Member of Time Stretched Pulses

Hideki Kawahara¹, Kohei Yatabe², Ken-Ichi Sakakibara³, Mitsunori Mizumachi⁴, Masanori Morise⁵, Hideki Banno⁶, Toshio Irino¹; ¹Wakayama University, Japan; ²Waseda University, Japan; ³HSUH, Japan; ⁴Kyutech, Japan; ⁵Meiji University, Japan; ⁶Meijo University, Japan Fri-A-S&T-1-1, Time: 16:00

Objective measurements of speech data acquisition and presentation processes are crucial for assuring reproducibility and reusability of experimental results and acquired materials. We introduce setting and measurement examples of those conditions using an interactive and real-time acoustic measurement tool based on an extended timestretched pulse. We also introduce supporting tools.

Save Your Voice: Voice Banking and TTS for Anyone

Daniel Tihelka, Markéta Řezáčková, Martin Grůber, Zdeněk Hanzlíček, Jakub Vít, Jindřich Matoušek; University of West Bohemia, Czechia Fri-A-S&T-1-2, Time: 16:00

The paper describes the process of automatic building of a personalized TTS system. The system was primarily developed for people facing the threat of voice loss; however, it can be used by anyone who wants to save his/her voice for any reason. Regarding the target group of users, the whole system is designed to be as simple to use as possible while still being fully autonomous.

NeMo (Inverse) Text Normalization: From Development to Production

Yang Zhang, Evelina Bakhturina, Boris Ginsburg; NVIDIA, USA

Fri-A-S&T-1-3, Time: 16:00

We introduce the NeMo Text Processing (NTP) toolkit — an opensource Python library for text normalization (TN) and inverse text normalization (ITN) based on weighted finite-state transducers (WF-STs). The English grammars provided within NTP can be seamlessly deployed to the C++ Sparrowhawk framework for production.

Lalilo: A Reading Assistant for Children Featuring Speech Recognition-Based Reading Mistake Detection

Corentin Hembise, Lucile Gelin, Morgane Daniel; Lalilo, France

Fri-A-S&T-1-4, Time: 16:00

Lalilo is a reading assistant intended to help kindergarten to second grade students to master their reading skills. Students progress at their own pace thanks to an adaptive learning system that differentiates instructions. Teachers can access data on their students' progression. Among other exercises, a read-aloud exercise is provided for students to practice their reading. This exercise uses a reading mistake detection system based on speech recognition to offer automatic feedback on the child's reading. Since speech recognition on children learning to read is highly challenging, we overcome potential inaccurate thus damageable feedback with an uncertainty estimation leading to a neutral feedback.

Automatic Radiology Report Editing Through Voice

Manh Hung Nguyen¹, Vu Hoang¹, Tu Anh Nguyen¹, Trung H. Bui²; ¹VinBrain, Vietnam; ²Independent Researcher, USA Fri-A-S&T-1-5, Time: 16:00

We present a system that allows radiologists to edit the radiology report through their voices. This is a function in our bigger system at VinBrain LLC that uses AI algorithms to assist radiologists with chest x-ray diagnosis, the system can suggest the abnormalities, then bases on the radiologist's confirmations or conclusions to automatically generate the report using predefined templates. We then allow the radiologist to freely edit the report using voice. The system combines two components, the first is the Speech Recognition System (SRS), and the second is the Natural Language Understanding System (NLUS) that executes the user's command. The user can delete, modify or add an arbitrary whole sentence. In addition, we successfully developed an SRS for such a non-mainstream language as Vietnamese and adapted it for the radiology domain.

WittyKiddy: Multilingual Spoken Language Learning for Kids

Ke Shi, Kye Min Tan, Huayun Zhang, Siti Umairah Md. Salleh, Shikang Ni, Nancy F. Chen; A*STAR, Singapore

Fri-A-S&T-1-6, Time: 16:00

We present *WittyKiddy*, a spoken language learning system for children, developed at the Institute for Infocomm Research (I2R), A*STAR, Singapore. Our system automatically evaluates a student's oral proficiency by scoring pronunciation, fluency and intonation of a spoken utterance. We demonstrate the technical capabilities of the system via reading aloud exercises and oral cloze tests in English and Malay. Both quantitative and qualitative feedback are given to the student. Our work helps support multilingual education for children.

Duplex Conversation in Outbound Agent System

Chunxiang Jin, Minghui Yang, Zujie Wen; Ant, China Fri-A-S&T-1-7, Time: 16:00

Intelligent outbound is a popular way to contact customers. The traditional outbound agents communicate with users in a simplex way. The user and the agent cannot speak at the same time, and the user cannot actively interrupt the conversation while the agent is playing audio generated by TTS. The traditional solution is based on the output of the VAD module, once the user voice is detected, the agent will immediately stop talking. However, the user sometimes expresses

NOTES

the short answer at will, not to interrupt the agent, and it will cause the agent to be frequently interrupted. In addition, when users say named entity nouns(numbers, locations, company names, etc), their speech speed is slow and the pause time between words is longer, and they may be interrupted by the agent unreasonably. We propose a method to identify user's interruption requests and discontinuous expressions by analyzing the semantic information of the user's utterance. As a result, fluency of the dialogue is improved.

Web Interface for Estimating Articulatory Movements in Speech Production from Acoustics and Text

Sathvik Udupa, Anwesha Roy, Abhayjeet Singh, Aravind Illa, Prasanta Kumar Ghosh; Indian Institute of Science, India

Fri-A-S&T-1-8, Time: 16:00

We release a web interface to visualise estimated articulatory movements in speech production from different modalities — acoustics and text. We allow the use of various trained models for this purpose. This tool also serves the purpose of comparing the predicted articulatory movements from different modalities and visually understanding the effect of noise in speech.

Author Index

Α

A., Arun Kumar	Wed-E-V-1-6	161
Abad, Alberto	Tue-M-SS-1-2	77
	Wed-E-V-2-9	164
Abarı, Kalman	Wed-E-V-4-7	169
Abdalali Ahmad	Inu-A-V-5-7	162
AbdelHady Mohamed	Fri-M-V-6-2	234
Abdullah. Badr M.	Fri-M-V-6-1	234
Abe, Masanobu	Tue-M-V-2-4	65
Abhishek	Wed-A-V-1-1	141
Abplanalp, Samuel	Fri-A-O-1-4	241
Abraham, Ajish K	Tue-A-V-2-4	85
Accou Bernd	Tue-M-V-3-4	81
Acero, Alex	Tue-E-V-4-4	111
Adda-Decker, Martine	Tue-E-O-2-2	101
Adi, Yossi	Thu-A-V-5-6	210
Adya, Saurabh	Fri-M-V-6-4	234
Alshan, Amper	Fri-M-V-3-11	162
	Wed-E-V-1-9	171
Ahlqvist-Björkroth, Sari	Thu-A-V-1-3	200
Ahmed, Beena	Thu-M-V-1-6	181
	Thu-A-V-6-7	212
Ahuja, Narendra	Tue-A-SS-2-6	97
Alchner, Robert	Wed-E-SS-I-I	255
Akata, Zevnen	Wed-F-O-3-5	160
Akbacak, Murat	Tue-M-V-4-1	69
Akeroyd, Michael	Tue-A-V-3-7	88
Akula, Jayaprakash	Wed-A-V-1-1	141
Alameda-Pineda, Xavier	Tue-M-O-3-1	62
Albanie, Samuel	Wed-E-O-3-5	1/13
Aldhafvan, Rawan	Wed-M-V-5-1	130
Aldholmi, Yahya	Wed-M-V-5-1	130
Alenin, Alexander	Wed-A-SS-1-1	155
Al-Ghezi, Ragheb	Wed-M-O-1-3	119
Ali, Ahmed	Wed-E-V-1-11	162
Ali Avesha	Fri-M-V-2-11	204
Ali, Murtiza	Thu-M-V-3-2	185
Alisamir, Sina	Wed-M-O-1-5	119
Alison, Laurence	Wed-A-V-1-2	141
Allauzen, Alexandre	Wed-M-O-1-5	119
Allauzen, Cyrli	Wed-M-V-6-4	228
Aloshban. Nujud	Wed-E-V-2-5	163
Aloufi, Ranya	Tue-A-SS-1-2	95
Alqahtani, Asma	Wed-M-V-5-1	130
Al-Radhi, Mohammed S.	Wed-A-V-5-8	151
Alsofyani, Huda	Wed-E-V-2-4	163
Alumäe Tanel	Fri-M-SS-1-6	240
Alwan, Abeer	Tue-E-V-5-3	112
	Thu-A-SS-1-1	214
Alyuz, Nese	Wed-E-V-2-10	164
Ambikairajah, E	Thu-A-V-6-7	212
Amin Preet P	Fri-A-V-2-5 Wed-M-V-4-11	129
Amiriparian, Shahin	Tue-M-SS-1-1	77
An, Keyu	Wed-A-V-3-6	146
An, Xiaochun	Fri-A-V-5-10	253
Ananthanarayana, T	Wed-A-V-6-13	155
Anastasopoulos, A	I Ue-M-V-5-2	124
Alluo, Atsusiii	Thu-M-V-5-4	189
Ando, Shintaro	Fri-A-V-1-6	244
André, Elisabeth	Tue-M-SS-1-12	79
	Thu-M-V-5-8	190
André-Obrecht Pégine	Fri-M-0-2-2 Wed-A-0-2-2	220 130
Andrés-Ferrer. Jesús	Wed-A-V-2-7	143
Andrusenko, Andrei	Fri-M-V-3-2	227
Anidjar, Or Haim	Thu-M-V-4-9	188
Annand, Colin T	Tue-A-V-2-7	86
Anvar, Aria	Fri-A-SS-2-2	257 118
Arai, Kenichi	Tue-M-V-3-5	67
Arai, Takayuki	Tue-A-S&T-1-3	<u>99</u>
· •	Thu-M-V-6-2	190
	Thu-M-V-6-4	191

Araki, Shoko	Tue-M-V-3-5 67 Wed-A-V-4-2 148
Arava Radhika	Eri.M.V.6.2 234
Arciuli Ioanne	Thu-A-V-6-7 212
Arefin, Md. Rifat	Thu-A-V-3-7 206
Arias-Vergara, T	Tue-M-SS-1-3 77
A	Thu-A-SS-2-2 217
Arivazhagan, Naveen	Wed-A-V-6-2 152
Arnrich Bert	Tue-E-0-3-3 102
Arora, Abhinav	Wed-A-V-2-1 142
Arora, Ashish	Thu-M-O-3-6 179
Arora, Siddhant	Tue-E-V-4-10 112
Arsikere, Harish	Thu-A-V-2-7 203
Ashihara, Takanori	Wed-M-V-5-12 132
	Wed-M-V-6-6 134
	Fri-M-V-3-6 228
Ashton, Elaine	Thu-M-V-1-8 181
Astanov Sorgoi	Wed-E-V-2-10 164
Astésano Corine	Wed-A-0-3-2 139
Atchayaram, Nalini	Thu-M-V-1-11 182
Audhkhasi, Kartik	Wed-M-V-2-12 126
	Wed-M-V-6-11 135
Audibort Nicolas	Wed-E-V-3-7 166
Auli. Michael	Tue-A-V-4-4 89
,	Wed-A-V-6-3 153
	Wed-E-V-1-3 161
Avgustinova, Tania	Fri-M-V-1-8 224
Avila Anderson R	Tue-M-V-5-10 72
Avila, Flavio	Tue-A-SS-2-11 98
Awasthi, Abhijeet	Fri-M-V-6-13 236
Azaria, Amos	Fri-M-V-5-1 231
В	
Baas, Matthew	Wed-M-V-2-6 124
BabaAli, Bagher	Tue-M-O-2-4 61
Bacchiani Michiel A II	Wed-A-V-3-11 147
Bäckström, Tom	Tue-A-V-3-5 87
	Thu-A-O-2-1 199
Badino, Leonardo	Tue-M-O-2-5 61
Bae, Handin	
Bae, Jae-Sung	Tue-M-V-2-2 64
	Wed-A-V-5-6 151
D 1. 41 .	Thu-A-V-5-5 209
Baevski, Alexei	Tue-A-V-4-4 89
	Wed-E-V-1-3 161
Baghel, Shikha	Wed-M-V-1-8 123
D . W	Fri-M-V-5-10 233
Ваі, Ye	Tue-M-V-5-1 70
	Wed-M-V-4-5 128
	Fri-M-V-3-1 227
Bailly, Gérard	Tue-M-O-1-3 60
Baird, Alice	Tue-M-SS-1-1 77
Bak Taeiun	Tue-M-V-2-2 64
bail, racjuir	Wed-A-V-5-6 151
	Thu-A-V-5-5 209
Baker, Elise	Thu-A-V-6-7 212
Bakhturina, Evelina	Wed-E-V-6-8 1/4
	Fri-A-S&T-1-3 260
Balagopalan, Aparna	Thu-A-SS-2-5 217
Balam, Jagadeesh	Wed-M-O-1-4 119
Baldridge, Jason	Thu-M-V-2-2 182
Bali, Kalika	Wed-E-V-1-7 161
Ballard, Kirrie J	Thu-M-V-1-6 181
B 111 1 1	Thu-A-V-6-7 212
Ballier, Nicolas	Tue-M-V-7-1 75
Bamman, David	Wed-E-V-4-3 108 Wed-E-V-2-2 163
Banbury, Colby	Fri-M-V-6-5 234
Bando, Yoshiaki	Tue-A-V-3-2 87
Bandon, John	Wed-A-O-3-1 139
Banno Hideki	Tue-M-SS-T-9 78 Fri-A-S&T-1-1 260
Baquero-Arnal, Pau	Thu-A-V-6-10 213
Barbier Florentin	Thu M SS 2 1 105
buibici, morenum	Thu-M-55-2-1 195

	14671137	00
	Tue-A-V-3-8	88
	Thu-M-V-1-8	181
	Thu-A-V-3-4	205
Barnes, Megan	Tue-E-V-4-4	111
Barra-Chicote Roberto	Tue-A-SS-1-9	96
barra enicote, Roberto .		120
	wea-w-v-4-10	129
	I hu-M-V-5-2	189
	Thu-M-V-5-6	189
	Fri-M-V-1-2	222
Barriuso Tavlor Anne	Wed-A-V-1-4	141
Partl Dolormy Vatrin D		102
Bai li-POKOIIIy, Kati III D.	Tue-E-O-3-3	102
Batliner, Anton	Tue-M-SS-1-1	- 77
Batsis, John A.	Thu-A-SS-2-3	217
Baumann. Timo	Thu-M-S&T-1-5	197
Baverl S P	Tue-M-SS-1-3	77
buyen, our minimum		217
Decels Coursels	1110-A-33-2-2	120
Beack, Seungkwon	wed-M-V-4-13	130
Bear, Helen L	Tue-A-V-1-9	- 83
Beaufays, Françoise	Tue-E-V-5-4	113
Behlau Mara	Tue-E-SS-1-7	118
Beiser David		110
Polioov Stopiolov		215
Bellaev, Stallislav	Inu-A-SS-1-10	215
Bell, Peter	Tue-M-V-5-3	71
	Tue-A-O-1-1	- 80
	Tue-A-V-4-3	89
	Wed-F-O-2-4	159
	Wod_E V 2 2	165
		170
D-1 - 17- 1 1	т nu-м-О-3-1	1/0
Belur, Yamini	Thu-M-V-1-11	182
Ben Amor, Imen	Tue-E-O-2-6	102
Benders. Titia	Thu-A-V-6-7	212
Reneš Karel	Wed-A-0-1-3	138
Ponetos Emmanouil		100
benetos, Emmanoum	Tue-A-V-1-9	00
	I hu-A-V-3-8	206
Bennasar Vázquez, J	Thu-A-V-3-5	205
Bensidi-Slimane, Y	Wed-A-V-1-7	142
Berdugo Baruch	Fri-Δ-SS-1-6	256
Borg Avol		226
beig, Axei	Fri-Wi-V-0-12	230
Bergler, Christian	Tue-M-SS-1-1	- 7 7
	Wed-E-O-3-2	159
Beringer, Grzegorz	Thu-M-V-5-2	189
8, - 8	Eri.M.V.1.2	222
Poricha Vicar	Tuo M V 2 12	68
berisila, visai		1 4 0
D 1.14 1.1	wed-A-O-3-4	140
Bernard, Mathieu	Wed-M-V-2-10	125
-		
Berry, James D	Fri-A-SS-2-2	257
Berry, James D Besacier, Laurent	Fri-A-SS-2-2 Wed-M-O-1-5	257 119
Berry, James D Besacier, Laurent	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1	257 119 158
Berry, James D Besacier, Laurent	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1	257 119 158 220
Berry, James D Besacier, Laurent	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1	257 119 158 220
Berry, James D Besacier, Laurent Beskow, Jonas	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-0-2-1 Wed-A-S&T-1-4	257 119 158 220 157
Berry, James D Besacier, Laurent Beskow, Jonas Bhandari, Nishchal	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9	257 119 158 220 157 204
Berry, James D Besacier, Laurent Beskow, Jonas Bhandari, Nishchal Bharadwaj, Abhinav	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7	257 119 158 220 157 204 174
Berry, James D Besacier, Laurent Beskow, Jonas Bhandari, Nishchal Bharadwaj, Abhinav Bharadwai, Samarth	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-M-V-2-1	257 119 158 220 157 204 174 123
Berry, James D Besacier, Laurent Beskow, Jonas Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7	257 119 158 220 157 204 174 123 161
Berry, James D Besacier, Laurent Beskow, Jonas Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Pharai Cauray	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fe: M V/E 11	257 119 158 220 157 204 174 123 161
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gauray	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-M-V-2-1 Wed-E-V-1-7 Fri-M-V-5-11	257 119 158 220 157 204 174 123 161 233
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-M-V-2-1 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1	257 119 158 220 157 204 174 123 161 233 96
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10	257 119 158 220 157 204 174 123 161 233 96 74
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhati, Shrirama Bhati, Saurabhchand Bhatiacharjee, Mrinmoy	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-V-6-10 Fri-M-V-5-10	$\begin{array}{c} 257\\ 119\\ 158\\ 220\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 96\\ 74\\ 233\\ \end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka .	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-M-V-2-1 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Thu-M-V-1-11	$\begin{array}{c} 257\\ 119\\ 158\\ 220\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 96\\ 74\\ 233\\ 182\\ \end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavar Arnan	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-M-V-2-1 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Thu-M-V-1-11 Tue-A-V-2-10	257 119 158 220 157 204 174 123 161 233 96 74 233 182 86
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale Swappil	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-10 Thu-M-V-1-11 Tue-A-SS-2-10	257 119 158 220 157 204 174 123 161 233 96 74 233 182 86 00
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhati, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Bioday Eggi	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-V-5-10 Thu-M-V-1-11 Tue-A-V-2-10 Tue-A-SS-2-10 C	$\begin{array}{c} 257\\ 119\\ 158\\ 220\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 96\\ 74\\ 233\\ 182\\ 86\\ 98\\ 952\\ \end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-M-V-2-1 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Thu-M-V-1-11 Tue-A-V-2-10 Tue-A-SS-2-11 Fri-A-SS-2-11	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 96\\ 74\\ 233\\ 182\\ 28\\ 86\\ 98\\ 259\\ 259\\ 259\\ 259\\ 259\\ 259\\ 259\\ 259$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Bie, Xiaoyu	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Fri-M-V-5-11 Tue-A-SS-2-11 Thu-M-V-1-11 Tue-A-V-2-10 Thu-M-V-1-11 Tue-A-SS-2-11 Tue-M-O-3-1	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 162\\ 233\\ 182\\ 86\\ 98\\ 259\\ 62\\ \end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhati, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-V-6-10 Fri-M-V-5-10 Thu-M-V-1-11 Tue-A-V-2-10 Thu-A-SS-2-10 Fri-A-SS-2-11 Tue-M-O-3-11 Thu-M-S&T-1-1	2577 119 158 2200 1577 2044 1233 161 2333 966 744 2333 1822 866 98 259 622 196
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-M-V-2-1 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-5-10 Thu-M-V-5-10 Thu-M-V-2-10 Tue-A-SS-2-10 Fri-A-SS-2-11 Tue-A-SS-2-11 Tue-M-O-3-1 Thu-M-S&T-1-5	2577 119 158 2200 1577 2044 123 161 2333 96 74 2333 1822 86 98 9259 62 259 62 196
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Birtham Lefferey	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Fri-M-V-5-11 Tue-A-SS-2-11 Tue-A-V-5-10 Thu-M-V-5-10 Thu-M-V-5-10 Thu-M-V-5-10 Thu-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Tue-A-SS-2-11 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Eri-A-SS-2-15	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 96\\ 74\\ 233\\ 182\\ 86\\ 98\\ 259\\ 62\\ 196\\ 62\\ 197\\ 260\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Bigham, Jefferey	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Tue-A-SS-2-11 Tue-M-O-3-11 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Fri-A-SS-2-15 Fri-A-SS-2-15	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 96\\ 74\\ 233\\ 182\\ 259\\ 62\\ 259\\ 62\\ 196\\ 197\\ 260\\ 260\\ 202\\ 202\\ 202\\ 202\\ 202\\ 202$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Bigham, Jefferey	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-V-5-10 Fri-M-V-5-10 Fri-M-V-5-10 Thu-M-V-1-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 96\\ 74\\ 233\\ 86\\ 98\\ 259\\ 62\\ 196\\ 197\\ 260\\ 203\\ 203\\ 203\\ 203\\ 203\\ 203\\ 203\\ 20$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Billa, Jayadev	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Thu-M-V-5-10 Thu-M-V-5-10 Thu-M-V-5-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Tue-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11	$\begin{array}{c} 2577\\ 119\\ 158\\ 220\\ 157\\ 204\\ 174\\ 123\\ 96\\ 74\\ 233\\ 182\\ 259\\ 62\\ 196\\ 62\\ 197\\ 260\\ 203\\ 167\\ \end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Bilgi, Raghavendra Billa, Jayadev Billington, Rosey	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 3161\\ 233\\ 366\\ 744\\ 233\\ 182\\ 259\\ 626\\ 197\\ 260\\ 203\\ 167\\ 227\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhatsar, Arpan Bhatsar, Arpan Bhasale, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Billgi, Raghavendra Billington, Rosey Birkholz, Peter	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Thu-M-V-1-11 Tue-A-V-2-10 Tue-A-SS-2-11 Tue-M-O-3-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-2-6	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 123\\ 161\\ 233\\ 96\\ 74\\ 233\\ 182\\ 259\\ 62\\ 196\\ 197\\ 260\\ 00\\ 203\\ 167\\ 227\\ 191\\ 1\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Bilgi, Raghavendra Billa, Jayadev Birkholz, Peter	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Thu-M-V-6-10 Fri-A-SS-2-10 Thu-M-S&T-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-1 SFri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-2-14 Thu-M-S&T-1-6 Thu-M-S&T-1-6	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 1577\\ 204\\ 174\\ 123\\ 161\\ 2333\\ 182\\ 233\\ 182\\ 259\\ 62\\ 196\\ 197\\ 260\\ 203\\ 167\\ 227\\ 191\\ 191\\ 196\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Billa, Jayadev Billa, Jayadev Birkholz, Peter	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Thu-M-V-6-10 Fri-M-V-5-10 Thu-M-V-5-10 Thu-M-V-5-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Tue-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-6-6 Thu-M-V-6-12	$\begin{array}{c} 2577\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 123\\ 161\\ 233\\ 161\\ 233\\ 161\\ 233\\ 161\\ 233\\ 161\\ 233\\ 161\\ 233\\ 161\\ 233\\ 161\\ 233\\ 161\\ 233\\ 162\\ 233\\ 167\\ 227\\ 196\\ 197\\ 227\\ 191\\ 196\\ 199\\ 199\\ 199\\ 199\\ 199\\ 199$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhatta	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-5-11 Tue-A-SS-2-10 Fri-M-V-5-10 Thu-M-V-1-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-S&T-1-2 Thu-M-S&T-1-2 Thu-M-S&T-1-2 Thu-M-S&T-1-2	$\begin{array}{c} 2577\\ 119\\ 1588\\ 2200\\ 1577\\ 204\\ 174\\ 1233\\ 161\\ 1233\\ 96\\ 74\\ 2333\\ 1822\\ 866\\ 988\\ 259\\ 622\\ 1966\\ 197\\ 2600\\ 2033\\ 167\\ 2277\\ 191\\ 1966\\ 1988\\ 248\\ 196\\ 1988\\ 248\\ 196\\ 1988\\ 248\\ 196\\ 1988\\ 248\\ 196\\ 1988\\ 248\\ 198\\ 248\\ 198\\ 248\\ 198\\ 198\\ 248\\ 198\\ 248\\ 198\\ 198\\ 198\\ 198\\ 198\\ 198\\ 198\\ 19$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bhara, Gaurav Bhati, Saurabhchand Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Biosale, Swapnil Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Billa, Jayadev Billington, Rosey Birkholz, Peter Bishop, Somer	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-2-14 Thu-M-V-2-12 Thu-M-V-1-2	$\begin{array}{c} 2577119\\ 15882200\\ 157720\\ 1744123\\ 161\\ 233396\\ 744\\ 2333\\ 1822\\ 866\\ 988\\ 2599\\ 622\\ 196\\ 197\\ 2600\\ 203\\ 167\\ 227\\ 191\\ 198\\ 180\\ 260\\ 203\\ 167\\ 227\\ 191\\ 198\\ 180\\ 203\\ 227\\ 191\\ 198\\ 180\\ 203\\ 227\\ 227\\ 191\\ 198\\ 180\\ 203\\ 203\\ 227\\ 227\\ 191\\ 198\\ 180\\ 203\\ 203\\ 203\\ 203\\ 203\\ 203\\ 203\\ 20$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Bilgi, Raghavendra Billa, Jayadev Billa, Jayadev Birkholz, Peter Bisshop, Somer Bisshop, Somer	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Thu-M-V-6-10 Fri-A-SS-2-10 Thu-M-V-1-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-15 Thu-M-S&T-1-15 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-2-12 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2	$\begin{array}{c} 2577\\ 119\\ 1588\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 161\\ 2333\\ 966\\ 744\\ 2333\\ 1822\\ 866\\ 988\\ 2599\\ 622\\ 1967\\ 2203\\ 167\\ 2277\\ 191\\ 196\\ 198\\ 180\\ 101 \end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharadwaj, Samarth Bhata, Shrirama Bhati, Saurabhchand Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Biadsy, Fadi Biemann, Chris Bigham, Jefferey Billa, Jayadev Billa, Jayadev Billington, Rosey Bishop, Somer Bissiri, Maria Paola Biswas, Ritujoy	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Thu-M-S&T-1-5 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-V-2-7 Wed-E-V-3-31 Fri-M-V-2-14 Thu-M-V-6-6 Thu-M-V-1-2 Thu-E-O-2-3 Tue-M-V-3-4	$\begin{array}{c} 257\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 161\\ 233\\ 182\\ 86\\ 98\\ 98\\ 259\\ 62\\ 196\\ 197\\ 2600\\ 203\\ 167\\ 191\\ 196\\ 198\\ 180\\ 011\\ 67\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Billa, Jayadev Billington, Rosey Birkholz, Peter Bisshop, Somer Bissiri, Maria Paola Biswas, Ritujoy Black, Alan W	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Thu-M-V-1-11 Tue-A-V-2-70 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-V-2-74 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-6-1-2 Thu-M-V-1-2 Thu-M-V-1-2 Tue-E-O-2-3 Tue-M-V-3-4 Tue-A-V-5-34	$\begin{array}{c} 2577\\ 119\\ 1588\\ 2200\\ 1577\\ 2204\\ 174\\ 123\\ 161\\ 233\\ 182\\ 259\\ 622\\ 196\\ 197\\ 260\\ 203\\ 167\\ 227\\ 191\\ 196\\ 198\\ 180\\ 101\\ 67\\ 92\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Biadsy, Fadi Biemann, Chris Bigham, Jefferey Bigham, Jefferey Bigli, Raghavendra Billa, Jayadev Billington, Rosey Birkholz, Peter Bissiri, Maria Paola Biswas, Ritujoy Black, Alan W	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Fri-M-V-5-10 Thu-M-V-1-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-6-6 Thu-M-V-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-V-1-2 Tue-E-O-2-3 Tue-A-V-5-8 Tue-E-V-3-4	$\begin{array}{c} 2577119\\ 15882200\\ 1577204\\ 1744123\\ 161\\ 23339\\ 744\\ 2333\\ 1822\\ 866\\ 988\\ 2599\\ 622\\ 196\\ 197\\ 2600\\ 203\\ 1677\\ 191\\ 196\\ 180\\ 101\\ 677\\ 921\\ 221\\ 112\\ 112\\ 112\\ 112\\ 112\\ 112$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharaj, Gaurav Bharadwaj, Samarth Bharadwaj, Samarth Bishop, Samer Biskop, Somer Biswas, Ritujoy Black, Alan W	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-SS-2-10 Thu-M-V-6-10 Fri-A-SS-2-10 Thu-M-V-5-10 Thu-A-V-2-10 Tue-A-SS-2-10 Fri-A-SS-2-11 Tue-A-SS-2-15 Fri-A-SS-2-15 Fri-A-SS-2-15 Fri-A-SS-2-15 Fri-A-SS-2-15 Fri-A-SS-2-15 Fri-A-SS-2-15 Thu-M-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-6-6 Thu-M-V-1-2 Thu-A-O-12 Thu-A-V-5-8 Tue-E-O-2-3 Tue-E-V-3-4 Tue-E-V-3-4	$\begin{array}{c} 2577119\\ 1582200\\ 157720\\ 174\\ 123\\ 161\\ 233\\ 96\\ 744\\ 233\\ 182\\ 86\\ 98\\ 259\\ 62\\ 233\\ 182\\ 259\\ 62\\ 203\\ 197\\ 197\\ 196\\ 197\\ 227\\ 227\\ 191\\ 196\\ 198\\ 180\\ 101\\ 67\\ 922\\ 112\\ 262\\ 203\\ 101\\ 101\\ 67\\ 102\\ 112\\ 262\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102\\ 10$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharaj, Gaurav Bhara, Samarth Bharaj, Gaurav Bharadwaj, Samarth Bharadwaj, Samarth Bhavsar, Arpan Bhavsar, Arpan Bhavsar, Arpan Bhavsar, Arpan Bhavsar, Arpan Bhavsar, Arpan Bishap, Fadi Bishop, Somer Bissiri, Maria Paola Biswas, Ritujoy Black, Alan W	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-5-11 Tue-A-SS-2-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-V-1-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-S&T-1-2 Thu-M-V-6-6 Thu-M-V-6-6 Thu-M-S&T-1-2 Thu-A-V-2-3 Tue-K-0-3-3 Tue-K-V-3-4 Tue-K-V-3-8 Tue-K-V-3-8 Tue-K-V-3-8 Tue-K-V-3-8	$\begin{array}{c} 257\\ 119\\ 158\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 96\\ 74\\ 233\\ 1826\\ 86\\ 98\\ 259\\ 62\\ 196\\ 197\\ 2600\\ 203\\ 167\\ 297\\ 191\\ 196\\ 198\\ 1800\\ 101\\ 67\\ 92\\ 112\\ 162\\ 112\\ 162\\ 112\\ 112\\ 112\\ 11$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhavsar, Arpan Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Billa, Jayadev Billa, Jayadev Billa, Jayadev Bishop, Somer Bissiri, Maria Paola Biswas, Ritujoy Black, Alan W	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Thu-M-V-1-11 Tue-A-V-2-70 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-S&T-1-2 Thu-M-V-2-14 Thu-M-V-2-12 Thu-M-V-2-12 Thu-M-V-1-2 Thu-M-V-3-4 Tue-A-V-5-8 Tue-E-V-3-10 Wed-E-V-1-10 Wed-E-V-1-10 Thu-M-V-1-7	$\begin{array}{c} 2577119\\ 15882200\\ 157720\\ 174123\\ 161\\ 233396\\ 744\\ 2333182\\ 868\\ 988\\ 2599\\ 622\\ 1966\\ 197\\ 2600\\ 2033\\ 1677\\ 2277\\ 191\\ 196\\ 198\\ 1800\\ 101\\ 167\\ 921\\ 122\\ 162\\ 181\\ 122\\ 162\\ 182\\ 122\\ 162\\ 182\\ 122\\ 162\\ 182\\ 122\\ 162\\ 182\\ 122\\ 162\\ 182\\ 122\\ 162\\ 182\\ 182\\ 182\\ 182\\ 182\\ 182\\ 182\\ 18$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Bie, Xiaoyu Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Billa, Jayadev Billa, Jayadev Billa, Jayadev Bishop, Somer Bissiri, Maria Paola Bisswas, Ritujoy Black, Alan W	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-V-5-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-6-6 Thu-M-V-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-0-1-2 Thu-A-0-1-2 Thu-A-0-1-2 Thu-A-0-1-2 Thu-A-V-3-4 Tue-E-V-4-10 Wed-E-V-1-10 Thu-M-V-1-7 Thu-A-SS-2-7	$\begin{array}{c} 2577119\\ 15882200\\ 1577204\\ 174123\\ 161\\ 2333\\ 964\\ 2333\\ 1822\\ 866\\ 989\\ 2259\\ 626\\ 1967\\ 2600\\ 2033\\ 1677\\ 2277\\ 191\\ 196\\ 180\\ 101\\ 677\\ 92\\ 112\\ 162\\ 181\\ 218\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bhara, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhavsar, Arpan Bhosale, Swapnil Biosale, Swapnil Biosale, Swapnil Biedsy, Fadi Biemann, Chris Bigham, Jefferey Bigham, Jefferey Billa, Jayadev Billa, Jayadev Billa, Jayadev Bishop, Somer Bissiri, Maria Paola Bissor, Alan W Blackburn, Daniel Blandin, Rémi	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Thu-M-V-1-11 Tue-A-V-2-10 Thu-M-S&T-1-5 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-S-2-15 Thu-A-S-2-12 Thu-A-S-1-2 Thu-A-V-1-2 Tue-E-V-3-11 Fri-M-V-2-14 Thu-M-S&T-1-2 Thu-A-O-1-2 Thu-A-SS-2-7 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2	$\begin{array}{c} 2577\\ 119\\ 1588\\ 2200\\ 1577\\ 204\\ 174\\ 123\\ 161\\ 233\\ 182\\ 233\\ 1822\\ 866\\ 988\\ 259\\ 62\\ 196\\ 203\\ 167\\ 227\\ 191\\ 196\\ 198\\ 198\\ 100\\ 101\\ 67\\ 92\\ 112\\ 181\\ 198\\ 198\\ 198\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bhara, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka Bhattacharjee, Tanuka Bhattacharjee, Tanuka Bhattacharjee, Tanuka Biadsy, Fadi Bie, Xiaoyu Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Bilgi, Raghavendra Billa, Jayadev Billa, Jayadev Bishop, Somer Bisshop, Somer Bissiri, Maria Paola Bisswas, Ritujoy Black, Alan W Blackburn, Daniel Blandin, Rémi Blatt, Alexander	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Thu-M-V-1-11 Tue-A-V-2-10 Tue-A-SS-2-10 Fri-A-SS-2-11 Tue-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-6-6 Thu-M-V-2-14 Thu-M-V-6-6 Thu-M-V-2-14 Thu-M-V-6-12 Thu-A-V-5-8 Tue-E-V-3-10 Wed-E-V-1-10 Thu-M-V-1-7 Thu-A-SS-2-7 Thu-A-O-12 Thu-M-SS-2-7 Thu-A-O-12 Thu-M-SS-2-7	$\begin{array}{c} 2577119\\ 15882200\\ 15772204\\ 174123\\ 161\\ 233396\\ 744\\ 23331822\\ 886\\ 988\\ 2599\\ 622\\ 196\\ 197\\ 2600\\ 2033\\ 167\\ 2277\\ 227\\ 191\\ 196\\ 198\\ 180\\ 101\\ 67\\ 922\\ 112\\ 162\\ 181\\ 1218\\ 198\\ 196\end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bharaj, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Biemann, Chris Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Billa, Jayadev Billa, Jayadev Birkholz, Peter Bisswas, Ritujoy Biaswas, Ritujoy Blackburn, Daniel Blandin, Rémi Blatt, Alexander Block, Aleese	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-V-5-10 Fri-M-V-5-10 Fri-M-V-5-10 Fri-M-V-5-10 Thu-M-V-6-10 Fri-M-V-5-10 Thu-M-V-1-11 Tue-A-V-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-S&T-1-2 Thu-M-V-1-2 Thu-A-V-2-8 Tue-E-V-3-10 Wed-E-V-1-10 Wed-E-V-1-10 Thu-M-SS-2-7 Thu-A-SS-2-7 Thu-A-SS-2-7 Thu-A-SS-2-7 Thu-A-SS-2-7 Thu-A-SS-2-7 Thu-A-SS-2-5 Tue-A-O-2-2	$\begin{array}{c} 2577119\\ 15882200\\ 157719\\ 1744123\\ 161\\ 233396\\ 744\\ 2333\\ 1822\\ 866\\ 989\\ 626\\ 1967\\ 191\\ 196\\ 196\\ 197\\ 191\\ 196\\ 198\\ 180\\ 101\\ 677\\ 91\\ 122\\ 162\\ 181\\ 198\\ 196\\ 81\\ \end{array}$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharadwaj, Samarth Bharaj, Gaurav Bharaj, Gaurav Bharadwaj, Samarth Bharadwaj, Samarth Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Bishop, Somer Bissiri, Maria Paola Bisswas, Ritujoy Black, Alan W Blackburn, Daniel Blandin, Rémi Blatt, Alexander Block, Aleese Bocklet T	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Thu-M-V-6-10 Fri-M-V-5-10 Thu-M-V-1-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-S&T-1-5 Fri-A-SS-2-15 Thu-A-S&T-1-15 Thu-M-S&T-1-12 Thu-M-S&T-1-12 Thu-M-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-V-1-22 Thu-M-V-1-22 Thu-M-V-1-20 Thu-M-V-1-20 Thu-M-V-1-20 Thu-A-O-2-20 Thu-A-O-2-20 Thu-A-O-2-20	$\begin{array}{c} 2577\\ 119\\ 1588\\ 2200\\ 157\\ 204\\ 174\\ 123\\ 161\\ 2333\\ 966\\ 744\\ 2333\\ 1822\\ 86\\ 98\\ 98\\ 2599\\ 622\\ 196\\ 203\\ 167\\ 2277\\ 191\\ 196\\ 197\\ 260\\ 203\\ 167\\ 227\\ 191\\ 196\\ 181\\ 198\\ 198\\ 198\\ 198\\ 198\\ 198\\ 198$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharadwaj, Samarth Bhara, Gaurav Bhati, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhatsar, Arpan Bhatsar, Arpan Bhatsar, Swapnil Bhatsar, Arpan Bhatsar, Arpan Bhatsar, Swapnil Biadsy, Fadi Bie, Xiaoyu Biemann, Chris Bigham, Jefferey Bigham, Jefferey Billa, Jayadev Billington, Rosey Bishop, Somer Bissiri, Maria Paola Bisswas, Ritujoy Black, Alan W Blackburn, Daniel Blandin, Rémi Blatt, Alexander Bock, Aleese Bockett, T Bochanti Samari	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Thu-M-V-1-11 Tue-A-V-2-10 Tue-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-S&T-1-2 Thu-A-V-2-3 Tue-H-V-3-4 Thu-M-S&T-1-2 Thu-A-V-5-8 Tue-E-V-3-10 Wed-E-V-1-10 Thu-M-SS-2-7 Thu-A-O-1-2 Thu-M-SS-2-7 Thu-A-O-2-2 Thu-M-SS-2-5 Tue-A-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-O-2-2 Tue-M-SS-1-3 T5-5	$\begin{array}{c} 2577119\\ 15882200\\ 15772204\\ 174123\\ 1611233\\ 96744\\ 23331822\\ 598\\ 62259\\ 62259\\ 62259\\ 62263\\ 1966\\ 19702600\\ 2033167\\ 22772191\\ 1966\\ 1980\\ 1011\\ 677\\ 922\\ 11218\\ 1986\\ 811\\ 777\\ 777\\ 7112\\ 11218\\ 1966\\ 811\\ 777\\ 777\\ 777\\ 777\\ 777\\ 777\\ 777$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bhara, Gaurav Bhat, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhattacharjee, Tanuka . Bhatsar, Arpan Bhattacharjee, Tanuka . Bhavsar, Arpan Biadsy, Fadi Biemann, Chris Bigham, Jefferey Bigham, Jefferey Billa, Jayadev Billa, Jayadev Billa, Jayadev Bishop, Somer Bissiri, Maria Paola Bisswas, Ritujoy Black, Alan W Black, Alan W Blandin, Rémi Blatt, Alexander Block, Aleese Bocklet, T Bodapati, Sravan	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-V-1-11 Thu-M-S&T-1-5 Fri-A-SS-2-15 Thu-A-V-2-7 Wed-E-V-3-11 Fri-M-V-2-14 Thu-M-S&T-1-2 Thu-M-V-2-14 Thu-M-V-2-14 Thu-M-V-2-14 Thu-M-V-2-12 Thu-M-V-3-4 Tue-E-V-3-10 Wed-E-V-1-10 Thu-M-V-1-7 Thu-A-SS-2-7 Thu-A-SS-2-7 Thu-A-O-1-2 Thu-M-SS-2-5 Tue-A-O-2-2 Thu-M-SS-1-3 Tue-E-V-5-10	$\begin{array}{c} 2577119\\ 15882200\\ 157720\\ 174123\\ 161\\ 233396\\ 744\\ 2333182\\ 866\\ 988\\ 2599\\ 622\\ 1966\\ 197\\ 2200\\ 203\\ 1677\\ 227\\ 191\\ 196\\ 198\\ 180\\ 101\\ 67\\ 922\\ 181\\ 198\\ 198\\ 196\\ 81\\ 77\\ 114\\ 198\\ 196\\ 81\\ 77\\ 114\\ 198\\ 196\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102$
Berry, James D Besacier, Laurent Bhandari, Nishchal Bharadwaj, Abhinav Bharadwaj, Abhinav Bharadwaj, Samarth Bharadwaj, Samarth Bharaj, Gaurav Bhata, Shrirama Bhati, Saurabhchand Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Mrinmoy Bhattacharjee, Tanuka . Bhavsar, Arpan Bhosale, Swapnil Biadsy, Fadi Biemann, Chris Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Bigham, Jefferey Billington, Rosey Birkholz, Peter Bisswas, Ritujoy Black, Alan W Black, Alan W Blatt, Alexander Block, Aleese Bock, Aleese Bocklet, T Bodapati, Sravan	Fri-A-SS-2-2 Wed-M-O-1-5 Wed-E-O-2-1 Fri-M-O-2-1 Wed-A-S&T-1-4 Thu-A-V-2-9 Wed-E-V-6-7 Wed-E-V-6-7 Wed-E-V-1-7 Fri-M-V-5-11 Tue-A-SS-2-1 Tue-M-V-6-10 Fri-M-V-5-10 Fri-M-V-5-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-10 Fri-A-SS-2-11 Thu-M-S&T-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-1 Thu-A-S-2-15 Fri-A-SS-2-15 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-O-1-2 Thu-A-SS-2-7 Thu-A-SS-2-7 Thu-A-O-1-2	$\begin{array}{c} 2577\\ 1199\\ 1588\\ 2200\\ 1577\\ 204\\ 174\\ 123\\ 396\\ 744\\ 233\\ 182\\ 866\\ 989\\ 626\\ 203\\ 182\\ 2599\\ 626\\ 1967\\ 197\\ 191\\ 196\\ 180\\ 101\\ 677\\ 92\\ 112\\ 181\\ 218\\ 1986\\ 811\\ 77\\ 114\\ 1933\\ 196\\ 811\\ 77\\ 114\\ 193\\ 196\\ 196\\ 196\\ 196\\ 196\\ 196\\ 196\\ 196$

INTERSPEECH 2021 262 BRNO | CZECHIA

Pooddolvor Christoph		205
boeudeker, Chiristophi	I nu-A-V-3-3	203
Boes, Wim	Wed-E-O-3-3	160
Boggust, Angie	Wed-M-V-2-12	126
	Thu-M-V-2-8	184
Bojar Ondřej	WedEO22	158
Dogar, Ondreg	T 5020	100
Bonastre, Jean-François	Tue-E-O-2-6	102
	Wed-A-O-2-6	139
	Thu-A-V-4-8	208
Ronneau Anne	Tue-M-V-7-3	75
Pönninghoff Ronodikt		1/18
Bommighon, Benedikt .	wea-A-v-4-2	140
Bonomi, Cyntia	Fri-A-V-1-4	243
Boogmans, Fleur	Thu-M-V-1-1	180
Borgström, Bengt I.	Tue-F-V-1-1	103
Borsdorf Marvin	Wed MO22	121
		222
	Fri-M-O-3-4	222
Bose, Deboshree	Fri-A-V-2-5	246
Botelho, Catarina	Tue-M-SS-1-2	77
	Wed-E-V-2-9	164
Rotros Rami	Wed-M-V-6-4	133
botros, Rum		240
	Fri-A-V-3-8	240
Botteldooren, Dick	Tue-M-V-6-1	73
Botten, Darren	Fri-A-SS-2-15	260
Bouafif Mansali, Mariem	Tue-A-V-3-5	87
Bourlard, Hervé	Thu-M-O-1-6	178
		220
Doutton Charles	FII-M-55-1-2	10
Bouveyron, Charles	Wed-E-O-2-3	159
Boves, Lou	Tue-A-O-2-4	81
	Tue- <u>A-O-2-5</u>	81
	Тие-А-О-2-6	82
Povco Suzanno		86
	Tue-A-v-Z-7	100
Boyer, Florian	Wed-E-V-2-1	163
Boyle, David	Tue-A-SS-1-2	95
Bozza, Bruno	Wed-F-V-1-5	161
Braccia Arianna	Tue-M-0-2-5	61
Progo Otovio		101
	<u>Inu-M-V-2-10</u>	104
Brand, Sophie	Tue-A-O-2-6	82
Branje, Sweeney	Wed-M-O-2-4	120
Braun, Bettina	Wed-M-V-5-9	132
braang becchia minim	Wod-E-V-4-9	160
		103
D. C. L. M.	Fri-M-V-1-3	223
Braun, Sebastian	Wed-E-SS-1-1	175
	Fri-A-SS-1-1	255
Bredin. Hervé	Thu-M-V-4-7	188
Bremerman Jacob	Thu-Δ-V-6-2	211
Proppor Michael D		257
bielillei, Michael F	FII-A-33-2-1	257
	Fri-A-SS-2-10	259
	Fri-A-SS-2-12	259
	Fri-A-SS-2-14	260
Brickell, Tracev A.	Tue-M-O-2-1	60
Bright Rebecca	Thu.M.V.1.8	181
Dright, Rebecca		04
Brossier, Baptiste	Tue-A-55-1-1	94
Broughton, Samuel J	Tue-A-V-6-1	92
Brown, Chloë	Tue-M-SS-1-1	77
Bruce. Stuart	Tue-A-V-2-6	85
Brümmer Niko	Tue-F-V-1-5	103
Prungant Douglas S		60
Brungart, Douglas 5	Tue-M-0-2-1	200
Brusco, Pablo	I hu-A-V-I-I I	202
Brutti, Alessio	Fri-M-O-1-4	220
Bu, Hui	Wed-E-V-6-4	173
	Thu-A-V-6-4	212
Ru Suliang	Wed-E-V-5-12	172
Pugiulong Catinga		117
Buchuleac, Catilica	Tue-E-55-1-4	117
Buet, François	Fri-A-O-2-5	242
Bui, Trung H	Fri-A-S&T-1-5	261
Bulyko, Ivan	Tue-A-SS-1-3	95
Runnell H Timothy	Tue-E-V-3-8	100
Dungot Lukéž		100
burget, LUKas	T 1 0 1 2	00
	Tue-A-O-1-3	80
	Tue-A-O-1-3 Tue- <u>A-O-1-5</u>	80 80
	Tue-A-O-1-3 Tue-A-O-1-5 Wed- <u>A-O-1-3</u>	80 80 138
	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4	80 80 138 158
	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M O 2 5	80 80 138 158 179
Durko William	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5	80 80 138 158 179
Burke, William	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7	80 80 138 158 179 142
Burke, William Burnham, Denis	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7	80 80 138 158 179 142 212
Burke, William Burnham, Denis Burroni, Francesco	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13	80 80 138 158 179 142 212 170
Burke, William Burnham, Denis Burroni, Francesco Busso, Carlos	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13 Thu-M-O-2-2	80 80 138 158 179 142 212 170 178
Burke, William Burnham, Denis Burroni, Francesco Busso, Carlos	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13 Thu-M-O-2-2 Fri-A-O-1-4	80 80 138 158 179 142 212 170 178 241
Burke, William Burnham, Denis Burroni, Francesco Busso, Carlos	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13 Thu-M-O-2-2 Fri-A-O-1-4 Tue-M-SS 1-0	80 80 138 158 179 142 212 170 178 241 78
Burke, William Burnham, Denis Burroni, Francesco Busso, Carlos Butler, Duncan	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13 Thu-M-O-2-2 Fri-A-O-1-4 Tue-M-SS-1-9	80 80 138 158 179 142 212 170 178 241 78
Burke, William Burnham, Denis Burroni, Francesco Busso, Carlos Butler, Duncan Bütlow, Andreas	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13 Thu-M-O-2-2 Fri-A-O-1-4 Tue-M-SS-1-9 Wed-E-O-3-1	80 80 138 158 179 142 212 170 178 241 78 159
Burke, William Burnham, Denis Burroni, Francesco Busso, Carlos Butler, Duncan Bütow, Andreas Butt, Miriam	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13 Thu-M-O-2-2 Fri-A-O-1-4 Tue-M-SS-1-9 Wed-E-O-3-1 Wed-E-V-4-12	$\begin{array}{r} 80 \\ 80 \\ 138 \\ 158 \\ 179 \\ 142 \\ 212 \\ 170 \\ 178 \\ 241 \\ 78 \\ 159 \\ 170 \end{array}$
Burke, William Burnham, Denis Burroni, Francesco Busso, Carlos Butler, Duncan Bütow, Andreas Butt, Miriam Byrd, Dani	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13 Thu-M-O-2-2 Fri-A-O-1-4 Tue-M-SS-1-9 Wed-E-V-4-12 Fri-M-V-2-1	80 80 138 158 179 142 212 170 178 241 78 159 170 224
Burke, William Burnham, Denis Burroni, Francesco Busso, Carlos Butler, Duncan Bütow, Andreas Butt, Miriam Byrd, Dani Bvun. Joon	Tue-A-O-1-3 Tue-A-O-1-5 Wed-A-O-1-3 Wed-E-O-1-4 Thu-M-O-3-5 Wed-A-V-1-7 Thu-A-V-6-7 Wed-E-V-4-13 Thu-M-O-2-2 Fri-A-O-1-4 Tue-M-SS-1-9 Wed-E-O-3-1 Wed-E-O-3-1 Wed-E-V-4-12 Fri-M-V-2-1 Wed-M-V-2-1	$\begin{array}{r} 80 \\ 80 \\ 138 \\ 158 \\ 179 \\ 142 \\ 212 \\ 170 \\ 178 \\ 241 \\ 78 \\ 159 \\ 170 \\ 224 \\ 130 \end{array}$

C

C		
Cai Bin	Fri-A-V-4-7	250
Cai Cong	Wed-F-V-2-8	164
Cai Danwei	Tue-E-V-1-2	103
Cai Shanging	Eri.A.SS.2.10	259
Cai Xingvu	Eri. A.V. 2.7	246
Calamaro Shira	Eri. A.V.1.1	240
Cámbara Guillermo	Thu.M.S&T.1.2	106
Campbell William	Tuo-M-V-4-8	70
Candido Ir Arnaldo		78
Calification JL, Alliando		211
Canzi Massimiliano	Wod E V 4 12	170
Cao Reiming	Tuo-A-V-2-10	86
Cao, Liangliang	Wed M V 6 10	124
Ca0, Lianghang		204
		204
		229
Cao Miao		124
Cao Songiun		224
Ca0, 3011gjuli	1 ue-A-V-4-1	122
	wea-m-v-1-4	242
Cao Viran		110
		210
Caper Charlie		157
Caplini Doborto	Wed-A-S&I-1-4	106
Carlozzi Noelle		110
Caranova Edrosson	Tue-E-SS-1-6	110
Casanova, Euresson	Tue-M-SS-1-4	211
Cassiali Laranza	Thu-A-V-5-12	211
Cascino Diamontino	Tue-A-55-2-2	122
Casello, Diamanuno	Wed-M-V-6-4	133
Cattiau, Julie	Fri-A-55-2-1	257
Cattoni Boldano	Fri-A-55-2-12	239
Cattolli, Koluallo	I nu-A-V-6-2	120
Cavallaro, Allurea	wed-M-V-4-14	130
Cave, Richard	Fri-A-55-2-1	207
Cornels Miles	Fri-A-55-2-12	259
Černak, milos	Wed-E-O-3-4	160
Cernocky, Jan	Tue-A-O-1-5	80
	Wed-M-O-3-1	121
	Wed-E-O-1-4	158
	Thu-M-O-3-5	1/9
	Thu-M-SS-2-2	195
	Thu-M-SS-2-5	196
Cerva, Petr	Wed-M-O-3-3	121
Cervone, Alessandra	Wed-E-V-2-6	163
Cettolo, Mauro	Wed-A-V-6-11	154
Cha, Sujeong	Fri-A-V-6-6	254
Chae, Gyeongsu	Wed-A-V-5-1	150
Chai, Li	Wed-E-V-5-6	1/1
Chai, Shuzhou	Thu-A-V-6-5	212
Chakraborty, Rupayan .	Tue-A-SS-2-10	98
Chakraborty, Titas	Tue-M-V-7-4	75
Chalamandaris, Aimilios	Wed-M-V-3-2	126
Chan, Bo-Cheng	Tue-M-V-1-3	63
Chan, Julian	Wed-M-V-6-3	133
	Wed-A-V-3-1	145
Chan, Wai-Yip	Tue-M-V-3-10	68
Chan, William	Thu-A-SS-1-3	214
	Thu-A-SS-1-11	216
Chanclu, Anaïs	Tue-E-O-2-6	102
	Thu-A-V-4-8	208
Chandorkar, Angad	Tue-E-V-5-4	113
Chandra, Vikas	Fri-A-V-3-10	249
Chandra-Shekar, Meena	Tue-E-O-1-4	100
Chang, Feng-Ju	Tue-M-V-5-7	71
Chang, Heng-Jui	Wed-E-V-3-5	166
Chang, Joon-Hyuk	Fri-M-V-3-7	228
Chang, Oscar	Wed-E-V-5-5	171
Chang, Peng	Tue-E-V-5-11	114
-	Thu-A-SS-1-1	214
Chang, Shiyu	Tue-E-V-5- <u>8</u>	113
Chang, Shuo-Yiin	Wed- <u>M-V-6-4</u>	133
Chang, Simyung	Fri-A-V-3-3	247
. , , , ,	Fri-A-V-3-9	249
Chang, Wonil	Fri-M-V-6-4	234
Chang, Xuankai	Tue-E-V-3-9	109
	Wed-M-V-2-8	125
	Thu-A-SS-1-2	214
	Thu-A-SS-1-9	215
Chang, Yueh-chin	Wed-E-V-4-6	169
0.	Fri-M-V-2-6	225
		-

	Tue-E-V-6-TU	110
Chapman, Kathy	Wed-A-O-3-4	140
Chau, Duen Horng	Tue-E-V-5-10	114
Chaudhari, Gunvant	Fri-M-V-5-2	231
Chaudhary Linisha	Wed-A-V-6-13	155
Chauhan Jagmohan	Tue-M-V-6-8	74
Chaunan, Jaginonan		77
Chahadi Asamaa		1/0
Chenaul, Assiliaa	wed-A-V-4-4	140
Chen, Binbin	Wed-E-V-5-4	1/1
Chen, Boxing	Wed-A-V-3-2	145
Chen, Brian	Wed-M-V-2-12	126
	Thu-M-V-2-8	184
Chen, Chia-Ping	Tue-M-V-1-3	63
, 8	Fri-M-SS-1-4	240
Chen Chia-Yu	Wed-F-V-3-12	167
Chen Fei	Tue-M-V-3-3	67
		84
Chan Eaiyang		174
Clieff, Felyalig	wea-E-V-6-6	1/4
Chen, Guoguo	I nu-A-V-6-5	212
Chen, Haiqing	Thu-M-V-7-5	193
Chen, Haitao	Tue-A-V-6-6	93
Chen, Hang	Tue-M-V-6-5	73
	Thu-M-V-2-7	183
Chen, Hangting	Wed-E-V-5-10	172
Chen, Hongsheng	Wed-E-SS-1-4	176
Chen. Jing	Tue-A-V-2-2	84
Chen lingdong		212
Chan Jinbui		116
Chen Jinlan	Tue-E-V-6-9	124
Chen, Jilikuli	Wed-M-V-6-8	154
Chen, John	I ue-M-V-4-8	70
Chen, Jun	Thu-A-SS-2-11	218
Chen, Junjie	Tue-A-V-6-4	93
Chen, Junkun	Wed-A-V-6-1	152
Chen, Junqi	Tue-M-V-5-6	71
Chen, Kai	Wed-E-SS-1-4	176
Chen. Kehan	Thu-M-V-7-5	193
Chen, Lei	Tue-A-V-1-11	84
Chen Li	Tue-M-V-3-7	67
Chen Lianwa	Tue-F-V-2-2	106
	Wod M SS 1 1	126
	Wed E SS 1 7	176
Cham Li Wai	wea-E-SS-1-7	170
Chen, Li-wei	Wed-M-V-2-8	125
	Thu-A-V-1-1	200
Chen, Long	Fri-A-V-4-1	249
Chen, Mengzhe	Tue-A-V-5-4	91
Chen. Minchuan	Wed-M-V-3-5	127
		230
,,	Fri-M-V-4-4	200
Chen, Mingqing	Fri-M-V-4-4 Tue-A-SS-1-4	95
Chen, Mingqing Chen, Nancy F	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8	95 244
Chen, Mingqing Chen, Nancy F	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6	95 244 261
Chen, Mingqing Chen, Nancy F	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11	95 244 261 216
Chen, Mingqing Chen, Nancy F Chen, Nanxin	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11	95 244 261 216 216
Chen, Mingqing Chen, Nancy F Chen, Nanxin	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tu-M-V-4-2	95 244 261 216 216 69
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2	95 244 261 216 216 69
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1	95 244 261 216 216 69 192
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6	95 244 261 216 216 69 192 235
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4	95 244 261 216 216 69 192 235 91
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6	95 244 261 216 216 69 192 235 91 111
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-1	95 244 261 216 216 69 192 235 91 111 184
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-1 Thu-M-V-3-9	95 244 261 216 216 69 192 235 91 111 184 186
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-3-9 Thu-M-V-6-8	95 244 261 216 216 216 69 192 235 91 111 184 186 191
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Thu-M-V-3-1 Thu-M-V-3-1 Thu-M-V-6-8 Wed-M-V-6-11	95 244 261 216 216 69 192 235 91 111 184 186 191 135
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Thu-M-V-3-9 Thu-M-V-6-8 Wed-M-V-6-11 Wed-A-SS-1-7	95 244 261 216 216 69 192 235 91 111 184 186 191 135 156
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Wei Chen, Weiguang	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Thu-M-V-3-1 Thu-M-V-3-9 Thu-M-V-6-8 Wed-M-V-6-11 Wed-A-SS-1-7 Wed-M-SS-1-7	95 244 261 216 216 235 91 111 184 186 191 135 156 137
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-6-8 Wed-M-V-6-11 Wed-A-SS-1-7 Wed-A-V-4-13	95 244 261 216 216 235 91 111 184 186 191 135 156 137 150
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-3-9 Thu-M-V-6-11 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12	235 95 244 261 216 69 192 235 91 111 184 186 191 135 156 137 150 233
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Wei Suang Chen, Weiguang	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Thu-M-V-6-6 Thu-M-V-6-11 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-13	235 244 261 216 216 69 192 235 91 111 184 186 191 135 156 137 1503 233 149
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang Chen, Xi	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Thu-M-V-3-1 Thu-M-V-3-9 Thu-M-V-6-11 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-A-V-4-10	235 95 244 261 216 69 192 235 91 111 184 186 191 135 156 137 150 233 149
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Gian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Wei Chen, Wei guang Chen, Xi Chen, Xi Chen, Xianzhao Chen, Xianzhao	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Tue-A-V-5-4 Thu-M-V-3-9 Thu-M-V-6-8 Wed-M-V-6-11 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-E-V-3-9 Tue-K-6	$\begin{array}{c} 250\\ 95\\ 244\\ 261\\ 216\\ 69\\ 192\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 115\\ \end{array}$
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang Chen, Xi Chen, Xianzhao Chen, Xiao	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-6-8 Wed-M-V-6-11 Wed-A-SS-1-7 Wed-A-V-4-10 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-5	235 244 261 216 69 192 2355 91 111 184 135 156 137 150 233 149 166 115
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang Chen, Xi Chen, Xianzhao	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-S&T-1-6 Thu-A-SS-1-10 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-6-11 Wed-A-V-6-11 Wed-A-V-6-11 Wed-A-V-6-11 Wed-A-V-6-12 Wed-A-V-4-10 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-5 Thu-M-V-1-10 Fri-M-V-10	$\begin{array}{c} 235\\ 95\\ 244\\ 261\\ 216\\ 69\\ 9192\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 115\\ 182\\ 237\\ 237\\ 149\\ 166\\ 115\\ 233\\ 149\\ 166\\ 115\\ 233\\ 149\\ 166\\ 115\\ 237\\ 182\\ 237\\ 182\\ 237\\ 182\\ 182\\ 182\\ 182\\ 182\\ 182\\ 182\\ 182$
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Tongzhou Chen, Wei Chen, Wei Chen, Wei guang Chen, Xi Chen, Xianzhao Chen, Xiao	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Thu-M-V-3-1 Thu-M-V-3-9 Thu-M-V-6-8 Wed-M-V-6-11 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-5 Thu-M-V-1-10 Fri-M-V-6-9 r	$\begin{array}{c} 235\\ 95\\ 244\\ 261\\ 216\\ 216\\ 69\\ 192\\ 235\\ 9\\ 192\\ 235\\ 192\\ 192\\ 192\\ 235\\ 192\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102\\ 10$
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Wei guang Chen, Xi Chen, Xianzhao Chen, Xiaoyu	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-1 Thu-M-V-6-8 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9	$\begin{array}{c} 95\\ 95\\ 244\\ 261\\ 216\\ 69\\ 91\\ 2235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 135\\ 182\\ 235\\ 228\\ 2235\\ 228\\ 228\\ 228\\ 228\\ 228\\ 228\\ 228\\ 22$
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang Chen, Xi Chen, Xianzhao Chen, Xiaoyu Chen, Xie	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Tue-A-V-5-4 Tue-A-V-5-4 Thu-M-V-3-9 Thu-M-V-6-8 Wed-M-V-6-11 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-5 Thu-M-V-10 Fri-M-V-3-5 Tue-A-V-4-10	95 244 261 216 69 9192 235 91 111 184 186 191 135 156 137 150 233 149 166 115 182 225 228 90 90 90 90 122
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Wei Chen, Wei guang Chen, Xi Chen, Xianzhao Chen, Xiao Chen, Xiao	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-6-7 Wed-A-V-4-10 Wed-A-V-4-10 Fri-M-V-3-9 Tue-E-V-6-9 Fri-M-V-3-14	$\begin{array}{c} 235\\ 95\\ 244\\ 261\\ 216\\ 69\\ 192\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 115\\ 182\\ 235\\ 228\\ 90\\ 167\\ \end{array}$
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang Chen, Xianzhao Chen, Xiaoyu Chen, Xiaoyu Chen, Xiaoyu	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-1 Thu-M-V-3-1 Thu-M-V-6-11 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-E-V-3-14 Tue-A-V-4-10 Wed-E-V-3-14 Tue-A-V-4-10 Wed-E-V-3-14 Tue-X-6-9	$\begin{array}{c} 95\\ 95\\ 244\\ 261\\ 216\\ 69\\ 91\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 115\\ 182\\ 235\\ 228\\ 90\\ 167\\ 116\end{array}$
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang Chen, Xi Chen, Xianzhao Chen, Xiaoyu Chen, Xunquan Chen, Yafeng	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-1 Thu-M-V-3-9 Thu-M-V-6-11 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-3-14 Tue-E-V-3-14 Tue-E-V-6-9 Fri-M-V-3-14 Tue-E-V-6-9 Fru-M-V-3-14 Tue-E-V-6-9 Tue-E-V-3-14 Tue-E-V-6-9 Tue-E-V-1-3	$\begin{array}{c} 95\\ 95\\ 244\\ 261\\ 216\\ 69\\ 91\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 61\\ 15\\ 182\\ 235\\ 228\\ 90\\ 167\\ 116\\ 103\\ \end{array}$
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Weiguang Chen, Xianzhao Chen, Xiaoyu Chen, Xiaoyu Chen, Xie Chen, Xie Chen, Xiaoyu Chen, Yafeng Chen, Yangbin	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-6-8 Wed-M-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-5 Tue-A-V-4-10 Wed-E-V-3-14 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-6-7	$\begin{array}{c} 95\\ 95\\ 244\\ 261\\ 216\\ 69\\ 9192\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 115\\ 182\\ 235\\ 228\\ 90\\ 167\\ 116\\ 103\\ 235\\ \end{array}$
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Wei Chen, Wei guang Chen, Xi Chen, Xianzhao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xie Chen, Xie Chen, Xie Chen, Yafeng Chen, Yangbin Chen, Yi-Chen	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-3-9 Thu-M-V-3-17 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-5 Thu-M-V-3-14 Tue-E-V-6-9 Tue-E-V-1-3 Fri-M-V-6-7 Thu-M-V-3-7 Thu-M-V-3-7	$\begin{array}{c} 235\\ 95\\ 244\\ 261\\ 216\\ 69\\ 192\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 115\\ 182\\ 235\\ 228\\ 90\\ 167\\ 116\\ 103\\ 235\\ 186\\ 186\\ 186\\ 186\\ 186\\ 186\\ 186\\ 186$
Chen, Mingqing Chen, Nancy F Chen, Nuo Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Tongzhou Chen, Wei Chen, Wei Chen, Wei Chen, Xi Chen, Xianzhao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xie Chen, Xie Chen, Yafeng Chen, Yangbin Chen, Youzheno	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Thu-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-1 Thu-M-V-3-1 Thu-M-V-6-11 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-6-12 Wed-A-V-4-10 Wed-E-V-3-12 Wed-A-V-4-10 Wed-E-V-3-14 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-1-3 Fri-M-V-6-7 Thu-M-V-3-7 Fri-M-V-3-7 Thu-M-V-3-7	$\begin{array}{c} 235\\ 95\\ 244\\ 261\\ 216\\ 69\\ 91\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 115\\ 182\\ 235\\ 228\\ 90\\ 167\\ 116\\ 103\\ 235\\ 186\\ 259\\ \end{array}$
Chen, Mingqing Chen, Nancy F. Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Wei Chen, Xianzhao Chen, Xiaoyu Chen, Xie Chen, Xie Chen, Yangbin Chen, Yu-Chen Chen, Yu-Chen Chen, Yu	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Thu-M-V-3-1 Thu-M-V-3-9 Thu-M-V-6-1 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-10 Wed-A-V-4-10 Wed-E-V-3-12 Thu-M-V-1-10 Fri-M-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-6-7 Thu-M-V-3-7 Fri-A-SS-2-11 Fri-M-V-6-7 Fri-A-SS-2-11 Fri-M-V-6-7	$\begin{array}{c} 95\\ 95\\ 244\\ 261\\ 216\\ 69\\ 91\\ 235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 166\\ 115\\ 182\\ 235\\ 228\\ 90\\ 167\\ 116\\ 103\\ 235\\ 186\\ 259\\ 230\end{array}$
Chen, Mingqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Taijing Chen, Tongzhou Chen, Wei Chen, Wei Chen, Wei Chen, Xianzhao Chen, Xianzhao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiaoyu Chen, Xie Chen, Xie Chen, Yafeng Chen, Yangbin Chen, Youzheng Chen, Yu Chen, Yu Chen, Yu	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-1 Thu-M-V-6-8 Wed-M-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-9 Fri-M-V-6-9 Fri-M-V-6-9 Fri-M-V-6-7 Thu-M-V-3-7 Fri-A-SS-2-11 Fri-M-V-4-5 Fri-M-V-4-5 Thu-M-V-3-7 Fri-M-V-3-7 Fri-M-V-3-7 Fri-M-V-3-7 Fri-M-V-4-5 Fri-M-V	$\begin{array}{c} 235\\ 95\\ 244\\ 261\\ 216\\ 69\\ 91\\ 2235\\ 91\\ 111\\ 184\\ 186\\ 191\\ 135\\ 156\\ 137\\ 150\\ 233\\ 149\\ 165\\ 182\\ 235\\ 228\\ 90\\ 167\\ 116\\ 103\\ 235\\ 186\\ 259\\ 230\\ 230\\ 230\\ 230\\ 230\\ 230\\ 230\\ 230$
Chen, Ningqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Taijing Chen, Tongzhou Chen, Tongzhou Chen, Wei Chen, Wei Chen, Wei Chen, Wiaguang Chen, Xianzhao Chen, Xianzhao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xie Chen, Xie Chen, Yangbin Chen, Yangbin Chen, Youzheng Chen, Yu Chen, Yu Chen, Yu Chen, Yu Chen, Yu	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-12 Tue-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-9 Thu-M-V-3-9 Thu-M-V-3-9 Thu-M-V-3-9 Thu-M-V-6-8 Wed-A-V-4-13 Fri-M-V-5-12 Wed-A-V-4-10 Wed-E-V-3-9 Tue-E-V-6-5 Thu-M-V-1-10 Fri-M-V-3-7 Tue-A-V-4-10 Wed-E-V-3-14 Tue-E-V-6-9 Tue-E-V-6-9 Tue-E-V-1-3 Fri-M-V-3-7 Fri-M-V-3-7 Fri-M-V-4-5 Fri-M-V-4-5 Fri-M-V-4-5 Fri-M-V-4-5 Fri-M-V-4-5	235 244 261 216 69 192 235 91 111 184 186 191 135 156 137 150 233 149 166 152 235 225 228 90 167 162 235 235 235 149 166 152 235 230 230 230 230 230 235 230 230 230 235 230 232 235 230 230 230 230 230 235 230 230 230 230 235 230 230 230 230 235 230 300 30
Chen, Ningqing Chen, Nancy F Chen, Nanxin Chen, Nuo Chen, Qian Chen, Qian Chen, Sanyuan Chen, Taijing Chen, Taijing Chen, Tongzhou Chen, Tongzhou Chen, Wei Chen, Wei Chen, Wei Chen, Wei Chen, Xianzhao Chen, Xianzhao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Xiao Chen, Yaiao Chen, Yafeng Chen, Yangbin Chen, Youzheng Chen, Yue Chen, Yue Chen, Zening	Fri-M-V-4-4 Tue-A-SS-1-4 Fri-A-V-1-8 Fri-A-S&T-1-6 Thu-A-SS-1-11 Thu-A-SS-1-12 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-7-1 Fri-M-V-6-6 Tue-A-V-5-4 Tue-E-V-4-6 Thu-M-V-3-1 Thu-M-V-3-9 Thu-M-V-6-11 Wed-A-SS-1-7 Wed-A-V-4-13 Fri-M-V-6-12 Wed-A-V-4-10 Wed-E-V-3-12 Wed-A-V-4-10 Wed-E-V-3-12 Tue-A-V-4-10 Wed-E-V-3-14 Tue-E-V-6-5 Thu-M-V-1-10 Fri-M-V-6-7 Thu-M-V-3-7 Fri-M-V-4-6 Fri-M-V-4-6 Fri-M-V-4-6 Fri-M-V-4-6 Fri-M-V-4-6 Fri-M-V-4-6 Fri-M-V-4-6	353

Chan Zhebuai	Tuo-A-V-4-7 00	Chung Joon Son		187	Daniel Morgane	Eri.M.O.1.5	220
Cileii, Zileiiuai	Tue-A-V-4-7 90	Chung, Joon Son		107	Damei, Morgane		220
	Fri-A-SS-2-11 259		Thu-M-V-4-5	187		Fri-A-S&1-1-4	261
Chen, Zhengyang	Wed-A-O-2-4 139		Thu-A-V-6-6	212	Dao, Mai Hoang	Fri-A-V-6-1	254
	Wed-A-SS-1-8 156	Chung, Minhwa	Fri-A-SS-2-13	259	Daoudi, Khalid	Tue-M-O-2-3	61
Chen. Zhuo	Wed-E-V-5-3 171	Chung Soo-Whan	Thu-Δ-V-6-6	212	d'Apolito, Sonia	Fri-M-V-2-12	227
,	$T_{M} M_{2} = 1.33$	Chung Vu An		62	Dae Amit	Wed MV62	133
	Thu M V 2 2 195	Chung, Tu-An		214	Das District		133
	Thu-M-V-5-5 105		I nu-A-SS-1-4	214		Tue-M-0-2-3	01
	Thu-M-V-3-9 186	Church, Kenneth	Thu-A-V-4-6	208	Das, Nilaksh	Tue-E-V-5-10	114
	Thu-A-V-2-2 202		Fri-A-V-2-7	246	Das, Partha Pratim	Fri-M-V-5-6	232
	Thu-A-V-4-3 207	Cicuta. Pietro	Tue-M-SS-1-1	77	Das, Rohan Kumar	Tue-A-SS-2-5	97
	Thu-A-V-6-4 212	Cieri Christopher	Tue-F-O-1-4	100	Dasgupta, Hirak	Thu-M-S&T-1-7	197
	$Eri_{A} = \frac{1}{243}$	elen, emiscopher mm	WedA022	128	Douwole Justin	Wod MV 1 1	122
Chop 7ivi			wea-A-O-2-2	100	Dauweis, Justin		144
	Tue-A-V-0-3 93		I hu-A-V-4-6	208	David, Robert	Fri-A-V-3-8	248
Cheng, Linjuan	Fri-A-SS-1-5 256	Ciesielski, Grzegorz	Tue-E-O-1-1	100	Dawalatabad, Nauman .	Thu-A-V-4-4	207
Cheng, Longbiao	Tue-E-V-2-3 106		Tue-E-O-1-2	100	Dawe-Lane, Erin	Tue-A-V-2-6	85
Cheng, Luyao	Thu-A-V-6-4 212	Civera. Jorge	Wed-A-V-6-10	154	Davalu, Praveen	Tue-E-SS-1-6	118
Cheng Mengli	Eri-A-V-3-2 247		Thu: A.V.6.10	213	de Bodt Marc S	Tue-F-SS-1-7	118
Chong Ning		Clarko Coorgia		164	Dobale Najim		74
Cheng, Ning			weu-E-v-2-7	104	Denak, Najini		
	Tue-E-V-3-4 108	Cmejla, J	Tue-A-S&T-1-7	99		Tue-A-O-T-2	80
	Wed-M-V-3-9 127	Cmejla, Roman	Tue-E-SS-1-7	118		Tue-E-V-1-9	104
Cheng, Xingliang	Fri-M-V-7-4 237	Codina-Filbà. Ioan	Thu-M-S&T-1-3	196		Thu-A-SS-1-11	216
Cheng. Yao-Fei	Wed-A-V-6-5 153	Cohen Israel	Eri-A-SS-1-6	256		Thu-A-SS-1-12	216
Cheng Yuan	Tue-E-V-1-7 104	Cohn Michelle		230 Q1		Thu-A-SS-2-10	218
Choon Sung Jun	Thu A V E 4 200	Colley Mett	Tue-A-0-2-2	61		Fri-M-V-7-10	238
Cheon, Sung Jun	Thu-A-V-5-4 209	Coler, Matt	I ue-M-O-I-4	60	Dobá Nicolo		200
	Fri-A-V-5-5 252	Collobert, Ronan	Tue-M-V-5-10	72	de la Frante Coffe		223
cneon, roungju	Wed-M-V-4-7 129		Tue-A-V-4-4	89	ue la ruellite, solla	Thu-A-SS-2-1	210
Cherry, Colin	Wed-A-V-6-2 152		Tue-A-V-4-8	90	Delcroix, Marc	Tue-E-V-2-10	107
Chestek, David	Tue- <u>A-SS-2-6</u> 97		Wed-E-V-1-3	161		Wed-M-O-3-1	121
Chettri, Bhusan	Wed-M-V-4-6 128	Conde, Pauline	Tue-A-V-2-6	85		Wed-M-V-6-6	134
Chetunalli Srikanth Dai	Tue A-SS 2.1 06	Condron Scott	Wed EV 2-7	164		Wed-A-V-4-2	148
спстираш, энканти кај		Cong Jien	Weu-E-V-Z-7	150		Thu-M-V-3-2	185
	rue-A-SS-2-11 98	Cong, Jian	wed-A-V-5-2	120			205
	Thu-A-V-4-1 207	_	Fri-A-V-5-4	252		Thu-A-V-3-3	203
Chhipa, Prakash C	Tue-A-SS-2-4 97	Conneau, Alexis	Wed-A-V-6-3	153		Thu-A-V-3-5	205
Chi, Po-Han	Tue-E-V-3-9 109		Wed-E-V-1-3	161		Thu-A-V-4-5	208
Chiba. Yuva	Tue-M-V-4-4 69	Cook. Darren	Wed-A-V-1-2	141	Delgado, Héctor	Fri-M-V-7-9	238
	Tue-M-V-4-5 69	Cooke Martin	Wed-M-V-5-4	131	Delpech, Estelle	Thu-M-SS-2-1	195
Chion Chung Ming	Tue-M-V-4-5 03	Cooper Erico		131	Del Rio Miguel	Thu.A.V.2.0	204
Chieff, Chung-Ming	Tue-A-V-6-7 94			237	Dolworth Natalio		201
Chien, Jen-Tzung	Wed-A-V-3-9 146	Copet, Jade	Thu-A-V-5-6	210	Denvir Matalle	1 nu-A-V-2-9	104
	Thu-M-V-7-3 192	Cordourier Maruri, H.A.	Wed-E-V-2-10	164	Demberg, vera	Wed-M-V-5-3	131
Chingacham, Anupama	Wed-M-V-5-3 131	Cornell, Samuele	Fri-M-O-1-4	220	Demopoulos, Carly	Wed-A-V-1-7	142
Chiu. Chung-Cheng	Wed-M-V-6-4 133	Cornish, Andrew	Wed-A-V-1-7	142	Demuth, Katherine	Thu-A-V-6-7	212
	Wed-M-V-6-10 134	Cox Felicity	Tue-F-SS-1-1	117	Demuvnck. Kris	Wed-A-SS-1-2	155
		cong reliency		117	Deng Chengyun	Thu-Δ-V-3-11	206
	Thu-A-55-1-5 214			212	Dong Jinjun		250
Chiebowski, Aurelie	Tue-M-V-7-1 75		Thu-A-V-6-7	212		FII-A-33-2-9	1239
Chng, Eng Siong	Wed-M-V-1-7 123	Cox, Trevor J	Tue-A-V-3-7	88	Deng, Keqi	Wed-M-V-1-4	122
	Fri-M-V-5-12 233	Coy, Thomas	Tue-M-V-7-7	76	Deng, Liqun	Tue-E-V-6-5	115
Cho. Hoon-Young	Tue-M-V-2-2 64	Cristia. Aleiandrina	Fri-Survey	240		Thu-M-V-1-10	182
.,	Wed-M-V-3-1 126	Csanó Tamás Gábor	Wed-A-0-3-6	140	Deng. Wesley	Wed-F-V-2-2	163
	Wod A V 5 6 151	Coupo, rumuo Gubor		151	Deng Yan	Tue-A-V-4-10	90
	wed-A-V-5-6 151		wea-A-V-5-8	100	Denicombro Alexander		110
	Thu-A-V-5-5 209	Cucchiarini, Catia	I hu-M-V-I-I	180	Definseriko, Alexander	wea-m-0-1-1	119
Cho, Hyunchang	Thu-A-V-6-12 213	Cuesta, Helena	Tue-A-SS-2-9	98	de Seyssel, Maureen	Wed-M-V-2-10	125
Cho, Jaejin	Thu-A-SS-2-10 218	Cui, Chenye	Wed-E-V-6-6	174	Deshmukh, Soham	Tue-A-V-1-12	84
Cho, Kiho	Tue-M-V-6-9 74	Cui. Xia	Wed-M-V-2-9	125	Deshpande, Gauri	Tue-A-SS-2-7	97
Cho, Sungiae	Wed-A-S&T-1-1 156	Cui Xiaodong	Wed-M-V-6-9	134	Desplangues, Brecht	Wed-A-SS-1-2	155
Cho Won Ik	Thu-A-V-6-12 213	cul, muouong	Wed E V 2 12	167			207
Chodorowalzi Joaguoo	Thu A V 2 12 207	Culling John F	Weu-E-V-3-12	107	Dhamual Hira		207
Choose Sector Choose Sector Se	THU-A-V-3-T3 207	Cuming, John F	Tue-A-V-3-7	00	Dhamyal, filla	F11-IVI-V-5-4	202
cnoe, soyeon	Thu-A-V-6-6 212	Cumani, Sandro	Fri-A-V-4-3	250	51 51 1	Fri-A-V-4-11	251
Choi, Byoung Jin	Thu-A-V-5-4 209	Cumbal, Ronald	Fri-A-V-1-12	245	Dharma Gita, A	Thu-M-SS-2-1	195
-	Fri-A-V-5-5 252	Cummins, Nicholas	Tue-A-V-2-6	85	Dhir, Chandra	Fri-M-V-6-4	234
Choi Funbi	Fri-M-V-4-2 230	Cunha, Conceição	Wed-M-O-2-2	120	DiCocco. Vince	Tue-A-V-5-2	91
Choi Hyuniin	Tue-A-V-5-0 02	Curado Francisco	Τμο.Δ-\/-2-0	86	Dighe, Pranav	Fri-M-V-6-4	234
Choi Kwang Dro	Wed My 4 2 120	Cutlor Poss	Mad AV 4 F	1/0	Diiketra Jaleka	Tuo M O 1 4	0.0
Chui, Kwalig Pyo	wea-M-V-4-3 128	cuuei, KOSS	wea-A-V-4-5	140	Dimitriadia Dimitria	Tue-M-0-1-4	
Choi, Min-Seok	Wed-E-V-5-8 172		Wed-E-SS-1-1	1/5	Dimuriadis, Dimitrios	Tue-E-V-6-3	115
Chojnacka, Roza	Tue-E-V-1-6 104		Fri-A-SS-1-1	255	Dinarelli, Marco	Wed-M-O-1-5	119
Cholin, Joana	Thu- <u>A-O-1-3</u> 198	Cvetkovic, Zoran	Tue-M-V-5-3	71	Dineley, Judith	Tue-A-V-2-6	85
Chorowski. Ian	Tue-E-O-1-1 100	-			Ding, Hongwei	Fri-M-V-1-6	223
,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	Tue-E-O-1-2 100	D			Ding, Wenbiao	Thu-A-V-1-2	200
Chouchane Oubaïda		B 1 1 - 1 1 - 1			Ding Yi-Yang	Tue-A-V-6-2	03
Choudham, Carryidhi		Dabral, Rishabh	Wed-A-V-1-1	141	Dinkol Hoinnich		. 33 75
chouunary, samriani	FII-A-V-6-8 255	Dai, Lingfeng	Wed-A-V-2-10	144	Dinami Aliala	Tue-M-V-6-TT	13
	Fri-A-V-6-9 255	Dai, Li-Rong	Tue-A-V-1-3	82	Dipani, Alisn	Thu-A-V-3-7	206
Chowdhury, Shammur A	Wed-E-V-1-11 162	,	Wed-M-V-1-2	122	Disch, Sascha	Tue-M-O-3-2	62
	Thu-A-V-2-11 204			183	Diskin-Holdaway, Chloé	Thu-A-V-6-7	212
Christensen, Heidi	Tue-Survey 79			221	Diwan, Anuj	Wed-M-V-2-1	123
. ,	Thu-M-V-1-7 181		Fri-M-V-4-7	231		Wed-F-V-1-7	161
	Thu M V 1 0 101		Fri-M-SS-1-3	239			202
		Dai, Suyang	Thu-M-V-7-5	193	Do Phat	Tue M-0 1-4	203
	Thu-A-SS-2-7 218	Dai, Xudong	Tue-M-V-2-5	65		тие-м-О-Т-4	00
Christensen, Mads G	Tue-E-V-3-3 108	d'Alessandro. C.	Tue-A-O-2-1	81	Do, Quynh	Tue-E-V-4-7	111
Chu, Fang	Fri-A-SS-2-11 259		Fri-M-O-2-5	221		Fri-A-V-6-3	254
Chu, Wei	Tue-E-V-5-3 112	Dalmia Siddharth	Tue E V 4 10	112	Dobel, Christian	Wed-M-V-5-11	132
,	Tue-E-V-5-11 114	Danna, Shundi III	1 ue-E-V-4-10	162	Dobson, Richard I.B.	Tue-A-V-2-6	85
	Thu-A-SS-1-1 214	Dama Er	wed-E-V-1-12	102	Doc. Jean-Bantiste	Thu-A-O-1-2	198
Chua Mason	Tuo E V 5 4 112	Dang, Feng	Wed-E-V-5-10	172	Doddinatla Dama		205
Chuong Chuon D	Thue May 2 = 100	Dang, Jianwu	Tue-M-V-2-1	64	Dogmin Charles	THU-A-V-3-4	203
Cnuang, Snun-Po	Thu-M-V-3-7 186		Tue-M-V-4-7	70	Dognin, Charles	Tue-A-SS-2-11	. 98
Chuang, Yung-Sung	Tue-E-V-3-9 109		Tue-E-V-1-11	105	Dominguez, Mónica	Wed-A-S&T-1-5	157
Chuangsuwanich, E	Wed-E-V-6-1 173		Fri-Δ-V-2-6	246	Dong, Joshua	Thu-A-V-2-9	204
Chukharev-Hudilainen F	Wed-A-V-1-4 141		$\frac{11.4 \cdot V \cdot 2 \cdot 0}{\text{Eri} \cdot 4 \cdot V \cdot 2 \cdot 10}$	247	Dong, Shuvan	Tue-E-V-3-9	109
Ching Clement	Fri-A-V-6-8 255	Dang Ting	Thu MAY 1.0	24/ 101	Dong Yuanije	Tue-E-V-1-4	103
Chung Hypneoung	Thu AVE 10 210	Dang, Ing	Thu-M-V-T-9	191	Dong, ruanjie		103
Chung, rivunseung	T NU-A-V-5-TU 210						

INTERSPEECH 2021 264 BRNO | CZECHIA

Doshi, Rohan	Fri-A-SS-2-11	259 65	Englert, Marina
Doutre. Thibault	Wed-M-V-6-10	134	Engwall, Olov
Dovrat, Shaked	Fri-M-O-3-1	221	Epps, Julien
Dovzhenko, Yuliya	Wed-M-O-1-4	119	Erdogan, Hakan
Drenhaus, Heiner	Fri-M-V-2-4	225	Ermis, Orhan
Drioli Carlo	Wed-M-0-3-4	121	Ernestus, Mirjam
	Wed-A-V-4-8	149	Escalante-B., Alberto N
Droppo, Jasha	Tue-A-V-4-2	89	,,, _,, _
	Tue-A-SS-1-8	96 96	Eshky, Aciel
	Tue-A-SS-1-9	96	Eskimez, Sefik Emre
	Wed-E-V-3-10	166	Fspin Juan M
	Thu-M-V-5-2	189	Esposito, Anna
	Thu-A-V-2-7	203	Espy-Wilson, Carol
	Fri-M-V-1-2	222	Fasid Clim
Drossos, Konstantinos .	Thu-A-V-1-3	200	Essiu, Siiii
Drude, Lukas	Wed-M-V-4-8	129	Loteve, runner
Drugman, Thomas	Thu-A-V-5-7	210	Estival, Dominique
	Fri-M-V-1-2	222	Evain, Solène
Du, Chenpeng	Thu-M-V-5-3	189	Evans, Nicholas
Du, Hongqiang	Tue-A-V-6-6	93	
	Tue-E-V-6-12	116	
Du, Jiayu	1 hu-A-V-6-5	212 73	
Du, juli	Wed-E-V-5-6	171	
	Wed-E-SS-1-8	176	Eyben, Florian
	Thu-M-V-2-7	183	r ·
	Thu-M-V-4-6	188	F
	Thu-A-V-4-3	207	Fabien, Maël
	Thu-A-V-6-4	212	Fadiga, Luciano
Du, Xingjian	Tue-M-V-3-7	67	Faerman, Afik
	Tue-M-V-6-1	73	Falaise. Achille
Duan, wenying	Thu-M-V-7-2	192 64	Falavigna, D.
	Fri-M-V-7-11	238	Falkowski-Gilski, P
Dubagunta, S.P	Wed-A-O-3-5	140	Fallgren, Per
Dubey, Harishchandra .	Wed-E-SS-1-1	175	Fan, Ruchao
Duckhorn, Frank	Tue-E-O-2-3	101	Fan, Zhiyun
Dugaii, Saran Dunbar Fwan	I UE-A-V-2-7	80 125	Fang, Lei
Dupoux, Emmanuel	Wed-M-V-2-10	125	Fang, Qiang
-	Thu-A-V-5-6	210	Eang Vin
Durda, Kevin	Tue-M-V-4-8	70	Failg, Alli
Duroselle, Raphael	Wed-M-V-1-2	122	Farris, Brian
Dvir. Amit	Thu-M-55-1-2	184	Farrús, Mireia
Dvoynikova, Anastasia .	Tue-M-SS-1-11	79	Fasoli, Andrea
Dzikowski, Jarosław	Tue-E-O-1-1	100	Fazer, Amm Feather Jenelle
	Tue-E-O-1-2	100	Federico, Marcello
Е			Félix, Simon
- Fami Vathmin I	Τ υο Δ \/ Ο Ζ	96	Feng, Jinwei
Edly, Kauliyii J Edler Bernd	Fri-M-0-3-3	222	
Edlund, Jens	Thu-A-V-6-8	213	Feng. Shulin
Edraki, Amin	Tue-M-V-3-10	68	Feng, Siyuan
Effendi, Johanes	Wed-A-V-6-6	153	Fenu, Gianni
Egas-Lopez, Jose vicente Eghbali Senehr	Tue-M-55-1-10 Fri-A-V-6-8	79 255	Feris, Rogerio
Egorova, Ekaterina	Thu-M-O-3-5	179	Fernandes Ir Ricardo
Ehlert, Hanna	Fri-M-O-1-3	219	Fernandez, Raul
Ehr, Heiko	Thu-M-SS-2-3	195	Ferragne, Emmanuel .
Ehrenfors, Johan	Wed-A-S&T-1-4	157	Ferreira, David
Elilielut, Marieke	Fri-M-V-1-3	223	Ferrer, Carlos A
Eisenberg, Aviad	Thu-A-V-3-1	204	Terrer, Lucialia
Eitel, Megan M	Tue-M-O-2-1	60	
El Amouri, Hussein	Tue-M-O-1-3	60	n . n 1
Elhilali, Mounya	Thu-Keynote	198 65	Fetaya, Ethan
Flibol Oguz	Tue-M-V-2-7	105	Finger, Marcelo
L11501, 05uL	Wed-E-V-3-10	166	·
	Fri-A-V-4-2	250	_
Ellie, Benjamin	Thu-M-SS-2-6	196	Fischer, Tim
Ellinas, Nikolaos Flmers Mikev	Wed-M-V-3-2	120 191	Fischer, Volker
Lincio, mikey	F <u>ri-M-V-2-4</u>	225	Flucha, Alexandre
El Shafey, Laurent	Fri-A-V-1-3	243	Fogerty, Daniel
Elyasi, Mahsa	Fri-M-V-5-11	233	Foglianti, Lorenzo
Emamian, Eshrat S			
Emaniidio anneos	Wed-E-V-6-3	173 61	Fohr Dominique
Emanuele, Marco Emond. Jesse	Wed-E-V-6-3 Tue-M-O-2-5 Tue-A-V-4-7		Fohr, Dominique
Emanuele, Marco Emond, Jesse	Wed-E-V-6-3 Tue-M-O-2-5 Tue-A-V-4-7 Wed-E-V-3-7	173 61 90 166	Fohr, Dominique Folarin, Amos A

1		110	Fontaino Mathiou		87
•••••	Tue-E-SS-T-7	110	Fontaine, Matineu	Tue-A-V-3-2	0/
	Fri-A-V-1-12	245	Foresti, Gian Luca	Wed-M-O-3-4	121
i		180	,	Wod-A-V-4-8	1/10
	1 HU-IVI-V-3-0	109		Weu-A-V-4-8	143
	Thu-A-V-6-7	212	Foubert-Samier, A	Tue-M-O-2-3	61
i	Thu M 1/ 2 2	185	Fraenkel Frnest	Eri-A-SS-2-2	257
	1110-101- 0-3-3	105	Fragman Stafan	111/(35 2 2	1 - 7
	Tue-A-SS-1-1	94	riagliel, stelall	wed-A-S&1-1-3	137
	Fri-A-O-2-3	242	Francart, Tom	Tue-A-O-2-3	- 81
	11/10/20	107	Francl Andrew	Wod-E-V-5-2	171
••••	wed-M-V-3-8	127		Weu L V J Z	1/1
) N.	Tue-M-O-3-2	62	Frank, Stefan L	Fri-A-O-2-3	242
	Tuo-A-V-2-1	86	Franzen, Ian	Fri-A-SS-1-4	256
	Tue-A-V-3-1	00	Erná Miogzko	5.4022	221
	Tue-A-V-2-8	86	rias, mieszko	Fri-M-0-3-2	221
. 1	Tue-E-V-6-3	115	French, Louis M.	Tue-M-O-2-1	60
	THEEVOS	171	Freyborg Keenan	Wed MO14	110
	Wed-E-V-5-3	171	Treyberg, Reenan	weu-wi-0-1-4	115
	Tue-A-SS-2-2	96	Fritsch, Julian	Tue-M-O-2-2	61
i	WedEVDE	162		Wed-A-O-3-5	140
	wed-E-v-2-5	105	Examp Davida		216
	Wed-E-V-2-11	164	Fromm, Davida	I nu-A-SS-2-1	210
i		246	Frv. Laura	Wed-A-V-1-7	142
	F11-A-V-2-0	240	Eu Voigi		244
	Thu-M-O-1-4	177	ги, кащі	Fri-A-V-1-9	244
1	Wed-M-0-1-5	119	Fu. Oiang	Tue-E-V-2-9	107
		10	Fu Puibo	Wod MV 4 F	128
	Wed-E-O-2-T	158		weu-ivi-v-4-5	120
	Thu-A-V-6-7	212	Fu, Szu-Wei	Tue-M-V-3-8	68
i	Wed MO15	110		Tue-M-V-3-9	68
	wed-IM-O-1-5	119	Fre Versela	T A 1/ 4 1	00
	Tue-A-SS-1-1	94	Fu, Yanzne	Tue-A-V-4-T	88
1	Tue-Δ-SS-2-2	96	Fu, Yihui	Thu-A-V-6-4	212
	T., F.) (110	105	Fuchs Susanne		101
	Tue-E-V-1-13	102	1 actio, 5usaitte	THU-IM-V-0-5	100
Í	Wed-E-O-1-2	158		Thu-A-O-1-6	199
		237	Fuegen, Christian	Tue-M-V-2-8	66
ļ	FTI-IMI-V-7-2	201	- acgen, emistian		110
	Fri-M-V-7-9	238		1ue-E-V-3-13	110
i	Fri-M-\/-7-12	230		Wed-M-V-4-12	130
	T 5	100			100
	1 ue-E-O-3-3	102		wed-M-V-6-3	133
				Wed-A-V-2-1	142
				Wod A V 2 1	1/5
				wea-A-V-3-I	140
	Tue A COT 1 4	00		Wed-A-V-3-14	147
	Iue-A-S&I-I-4	99		Fri-A-V-3-6	248
	Tue-M-O-2-5	61			100
		140	Fujie, Shinya	Thu-M-V-7-4	192
	wed-A-V-I-7	142	Fujimoto, Masakivo	Tue-M-V-5-4	71
	Fri-M-V-5-2	231	Fuitala Talana		240
	Wed EV/ 4.2	168	Fujioka, Такиуа	Fri-A-V-2-4	246
	wed-E-V-4-3	100	Fujita, Kenichi	Thu-M-V-5-4	189
	Fri-M-O-1-2	219	Fujita Vusulta		107
i	Wed-A-V-4-1	148	Fujita, rusuke	<u>т nu-м-v-4-4</u>	107
	Weu A V 4 I	140		Thu-M-V-4-8	188
	Thu-A-V-6-8	213	Fujita Vuva	Wod-M-V-2-8	125
	Tue-F-V-5-3	112	rujita, ruya	weu-wi-v-2-o	125
		214		Thu-A-SS-1-6	215
	I nu-A-SS-I-I	214		Thu-Δ-SS-1-9	215
	Wed-M-V-1-5	122	Technolo, Telesslei		213
i	Wod M V 1 2	122	Fukuda, Takashi	I hu-A-V-2-6	203
•••••	weu-wi-v-1-5	122	Fukumori. Takahiro	Fri-M-V-5-9	233
	Tue-A-V-2-5	85	Fulford Danial	Thu M O 2 2	170
i	Thu-M-V-2-4	183	runoru, Daniel	тпи-м-0-2-2	170
		105		Fri-A-O-1-4	241
	Fri-M-SS-1-3	239	Fung Dascale	Wed-Keynote	137
	Wed-A-0-3-2	139		wea keynote	157
		166	Futamata, Kosuke	Thu-M-V-5-1	188
	Wed-E-V-3-6	100			
	Thu-M-S&T-1-3	196	C		
i	Wod E V 2 12	167	G		
•••••	weu-E-v-5-12	107			100
	Tue-A-SS-1-9	96	Gabrys, Adam	Wed-M-V-4-10	129
1	Wed-F-V-5-2	171	Gaddam, Navaneetha	Tue-A-V-2-3	85
		100	Cadail Swannil		101
· · · ·	1 nu-M-V-5-6	109	Gaugii, Swapiili	I NU-M-V-I-8	101
	Thu-A-O-1-2	198	Gales, Mark I.F.	Tue-M-V-2-6	65
		80	Camage Amila	Wod M V 2 0	125
· · · · ·	Tue-A-V-3-T0	00	Gamage, Amila	weu-m-v-z-9	120
	Wed-M-SS-1-7	137	Gamboa Gamboa, J.E	Tue-A-SS-1-1	94
i		207	Gamper Hannes	Wod-E-SS 1-1	175
	Thu-A-V-4-2	201	Sumper, names	weu-E-33-1-1	113
	Wed-A-V-3-8	146		Fri-A-SS-1-1	255
i	Wed-M-V-2-2	124	Gan, Wendong	Tue-A-V-6-6	93
	Wed Mr V Z Z	100	Composition Contraction	T	00
	wed-A-O-2-3	128	Ganapathy, Sriram	Tue-A-SS-2-1	96
	Wed-M-V-2-12	126		Tue-A-SS-2-11	- 98
		184		Wed-M-SS-1-6	137
1 ~ !	т пи-м-v-2-8	104		wed-W-55-1-6	107
10 C.	Tue-M-SS-1-4	78		Wed-Survey	137
i	Fri-A-V-5-11	253		Thu-A-V-4-1	207
	T. 5020	100		THU-A-V-4-1	207
ı	Tue-E-O-2-6	102		Thu-A-V-4-6	208
İ	Tue-A-V-2-9	86	Gandhi Vineet	Fri-4-1/-5-2	252
		110	Conhotre Istin		111
•••••	Tue-E-SS-T-7	110	Gaimoura, Jaun	Tue-E-V-4-8	111
	Tue-E-V-1-5	103	Gannot, Sharon	Tue-E-V-2-6	106
		201	,		204
ļ	Thu-A-V-T-7	201	o 1	Thu-A-V-3-1	204
	Thu-A-SS-2-4	217	Ganzeboom, Mario	Thu-M-V-1-1	180
i	Fri-A-V-1-4	243	Gao Dongij	Thu MO26	170
	T 5 1 2 5	106		THU-M-0-3-6	110
	Tue-E-V-2-6	100	Gao, Fan	Tue-E-V-5-4	113
. İ	Tue-M-SS-1-4	78	Gao, Heting	Tue-E-V-5-8	113
	Wod E CC 1 -	175	Cao Barrer		221
· · · ·	Wed-E-SS-1-3	175	Gao, Jiameng	Fri-M-O-2-3	221
	Thu-M-O-1-3	177	-	Fri-M-0-2-4	221
ł		256	Cao Jianfong		110
ļ	FTI-A-SS-1-4	200	Gao, Jianieng	Tue-E-V-4-1	110
	Thu-M-S&T-1-1	196	Gao, Xibin	Fri-M-V-6-2	234
	Eri A O 1 E	242	Cao Vang		221
· · · · · !	FII-A-0-1-5	120	GaU, 1 ally	Fri-IVI-V-5-1 I	233
	Wed-A-O-2-2	138	Gao, Yingbo	Wed-A-O-1-4	138
İ	Тие-А-S&T-1-5	99	Gao Vongya	Wod A V 2 A	1/1
		60		weu-A-V-2-9	144
	Tue-M-V-3-T0	60	Gao, Yuan	Fri-A-V-2-6	246
İ	Eri-M-O-2-3	221	Gao Theng	Fri-M-V-6-2	234
			Guo, Lucuz		204
	F-1402	221			
	Fri-M-O-2-4	221	Gao, Zhifu	Fri-A <u>-V-3-5</u>	248
i	Fri-M-O-2-4 Tue-A-V-3-9	221 88	Gao, Zhifu Garcés Díaz-Munío G-V	Fri-A-V-3-5	248
	Fri-M-O-2-4 Tue-A-V-3-9	221 88 138	Gao, Zhifu Garcés Díaz-Munío, G.V.	Fri-A-V-3-5 Thu-A-V-6-10	248 213
	Fri-M-O-2-4 Tue-A-V-3-9 Wed-A-O-1-1	221 88 138	Gao, Zhifu Garcés Díaz-Munío, G.V.	Fri-A-V-3-5 Thu-A-V-6-10	248 213
	Fri-M-O-2-4 Tue-A-V-3-9 Wed-A-O-1-1 Tue-A-V-2-6	221 88 138 85	Gao, Zhifu Garcés Díaz-Munío, G.V.	Fri-A-V-3-5 Thu-A-V-6-10	248 213
	Fri-M-O-2-3 Fri-M-O-2-4 Tue-A-V-3-9 Wed-A-O-1-1 Tue-A-V-2-6 Tue-A-SS-2-2	221 88 138 85 96	Gao, Zhifu Garcés Díaz-Munío, G.V.	Fri-A-V-3-5 Thu-A-V-6-10	248 213

-		1/1
n L	Fri-A-0-2-3	242
1	Fri-A-SS-1-4	256
	Fri-M-O-3-2	221
is M	Tue-M-O-2-1	60
enan	Wed-M-O-1-4	119
an	Tue-M-O-2-2	61
	Wed-A-O-3-5	140
ida	Thu-A-SS-2-1	216
	Wed-A-V-1-7	142
	Fri-Δ-V-1-9	244
		107
	Tue-E-V-2-9	107
• • • • • • • • • • • • • • • • •	<u>wed-M-V-4-5</u>	120
•••••	I ue-M-V-3-8	68
	Tue-M-V-3-9	68
	Tue-A-V-4-1	88
	Thu-A-V-6-4	212
nne	Thu-M-V-6-5	191
	Thu-A-O-1-6	199
istian	Tue-M-V-2-8	66
	Tue-F-V-3-13	110
	Wed-M-V-4-12	130
	Wod-M-V-6-3	133
	Wed-W-V-0-3	1/12
	wed-A-V-2-1	144
	wed-A-V-3-1	140
	Wed-A-V-3-14	147
	Fri-A-V-3-6	248
a	Thu-M-V-7-4	192
asakiyo	Tue-M-V-5-4	71
	Fri-A-V-2-4	246
hi	Thu-M-V-5-4	189
ke	Thu-M-V-4-4	187
	Thu-M-V-4-8	188
	Wed-M-V-2-8	125
		215
	Thu A SS 1 0	215
achi	Thu-A-55-1-9	213
asiii	Inu-A-V-2-6	203
akaniro	Fri-M-V-5-9	233
nel	Thu-M-O-2-2	178
	Fri-A-O-1-4	241
P		
	Wed-Keynote	137
osuke	Wed-Keynote Thu-M-V-5-1	137 188
m	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10	137 188 129
m vaneetha	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10 Tue-A-V-2-3	137 188 129 85
m vaneetha	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10 Tue-A-V-2-3 Thu-M-V-1-8	137 188 129 85 181
m vaneetha J.F	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10 Tue-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-6	137 188 129 85 181 65
m vaneetha J.F ila	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10 Tue-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-6 Wed-M-V-2-9	137 188 129 85 181 65 125
m vaneetha pnil J.F nila mboa, I.F.	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10 Tue-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1	137 188 129 85 181 65 125 94
m vaneetha onil J.F nboa, J.E nboa, J.E	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10 Tue-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1	137 188 129 85 181 65 125 94 175
m vaneetha onil J.F illa mboa, J.E nnes	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10 Tue-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Fri-A-SS-1-1	137 188 129 85 181 65 125 94 175 255
m vaneetha onil J.F nila nboa, J.E nnes	Wed-Keynote Thu-M-V-5-1 Wed-M-V-3-10 Thu-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Tue-A-V-6-6	137 188 129 85 181 65 125 94 175 255 03
m vaneetha mil J.F nboa, J.E nboa, J.E nnes	Wed-Keynote Thu-M-V-5-1 Tue-A-V-2-3 Thu-M-V-1-8 Tue-A-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Fri-A-SS-1-1 Tue-A-V-6-6	137 188 129 85 181 65 125 94 175 255 93
m vaneetha mil J.F nboa, J.E nnes ng Sriram	Wed-Keynote Thu-M-V-5-1 Wed-M-V-2-3 Thu-A-V-2-3 Thu-M-V-1-8 Tue-A-V-2-9 Wed-M-V-2-9 Tue-A-SS-1-1 Fri-A-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-1	137 188 129 85 181 65 125 94 175 255 93 96
m vaneetha onil J.F. ila nboa, J.E nnes Sriram	Wed-Keynote Thu-M-V-5-1 Wed-M-V-4-10 Tue-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-11 Wed-E-SS-11 Tue-A-SS-11 Tue-A-SS-2-1 Tue-A-SS-2-11	137 188 129 85 181 65 125 94 175 255 93 96 98
m vaneetha onil J.F nila nboa, J.E nnes Sriram	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-2-3 Thu-M-V-2-3 Tue-A-V-2-9 Tue-A-V-2-9 Tue-A-SS-1-1 Fri-A-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-11 Wed-M-SS-2-11 Wed-M-SS-1-6	137 188 129 85 181 65 125 94 175 255 93 96 98 137
m vaneetha onil J.F nila nboa, J.E nnes ng Sriram	Wed-Keynote Thu-M-V-5-1 Tue-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Tue-A-SS-2-1 Tue-A-SS-2-1 Tue-A-SS-2-1 Wed-M-SS-1-6 Wed-Survey	$137 \\ 188 \\ 129 \\ 85 \\ 181 \\ 65 \\ 125 \\ 94 \\ 175 \\ 255 \\ 93 \\ 96 \\ 98 \\ 137 \\ 137 \\ 137 \\$
m vaneetha mil J.F nila nboa, J.E nnes ng Sriram	Wed-Keynote Thu-M-V-5-1 Tue-A-V-2-3 Thu-M-V-2-3 Tue-A-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Fri-A-SS-1-1 Tue-A-SS-2-1 Tue-A-SS-2-11 Wed-M-SS-1-1 Wed-SS-1-1 Wed-SS-2-11	$137 \\ 188 \\ 129 \\ 85 \\ 181 \\ 65 \\ 125 \\ 94 \\ 175 \\ 255 \\ 93 \\ 96 \\ 98 \\ 137 \\ 137 \\ 207 \\$
m vaneetha onil J.F nboa, J.E nnes ng Sriram	Wed-Keynote Thu-M-V-5-1 Thu-M-V-3-10 Tue-A-V-2-3 Thu-M-V-1-8 Tue-M-V-2-9 Wed-M-V-2-9 Tue-A-SS-1-1 Fri-A-SS-1-1 Tue-A-SS-2-11 Tue-A-SS-2-11 Wed-Survey Thu-A-V-4-6	137 188 129 85 181 65 24 175 255 93 96 98 137 137 207 208
m vaneetha onil J.F nboa, J.E nnes ng Sriram	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-2-3 Thu-M-V-2-3 Thu-M-V-2-9 Tue-A-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-6 Fri-A-V-5-6	137 188 129 85 181 65 125 94 175 255 93 96 98 137 137 207 208 252
m vaneetha onil J.F nila nboa, J.E nnes Sriram Sriram	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-2-3 Thu-M-V-2-3 Thu-M-V-2-9 Tue-A-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Fri-A-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-1 Fri-A-V-5-3 Tue-E-V-4-8	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 125\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 137\\ 207\\ 208\\ 252\\ 111 \end{array}$
m vaneetha mil nboa, J.E nboa, J.E nnes Sriram eet	Wed-Keynote Thu-M-V-5-1 Tue-A-V-2-3 Thu-M-V-2-3 Thu-M-V-2-3 Tue-A-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Tue-A-SS-2-1 Tue-A-SS-2-1 Tue-A-SS-2-1 Tue-A-SS-2-1 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-6 Fri-A-V-5-3 Tue-E-V-4-8 Tue-E-V-4-8	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 125\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\end{array}$
m vaneetha mil J.F nboa, J.E nboa, J.E nnes Sriram Sriram	Wed-Keynote Thu-M-V-5-1 Tue-A-V-2-3 Thu-M-V-2-3 Thu-M-V-2-6 Wed-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Fri-A-SS-1-1 Tue-A-SS-2-1 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-6 Fri-A-V-5-3 Tue-E-V-4-8 Tue-E-V-2-6 Thu-A-V-2-6	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204 \end{array}$
m vaneetha onil J.F nboa, J.E nboa, J.E nnes ng Sriram Sriram Marie	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-2-3 Thu-M-V-2-3 Thu-M-V-2-9 Tue-A-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-11 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-3 Thu-A-V-4-3 Tue-E-V-4-8 Tue-E-V-2-6 Thu-A-V-3-1 Thu-A-V-3-1	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204\\ 180\end{array}$
m vaneetha onil J.F nla nboa, J.E nnes ng Sriram sriram tiin	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-2-3 Thu-M-V-2-3 Thu-M-V-2-9 Tue-A-V-2-9 Tue-A-SS-1-1 Wed-M-V-2-9 Tue-A-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-6 Fri-A-V-5-3 Tue-E-V-4-8 Tue-E-V-2-6 Thu-A-V-3-1 Thu-M-V-1-1 Thu-M-V-1-1	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 125\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204\\ 180\\ 0\\ 204\\ 170\\ \end{array}$
m vaneetha mil nboa, J.E nboa, J.E nnes ng Sriram sriram ttin on	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-A-V-2-3 Thu-M-V-2-3 Thu-M-V-2-9 Tue-A-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Fri-A-SS-2-1 Tue-A-V-6-6 Tue-A-SS-2-1 Tue-A-SS-2-1 Wed-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-1 Fri-A-V-5-3 Tue-E-V-2-6 Thu-A-V-3-1 Thu-M-V-1-1 Thu-M-V-1-1	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 125\\ 94\\ 175\\ 255\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204\\ 180\\ 179\\ 179\\ 122\\ 112\\ 106\\ 204\\ 180\\ 179\\ 122\\ 122\\ 122\\ 122\\ 122\\ 122\\ 122\\ 12$
m vaneetha onil J.F nboa, J.E nboa, J.E nnes ng Sriram sriram y Mario	Wed-Keynote Thu-M-V-5-1 Tue-A-V-2-3 Thu-M-V-2-3 Thu-M-V-2-6 Wed-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Tue-A-SS-2-11 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-6 Fri-A-V-5-11 Wed-M-SS-1-16 Wed-Survey Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-4-10 Thu-A-V-3-10 Thu-A-V-3-10 Thu-M-O-3-6 Tue-E-V-5-4	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 255\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204\\ 180\\ 179\\ 113\\ 207\\ 204\\ 180\\ 179\\ 113\\ 204\\ 180\\ 179\\ 113\\ 207\\ 204\\ 180\\ 179\\ 113\\ 207\\ 204\\ 180\\ 179\\ 113\\ 207\\ 208\\ 204\\ 180\\ 179\\ 113\\ 207\\ 208\\ 208\\ 108\\ 108\\ 108\\ 108\\ 108\\ 108\\ 108\\ 1$
m vaneetha onil J.F nboa, J.E nnes ng Sriram ttin , Mario	Wed-Keynote Thu-M-V-5-1 Tue-A-V-2-3 Thu-M-V-2-3 Thu-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Tue-A-SS-1-1 Tue-A-SS-2-1 Tue-A-SS-2-1 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-8 Tue-E-V-5-3 Tue-E-V-5-4 Tue-E-V-5-8	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 94\\ 175\\ 255\\ 93\\ 96\\ 8\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204\\ 180\\ 179\\ 113\\ 113\\ 113\\ 207\\ 204\\ 180\\ 179\\ 113\\ 113\\ 207\\ 204\\ 180\\ 100\\ 100\\ 100\\ 100\\ 100\\ 100\\ 100$
m vaneetha onil J.F nboa, J.E nnes ng Sriram fon , Mario g	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-2-3 Thu-M-V-2-3 Thu-M-V-2-9 Tue-A-V-2-9 Tue-A-SS-1-1 Wed-M-V-2-9 Tue-A-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-8 Tue-E-V-4-8 Tue-E-V-4-8 Tue-E-V-4-8 Tue-E-V-4-8 Tue-E-V-4-8 Tue-E-V-4-8 Tue-E-V-4-8 Tue-E-V-5-8 Fri-M-O-2-3	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 125\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204\\ 180\\ 179\\ 113\\ 221\\ \end{array}$
m vaneetha onil J.F nila nboa, J.E nnes ng Sriram sriram y Mario g	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-2-3 Thu-M-V-2-3 Thu-M-V-2-3 Tue-A-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Wed-E-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-11 Tue-A-SS-2-11 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-1 Fri-A-V-5-3 Tue-E-V-4-8 Tue-E-V-2-6 Thu-A-V-3-1 Thu-M-V-1-1 Thu-M-V-3-1 Thu-M-V-3-1 Thu-M-V-3-1 Thu-E-V-5-4 Tue-E-V-5-4 Fri-M-O-2-3 Fri-M-O-2-4	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 125\\ 94\\ 175\\ 255\\ 94\\ 175\\ 255\\ 94\\ 175\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204\\ 180\\ 179\\ 113\\ 121\\ 221\\ 221\\ \end{array}$
m vaneetha onil nboa, J.E nboa, J.E ng Sriram eet tiin y Mario g g	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-A-V-2-3 Thu-M-V-2-3 Thu-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Fri-A-SS-1-1 Tue-A-SS-2-1	137 188 129 85 181 65 255 94 175 255 93 96 98 137 137 207 208 252 211 106 204 180 179 113 221 211
m vaneetha onil J.F nboa, J.E nboa, J.E ng Sriram Sriram g g	Wed-Keynote Thu-M-V-5-1 Tue-A-V-2-3 Thu-M-V-2-3 Thu-M-V-2-3 Wed-M-V-2-6 Wed-M-V-2-9 Tue-A-SS-1-1 Wed-E-SS-1-1 Tue-A-SS-2-1 Tue-A-SS-2-1 Tue-A-SS-2-11 Wed-M-SS-1-6 Wed-Survey Thu-A-V-6-6 Fri-A-V-5-3 Tue-E-V-4-8 Tue-E-V-2-6 Thu-A-V-3-1 Thu-M-V-1-1 Thu-M-V-1-1 Thu-M-V-3-6 Tue-E-V-5-8 Fri-M-O-2-3 Fri-M-O-2-4 Tue-E-V-4-2	137 188 129 85 181 65 94 175 255 93 96 8 8 137 207 208 252 111 106 204 180 179 113 123 221 110 234
m vaneetha onil J.F nboa, J.E nboa, J.E nnes ng Sriram g g g	Wed-Keynote Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-5-1 Thu-M-V-2-3 Thu-M-V-2-3 Thu-M-V-2-9 Tue-A-V-2-9 Tue-A-SS-1-1 Wed-M-SS-1-1 Tue-A-V-6-6 Tue-A-SS-2-11 Wed-Survey Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-1 Thu-A-V-4-3 Tue-E-V-4-8 Tue-E-V-4-8 Tue-E-V-4-8 Tue-E-V-5-8 Fri-M-O-2-3 Fri-M-V-6-2 Fri-M-V-6-2	$\begin{array}{c} 137\\ 188\\ 129\\ 85\\ 181\\ 65\\ 125\\ 94\\ 175\\ 255\\ 93\\ 96\\ 98\\ 137\\ 137\\ 207\\ 208\\ 252\\ 111\\ 106\\ 204\\ 180\\ 179\\ 113\\ 113\\ 221\\ 1221\\ 110\\ 234\\ 233\\ \end{array}$

INTERSPEECH 2021

García Perera, Leibny P.	Wed-M-V-1-1	122	Goel, Sri
	Thu-M-O-3-6	179	Goginen
	Thu-M-V-4-4	187	GOK, All Gölge F
Gard David	Thu-M-0-2-2	178	Gonge, L
Guru, Duriu	Fri-A-O-1-4	241	Gómez.
Garg, Shefali	Tue-A-V-4-6	89	Gomez,
Garg, Vineet	Fri-M-V-6-4	234	Gomez-A
Garner, Philip N	Thu-M-O-3-4	179	Gong, Cl
Garrison, Jake	Tue-E-V-3-11	110	6 D
Gaspers, Judith	Tue-E-V-4-7	111	Gong, Ro
Gauder Lara	Thu-A-SS-2-4	234	Cong Vi
Gaur. Neerai	Wed-F-V-3-6	166	Goilg, 11
Gaur, Yashesh	Thu-A-V-2-2	202	
	Fri-A-V-1-2	243	
Gauvain, Jean-Luc	Thu-M-SS-2-6	196	
Gauvain, Jodie	Thu-M-SS-2-6	196	
Gaydecki, Patrick	Thu-A-O-2-3	200	Cong Vi
Ge, Chunyu	Thu-Δ-V-3-12	206	Gonzale
Ge. Wanving	Fri-M-V-7-13	239	Gonzále
Ge, Zhenhao	Tue-M-V-2-10	66	Gonzale
Geislinger, Robert	Thu-M-S&T-1-5	197	Gopal, V
Gelin, Lucile	Fri-M-O-1-5	220	Gopalak
Conduct Cádric	Fri-A-S&T-1-4	261	Gope, Di
Genarol, Cearle	1 ue-E-O-2-6	102	Gordon,
Geng Mengzhe	Fri-A-SS-2-4	257	Gorthi F
	Fri-A-SS-2-6	258	Gorthi S
	Fri-A-SS-2-9	259	Gósy, Ma
Geng, Wang	Thu-M-SS-1-3	194	Gosztoly
Georges, Marc-Antoine .	Thu-A-O-1-5	199	
Georgiou, Effhymios	<u>Thu-M-O-2-3</u>	1/8	Graetzei
Georgio Dhilip	Fri-A-55-2-15	200	Graham,
Gerazov Branislav	Wed-F-V-4-4	168	Gran Te
Gerazov, Branslav	Thu-M-V-6-6	191	Green. I
Gerczuk, Maurice	Tue-M-SS-1-1	77	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
German, James S	Wed-M-O-2-4	120	
Gerstenberger, A	Wed-A-O-1-4	138	
Getman, Yaroslav	Wed-M-O-1-3	119	
Ghahremani Pegah	Fri-A-U-1-5	242	Gretter
Ghio. Alain	Fri-A-V-4-8	251	Gris, Luc
Ghodsi, Mohammadreza	Tue-A-V-4-7	90	Grivolla,
Ghosh, Prasanta Kumar	Tue-A-V-2-3	85	Gromog
	Tue-A-SS-2-1	96	Grondin
	Tue-E-V-3-1	107	Gruber,
	Wed-F-V-1-7	161	Gu, bili
	Thu-M-V-1-11	182	Gu, Feng
	Fri-A-S&T-1-8	261	Gu, Jian
Gibson, Andy	Tue-E-SS-1-1	117	Gu, Jiata
Ciberry Mark	Tue-E-SS-1-2	117	Gu, Nan
GIDSON, Mark	Fri-M-V-2-7	220	Gu, Ron
Gili Fivela Barbara	Thu-A-V-1-3	180	
	Fri-M-V-2-12	227	Gu. Yile
Giller, Maximilian	Wed-A-S&T-1-3	157	Gudupu
Gillick, Jon	Wed-E-V-2-2	163	Gudur, O
Gilmartin, Emer	1ue-A-V-5-10	92 1 = 4	Guevara
Gillellez Pastol, Aulia .	wea-A-V-6-10	213	Guna, Pr
Gimeno, Pablo	Fri-A-0-1-1	241	Guilly A
Ginsburg, Boris	Wed-M-O-1-4	119	Guily, H
	Wed-E-V-6-8	174	
	Thu-A-SS-1-10	215	Guncele
	Fri-A-V-1-13	245	Gundog
Giollo Manuel	Wed-F-V-1-1	160	Guo, Ch
Giri. Ritwik	Tue-E-V-2-5	106	Guo, Lia
Girin, Laurent	Tue-M-O-3-1	62	Guo. Per
	Thu-A-O-1-5	199	Guo, Sha
Ciuliotti Cimere	Fri-M-O-2-1	220	Guo, Wu
Glass James	Tue-M-O-2-5	01 82	0 7
G1055, Juliito	Wed-M-V-2-12	126	Guo, Zho
	Thu-M-V-1-12	182	Guruju,
	Thu-M-V-2-8	184	Gutierre
	Thu-M-V-7-7	193	Gutierre
	Thu-A-V-6-1	211	Gutierre
Cluchko Alakaandr	Thu-A-SS-1-4	214	
Giusiiko, Aleksaflar Gmyr Robert	Tue-E-V-6-3	$\frac{111}{115}$	Cost C
Gobl. Christer	Wed-E-V-4-10	169	Guizman
Goderre, Andrew	Fri-M-O-1-1	219	Gwon V
Goel. Nagendra	Thu-M-O-3-6	179	5511, 1

Cool Srighti		115
	Tue-E-V-0-7	115
Gogineni, K	Wed-A-S&T-1-7	157
Gok Alican	Eri-A-0-2-2	242
		011
Golge, Eren	Thu-A-V-5-12	211
Gomez Angel M	Tue-A-SS-2-2	96
Cáman Emilia	Tuc / 05 2 2	00
Gomez, Emilia	Tue-A-SS-2-9	98
Gomez. Juan Zuluaga	Thu-M-SS-2-5	196
Comoz Alania A	Tue A (2 2 2	06
Gomez-Alams, A.	Tue-A-55-2-2	90
Gong. Cheng	Tue-M-V-2-1	64
8, 8		65
-	Tue-IVI-V-2-5	05
Gong, Rong	Fri-M-O-1-1	219
Cong Yun	Tuo-E-\/-5-2	112
	Tue-E-V-3-2	112
Gong, Yifan	Tue-A-V-4-10	- 90
0.	Tuo-E-\/-5-0	114
	Tue-E-V-5-9	117
	Wed-M-V-6-2	133
	Wod-M-V-6-5	134
	weu-w-v-o-s	107
	Wed-E-V-3-14	167
	Thu.A.V.2.3	203
		201
	1hu-A-V-2-10	204
Gong. Yuan	Tue-A-V-1-7	83
Congolog Dachol		241
GOIIZalez, Racilei	Fri-A-0-1-4	241
González Hautamäki, R.	Wed-M-V-4-6	128
Conzoloz Lonoz Joco A		06
GOIIZalez-Lopez, Jose A.	Tue-A-33-2-2	90
Gopal, Vishak	Wed-E-SS-1-1	175
Conalakrishnan Kailash	Wod E V 2 12	167
Soparaki isinian, Kanash	weu-E-v-5-12	107
Gope, Dipanjan	Thu-M-V-1-11	182
Gordon Matthew	Wed-E-V-4-1	168
Common T 1	- Weu E V 4 1	100
Gorman, Kyle	Fri-A-V-1-13	245
Gorthi, Rama Krishna SS	Thu-4-0-2-2	190
Contra Coloritor		100
Gorthi, Subrahmanyam	Thu-A-O-2-2	199
Gósy Mária	Wed-F-V-4-7	169
		105
Gosztolya, Gabor	Iue-M-SS-I-I0	-79
	Wed-A-O-3-6	140
C		110
Graetzer, Simone	I ue-A-V-3-7	- 88
Graham Calbert	Fri-M-V-1-11	224
Current and a standard and a standard and a standard and a standard a s	T	
Grammenos, Andreas	Iue-M-SS-I-I	- 7 7
Grau. Teresa	Tue-A-SS-2-2	- 96
Croon Jordan P	Eri A CC 2 1	257
Green, Joruan K	FIFA-33-2-1	231
	Fri-A-SS-2-2	257
	Eri_A_55_7_5	258
		250
	Fri-A-SS-2-10	259
	Fri-A-SS-2-12	259
		260
	FII-A-55-2-14	200
Gretter, R	Fri-M-O-1-2	219
Gretter, R Gris Lucas Rafael S	Fri-A-55-2-14 Fri-M-O-1-2	219 78
Gretter, R Gris, Lucas Rafael S	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4	219 78
Gretter, R Gris, Lucas Rafael S Grivolla, Jens	Fri-A-55-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3	219 78 196
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V	FTI-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4	219 78 196
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4	219 78 196 197
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4	219 78 196 197 207
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grůber. Martin	Fri-A-35-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2	219 78 196 197 207 260
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grüber, Martin	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2	219 78 196 197 207 260
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grůber, Martin Gu, Bin	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-4 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-M-V-1-2	219 78 196 197 207 260 63
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grůber, Martin Gu, Bin	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-M-V-1-2 Tue-E-V-1-3	219 78 196 197 207 260 63 103
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Guber, Martin Gu, Bin	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-M-V-1-2 Tue-E-V-1-3 Wood EV F 11	219 78 196 197 207 260 63 103
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grüber, Martin Gu, Bin Gu, Feng	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11	219 78 196 197 207 260 63 103 172
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Guber, Martin Gu, Bin Gu, Feng Gu, Jianjun	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-M-V-1-2 Tue-E-V-5-11 Tue-E-V-2-3	219 78 196 197 207 260 63 103 172 106
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grůber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wood A V 6 4	219 78 196 197 207 260 63 103 172 106
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grůber, Martin Gu, Bin Gu, Feng Gu, Jianjun Gu, Jiatao	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4	219 78 196 197 207 260 63 103 172 106 153
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grůber, Martin Gu, Bin Gu, Feng Gu, Jianjun Gu, Jiatao Gu, Nan	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5	219 78 196 197 207 260 63 103 172 106 153 85
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grüber, Martin Gu, Bin Gu, Feng Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Ronozzhi	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-0-1-5	219 78 196 197 207 260 63 103 172 106 153 85 80
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Guðber, Martin Gu, Bin Gu, Feng Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-M-V-1-2 Tue-E-V-5-11 Tue-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5	219 78 196 197 207 260 63 103 172 106 153 85 85
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromdin, François Grûber, Martin Gu, Bin Gu, Feng Gu, Jianjun Gu, Jianjun Gu, Nan Gu, Rongzhi	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-B-V-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Wed-E-O-1-4	219 78 196 197 207 260 63 103 172 106 153 85 80 158
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grúber, Martin Gu, Bin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Tue-A-V-5-11 Tue-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-A-V-2-5 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6	219 78 196 197 207 260 63 103 172 106 153 85 80 158 235
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Feng Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-E-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6	219 78 196 197 207 260 63 103 172 106 153 85 80 158 235
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-6-4 Tue-A-V-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3	219 78 196 197 207 260 63 103 172 106 153 85 80 158 235 95
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gu, Bin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gu, Yile Gudupudi, Leela K	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3	219 78 196 197 207 260 63 103 172 106 153 85 80 158 235 95 159
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gu, Yile Gudupudi, Leela K Gudupudi, Leela K	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-0-1-5 Wed-E-0-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-0-2-3 Wed-E-0-2-3	2199 78 1966 1977 2007 2600 63 103 1722 1066 1533 85 800 1588 2355 95 1599 1255
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-B-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-0-1-5 Wed-E-0-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-0-2-3 Wed-E-0-2-3	2199 78 1966 1977 2070 2070 2070 63 1097 1006 1533 85 800 1588 2355 959 1255
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gu, Bin Gu, Bin Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gu, Yile Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-5-11 Tue-E-V-5-11 Tue-E-V-5-11 Tue-E-V-5-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-M-V-2-11 Fri-M-V-2-9	2199 78 196 197 207 260 63 103 172 106 153 85 80 158 235 95 159 125 226
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudupudi, Leela K Guevara-Rukoz, Adriana Guba Prithwiiit	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Wed-E-O-1-4 Fri-M-V-6-4 Fri-M-V-2-11 Fri-M-V-2-9 Fri-M-V-5-10	2199 78 196 197 207 260 63 103 172 106 153 85 80 158 80 158 235 225 226 226 223
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Guber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gu, Rongzhi Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guhati Acaral	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-3-11 Tue-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Wed-E-O-2-3 Wed-B-O-2-3 Wed-M-V-2-11 Fri-M-V-5-10	2199 78 1966 1977 2077 2600 63 103 1722 1066 1533 85 800 1588 2355 255 2266 2333 2256 2236 2332 2260 2333 2255 2260 2332 235 2260 2332 235 2260 235 2260 235 235 2260 235 235 2260 235 235 235 235 235 235 235 235 235 235
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromdin, François Grûber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-R-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-M-V-2-11 Fri-M-V-2-10 Wed-M-V-6-4	2199 78 196 197 207 260 63 103 172 106 153 85 80 153 85 85 85 85 85 85 85 85 85 85 85 85 85
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Guber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Gulat, Amelia L	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-M-V-2-11 Fri-M-V-2-9 Fri-M-V-2-10 Wed-M-V-6-4 Tue-M-V-6-4	$2199 \\ 78 \\ 78 \\ 196 \\ 197 \\ 207 \\ 260 \\ 63 \\ 103 \\ 172 \\ 207 \\ 106 \\ 153 \\ 85 \\ 80 \\ 158 \\ 235 \\ 95 \\ 159 \\ 125 \\ 226 \\ 233 \\ 133 \\ 76 \\ 76 \\ 76 \\ 76 \\ 76 \\ 76 \\ 76 \\ $
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gu, Bin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gully, Amelia J	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-A-V-2-5 Tue-A-V-2-5 Tue-A-V-2-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-	2199 78 1966 1977 2007 2600 633 103 1722 1066 1533 808 1588 2055 2266 2333 1336 766 2333 766
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Guðber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gulati, Anmol Gully, Amelia J	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-0-1-5 Wed-A-V-6-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-1-4 Fri-M-V-6-10 Wed-E-O-2-3 Wed-M-V-2-11 Fri-M-V-2-10 Wed-M-V-2-10 Wed-M-V-5-10 Wed-M-V-7-7 Wed-M-V-5-8	2199 78 1966 197 207 2600 63 103 172 1066 153 85 800 158 235 2266 233 159 125 2262 233 376 132
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gu, Bin Gu, Bin Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gu, Rongzhi Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gully, Amelia J	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-5-11 Tue-E-V-5-11 Wed-E-V-5-11 Wed-E-O-2-3 Wed-M-V-2-11 Fri-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-8 Fri-M-V-5-8 Fri-M-V-5-8	2199 78 196 197 2077 22600 63 103 172 106 153 85 80 0 158 82 355 95 159 125 226 233 133 76 132 226
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromdin, François Grondin, François Gruber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gulati, Anmol Gully, Amelia J	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-R-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-3 Wed-A-V-6-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-1-4 Fri-M-V-5-10 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-O-2-3 Wed-B-V-5-10 Wed-M-V-6-4 Tue-M-V-5-8 Fri-M-V-2-8 Fri-M-V-2-8	2199 78 196 197 207 260 63 103 172 207 260 63 103 153 158 235 95 526 233 159 125 226 233 133 76 132 226 6 233 133 76 132 226 6 233 133 76 132 226 6 233 133 76 132 226 226 233 133 76 132 226 226 233 133 76 26 226 226 236 236 226 226 237 26 26 26 26 27 26 26 26 27 26 26 26 27 26 26 26 27 26 26 27 26 26 26 27 26 26 26 27 26 26 26 27 26 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 27 26 27 26 27 27 27 26 27 27 26 27 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 27 26 26 27 27 26 27 26 27 26 27 26 27 26 27 26 27 27 26 27 27 26 27 27 26 27 27 26 27 27 26 27 27 26 27 27 26 27 27 26 27 27 27 26 27 27 27 26 27 27 27 27 27 27 27 27 27 27 27 27 27
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Guber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gulati, Anmol	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-M-V-2-11 Fri-M-V-2-10 Wed-M-V-2-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-8 Fri-M-V-2-8 Fri-M-V-2-8	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 2207\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 80\\ 158\\ 235\\ 226\\ 233\\ 133\\ 76\\ 132\\ 226\\ 160\\ 160\\ \end{array}$
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Rongzhi Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gully, Amelia J Gunceler, Deniz Gundogdu, Batuhan	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-6-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-1-4 Fri-M-V-2-9 Fri-M-V-2-11 Fri-M-V-2-9 Fri-M-V-5-8 Fri-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 85\\ 85\\ 235\\ 95\\ 159\\ 225\\ 223\\ 133\\ 76\\ 132\\ 226\\ 60\\ 160\\ 242\\ \end{array}$
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Guber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Jiatao Gu, Nan Gu, Nan Gu, Nan Gu, Nan Gudupudi, Leela K Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gudur, Gautham J Gulati, Anmol Gully, Amelia J Gunceler, Deniz Gundogdu, Batuhan	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Wed-A-V-6-4 Fri-M-V-6-4 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-M-V-2-11 Fri-M-V-2-10 Wed-M-V-6-4 Tue-M-V-5-8 Fri-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-2-2	2199 78 196197 2077207 22070 633103 1722 106153 855 800 1582 235295 15991255 2266233 133376 13222661602 2422662 2422662
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gunceler, Deniz	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-V-2-11 Fri-M-V-2-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4	2019 78 196 197 2077 2600 63 103 172 106 153 85 800 158 235 95 125 2260 233 133 766 1322 2260 2260 2333 1333 766 1226 2260 2
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grůber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Jiatao Gu, Nan Gu, Nan Gu, Nan Gu, Nan Gudupudi, Leela K Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gudur, Gautham K Gudur, Gautham J Gulati, Anmol Gully, Amelia J Gundogdu, Batuhan Guo, Chenkai Guo, Chenkai	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-5-31 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-A-V-6-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-1-4 Fri-M-V-2-11 Fri-M-V-2-11 Fri-M-V-2-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-11 Fri-A-O-2-2 Tue-M-V-1-4 Wed-M-SS-1-1	2019 78 196 197 2070 2260 63 103 172 106 153 855 858 805 235 2266 2333 76 132 2263 1333 76 132 2266 1600 2422 633 136
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gu, Bin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gulati, Anmol Gulati, Anmol Gulati, Anmol Gunceler, Deniz Gundogdu, Batuhan Guo, Chenkai Guo, Liang	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-M-V-2-11 Fri-M-V-2-10 Wed-M-V-6-4 Tue-M-V-7-7 Wed-M-V-5-8 Fri-M-V-2-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4 Wed-M-V-1-4	2019 78 196 197 207 260 63 103 172 260 63 103 172 106 153 85 80 158 235 59 159 125 226 233 133 76 61 32 226 233 160 125 226 233 133 76 122 226 233 133 76 122 226 233 133 226 226 233 133 226 226 233 133 226 226 226 233 133 226 226 226 233 125 226 226 226 226 226 226 226 226 226 2
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Nan Gu, Nan Gu, Nan Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gulati, Anmol Gulati, Anmol Gully, Amelia J Gundogdu, Batuhan Guo, Chenkai	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-H-V-5-10 Wed-M-V-5-10 Wed-M-V-5-8 Fri-M-0-22 Fri-M-0-22 Fri-M-0-22 Fri-M-0-22 Fri-M-0-22 Fri-M-0-22 Fri-M-0-22 Fri-M-0-21 Fr	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 235\\ 235\\ 235\\ 226\\ 63\\ 133\\ 76\\ 132\\ 226\\ 61\\ 102\\ 242\\ 26\\ 136\\ 136\\ 126\\ 242\\ 242\\ 242\\ 136\\ 126\\ 126\\ 126\\ 126\\ 126\\ 126\\ 126\\ 12$
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gu, Bin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Jiatao Gu, Nan Gu, y, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gulati, Anmol Gunceler, Deniz Gundogdu, Batuhan Guo, Chenkai Guo, Liang	Fri-A-S3-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-M-V-2-11 Fri-M-V-2-9 Fri-M-V-2-11 Fri-M-V-2-9 Fri-M-V-2-11 Fri-M-V-2-9 Fri-M-V-2-11 Fri-M-V-2-8 Wed-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4 Wed-E-SS-1-7 Wed-E-SS-1-7	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 80\\ 158\\ 235\\ 959\\ 159\\ 2266\\ 233\\ 766\\ 132\\ 226\\ 132\\ 136\\ 132\\ 226\\ 132\\ 132\\ 136\\ 132\\ 132\\ 132\\ 132\\ 132\\ 132\\ 132\\ 132$
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Grondin, François Gu, Bin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gulati, Anmol Gulati, Anmol Gully, Amelia J Gunceler, Deniz Gundogdu, Batuhan Guo, Chenkai Guo, Pengcheng	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-6-4 Tue-A-V-6-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-M-V-5-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-4 Wed-M-V-1-4 Wed-M-SS-1-1 Wed-E-SS-1-7 Thu-A-SS-1-2	2019 78 196 197 2070 260 63 103 172 2060 63 103 172 106 153 85 85 85 159 95 125 226 233 133 76 132 226 60 122 226 63 103 172 105 153 85 85 159 95 125 226 233 133 76 122 226 226 226 105 153 105 155 159 125 226 226 105 155 159 125 226 226 105 155 159 125 226 226 105 155 159 125 226 226 226 155 159 125 226 226 226 155 159 125 226 226 226 155 159 125 226 226 226 155 159 125 226 226 226 226 155 226 226 226 155 226 226 226 226 226 155 226 226 226 226 226 226 155 226 226 226 226 226 226 226 226 226 2
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gubter, Martin Gu, Bin Gu, Jianjun Gu, Jiatao Gu, Jiatao Gu, Vile Gu, Nan Gu, Nan Gu, Nan Gu, Nan Gu, Vile Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gulati, Anmol Gulati, Anmol Gully, Amelia J Gunceler, Deniz Gunceler, Deniz Guo, Chenkai Guo, Liang Guo, Shaotong	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Wed-A-V-6-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-M-V-2-11 Fri-M-V-2-9 Fri-M-V-2-10 Wed-M-V-6-4 Tue-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4 Wed-M-SS-1-2 Tue-M-V-2-1	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 2207\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 80\\ 158\\ 235\\ 235\\ 235\\ 235\\ 226\\ 133\\ 76\\ 132\\ 226\\ 136\\ 160\\ 242\\ 263\\ 136\\ 176\\ 214\\ 64\\ 214\\ 64\\ 242\\ 246\\ 246$
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gruber, Martin Gu, Bin Gu, Feng Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gulati, Anmol Gundogdu, Batuhan Guo, Chenkai Guo, Liang Guo, Shaotong Guo, Wu	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-5-3 Wed-E-V-5-3 Wed-E-V-5-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-1-4 Fri-M-V-5-10 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Tue-M-V-5-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-2-12	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 2260\\ 63\\ 103\\ 172\\ 2260\\ 63\\ 103\\ 172\\ 106\\ 158\\ 80\\ 158\\ 80\\ 158\\ 235\\ 95\\ 159\\ 95\\ 125\\ 226\\ 233\\ 133\\ 76\\ 132\\ 226\\ 63\\ 136\\ 136\\ 126\\ 242\\ 63\\ 136\\ 214\\ 64\\ 63\end{array}$
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gu, Feng Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Jiatao Gu, Nan Gu, Nan Gu, Nan Gu, Vile Gudupudi, Leela K Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gudur, Gautham K Gulati, Anmol Gulati, Anmol Gully, Amelia J Gundogdu, Batuhan Guo, Chenkai Guo, Liang Guo, Shaotong Guo, Wu	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-5-31 Wed-E-V-5-31 Wed-E-V-5-31 Wed-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-A-V-6-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-1-4 Fri-M-V-2-11 Fri-M-V-2-9 Fri-M-V-2-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-11 Fri-A-O-2-2 Tue-M-V-1-4 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-2-11 Tue-M-V-2-11	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 2207\\ 2207\\ 2207\\ 2307\\ 103\\ 103\\ 172\\ 106\\ 158\\ 85\\ 85\\ 85\\ 85\\ 85\\ 85\\ 235\\ 235\\ 235\\ 159\\ 125\\ 226\\ 133\\ 76\\ 132\\ 226\\ 133\\ 76\\ 132\\ 226\\ 133\\ 136\\ 176\\ 214\\ 64\\ 64\\ 63\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102\\ 102$
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Grondin, François Gu, Bin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Nan Gu, Nan Gu, Rongzhi Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Guevara-Rukoz, Adriana Guha, Prithwijit Gulati, Anmol Gulati, Anmol Gundogdu, Batuhan Guno, Chenkai Guo, Liang Guo, Pengcheng Guo, Wu	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-1-4 Fri-M-V-2-11 Fri-M-V-2-10 Fri-M-V-2-11 Fri-M-V-5-10 Wed-M-V-6-4 Tue-M-V-7-7 Wed-M-V-5-8 Fri-M-V-2-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-1-2 Tue-M-V-1-2 Tue-M-V-1-2	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 260\\ 63\\ 103\\ 172\\ 216\\ 103\\ 85\\ 80\\ 158\\ 80\\ 158\\ 235\\ 959\\ 125\\ 226\\ 233\\ 133\\ 66\\ 126\\ 214\\ 63\\ 136\\ 6214\\ 63\\ 103\\ 226\\ 160\\ 214\\ 63\\ 103\\ 226\\ 160\\ 214\\ 212\\ 226\\ 160\\ 214\\ 226\\ 160\\ 214\\ 226\\ 160\\ 214\\ 226\\ 160\\ 214\\ 226\\ 160\\ 214\\ 226\\ 160\\ 214\\ 226\\ 160\\ 214\\ 226\\ 226\\ 160\\ 214\\ 226\\ 160\\ 214\\ 226\\ 226\\ 226\\ 226\\ 226\\ 226\\ 226\\ 22$
Gretter, R Gris, Lucas Rafael S Grivolla, Jens Gromoglasov, V Gromodin, François Grůber, Martin Gu, Bin Gu, Jianjun Gu, Jianjun Gu, Jiatao Gu, Jiatao Gu, Nan Gu, Nan Gu, Nan Gu, Nan Gu, Nan Gudupudi, Leela K Gudupudi, Leela K Gudur, Gautham K Gudur, Gautham K Gudur, Gautham K Gudur, Gautham K Gudur, Gautham S Gudur, Amelia J Gulati, Anmol Gully, Amelia J Gundogdu, Batuhan Guo, Chenkai Guo, Shaotong Guo, Shaotong Guo, Zhe-chen	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-3 Wed-A-V-6-4 Fri-M-V-2-15 Wed-E-O-1-5 Wed-E-O-1-5 Wed-E-O-1-5 Wed-E-O-1-5 Wed-E-O-1-5 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-V-2-10 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-2-10 Fri-M-V-2-10	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 235\\ 95\\ 159\\ 226\\ 133\\ 76\\ 132\\ 226\\ 136\\ 136\\ 242\\ 63\\ 136\\ 214\\ 64\\ 3\\ 133\\ 226\end{array}$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-3 Thu-M-S&T-1-3 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Wed-A-V-6-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-1-4 Fri-M-V-2-11 Fri-M-V-2-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-11 Fri-A-O-2-2 Tue-M-V-1-4 Wed-M-SS-1-1 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-2-10 Wed-M-V-2-10 Wed-M-V-2-10	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 2260\\ 63\\ 103\\ 172\\ 210\\ 106\\ 153\\ 85\\ 80\\ 158\\ 235\\ 959\\ 125\\ 2266\\ 233\\ 76\\ 132\\ 2266\\ 132\\ 122\\ 122\\ 122\\ 122\\ 122\\ 122\\ 122$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-5-31 Wed-E-V-5-31 Wed-A-V-6-4 Tue-A-V-6-4 Tue-A-V-6-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Tue-A-V-5-10 Wed-M-V-5-10 Wed-M-V-2-10 Tue-E-V-1-3 Fri-M-V-2-10 Wed-A-S&T-16 Wed-A-S&T-16 Wed-A-S&T-16	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 85\\ 235\\ 95\\ 159\\ 226\\ 233\\ 133\\ 76\\ 132\\ 226\\ 160\\ 214\\ 63\\ 103\\ 226\\ 157\\ 621\\ 46\\ 3103\\ 226\\ 157\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 157\\ 103\\ 226\\ 103\\ 103\\ 226\\ 103\\ 103\\ 226\\ 103\\ 103\\ 226\\ 103\\ 103\\ 226\\ 103\\ 103\\ 226\\ 103\\ 103\\ 103\\ 226\\ 103\\ 103\\ 103\\ 103\\ 103\\ 103\\ 103\\ 103$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-K-V-4-4 Fri-A-S&T-1-2 Tue-E-V-5-11 Tue-E-V-5-11 Wed-E-V-5-11 Wed-E-V-5-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Wed-A-V-6-4 Fri-M-V-6-6 Tue-A-V-2-5 Wed-A-V-2-13 Fri-M-V-2-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-12 Tue-M-V-1-2 Tue-M-V-1-2 Tue-M-V-1-2 Tue-M-V-1-2 Tue-M-V-1-2 Tue-M-V-1-3 Fri-M-V-2-10 Wed-A-SS-1-3	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 206\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 85\\ 85\\ 85\\ 235\\ 235\\ 226\\ 133\\ 76\\ 132\\ 226\\ 133\\ 76\\ 132\\ 226\\ 133\\ 132\\ 226\\ 136\\ 176\\ 214\\ 64\\ 63\\ 103\\ 226\\ 157\\ 155\\ 155\\ 157\\ 155\\ 157\\ 155\\ 157\\ 155\\ 157\\ 155\\ 157\\ 155\\ 157\\ 155\\ 157\\ 155\\ 157\\ 157$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-5-3 Wed-E-V-5-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-1-4 Fri-M-V-2-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-12 Tue-M-V-2-1 Tue-M-V-2-1 Tue-M-V-2-1 Tue-M-V-2-10 Wed-A-S&T-1-6 Wed-A-S&T-1-6 Wed-A-S&T-13 Fri-M-V-2-10	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 2260\\ 63\\ 103\\ 172\\ 2260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 80\\ 158\\ 80\\ 158\\ 235\\ 95\\ 159\\ 226\\ 233\\ 133\\ 76\\ 132\\ 226\\ 136\\ 136\\ 214\\ 64\\ 63\\ 103\\ 226\\ 157\\ 155\\ 250\\ \end{array}$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-M-S&T-1-2 Tue-M-S-1-2 Tue-M-V-1-2 Tue-E-V-5-11 Tue-E-V-5-31 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-O-1-5 Wed-E-O-1-3 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-V-2-5 Wed-E-O-1-3 Wed-E-O-1-4 Fri-M-V-2-10 Wed-M-V-2-11 Fri-M-V-2-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-22 Tue-M-V-1-2 Tue-M-V-1-2 Tue-M-V-1-2 Tue-M-V-2-13 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-M-SS-1-3 Fri-M-V-2-10 <t< td=""><td>$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 2207\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 85\\ 85\\ 235\\ 216\\ 159\\ 125\\ 226\\ 133\\ 76\\ 132\\ 226\\ 133\\ 76\\ 132\\ 226\\ 133\\ 136\\ 176\\ 214\\ 64\\ 63\\ 103\\ 226\\ 157\\ 155\\ 259\\ 259\\ 212\\ 157\\ 155\\ 259\\ 259\\ 226\\ 125\\ 125\\ 259\\ 212\\ 125\\ 259\\ 226\\ 125\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 259\\ 226\\ 125\\ 259\\ 259\\ 226\\ 125\\ 259\\ 259\\ 259\\ 259\\ 259\\ 259\\ 259\\ 2$</td></t<>	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 2207\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 85\\ 85\\ 235\\ 216\\ 159\\ 125\\ 226\\ 133\\ 76\\ 132\\ 226\\ 133\\ 76\\ 132\\ 226\\ 133\\ 136\\ 176\\ 214\\ 64\\ 63\\ 103\\ 226\\ 157\\ 155\\ 259\\ 259\\ 212\\ 157\\ 155\\ 259\\ 259\\ 226\\ 125\\ 125\\ 259\\ 212\\ 125\\ 259\\ 226\\ 125\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 226\\ 125\\ 259\\ 259\\ 226\\ 125\\ 259\\ 259\\ 226\\ 125\\ 259\\ 259\\ 259\\ 259\\ 259\\ 259\\ 259\\ 2$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-V-2-5 Wed-E-O-1-4 Fri-M-V-2-10 Wed-E-O-1-3 Wed-E-O-1-4 Fri-M-V-2-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-11 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-2-10 Wed-A-SS-1-3 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-A-SS-2-9 Wed-A-O-3-1	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 2260\\ 63\\ 103\\ 172\\ 2260\\ 153\\ 85\\ 80\\ 158\\ 80\\ 158\\ 80\\ 158\\ 235\\ 959\\ 125\\ 226\\ 233\\ 766\\ 132\\ 226\\ 133\\ 766\\ 124\\ 63\\ 103\\ 226\\ 157\\ 155\\ 259\\ 139\\ 139\\ \end{array}$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-3 Wed-A-V-6-4 Fri-M-V-6-4 Fri-M-V-6-4 Fri-M-V-6-4 Fri-M-V-6-4 Fri-M-V-6-4 Fri-M-V-2-13 Wed-E-O-1-5 Wed-E-O-1-5 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-V-1-1 Fri-M-V-2-10 Wed-M-V-5-8 Fri-M-V-2-11 Tue-M-V-1-2 Tue-M-V-1-12 Tue-M-V-1-13 Fri-M-V-2-11 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-1-2 Tue-M-V-2-10 Wed-A-S&T-1-6 Wed-A-S&T-1-6 Wed-A-S&T-1-6 Wed-A-O-3-1	$\begin{array}{c} 219\\ 78\\ 196\\ 197\\ 207\\ 2207\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 85\\ 85\\ 85\\ 159\\ 215\\ 226\\ 133\\ 76\\ 132\\ 226\\ 136\\ 136\\ 136\\ 214\\ 64\\ 31\\ 3226\\ 157\\ 155\\ 259\\ 139\\ 249\\ 155\\ 259\\ 139\\ 94\end{array}$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-E-O-1-4 Fri-M-V-2-10 Wed-M-V-5-10 Wed-A-SS-1-1 Wed-A-SS-1-1 Wed-A-SS-1-1 SFri-M-V-2-10 Wed-A-SS-1-1	2219 78 196 197 2207 2600 63 103 1722 1006 153 858 800 1588 2355 959 1252266 2333 1337 61322266 2333766 1252266 2333766 1252266 242466 12666 12666 12666 126666 126666 126666 1266666 12666666 1266666666 12666666666666666666666666666666666666
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-M-S&T-1-4 Thu-M-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-13 Wed-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-2-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-V-5-10 Wed-E-O-2-3 Wed-E-O-2-3 Wed-H-V-2-10 Wed-M-V-2-10 Wed-M-V-2-12 Tue-M-V-1-4 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-1-4 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-2-10 Wed-A-SS-1-2 Tue-M-V-2-10 Wed-A-SS-1-3 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-M-V-2-10 Wed-A-SS-2-9 Wed-A-O-3-1-4 <	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 2267\\ 206\\ 31\\ 197\\ 206\\ 103\\ 103\\ 103\\ 103\\ 158\\ 85\\ 85\\ 85\\ 85\\ 159\\ 95\\ 125\\ 226\\ 233\\ 133\\ 76\\ 132\\ 226\\ 160\\ 214\\ 63\\ 136\\ 214\\ 63\\ 103\\ 226\\ 157\\ 155\\ 259\\ 139\\ 94\\ 141 \end{array}$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-5-31 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-A-V-2-5 Wed-A-V-2-11 Fri-M-V-2-0 Wed-M-V-2-11 Fri-M-V-2-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-2-2 Tue-M-V-1-4 Wed-M-SS-1-2 Tue-M-V-1-2 Tue-M-V-1-3 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-A-SS-1-2 Tue-M-V-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-1-3 Fri-A-SS-2-9 Wed-A-O-3-1 Tue-A-V-6-8 Wed-A-V-1-4 Thu-M-V-1-4	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 260\\ 63\\ 103\\ 172\\ 216\\ 103\\ 158\\ 80\\ 158\\ 235\\ 959\\ 125\\ 2266\\ 2333\\ 766\\ 132\\ 2266\\ 160\\ 242\\ 63\\ 136\\ 160\\ 244\\ 63\\ 103\\ 2266\\ 157\\ 155\\ 259\\ 139\\ 94\\ 141\\ 181\\ \end{array}$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-A-V-4-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-3 Wed-A-V-6-4 Fri-M-V-2-5 Wed-A-V-6-4 Fri-M-V-2-5 Wed-E-O-1-4 Fri-M-V-2-5 Wed-E-O-1-4 Fri-M-V-2-10 Wed-E-O-1-4 Fri-M-V-2-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-10 Wed-M-V-5-11 Tue-M-V-2-10 Wed-A-SS-1-2 Tue-M-V-2-10 Wed-A-SS-1-3 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-A-SS-29 Wed-A-O-3-11 Tue-A-V-6-8 Wed-A-V-1-4 Thu-M-V-1-6	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 2260\\ 63\\ 103\\ 172\\ 2260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 80\\ 158\\ 80\\ 158\\ 235\\ 95\\ 159\\ 125\\ 226\\ 233\\ 133\\ 76\\ 132\\ 226\\ 63\\ 136\\ 176\\ 6214\\ 64\\ 63\\ 103\\ 226\\ 614\\ 157\\ 155\\ 259\\ 94\\ 141\\ 181\\ 181\\ 258\\ 94\\ 141\\ 181\\ 181\\ 181\\ 181\\ 181\\ 181\\ 18$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-M-S&T-1-4 Thu-M-S&T-1-2 Tue-R-V-1-3 Wed-E-V-5-11 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-1-3 Wed-E-V-5-31 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-O-1-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-V-2-5 Wed-E-O-1-3 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-V-2-3 Wed-E-O-1-3 Wed-E-O-1-4 Fri-M-V-2-10 Wed-M-V-5-8 Fri-M-V-2-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-E-V-1-1 Fri-A-O-22 Tue-M-V-1-2 Tue-M-V-2-1 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-2-13 Fri-M-V-2-10 Wed-A-SS-1-2 Tue-M-V-1-2 Tue-S-1-2 Wed-A-SS-1-3 F	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 2207\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 235\\ 216\\ 233\\ 76\\ 132\\ 226\\ 133\\ 76\\ 132\\ 226\\ 136\\ 176\\ 214\\ 64\\ 313\\ 103\\ 226\\ 157\\ 155\\ 259\\ 139\\ 94\\ 141\\ 181\\ 258\\ \end{array}$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Fri-A-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-2-3 Wed-E-V-5-11 Tue-A-V-2-3 Wed-E-V-5-11 Tue-A-V-2-5 Tue-A-V-2-5 Tue-A-V-2-5 Wed-E-O-1-4 Fri-M-V-6-6 Tue-A-SS-1-3 Wed-E-O-2-3 Wed-H-V-2-11 Fri-M-V-2-5 Tue-A-SS-1-3 Wed-H-V-2-11 Fri-M-V-2-8 Wed-M-V-5-8 Fri-M-V-2-8 Wed-M-V-5-10 Wed-M-V-5-11 Fri-A-O-2-2 Tue-M-V-1-4 Wed-E-SS-1-7 Thu-A-SS-12 Tue-M-V-1-4 Wed-A-SS-1-1 Wed-A-SS-1-2 Tue-M-V-2-10 Wed-A-SS-1-3 Fri-A-SS-2-13 Fri-A-SS-2-13 Fri-A-SS-2-5	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 260\\ 63\\ 103\\ 172\\ 260\\ 63\\ 103\\ 172\\ 216\\ 105\\ 85\\ 80\\ 158\\ 80\\ 158\\ 80\\ 158\\ 80\\ 158\\ 235\\ 95\\ 125\\ 226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 233\\ 76\\ 2226\\ 223\\ 226\\ 222\\ 226\\ 222\\ 226\\ 225\\ 9\\ 94\\ 111\\ 181\\ 258\\ 2248 \end{array}$
Gretter, R	Fri-A-SS-2-14 Fri-M-O-1-2 Tue-M-SS-1-4 Thu-M-S&T-1-3 Thu-M-S&T-1-4 Thu-M-S&T-1-4 Thu-M-S&T-1-2 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-1-3 Wed-E-V-5-11 Tue-E-V-1-3 Wed-E-V-5-31 Wed-E-V-2-3 Wed-A-V-6-4 Tue-A-V-2-5 Tue-A-V-1-5 Wed-E-O-1-5 Wed-E-O-1-3 Wed-E-O-2-3 Wed-E-O-2-3 Wed-H-V-2-10 Fri-M-V-2-10 Wed-M-V-5-8 Fri-M-V-2-8 Wed-M-V-5-8 Fri-M-V-2-10 Wed-M-SS-1-2 Tue-M-V-1-4 Wed-SS-1-1 Wed-E-SS-1-7 Thu-A-SS-1-2 Tue-M-V-1-2 Tue-N-V-2-10 Wed-A-SS-1-2 Tue-M-V-2-10 Wed-A-SS-1-3 Fri-M-V-2-10 Wed-A-SS-1-3 Fri-A-SS-2-9 Wed-A-O-3-1 Tue-A-V-6-68	$\begin{array}{c} 219\\ 219\\ 78\\ 196\\ 197\\ 207\\ 2207\\ 260\\ 63\\ 103\\ 172\\ 106\\ 153\\ 85\\ 85\\ 85\\ 85\\ 235\\ 95\\ 159\\ 95\\ 125\\ 226\\ 233\\ 133\\ 76\\ 132\\ 226\\ 160\\ 214\\ 64\\ 313\\ 226\\ 157\\ 155\\ 259\\ 94\\ 141\\ 181\\ 258\\ 248\\ 248\\ 248\\ 248\\ 248\\ 248\\ 248\\ 24$

Η

Habberstad, Doug	Wed-A-V-1-7	142
Haddadi Hamed	Tue-A-SS-1-2	95
Hach Umbach Boinhold	Thu A V 2 2	205
Haeb-Ollibach, Kellillolu	1 nu-A-v-5-5	203
Hagberg, Nils	Wed-A-S&T-1-4	157
Haghani, Parisa	Wed-E-V-3-6	166
Haidar, Md. Akmal	Wed-A-V-3-13	147
Haider Fasih		216
	111u-A-55-2-1	210
	Fri-A-V-2-1	245
Hain, Thomas	Fri-M-V-3-10	229
Hair Adam	Thu-M-V-1-6	181
Haiai Chan		100
Hajaj, Chen	ти-м-v-4-9	188
Hajek, John	Fri-M-V-1-1	222
Hakkani Tür Dilek	Wed-A-V-2-13	145
Hamann Silko		76
	1 ue-Ivi-v-7-0	105
Hamilakis, Nicolas	Wed-M-V-2-10	125
Han, Ambyera	Fri-M-SS-1-5	240
Han Asriel	Fri-M-V-5-2	231
Hon Ding		156
	wea-A-55-1-6	100
Han, Cong	Thu-M-V-3-3	185
	Thu-A-V-3-2	205
	Thu.A.V.3.6	205
Han David V		205
Hall, David K.	Tue-A-v-I-I	02
Han, Eunjung	Thu-M-V-4-1	187
Han. leong-Im	Fri-M-V-1-1	222
Han Jiangai	Wod M SS 1 4	136
	weu-Ivi-55-1-4	150
Han, Jing	I ue-M-SS-I-I	11
	Thu-M-V-1-9	181
Han. Iiqing	Tue-E-V-1-8	104
····, J-·I-···B	Wed E V 2.4	165
	wea-E-V-3-4	102
	Thu-A-V-1-4	200
Han. Kvu I	Wed-M-V-6-14	135
,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	Thu.A.V.1.11	202
Han Mai		172
	wea-E-V-5-12	172
Han, Sangwook	Wed-M-V-4-7	129
Han. Seungu	Wed-M-V-4-1	128
Han Soveon Caren	Eri-A-V-6-10	255
Han, Soycon curch		110
Hall, 1111g	Tue-E-V-4-9	112
Han, Wei	Wed-M-V-6-10	134
	Thu-A-V-2-8	204
Han Zhuoda	Tue-M-SS-1-9	78
Hangoolz Edwin D		116
Hancock, Euwin R	Tue-E-V-6-9	110
Hanley, Terry	Wed-M-V-2-9	125
Hansen. John H.L.	Tue-E-O-1-4	100
, 5	Wed-M-O-3-5	121
Uantka Simona		102
	Tue-E-O-3-3	102
Hanzlicek, Zdenek	Fri-A-S&T-1-2	260
Hao, Jie	Tue-M-V-3-1	66
Hao Shuai	Thu-Δ-\$\$-2-12	219
Hara Sunao		65
	Tue-IVI-V-2-4	05
Haro, Josep Maria	I ue-A-V-2-6	85
Harrington, Jonathan	Wed-M-O-2-2	120
Harris Jennifer M	Thu-A-\$\$-2-7	218
Harrison Dotor MC		220
	F11-MI-0-2-2	220
Harrison, Philip	Tue-M-V-7-7	76
	Wed-M-V-5-8	132
Harvill, John	Tue-A-SS-2-6	97
Harwath David	Wed M V 2 12	126
	weu-w-v-2-12	120
	1 nu-M-V-2-8	184
Hasegawa-Johnson, M	Tue-A-SS-2-6	97
	Tue-E-V-5-8	113
Hasthanasombat A	Tue-M-SS-1-1	77
Hauley P'mani	Thu M V 1 12	182
Haurcy, Killani		102
Hauptmann, Alexander	I hu-A-O-I-4	199
Haws, David	Wed-M-V-6-9	134
	Wed-A-V-2-11	144
	$Eri_A = \sqrt{-5} = 11$	253
TT 1		200
Hayakawa, Tomoaki	wed-E-V-1-4	101
Hayashi, Tomoki	Tue-A-V-1-13	84
Hdez-Díaz, María E	Tue-F-SS-1-7	118
He Di	Τυο-Δ-\/-4-2	- 80
	Tue-A-V-4-2	00
He, Lei	Iue-A-V-4-10	90
	Fri-A-V-5-8	253
He, Liang	Fri-A-V-6-7	255
He. Maokui	Thu-M-V-4-6	188
, Piuokui		207
	Thu-A-V-4-3	207
не, Qing	Tue-M-V-2-8	66
He, Shulin	We <u>d-E-SS-1-6</u>	176
He Weineng	Тие-А-О-1-4	80
Uo Vianghong		00
	Tue-A-V-6-4	33
Не, Хіаохі	Thu-M-V-7-2	192
He, Yanzhang	Wed-M-V-6-4	133
,	Fri-M-V-2-0	220
		224
	Fri-M-V-6-3	234
		o · · ·
	Fri-A-V-3-8	248
Не, Үі	Fri-A-V-3-8 Wed-E-V-3- <u>9</u>	$\begin{array}{c} 248 \\ 166 \end{array}$
He, Yi He, Yuxuan	Fri-A-V-3-8 Wed-E-V-3-9 Fri <u>-M-V-6-11</u>	248 166 236

Heelrow Decoel		
Hecker, Pascal	Tue-E-O-3-3	102
Hoorings Wilbort		227
ilectiliga, wilbert	F11-IVI-V-2-13	221
Hegde, Sindhu B	Thu-A-V-6-11	213
Heikinheimo Hannes	Wed-A-O-1-5	138
Hoinó Míča		160
11ejiia, misa	weu-E-V-4-2	100
	Fri-M-V-2-3	225
Helmke, Hartmut	Thu-M-SS-2-3	195
Hombico Corontin	Eri A COT 1 4	261
Hembise, Corentin	FII-A-3@1-1-4	201
Henderson, Alice	Wed-E-V-4-3	168
Henriques Ioão F	Wed-E-0-3-5	160
11cm1ques, joao 1	Weu-E-O-3-3	100
нео, нее-500	Thu-M-V-4-2	187
	Thu-M-V-4-5	187
		212
	1 nu-A-v-0-0	212
Hermann, Enno	Fri-A-SS-2-3	257
Hermansky, Hynek	Wed-M-O-1-2	119
Horeboy John D		105
Hersney, John R.	I NU-IVI-V-3-3	102
Heymann, Jahn	Wed-M-V-4-8	129
Heywood Rus	Eri.A.SS.2.1	257
11cy wood, 1kus	111/4 35 2 1	257
	Fri-A-SS-2-12	259
Hifny, Yasser	Thu-A-V-2-11	204
Higashinaka Dynichiro	Tuo-M-V-4-4	60
ingasimiaka, Kyuicimo .	100-101-0-4-4	03
Higuchi, Yosuke	Tue-A-V-4-5	89
Hildén, Raili	Wed-M-O-1-3	119
Hioko Vuouko		174
	wed-E-V-6-5	1/4
Hıramura, Takekatsu	Wed-A-V-1-3	141
Hirova, Sadao	Wed-E-V-2-3	163
Hirschharg Julio		102
Thischberg, Julia	Tue-E-O-3-4	105
Hoang, Vu	Fri-A-S&T-1-5	261
Hoi, Steven	Wed-E-V-1-8	162
Hong Mi		102
11011g, MI	TUE-E-V-T-2	103
Hong, Qingyang	Tue-M-V-1-8	64
5. 5. 5. 5	Thu-M-SS-1-1	193
	1110-101-33-1-1	104
	Thu-M-SS-1-6	194
	Thu-A-SS-1-8	215
		251
	FII-A-V-4-9	251
	Fri-A-V-4-10	251
Hong, YW. Peter	Tue-E-V-3-6	108
Hong Thonhou		200
Hong, Zheimou	1 riu-A-v-5-1	209
Hori, Chiori	Tue-A-V-1-10	83
	Wed-A-V-3-12	147
Hori Takaaki		83
11011, 1 akaaki	Tue-A-V-1-10	05
	I ue-A-V-4-5	89
	Wed-M-V-6-13	135
	Wed-A-V-3-12	147
Hanimu alai Chasta		197
	I III-IVI-V-4-4	107
Horiguchi, Shota	111011111	
Horiguchi, Shota	Thu-M-V-4-8	188
Horigucii, Snota	Thu-M-V-4-8 Tue-A-V-2-6	188 85
Horiguchi, Shota	Thu-M-V-4-8 Tue-A-V-2-6	188 85
Hotopf, Matthew Hou, Feng	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11	188 85 68
Hotopf, Matthew Hou, Feng Hou, Junfeng	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8	188 85 68 134
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6	188 85 68 134 254
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6	188 85 68 134 254
Horguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1	188 85 68 134 254 202
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1	188 85 68 134 254 202 73
Horguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Iulian	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6	188 85 68 134 254 202 73 142
Horguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6	188 85 68 134 254 202 73 142
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9	188 85 68 134 254 202 73 142 218
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6	188 85 68 134 254 202 73 142 218 169
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6	188 85 68 134 254 202 73 142 218 169 225
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-2-6	188 85 68 134 254 202 73 142 218 169 225 68
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8	188 85 68 134 254 202 73 142 218 169 225 68
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Tri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9	$188 \\ 85 \\ 68 \\ 134 \\ 254 \\ 202 \\ 73 \\ 142 \\ 218 \\ 169 \\ 225 \\ 68 \\ 68 \\ 68 \\ 68 \\ 68 \\ 68 \\ 68 \\ 6$
Horguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4	$\begin{array}{r} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6	$188 \\ 85 \\ 68 \\ 134 \\ 254 \\ 202 \\ 73 \\ 142 \\ 218 \\ 169 \\ 225 \\ 68 \\ 68 \\ 89 \\ 210 \\$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-A-V-1-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-A-V-4-4 Thu-A-V-5-6	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-6	188 85 68 134 254 202 73 142 218 169 225 68 68 89 210 210
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 114 \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Ke	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Na	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Eri-A-V-3-4	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Ke Hu, Na	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-4	188 85 68 134 254 202 73 142 218 169 225 68 68 89 210 210 210 114 145 252
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Ke Hu, Na Hu, Peng	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-4 Wed-A-SS-1-6	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 156\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Yuanbo Hou, Yuanbo Hough, Julian Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Na Hu, Peng Hu, Pengfei	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-A-V-3-4 Fri-A-V-5-9 Wed-A-V-3-4 Fri-A-V-5-4 Wed-A-SS-1-6 Tue-M-V-5-9	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 218\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 68\\ 68\\ 72\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Ke Hu, Na Hu, Peng Hu, Pengfei	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-SS-1-6	188 85 68 134 2254 202 73 142 218 169 225 68 68 89 210 210 114 145 252 156 72 240
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hui, Feng-fan Hu, Cheng-Hung Hu, Chenxu Hu, Na Hu, Peng Hu, Peng Hu, Peng	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 899\\ 210\\ 210\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 156\\ 72\\ 240\\ 220\\ 102\\ 240\\ 220\\ 240\\ 220\\ 220\\ 220\\ 220\\ 2$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Chenxu Hu, Na Hu, Peng Hu, Pengfei Hu, Qian	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V2-1 Tue-M-V-2-1 Tue-M-V-2-1 Wed-A-V-1-6 Thu-A-V2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-4 Wed-A-SS-1-6 Tue-M-V-5-9 Fri-M-SS-1-5 Fri-M-V-6-2	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 156\\ 72\\ 240\\ 234\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Ke Hu, Na Hu, Peng Hu, Pengfei Hu, Qian Hu, Shell Xu	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-SS-1-5 Fri-M-V-6-2 Thu-A-V-3-7	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 2206 \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hough, Julian Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Na Hu, Peng Hu, Peng fei Hu, Shell Xu Hu Shih-Wei	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Wed-A-V-3-4 Wed-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hough, Julian Hough, Julian Husieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Na Hu, Peng Hu, Pengfei Hu, Shell Xu Hu, Shih-Wei Hu Shih-Wei	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-S2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-4 Wed-A-SS-1-6 Tue-M-V-5-9 Fri-M-SS-1-5 Fri-M-V-6-2 Thu-A-V-3-7 Wed-M-V-4-2 Cn-4-6-2	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 273\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 206\\ 128\\ 240\\ 234\\ 206\\ 237\\ 27\\ 240\\ 234\\ 206\\ 237\\ 27\\ 240\\ 234\\ 206\\ 237\\ 27\\ 27\\ 240\\ 234\\ 206\\ 27\\ 27\\ 27\\ 27\\ 240\\ 234\\ 206\\ 27\\ 27\\ 27\\ 27\\ 240\\ 234\\ 206\\ 27\\ 27\\ 27\\ 27\\ 240\\ 234\\ 206\\ 27\\ 27\\ 27\\ 27\\ 240\\ 234\\ 206\\ 27\\ 27\\ 27\\ 27\\ 27\\ 27\\ 27\\ 27\\ 27\\ 27$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Ke Hu, Na Hu, Peng Hu, Pengfei Hu, Shell Xu Hu, Shoukang	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-3-9 Tue-M-V-3-9 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Wed-A-V-3-4 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-V-6-2 Thu-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-4	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 156\\ 72\\ 240\\ 234\\ 226\\ 128\\ 257\\ \end{array}$
Horiguchi, ShotaHotopf, MatthewHou, FengHou, JunfengHou, WangruiHou, WangruiHou, WangruiHou, WangruiHou, WangruiHou, YuanboHough, JulianHough, JulianHsieh, Feng-fanHsieh, Tsun-AnHsu, Wei-NingHu, Cheng-HungHu, ChensuHu, ReHu, PengHu, PengfeiHu, Shell XuHu, Shoukang	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-4 Wed-A-SS-1-5 Fri-M-V-5-9 Fri-M-SS-1-5 Fri-M-V-6-2 Thu-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-9	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 206\\ 128\\ 257\\ 259\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Chensu Hu, Re Hu, Peng Hu, Pengfei Hu, Shell Xu Hu, Shoukang	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-SS-1-6 Tue-M-V-5-9 Fri-M-SS-1-6 Tue-M-V-5-9 Fri-M-SS-1-6 Tue-M-V-5-9 Fri-M-SS-1-6 Tue-M-V-5-9 Fri-M-SS-2-4 Fri-A-SS-2-9 Fri-A-SS-2-9 Fri-A-SS-2-9 Fri-A-SS-2-9	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 206\\ 128\\ 257\\ 229\\ 221\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Re Hu, Peng Hu, Peng Hu, Pengfei Hu, Shell Xu Hu, Shoukang	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-3-8 Tue-M-V-3-9 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-2 Thu-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-9 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-O-2-3	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 156\\ 72\\ 240\\ 234\\ 256\\ 221\\ 128\\ 257\\ 259\\ 221\\ 221\\ 221\\ 221\\ 222\\ 221\\ 222\\ 221\\ 222$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hough, Julian Hu, Yuanbo Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Peng fei Hu, Pengfei Hu, Shell Xu Hu, Shoukang Hu, Vivian	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Wed-A-V-3-4 Wed-A-V-3-4 Wed-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-4 Fri-A-SS-2-4 Fri-A-SS-2-4 Fri-A-SS-2-4 Fri-A-SS-2-4 Fri-A-SS-2-4	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 232\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 210\\ 234\\ 206\\ 68\\ 89\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 252\\ 211\\ 221\\ 221\\ 221\\ 221\\ 22$
Horiguchi, ShotaHotopf, MatthewHou, FengHou, JunfengHou, WangruiHou, WangruiHou, WenxinHou, WenxinHou, YuanboHou, YuanboHough, JulianHsieh, Feng-fanHsieh, Feng-fanHsieh, Tsun-AnHsu, Wei-NingHu, Cheng-HungHu, Cheng-HungHu, ChenxuHu, ReHu, PengHu, PengfeiHu, Shell XuHu, ShoukangHu, VivianHu, Wenchao	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Wed-A-V-3-4 Wed-A-V-3-4 Wed-A-V-3-7 Wed-M-V-5-9 Fri-M-SS-1-6 Tue-M-V-5-9 Fri-M-SS-1-6 Tue-M-V-5-9 Fri-M-SS-1-6 Tue-M-V-5-9 Fri-M-SS-2-4 Fri-A-SS-2-9 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-O-2-3	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hough, Julian Hough, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Chenxu Hu, Chenxu Hu, Peng Hu, Peng fei Hu, Shell Xu Hu, Shoukang Hu, Vivian Hu, Wenchao Hu, Xinhui	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Wed-A-V-3-4 Wed-A-V-3-1 Wed-A-V-3-1 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-A-SS-2-9 Fri-A-SS-2-9 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-V-6-9 Thu-M-SS-1-4	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 210\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 210\\ 221\\ 257\\ 221\\ 225\\ 221\\ 225\\ 221\\ 235\\ 194 \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hough, Julian Hu, Yeng-fan Hsieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Peng fei Hu, Peng fei Hu, Shell Xu Hu, Shoukang Hu, Vivian Hu, Wenchao Hu, Xinhui	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-V-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-4 Wed-A-V-3-7 Wed-A-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-9 Fri-M-0-2-3 Fri-M-0-2-3 Fri-M-0-2-4 Fri-M-0-2-4 Fri-M-0-2-4 Fri-M-V-6-9 Thu-M-S-1-4 Wood-ESC-1-4	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 212\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 206\\ 6128\\ 257\\ 2240\\ 234\\ 206\\ 128\\ 257\\ 221\\ 221\\ 221\\ 235\\ 194\\ 176\\ 68\\ 128\\ 257\\ 221\\ 211\\ 235\\ 194\\ 176\\ 176\\ 176\\ 176\\ 176\\ 176\\ 176\\ 176$
Horiguchi, ShotaHotopf, MatthewHou, FengHou, JunfengHou, WangruiHou, WangruiHou, WenxinHou, WenxinHou, YuanboHou, YuanboHough, JulianHsieh, Feng-fanHsieh, Feng-fanHsieh, Tsun-AnHsu, Wei-NingHu, Cheng-HungHu, Cheng-HungHu, ReHu, NaHu, PengHu, Shell XuHu, ShoukangHu, VivianHu, WenchaoHu, XinhuiHu, Yanxin	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-3-9 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-2 Thu-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-4 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2 Fri-M-V-6-2	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 211\\ 415\\ 252\\ 156\\ 72\\ 240\\ 234\\ 257\\ 224\\ 128\\ 257\\ 225\\ 221\\ 225\\ 194\\ 176\\ 68\\ 257\\ 221\\ 221\\ 221\\ 221\\ 221\\ 221\\ 221\\ 22$
Horiguchi, ShotaHotopf, MatthewHou, FengHou, JunfengHou, WangruiHou, YuanboHsieh, Feng-fanHsieh, Feng-fanHsieh, Tsun-AnHsu, Wei-NingHu, Cheng-HungHu, ChenxuHu, ChenxuHu, Reng-HungHu, PengHu, PengfeiHu, Shell XuHu, ShoukangHu, ShoukangHu, VivianHu, WenchaoHu, XinhuiHu, Yanxin	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-A-V-3-9 Tue-A-V-3-4 Wed-A-V-3-4 Wed-A-V-3-7 Wed-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-6-2 Fri-M-V-6-9 Thu-M-V-6-9 Thu-A-V-6-4	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 254\\ 202\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Chensu Hu, Peng Hu, Peng fei Hu, Pengfei Hu, Shih-Wei Hu, Shoukang Hu, Vivian Hu, Wenchao Hu, Xinhui Hu, Yanxin	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-V2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-4 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-9 Fri-M-O2-4 Fri-A-SS-2-9 Thu-M-V-6-9 Thu-M-V-6-9 Thu-M-V-6-9 Thu-M-V-6-9 Thu-M-V-6-4 Fri-A-V-6-4 Fri-A-V-6-4	$\begin{array}{c} 188\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 240\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 206\\ 128\\ 257\\ 221\\ 234\\ 206\\ 128\\ 257\\ 221\\ 221\\ 235\\ 194\\ 176\\ 212\\ 2256 \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hough, Julian Hough, Julian Husieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Sholkang Hu, Shoukang Hu, Vivian Hu, Yanxin Hu, Yanxin	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-A-SS-2-4 Fri-A-SS-2-9 Fri-M-V-6-9 Thu-M-SS-1-3 Fri-M-V-6-4 Fri-A-VS-1-3 Thu-A-V-6-4 Fri-A-V-6-4 Fri-A-V-6-4 Fri-A-V-6-4	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 254\\ 202\\ 273\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 211\\ 415\\ 252\\ 156\\ 72\\ 240\\ 234\\ 257\\ 259\\ 2211\\ 2235\\ 128\\ 257\\ 259\\ 2211\\ 221\\ 235\\ 194\\ 176\\ 212\\ 256\\ 212\\ 185\\ 185\\ 185\\ 185\\ 185\\ 185\\ 185\\ 185$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Husieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Chenxu Hu, Chenxu Hu, Peng Hu, Peng fei Hu, Shell Xu Hu, Shoukang Hu, Vivian Hu, Wenchao Hu, Yanxin Hu, Ying	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Wed-A-SS-1-6 Tue-M-V-5-9 Fri-M-SS-1-5 Fri-M-V-6-9 Thu-A-V-6-9 Thu-A-V-6-9 Thu-A-V-6-9 Thu-A-V-6-9 Thu-M-V-6-9 Thu-M-V-6-9 Thu-M-V-6-9 Thu-M-V-6-9 Thu-M-V-6-9 Thu-A-V-6-1 Wed-E-SS-1-5	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 206\\ 128\\ 2257\\ 221\\ 221\\ 225\\ 194\\ 176\\ 212\\ 256\\ 185\\ 256\\ 185\\ 256\\ 212\\ 2256\\ 212\\ 2256\\ 212\\ 2256\\ 225\\ 225\\ 221\\ 2256\\ 225\\ 225\\ 221\\ 2256\\ 225\\ 225\\ 225\\ 225\\ 225\\ 225\\ 22$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Chensu Hu, Peng Hu, Pengfei Hu, Shell Xu Hu, Shoukang Hu, Vivian Hu, Wenchao Hu, Xinhui Hu, Yanxin	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-S2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-4 Wed-A-V-3-4 Fri-A-V5-16 Tue-M-V-5-9 Fri-M-V5-15 Fri-M-V-6-2 Thu-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-1-5 Fri-M-O2-4 Fri-A-SS-1-3 Fri-M-O2-4 Fri-A-SS-1-5 Thu-M-V-6-9 Thu-M-V-6-5 Thu-M-V-3-5	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wenxin Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hou, Yuanbo Hough, Julian Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Chenxu Hu, Peng Hu, Peng fei Hu, Pengfei Hu, Shell Xu Hu, Shoukang Hu, Shoukang Hu, Vivian Hu, Vinyan Hu, Yanxin Hu, Ying Hu, Ying Hu, Yu	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-2-1 Tue-M-V-2-1 Wed-A-V-1-6 Thu-A-V-2-6 Tue-M-V-3-8 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Wed-A-V-3-4 Wed-A-V-3-1 Wed-A-V-3-1 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-A-SS-1-5 Fri-M-V-6-2 Thu-M-V-3-5 Thu-M-V-6-3 Thu-M-V-3-5 Tue-A-V-6-3 Thu-M-V-2-3	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 210\\ 234\\ 257\\ 221\\ 225\\ 156\\ 72\\ 240\\ 234\\ 257\\ 259\\ 221\\ 225\\ 128\\ 257\\ 225\\ 128\\ 128\\ 257\\ 259\\ 221\\ 176\\ 212\\ 256\\ 194\\ 176\\ 212\\ 256\\ 193\\ 183\\ 3182\\ 3$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hu, Yeng-fan Hu, Cheng-Hung Hu, Cheng-Hung Hu, Chenxu Hu, Chenxu Hu, Chenxu Hu, Peng Hu, Peng fei Hu, Shell Xu Hu, Shoukang Hu, Vivian Hu, Wenchao Hu, Yanxin Hu, Ying Hu, Yu Huang Che-Wei	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-4 Fri-A-SS-2-4 Fri-A-SS-2-4 Fri-A-SS-2-9 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-O-2-3 Fri-M-V-6-9 Thu-A-V-6-9 Thu-A-V-6-3 Thu-A-V-6-3 Thu-M-V-3-5	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 206\\ 68\\ 89\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 257\\ 221\\ 221\\ 235\\ 194\\ 176\\ 222\\ 256\\ 185\\ 93\\ 183\\ 80\\ 89\\ 210\\ 201\\ 201\\ 201\\ 201\\ 201\\ 201\\ 201$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Wenxin Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hou, Yuanbo Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Chensu Hu, Re Hu, Peng Hu, Pengfei Hu, Shell Xu Hu, Shoukang Hu, Vivian Hu, Wenchao Hu, Xinhui Hu, Yanxin Hu, Ying Hu, Ying Huang, Che-Wei	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-3-9 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-A-SS-1-5 Fri-M-V-6-2 Fri-M-V-6-3 Fri-M-V-6-3 Thu-A-V-6-3 Thu-A-V-6-3 Thu-M-V-2-7 Tue-A-V-6-3 Thu-M-V-2-7 Tue-A-V-6-3	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 211\\ 45\\ 252\\ 156\\ 72\\ 240\\ 234\\ 257\\ 259\\ 221\\ 2256\\ 128\\ 257\\ 259\\ 221\\ 2256\\ 194\\ 176\\ 212\\ 256\\ 193\\ 183\\ 89\\ 89\\ 3183\\ 89\\ 89\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210$
Horiguchi, ShotaHotopf, MatthewHou, FengHou, JunfengHou, WangruiHou, WenxinHou, WangruiHou, WangruiHou, WangruiHou, WangruiHou, WangruiHou, WangruiHou, WangruiHou, WangruiHou, WenxinHou, YuanboHusieh, Feng-fanHsieh, Tsun-AnHsu, Wei-NingHu, Cheng-HungHu, ChenxuHu, ChenxuHu, ReHu, PengHu, Peng feiHu, Shell XuHu, ShoukangHu, ShoukangHu, VivianHu, WenchaoHu, YingHu, YingHuang, Che-WeiHuang, Chien-yu	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-9 Fri-M-O-2-3 Thu-M-V-6-3 Thu-M-V-6-3 Thu-M-V-2-7 Tue-A-V-4-2 Wed-E-V-6-9	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 8\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Yuanbo Hisieh, Feng-fan Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Chensu Hu, Peng Hu, Peng fei Hu, Peng fei Hu, Shell Xu Hu, Shoukang Hu, Shoukang Hu, Vivian Hu, Ying Hu, Ying Hu, Ying Huang, Che-Wei Huang, Chien-yu Huang, Chonexuan	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-V-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-6 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-A-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Wed-M-V-3-7 Fri-M-V6-2 Fri-M-V6-2 Fri-M-V6-9 Fri-M-0-2-3 Fri-M-0-2-4 Fri-A-V-6-9 Thu-M-V-3-5 Tue-A-V-6-3 Thu-A-V-6-3 Thu-A-V-6-4 Thu-A-V-6-3 Thu-A-V-6-3 Thu-A-V-6-3 Thu-A-V-6-3 Thu-A-V-6-3 Thu-A-V-	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 224\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 240\\ 234\\ 257\\ 2240\\ 234\\ 206\\ 68\\ 89\\ 210\\ 212\\ 234\\ 257\\ 221\\ 221\\ 235\\ 194\\ 176\\ 183\\ 89\\ 93\\ 183\\ 89\\ 91\\ 112\\ 256\\ 183\\ 89\\ 91\\ 112\\ 256\\ 183\\ 183\\ 183\\ 183\\ 112\\ 112\\ 112\\ 112\\ 112\\ 112\\ 112\\ 11$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hough, Julian Hough, Julian Husieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Chenxu Hu, Chenxu Hu, Chenxu Hu, Chenxu Hu, Peng Hu, Peng Hu, Shell Xu Hu, Shoukang Hu, Shoukang Hu, Vivian Hu, Ying Hu, Ying Huang, Che-Wei Huang, Chongxuan Huang, Chongxuan	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-3-8 Tue-M-V-3-9 Tue-M-V-3-9 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-E-V-6-1 Wed-A-V-3-4 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-6-2 Thu-M-V-3-7 Wed-M-V-4-2 Fri-A-SS-1-3 Thu-M-V-2-7 Tue-A-V-6-3 Thu-M-V-2-7 Tue-A-V-6-9 Thu-M-V-2-7 Tue-A-V-6-9 Thu-M-V-2-7 Tue-A-V-6-9 Thu-M-V-2-7	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 211\\ 45\\ 252\\ 156\\ 72\\ 240\\ 234\\ 256\\ 221\\ 225\\ 259\\ 221\\ 225\\ 235\\ 122\\ 122\\ 122\\ 156\\ 72\\ 259\\ 221\\ 122\\ 123\\ 58\\ 93\\ 174\\ 112\\ 256\\ 272\\ 72\\ 72\\ 72\\ 72\\ 72\\ 72\\ 72\\ 72\\ 7$
Horiguchi, ShotaHotopf, MatthewHou, FengHou, JunfengHou, WangruiHou, YuanboHsieh, Feng-fanHsieh, Tsun-AnHsu, Wei-NingHu, Cheng-HungHu, ChensuHu, ChenxuHu, ReHu, PengHu, Peng feiHu, Shell XuHu, ShoukangHu, ShoukangHu, VivianHu, WenchaoHu, YingHuang, Che-WeiHuang, Chien-yuHuang, ChongxuanHuang, Hao	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-2-1 Tue-M-V-2-1 Tue-M-V-2-1 Wed-A-V-1-6 Thu-A-V-2-9 Wed-E-V-4-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-A-V-3-4 Wed-A-SS-1-6 Tue-M-V-5-9 Fri-M-SC-1-5 Fri-M-SC-1-5 Fri-M-SC-1-5 Fri-M-V-6-2 Thu-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-9 Fri-M-O-2-3 Fri-M-O-2-4 Fri-A-SS-1-3 Thu-M-V-3-5 Tue-A-V-6-3 Thu-M-V-2-7 Tue-A-V-4-2 Wed-E-V-6-9 Tue-E-V-4-9 Tue-E-V-4-9 Tue-B-V-5-9	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210\\ 210$
Horiguchi, ShotaHotopf, MatthewHou, FengHou, JunfengHou, WangruiHou, YuanboHough, JulianHsieh, Feng-fanHsieh, Feng-fanHsieh, Tsun-AnHsu, Wei-NingHu, ChenzuHu, ChenzuHu, ChenzuHu, ReHu, NaHu, PengHu, PengfeiHu, Shell XuHu, ShoukangHu, ShoukangHu, VivianHu, YingHu, YanxinHu, YingHuang, Che-WeiHuang, ChongxuanHuang, Hao	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-2-1 Tue-M-V-2-1 Wed-A-V-1-6 Fri-M-V-2-6 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Fri-A-V-5-9 Tue-A-V-3-4 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-2-7 Thu-A-V-3-1 Fri-M-V-2-7 Thu-A-V-6-3 Thu-A-V-6-3 Thu-M-V-2-7 Tue-A-V-4-2 Wed-E-V-6-9 Thu-M-V-2-7 Tue-A-V-4-9 Tue-K-V-9 Tue-K-V-9 Tue-K-V-9 Wed-M-V-1-7	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 254\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 156\\ 68\\ 89\\ 210\\ 210\\ 114\\ 145\\ 252\\ 156\\ 72\\ 2240\\ 128\\ 257\\ 2240\\ 128\\ 257\\ 224\\ 128\\ 188\\ 89\\ 221\\ 176\\ 212\\ 256\\ 188\\ 89\\ 93\\ 174\\ 112\\ 272\\ 123\\ \end{array}$
Horiguchi, Shota Hotopf, Matthew Hou, Feng Hou, Junfeng Hou, Wangrui Hou, Wangrui Hou, Wenxin Hou, Wenxin Hou, Yuanbo Hou, Yuanbo Houdy, Julian Hough, Julian Hsieh, Feng-fan Hsieh, Tsun-An Hsu, Wei-Ning Hu, Cheng-Hung Hu, Chensu Hu, Chensu Hu, Re Hu, Peng Hu, Pengfei Hu, Shell Xu Hu, Shoukang Hu, Shoukang Hu, Vivian Hu, Yanxin Hu, Yanxin Hu, Ying Hu, Ying Huang, Che-Wei Huang, Chongxuan Huang, Hao	Thu-M-V-4-8 Tue-A-V-2-6 Tue-M-V-3-11 Wed-M-V-6-8 Fri-A-V-6-6 Thu-A-V-2-1 Tue-M-V-6-1 Wed-A-V-1-6 Thu-A-SS-2-9 Wed-E-V-4-6 Fri-M-V-2-6 Tue-M-V-3-8 Tue-M-V-3-9 Tue-A-V-4-4 Thu-A-V-5-9 Tue-A-V-4-4 Wed-A-V-5-9 Tue-E-V-6-1 Wed-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-5-9 Fri-M-V-6-2 Thu-A-V-3-7 Wed-M-V-4-2 Fri-A-SS-2-9 Fri-M-0-2-3 Fri-M-0-2-3 Fri-M-0-2-3 Fri-M-0-2-3 Fri-M-0-2-4 Fri-A-SS-1-3 Thu-M-V-5-7 Tue-A-V-6-3 Thu-M-V-2-7 Tue-A-V-4-2 Wed-E-V-6-9 Tue-M-V-5-9 Wed-M-V-1-7 Thu-M-V-1-7 Thu-M-V-1-7	$\begin{array}{c} 188\\ 85\\ 68\\ 85\\ 68\\ 134\\ 202\\ 202\\ 73\\ 142\\ 218\\ 169\\ 225\\ 68\\ 68\\ 89\\ 210\\ 210\\ 210\\ 114\\ 145\\ 252\\ 156\\ 72\\ 240\\ 234\\ 257\\ 221\\ 126\\ 234\\ 257\\ 221\\ 221\\ 235\\ 128\\ 89\\ 174\\ 112\\ 72\\ 2123\\ 185\\ \end{array}$

Huang, Jing Huang, Jinwen Huang, Jun		~~-
Huang, Jing Huang, Jinwen Huang, Jun	Eri.M.V.2.6	225
Huang, Jinwen Huang, Jun	111101-0-2-0	223
Huang, Jun	Thu-M-SS-1-3	194
maang, jun	Fri-A-V-3-2	247
Unong Inneri		220
Huang, Junrui	Fri-IVI-V-7-7	230
Huang, Leichao	Fri-M-V-7-7	238
Huang Liang	Wod-A-V-6-1	152
fluang, Liang	Weu A V O T	132
	Fri-A-V-2-7	240
Huang. Lu	Wed-M-V-6-8	134
Lineng Michelle		204
Hualig, Michelle	I NU-A-V-2-9	204
Huang, Qiushi	Wed-A-V-2-8	144
Huang Rongije	Wed-E-V-6-6	174
Indiana, Rongjie		170
Huang, Ruizne	Тпи-м-О-3-6	179
Huang, Shen	Tue-M-V-5-9	72
0.	Eri-M-SS-1-5	240
Iliana Coma Fora		100
Huang, Sung-Feng	<u>I nu-M-V-3-7</u>	100
Huang, Tzu-Hsien	Tue-E-V-3-9	109
Huang Weilong	Tuo-A-\/-2-10	88
indding, wellong	146 4 9 10	10-
	Wed-M-SS-1-7	137
	Thu-A-V-4-2	207
Huang Wan Chin		114
mualig, well-chill	Tue-E-V-0-2	114
	Thu-A-V-5-9	210
Huang W Ronny	Wed-A-V-2-6	143
Iliana Vina		74
Huang, Aing	I ue-M-V-6-9	74
Huang, Yan	Tue-E-V-5-9	114
Huang Vinghui	Tuo-A-V-4-7	00
fluang, finghui	1ue-A-V-4-7	100
	Wed-E-V-3-7	100
Huang, Yiqing	Tue-A-SS-2-2	96
Huang Viteng	Eri M V G 2	224
	FTI-IVI-V-0-3	204
Huang, Yu-Han	Wed-A-V-3-9	146
Huang, Yukai	Thu-A-V-6-13	214
Uuong Vu Lin		100
ilualig, IU-LIII	тие-Е-V-3-6	100
Huang, Zhihua	Wed-M-V-3-7	127
-	Thu-M-V-3-5	185
		100
_	Fri-M-V-4-9	231
Huang, Zifang	Fri-A-SS-2-15	260
Huang 7ili	Tue-E-V-3-9	109
Tiuding, Zin		105
	Thu-A-V-4-3	207
Huckvale, Mark	Tue-E-SS-1-4	117
Hueber Thomas		60
flueber, fliolitas	Tue-M-0-1-5	00
	Tue-M-O-3-1	62
	Thu-A-O-1-5	199
	Eri_M_O_2_1	220
TT . 1	FII-M-O-2-1	220
Hugnes, vincent	I ue-M-V-7-2	75
	Tue-M-V-7-7	76
Hui CT Justine	Wod-E-V-6-5	174
	Weu-L-V-0-3	1/1
Huo, Zhouyuan	I ue-A-V-4-6	89
Hussain. Sarmad	Tue-M-V-7-5	75
Hussein Amir	Wod-E-V-1-11	162
nusseni, Anni	Weu-L-V-1-11	102
	Thu-A-V-2-11	204
Hustad. Katherine C	Wed-A-O-3-4	140
	Tuo E 0.2.2	110
Hutin Mathilde		101
Hutin, Mathilde		101
Hutin, Mathilde Hwang, Dongseong	Tue-A-V-4-6	101 89
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol	Tue-A-V-4-6 Wed-A-V-5-1	101 89 150
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol	Tue-A-V-4-6 Wed-A-V-5-1	101 89 150
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11	101 89 150 152
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2	101 89 150 152 209
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7	101 89 150 152 209 129
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7	101 89 150 152 209 129
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1	101 89 150 152 209 129 165
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-1 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1	101 89 150 152 209 129 165
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1	101 89 150 152 209 129 165
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1	140 101 89 150 152 209 129 165
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu	Tue-A-V-4-6 Wed-A-V-5-11 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1	140 101 89 150 152 209 129 165 73
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1	101 89 150 152 209 129 165 73 131
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori Mana	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-4	101 89 150 152 209 129 165 73 131
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Songoong Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-11 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6	101 89 150 152 209 129 165 73 131 74
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5	140 101 89 150 152 209 129 165 73 131 74 91
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-11 Wed-A-V-5-2 Wed-M-V-5-2 Wed-M-V-6-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-A-V-5-5 Wed-E-V-3-13	140 101 89 150 152 209 129 165 73 131 74 91 167
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6	140 101 89 150 152 209 129 165 73 131 74 91 167 228
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-1 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6	140 101 89 150 152 209 129 165 73 131 74 91 167 228
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-11 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-6-2 Wed-E-V-3-11 Tue-M-V-6-4 Wed-M-V-5-5 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11	140 101 89 150 152 209 129 165 73 131 74 91 167 228 245
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-1	140 101 89 150 152 209 129 165 73 131 74 91 167 228 245 190
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Girag Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-4	140 101 89 150 152 209 129 165 73 131 74 91 167 228 245 190
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-13 Fri-M-V-3-13 Fri-M-V-3-11 Thu-M-V-6-1 Thu-M-V-6-4	1101 101 189 150 152 209 129 165 73 131 74 91 167 228 2450 191
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Sung Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana Ijima, Yusuke	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-M-V-3-6 Tri-A-V-1-11 Thu-M-V-6-4 Tue-M-V-2-4	1101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-4 Tue-M-V-2-3	101 89 150 152 209 165 165 73 131 74 91 167 228 245 190 191 65 163
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-2-3	 1101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65 163 189
Hutin, Mathilde Hwang, Dongseong Hwang, Ceumbyeol Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Ijima, Yusuke	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-6-4 Wed-E-V-3-13 Fue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-2-3	1101 89 150 152 209 129 165 131 74 91 167 228 245 190 191 65 163 189 227
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-A-V-5-11 Thu-A-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Wed-E-V-3-13 Fri-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-3 Thu-M-V-5-4 Fri-M-V-5-4 Fri-M-V-3-2	1101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65 163 189 227
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana Ijima, Yusuke Ilin, Aleksei Illa, Aravind	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-2 Tue-E-V-3-2	1101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65 163 189 227 107
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Ijima, Yusuke Ilin, Aleksei Illa, Aravind	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Wed-E-V-3-1 Fri-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-1 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-M-V-3-2 Tue-E-V-3-1 Fri-A-S&T-1.8	1101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65 163 189 927 107 261
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Ilina Irina	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-A-V-5-11 Thu-A-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Wed-E-V-3-13 Fri-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-M-V-5-4 Fri-A-S&T-1-8 Wood M-V-1-8	1101 89 150 152 209 165 165 165 165 165 167 228 245 190 191 65 163 189 227 107 261
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-M-V-3-2 Tue-E-V-3-1 Fri-A-S&T-1-8 Wed-M-V-1-2	1101 89 150 152 209 129 129 165 73 131 74 91 167 228 245 190 191 65 163 189 227 107 261 1222
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Wed-M-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-1 Thu-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-S&T-1-8 Wed-M-V-1-2 Wed-A-O-1-1	1101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65 163 189 227 107 261 122 138
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-2 Tue-E-V-3-1 Fri-A-S&T-1-8 Wed-M-V-1-2 Wed-A-O-1-1 Thu-M-SS-1-2	1101 89 150 152 209 129 129 165 73 131 74 91 167 228 245 190 191 653 189 227 107 261 122 2138
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium Stoffen	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-1-11 Fri-A-S&T-1-8 Wed-M-V-1-2 Wed-A-O-1-1 Thu-M-SE-1-2 Tue-SE-1-2	1101 89 150 152 209 129 165 73 131 74 91 165 738 245 190 191 655 163 189 227 107 261 122 138 194 78
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Min-Jae Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Wed-E-V-3-13 Fri-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-6-1 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-S&T-1-8 Wed-M-V-1-2 Wed-A-O-1-1 Thu-M-V5S-1-2 Tue-M-SS-1-2	1101 89 150 152 209 129 129 165 73 131 74 91 165 73 131 74 91 165 131 74 91 165 163 189 227 163 189 227 163 189 227 107 261 138 194 78
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-M-V-3-2 Tue-E-V-3-1 Fri-A-S&T-1-8 Wed-M-V-1-2 Wed-A-O-1-1 Thu-M-SS-1-2 Tue-M-SS-1-7	101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65 163 189 227 107 261 122 2138 189 227 707 261 122 8 94 78 78
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-5-2 Wed-E-V-3-13 Fri-A-V-5-5 Wed-E-V-3-13 Fri-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-2-3 Thu-M-V-5-4 Fri-A-S&T-1-8 Wed-M-V-1-2 Wed-A-O-1-1 Thu-M-SS-1-2 Tue-M-SS-1-7 Tue-M-SS-1-7	141 101 89 150 152 209 129 165 73 131 74 91 167 228 245 163 189 227 107 261 122 138 194 78 194 78 194 78 194 195 195 195 195 195 195 195 195
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Tue-M-V-6-4 Wed-M-V-5-5 Tue-M-V-3-13 Fri-M-V-3-0 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-1-2 Wed-M-V-1-2 Wed-M-V-1-2 Tue-M-SS-1-7 Thu-M-V-6-4 Ward-M-V-6-4 Ward-M-V-6-4	101 89 150 152 209 129 129 165 73 131 74 91 167 228 245 190 191 165 228 245 190 191 165 227 107 261 189 227 107 261 122 219 219 129 129 129 129 129 129 12
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Goojoong Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia Inaguma, Hirofumi	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-3 Thu-M-V-2-4 Wed-E-V-3-1 Fri-A-S&T-1-8 Wed-M-V-1-2 Wed-A-O-1-1 Thu-M-SS-1-2 Tue-M-SS-1-7 Thu-M-V-6-4 Wed-M-V-6-12	101 89 150 152 209 129 129 165 73 131 74 91 167 228 245 190 191 65 163 189 227 261 122 138 194 78 194 135 65
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Goojoong Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Illin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia Inaguma, Hirofumi	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-6-4 Fri-A-V-2-3 Thu-M-V-5-4 Fri-A-S&T-1-8 Wed-M-V-2-12 Wed-A-O-1-1 Thu-M-SS-1-2 Tue-M-SS-1-7 Thu-M-V-6-12 Fri-M-V-6-12 Fri-M-V-6-12 Fri-M-V-6-4	141 101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65 163 189 227 107 261 122 138 194 788 191 122 129 129 129 129 129 129
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia Inan, Omer T.	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-1-1 Fri-A-V-1-2 Wed-A-O-1-1 Thu-M-S-1-2 Tue-M-SS-1-5 Tue-M-SS-1-7 Thu-M-V-6-12 Fri-M-V-3-4 Wed-M-V-6-12 Fri-M-V-3-2 Tue-A-V-2-10	101 89 150 152 209 129 129 165 73 131 74 91 167 228 245 190 191 167 228 2245 190 191 165 227 107 261 189 227 107 261 189 227 107 261 189 227 199 199 199 199 199 199 199 199 199 19
Hutin, Mathilde Hwang, Dongseong Hwang, Ceumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Jida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia Inan, Omer T. Ionescu, Radu Tudor	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-3-12 Wed-A-O-1-1 Thu-M-SS-1-5 Tue-M-SS-1-5 Tue-M-SS-1-5 Tue-M-SS-1-5 Tue-M-SS-1-7 Thu-M-V-6-4 Wed-M-V-6-12 Fri-M-V-3-4 Tue-A-V-2-10 Thu-M-0-12	101 89 150 152 209 129 165 73 131 74 91 167 228 245 190 191 65 163 89 227 261 107 261 122 2138 8194 78 191 135 228 826 177
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Illin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia Inan, Omer T. Ionescu, Radu Tudor	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-M-V-5-2 Wed-M-V-5-2 Wed-E-V-3-13 Fri-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-4 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-1-11 Thu-M-V-6-4 Wed-M-V-1-2 Wed-A-O-1-1 Thu-M-V-6-4 Wed-M-V-6-12 Tue-M-SS-1-7 Thu-M-V-6-4 Wed-M-V-6-12 Fri-M-V-3-4 Tue-A-V-2-10 Thu-M-V-6-11 Thu-M-V-6-12	1101 89 150 152 209 129 165 73 131 74 91 165 733 131 74 91 165 163 189 227 107 261 122 138 194 788 78 191 135 228 866 177
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia Inan, Omer T. Ionescu, Radu Tudor	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-6-1 Thu-M-V-6-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-1-2 Wed-E-V-3-1 Fri-A-S&T-1-8 Wed-M-V-1-2 Wed-A-O-1-1 Thu-M-SS-1-2 Tue-M-SS-1-7 Thu-M-V-6-4 Wed-M-V-6-12 Fri-M-V-3-4 Tue-A-V-2-10 Thu-M-V-3-6	1101 89 150 152 209 129 165 73 131 74 91 165 73 131 74 91 165 163 189 227 107 261 122 138 194 78 191 135 228 86 177 1799
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Goojoong Hwang, Soojoong Hwang, Soojoong Hwang, Song Ju Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia Inan, Omer T. Ionescu, Radu Tudor Iqbal, Zikra Iranzo-Sánchez, Javier	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-M-V-5-2 Wed-E-V-3-1 Wed-E-V-3-1 Wed-E-V-3-1 Fri-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-4 Wed-E-V-2-3 Thu-M-V-6-4 Wed-A-O-1-1 Thu-M-SS-1-2 Tue-M-SS-1-5 Tue-M-SS-1-5 Tue-M-SS-1-5 Tue-M-SS-1-5 Tue-M-SS-1-7 Thu-M-V-6-4 Wed-A-V-6-12 Fri-M-V-3-4 Tue-A-V-2-10 Thu-M-O-3-6 Wed-A-V-6-10	$\begin{array}{c} 101\\ 89\\ 150\\ 152\\ 209\\ 129\\ 129\\ 165\\ 73\\ 131\\ 74\\ 91\\ 167\\ 228\\ 245\\ 190\\ 191\\ 65\\ 163\\ 189\\ 227\\ 161\\ 122\\ 218\\ 84\\ 66\\ 177\\ 154\\ 179\\ 154\\ \end{array}$
Hutin, Mathilde Hwang, Dongseong Hwang, Geumbyeol Hwang, Soojoong Hwang, Soojoong Hwang, Soojoong Hwang, Sung Ju I Ichikawa, Osamu Ide, Yuta Ihori, Mana Iida, Akiyoshi Ijima, Yusuke Ilin, Aleksei Illina, Irina Illium, Steffen Inaam, Rafia Inan, Omer T. Ionescu, Radu Tudor Iqbal, Zikra Iranzo-Sánchez, Javier	Tue-A-V-4-6 Wed-A-V-5-11 Thu-A-V-5-2 Wed-M-V-5-11 Thu-A-V-5-2 Wed-M-V-4-7 Wed-E-V-3-1 Tue-M-V-6-6 Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-6-13 Fri-A-V-1-11 Thu-M-V-6-14 Tue-M-V-2-4 Wed-E-V-2-3 Thu-M-V-5-4 Fri-A-V-1-12 Wed-M-V-1-2 Wed-M-V-1-2 Wed-M-V-1-2 Wed-M-V-1-2 Tue-M-SS-1-7 Thu-M-V-6-12 Fri-M-V-3-6 Thu-M-V-6-10 Thu-M-0-1-1 Thu-M-0-6-10 Wed-A-0-10	141 101 89 150 152 209 209 129 165 73 131 74 91 165 73 131 74 91 165 131 74 91 165 165 190 191 165 165 189 227 107 261 122 138 194 78 78 78 194 78 228 86 1779 154 213 213

Irino, Toshio	Tue-M-V-3-5 67 Thu-M-V-6-9 192
Isaieva Karvna	Fri-A-S&T-1-1 260
Ishi, Carlos Toshinori	Wed-A-V-1-8 142
Ishizuka, Kenkichi	Wed-A-V-1-3 141
Isik, Umut	Tue-E-V-2-5 106
Itto. Akinori	Tue-M-V-4-5 69
	Fri-A-V-1-14 245
Ito, Koichiro	Fri-A-V-2-4 246
Itoyama, Katsutoshi	Wed-A-V-4-9 149
Ivan. Alina	Tue-A-V-2-6 85
Ivry, Amir	Fri-A-SS-1-6 256
T	
J Jacoba Christiaan	M(1843) 2 F 124
Jacoby, Chilistiaan	Fri-M-0-2-2 220
Jahchan, Nataly	Thu-M-SS-2-1 195
Jain, Abhinav	Wed-E-V-1-9 162
Jain, Mahaveer	Wed-M-V-6-3 133
Jakob. Moritz	Wed-E-V-4-9 169
Jalal, Md. Asif	Tue-A-V-6-1 92
Iolilnour Marael M	Fri-M-V-3-10 229
Jampour Monesi, M Janaliyeya Aigerim	Wed-E-V-6-10, 174
Jang, Inseon	Wed-M-V-4-7 129
Jang, Won	Wed-A-V-5-7 151
Jansen, Anne	Wed-A-S&T-1-4 157
Jansky, J Jaramillo Alfredo F	Tue-F-V-3-3 108
Jatteau, Adèle	Tue-E-O-2-2 101
Jawahar, C.V.	Thu-A-V-6-11 213
Jayasuriya, Suren	lue-M-V-3-13 68
Jensen, Jesper	Tue-M-V-3-10 68
Jeong, Kihyuk	Thu-A-V-5-2 209
Jeong, Myeonghun	Thu-A-V-5-4 209
Jęsko, waldemar	Ihu-M-V-1-3 180 Wed-E-V-4-2 168
Jessen, Michael	Tue-M-V-7-8 76
Jetté, Miguel	Thu-A-V-2-9 204
Jl, Youna	Wed-E-V-5-8 172
Jia, Jie	Wed-E-V-5-4 171
Jia, Yan	Fri-M-V-6-10 235
Jia, i e	Tue-M-V-2-9 66
Jiang, Dongcheng	Wed-E-V-3-15 168
Jiang, Dongwei	Wed-M-V-2-4 124
Jiang, i ci	Fri-M-V-7-11 238
Jiang, Liyang	Fri-A-SS-2-11 259
Jiang, Pan-Pan	Fri-A-SS-2-1 257
Jiang, Xinyi	Fri-M-V-5-2 231
Jiang, Yidi	Fri-A-V-6-4 254
Jiao, rumong Iimerson Robbie	wea-M-V-4-10 129 Fri-M-SS-1-7 240
Jin, Chunxiang	Fri-A-S&T-1-7 261
Jin, Mingjie	Thu-A-V-6-5 212
Jin, Qin Iin Zengrui	Fri-A-V-2-3 246
Jiii, Zengrui	Fri-A-SS-2-6 258
Jing, Zijun	Tue-M-V-6-5 73
Jo, Jaecnoon	Fri-M-V-6-8 235
Joglekar, Aditya	Tue-E-O-1-4 100
Joglekar, Sachin	Tue-E-V-3-11 110
Johnson, Khia A	Tue-M-V-7-10 76 Wed-E-V-4-14 170
Johnson, Moris	Thu-M-O-3-6 179
Joly, Arnaud	Thu-A-V-5-7 210
Jones, Lilon	Thu-A-SS-2-7 218
Joo, Young-Sun	Thu-A-V-5-5 209
Jorge, Javier	Wed-A-V-6-10 154
Ioshi. Dhirai	Thu-A-V-6-10 213 Wed-M-V-2-12 126
Joshi, Sachindra	Tue-E-V-4-8 111
Joshi, Sonal	Thu-A-SS-2-10 218
Ioshi. Vikas	Wed-M-V-6-2 133
Jouvet, Denis	Wed-M-V-1-2 122
	Thu-M-SS-1-2 194

Ju, Chelsea JT	Fri-A-V-4-2	250	Kawahara, Hideki	Thu-M-V-6-9	192	Kingsbury, Brian	Tue-E-V-4-8	111
Juan, Alfons	Wed-A-V-6-10	154		Fri-A-S&T-1-1	260		Wed-M-V-2-12	126
I	Thu-A-V-6-10	213	Kawahara, Tatsuya	Wed-M-V-6-12	135		Wed-M-V-6-9	134
Juang, Bling-Hwang	Tue-M-V-4-3	09 111	Kawai Hicachi	Fri-M-V-3-4	228		Wed-A-V-2-11	144
Ιιιησ Ηνιιη	Fri-A-V-6-6	254	Kawai, msasin Kava, Hevsem	Tue-M-SS-1-1	77		Wed-F-V-3-12	145
Jung, Jee-weon	Wed-E-O-1-2	158	11a/a, 11c/0cm 11111	Tue-M-SS-1-8	78		Thu-M-V-2-8	184
5 0,5	Thu-M-V-4-2	187	Kazak, Ekaterina	Wed-M-V-5-9	132	Kinnunen, Tomi	Wed-M-V-4-6	128
	Thu-M-V-4-5	187	Ke, Dengfeng	Fri-A-V-1-9	244		Fri-M-V-7-9	238
Jv, Yukai	Thu-A-V-6-4	212	Ke, Yuxuan	Tue-M-V-3-6	67	Kinoshita, Keisuke	Tue-M-V-3-5	67
Jyothi, Preethi	Wed-M-V-2-1	123	Keesing, Aaron	Thu-A-V-1-10	202		Tue-E-V-2-10	107
	Wed-A-V-1-1	141	Kelly Barbara	Fri-A-V-3-1	247		Wed-A-V-4-2	140
	Thu-A-V-2-5	203	Keren Gil	Wod.M.V.4.12	130		Thu-M-V-3-3	205
l l	Thu A V 2 J	205		Wed-M-V-6-3	133		Thu-A-V-3-5	205
K			Kershenbaum, Benji	Tue-A-V-4-3	89		Thu-A-V-4-5	208
Kaandorn Casper S	Tue-M-SS-1-1	77	Kerz, Elma	Thu-A-SS-2-6	217	Kirchhoff, Katrin	Tue-A-V-5-1	90
Rumuorp, cusper 5	Tue-M-SS-1-8	78	<i>V</i> · F	Fri-A-V-1-10	244	Vishinger Tabashi	<u>Thu-M-V-7-8</u>	193
Kachare, Pramod H	Thu-M-S&T-1-7	197	Kesim, Ege	Wed-M-V-3-8	127	Kisniyama, Takeshi	Wed-M-V-5-2	130
Kachkovskaia, Tatiana .	Wed-A-V-1-5	141	Khare Shreva	Wod-M-V-2-1	123	Klabbers, Esther	Tue-M-0-1-4	60
Kafentzis, George P	Tue-E-SS-1-3	117	Kildre, Shi'eyu	Wed-F-V-1-7	161	Klakow, Dietrich	Wed-M-V-5-3	131
Kann, Jacob	Tue-M-V-5-10	72	Kharitonov, Eugene	Wed-M-V-2-10	125		Thu-M-SS-2-5	196
	Tue-A-V-4-4 Tue-Δ-V-4-8	90		Thu-A-V-5-6	210		Fri-M-V-1-8	224
Kajarekar, Sachin	Fri-A-SS-2-15	260	Khassanov, Yerbolat	Wed-E-V-6-10	174	Vlainant Matthiaa	Fri-M-V-6-1	234
Kakegawa, Naoto	Tue-M-V-2-4	65	Khelif, Khaled	Thu-M-SS-2-1	195	Kleich Ondřej	<u>I nu-M-SS-2-3</u>	195
Kaland, Constantijn	Wed-E-V-4-1	168	Knokniov, Yuri	Fri-M-V-3-2	122	Kiejen, ondrej	Thu-M-O-3-1	178
Kalgaonkar, Kaustubh	Wed-M-V-4-12	130	Khorrami Khazar		122	Klementiev. Anita	Wed-E-V-2-7	164
Kalinli, Ozlem	Tue-M-V-2-8	66	Khosravani. Abbas	Thu-M-0-3-4	179	Klimkov, Viacheslav	Wed-M-V-4-10	129
	Wed-M-V-6-3	133	Khudanpur, Sanjeev	Wed-M-V-1-1	122	Klinke, Piotr	Fri-A-O-2-1	242
	Wed-A-V-2-1	142	1 0	Wed-E-O-1-1	157	Klumpp, P	Tue-M-SS-1-3	77
	Wed-A-V-3-14	147		Thu-M-O-3-6	179	Vill V	Thu-A-SS-2-2	217
	Fri-A-V-3-6	248		Thu-A-V-3-9	206	Knill, K		219
	Fri-A-V-3-10	249	Vilgour Vouin	Thu-A-V-6-5	212	Ko Hanseok	Tuo.4.V-1.1	233
Kalluri, Shareef Babu	Tue-A-SS-2-8	98	Kilgour, Kevin Kilpikoski Juho	Fri-M-V-6-13	230	Ko, Tom	Wed-A-V-2-8	144
Kamble, Madhu R	Tue-A-SS-1-1	94	Kim Bongwan	Wed-A-V-5-7	150	,	Fri-M-V-6-7	235
Vamaaka Hirokazu	Tue-A-SS-2-2	96	Kim, Byeonggeun	Fri-A-V-3-3	247		Fri-M-V-6-11	236
Kamo Naoyuki	Tue-E-V-0-8	115	Kim, Daeyoung	Fri-A-V-5-1	251	Ko, Yuka	Wed-A-V-6-9	154
Kamper, Herman	Wed-M-V-2-3	124	Kim, Dalhyun	Wed-A-V-5-1	150	Kobayashi, Akio	Tue-M-V-6-7	74
maniper, merman merma	Wed-M-V-2-5	124	Kim, Eesung	Wed-E-V-5-13	173	Vohavashi Vazuhiro	Wed-E-V-1-4	161
	Wed-M-V-2-6	124	Kim, Eunggyun	Fri-M-V-6-8	235	Kobayashi Tetsunori		114
	Thu-M-V-2-1	182	Kim, Eunhwa	Fri-M-V-6-8	235	Kobayasin, retsunori	Thu-M-V-7-4	192
	Thu-M-V-2-5	183	Kim, Gwantae	Tue-A-V-1-1	82	Kocharov. Daniil	Wed-A-V-1-5	141
Kanda, Naoyuki	Wed-M-V-6-5	134	Kim Hwa Voon	Wed-E-V-3-1	702	Kocour, Martin	Thu-M-SS-2-2	195
	Wed-E-V-3-14	167	Kim Hyeongiu	Thu-A-V-5-4	209	-	Thu-M-SS-2-4	195
	Thu-M-V-3-9	202	Kim. Jae-Min	Wed-A-V-5-11	152		Thu-M-SS-2-5	196
	Thu-A-V-2-2 Thu-A-V-2-3	203	, 5	Fri-M-V-4-2	230	Koehler, Thilo	Tue-M-V-2-8	66
	Fri-A-V-1-2	243	Kim, Jaeyoung	Thu-A-V-2-4	203	Koenig, Laura L	Thu-A-O-1-6	199
Kang, Haimei	Fri-A-V-6-5	254	Kim, Jangho	Fri-A-V-3-9	249	Koepke, A. Sopilia	Wed-E-O-3-5	100
Kang, Hong-Goo	Thu-A-V-5-2	209	Kim, Ji-Hoon	Wed-A-V-5-5	151	Koishida Kazuhito	Wed-F-V-5-5	171
Kang, Jian	Tue-M-V-5-9	72	Kim, Ji-Hwan	Tue-A-V-1-8	83	Robindu, Ruzunito	Wed-E-SS-1-1	175
Vara - En en	Fri-M-SS-1-5	240	Kim long lin	Fri-M-V-4-2	250	Kojima, Atsushi	Thu-M-O-1-2	177
Kang, Jingu		215	Kim Joo-Yeon	Fri-M-V-1-1	222	Kojima, Ryosuke	Wed-A-V-4-9	149
	Thu-A-SS-1-12	216	Kim, Juntae	Wed-A-V-5-Z	151	Koldovský, Z	Tue-A-S&T-1-7	99
Kang, Tae Gyoon	Wed-F-V-3-1	165	Kim, Kwangyoun	Wed-M-V-6-14	135	Kolossa, Dorothea	Tue-A-SS-1-5	95
Kang, Woo Hyun	Wed-A-SS-1-4	155	Kim, Kyuhong	Tue-M-V-6-9	74	Komatani Kazana	Wed-A-V-4-2	148
Kang, Yueteng	Tue-A-V-4-1	88	Kim, Minchan	Fri-A-V-5-5	252	Komatsu Tatsuwa	Fri-M-V-5-7	232
	Fri-A-V-1-5	243	Kim, Minje	Wed-E-V-5-1	170	кошасоа, гасойуа	Thu-A-SS-1-5	215
Kansara, Seemran	Wed-A-V-1-7	142	ыт, Nam Soo	Wed-A-SS-1-4	155	Kong, Jiaming	Fr <u>i-M-SS-1-6</u>	$\frac{1}{240}$
Kantnaraju, Pavan	I ue-M-O-3-2	62 174		Thu-A-V-5-4	209	Kong, Qiuqiang	Tue-M-V-3-7	67
Kanja Enkeleida	Wed-E-V-6-9	174 120		Fri-A-V-5-5	252	Kong, Tianlong	Thu-M-SS-1-3	194
Kapoor, Parul	Thu-A-V-6-11	213	Kim, Seok Min	Thu-A-V-6-12	213	Kong, Yuxiang	Thu-A-V-6-4	212
Karanasou. Pennv	Thu-A-V-5-7	210	Kim, Seong-Hu	Tue-M-V-1-1	62	Vongthorner T	Fri-A-SS-1-3	256
Karas, Vincent	Tue-A-SS-2-3	97	Kim, So Hyun	Thu-M-V-1-2	180	Kongthaworn, 1	Wed-E-V-6-1	1/3
Karita, Shigeki	Wed-A-V-3-11	147	Kim, Sunhee	Fri-A-SS-2-13	259	Konnaranu Sunil Kumar	Tue-M-SS-1-8	79 08
	Fri-M-V-4-1	229	KIM, SUNWOO	Wed-E-V-5-1	1/0	Koppelmann Timm	Tue-A-SS-1-5	90
Karlapati, Sri	Thu-A-V-5-7	210	кші, зиуоші	Wed-M-V-6-3	133 142	Korenevskava, Mariva	Fri-M-V-3-2	227
Karpov, Alexey	Tue-M-SS-1-11	79 212	Kim. Tae-Woo	Wed-M-V-3-1	126	Korenevsky, Maxim	Fri-M-V-3-2	227
Karnov Nikolav	Thu-A-V-6-9 Wed-M-O-1-1	213 119	Kim, Ui-Hyun	Tue-M-V-6-2	73	Koriyama, Tomoki	Tue-M-V-2-3	65
Karra, Kiran	Thu-A-V-4-9	208	Kim, Yelin	Thu-A-V-1-6	201		Wed-M-V-3-6	127
Kashino, Makio	Wed-M-V-5-12	132	Kim, You Jin	Thu-M-V-4-5	187	Konzolaure Derrich	Wed-A-V-5-4	151
Kashiwagi, Yosuke	Tue-M-V-5-8	72	T 71 T 7 T 1	Thu-A-V-6-6	212	Korzekwa, Daniel	Wed-M-V-4-10	129
Kashyap, Madhav M	Wed-E-V-5-9	172	Kim, Young-Ik	Tue-M-V-2-2	64		Fri-M-V-1-2 Eri-A-V-1-1	222
Kaszuba-Miotke, K	Fri-A-O-2-1	242		Wed-M-V-3-1	120 151	Kosgi, Saiteia	Fri-A-V-5-3	252
Kataria, Saurabh	Tue-E-V-1-9	104	Kim, Young-Kyung		180	Koshizuka, Takeshi	Tue-E-V-6-4	115
Katavama Uavata	Fri-A-V-4-5	25U 102	King, Brian	Fri-A-V-3-4	248	Kostek, Bozena	Fri-M-V-1-2	222
Katayama, nayato Katsurada Kowichi	Tuo-E-V-6-4	192 115	King, Jeff	Tue-M-V-4-8	70	·· .1 ·· ··	Fri-A-V-1-1	243
Katourada, Koultill	Thu-M-V-6-3	190	King, Simon	Wed-E-V-6-2	173	Kothare, Hardik	Wed-A-V-1-7	142
Katz, Boris	Thu- <u>A-V-6-1</u>	211	-	Fri-M-O-2-3	221	Kotov Denie	Fri-A-SS-2-2	257
Kaushik, Lakshmish	Tue-M-V-2-10	66		Fri-M-O-2-4	221	KOLOV, DEIIIS	Tue-M-55-1-11	79
Kaushik, Vishesh	Wed-M-V-5-10	132						

INTERSPEECH 2021

Koul, Ashwani	Thu-M-V-3-2	185
Kounovský, T	Tue-A-S&T-1-7	99
Kovács, György	Tue-A-SS-2-4	97
Kowalczyk, Konrad	Tue-A-O-1-2	80
	Fri-M-O-3-2	221
Kraljevski, Ivan	Tue-E-O-2-3	101
Krikun. Maxim	Wed-A-V-6-2	152
Krishnamohan, Venkat	Thu-A-V-4-1	207
,,	Thu-A-V-4-6	208
Krishnan Aravind	Тие-А-S&T-1-4	- 99
Krishnan Prashant	Tue-A-SS-2-1	96
Krishnaswamy Arvindh	Tue-F-V-2-5	106
Krug Paul Konstantin	Thu.M.V.6.6	191
Kubo Votaro	Wed-A-V-3-11	147
Kuchaiev Oleksii	Wed-M-O-1-4	119
Kuchinsky Stefanie F	Tue-M-0-2-1	60
Kucsko Georg	Wed-M-O-1-4	119
Kudera Jacek	Fri.M.V.1.8	224
Kuehne Hilde	Wod-M-V-2-12	126
Ruenne, mue	Thu-M-V-2-12	184
Kuhnert Barhara	Tuo-A-O-2-1	81
Kumar Aiswarya Vinod	Fri. A.V. 4.11	251
Kumar, Aiswarya vinou	Tuo E V 2 12	110
Kumar Anach		102
Kullial, Ayusii		195
Kullidi, C. Siliva	Wed-M-V-1-6	123
Kumar Mari Canash	Wod EV 1 C	229 161
Kumar Nooraj		115
Kumar Dupoot	Wod M V 5 10	120
Kumar Dobit		132
Kullai, KUllit	Tue-A-SS-2-1	90 09
	Tue-A-SS-Z-TT	90 127
Vumar Saleat	Tue M1/ 2 10	137
Kullidi, Saket		142
Kullial, Shalikal		143
Kullial, SilaSili	1 ue-E-V-5-5	115
Vumatani Vanishi	wea-E-V-5-7	115
Kullididili, Kellicili		201
Kullidwal, POOJa		201
Kumm, Jocnen	Fri-A-SS-2-2	257
Kullz, Illilo P	Thu-M-S&T-T-4	197
Kunzmann, Slegiried	Iue-E-V-4-2	110
Kuo, Cheng-Hao	Thu-A-V-1-6	201
Kuo, Hong-Kwang J	Tue-E-V-4-8	111
	Fri-A-V-6-6	254
L'invoto l'obrito		1/1/1
Kurata, Gakuto	wea-A-v-2-11	177
Kuriakose, Jom	Wed-A-V-2-11 Wed-E-V-1-6	161
Kurata, Gakuto Kuriakose, Jom Kurimo, Mikko	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3	161 119
Kuriaka, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2	161 119 102
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12	161 119 102 154
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9	161 119 102 154 249
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6	161 119 102 154 249 212
Kurata, Gakuto Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia . Kwak, Nojun Kwon, Yoohwan Kwon, Young D.	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8	161 119 102 154 249 212 74
Kurata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia . Kwak, Nojun Kwon, Yoohwan Kwon, Young D Kwon, Youngki	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2	161 119 102 154 249 212 74 187
Kurata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia . Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-5	161 119 102 154 249 212 74 187 187
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D Kwon, Youngki	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-5 Thu-M-V-4-5 Thu-A-V-6-6	144 161 119 102 154 249 212 74 187 187 212
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D Kwon, Youngki	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-6 Fri-M-O-2-5	144 161 119 102 154 249 212 74 187 187 212 221
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia . Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-A-V-6-8 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3	144 161 119 102 154 249 212 74 187 187 212 221 121
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia . Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D Kwon, Young Ri Ky, Nguyen Hoang	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3	144 161 119 102 154 249 212 74 187 187 212 221 121
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia . Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D Kwon, Young Ri Ky, Nguyen Hoang Kynych, Frantisek	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3	144 161 119 102 154 249 212 74 187 187 212 221 121
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D. Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek L Laali, Maiid	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-5 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3	144 161 119 102 154 249 212 74 187 187 212 221 121
Kuriata, Gakuto Kurimo, Mikko Kuznetsova, Anastasia Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Young D. Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek L Laali, Majid Lachiri, Zied	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-0-2-5 Wed-M-0-3-3 Tue-M-V-4-8 Tue-M-V-4-8	144 161 119 102 154 249 212 74 187 187 212 221 121 70 87
Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D Kwon, Young D Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-5 Thu-M-V-4-5 Thu-M-V-4-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-1	144 161 119 102 154 249 212 74 187 187 212 221 121 70 87 257
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoong D Kwon, Young D Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A.	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-5 Thu-M-V-4-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5- Fri-A-SS-2-1 Fri-A-SS-2-1	144 161 119 102 154 249 212 74 187 187 187 221 221 121 70 87 257 259
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D. Kwon, Young D. Kyon, Youngki Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2	144 161 119 102 154 249 212 74 187 187 212 221 121 70 87 257 259 180
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Yoong D. Kwon, Young D. Kwon, Young D. Kwon, Youngki Kynych, Frantisek L Laali, Majid Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-8 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4	144 141 119 102 154 249 212 74 187 187 212 221 121 70 87 257 259 180 159
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D. Kwon, Young B. Kwon, Young D. Kwon, Youngki Kyon, Youngki Kurisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lai, Catherine	Wed-A-V-2-11 Wed-F-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-A-V-6-6 Fri-M-V-4-5 Thu-A-V-6-6 Fri-M-V-4-5 Thu-A-V-6-6 Fri-M-V-4-8 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-11 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4	141 119 102 154 249 212 74 187 187 212 221 121 70 87 257 259 180 159 109
Kuriata, Gakuto Kurimo, Mikko Kuznetsova, Anastasia Kuznetsova, Anastasia Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Young D. Kwon, Youngki Kyon, Youngki Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lainer, Rimita Lai, Catherine Lai, Cheng-I Jeff	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-5 Thu-M-V-4-5 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-1-1	141 119 102 154 249 212 74 187 212 221 121 70 87 257 259 180 159 219
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwak, Nojun Kwon, Yoong D. Kwon, Young D. Kwon, Young D. Kwon, Youngki Kynych, Frantisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Laínez, José Lakhotia, Kushal	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-1-1 Tue-E-V-3-9	141 119 102 154 249 212 274 187 187 212 221 121 70 87 257 259 180 159 219 109
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kuvak, Nojun Kuvon, Yoohwan Kwon, Young D. Kwon, Youngki Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek L Laali, Majid Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lai, Cheng-I Jeff Lakhotia, Kushal	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-8 Thu-M-V-4-8 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-1-1 Tue-E-V-3-9 Thu-A-V-5-6	141 119 102 154 249 212 241 187 187 212 221 121 70 87 257 259 180 159 109 210
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Yooung D. Kwon, Young D. Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lai, Cheng-I Jeff Lainez, José Lahkotia, Kushal	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-2 Thu-M-V-4-8 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-3-9 Fri-A-SS-2-12 Thu-M-V-1-11 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-11	144 119 102 154 249 212 74 187 187 212 221 121 70 87 257 259 109 210 210 210 77
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoohwan Kwon, Young D. Kwon, Young D. Kwon, Youngki Kynych, Frantisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lai, Catherine Lai, Cheng-I Jeff Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy	Wed-A-V-2-11 Wed-E-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-2 Thu-M-V-4-8 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-3 Fri-A-SS-2-12 Thu-M-V-1-3-9 Fri-M-O-1-11 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-11 Tue-E-V-3-1	144 119 102 154 249 212 249 214 187 187 212 221 121 70 87 257 259 180 159 109 219 109 210 77 115
Kuriata, Gakuto Kurimo, Mikko Kuznetsova, Anastasia Kuznetsova, Anastasia Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D. Kwon, Youngki Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lai, Catherine Lai, Cheng-I Jeff Laínez, José Lahiri, Rimita Lainez, José Lahkotia, Kushal Lalhminghlui, Wendy Lall, Brejesh	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-V-3-9 Fri-M-O1-11 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-11 Tue-E-V-6-7 Thu-M-V-7-7	141 119 102 154 249 212 74 187 212 212 212 212 221 121 70 87 257 259 180 109 219 109 219 109 210 77 115 193
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D. Kwon, Youngki Kwon, Youngki Kyon, Youngki Kyon, Youngki Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lainez, José Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lalwani, Garima Lawani, Garima	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-0-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-0-1-1 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-11 Tue-E-V-3-7 Tue-M-V-7-7	141 119 102 154 249 212 74 187 212 74 187 212 74 187 221 121 70 87 257 259 180 109 219 109 210 77 115 193 72
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D Kwon, Young K Kwon, Young K Kwon, Young K Kyon, Young K Kyon, Young K Lalai, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lam, Garima Lam, Tsz Kin	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-5 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-1-1 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-11 Tue-E-V-6-7 Thu-M-V-7-7 Tue-M-V-5-11	141 119 102 154 249 212 74 187 187 221 121 221 121 70 87 257 259 180 159 210 77 193 72 113
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kuvon, Yoohwan Kwon, Young D Kwon, Young K Kwon, Young K Kwon, Young Maine Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Laínez, José Lakhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lama, Max W.Y. Lama, Jatin	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-6-8 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-8 Tue-A-V-3-5 Wed-M-O-3-3 Wed-E-O-2-4 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-11 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-11 Tue-E-V-5-7 Tue-M-V-5-11 Tue-E-V-5-7 Wed-A-V-1-1	141 119 102 154 249 212 74 187 187 221 121 221 121 70 87 257 259 180 159 210 77 115 193 72 113 141
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kuvon, Yoohwan Kwon, Young D. Kwon, Young Ri Kwon, Young Ri Kwon, Young Ri Ky, Nguyen Hoang Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lai, Cheng-I Jeff Lainez, José Lakhotia, Kushal Lall, Brejesh Lalwani, Garima Lam, Max W.Y. Lamba, Jatin	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-M-V-4-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O1-1 Tue-E-V-3-9 Thu-A-V-5-6 Tue-H-V-7-11 Tue-E-V-6-7 Tue-M-V-7-7 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1	141 119 102 154 249 74 187 212 241 187 212 212 211 121 70 87 257 259 180 159 109 210 77 115 193 72 113 141 101
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D Kwon, Young D Kwon, Youngki Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lawani, Garima Lam, Max W.Y. Lamel, Lori	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-M-V-4-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-5-6 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-1-1 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-7 Tue-M-V-5-77 Wed-A-V-11 Tue-E-V-5-7 Wed-A-V-11 Tue-E-O-2-2 Thu-M-SS-2-12	141 119 102 154 249 74 187 187 212 221 121 70 87 257 259 180 159 109 210 77 115 193 72 113 141 101 196
Kuriata, Gakuto Kurimo, Mikko Kuznetsova, Anastasia Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D. Kwon, Young D. Kwon, Youngki Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lainez, José Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lalkani, Garima Lawani, Garima Lampa, Jatin Lammert, Adam	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-0-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-6-6 Fri-M-02-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-01-1 Tue-E-V-3-9 Fri-M-0-1-1 Tue-E-V-3-9 Thu-M-V-7.7 Tue-M-V-5-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Tue-M-V-5-26 Thu-M-V-6-68 Thu-M-V-6-68	141 119 102 154 249 212 74 187 212 212 74 187 212 221 121 70 87 257 259 180 109 219 109 219 109 219 109 210 77 115 193 72 113 141 101 196 191
Kuriata, Gakuto Kurimo, Mikko Kuznetsova, Anastasia Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D. Kwon, Young D. Kwon, Young R. Kwon, Young Maile Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Lachiri, Zied Ladewig, Marilyn A. Lainez, José Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lalwani, Garima Lammert, Adam Lammert, Adam	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-5 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-0-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-0-1-1 Tue-E-V-3-9 Fnu-M-V-7-11 Tue-E-V-3-9 Thu-M-V-7-7 Tue-M-V-5-6 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-2 Thu-M-V-6-8 Tue-M-V-3-8	141 119 102 154 249 212 74 187 212 74 187 212 74 187 212 74 187 212 212 121 70 87 257 259 180 109 210 77 115 193 72 113 141 106 191 66
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Young D Kwon, Young K Kwon, Young D Kwon, Young K Kynych, Frantisek Laali, Majid Lachiri, Zied Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lam, Garima Lam, Tsz Kin Lammert, Adam Lan, Tian	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-0-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-6-6 Fri-M-0-2-5 Wed-M-O-3-3 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-0-1-1 Tue-E-V-3-9 Fni-M-O-1-1 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-71 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-2 Thu-M-S2-6 Thu-M-S2-6 Thu-M-S2-6 Thu-M-S2-6 Thu-M-S2-6 Thu-M-S2-7 Wed-A-V-1-1 Tue-E-V-5-3 Thu-M-S2-6 Thu-M-V-6-8 Tue-A-V-3	141 119 102 154 249 212 74 187 212 74 187 212 212 74 187 212 212 121 70 87 257 259 180 159 109 210 77 193 72 113 141 101 196 206
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Young D Kwon, Young K Kwon, Young K Kwon, Young K Kwon, Young K Kynych, Frantisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lam, Max W.Y. Lamel, Lori Lammert, Adam Lan, Tian Lan, Tian	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-M-O-2-5 Wed-E-O-2-5 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-11 Tue-E-V-3-9 Fri-M-O-11 Tue-E-V-3-9 Thu-A-V-5-6 Tue-H-V-5-61 Tue-H-V-5-71 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Thu-M-V-5-8 Tue-M-V-5-11 Tue-E-V-5-2 Thu-M-V-5-8 Tue-M-V-3-10 Tue-E-O-2-2 Thu-A-V-3-10	141 119 102 154 249 212 74 187 212 211 121 70 87 257 259 180 159 219 1009 210 77 193 72 113 141 101 196 66 206 100
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kuvon, Yoohwan Kwon, Young D Kwon, Young Ki Kwon, Young Main Kwon, Young Main Kwon, Young Main Kwon, Young Main Ky, Nguyen Hoang Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lai, Cheng-I Jeff Laikhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lam, Max W.Y Lamba, Jatin Lamert, Adam Lan, Tian Lan, Tian Lancucki, Adrian	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-M-V-4-6 Fri-A-SS-2 Wed-A-V-3-5 Thu-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-1-1 Tue-E-V-3-9 Thu-A-V-5-6 Tue-E-V-3-9 Thu-M-V-7-7 Tue-E-V-5-7 Wed-A-V-11 Tue-E-V-5-7 Wed-A-V-11 Tue-E-O-2-2 Thu-M-V-6-8 Tue-M-V-5-11 Tue-E-O-2-2 Thu-A-V-6-8 Tue-M-V-3-10 Tue-E-O-2-2 Thu-A-V-6-8 Tue-M-V-3-10 Tue-E-0-1-2	141 119 102 154 249 212 74 187 212 211 121 70 87 257 259 180 159 109 210 77 115 193 72 113 141 101 96 206 100 100
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kuvon, Yoohwan Kwon, Young D. Kwon, Young D. Kwon, Youngki Kwon, Youngki Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lai, Cheng-I Jeff Lainez, José Lakhotia, Kushal Lall, Brejesh Lalwani, Garima Lam, Max W.Y. Lamet, Adam Lan, Tian Lan, Tian Lan, Tian Lan, Tian Lan, Tian Lange, Rael T.	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-M-V-4-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-6 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O1-11 Tue-E-V-3-9 Fri-M-O1-11 Tue-E-V-3-9 Thu-A-V-5-61 Tue-M-V-7-71 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-O-2-2 Thu-M-S2-66 Thu-M-S2-66 Tue-E-O-5-7 Wed-A-V-11 Tue-E-O-5-7 Wed-A-V-11 Tue-E-O-12 Thu-M-V-6-8 Tue-E-O-11 Tue-E-O-12 Tue-E-O-12	141 119 102 154 249 212 74 187 212 74 187 212 211 121 70 87 257 259 1809 219 109 210 77 115 193 72 191 66 206 100 60
Kuriata, Gakuto Kurimo, Mikko Kuznetsova, Anastasia Kuznetsova, Anastasia Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D. Kwon, Young D. Kwon, Youngki Kwon, Youngki Ky, Nguyen Hoang Kynych, Frantisek L Laali, Majid Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lai, Catherine Lai, Catherine Lainez, José Lakhotia, Kushal Lall, Brejesh Lalwani, Garima Lalwani, Garima Lammert, Adam Lan, Tian Lan, Tian Lan, Tian Lange, Rael T. Langidot, Itshak	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-02-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-01-1 Tue-E-V-3-9 Fnu-A-V-5-6 Tue-M-V-7.11 Tue-E-V-3-9 Thu-M-V-7.7 Tue-M-V-5-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Thu-M-V-5-8 Tue-M-V-5-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Thu-M-V-6-8 Tue-M-V-3-10 Tue-E-O-1-2 Thu-M-V-6-3 Thu-M-V-6-12 Thu-M-V-4-3-10	141 119 102 154 249 212 74 187 212 74 187 212 212 121 121 70 87 257 259 180 109 219 109 219 109 219 109 219 109 210 77 115 193 72 113 141 191 66 206 1000 60 188
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Yoong D. Kwon, Young D. Kwon, Young X Kwon, Young D. Kwon, Young D. Kwon, Young X Kwon, Young X Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Lachiri, Zied Lachiri, Zied Ladewig, Marilyn A. Lainez, José Lainez, José Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lawani, Garima Lammert, Adam Lan, Tian Lan, Tian Lan, Tian Lan, Tian Lange, Rael T. Lapidot, Itshak Laprie, Yves	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-0-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-A-SS-2-11 Tue-E-V-3-9 Fri-M-O-1-1 Tue-E-V-3-9 Fri-M-O-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Wed-A-V-1-11 Tue-E-V-5-7 Tue-M-V-5-8 Tue-M-V-3-10 Tue-E-O-1-2	141 119 102 154 249 212 74 187 212 74 187 212 74 187 212 74 187 212 212 121 70 87 257 259 180 109 210 77 115 193 72 113 141 106 206 100 106 206 100 108 198
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kwon, Yoohwan Kwon, Young D. Kwon, Young D. Kwon, Young Maile Kwon, Young Maile Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lainez, José Lakhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lammert, Adam Lan, Tian Lan, Tian Lan, Tian Lange, Rael T Langidot, Itshak Laprie, Yves	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-5 Thu-M-V-4-5 Thu-A-V-6-6 Fri-M-O-2-5 Wed-M-O-3-3 Tue-M-V-4-8 Tue-A-V-3-5 Fri-A-SS-2-12 Thu-M-V-1-2 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-1-1 Tue-E-V-3-9 Fri-M-O-1-1 Tue-E-V-3-9 Thu-A-V-5-6 Tue-H-V-7-7 Tue-H-V-5-11 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-V-5-7 Thu-M-V-6-8 Tue-M-V-3-20 Thu-M-V-3-10 Tue-E-O-1-2 Tue-M-V-2-11 Tue-E-O-1-2 Tue-M-O-2-1 Thu-A-V-3-11	141 119 102 154 249 212 74 187 212 212 74 187 212 212 211 121 70 87 257 259 180 159 219 109 210 77 193 72 193 72 113 141 101 196 206 1000 60 200 100 66 100 68 198 227
Kuriata, Gakuto Kuriakose, Jom Kurimo, Mikko Küster, Dennis Kuznetsova, Anastasia Kwak, Nojun Kuvon, Yoohwan Kwon, Young D. Kwon, Young D. Kwon, Young D. Kwon, Young Mailer Kwon, Young Mailer Ky, Nguyen Hoang Kynych, Frantisek Laali, Majid Lachiri, Zied Lachiri, Zied Ladewig, Marilyn A. Lahiri, Rimita Lai, Catherine Lai, Catherine Lai, Catherine Lai, Catherine Laikhotia, Kushal Laikhotia, Kushal Laikhotia, Kushal Laikhotia, Kushal Lalhminghlui, Wendy Lall, Brejesh Lama, Garima Lam, Tsz Kin Lamet, Lori Lammert, Adam Lan, Tian Lan, Tian Lange, Rael T. Laptie, Yves Laptie, Yves Laptie, Yves	Wed-A-V-2-11 Wed-K-V-1-6 Wed-M-O-1-3 Tue-E-O-3-2 Wed-A-V-6-12 Fri-A-V-3-9 Thu-A-V-6-6 Tue-M-V-4-8 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-2 Thu-M-V-4-8 Tue-A-V-3-5 Fri-M-O-2-5 Wed-E-O-2-4 Tue-E-V-3-9 Fri-M-O-11 Tue-E-V-3-9 Fri-M-O-11 Tue-E-V-3-9 Thu-A-V-5-6 Tue-M-V-7-71 Tue-B-V-5-71 Tue-E-V-5-7 Wed-A-V-1-1 Tue-E-O-2-2 Thu-M-V-3-10 Tue-E-O-1-2 Tue-M-V-3-10 Tue-E-O-1-2 Tue-M-O-2-1 Thu-A-V-4-9 Thu-A-V-4-9 Thu-A-V-3-2 Tue-A-S&T-1	141 119 102 154 249 212 74 187 212 212 74 187 212 211 121 70 87 257 259 180 159 219 100 72 193 72 113 141 101 196 66 206 100 60 100 60 100 60 100 60 198 227 99

Lardy, Thomas		
Laruy, momas		04
	Tue-A-33-1-1	34
Larrouv-Maestri, Pauline	Fri-M-O-2-2	220
Laurent Antoine		188
Laurent, Antonie	1 HU-IVI- V-4-7	100
Lavelle, Grace	Tue-A-V-2-6	85
Lavrukhin Vitaly	Wod MO14	110
Laviukiiii, vitaiy	weu-w-0-1-4	115
	Wed-E-V-6-8	174
Lawson Jenna	Tue-M-SS-1-9	78
		170
Lazaridis, Alexandros	Thu-M-O-3-4	179
Le Duc	Wod-M-V-6-3	133
LC, Duc	weu-w-v-o-s	1.10
	Wed-A-V-2-1	142
	Wod-A-V-2-1	145
	WeuAVJI	1 1 7
	Wed-A-V-3-14	147
	Fri-A-V-3-6	248
T . TT		110
Le, Hang	Wed-M-O-1-5	119
Le Xiaohuai	Wed-E-SS-1-4	176
Les Calla		200
Lea, Colin	Fri-A-SS-2-15	260
Leal. Isabel	Wed-F-V-3-6	166
Leak Managerst		210
Lech, Margaret	I hu-A-SS-2-8	218
Leclere Justine	Thu-A-O-1-1	198
Lectere, Justine minin		110
Lecouleux, Benjamin	wed-M-O-1-5	119
Lee. Ann	Tue-A-V-4-4	89
Loo Pong Jin		107
Lee, boilg-Jill	I NU-IVI-V-4-2	107
	Thu-M-V-4-5	187
		212
	111u-A-v-0-0	212
Lee, Chi-Chun	Tue-E-V-3-6	108
Lee, Chin-Hui	Tue-M-V-6-5	73
, inn inn		6-
	Tue-A-SS-T-6	95
	Wed-E-V-5-6	171
	Thu M V 2 7	100
	Thu-M-V-2-7	103
	Thu-M-V-4-6	188
Lee Chul		187
	1 11u-1vi- v-4- 1	107
Lee, Chwee Beng	Thu-A-V-6-7	212
Lee Daniel	Fri-M-V-6-8	235
Lee, Dunier		200
Lee, Dongnyun	Iue-A-V-I-8	83
Lee. Dongvuh	Fri-M-V-6-8	235
Lee, Dong, up theor		100
Lee, Gyeong-Hoon	wed-IVI-V-3-1	120
Lee, Hoshik	Wed-E-V-3-1	165
Lee Hung-Shin	Wed-A-V-6-5	153
Lee, mang binn minin		100
	I hu-M-V-3-8	180
	Thu-A-V-5-9	210
		04
Lee, mulig-yr	Tue-A-v-0-7	34
	Tue-E-V-3-9	109
	Wed-F-V-3-5	166
	WedEVCO	174
	weu-L-v-0-5	100
	Thu-M-V-3-7	186
	Fri-M-V-6-11	236
		220
	FFI-IVI-V-7-0	230
Lee, Hyunjae	Tue-A-V-5-9	- 92
Lee Jaesong	Thu.A.SS.1.7	215
Lee, Jaesong	Thu-A-SS-1-7	215
Lee, Jaesong Lee, Jaeyun	Thu-A-SS-1-7 Tue-M-V-6-9	215 74
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5	215 74 151
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5	215 74 151
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1	$215 \\ 74 \\ 151 \\ 165 \\ 201 \\$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Jinkyu	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3	215 74 151 165 247
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Jinkyu	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Word M-V-4-1	215 74 151 165 247 128
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Jinkyu Lee, Junhyeok	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1	215 74 151 165 247 128
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Junhyeok Lee, Junhyeok	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1	215 74 151 165 247 128 251
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Jinkyu Lee, Keon Lee, Keon Aik	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Jinkyu Lee, Junhyeok Lee, Keon Lee, Kong Aik	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 $
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Junhyeok Lee, Keon Lee, Kong Aik	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9 Tue-E-V-1-11	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ $
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Jinkyu Lee, Junhyeok Lee, Keon Lee, Kong Aik	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9 Tue-E-V-1-11 Tue-E-V-1-12	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Keon Lee, Kong Aik	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9 Tue-E-V-1-12 Fri-M-V-7-9 Fri-M-V-7-9	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Junhyeok Lee, Kong Aik	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9 Tue-E-V-1-11 Tue-E-V-1-12 Fri-M-V-7-9	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 100 \\$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Jinkyu Lee, Junhyeok Lee, Keon Lee, Kong Aik	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9 Tue-E-V-1-12 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 109 \\ 109 \\ 109 \\ 100 \\ $
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Ko-tik Lee, Lin-shan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9 Tue-E-V-1-11 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-9	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Ko-tik Lee, Lin-shan Lee, Lukas	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Tue-M-V-1-9 Tue-E-V-1-11 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-5 Wed-E-V-3-5	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Kortik Lee, Lin-shan Lee, Lukas	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-19 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-5 Wed-E-V-5-8	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Keon Lee, Kong Aik Lee, Ko-tik Lee, Lin-shan Lee, Lukas Lee, Minjae	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-M-V-1-9 Tue-E-V-1-11 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-9 Wed-E-V-5-8 Wed-E-V-5-8	$215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 172 \\ 172 \\$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Kortik Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Min-Ii	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-19 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-5-8 Wed-E-V-5-8 Wed-M-V-3-9	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 172 \\ 126 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lukas Lee, Minjae Lee, Minjae Lee, Minjac	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-3-1 Wed-E-V-3-1	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 172 \\ 126 \\ 165 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lukas Lee, Minjae Lee, Min-Ji Lee, Min-Joong	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Tue-M-V-1-9 Tue-E-V-1-11 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-1	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 172 \\ 126 \\ 165 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koon Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Mun-Hak	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-1 Wed-E-V-3-7	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 172 \\ 126 \\ 165 \\ 228 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lukas Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Mun-Hak Lee, Sang-Hoon	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Tue-E-V-1-11 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-5	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jihyun Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Lukas Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Mun-Hak Lee, Sang-Hoon	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-A-V-5-9 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-5	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 122\\ 8\\ 151\\ 122\\ 126\\ 125\\ 122\\ 126\\ 122\\ 126\\ 122\\ 122\\ 122\\ 122$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lukas Lee, Minjae Lee, Minjae Lee, Min-Joong Lee, Mun-Hak Lee, Sang-Hoon	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-5 Wed-E-V-3-5 Wed-E-V-5-8 Wed-A-V-5-10	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Mun-Hak Lee, Sang-Hoon Lee, Seong-Whan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Tue-E-V-1-11 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-1 Fri-M-V-3-1 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-10	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 105\\ 238\\ 109\\ 166\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Junhyeok Lee, Koon Lee, Koon Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Mun-Hak Lee, Sang-Hoon	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-5-8 Wed-E-V-3-1 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koon Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Ji Lee, Mun-Hak Lee, Sang-Hoon Lee, Seong-Whan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-5 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-5-8 Wed-A-V-5-10 Wed-A-V-5-10	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 60\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 150\\ 150\\ 150\\ 150\\ 150\\ 150\\ 150\\ 1$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Lukas Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Sang-Hoon Lee, Seong-Whan Lee, Soo-Young	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-3 Wed-M-V-4-1 Fri-A-V-3-3 Wed-M-V-4-1 Tue-E-V-1-12 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-S&T-1-1	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ 156\\ \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Sang-Hoon Lee, Seong-Whan Lee, Soo-Young Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Tue-E-V-1-19 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-3-1 Wed-E-V-3-1 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-SST-11- Thu-M-V-1-S	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 156\\ 180\\ \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Sang-Hoon Lee, Seong-Whan Lee, Soo-Young Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Tue-E-V-1-11 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-3-1 Fri-M-V-3-7 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-1-5	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 156\\ 180\\ 0\end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Junhyeok Lee, Koon Lee, Koon Aik Lee, Koon Aik Lee, Koon Aik Lee, Koon Aik Lee, Koon Aik Lee, Koon Aik Lee, Lin-shan Lee, Min-Jae Lee, Min-Ji Lee, Min-Joong Lee, Mun-Hak Lee, Sang-Hoon Lee, Seong-Whan Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-3-8 Wed-E-V-3-8 Wed-E-V-3-1 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-1-5 Thu-M-V-5-7	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 126 \\ 165 \\ 228 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 150 \\ 190 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koon Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Lukas Lee, Minjae Lee, Min-Ji Lee, Min-Ji Lee, Min-Ji Lee, Sang-Hoon Lee, Seong-Whan Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-3-8 Wed-E-V-3-1 Fri-M-V-3-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-5-7 Fri-M-V-7-6	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 1$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Kong Aik Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Sang-Hoon Lee, Seong-Whan Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-3-8 Wed-E-V-3-8 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-70 Wed-A-V-5-70 Thu-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-5-7 Fri-M-V-7-6 Fri-M-V-7-6	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 156\\ 180\\ 190\\ 237\\ 253\\ \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Lin-shan Lee, Lin-shan Lee, Min-Jiae Lee, Min-Ji Lee, Min-Joong Lee, Mun-Hak Lee, Sang-Hoon Lee, Soo-Young Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Fri-A-V-5-10 F	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 126 \\ 165 \\ 228 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 156 \\ 180 \\ 190 \\ 237 \\ 253 \\ 172 \\ \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Koon Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Ji Lee, Sang-Hoon Lee, Soong-Whan Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-5 Wed-E-V-3-5 Wed-E-V-3-8 Wed-E-V-3-8 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-5-7 Fri-M-V-7-6 Fri-M-V-7-6 Fri-M-V-7-9 Thu-M-V-5-9 Thu-A-V-5-9 Thu-A-V-5-9	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 172 \\ 126 \\ 165 \\ 228 \\ 151 \\ 210 \\ 151 \\ 223 \\ 210 \\ 151 \\ 223 \\ 178 \\ 151 \\ 210 \\ 210 \\ 2$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Min-Joong Lee, Min-Joong Lee, Sang-Hoon Lee, Seong-Whan Lee, Tan Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-10 Wed-A-V-5-55 Thu-A-V-5-10 Wed-A-S&T-1-1 Wed-A-S&T-1-1 Thu-M-V-1-5 Thu-M-V-5-7 Fri-M-V-5-9 Thu-M-V-5-9 Thu-M-V-5-9 Thu-M-O-22 Tue-M-SS-1-1	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 172 \\ 126 \\ 165 \\ 228 \\ 151 \\ 121 \\ 151 \\ 210 \\ 156 \\ 180 \\ 190 \\ 237 \\ 253 \\ 178 \\ 77 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Min-Jaa Lee, Mun-Hak Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Leem, Seong-Gyun Lefter, Julia Leefter, Julia	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-3 Wed-E-V-3-3 Wed-E-V-3-5 Wed-E-V-3-5 Wed-E-V-3-1 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Thu-M-V-5-7 Fri-M-V-7-6 Fri-A-V-5-9 Thu-M-V-5-9 Thu-M-V-5-11 Tue-M-SS-1-1	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 27\\ 76\\ 253\\ 77\\ 62\end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Koon Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Min-Joong Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Lee, Tan	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-5 Wed-E-V-3-8 Wed-E-V-3-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-A-V-5-70 Wed-A-V-5-50 Wed-A-V-5-50 Wed-A-V-5-50 Wed-A-V-5-50 Wed-A-V-5-50 Wed-A-V-5-50 Wed-A-V-5-50 Wed-A-V-5-50 Wed-A-V-5-70 Wed-A-V-5-70 Fri-M-V-7-6 Fri-A-V-5-9 Thu-M-V-2-2 Tue-M-03-1	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ 156\\ 180\\ 190\\ 237\\ 223\\ 178\\ 77\\ 62\\ 233\\ 178\\ 77\\ 62\\ 180\\ 180\\ 180\\ 180\\ 180\\ 180\\ 180\\ 180$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Lin-shan Lee, Lin-shan Lee, Min-Ji Lee, Min-Ji Lee, Min-Joong Lee, Min-Joong Lee, Mun-Hak Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Lee, Tan Leeglaive, Simon Lehnen, Patrick	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-1 Wed-E-V-3-1 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Thu-A-V-5-10 Wed-A-V-5-10 Thu-A-V-5-10 Wed-A-V-5-10 Thu	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 156\\ 180\\ 190\\ 237\\ 253\\ 178\\ 8\\ 77\\ 62\\ 111 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Koon Lee, Koon Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lukas Lee, Min-Ji Lee, Min-Ji Lee, Min-Ji Lee, Min-Ji Lee, Sang-Hoon Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Lee, Tan Leglaive, Simon Lehnen, Patrick	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-3 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-3-5 Wed-E-V-3-7 Wed-A-V-5-8 Wed-A-V-5-8 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-5-7 Fri-M-V-7-6 Fri-A-V-5-9 Thu-M-O-2-2 Tue-M-O-3-1 Tue-M-O-3-1 Tue-E-V-4-7	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 1$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Jinkyu Lee, Koon Lee, Koon Aik Lee, Koon Aik Lee, Koon Aik Lee, Koon Aik Lee, Koon Aik Lee, Lin-shan Lee, Lin-shan Lee, Min-Jae Lee, Min-Ja Lee, Min-Joong Lee, Min-Joong Lee, Min-Joong Lee, Sang-Hoon Lee, Seong-Whan Lee, Tan Lee, Tan Lee, Tan Leelaive, Simon Lehnen, Patrick	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-A-V-5-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-1 Fri-M-V-3-7 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-1-5 Thu-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-1-5 Thu-M-V-5-7 Fri-A-V-5-9 Thu-M-O-2-2 Tue-M-SS-1-1 Tue-M-O-3-1 Tue-E-V-4-7 Fri-A-V-6-3	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 172 \\ 126 \\ 165 \\ 228 \\ 151 \\ 210 \\ 156 \\ 180 \\ 190 \\ 237 \\ 253 \\ 151 \\ 210 \\ 156 \\ 180 \\ 191 \\ 253 \\ 77 \\ 62 \\ 111 \\ 254 \\ 174$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Lin-shan Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Ji Lee, Sang-Hoon Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Lee, Tan Leelaive, Simon Lehnen, Patrick	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Thu-M-V-5-7 Fri-M-V-7-6 Fri-A-V-6-9 Thu-M-0-3-11 Tue-E-V-4-7 Fri-A-V-6-3 Wed-E-V-6-6	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 126 \\ 172 \\ 126 \\ 165 \\ 228 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 228 \\ 151 \\ 210 \\ 151 \\ 254 \\ 172 \\ 172 \\ 254 \\ 172 \\ 274 \\ 2$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Junhyeok Lee, Koon Lee, Koon Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Minjae Lee, Min-Ji Lee, Min-Joong Lee, Min-Joong Lee, Sang-Hoon Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Lee, Tan Leglaive, Simon Lehnen, Patrick Lei, Ming	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-5-8 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-7 Thu-M-V-5-7 Fri-M-V-7-6 Fri-A-V-5-9 Thu-M-V-7-7 Fri-M-V-7-6 Fri-A-V-6-3 Wed-E-V-4-7 Fri-A-V-6-3 Wed-E-V-4-2	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 228\\ 171\\ 254\\ 174\\ 207\\ \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Lin-shan Lee, Lin-shan Lee, Min-Ji Lee, Min-Ji Lee, Min-Ji Lee, Min-Joong Lee, Min-Joong Lee, Mun-Hak Lee, Sang-Hoon Lee, Seong-Whan Lee, Tan Lee, Tan Leelaive, Simon Leelaive, Simon Lei, Ming	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-3-7 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Thu-M-V-5-7 Fri-M-V-7-6 Fri-A-V-5-9 Thu-M-V-2-2 Tue-M-O-3-1 Tue-E-V-4-7 Fri-A-V-6-3 Wed-E-V-6-6 Thu-A-V-4-2 Fri-A-V-3-9	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 228\\ 151\\ 210\\ 151\\ 210\\ 156\\ 151\\ 210\\ 156\\ 180\\ 190\\ 237\\ 253\\ 178\\ 8\\ 77\\ 62\\ 111\\ 254\\ 174\\ 207\\ 248 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Koong Aik Lee, Lukas Lee, Lukas Lee, Min-Jae Lee, Min-Ji Lee, Min-Joong Lee, Sang-Hoon Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Lee, Tan Lee, Tan Leehnen, Patrick Lei, Ming	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-3 Wed-E-V-3-5 Wed-E-V-3-5 Wed-E-V-3-7 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-M-V-5-7 Fri-M-V-7-6 Fri-A-V-5-9 Thu-M-O-3-1 Tue-M-O-3-1 Tue-M-O-3-1 Tue-M-O-3-1 Tue-M-O-3-1 Tue-M-O-3-1 Tue-M-O-3-1 Tue-M-O-3-3 Wed-E-V-6-6 Thu-A-V-4-2 Fri-A-V-3-5	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 228\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 2$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Jinkyu Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Kong Aik Lee, Lin-shan Lee, Lin-shan Lee, Minjae Lee, Minjae Lee, Min-Jo ong Lee, Min-Joong Lee, Min-Joong Lee, Min-Joong Lee, Sang-Hoon Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Lee, Tan Lee, Simon Lefter, Iulia Lefter, Iulia Lehnen, Patrick Lei, Ming	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-5-7 Fri-M-V-7-6 Fri-A-V-5-9 Thu-M-O-2-2 Tue-M-O-3-1 Tue-E-V-4-7 Fri-A-V-6-3 Wed-E-V-6-6 Thu-A-V-5-1	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 156\\ 151\\ 210\\ 156\\ 180\\ 190\\ 237\\ 2253\\ 178\\ 77\\ 62\\ 111\\ 254\\ 174\\ 207\\ 248\\ 112\\ 112\\ 112\\ 112\\ 112\\ 112\\ 112\\ 11$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Min-Jae Lee, Min-Ji Lee, Min-Ji Lee, Sang-Hoon Lee, Soo-Young Lee, Tan Lee, Tan Leen, Seong-Gyun Lefter, Iulia Leelaive, Simon Lein, Ming Lei, Xiaoning Lei, Xia	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-5 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Thu-A-V-5-10 Wed-A-V-5-50 Thu-A-V-5-10 Wed-A-V-5-50 Thu-A-V-5-10 Wed-A-V-5-50 Thu-A-V-5-10 Wed-A-V-5-50 Thu-A-V-5-10 Wed-A-V-5-50 Thu-A-V-5-10 Wed-A-V-5-50 Thu-A-V-5-10 Wed-A-V-5-50 Thu-A-V-5-10 Wed-A-V-5-70 Fri-A-V-5-9 Thu-M-V-5-7 Fri-A-V-6-9 Thu-M-O-3-11 Tue-E-V-4-7 Fri-A-V-6-3 Wed-E-V-6-6 Thu-A-V-3-5 Tue-E-V-5-5	$\begin{array}{c} 215 \\ 74 \\ 151 \\ 165 \\ 247 \\ 128 \\ 251 \\ 64 \\ 105 \\ 238 \\ 109 \\ 166 \\ 172 \\ 126 \\ 165 \\ 228 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 210 \\ 151 \\ 228 \\ 177 \\ 248 \\ 112 \\ 228 \end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinhyun Lee, Jinhyuok Lee, Junhyeok Lee, Koong Aik Lee, Min-Ji Lee, Min-Ji Lee, Min-Ji Lee, Sang-Hoon Lee, Seong-Whan Lee, Soo-Young Lee, Tan Lee, Soo-Young Lee, Tan Lee, Simon Leelaive, Simon Leehnen, Patrick Lei, Ming Lei, Xiaoning Lei, Xin Lei, Xin Lei, Xin	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-3 Wed-E-V-3-3 Wed-E-V-3-5 Wed-E-V-3-5 Wed-E-V-3-7 Fri-M-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-M-V-5-5 Thu-M-V-5-7 Fri-M-V-7-6 Fri-A-V-5-9 Thu-M-V-5-9 Thu-M-V-5-11 Tue-M-O-3-1 Tue-M-O-3-1 Tue-M-O-3-1 Tue-M-O-3-1 Tue-K-V-5-1 Fri-A-V-6-6 Thu-A-V-5-1 Fri-A-V-6-6 Thu-A-V-5-1 Fri-A-V-6-5	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 165\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 174\\ 207\\ 248\\ 112\\ 228\\ 85\\ 85\end{array}$
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Min-Joong Lee, Min-Joong Lee, Min-Joong Lee, Min-Joong Lee, Sang-Hoon Lee, Seong-Whan Lee, Soo-Young Lee, Tan Lee, Seong-Gyun Lefter, Iulia Lei, Ming Lei, Ming Lei, Xiaoning Lei, Xia Lei, Xia Leightley, Daniel	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-9 Wed-E-V-3-9 Wed-E-V-3-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Wed-E-V-5-8 Thu-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-V-5-10 Wed-A-S&T-1-1 Thu-M-V-5-7 Fri-M-V-5-7 Fri-M-V-5-9 Thu-M-V-5-7 Fri-A-V-5-9 Thu-M-V-5-7 Fri-A-V-5-9 Thu-M-V-5-7 Fri-A-V-5-9 Thu-M-V-2-2 Tue-M-S3-11 Tue-E-V-4-7 Fri-A-V-6-3 Wed-E-V-6-6 Thu-A-V-3-5 Tue-A-V-	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 228\\ 151\\ 210\\ 151\\ 228\\ 177\\ 62\\ 111\\ 254\\ 174\\ 207\\ 248\\ 112\\ 228\\ 88\\ 88\\ 88\\ 88\\ 88\\ 88\\ 88\\ 88\\ $
Lee, Jaesong Lee, Jaeyun Lee, Ji-Hyun Lee, Jinkyu Lee, Jinkyu Lee, Junhyeok Lee, Koong Aik Lee, Min-Jaong Lee, Min-Joong Lee, Min-Joong Lee, Sang-Hoon Lee, Seong-Whan Lee, Soo-Young Lee, Tan Lee, Soo-Young Lee, Tan Lee, Seong-Gyun Leefure, Iulia Leehnen, Patrick Lei, Ming Lei, Xiaoning Lei, Xiaoning Lei, Xin Leightley, Daniel Lee Moine, Clément	Thu-A-SS-1-7 Tue-M-V-6-9 Wed-A-V-5-5 Wed-E-V-3-1 Fri-A-V-3-3 Wed-M-V-4-1 Fri-A-V-5-1 Tue-E-V-1-12 Fri-M-V-7-9 Tue-E-V-1-12 Fri-M-V-7-9 Wed-E-V-3-3 Wed-E-V-3-5 Wed-E-V-3-5 Wed-E-V-3-7 Wed-E-V-3-1 Wed-E-V-3-1 Wed-E-V-3-1 Wed-E-V-3-1 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-V-5-5 Thu-A-V-5-10 Wed-A-SS-1-1 Thu-M-V-5-7 Fri-A-V-5-9 Thu-M-V-5-9 Thu-M-V-5-9 Thu-M-V-5-9 Thu-M-V-5-9 Thu-M-V-5-1 Tue-E-V-4-7 Fri-A-V-6-3 Wed-E-V-5-1 Tue-E-V-4-7 Fri-A-V-6-3 Wed-E-V-5-1 Tue-E-V-5-1 Fri-A-V-2-6 Thu-A-V-2-6 Thu-A-V-2-6 Thu-M-O-2-1	$\begin{array}{c} 215\\ 74\\ 151\\ 165\\ 247\\ 128\\ 251\\ 64\\ 105\\ 238\\ 109\\ 166\\ 172\\ 172\\ 126\\ 228\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 210\\ 151\\ 228\\ 178\\ 77\\ 2253\\ 178\\ 112\\ 248\\ 112\\ 248\\ 112\\ 228\\ 85\\ 178\end{array}$

Leong, C.W.	Fri-M-O-1-2 219
Leow, Chee Siang	Tue-M-V-6-7 74
	Wed-E-V-1-4 161
Le Roux, Jonathan	Tue-A-V-1-10 83
, 0	Tue-A-V-4-5 89
	Wed-M-V-6-13 135
	Wed-A-V-3-12 147
Le-The Quoc-Nam	Wed-M-V-6-4 133
Leung Teun-Vat	Thu $A - V - 4 - 7 = 208$
Loung Wai Vim	Fr: M V 1 10 224
Leuiig, wai-Kiiii	Fri-IM-V-1-10 224
Levis, John	Wed-A-V-1-4 141
Levitan, Sarah Ita	Tue-E-O-3-4 103
Levow, Gina-Anne	Wed-A-V-2-5 143
Leykum, Hannah	Tue-E-O-2-1 101
Li, Andong	Tue-M-V-3-6 67
	Wed-E-SS-1-2 175
Li, Bo	Wed-M-V-6-4 133
	Fri-M-V-3-8 229
Li. Bohan	Fri-A-V-5-6 252
Li, Chenda	Thu-M-V-2-11 184
,	Thu-M-V-3-3 185
Li Chengfei	Tue-E-V-5-6 113
Li, Cuangzhi	$Fri A V_5 4 252$
Li, Oudingzini	
Li, Hai	Tue-A-V-0-0 95
	Tue-M-V-1-9 04
	Tue-M-V-3-2 00
	Tue-A-V-4-9 90
	Tue-A-V-6-2 93
	Tue-A-SS-2-5 97
	1ue-E-V-1-12 105
	Tue-E-V-6-11 116
	Wed-M-O-3-2 121
	Thu-A-V-3-12 206
	Fri-M-O-3-4 222
	Fri-A-V-5-2 252
I. Hawa	Fri-A-V-6-4 254
	Inu-A-V-I-2 200
Li, Hao	wea-E-SS-1-6 170
Li, Haoqi	Thu-A-V-1-6 201
Li, Huichung	Fri-A-V-6-10 233
Li, Jidiidiig	Fri-A-V-2-9 247
Li, Jianchen	104 104
Li, Jillieng	wea-A-V-2-3 143
LI, JIIIg	
Li Jinghei	Tue-A-V-4-10 231
	Fri-Δ-V-5-7 252
Li. IingWei	Wed-A-V-4-6 149
Li. lingvu	Thu-M-V-1-5 180
Li, Jinyu	Tue-A-V-4-10 90
,	Tue-E-V-5-9 114
	Wed-M-V-6-2 133
	Wed-M-V-6-5 134
	Wed-A-V-2-2 143
	Wed-E-V-3-14 167
	Thu-M-V-3-1 184
	Thu-M-V-3-9 186
	Thu-A-V-2-3 203
	Thu-A-V-2-10 204
Li, Juan	Wed-A-V-4-6 149
Li, Juncheng	Wed-E-V-1-10 162
Li, Junfeng	Tue-E-V-2-3 106
Li, Junjie	Fri-M-V-4-4 230
Li, Ke	Thu-A-V-6-13 214
Li. Kun	Fri-M-V-1-10 224
Li, Lei	Wed-A-V-6-8 154
Li, Li Erran	Wed-A-V-2-13 145
Li, Lin	Tue-M-V-1-8 64
-	Thu-M-SS-1-1 193
	Thu-M-SS-1-6 194
	Thu-A-SS-1-8 215
	Fri-A-V-4-9 251
	Fri-A-V-4-10 251
Li, Lin	Tue-E-V-4-4 111
Li, Meng	Wed-M-V-1-5 122
Li, Ming	Tue-M-V-1-5 63
	Tue-E-V-1-2 103
	Wed-A-SS-1-5 156
	Wed-E-V-6-4 173
	Fri-M-V-6-10 235
L1, Ping	Fri-M-V-7-7 238
Li, Qiujia	Fri-M-V-3-8 229
	Fri-M-V-3-9 229
Li, Ruichen	Fri-A-V-2-3 246
Li, Ruirui	Tue-E-V-1-10 105
Li Carah D	Fri-A-V-4-2 250
ы, баган К	Tue-A-V-2-7 86

INTERSPEECH 2021 269 BRNO | CZECHIA

Li, Shang-Wen	Tue-E-V-3-9	109	Ling, Zhen-Hua	Tue-A-V-6-3	93	Loide,
Li Shang	Thu-M-V-7-7	193		Fri-M-V-4-6	230	Long, 1
Li, Sheng	Thu-M-SS-1-4	194	Linnhoff-Popien, C	Tue-M-SS-1-7	78	Loose,
Li, Song	Thu-A-SS-1-8	215	Lippa, Sara M	Tue-M-O-2-1	60	Łopatk
	Fri-A-V-4-9	251 251	Liscombe, Jackson	Wed-A-V-1-7 Fri-A-SS-2-2	142 257	Lopes,
Li, Tao	Tue-A-V-6-6	93	Liss, Julie	Wed-A-O-3-4	140	LOPCZ
Li, Tingle	Tue-E-V-6-1	114	Liu, Alexander H	Thu-A-SS-1-4	214	Tand
Ll, wei	Wed-M-V-6-4 Fri-A-V-3-8	248	Liu, Andy T	Wed-E-V-2-8	164	Lora, C
Li, Wubo	Wed-M-V-2-4	124	Liu, Bing	Tue-A-V-4-10	90	Loren
Li, Xiang	Fri-A-V-5-7	252	Liu, Ching-Feng	Tue-E-V-6-2	114	Lou V
	Wed-M-V-2-4	124	Liu, Da-Kong	Thu-M-V-3-7	109	Lou, A
	Thu-A-V-6-5	212	Liu, Haohe	Tue-M-V-3-7	67	T CI
Li, Xiangrui	Thu-M-V-3-4	185 67	Liu, Hexin	Wed-M-V-1-1	122	Lu, Ch
	Wed-E-SS-1-2	175		Fri-A-V-2-2	245	Lu, Ha
T · T/2 0 ·	Fri-A-SS-1-5	256	Liu, Jiaxing	Fri-A-V-2-6	246	Lu, Hu
Li, Xiaofei	Tue-A-V-3-3	87 103	Liu lie	Fri-A-V-2-10	247	Lu Iin
Li, Xinhui	Thu-A-V-5-11	211	Liu, Jing	Wed-A-V-2-4	143	Lu, Li
Li, Xinjian	Tue-M-V-5-2	70	Liu, Jinglin	Wed-E-V-6-6	174	Lu, Lia
	Wed-E-V-1-10	162 211	Liu, Jinjiang	Wed-M-SS-1-5	136 187	
Li, Xiyun	Tue-E-V-2-4	106	Liu, Li	Tue-A-V-2-5	85	Lu, Xu
Li, Xu	Thu-A-SS-1-13	216	T' T'T	Thu-M-V-2-4	183	Lu Vo
	Fri-M-V-7-6 Fri-M-V-7-12	237	Liu, Li-Juan	Tue-A-V-6-3	93 82	Lu, re Lu, Yit
Li, Xuewei	Tue-A-V-2-5	85	Liu, Liii	Wed-M-V-1-3	122	Lu, Yiz
Li Voving	Thu-M-V-2-4	183	Liu, Meng	Tue-E-V-1-11	105	Lu, Zh
Li, Yinghao Aaron	Tue-E-V-1-4	115	Liu, Mini	Wed-A-SS-1-5 Wed-A-V-2-10	130	Lüdtke
Li, Yuepeng	Wed-M-SS-1-2	136	Liu, Qiao	Thu-A-V-3-10	206	Ludus
Li, Zezhong	Thu-M-V-7-5	193	Liu, Qing-Feng	Fri-M-V-4-6	230	Lulich
LI, ZHENg	Thu-M-SS-1-1	193	Liu, Rongliang	Wed-A-V-4-10	149	Lunda Luo, C
	Thu-M-SS-1-6	194	Liu, Rui	Fri-A-V-5-2	252	Luo, H
Li, Zhengyang	Thu-M-O-1-3	177	Liu, Shansong	Fri-A-SS-2-4	257	Luo, Ji
ы, лии	Fri-M-V-1-9	224		Fri-A-SS-2-9	259	Luo, L
Lian, Jiachen	Fri-A-V-4-11	251	Liu, Shichao	Thu-M-V-5-5	189	Luo, Y
Lian, Jie	Tue-E-V-6-9	116 146	Liu, Shujie	Thu-M-V-3-1	184 186	
Liang, Qiao	Wed-M-V-6-4	133		Thu-A-V-2-10	204	
- -	Fri-M-V-6-3	234	Liu, Shuo	Fri-M-V-5-5	232	Luo, Y
Liang, Xia	<u>Tue-M-V-6-1</u> Thu-Δ-SS-2-3	73 217	Liu, Songxiang	Fri-A-SS-2-8 Fri-A-V-5-6	258 252	Luong
Liao, Dexin	Thu-A-SS-1-8	215	Liu, Wenzhe	Tue-M-V-3-6	67	Lupea,
Liboratora Christophan	Fri-A-V-4-10	251	Liu Viengu	Wed-E-SS-1-2	175	Luu, C
Liberman. Mark	Thu-A-V-4-6	94 208	Liu, Xubo	Wed-A-V-2-3	145	Luz, Si
Liebling, Daniel J	Tue-A-V-5-3	91	Liu, Xuefei	Tue-M-V-5-1	70	Lv, Shu
Likhomananko Tatiana	Fri-A-V-4-6	250 72	Liu Yunving	Wed-E-V-2-8	164 115	
Likiloinanciiko, Tatiana	Tue-A-V-4-4	89	Liu, Auriying	Thu-M-V-1-10	182	Lv, Zhi
I iller. Is see	Tue-A-V-4-8	90		Thu-A-SS-1-13	216	Lyu, K
Lilley, Jason	Fri-A-SS-2-10	259		Fri-M-V-7-12 Fri-A-SS-2-4	238 257	Lyu, SI
Lim, Dan	Wed-A-V-5-7	151		Fri-A-SS-2-6	258	Dyu, I
Lin, Binghuai	Tue-A-V-5-6	91 131		Fri-A-SS-2-7	258 258	IM
	Wed-A-V-3-10	146		Fri-A-SS-2-9	259	M., Gu
	Fri-M-V-1-6	223	Liu, Yang	Tue-M-V-4-8	70 64	Ma, Gu Ma, Ha
	Fri-M-V-1-7 Fri-M-V-1-9	223		Thu-M-SS-1-6	04 194	1.100,110
	Fri-A-V-1-5	243		Fri-A-V-4-9	251	Ma, Ju
	Fri-A-V-1-7	244	Liu, Yanqing	Wed-A-V-2-2	143	Ma, Ka
Lin, Gang-Xuan	Wed-M-V-4-2	128	Liu, Yichen	Tue-E-V-6-1	114	Ma, Lo
Lin, Guan-Ting	Tue-E-V-3-9	109	Liu, Yi Chieh	Thu-M-V-4-1	187	
Lin, Haitao Lin Haopeng	Fri-A-V-6-2	254 189	Liu, Yi-Fen	Tue-A-V-2-1	84 116	
Lin, Hung-Pang	Fri-M-SS-1-4	240		Wed-M-V-3-4	126	Ma, Lu
Lin, Jheng-hao	Tue-A-V-6-7	94	Liu, Yulan	Tue-A-SS-1-9	96 160	Ma, MI
Lin, Ju Lin, Pingvuan	wed-м-V-4-12 Тue-E-V-6-9	130 116	Liu, Zhang	Tue-E-V-2-9	107	Ma, Pi
Lin, Qingjian	Tue-E-V-1-2	103	Liu, Zhengchen	Wed-M-V-3-5	127	Ma, Sh
Lin, Shilun	Thu-A-V-5-11	211	Liu, Zhengxi	Wed-A-V-5-10	152 225	Ma, Yo
Lin, Shouleng		149 94	Liu, Zitao	Thu-A-V-1-2	200	Ma, Ze
-,	Tue-E-V-3-9	109	Livescu, Karen	Thu-Survey	197	
Lin Yue	Wed-E-V-6-9	174 126	Lleida, Eduardo	Wed-E-O-1-3	158 241	
Lin, Zhouhan	Wed-A-V-2-3	143	Lo, Justin J.H.	Tue-M-V-7-9	76	Ma, Zh
Ling, Hefei	Fri-M-V-7-7	238	Locqueville, Grégoire	Tue-A-O-2-1	81	maas,
			Lonrenz, 11mo	Thu-M-O-1-3	177	

oide, Markus	Fri-A-SS-1-1	255
ong, Siqu	Fri-A-V-6-10	255
ong, Yanhua	Wed-M-SS-1-4	136
oose, Zdeněk	Wed-A-S&T-1-2	156
opatka, Kuba	Fri-A-O-2-1	242
opes, José	Fri-A-V-1-12	245
opez Moreno, Ignacio .	Tue-M-V-6-3	73
1	Tue-E-V-1-6	104
	Fri-A-V-4-4	250
ord. Catherine	Thu-M-V-1-2	180
orenzo-Trueba. Jaime .	Thu-A-V-5-7	210
, ,	Fri-M-V-1-2	222
	Fri-A-V-1-1	243
ou, Xin	Wed-A-V-4-6	149
oweimi, Erfan	Tue-M-V-5-3	71
	Wed-E-V-3-3	165
u, Chung-Li	Tue-M-V-1-3	63
u, Chun-Shien	Wed-M-V-4-2	128
u, Han	Thu-A-V-2-4	203
u, Hui	Thu-A-SS-1-13	216
	Fri-M-V-7-12	238
u, Jing	Wed-E-SS-1-4	176
u, Li	Thu-A-V-5-11	211
u, Liang	Wed-M-V-6-5	134
	Wed-E-V-3-14	167
	Thu-A-V-2-3	203
u, Xugang	Tue-M-V-3-8	68
	Tue-M-V-3-9	68
u, Yen-Ju	Wed-M-V-4-2	128
u, Yiting	Tue-M-V-2-6	65
u, Yizhou	Tue-E-V-5-2	112
u, Zhiyun	Thu-A-V-2-8	204
uckenbaugh, Jarrod	Fri-A-O-1-4	241
üdtke, Ulrike	Fri-M-O-1-3	219
udusan, Bogdan	Tue-E-O-2-4	101
ulich, Steven M.	Tue-E-SS-1-5	117
umban Tobing, Patrick	Wed-A-V-5-9	152
uo, Chong	Thu-A-V-5-3	209
uo, Hongyin	Thu-M-V-7-7	193
uo, Jian	Tue-E-V-3-4	108
uo, Liuping	Fri-A-V-4-7	250
	Wed-E-SS-1-2	1/5
uo, Y1	Thu-M-V-3-3	185
	<u>Ihu-M-V-3-10</u>	186
	Thu-A-V-3-2	205
Vivu		105
uo, 11yu	1 ue-E-V-2-1	140
uo, ziidojie	wea-A-V-4-11	149
uong, Mann	Tue-A-V-6-10	102
upea, Daviu		105
uu, Clidu		216
		245
y Shubo	Fri-A-V-Z-I Wod E SS 1 E	176
v, 511ub0		212
	Fri-A-SS-1-2	256
v Zhigiang	Fri-M-SS-1-5	240
ni Kexin	Fri-M-V-1-4	223
vii Shao-ren	Wed-E-V-4-8	169
vii. Yilan	Thu-A-1/-3-10	206
, u, 111111		-00
1		
Cumunath Daddy		222
., Gurunani Keday	Fri-M-V-5-6	232
	<u>1ue-M-V-5-9</u>	12

,		
a, Guodong	Tue-M-V-5-9	72
a, Haoxin	Tue-A-SS-1-7	95
	Wed-M-V-4-5	128
ı, Jun	Wed-M-V-3-5	127
	Fri-M-V-4-4	230
ı, Kaili	Thu-M-V-2-3	183
a, Long	Tue-A-V-4-1	88
_	Wed-M-V-1-4	122
	Fri-A-V-1-5	243
	Fri-A-V-3-11	249
ı, Lu	Tue-E-V-5-6	113
a, Mingbo	Wed-A-V-6-1	152
a, Ning	Tue-A-V-3-8	88
, Pingchuan	Thu-M-V-2-9	184
, Shiqian	Thu-A-V-3-11	206
a, Yong	Wed-A-SS-1-5	156
, Youxuan	Fri-M-V-5-3	232
, Zejun	Tue-M-V-6-1	73
	Wed-M-V-6-8	134
	Wed-E-V-3-9	166
	Thu-M-V-5-5	189
a, Zhizhong	Tue-M-V-3-11	68
as, Roland	Tue-A-V-4-2	89
-	Tue-A-SS-1-9	96

MacDonald, Robert L	Fri-A-SS-2-1	257	McCre
Macháčak Dominik	Fri-A-SS-2-12	259	McDe
Maciejewski Matthew	wed-E-O-2-2 Thu-Δ-V-3-9	206	McGo
Macoskey, Jonathan	Wed-F-V-3-2	165	MCOR
	Fri-A-V-3-7	248	McKe
MacWhinney, Brian	Thu-A-SS-2-1	216	McKe
Madhavi, Maulik	Tue-A-SS-2-5	97	McLo
Madilani Cuilanth	Fri-A-V-6-4	254	
Maulkeri, Srikanun	I NU-M-O-1-6	230	McNa
	Fri-A-O-1-3	241	Mdha
Madsen, Julia	Wed-E-V-6-3	173	Md. S
Maekaku, Takashi	Wed-M-V-2-8	125	Medd
Magimai-Doss, Mathew .	Tue-M-O-2-2	61	Mede
	Tue-M-O-3-3	62 140	Medir
	Wed-A-O-3-5	257	Meew
Mahadeokar, Jay	Wed-M-V-6-3	133	Mehr
, 3,	Wed-A-V-3-1	145	Mehri
	Wed-A-V-3-14	147	Mehta
	Fri-A-V-3-6	248	Meign
Mahoor, Mohammad H.	Wed-E-V-6-3	173	Meiss
Manr, Iristan J.	Wed-A-O-3-4	140	Meist
Malei, Anureas	Wed-F-O-3-2	159	Meisu Máli
Mainar, Pablo	Wed-E-O-3-4	160	Melor
Majumdar, Somshubra .	Wed-M-O-1-4	119	Meng
Mak, Man-Wai	Tue-M-V-1-6	63	Meng
Makarov, Rostislav	Wed-A-SS-1-1	155	
Makino, Shoji	Wed-A-V-1-3	141	
Makisiiiiia, Naoki	Tue-M-V-5-5	91	
	Wed-E-V-3-13	167	
	Fri-M-V-3-6	228	
	Fri-A-V-1-11	245	
Malek, Jiri	Tue-A-S&T-1-7	99	
Malik John	Wed-M-O-3-3	145	
Mallela Ihansi	Thu-M-V-1-11	182	Meng
Mallidi. Sri Harish	Tue-A-V-4-2	89	Meng
Mallol-Ragolta, Adria	Tue-A-SS-2-9	98	Meng
Mamontov, Danila	Tue-M-SS-1-11	79	0
Mane, Vishal	Thu-M-S&T-1-7	197	
Maniati, Georgia	Wed-M-V-3-2	126	
Manohar N	Tue-M-V-5-5	71 85	
Manohara, K.	Wed-E-V-5-9	172	Mens
Mansbach, Noa	Fri-M-V-5-1	231	Merbo
Mansfield, Courtney	Wed-A-V-2-5	143	
Mao, Hongda	Fri-A-V-4-2	250	Merky
Marcinek, Luboš	Thu-A-O-2-3	200	Merte
Marko, Alexandra	Wed-A-O-3-6	140	
Markuson Sara	Wed-M-V-3-2	142	Mesga
Marlo. Michael R.	Tue-M-V-5-2	70	8
Maron, Haggai	Tue-E-V-2-6	106	
Marras, Mirko	Wed-A-O-2-3	139	
Martin, Rainer	Tue-A-SS-1-5	95	Maala
Martin, Vincent P	Wed-E-V-2-1	163	Mesng
Marver Picard	Tuo-E-O-1-1	100	Metze
	Tue-E-O-1-2	$100 \\ 100$	
Masaki, Shungo	Wed-A-V-4-9	149	
Mascolo, Cecilia	Tue-M-V-6-8	74	Meun
	Tue-M-SS-1-1	77	Meyer
Maskall Simon	I hu-M-V-I-9	101	Miao,
Mast T Douglas	Tue-A-V-2-7	86	Micha
Masumura. Rvo	Tue-M-V-6-6	74	Miche
, ,	Tue-A-V-5-5	91	
	Wed-M-V-6-6	134	Migue
	Wed-E-V-3-13	167	M;11
	Fri-M-V-3-6	228 245	Mikol
Matassoni. Marco	Fri-M-0-1-2	219	Milan
	Fri-M-O-1-4	220	Milde
Matcham, Faith	Tue-A-V-2-6	85	·
Mateju, Lukas	Wed-M-O-3-3	121	Milhé
Mathad, Vikram C	Wed-A-O-3-4	140	Millin
Matoušek Jindřich	۲ue-A-SS-T-4 Fri-۸-S&T-1-2	95 260	Millm
Matrouf, Driss	Wed-A-0-2-6	139	Miner
Matsui, Toshie	Thu-M-V-6-9	192	Minor
Matthews, Iain	Thu-A-O-1-4	199	Minke
Mau, Ted	Tue-A-V-2-10	86	Minki
Mavandadi, Sepand	Wed-A-V-3-4	145	Mira,
wazumuer, Mark	Eri-M-V-6-5	234	

McCree, Alan	Thu-A-V-4-9	208
McDermott, Josh H	Wed-E-V-5-2	171
McGowan, Ross	Tue-A-V-5-2	91
McGraw Ian	Fri-M-V-3-9	229
	Eri.M.V.6.3	234
McKoon Cristina		101
McKean, Clistina		101
MCKenna, Joseph P	Fri-A-V-6-9	200
McLoughlin, lan	Tue-A-V-1-3	82
	Wed-M-V-1-3	122
	Fri-A-V-3-5	248
McNamara. Ouinten	Thu-A-V-2-9	204
Mdhaffar, Salima	Wed-M-O-1-5	119
Md Salleh Siti Umairah	Fri-Δ-S&T-1-6	261
Medda Ciacomo	Wod-A-0-2-3	130
Medennikov Ivan		227
Medina Calvadar		100
Medina, Salvador	<u>I nu-A-O-I-4</u>	199
Meewis, Floor	I ue-M-SS-I-8	/8
Meghawat, Aaksha	Wed-A-V-2-12	144
Mehmood, Hadi	Thu-A-V-6-7	212
Mehrish, Ambuj	Tue-A-S&T-1-5	-99
Mehta, Rupesh R	Wed-M-V-6-2	133
Meignier Sylvain	Τμο-Δ-S&T-1-5	99
Meissner Wassilios C	Tuo-M-0-2-3	61
Meister Finar	Eri_M_V_2_11	226
Moiston Luo		220
		100
Mell, Adrien	Wed-E-V-4-3	108
Meloni, Giacomo	Wed-A-O-2-3	139
Meng, Hao	Fri-M-V-5-5	232
Meng, Helen	Tue-A-V-6-9	94
	Tue-E <u>-V-6-5</u>	115
	Thu-M-V-1-10	182
	Thu-A-SS-1-13	216
	Fri-M-V-1-10	224
	Fri-M-V-7-12	238
	Eri-A-1/-5-7	250
		252
	Fri-A-55-2-4	201
	Fri-A-SS-2-6	200
	Fri-A-55-2-8	258
	Fri-A-SS-2-9	259
Meng, Li	Thu-A-V-5-11	211
Meng, Yixiong	Tue-A-SS-1-9	96
Meng, Zhong	Tue-A-V-4-10	90
	Wed-E-V-3-14	167
	Thu-A-V-2-2	202
	Thu-A-V-2-3	203
	Thu-Δ-V-2-10	204
		2/2
Monchikova Alla		1/1
Mellslikova, Alla	wea-A-V-1-5	141
Merbolat, Allare	Wed-A-V-3-3	145
	Thu-M-O-3-3	179
Merkx, Danny	Fri-A-O-2-3	242
Mertes, Silvan	Tue-M-SS-1-12	79
	Thu-M-V-5-8	190
	Fri-M-O-2-2	220
Mesgarani, Nima	Tue-E-V-6-6	115
0	Thu-M-V-3-3	185
	Thu-M-V-3-10	186
	Thu-Δ-V-3-2	205
	Thu.A.V.3.6	205
Meshai Kourosh	Eri-M-V-1-5	223
Motzo Elorian		02
Metze, Monali	Tue-A-V-3-6	112
	1ue-E-V-4-10	162
	Wed-E-V-1-10	102
	Wed-E-V-I-I2	162
Meunier, Christine	Fri-A-V-4-8	251
Meyer, Josh	Fri-M-V-6-5	234
Miao, Chenfeng	Wed-M-V-3-5	127
Miao, Xiaoxiao	Fri-A-V-4-7	250
Michael, Thilo	Wed-E-O-3-1	159
Michel. Wilfried		145
	Wed-A-V-3-3	
	Wed-A-V-3-3 Thu-M-O-1-5	177
Miguel Antonio	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3	177
Miguel, Antonio	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3	177 158 241
Miguel, Antonio	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1	177 158 241 78
Miguel, Antonio Mikolajczyk, Krystian	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Eri Kourget	177 158 241 78
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote	177 158 241 78 241
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-M-O-1-1	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 178 \\ 219 \\ 178 \\ 178 \\ 178 \\ 178 \\ 178 \\ 178 \\ 188 \\$
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 196$
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5	177 158 241 78 241 219 196 197
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S.	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Tue-M-O-2-3	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 196 \\ 197 \\ 61$
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S. Milling, Manuel	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Tue-M-O-2-3 Thu-M-V-5-3	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 196 \\ 197 \\ 61 \\ 190 \\$
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S. Milling, Manuel Millman, Rebecca	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Tue-M-O-2-3 Thu-M-V-5-8 Thu-A-O-2-3	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 196 \\ 197 \\ 61 \\ 190 \\ 200 \\$
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S. Milling, Manuel Millman, Rebecca Minematsu Nobuaki	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Tue-M-O-2-3 Thu-A-O-2-3 Fri-A-V-1-6	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 196 \\ 197 \\ 61 \\ 190 \\ 200 \\ 244 \\$
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S. Milling, Manuel Milliman, Rebecca Minematsu, Nobuaki Miner Jennifor Ann	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Tue-M-O-2-3 Thu-M-V-5-8 Thu-A-O-2-3 Fri-A-V-1-6 Tue-E-SS-1-6	177 158 241 78 241 219 196 197 61 190 200 244 118
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S. Milling, Manuel Millman, Rebecca Minematsu, Nobuaki Miner, Jennifer Ann Mingote Victoria	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Tue-M-O-2-3 Thu-M-V-5-8 Thu-A-O-2-3 Fri-A-V-1-6 Tue-E-SS-1-6 Wed-E-O-1-2	177 158 241 78 241 219 196 197 61 190 200 244 118 158
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S. Milling, Manuel Milling, Manuel Minematsu, Nobuaki Minematsu, Nobuaki Minegote, Victoria Minkar Wolfgarg	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Tue-M-O-2-3 Thu-M-V-5-8 Thu-A-O-2-3 Fri-A-V-1-6 Tue-E-SS-1-6 Wed-E-O-1-3 Tuo-M-52-1	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 196 \\ 197 \\ 61 \\ 190 \\ 200 \\ 244 \\ 118 \\ 158 \\ 70 \\ 70 \\ 10$
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S. Milling, Manuel Milman, Rebecca Minematsu, Nobuaki Miner, Jennifer Ann Mingote, Victoria Mingote, Victoria	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-1 Thu-M-S&T-1-5 Tue-M-O2-3 Thu-A-O-2-3 Thu-A-O-2-3 Fri-A-V-1-6 Tue-E-SS-1-6 Wed-E-O-1-3 Tue-M-SS-1-11	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 196 \\ 197 \\ 61 \\ 190 \\ 200 \\ 244 \\ 118 \\ 158 \\ 79 \\ 116 \\ 1$
Miguel, Antonio Mikolajczyk, Krystian Mikolov, Tomáš Milanović, Ljubomir Milde, Benjamin Milhé de Saint Victor, S. Milling, Manuel Milling, Manuel Milman, Rebecca Minematsu, Nobuaki Miner, Jennifer Ann Mingote, Victoria Minker, Wolfgang Minkin, Fedor	Wed-A-V-3-3 Thu-M-O-1-5 Wed-E-O-1-3 Fri-A-O-1-1 Tue-M-SS-1-9 Fri-Keynote Fri-M-O-1-1 Thu-M-S&T-1-5 Tue-M-O-2-3 Thu-M-V-5-8 Thu-A-V-1-6 Tue-E-SS-1-6 Wed-E-O-1-3 Tue-M-SS-1-11 Wed-M-O-1-1	$177 \\ 158 \\ 241 \\ 78 \\ 241 \\ 219 \\ 196 \\ 197 \\ 61 \\ 190 \\ 200 \\ 244 \\ 118 \\ 158 \\ 79 \\ 110 \\ 110 \\ 1$

Mirheidari, Bahman		101
	Thu-M-V-1-7	181
	Thu-A-SS-2-7	218
Mirzaei, Maryam Sadat .	Fri-M-V-1-5	223
Mirzakhmetov, Almas	Wed-E-V-6-10	174
Misra. A.	Fri-M-O-1-2	219
Misra, Ananya	Tue-A-V-4-6	89
Mitra Vikramiit	Fri-Δ-SS-2-15	260
Mitrofanov Anton		227
Mittag Cabriel		110
Mittag, Gabriel	wed-A-V-4-4	140
	Wed-E-O-3-1	159
Mittal, Ashish	Wed-M-V-2-1	123
	Wed-E-V-1-7	161
Mittal, Deepak	Tue-A-SS-2-11	98
Miyazaki, Koichi	Tue-A-V-1-13	84
Mizumachi. Mitsunori	Fri-A-S&T-1-1	260
Mizuta, Kazuki	Wed-A-V-5-4	151
Möbius Bernd	Thu-M-V-6-5	191
hiobius, being minim	Thu-M-V-6-7	191
		224
		221
		223
Maall Diaman	Fri-M-V-6-1	234
Moell, Birger	Fri-A-V-1-12	245
Mohamed, Abdelrahman	Tue-E-V-3-9	109
	Wed-E-V-1-3	161
	Thu-A-V-5-6	210
Mohamed, Thahir	Fri-M-V-6-2	234
Mohammadamini, M	Wed-A-O-2-6	139
Mohan, Devang S. Ram .	Fri-M-O-2-3	221
	Fri-M-O-2-4	221
Mohri Mehrvar	Tue-A-SS-1-4	95
Moinet Alevis	Thu-A-V-5-7	210
Mole Deggy	Eri.M.V.2.5	225
Molthogi Defuee		223
Moknosi, Refude	Inu-A-V-3-10	200
Moller, Sebastian	Wed-A-V-4-4	148
	Wed-E-O-3-1	159
	Wed-E-V-6-7	174
Moniz, Joel Ruben A	Tue-E-V-4-4	111
Moore, Roger K	Tue-A-V-6-1	92
Morais, Edmilson	Fri-A-V-6-6	254
Mordido, Goncalo	Fri-A-V-3-1	247
Moreno, Pedro I.	Tue-A-V-4-7	90
	Wed-M-V-6-11	135
	Wed-F-V-3-6	166
	Fri-Δ-SS-2-11	259
Morfi Veronica	Τιιο-Δ-\/-1-9	83
Morgenshtern VI		62
Morgensniterin, v.n	Tue-MP0-3-2	110
Mori Uirolzi		
Mori, Hiroki	Tue-E-V-3-12	110
Mori, Hiroki Morimoto, Yoshiya	Tue-E-V-3-12 Wed-A-V-4-9	110
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9	110 149 192
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1	110 149 192 260
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5	110 149 192 260 89
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13	110 149 192 260 89 135
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12	110 149 192 260 89 135 147
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10	110 149 192 260 89 135 147 107
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12	110 149 192 260 89 135 147 107 132
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-6-12	110 149 192 260 89 135 147 107 132 134
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-6-6 Fri-M-V-3-6	110 149 192 260 89 135 147 107 132 134 228
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-6-6 Fri-M-V-3-6 Tue-M-V-6-10	110 149 192 260 89 135 147 107 132 134 228 74
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-6-10 Fri-M-V-3-6 Tue-E-V-1-9	110 149 192 260 89 135 147 107 132 134 228 74 104
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L	Tue-E-V-3-12 Wed-A-V-4-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2	110 149 192 260 89 135 147 107 132 134 228 74 104 124
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12	110 149 192 260 89 135 147 107 132 134 228 74 104 124 216
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-10 Tue-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-2-10	110 149 192 260 89 135 147 107 132 134 228 74 104 124 216 218
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris Ethan	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-M-V-6-10 Tue-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12	110 149 192 260 89 135 147 107 132 134 228 74 104 124 216 218 240
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-2-10 Fri-M-SS-1-7 Wed-E-V-2-7	110 149 192 260 89 135 147 107 132 134 228 74 104 124 216 218 240
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morrise-Kopp, Daniela Mortenen David P	Tue-E-V-3-12 Wed-A-V-4-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-6 Fri-M-V-3-6 Tue-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-2-10 Fri-M-SS-1-7 Wed-E-V-2-7 Tue-M-V-5-7	110 149 192 260 89 135 147 107 132 134 228 74 104 124 218 240 164
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-7 Wed-E-V-2-7 Tue-M-V-5-2	110 149 192 2600 89 135 147 107 132 134 2288 74 104 124 216 218 240 164 216 218 240 016 216 218 240 016 218 240 260 260 27 27 260 289 260 260 260 260 260 260 260 260 260 260
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-12-10 Fri-M-SS-12-7 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-1-12	1100 1499 1922600 899 1355 1477 107 1322 1344 2288 74 104 216 2188 2400 1642 700 1622
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-M-V-6-10 Tue-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-7 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-1-12 Thu-A-V-6-3	110 149 192 2600 89 135 147 107 132 134 228 74 104 124 216 218 240 164 70 162 211
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morres-Kopp, Daniela Mortensen, David R	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-3-12 Wed-M-V-3-12 Wed-M-V-3-12 Tue-E-V-1-9 Wed-M-V-3-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-7 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-2-7 Tue-M-V-5-3 Fri-M-V-6-1 Fri-M-V-6-1	1100 1499 1922 2600 899 1355 1477 1322 1344 2288 744 1044 1244 2166 2188 2400 1644 700 1622 2111 2344
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Fri-M-SS-1-7 Wed-E-V-2-7 Thu-A-SS-2-10 Fri-M-SS-1-7 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-1-12 Thu-A-V-6-3 Fri-M-V-6-1 Tue-A-O-1-4	1100 1499 1922600 899 1355 1477 1321 344 2288 744 1044 2218 2400 1642 2111 2344 800
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-A-SS-1-12 Thu-A-SS-2-10 Fri-M-SS-1-7 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-1-2 Thu-A-V-6-3 Fri-M-V-6-14 Tue-A-S&T-1-4	110 149 192 2600 89 135 147 107 132 134 228 74 104 124 218 2400 164 70 162 211 234 80 99
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-2-10 Fri-M-SS-12 Thu-A-SS-2-10 Fri-M-V-5-2 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-1-12 Thu-A-V-6-3 Fri-M-V-6-1 Tue-A-O-1-4 Tue-A-S&T-1-4 Wed-A-O-3-5	110 149 192260 89 135 147 107 132 134 228 74 104 124 218 240 162 211 234 80 99 1400
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-17 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-1-12 Thu-A-V-6-3 Fri-M-V-6-1 Tue-A-O-1-4 Tue-A-O-3-5 Thu-M-SS-2-3	1100 1499 1922600 899 1355 1477 1077 1322 1344 2288 744 104 2288 744 1244 2266 2188 2400 1642 702 1622 2111 2344 800 999 1400 195
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-1-19 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-17 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-1-12 Thu-A-V-6-3 Fri-M-V-6-1 Tue-A-O-1-4 Tue-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-4	110 149 192 260 89 135 147 107 132 134 228 2134 228 2134 228 240 164 70 162 211 234 80 99 140 195 195
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-10 Tue-E-V-19 Wed-M-V-6-10 Tue-A-V-6-10 Tue-A-V-6-10 Fri-M-V-6-10 Fri-M-V-6-10 Fri-M-V-6-10 Fri-M-SS-2-10 Fri-M-V-6-3 Fri-M-V-6-3 Fri-M-V-6-3 Fri-M-V-6-3 Fri-M-V-6-3 Fri-M-V-6-3 Fri-M-V-6-3 Fri-M-V-6-3 Thu-A-S&2-3 Thu-M-SS-2-3	$\begin{array}{c} 110\\ 149\\ 192\\ 260\\ 89\\ 135\\ 147\\ 107\\ 107\\ 132\\ 134\\ 228\\ 74\\ 104\\ 124\\ 218\\ 240\\ 164\\ 218\\ 240\\ 164\\ 218\\ 240\\ 162\\ 211\\ 234\\ 80\\ 99\\ 9140\\ 0195\\ 195\\ 196\end{array}$
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-A-V-6-10 Tue-A-V-6-10 Fri-M-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-14 Wed-E-V-1-4 Tue-A-O-1-4 Tue-A-S&T-1-4 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Fri-M-SS-1-2	110 149 192260 89 135 1477 132 228 74 104 228 74 104 218 240 164 218 240 164 211 234 80 99 1400 1955 239
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-6-6 Fri-M-V-3-6 Tue-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-2-10 Fri-M-SS-12 Thu-A-SS-2-10 Fri-M-V-6-1 Tue-M-V-5-2 Wed-E-V-1-12 Thu-A-S&T-1-4 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-5 Fri-M-SS-1-3	1100 1499 192260 899 1355 1477 1322 1344 2288 744 104 2134 2288 744 104 2142 2162 211 2344 800 999 1400 1955 1956 2399 2391
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-S&T-1-4 Wed-E-V-1-12 Thu-A-S&T-1-4 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Fri-M-SS-12 Fri-M-SS-12 Fri-A-O-13 Tue-M-V-5-7	1100 1499 192260 899 1355 1477 1322 1344 2288 744 104 1234 2288 744 104 124 213 2400 162 211 234 800 999 1400 1955 1956 2399 2411 71
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr Mouchtaris, Athanasios	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-A-SS-1-12 Thu-A-SS-2-10 Fri-M-SS-2-10 Fri-M-SS-2-10 Fri-M-V-6-1 Tue-A-V-6-3 Fri-M-V-6-1 Tue-A-S&T-1-4 Wed-E-V-1-12 Thu-A-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-5 Fri-M-SS-1-2 Fri-A-O-1-3 Tue-A-VS-7 Tue-A-SS-1-2	110 149 192 260 89 135 147 107 132 134 228 74 104 218 240 164 218 240 164 211 234 80 99 9140 195 196 239 241 71 96
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-A-V-6-10 Tue-A-V-6-10 Fri-M-V-6-10 Fri-M-V-6-10 Fri-M-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-14 Wed-E-V-2-7 Tue-A-SS-1-4 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-12 Fri-A-O-1-3 Tue-A-SS-1-18 Tue-A-SS-1-18 Tue-A-SS-1-18	110 149 192 260 89 135 147 107 132 218 240 164 218 240 162 211 234 80 99 91 400 195 195 195 195 239 241 71 71 71 71 75 195 195 195 195 195 195 195 19
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-14 Wed-E-V-1-4 Tue-A-S&T-1-4 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-5 Fri-M-SS-1-8 Thu-M-SS-1-8 Tue-A-SV-1-8 Tue-A-SV-1-8	$\begin{array}{c} 110\\ 149\\ 192\\ 260\\ 89\\ 135\\ 147\\ 107\\ 132\\ 134\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 216\\ 218\\ 240\\ 162\\ 211\\ 234\\ 80\\ 99\\ 140\\ 195\\ 195\\ 239\\ 241\\ 71\\ 96\\ 6110\\ 143\\ 31\\ 96\\ 110\\ 143\\ 31\\ 96\\ 110\\ 143\\ 31\\ 143\\ 140\\ 143\\ 140\\ 143\\ 140\\ 143\\ 140\\ 143\\ 140\\ 140\\ 140\\ 140\\ 140\\ 140\\ 140\\ 140$
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-2-10 Fri-M-SS-12 Thu-A-SS-2-10 Fri-M-V-6-1 Tue-A-O-1-4 Tue-A-O-1-4 Tue-A-S&T-1-4 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-5 Fri-M-SS-1-2 Fri-A-O-1-3 Tue-A-S-1-8 Tue-A-S-1-8 Tue-A-S-1-8 Tue-A-S-1-8 Tue-A-V-2-4 Wed-A-V-2-4 Fri-A-V-2-4	1100 1499 192260 899 1355 1477 1322 1344 2288 744 104 2134 2288 744 104 2142 2111 2344 800 999 1400 1955 1966 2399 2411 711 966 1100 1433 2488
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-13 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-1-13 Tue-A-V-3-18 Tue-A-SS-1-8 Tue-A-SS-1-8 Tue-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4	$\begin{array}{c} 110\\ 149\\ 192\\ 260\\ 89\\ 135\\ 147\\ 107\\ 107\\ 132\\ 134\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 228\\ 74\\ 104\\ 124\\ 80\\ 99\\ 9140\\ 195\\ 196\\ 239\\ 241\\ 71\\ 96\\ 196\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 71\\ 96\\ 100\\ 239\\ 241\\ 71\\ 71\\ 228\\ 74\\ 70\\ 70\\ 70\\ 70\\ 70\\ 70\\ 70\\ 70\\ 70\\ 70$
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr Mouchtaris, Athanasios	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-E-V-19 Wed-M-V-6-10 Tue-A-V-6-10 Fri-M-V-6-10 Fri-M-SS-2-10 Fri-M-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-S-2-3 Fri-M-V-6-1 Tue-A-O-1-4 Tue-A-O-1-4 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Tue-A-SS-1-8 Tue-E-V-4-2 Wed-A-V-2-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4	1100 1499 192260 899 1355 1477 107 107 1322 134 2288 74 104 218 2400 1622 211 2344 800 999 1400 1955 1966 2399 22411 711 960 10143 2255 1967 118
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr Mouchtaris, Athanasios Mower Provost, Emily	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-14 Wed-E-V-1-2 Thu-A-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-1-8 Tue-A-S-1-8 Tue-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4	1100 1499 192260 899 1355 1477 107 1322 8 74 104 2288 74 104 2288 74 104 218 2400 1642 2111 2344 80 999 1400 1955 239 2411 711 96 6239 2411 719 248 2400 1955 239 2411 719 241 248 2400 1955 2400 1957 2000 1000 1000000000000000000000000000
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Motlicek, Petr Mouchtaris, Athanasios Mower Provost, Emily	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-2-10 Wed-M-V-6-6 Fri-M-V-3-6 Tue-M-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-17 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-2-7 Tue-M-V-5-2 Wed-E-V-1-12 Thu-A-SS-1-14 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-1-8 Tue-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-4 Fri-A-V-3-1 Wed-A-O-3-1 Wed-A-O-3-1	1100 1499 192260 899 1355 1477 1322 1344 2288 744 104 2288 744 247 201 1642 2111 2344 80 999 1400 1955 1955 1956 239 2411 711 996 61100 1433 2488 2555 1189 1399 2411 711 719 2481 2481 2481 2481 2491 2491 2491 2491 2491 2491 2491 249
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Mosbach, Marius Motlicek, Petr Mouchtaris, Athanasios Mower Provost, Emily Mu, Tingting	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Tue-A-SS-2-10 Fri-M-SS-12 Thu-A-SS-2-10 Fri-M-SS-12 Wed-E-V-1-12 Thu-A-SS-2-10 Fri-M-V-6-1 Tue-A-O-1-4 Wed-E-V-1-12 Thu-A-S&T-1-4 Wed-E-V-2-7 Tue-A-S&T-1-4 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-4 Thu-M-SS-2-4 Thu-M-SS-1-8 Tue-M-V-5-7 Tue-A-SS-1-8 Tue-M-V-5-7 Tue-A-SS-1-8 Tue-M-V-5-7 Tue-A-SS-1-8 Tue-E-V-4-2 Wed-A-O-3-1 Wed-A-O-3-1 Wed-A-O-3-1 Wed-A-O-3-1	1100 1499 192260 899 1355 1477 1322 1344 2288 744 104 218 2400 1642 218 2400 1642 700 1622 211 2344 800 999 1400 1955 1966 2399 2411 711 966 2199 241 719 248 248 248 248 1995 1955 1965 239 241 719 248 248 248 248 248 249 248 249 249 249 249 249 249 249 249 249 249
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Morse-Kopp, Daniela Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Mosbach, Marius Motlicek, Petr Mouchtaris, Athanasios Mower Provost, Emily Mu, Tingting Mueller, Markus	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-6-10 Tue-E-V-1-9 Wed-M-V-6-10 Fri-M-SS-1-7 Wed-E-V-2-2 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-13 Fri-M-V-6-1 Tue-A-S-1-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Tue-A-SS-1-8 Tue-E-SS-1-6 Wed-A-O-3-1 Wed-M-V-2-9 Tue-A-SS-1-8	$\begin{array}{c} 110\\ 149\\ 192\\ 260\\ 89\\ 135\\ 147\\ 107\\ 107\\ 132\\ 134\\ 228\\ 74\\ 104\\ 124\\ 218\\ 240\\ 164\\ 218\\ 240\\ 164\\ 218\\ 240\\ 164\\ 218\\ 240\\ 164\\ 218\\ 240\\ 164\\ 218\\ 240\\ 164\\ 218\\ 240\\ 164\\ 211\\ 234\\ 80\\ 99\\ 99\\ 140\\ 195\\ 196\\ 239\\ 241\\ 195\\ 196\\ 239\\ 241\\ 195\\ 196\\ 239\\ 241\\ 101\\ 143\\ 248\\ 255\\ 118\\ 139\\ 125\\ 96\\ 29\\ 26\\ 110\\ 110\\ 143\\ 248\\ 255\\ 118\\ 139\\ 125\\ 96\\ 26\\ 26\\ 26\\ 26\\ 26\\ 26\\ 26\\ 26\\ 26\\ 2$
Mori, Hiroki Morimoto, Yoshiya Morise, Masanori Moritz, Niko Moriya, Takafumi Moro-Velázquez, L Moro-Velázquez, L Morris, Ethan Morse-Kopp, Daniela Mortensen, David R Mosbach, Marius Mosbach, Marius Motlicek, Petr Mouchtaris, Athanasios Mower Provost, Emily Mu, Tingting Muguli, Ananya	Tue-E-V-3-12 Wed-A-V-4-9 Thu-M-V-6-9 Fri-A-S&T-1-1 Tue-A-V-4-5 Wed-M-V-6-13 Wed-A-V-3-12 Tue-E-V-2-10 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-12 Wed-M-V-5-10 Tue-A-V-6-10 Tue-E-V-1-9 Wed-M-V-2-2 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-12 Thu-A-SS-1-14 Wed-E-V-2-7 Tue-A-SS-1-4 Wed-A-O-3-5 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Thu-M-SS-2-3 Tue-A-SS-1-8 Tue-E-V-4-2 Wed-A-O-3-5 Tue-A-SS-1-6 Wed-A-O-3-1 Wed-A-O-3-1 Wed-A-O-3-1 Tue-A-SS-1-6 Wed-A-O-3-1 Wed-A-O-3-1 Tue-A-SS-1-6 Wed-A-O-3-1 Wed-A-O-3-1	1100 149 192 260 89 135 147 107 132 134 228 74 104 228 74 104 228 74 104 228 240 162 211 234 80 99 91 105 239 241 71 71 70 71 228 240 162 214 74 218 240 162 219 234 200 162 211 234 80 99 91 105 239 241 71 71 71 71 72 74 74 74 74 74 74 74 74 74 74

Muhlack. Beeke	Thu-M-V-6-7	191
	Fri-M-V-2-4	225
Mukherjee, Sankar	Thu-A-V-5-8	210
Mukhopaunyay, K Mulder Kimberley	Tue-A-V-6-11	215
Mulimani, Manjunath	Tue-A-V-1-5	82
Müller, Markus	Wed-A-V-2-4	143
Müller, Nicolas Michael .	Thu-A-V-5-12	211
Muller, Robert	Iue-M-SS-1-5 	78 78
Mumtaz, Benazir	Tue-M-V-7-5	75
·	Wed-E-V-4-12	170
Muniyappa, Thejaswi	Tue-A-V-5-2	91
Munkhualai, 1	<u>IUE-E-V-5-4</u> Τμε-Δ-V-5-3	91
Muralidharan, Deepak	Tue-E-V-4-4	111
Murauskas, Šarūnas	Thu-M-SS-2-3	195
Murthy, Aditya	Tue-A-V-2-3	85
Murtny, Hema A	Wed-E-V-1-6	101
Murty, K. Sir Kana Mussakhojaveva, Saida	Wed-F-V-6-10	174
Muxika, Oihane	Fri-M-V-2-7	226
N		
IN		-
N., Krishna D	Wed-E-V-1-5	161
Na, riviuolig Na. Yuevile	Tue-F-V-2-9	207
Nachman, Lama	Wed-E-V-2-10	164
Nachmani, Eliya	Fri-M-O-3-1	221
Naderi, Babak	Wed-A-V-4-4	148
	Wed-A-V-4-5	148
Nagamatsu, Kenji	Thu-M-V-4-4	187
	Thu-M-V-4-8	188
M	Fri-A-V-2-4	246
Nagano, Mizuki	Wed-E-V-2-3	103
Nagaraja, varun	Fri-A-V-3-10	249
Naijo, Satsuki	Fri-A-V-1-14	245
Nair, Arun	Wed-E-SS-1-1	175
Nakadal, Kazuhiro	Wed-A-V-4-9	149
Nakamura. Satoshi	Wed-M-V-2-7	$105 \\ 125$
	Wed-A-V-6-6	153
	Wed-A-V-6-7	153
	Wed-A-V-6-9	154
Nakamura, Taiki	Tue-M-V-2-3	65
Nakano, Kaito	Tue-M-V-6-4	73
Nakano, Teppei	Wed-M-V-5-5	131
	Wed-A-V-4-2	148
Nakayama, Takahiro	Tue-M-V-6-4	73
Nalamalapu, Anil Kumar	Tue-M-V-4-1	69
Nam, Gyuhyeon	Wed-A-V-5-1	150
Namboodiri. Vinav	Thu-A-V-6-11	213
Namyalo, Saudah	Wed-M-O-2-3	120
Nanavati, Jai	Wed-E-V-1-7	161
Nanavati, Raoul	Wed-E-V-1-7	161
Nandi. Ritika	Tue-A-33-2-1	82
Naowarat, Burin	Wed-E-V-6-1	173
Narang, Ankur	Tue-E-V-6-7	115
Narayan, Vaibhav A	Tue-A-V-2-6	85
Narayanan, Arun	Wed-M-V-6-4	133
	Fri-M-V-6-3	234
Narayanan, Shrikanth S.	Thu-M-V-1-2	180
	Inu-A-V-1-6 Fri-M-V-2-1	201
Narayanan Sundara., M.	Thu-M-V-7-6	193
Narayanaswamy, V.S	Wed-A-V-4-12	150
Nareddula, Santhan K.R.	Thu-A-O-2-2	199
Nairseuy, Chaitanya Nasir. Md	Tue-M-V-5-8	180
Nasreen, Shamila	Wed-A-V-1-6	142
Nataraj, K.S	Thu-M-S&T-1-7	197
Natarajan, S	Wed-E-V-5-9	172
INAUIWAIII, NAIAII	<u>— тие-м-v-з-4</u> Thu- <u>M-V-3-2</u>	185
Nautsch, Andreas	Tue-E-V-1-13	105
	Wed-A-O-2-6	139
Navar Indu	Fri-M-V-7-9 Fri-A-SS-2-2	238 257
Navem, Khandokar Md.	Tue-M-V-3-12	68
, , ,		-

Nechanský, Tomáš Negri, Matteo	Wed-M-O-2-1 Wed-A-V-6-11	120 154
	Thu-A-V-6-2	211
Neiterman, Evgeny H	Fri-M-V-5-1	231
Nelson, Philip C	Fri-A-SS-2-1	257
	Fri-A-55-2-10	259
Nelus Alexandru	Tuo.A.SS.1.5	239
Németh. Géza	Wed-A-V-5-8	151
Ness, Steven R.	Wed-E-O-3-2	159
Nessler, Natalia	Wed-E-O-3-4	160
Neubig, Graham	Tue-M-V-5-2	70
Neumann, Michael	Wed-A-V-1-7	142
	Fri-A-SS-2-2	257
Ney, Hermann	Tue-Keynote	100
	Wed-A-O-I-4	138
	Thu-M-O-1-5	145
	Thu-M-O-3-2	179
	Thu-M-O-3-3	179
Ng, Cymie Wing-Yee	Thu-M-V-1-5	180
Ng, Edwin G.	Thu-A-SS-1-3	214
Ng, Sara	Wed-A-V-2-5	143
Ng, Si-Ioi	Thu-M-V-1-5	180
Nguyen, Dat Quoc	Fri-A-V-6-1	254
Nguyen, Ha	Wed-M-O-1-5	119
Nguyen Hoang H	Wed-E-O-2-1	130
Nguyen Hoang Long	Tue-M-V-4-1	69
Nguyen, Huu-Kim	Thu-A-V-5-2	209
Nguyen, Huyen	Tue-E-O-3-4	103
Nguyen, Manh Hung	Fri-A-S&T-1-5	261
Nguyen, Thai-Son	Wed-M-V-6-1	133
Nguyen, Tuan-Nam	Wed-E-V-1-2	160
Nguyen, IuAnh	Wed-M-V-2-10	125
Nguyen Viet-Nhat	Fri-A-3&1-1-5	201
Ni. Hao	Wed-F-V-3-9	166
Ni, Junrui	Tue-E-V-5-8	113
Ni, Shikang	Fri-A-S&T-1-6	261
Nicmanis, Dāvis	Thu-M-S&T-1-6	197
Nicolson, Aaron	Tue-M-V-3-2	66
Nie, Yuting	Thu-M-SS-1-5	194
Nielsen Jesper Kiær	Tue-M-O-3-2	108
Nigmatulina, Julija	Thu-M-SS-2-4	195
Nikolaidou, Konstantina	Tue-A-SS-2-4	97
Nishida, Kenji	Wed-A-V-4-9	149
Nishizaki, Hiromitsu	Tue-M-V-6-7	74
Nine Mineman	Wed-E-V-1-4	161
Niu, Miligyue		104
Noé Paul-Gauthier	Wed-A-O-2-6	139
Nonaka. Yuto	Tue-M-V-6-7	74
Norel, Raquel	Fri-A-SS-2-2	257
Noroozi, Vahid	Wed-M-O-1-4	119
Norouzi, Mohammad	Thu-A-SS-1-11	216
Nortje, Leanne	Wed-M-V-2-6	124
Noso Takashi		182
	Fri-A-V-1-14	245
Nöth. Elmar	Tue-M-O-2-2	61
,	Tue-M-SS-1-3	77
	Tue-E-SS-1-7	118
	Wed-E-O-3-2	159
Novak Josef		217
Novikova, Jekaterina	Thu-A-SS-2-5	217
Novitasari, Sashi	Fri-M-V-4-8	231
Novoselov, Sergey	Wed-A-SS-1-3	155
Novotney, Scott	Tue-A-SS-1-3	.95
Novotny, Ondřej	Thu-M-SS-2-2	195
Nozaki, Jumon	Thu-A-SS-1-5	215
Nugraha, Aditya Arie	Tue-A-V-3-2	190
Nwogu, Ifeoma	Wed-A-V-6-13	155
•		- 0
U		
Oard, Douglas W	Thu-A-V-6-2	211
O'Prion Ponjamin	Thu-M-O-2-1	178
V DUCH, DEHIMINI	Thu-A-V-4-8	200

Oard, Douglas W	Thu-A-V-6-2	211
Obin, Nicolas	Thu-M-O-2-1	178
O'Brien, Benjamin	Thu-A-V-4-8	208
	Fri-A-V-4-8	251
Obyat, Abdelrahman	Thu-A-SS-2-3	217
Ochiai, Tsubasa	Tue-E-V-2-10	107
	Wed-M-V-6-6	134
	Wed-A-V-4-2	148
	Thu-A-V-3-5	205

O'Connor, Mark	Fri-M-V-6-12	236
Odobez, Jean-Marc	Tue-A-O-1-4	80
Octomorph Corolin		00
Oetzmann, Caronn	1 ue-A-V-2-6	00
Ogawa, Atsunori	Tue-M-V-3-5	67
Ogawa Tetsuii	Wed-M-V-5-5	131
Ogawa, recouji		105
	I NU-IVI-V-3-6	102
Oh, Miran	Fri-M-V-2-1	224
Oh Yoo Rhee	Tuo-A-S&T-1-6	99
	10e-A-3@1-1-0	115
Ohmura, Hidefumi	Tue-E-V-6-4	115
	Thu-M-V-6-3	190
Obmaiaan Olivar		100
Onneiser, Onver	<u>ппп-м-22-3</u>	195
Ohta. Mavumi	Tue-E-V-5-7	113
Okomura Daiki	Wed F V 2 12	167
	weu-E-v-5-15	107
Okhotnikov, Anton	Wed-A-SS-1-1	155
Olaleve Kavode	Thu-M-V-2-5	183
		211
Oliveira, Frederico S. de	I hu-A-V-5-12	211
Ollerenshaw. Anna	Fri-M-V-3-10	229
Omachi Motoi		215
	1 flu-A-55-1-0	215
O'Malley, Ronan	Thu-M-V-1-7	181
Omologo Maurizio	Tue-M-V-5-7	71
omologo, Maarizio		1 4 2
	Wed-A-V-2-4	143
Omote. Masanori	Tue-M-V-2-10	66
Oncoscu Androop Maria		160
Oncescu, Anureca-Maria	wea-E-O-3-3	100
O'Neill, Patrick K	Wed-M-O-1-4	119
Önen Melek	Tuo.A.SS.1.1	94
Ormala India Dalla		170
Onnela, Jukka-Pekka …	Thu-M-O-2-2	178
Oostermeiier. Koen	Wed-E-SS-1-8	176
Onala Matouez		100
Opaia, mateusz	Tue-E-O-1-1	100
	Tue-E-O-1-2	100
Orihashi Shota	Tue-M-V-6-6	74
crimani, mora		01
	Tue-A-V-5-5	91
	Wed-E-V-3-13	167
		228
	FTI-IVI-V-3-0	220
	Fri-A-V-1-11	245
Orozco-Arrovave IR	Tue-M-0-2-2	61
010200711109470, 5.10. 1.		77
	Tue-M-SS-T-3	11
	Thu-A-SS-2-2	217
Ortega Alfonso	Wed-E-0-1-3	158
Ortega, Anonso	Weu L O I S	130
	Fri-A-O-1-1	241
Ostapenko, Alissa	Tue-E-V-4-10	112
Octopdorf Mari		142
Ostelluoli, Mali	weu-A-v-2-5	145
	Wod_F_V_4_5	168
	Weu L V T J	
Ostermann, Jörn	Fri-M-O-1-3	219
Ostermann, Jörn	Fri-M-O-1-3	219
Ostermann, Jörn Ottl, Sandra	Fri-M-O-1-3 Tue-M-SS-1-1	219 77
Ostermann, Jörn Ottl, Sandra Ou, Wenjie	Fri-M-O-1-3 Tue-M-SS-1-1 Wed-M-V-3-3	219 77 126
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian	Fri-M-O-1-3 Tue-M-SS-1-1 Wed-M-V-3-3	219 77 126
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian	Fri-M-O-1-3 Tue-M-SS-1-1 Wed-M-V-3-3 Wed-A-V-3-6	219 77 126 146
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei	Fri-M-O-1-3 Tue-M-SS-1-1 Wed-M-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8	219 77 126 146 215
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei	Fri-M-O-1-3 Tue-M-SS-1-1 Wed-M-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8	219 77 126 146 215
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P	Fri-M-O-1-3 Tue-M-SS-1-1 Wed-M-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8	219 77 126 146 215
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P	Fri-M-O-1-3 Tue-M-SS-1-1 Wed-M-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8	219 77 126 146 215
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield Dirk	Wed A V/6 2	219 77 126 146 215
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk	Wed-A-V-6-2	219 77 126 146 215
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk	Wed -A-V-6-2 Fri-M-A-SS-1-1 Wed-M-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6	219 77 126 146 215 152 250
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M	Wed -A-V-6-2 Fri-A-V-4-6 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4	219 77 126 146 215 152 250 97
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Jasaph	Wed -A-V-6-2 Fri-A-V-6-2 Fri-A-V-4-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4	219 77 126 146 215 152 250 97
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph	Wed -A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Thu-A-SS-1-1 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9	219 77 126 146 215 152 250 97 204
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti	Wed -A-V-6-2 Fri-A-V-6-2 Fri-A-V-6-2 Fri-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8	219 77 126 146 215 152 250 97 204 92
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaz, Dimitri	Wed -A-V-6-2 Fri-A-V-6-2 Fri-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-F-V-2-7	219 77 126 146 215 152 250 97 204 92 164
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Palfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaz, Ashuizi	Wed -A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7	219 77 126 146 215 152 250 97 204 92 164
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaz, Dimitri Palekar, Ashwini	Wed -A-V-6-2 Fri-A-V-6-2 Fri-A-V-6-2 Fri-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15	219 77 126 146 215 152 250 97 204 92 164 260
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaskar, Ashwini Palekar, Ashwini Palmer, Ian	Wed -A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-2	219 77 126 146 215 152 250 97 204 92 164 260 211
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaz, Dimitri Palekar, Ashwini Pan Changije	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-3-3	219 77 126 146 215 152 250 97 204 92 164 260 211 67
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaskar, Shruti Palekar, Ashwini Palmer, Ian Pan, Changjie	Wed -A -V-3-3 Tue-M-SS-1-1 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-6-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-M-V-3-3	219 77 126 146 215 152 250 97 204 92 164 260 261 167
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaskar, Shruti Palaekar, Ashwini Palekar, Ashwini Palmer, Ian Pan, Changjie Pan, Ho-hsien	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-8 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-8	219 77 126 146 215 152 250 97 204 92 164 260 211 67 169
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palaskar, Shruti Palaz, Dimitri Palekar, Ashwini Palmer, Ian Pan, Changjie Pan, Ho-hsien Pan, Jing	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-A-S-3 Wed-E-V-4-8 Thu-A-V-6-1	219 77 126 146 215 152 250 97 204 92 164 260 211 67 169 202
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaskar, Shruti Palekar, Ashwini Palmer, Ian Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Jing	Wed -A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-6-1 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-4-6 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-A-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11	219 77 126 146 215 152 250 97 204 92 164 260 211 67 202
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M. Palakapilly, Joseph Palakar, Shruti Palaz, Dimitri Palaz, Dimitri Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Paner, Ian Pan, Changjie Pan, Jing Pan, Jing	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4	219 77 126 146 215 152 2500 97 204 92 164 2600 211 67 169 202 211
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaskar, Shruti Palekar, Ashwini Palekar, Ashwini Palmer, Ian Pan, Changjie Pan, Jing Pan, Jing Pan, Shifeng	Wed -A -V-6-2 Fri-A-V-6-2 Tue-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-A-V-6-3 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-4 Fri-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8	219 77 126 146 215 250 97 204 92 164 260 211 67 169 202 111 253
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palaskar, Shruti Palaskar, Ashwini Palmer, Ian Palmer, Ian Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Shifeng Pan, Yilin	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tri-A-SS-1-8 Wed-A-V-3-9 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-V-5-8	219 77 126 146 215 152 250 97 204 92 164 260 211 67 169 202 111 253 181
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palakar, Shruti Palekar, Ashwini Palmer, Ian Palmer, Ian Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Shifeng Pan, Yilin	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tri-A-V-4-6 Fri-A-V-4-7 Thu-A-V-2-9 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-6-1 Tue-A-V5-8 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-V-5-8 Thu-A-V-5-8 Thu-A-V-5-7	219 77 126 146 215 250 97 204 92 164 260 211 67 169 202 111 253 181
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palaskar, Shruti Palaskar, Shruti Palekar, Ashwini Palekar, Ashwini Palmer, Ian Pan, Changjie Pan, Jing Pan, Jing Pan, Shifeng Pan, Yilin	Wed -A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-2-7 Fri-A-V-4-6 Tue-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-M-V-1-7 Thu-A-SS-2-7	219 77 126 146 215 152 250 97 204 92 164 260 211 67 169 202 2111 253 181 218
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M. Palakapilly, Joseph Palakar, Shruti Palakar, Shruti Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Paner, Ian Pan, Changjie Pan, Jing Pan, Jing Pan, Shifeng Pan, Yilin Pan, Yiqian	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-6-1 Tue-M-V-3-3 Wed-E-V-2-7 Fri-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-SS-17	219 77 126 146 215 152 250 97 204 260 211 67 169 202 111 253 181 218 156
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palaskar, Shruti Palaskar, Shruti Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palmer, Ian Pan, Changjie Pan, Jing Pan, Jing Pan, Jing Pan, Shifeng Pan, Yilin Pan, Yiqian Pan, Yiqian	Wed -A -V-6-2 Fri-A-V-6-2 Tue-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-6-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-A-V-6-3 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-B-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-M-V-1-7 Thu-A-SS-2-7 Wed-A-V-5-8 Thu-M-V-1-7 Thu-A-SS-2-7 Wed-A-SS-1-7 Fri-A-W-5-8 Thu-M-V-1-7 Wed-A-SS-1-7 Fri-A-W-5-8 Thu-M-Y-1-7 Wed-A-SS-1-7 Fri-A-W-5-8 Thu-M-Y-1-7 Wed-A-SS-1-7 Fri-A-W-5-8	219 77 126 215 152 250 97 204 92 164 260 211 67 169 202 111 253 181 218 156 202
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakar, Shruti Palaskar, Shruti Palaskar, Shruti Palekar, Ashwini Palekar, Ashwini Palmer, Ian Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing jing Pan, Shifeng Pan, Yilin Pan, Yiqian Pan, Pan Pan, Michele	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-2 Fri-A-V-4-6 Tue-A-V-2-9 Tue-A-V-2-9 Tue-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-SS-2-17 Thu-A-V-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Fri-M-V-7-13	219 77 126 146 215 152 250 97 204 92 202 164 260 211 67 169 202 111 253 181 218 156 239 97
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palakar, Shruti Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Shifeng Pan, Yilin Pan, Yiqian Panda, Rameswar	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-V-5-8 Thu-A-V-5-7 Thu-A-SS-2-7 Wed-A-V-5-17 Thu-A-SS-17 Wed-A-V-7-12 Wed-M-V-2-12	219 77 126 215 152 250 97 204 92 204 92 164 260 211 67 169 202 111 253 181 218 56 209 126
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palakar, Shruti Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Jing Pan, Jing Pan, Jing Pan, Shifeng Pan, Yilin Panariello, Michele Panda, Rameswar	Wed -A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-1 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Thu-A-V-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-111 Tue-E-V-4-4 Fri-A-V-5-8 Thu-M-V-1-7 Thu-A-SS-2-7 Wed-A-SS-1-7 Fri-M-V-7-13 Wed-M-V-2-12 Thu-M-V-2-12	219 77 126 215 250 97 204 92 260 211 67 169 202 111 253 181 218 56 239 926 184
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M. Palakapilly, Joseph Palakar, Shruti Palakar, Shruti Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Jing Pan, Yilin Pan, Yiqian Panday, Abbiabali	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-8 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-SS-17 Fri-M-V-7-13 Wed-M-V-2-12 Thu-M-V-2-18	219 77 126 215 250 97 204 92 204 92 204 92 204 164 260 202 111 253 81 125 239 126 181 215 239 126 146 215
Ostermann, JörnOttl, SandraOu, WenjieOu, ZhijianOuyang, BeibeiOuyang, BeibeiPPadfield, DirkPahlavan Nodeh, M.Palakapilly, JosephPalaskar, ShrutiPalaz, DimitriPalekar, AshwiniPalker, AshwiniPalmer, IanPan, ChangjiePan, Ho-hsienPan, JingPan, ShifengPan, YiqianPanda, RameswarPandey, Abhishek	Wed -A -V-6-2 Trie-M-SS-1-1 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-46 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-6-1 Tue-A-SS-2-15 Thu-A-V-6-1 Tue-B-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-SS-1-7 Fri-A-V-7-13 Wed-M-V-2-12 Thu-M-V-2-8 Tue-F-V-5-5	219 77 126 215 152 250 97 204 92 164 260 211 67 69 202 111 253 181 218 239 126 239 126 184 113
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palakar, Shruti Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Jing Pan, Jing Pan, Jing Pan, Shifeng Pan, Yilin Panariello, Michele Pandey, Abhishek	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-2-9 Tue-A-V-2-9 Tue-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-SS-2-7 Wed-A-SS-1-7 Thu-A-SS-2-7 Wed-A-SS-1-7 Fri-M-V-7-13 Wed-M-V-2-12 Thu-A-V-2-8 Tue-E-V-5-9	219 77 126 215 250 97 204 260 201 167 169 202 211 67 169 202 211 153 181 218 156 2399 126 818 151 218 151 218 151 218 151 218 151 218 200 201 201 201 201 201 201 201 201 201
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palakar, Shruti Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Shifeng Pan, Yilin Pan, Yilin Pandey, Abhishek Pandey, Prem C	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-46 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-6-1 Tue-A-V-5-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-V-5-8 Thu-A-SS-2-7 Wed-A-V-1-7 Thu-A-SS-2-7 Wed-A-V-2-12 Thu-M-V-2-13 Wed-M-V-2-12 Thu-M-V-2-13 Wed-E-V-5-5 Wed-E-V-1-7	219 77 126 215 250 97 204 92 204 92 216 4 260 211 67 169 202 111 253 181 1218 156 239 126 184 156 239 126 184 152 219 181 219 202 111 203 204 201 204 201 204 205 204 204 205 204 205 204 205 204 205 204 205 204 205 204 205 204 205 204 205 204 205 204 205 205 204 205 204 205 205 205 205 205 205 205 205 205 205
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M. Palakapilly, Joseph Palaskar, Shruti Palaz, Dimitri Palaekar, Ashwini Palner, Ian Pan, Changjie Pan, Jing Pan, Jing Pan, Jing Pan, Shifeng Pan, Yiqian Pandey, Abhishek Pandey, Prem C.	Wed -A-V-6-2 Fri-M-0-1-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-1 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-111 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-V-17 Thu-A-SS-2-7 Wed-A-SS-1-7 Fri-M-V-7-13 Wed-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-5-5 Wed-E-V-1-9 Thu-M-S&T-1-7	219 77 126 215 250 97 204 92 260 211 67 169 202 211 17 253 181 218 156 239 9126 184 131 67 169 202 211 17 253 181 258 209 204 260 201 27 204 260 201 204 260 201 204 260 201 204 260 201 204 260 201 204 205 200 204 201 204 205 200 204 201 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 205 200 204 202 201 205 200 204 205 200 201 205 201 205 205 205 205 205 205 205 205 205 205
Ostermann, JörnOttl, SandraOu, WenjieOu, ZhijianOuyang, BeibeiOuyang, BeibeiPPadfield, DirkPalakapilly, JosephPalakapilly, JosephPalakapilly, JosephPalakar, ShrutiPalekar, AshwiniPalekar, AshwiniPalmer, IanPan, ChangjiePan, JingPan, JingPan, ShifengPan, YiqianPandey, AbhishekPandey, Prem C.Pang, Ruoming	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-6-2 Fri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V5-8 Thu-A-SS-2-7 Wed-A-SS-17 Fri-M-V-7-13 Wed-M-V-2-12 Thu-M-V-2-8 Tue-E-V-5-5 Wed-E-V-19 Thu-M-S&T-17 Wed-M-V-6-4	219 77 126 215 250 97 204 92 204 92 204 92 204 164 260 211 67 169 202 111 253 81 121 239 126 414 209 202 111 253 81 125 239 204 97 204 111 253 81 12 23 97 126 113 127 127 127 127 127 127 127 127 127 127
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palakar, Shruti Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Shifeng Pan, Yiqian Pandey, Abhishek Pandey, Prem C Pang, Ruoming	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Thu-A-V-6-1 Tue-A-V-5-3 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-V-5-7 Thu-A-SS-2-7 Wed-A-SS-17 Fri-A-V-7-13 Wed-A-SS-17 Fri-N-V-7-13 Wed-M-V-2-12 Thu-M-SS-1-7 Wed-E-V-1-9 Thu-M-S&T-1-7 Wed-E-V-1-9 Thu-M-S&T-1-7 Wed-M-V-6-4 Wed-M-V-6-4	219 777 126 146 215 152 2500 97 204 92 260 211 67 260 211 67 260 211 169 202 111 218 152 253 181 263 204 164 260 211 67 169 202 111 167 215 250 97 204 167 204 215 215 215 215 215 215 215 215 215 215
Ostermann, JörnOttl, SandraOu, WenjieOu, ZhijianOuyang, BeibeiOuyang, BeibeiPPadfield, DirkPalakapilly, JosephPalakapilly, JosephPalakapilly, JosephPalakapilly, JosephPalakar, ShrutiPalekar, AshwiniPalekar, AshwiniPalekar, AshwiniPalekar, IanPaner, IanPan, ChangjiePan, JingPan, JingPan, ShifengPan, YilinPanariello, MichelePanda, RameswarPandey, AbhishekPandey, Prem C.Pang, Ruoming	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-SS-2-7 Wed-E-V-2-7 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-SS-2-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Fri-M-V-7-13 Wed-M-V-2-12 Thu-M-V-2-8 Tue-E-V-5-5 Wed-E-V-1-9 Thu-M-S&T-17 Wed-M-V-6-4 Wed-M-V-6-40	219 77 126 2250 97 204 260 260 202 111 67 169 202 211 67 169 202 211 153 181 218 156 2399 126 81413 181 218 156 239 126 169 202 211 3181 218 200 201 201 201 201 201 201 201 201 201
Ostermann, JörnOttl, SandraOu, WenjieOu, ZhijianOuyang, BeibeiOuyang, BeibeiPPadfield, DirkPahlavan Nodeh, M.Palakapilly, JosephPalakapilly, JosephPalakan, ShrutiPalaz, DimitriPalekar, AshwiniPalmer, IanPan, ChangjiePan, Ho-hsienPan, JingPan, JingPan, ShifengPan, YilinPanariello, MichelePandey, AbhishekPandey, Prem C.Pang, Ruoming	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Fri-A-V-46 Fri-A-V-47 Fri-A-V-48 Wed-E-V-2-7 Fri-A-V-5-8 Wed-E-V-2-7 Fri-A-V-6-1 Tue-A-V-5-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-27 Wed-A-SS-17 Fri-M-V-7-13 Wed-ASS-17 Fri-M-V-2-12 Thu-M-SZ-7 Wed-A-V-1-7 Wed-M-V-2-12 Thu-M-SS-10 Wed-M-V-2-17 Wed-M-V-2-18 Tue-E-V-5-5 Wed-M-V-6-10 Fri-M-V-6-10 Fri-M-V-6-10	219 77 126 215 250 97 204 92 204 92 216 4 260 211 67 169 202 111 253 81 1218 156 239 126 184 156 239 126 184 131 162 219 77 204 92 204 205 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 201 204 202 204 201 204 201 202 204 204
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M. Palakapilly, Joseph Palakar, Shruti Palaz, Dimitri Palaz, Dimitri Palekar, Ashwini Palner, Ian Pan, Changjie Pan, Jing Pan, Jing Pan, Jing Pan, Shifeng Pan, Yiqian Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-M-V-1-17 Wed-A-SS-2-7 Wed-A-V-2-12 Thu-M-V-2-13 Wed-M-V-2-14	219 77 126 2250 97 204 2250 97 204 260 260 202 111 67 169 202 211 67 169 202 211 67 169 202 111 153 181 218 156 239 912 66 184 133 162 197 204 201 201 201 201 201 201 201 201 201 201
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palex, Ashwini Pan, Ho-hsien Pan, Jingjing Pan, Jingjing Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Prem C. Pantazis, Yannis Pantazis, Yannis	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V5-8 Wed-E-V-2-7 Fri-A-V-5-8 Thu-A-V-1-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-SS-17 Fri-M-V-7-13 Wed-A-SS-19 Thu-M-V-2-8 Tue-E-V-5-5 Wed-E-V-19 Thu-M-S&T-17 Wed-M-V-6-10 Fri-M-V-6-10	219 77 126 215 250 97 204 92 204 92 202 164 260 211 67 169 202 111 253 81 12 81 239 126 411 253 181 215 239 126 4146 215 201 97 204 92 204 111 253 81 12 239 126 127 133 181 12 239 126 129 120 120 111 213 181 12 12 13 13 12 12 12 12 12 12 12 12 12 12 12 12 12
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakan Nodeh, M. Palakapilly, Joseph Palaskar, Shruti Palaz, Dimitri Palekar, Ashwini Palekar, Ashwini Palner, Ian Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Shifeng Pan, Yiqian Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Pantic, Maja	Wed-A-V-6-2 Fri-A-V-3-6 Thu-A-SS-1-1 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Thu-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-V-5-17 Thu-A-SS-2-7 Wed-A-SS-1.7 Fri-A-V-5-8 Thu-A-SS-2.7 Wed-A-SS-1.7 Fri-M-V-7-13 Wed-M-V-2-12 Thu-M-SS-1.7 Wed-A-SS-1.7 Fri-M-V-7-13 Wed-M-V-2-12 Thu-M-S&T-1.7 Wed-M-V-2-19 Thu-M-S&T-1.7 Wed-M-V-6-4 Wed-M-V-6-4 Wed-M-V-6-4 Wed-M-V-6-5-8 Thu-A-V-5-8 Thu-A-V-5-8 Thu-M-S&T-1.7 Wed-M-V-6-40 Wed-M-V-6-2.9	219 777 126 219 777 126 216 2250 97 204 92 2250 97 204 260 211 67 260 211 67 260 211 253 181 253 202 111 253 261 261 275 204 97 204 261 261 261 261 261 261 261 261 261 261
Ostermann, JörnOttl, SandraOu, WenjieOu, ZhijianOuyang, BeibeiOuyang, BeibeiPPadfield, DirkPalakapilly, JosephPalakapilly, JosephPalakapilly, JosephPalakar, ShrutiPalexar, AshwiniPalekar, AshwiniPalekar, AshwiniPalekar, AshwiniPalekar, AshwiniPaner, IanPan, ChangjiePan, JingPan, JingPan, JingPan, ShifengPan, YilinPanariello, MichelePandey, AbhishekPandey, Rem C.Pantazis, YannisPantic, MajaPantazis, Vasileios	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-6-11 Tue-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-SS-2-7 Wed-E-V-2-7 Wed-A-SS-1-7 Fri-M-V-7-13 Wed-A-SS-1-7 Fri-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-13 Wed-E-V-1-19 Thu-M-S&T-1-7 Wed-M-V-6-40 Fri-M-V-6-41 Thu-M-V-2-8 Thu-M-V-2-9 Wed-A-V-2-4 Wed-A-V-2-4	2199 777 1266 2157 204 92 204 92 204 92 204 92 204 92 204 92 204 111 253 81 218 156 2399 1266 181 112 213 181 218 156 2399 210 184 1462 215 200 97 204 111 253 81 215 200 201 111 253 81 215 200 201 111 253 81 215 200 201 111 253 81 215 200 201 111 253 81 202 211 202 204 202 201 111 202 204 202 204 209 204 202 202
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palakapilly, Joseph Palaskar, Shruti Palaz, Dimitri Palekar, Ashwini Palmer, Ian Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Yilin Pandey, Abhishek Pandey, Prem C. Pantazis, Yannis Pantic, Maja Pantazis, Vasileios Pandourakis, Vasileios	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tri-A-V-4-6 Fri-A-V-4-7 Fri-A-V-4-8 Wed-E-V-2-7 Fri-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Thu-A-V-1-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2.7 Wed-A-SS-1.7 Fri-M-V-7-13 Wed-A-SS-1.7 Fri-M-V-2-12 Thu-M-SS-2.7 Wed-A-SS-1.7 Fri-M-V-2-13 Wed-M-V-2-12 Thu-M-SS-1.7 Wed-M-V-2-13 Wed-M-V-2-12 Thu-M-SS-1.7 Wed-M-V-2-13 Wed-M-V-2-14 Thu-M-SS-2.17 Wed-M-V-6-10 Fri-M-V-6-10 Fri-M-V-6-10 Fri-M-V-2-8 Thu-A-V-5-8 Thu-A-V-5-8 <	219 77 126 215 250 97 204 92 204 92 200 211 67 169 202 111 253 81 121 239 126 184 113 169 202 111 253 181 219 202 111 259 204 97 204 97 204 92 204 97 204 92 204 97 204 92 204 205 111 258 209 204 111 258 209 204 111 258 209 204 111 258 209 204 111 258 209 204 111 258 209 204 111 258 209 204 111 258 209 204 111 258 209 204 111 258 209 204 111 258 209 126 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 10 10 11 202 202 111 215 202 10 10 10 202 111 202 202 10 10 202 10 10 202 10 10 10 10 202 10 10 10 10 10 10 10 202 10 10 10 10 10 202 10 10 10 10 10 10 10 10 10 10 10 10 10
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palakapilly, Joseph Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palaz, Dimitri Palekar, Ashwini Palner, Ian Pan, Changjie Pan, Changjie Pan, Jing Pan, Jing Pan, Jingjing Pan, Yiqian Pan, Yilin Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Papadourakis, Vasileios Papagagari, Raghavendra	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-1 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-B-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V5-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V5-8 Thu-M-V-1-13 Wed-A-V-2-12 Thu-M-V-2-13 Wed-M-V-2-12 Thu-M-V-2-8 Tue-E-V-5-5 Wed-M-V-6-4 Wed-M-V-6-4 Wed-M-V-6-10 Fri-M-V-4-17 Wed-M-V-6-10 Fri-M-V-4-12 Thu-M-V-2-9 Wed-A-V-2-4 Thu-A-V-5-8 Thu-M-V-2-9 Wed-A-V-2-4	219 77 126 2250 97 204 2250 97 204 260 260 202 111 67 169 202 211 67 169 202 211 173 3181 218 156 239 9126 84 133 162 197 133 3134 229 210 184 146 239 200 184 146 239 200 126 200 127 200 127 200 126 200 126 200 127 200 126 200 126 200 200 126 200 200 126 200 200 126 200 200 200 200 200 200 200 200 200 2
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palex, Ashwini Palex, Ashwini Pan, Ho-hsien Pan, Jingjing Pan, Jingjing Pan, Yiqian Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Papadourakis, Vasileios Papadourakis, Vasileios Papagari, Raghavendra <td>Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-5-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-1 Thu-A-V-1-11 Tue-M-V-3-3 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-SS-17 Fri-M-V-7-13 Wed-A-SS-17 Fri-M-V-7-13 Wed-A-SS-10 Fri-M-V-6-10 Fri-M-V-5-8 Thu-A-SS-2-10 Fri-M-V-5-5 </td> <td>219 77 126 215 250 97 204 92 204 92 202 164 260 201 169 202 111 253 81 128 239 126 181 219 123 181 219 202 111 253 181 219 202 111 253 181 215 209 204 92 204 111 253 181 215 204 92 204 111 253 181 215 204 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 202 111 215 204 111 215 202 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 211 215 204 211 215 204 211 215 204 211 215 215 215 215 215 215 215 215 215</td>	Wed -A -V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-5-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-1 Thu-A-V-1-11 Tue-M-V-3-3 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-SS-17 Fri-M-V-7-13 Wed-A-SS-17 Fri-M-V-7-13 Wed-A-SS-10 Fri-M-V-6-10 Fri-M-V-5-8 Thu-A-SS-2-10 Fri-M-V-5-5	219 77 126 215 250 97 204 92 204 92 202 164 260 201 169 202 111 253 81 128 239 126 181 219 123 181 219 202 111 253 181 219 202 111 253 181 215 209 204 92 204 111 253 181 215 204 92 204 111 253 181 215 204 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 202 111 215 204 111 215 202 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 111 215 204 211 215 204 211 215 204 211 215 204 211 215 215 215 215 215 215 215 215 215
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakan Nodeh, M. Palakapilly, Joseph Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palaz, Dimitri Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Jing Pan, Yiqian Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Papadourakis, Vasileios Pappagari, Raghavendra Parada-Cabaleiro, Emilia Parada-Cabaleiro, Emilia	Wed -A-V-6-2 Fri-M-O-1-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-1 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-1 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-111 Tue-E-V-4-8 Thu-A-V-17 Thu-A-V-17 Thu-A-V-17 Thu-A-V-17 Thu-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-12 Thu-M-V-2-17 Wed-M-V-6-4 Wed-M-V-6-10 Fri-M-V-7-13 Wed-M-V-6-10 Fri-M-V-2-9 Wed-A-V-2-4 Thu-A-V-5-8 Thu-A-V-5-8 Thu-M-V-2-9 Wed-A-V-2-4 Thu-A-SS-210 Fri-M-V-2-5	219 77 126 146 215 152 250 97 204 92 260 211 67 169 202 211 67 169 202 211 17 3 3134 218 152 250 97 204 260 211 67 169 202 211 167 169 202 211 178 215 250 97 204 260 211 275 260 97 204 260 211 275 260 204 260 211 275 260 204 260 201 204 260 201 204 260 201 205 204 205 204 205 204 205 205 204 205 205 205 205 205 205 205 205 205 205
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palakar, Ashwini Palekar, Ashwini Palexar, Ashwini Palexar, Ashwini Paner, Ian Pan, Jing Pan, Jing Pan, Yiqian Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Papagari, Raghavend	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-6-11 Tue-E-V-4-4 Fri-A-SS-2-12 Thu-M-V-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Fri-M-V-7-13 Wed-A-V-2-12 Thu-M-V-2-12 Thu-M-V-2-13 Wed-M-V-6-10 Fri-M-V-6-10 Fri-M-V-6-10 Fri-M-V-2-17 Wed-M-V-6-10 Fri-M-V-2-9 Wed-A-V-2-4 Thu-M-S&2-10 Fri-M-V-2-5 Thu-A-SS-2-10 Fri-M-V-5-5 Thu-M-0-2-3 <	2199 777 1266 215 2500 97 204 92 204 92 204 92 204 92 204 97 204 92 202 111 253 81 202 211 67 169 202 211 113 215 239 216 239 216 239 216 215 201 97 204 97 204 97 204 92 204 97 204 97 204 92 204 97 204 111 253 81 215 202 211 113 215 202 211 113 215 202 211 113 215 202 211 113 215 202 211 113 215 202 211 215 215 215 215 215 215 215 215 21
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palakar, Shruti Palekar, Ashwini Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Yilin Pan, Yilin Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Papadourakis, Vasileios<	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Fri-A-V-46 Fri-A-V-47 Thu-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-8 Thu-A-V-1-7 Thu-A-SS-2-7 Wed-A-SS-17 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-SS-17 Fri-M-V-7-13 Wed-M-V-2-12 Thu-A-SS-2-7 Wed-A-SS-1-7 Fri-M-V-2-13 Wed-M-V-2-12 Thu-M-SS-10 Fri-M-V-5-10 Fri-M-V-5-10 Fri-M-V-2-10 Fri-M-V-2-10 Fri-M-V-2-10 Fri-M-V-2-10 Fri-M-V-2-10 Fri-M-V-2-10 Fri-M-V-2-10 Fri-M-V-2-2 Wed-A-V-2-3 Thu-A-SS-2-10 Fri-M-V-2-3	219 777 126 146 215 152 2500 97 204 92 200 211 67 260 211 67 260 211 67 260 211 169 202 111 213 181 253 181 253 181 253 200 211 67 169 202 111 215 250 201 204 201 201 204 201 201 201 201 201 201 201 201 201 201
Ostermann, JörnOttl, SandraOu, WenjieOu, ZhijianOuyang, BeibeiOuyang, BeibeiPPadfield, DirkPalakapilly, JosephPalakapilly, JosephPalakar, AshwiniPalekar, AshwiniPalekar, AshwiniPalekar, AshwiniPalekar, AshwiniPalekar, AshwiniPalekar, AshwiniPan, IingPan, JingPan, JingPan, JingPan, JingPan, YiqianPanariello, MichelePanda, RameswarPandey, AbhishekPandey, Prem C.Pang, RuomingPantazis, YannisPantic, MajaPapadourakis, VasileiosPapagari, RaghavendraParada-Cabaleiro, EmiliaParaskevopoulos, G.Parcollet, Titouan	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-M-V-1-13 Wed-A-V-2-12 Thu-A-V-5-13 Wed-M-V-2-12 Thu-M-V-2-13 Wed-M-V-2-12 Thu-M-V-2-13 Wed-M-V-2-12 Thu-M-V-2-13 Wed-M-V-2-14 Thu-M-V-2-17 Wed-M-V-6-10 Fri-M-V-5-8 Thu-M-V-2-9 Wed-A-V-2-4 Thu-A-SS-2-10 <t< td=""><td>219 77 126 215 250 97 204 260 260 260 202 111 253 81 260 202 111 213 181 218 156 239 126 81 81 133 134 129 219 133 134 219 219 219 219 219 204 260 201 204 260 202 111 253 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 204 200 201 113 181 215 200 202 111 215 200 202 211 67 202 211 67 202 211 67 202 211 67 202 211 215 202 211 173 215 202 211 215 211 215 212 202 211 215 212 212 212 212 212 212 212 21</td></t<>	219 77 126 215 250 97 204 260 260 260 202 111 253 81 260 202 111 213 181 218 156 239 126 81 81 133 134 129 219 133 134 219 219 219 219 219 204 260 201 204 260 202 111 253 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 169 202 211 67 204 200 201 113 181 215 200 202 111 215 200 202 211 67 202 211 67 202 211 67 202 211 67 202 211 215 202 211 173 215 202 211 215 211 215 212 202 211 215 212 212 212 212 212 212 212 21
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palakapilly, Joseph Palaskar, Shruti Palaexa, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Jing Pan, Jing Pan, Jing Pan, Yiqian Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Papadourakis, Vasileios <	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-9 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-13 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-A-SS-17 Fri-M-V-7-13 Wed-A-SS-17 Fri-M-V-7-13 Wed-A-SS-10 Fri-M-V-6-10 Fri-M-V-6-10 <t< td=""><td>219 77 126 146 215 250 97 204 92 204 92 202 164 202 111 253 81 125 239 126 125 209 202 111 253 81 125 239 126 146 239 126 127 202 111 253 81 125 209 204 92 204 111 253 81 125 209 204 111 253 81 125 209 204 111 253 81 125 209 204 111 253 181 229 206 111 253 181 245 209 126 127 204 111 253 181 222 106 111 233 1134 229 210 113 213 215 215 215 215 215 215 215 215 215 215</td></t<>	219 77 126 146 215 250 97 204 92 204 92 202 164 202 111 253 81 125 239 126 125 209 202 111 253 81 125 239 126 146 239 126 127 202 111 253 81 125 209 204 92 204 111 253 81 125 209 204 111 253 81 125 209 204 111 253 81 125 209 204 111 253 181 229 206 111 253 181 245 209 126 127 204 111 253 181 222 106 111 233 1134 229 210 113 213 215 215 215 215 215 215 215 215 215 215
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Pahlavan Nodeh, M. Palakapilly, Joseph Palakapilly, Joseph Palakapilly, Joseph Palakar, Shruti Palaz, Dimitri Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Palekar, Ashwini Pan, Changjie Pan, Changjie Pan, Ho-hsien Pan, Jing Pan, Jing Pan, Shifeng Pan, Yiqian Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Abhishek Pandey, Ragnayaning Pantazis, Yannis Papagari, Raghavendra Parada-Cabaleiro, Emilia Paracollet, Titouan	Wed-A-V-6-2 Fri-M-O-1-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-1 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-4 Fri-A-V-5-8 Thu-A-V-111 Tue-E-V-4-4 Fri-A-V-5-8 Thu-M-V-17 Thu-A-V-17 Wed-E-V-2-7 Fri-M-V-7-13 Wed-M-V-2-12 Thu-M-V-2-8 Tue-E-V-5-5 Wed-M-V-6-4 Wed-M-V-6-4 Wed-M-V-6-4 Wed-M-V-2-9 Wed-A-V-2-9 Wed-A-V-2-9 Wed-A-V-2-9 Wed-A-V-2-4 Thu-A-SS-2-10 Fri-M-V-5-5 Thu-A-0-2-3 Wed-M-0-1-3 Wed-M-0-1-4	219 77 126 146 215 152 250 97 204 92 260 211 67 169 202 211 67 169 202 211 17 253 181 218 152 260 211 67 169 202 211 167 169 202 211 17 215 250 97 204 260 211 167 260 211 167 260 201 215 250 97 204 260 201 167 260 201 201 201 201 201 201 201 201 201 20
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palakar, Ashwini Palekar, Ashwini Palexar, Ashwini Palexar, Ashwini Pander, Ian Pan, Jingjing Pan, Yiqian Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Papagari, Raghavendra Pa	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-SS-2-15 Thu-A-V-6-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-6-11 Tue-E-V-4-4 Fri-A-SS-2-17 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Fri-M-V-7-13 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-A-SS-1-7 Wed-M-V-2-12 Thu-M-V-2-13 Wed-B-V-2-12 Thu-M-V-2-13 Wed-M-V-2-14 Thu-M-S&T-17 Wed-M-V-6-40 Wed-M-V-6-10 Fri-M-V-2-30 Wed-A-V-2-41 Thu-M-S&Z-210 Fri-M-V-5-55	2199 777 1266 2157 2250 97 204 92 204 92 202 211 67 169 202 211 67 169 202 211 171 253 81 218 156 2399 1266 184 1152 219 210 181 215 215 209 77 204 97 204 97 204 97 204 97 204 97 204 97 204 97 204 201 1152 200 97 204 201 1152 200 97 204 201 1152 200 97 204 201 1152 200 97 204 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1152 200 201 1153 201 202 201 1153 202 202 1111 203 1154 202 201 1153 202 202 1153 1153 1154 202 202 1153 1154 202 202 1153 1154 202 201 1153 1154 202 201 1153 1154 202 201 1153 1154 202 201 1153 1154 202 201 1153 1154 202 201 1153 1154 202 201 1153 1154 202 201 1157 202 201 1157 202 201 1157 202 202 202 1157 1157 202 202 202 1157 1157 202 202 1157 1154 202 202 1157 1154 202 202 1157 1154 202 201 1157 202 201 201 201 201 201 201 201 201 201
Ostermann, Jörn Ottl, Sandra Ou, Wenjie Ou, Zhijian Ouyang, Beibei Ouyang, Beibei P Padfield, Dirk Palakapilly, Joseph Palakapilly, Joseph Palaskar, Shruti Palaz, Dimitri Palexar, Ashwini Palekar, Shruti Palekar, Shruti Palekar, Shifeng Pan, Jing Pan, Jingjing Pan, Shifeng Pan, Yiqian Pan, Yiqian Pandey, Abhishek Pandey, Abhishek Pandey, Prem C. Pang, Ruoming Pantazis, Yannis Papadourakis, Vasileios Pappagari, Raghavendra Parada-Cabaleiro, Emilia Paraskevopoulos, G. Parcollet, Titouan Parida, Shantipriva <td>Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Thu-A-V-1-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-S-S-17 Fri-M-V-7-13 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-M-V-2-12 Thu-M-SS-2-70 Wed-M-V-2-12 Thu-M-SS-2-70 Wed-M-V-2-12 Thu-M-SS-2-10 Fri-M-V-5-58 Thu-M-V-6-10 Fri-M-V-5-13 Wed-M-V-6-10 Fri-M-V-5-58 Thu-M-O-2-3</td> <td>219 77 126 219 77 126 215 250 97 204 92 200 211 67 169 202 111 253 81 12 239 126 181 239 126 181 239 126 181 215 239 126 181 215 209 77 204 92 202 111 250 202 111 253 202 112 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 10 112 202 10 112 202 10 112 202 10 112 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 10 202 10 10 202 10 10 10 202 10 10 10 10 10 10 10 10 10 10 10 10 10</td>	Wed-A-V-6-2 Fri-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-3 Wed-A-V-3-6 Thu-A-SS-1-8 Wed-A-V-3-6 Tri-A-V-4-6 Tue-A-SS-2-4 Thu-A-V-2-9 Tue-A-V-5-8 Wed-E-V-2-7 Fri-A-V-5-8 Thu-A-V-1-11 Tue-M-V-3-3 Wed-E-V-4-8 Thu-A-V-1-11 Tue-E-V-4-4 Fri-A-V-5-8 Thu-A-SS-2-7 Wed-S-S-17 Fri-M-V-7-13 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-A-SS-2-7 Wed-M-V-2-12 Thu-M-SS-2-70 Wed-M-V-2-12 Thu-M-SS-2-70 Wed-M-V-2-12 Thu-M-SS-2-10 Fri-M-V-5-58 Thu-M-V-6-10 Fri-M-V-5-13 Wed-M-V-6-10 Fri-M-V-5-58 Thu-M-O-2-3	219 77 126 219 77 126 215 250 97 204 92 200 211 67 169 202 111 253 81 12 239 126 181 239 126 181 239 126 181 215 239 126 181 215 209 77 204 92 202 111 250 202 111 253 202 112 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 111 215 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 117 212 202 10 112 202 10 112 202 10 112 202 10 112 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 202 10 10 10 202 10 10 202 10 10 10 202 10 10 10 10 10 10 10 10 10 10 10 10 10

Park, Byeongseon Thu-M-V-5-1 188 Park, Hyoungmin Wed-M-V-3-2 126

INTERSPEECH 2021 272 BRNO | CZECHIA

Park, Hyun-Jin	Tue-M-V-6-3	73	Pinto, Lancelot	Tue-A-SS-2-1 96	Qiu, David	Fri-M-V-3-9	229
Park, Kiyoung	Tue-A-S&T-1-6	99	Pirogova, Elena	Thu-A-SS-2-8 218	Qiu, Yuanhang	Tue-M-V-3-11	68
Park, Kyumin	Fri-A-V-5-1	251	Pitkow, Xaq	Thu-A-V-3-7 206	Qu, Xiaoyang	Tue-A-O-1-5	80
Park. Sung-Un	Tue-M-V-6-9	74	Plakal, Manoj	Fri-A-SS-2-14 260		Tue-A-V-1-11	84
Park. Yong-Hwa	Tue-M-V-1-1	62	Plantinga, Peter	Tue-M-V-3-9 68		Wed-M-V-3-9	127
Park, Youngcheol	Wed-M-V-4-13	130	Plaut. Florian	Tue-A-S&T-1-5 99		Wed-E-O-1-4	158
Park. Youngo	Wed-M-V-4-3	128	Podluzhny. Ivan	Fri-M-V-3-2 227		Thu-A-V-5-1	209
Parnamaa, Tanel	Fri-A-SS-1-1	255	Pokorny, Florian B.	Tue-E-O-3-3 102	Quatieri, Thomas F	Tue-M-O-2-1	60
Parrell Benjamin	Thu-M-V-6-8	191	Polyak, Adam	Thu-A-V-5-6 210	Quillen, Carl	Fri-M-O-1-1	219
Parry Jack	Wed-F-V-2-7	164	Ponti Moacir Antonelli	Tue-M-SS-1-4 78	,		
Patel Rita R	Tue-F-SS-1-5	117	rond, Mouen Antonem I	Thu-A-V-5-12 211	R		
Patel Shwetak	Tue-F-V-3-11	110	Poon, Iosiah	Fri-A-V-6-10 255	D. Niemala	T A ((2 1	00
Pathak Sheetal K	Thu-M-S&T-1-7	197	Pooriam, Amir H.	Tue-A-SS-2-11 98	K., Nirmaia	Tue-A-SS-2-1	90
Patil Hemant A	Wed-M-V-4-11	129	Popien Claudia-I	Tue-M-SS-1-5 78	Radiar, Martin	Tue-M-V-5-7	110
Patil Vaishali	Tue-M-V-7-4	75	Porter Eszter	Τμε-Δ-V-3-7 88	Dodová Vlasta	IUE-E-V-4-2	110
Patino, Jose	Tue-A-SS-1-1	94	Portera, Saige	Wed-A-V-1-7 142	Raduova, viasta	wed-A-5&1-1-2	161
rumo, jose	Tue-A-SS-2-2	96	Portet François	Wed-M-0-1-5 119	Ragilavali, Sriilivasa	Wed-E-V-I-7	101
	Tue-F-V-1-13	105	Post Matt	Thu: A.V.6.2 211	Kaj, Bliikslia	1ue-A-V-1-12	04
	Wed-F-O-1-2	158	Potamianos Alexandros	Thu-M-0-2-3 178	Dai Daab	Fri-A-V-4-11	231
	Fri-M-V-7-2	237	Potdar Nihal	Tue-F-V-4-3 110	kaj, Desn	Wed-M-O-3-1	
	Fri-M-V-7-13	239	Potharaju Srividya P	Tue-M-V-4-1 69		Wed-E-O-T-T	137
Patrocinio. Danilo	Thu-M-V-1-4	180	Pounlier Marianne	Eri-M-V-2-7 226		Thu-M-O-3-6	179
Pattee. Garv L.	Fri-A-SS-2-2	257	Poupon Nicolas	Tue-A-S&T-1-5 99	Doi D.C. Drithri	Inu-A-V-4-3	207
Paul. Dipivoti	Thu-A-V-5-8	210	Pouthier Bantiste	Wod-E-0-2-2 159	Raj, K.G. Pfiuivi	Wed-M-SS-1-6	157
Paul. Soumava	Fri-M-V-5-6	232	Povey Daniel	Thu-A-V-6-5 212	Rajpuria, Akasn	Wed-E-V-I-7	101
Paulus, Jouni	Fri-M-O-3-3	222	rovey, Damer	Thu A V C 12 214	Raju, Anirudh	Thu-A-V-2-7	203
Pautler, David	Wed-A-V-1-7	142	Drabbayallar Dobit	1 nu-A-V-0-15 214	Rallabandi, Sai Sirisha	Wed-E-V-6-7	1/4
,	Fri-A-SS-2-2	257		Eri-A-V-3-6 248	Ramabhadran, Bhuvana	I ue-A-V-4-7	90
Pavy-Le Traon, Anne	Tue-M-O-2-3	61	Prajapati Cauri P	Wod-M-V-4-11 120		Wed-M-V-6-11	135
Pede. Silvia	Thu-M-V-1-4	180	Prakash Anusha	Wod E V 1 6 161		Wed-E-V-3-6	166
Peinado, Antonio M	Tue-A-SS-2-2	96	Prandoni Paolo	Wod E O 2 4 160		Wed-E-V-3-7	166
Peiró-Lilia. Alex	Thu-M-S&T-1-3	196	Presed Amputhe	Thu M SS 2 4 105	Demelaishaan Couseh	Fri-A-SS-2-11	259
Peirolo, Morgane	Wed-M-O-2-4	120	Prasad Lodagala VSVD	Wod E V 1 6 161	Ramakrisnnan, Ganesn .	Wed-A-V-I-I	141
Pelecanos, Jason	Tue-F-V-1-6	104	Pracad Manaca	Wod E V 2 6 166	Raman, B.	Wed-M-V-5-10	132
Terecarios, juson mini	Fri-A-V-4-4	250	Pracad DaviShankar	Tuo M O 2 2 62	Kamanarayanan, vikram	Wed-A-V-I-7	142
Pelemans, Ioris	Tue-M-V-4-1	69	Prasanna S.P. Mahadaya	Wod M V 1 8 123	Demash C	Fri-A-SS-2-2	257
Pellegrini, Thomas	Tue-M-SS-1-6	78	Flasainia, S.K. Manaueva		Ramesn, G	Wed-M-V-1-6	123
,	Fri-M-O-1-5	220	Pratan Vineel	$T_{\rm Ho}M_{\rm V}$	Ramirez-de-la-Rosa, G	Wed-A-O-3-5	140
Peñarrubia-María, M.T.	Tue-A-V-2-6	85		Tue A V 4 4 80	Ramoji, Snreyas	1ue-A-SS-2-1	96
Peng. Baolin	Tue-F-V-4-1	110	Dražál Aloč	IUE-A-V-4-4 09	Ranjan, Yatharth	Tue-A-V-2-6	85
Peng Junyi	Tue-A-O-1-5	80	riazak, Ales		Rao, Hong	Thu-M-V-7-2	192
i eng, juniji i i i i i i i i i i i i i i i i i i	Wed-F-O-1-4	158	Precioso Frederic	Wod-E-0-2-2 150	Rao, K. Sreenivasa	Fri-M-V-5-6	232
Peng, Linkai	Fri-A-V-1-9	244	Dreet Supreet	Thu-M-0-2-6 170	Rao, Milind	<u>Thu-A-V-2-7</u>	203
Peng Renhua	Fri-A-SS-1-5	256	Proctor Michael	Tuo E-SS-1-2 117	Rao, Preeti	<u>Tue-M-V-7-4</u>	120
Peng Wei	Tue-F-V-4-9	112	Prokonalo Veyhenii	Tuo-A-S&T-1-5 00	Rao, Wei	Wed-M-SS-1-4	136
Peng Vivian	Wed-E-V-5-4	171	Prom-on Santitham	Thu M V 6 6 101	Rasanen, Okko	<u>Thu-M-V-2-6</u>	183
Peng Vizhou	Wed-M-V-1-7	123	Prud'hommeaux Emily				200
Peng Yu-Huai	Tue-E-V-6-2	114	Deutka Josef	Wod A S&T 1 2 156	Rascol, Olivier	<u>Iue-M-O-2-3</u>	01
reng, ru muu	Thu-M-V-2-2	186	Peutka, Josef V	Wed A S&T 1 2 156	Rashid, Zulqarnain	Tue-A-V-2-6	85
	Thu-M-V-5-0	210	1 Sutka, JOSEI V	Eri-A-0-2-4 242	Rastrow, Ariya	Iue-A-V-4-2	89
Peng Zhendong	Fri-M-V-3-5	228	Pu lie	Tue-E-V-1-10 105		IUE-E-V-4-2	
Peng Zhivuan	Eri-M-V-7-6	237	Pucher Michael	Tue-M-0-1-1 60		Wed-E-V-3-2	105
Penney Joshua	Tuo.F.SS.1.1	117	Puertolas Adrien	Τμο-Δ-S&T-1-5 99		Inu-A-V-2-7	203
Tenney, Joshua	Tue-E-SS-1-7	117	Pugliatti Maura	Tue-M-0-2-5 61	Dath Chalti D	Fri-A-V-3-7	248
Penerkamn Sharon	Eri.M.V.2.0	226	Pulman Stenhen	Tue-F-V-4-4 111	Kauli, Sllaku P	Tue-E-V-5-5	113
Penino Leonardo	Thu-A-V-1-7	201	Purin Mariu	Fri-A-SS-1-1 255	Dathod Alcohada	Wea-E-V-5-7	107
repino, reonardo	Thu-A-SS-2-4	217	Purushothaman A	Wed-M-SS-1-6 137	Ratilou, AKSildud		197
Penlinski Jacob	Tue-F-V-3-11	110	Purver Matthew	Wed A.V.1.6 142	Ratilal ajali, Alitoli		
Perenii Satheesh Kumar	Wed-M-V-2-1-1	125	i ai ver, maturette	Thu-A-SS-2.0 218	Kavanem, Mirco		00 207
Derez Matthew		118	Pusateri Frnest	Wed-A-V-2-12 144		Thu-A-V-4-4	207
ו נוכב, שומנוווכיי	Wod-A-O-2-1	130	Pusz Piotr	Tue-E-0-1-1 100	Davichandran V		249
Pérez-Conzález-de-M A	Wed A V 6 10	157	1432,1104	Tue-E-0-1-2 100	Day Swayambhu Nath	Thu AV 2 7	249
Terez Gonzalez de M, A	Thu-A-V-6-10	213	Putze, Felix	Tue-E-O-3-2 102	Ray, Swayambinu Natii Daza Agba Ali		203
Pérez-Toro PA	Tue-M-SS-1-3	77	Pvlkkönen. Janne	Wed-A-O-1-5 138	Raza, Agna An Poddi Vijev Jenene	Fri-M-V-6-5	234
	Thu-A-SS-2.2	217			Reddy Chandan V A	Wod E SS 1 1	175
Pérez Zarazaga Pablo	Tue-A-V-3-5	87	0		Reddy, Dradoon	Thu M V 1 11	182
Pernkonf. Franz	Wed-A-S&T-1-3	157		F 1444 - 0000	Dehman Juana	Mod A-V 1-4	1/1
remopi, mult mini	Fri-A-SS-1.2	256	Qazi, insan Ayyub	Fri-M-V-5-4 232	Reinhall, IValla	Tuo E-O 2-2	141
Perrotin Olivier	Tue-M-O-1-3	60	Qenaro, Lorena	Thu-M-V-1-9 181	Remus Staffer	Thu M CET 1-1	102
Peterson, Kav	Fri-M-SS-1-1	239		Tue-M-V-1-2 63	Ren Vinlei	Wod M SS 1-1-1	130
Petridis. Stavros	Thu-M-V-2.0	184	QI, JINZI	Wed-A-O-3-3 140		Wod E SS 1 -7	176
Pevser, Cal	Wed-A-V-2-6	143	Qiali, Fafi	Thu-A-V-1-4 200	Ren Vi	Wed-M-V-4-4	128
Pfeifenberger, Lukas	Wed-A-S&T-1-3	157	Qiali, KalZili	TUE-E-V-5-8 113	iccii, ii	Wed-E-V-6-6	174
Treaction Serger, Eukuo	Fri-A-SS-1-2	256	Oian Varmin	Wed-E-V-5-2 1/1	Ren 7hao	Tuo-E-O-2-2	102
Pham. Ngoc-Quan	Wed-F-V-1-2	160	Qian, ranmin	Tue-E-V-5-2 112	Ren Zongze	Fri-M-V-5-2	232
Pham. Van Tung	Wed-M-V-1-7	123		wed-A-O-2-4 139	Renals Steve		71
	Fri-M-V-5-12	233		wed-A-V-5-10 152	Kenuis, steve	Τμο-Λ-Ο-1-1	20
Philip, Pierre	Wed-E-V-2-1	163		wed-A-SS-1-8 156		Τμε-Δ-V-2-9	86
Phung, Mv	Fri-A-V-6-6	254	Oian Vurin	Thu-M-V-2-11 184		Wed-E-V-2-8	165
Picheny, Michael	Wed-M-V-2-12	126	Qiali, ruxin	Thu-A-V-3-10 206	Renkens Vincent	Tue-M-V-4-1	60
, menuel	Thu-M-V-2-9	184	Qiau, 1u		Řezáčková Markóta		60
	Fri-A-V-6-6	254	Oin James	FIT-A-V-T-T0 244	NCZackova, Maikela	Fri-A-S&T-1-2	260
Picone Jordan	Thu-A-V-6-3	211	Qiii, Jaiiits Oin Tao	Thu A V 2 1 202	Rezagholizadeh Mehdi	Wed-A-V-2-12	147
Pilati, Laurent	Wed-E-O-2-3	159	QIII, 1au		Ribeiro Manuel Sam		141 AR
Pillot-Loiseau Claire	Tue-A-0-2-1	81	Oin Viacui	FII-A-V-5-6 232	Ribeiro Vinicius	Thu-A-V-2-8	100
Pino. Juan	Wed-A-V-6-3	153	дш, лаоуг	Wod A SS 1 5 156	Riccardi Ciusenne	Wod-E-V-2-6	162
, , , , , , , , , , , , , , , , ,	Wed-A-V-6-4	153		wea-A-55-1-5 150	Mecanal, Gluseppe	weu-E-V-2-6	103
Pinquier Julien	Fri-M-0-1-5	220	Oin Ving	Fri-M-V-6-10 235			
i inquici, junch			үш, тш <u>у</u>	Thu-M-V-5-7 190			

INTERSPEECH 2021 273 BRNO | CZECHIA

Richmond, Korin	Tue-A-V-2-8 86	Sailor, Hardik	Wed-E-V-1-9	162	Schultz, Tanja	Tue-E-O-3-2	102
Richter. Richard	Wed-E-V-6-11 175 Wed-A-O-3-1 139	Sainath, Tara N	Wed-M-V-6-4 Wed-A-V-2-6	133		Wed-M-O-3-2 Wed-E-V-2-9	164
Riedhammer, K	Thu-A-SS-2-2 217		Wed-A-V-3-4	145		Fri-M-O-3-4	222
Riera, Pablo	Thu-A-V-1-7 201		Wed-E-V-3-8	166	Schuster, M	Thu-A-SS-2-2	217
Riezler. Stefan	Tue-E-V-5-7 113	Saito. Daisuke	Fri-A-V-3-8	248	Schwab, Didler	wed-м-О-1-5 Tue-A-V-2-7	86
Rigal, Rémi	Thu-A-V-3-13 207	Saito, Susumu	Wed-M-V-5-5	131	Schwartz, Boaz	Thu-A-V-3-1	204
Rikhye, Rajeev	Fri-M-V-6-3 234	Saito, Yuki	Wed-M-V-3-6	127	Schwartz, Jean-Luc	Thu-A-O-1-5	199
Riley, Michael	Fri-M-V-3-3 228	Sakakibara Ken-Ichi	<u>Ihu-A-V-2-4</u> Thu-M-V-6-9	203	Schwarz, Andreas	Wed-M-V-4-8	129
Rilliard. Albert	Fri-M-O-2-5 221	Sukukiburu, Ken Iem	Fri-A-S&T-1-1	260	Schymura, Christopher .	Wed-A-V-4-2	148
Ringeval, Fabien	Wed-M-O-1-5 119	Sakallah, Alaa	Wed-A-V-1-7	142	Sculley, D	Fri-A-SS-2-10	259
Ristea, Nicolae-Cătălin	<u>Thu-M-O-1-1</u> 177	Sakamoto, Shoki	Tue-E-V-6-8	115	Seaver, Katie	Fri-A-SS-2-1	257
Riverin-Coutlée. Iosiane	Wed-M-O-2-2 120	Sukti, Sukrium	Wed-A-V-6-6	153		Fri-A-SS-2-10	259
Rivière, Morgane	Wed-M-V-2-10 125		Wed-A-V-6-7	153	Sebkhi, Nordine	Tue-A-V-2-10	86
Rizos, Georgios	Tue-M-SS-1-9 78		Wed-A-V-6-9	154	SedImeier, Andreas	Tue-M-SS-1-5	78
Ro Jae	Tue-A-V-6-4 93	Sakuma, Jin	Thu-M-V-7-4	192	Sekiguchi, Kouhei	Tue-A-V-3-2	87
Roberts, Angela	Tue-E-SS-1-6 118	Salakhutdinov, Ruslan .	Tue-A-V-5-8	92	Seltzer, Michael L	Wed-M-V-6-3	133
	Wed-A-O-3-1 139	Salesky, Elizabeth	Thu-A-V-6-2	211		Wed-A-V-2-1	142
Rocholl, Johann C	Tue-A-V-5-3 91	Sallustio, Vincenzo	ти-м-5&т-т-6 Тhu-M-V-1-4	197		Wed-A-V-3-1 Wed-A-V-3-14	145 147
KUCDEI, AXEI	Thu-M-O-2-1 178	Salvati, Daniele	Wed-M-O-3-4	121		Fri-A-V-3-6	248
Roesler, Oliver	Wed-A-V-1-7 142	Companyla Toll's	Wed-A-V-4-8	149	Conorrinoter NT- 1	Fri-A-V-3-10	249
Pohanian Mortoza	Fri-A-SS-2-2 257	Samarakoon, Lahiru	<u>Πημ-Α-Λ-Λ-4-7</u> Τμε-Δ-Λ-2-10	208 86	Seneviratne, Nadee	Wed-E-V-2-11 Fri-A-V-2-8	164 246
Rohdin, Johan	Tue-A-0-1-3 80	Sanabria, Ramon	Thu-M-V-2-10	182	Sensi, Mariachiara	Tue-M-O-2-5	61
Rohmatillah, Mahdin	Thu-M-V-7-3 192	Sanchez, Enrique	Thu-A-V-1-5	201	Seo, Hyeji	Wed-E-V-5-13	173
Rolland, Thomas	Tue-M-SS-1-2 77	Sanchis, Albert	Wed-A-V-6-10	154 213	Seo, Soonshin	Tue-A-V-1-8 Fri-M-V-1-2	83 222
Romana, Amrit	Tue-E-SS-1-6 118	Sancinetti, Marcelo	Fri-A-V-0-10	243	Serrano, Mauricio	Wed-E-V-3-12	167
	Wed-A-O-3-1 139	Sandler, Mark	Thu-A-V-3-8	206	Seshadri, Vivek	Wed-E-V-1-7	161
Romanenko, Aleksei	Fri-M-V-3-2 227	Sankaranarayanan, K	Wed-E-V-1-7	161	Seth, Ashish	Wed-E-V-1-6	161
Ronanki, Srikantn	Thu-M-V-7-8 193	Santoso, Jennier	Wed-A-V-1-3	141	Settiu, viunyasanaran	I nu-A-V-6-7 Fri-A-V-2-5	212
Rose, Richard	Thu-M-V-2-10 184		Wed-M-V-6-9	134	Sevastjanova, Rita	Wed-M-V-5-9	132
Roselló, Nahuel	Thu-A-V-6-10 213		Wed-A-V-2-11	144	Seyfarth, Scott	Tue-A-V-5-1	90
Rosenberg, Andrew	Tue-A-V-4-7 90		Wed-A-V-3-5	145 167	Slaklanaki, Anna	1ue-E-SS-1-3	206
Rosindell, James	Tue-M-SS-1-9 78	Saraclar, Murat	Fri-A-O-2-2	242	Shafran, Izhak	Fri-A-V-1-3	243
Rossato, Solange	Wed-M-O-1-5 119	Saraf, Yatharth	Wed-M-V-6-3	133	Shah, Muhammad A	Tue-A-SS-1-8	96
Roth, Julian	Thu-M-V-5-2 189	Sarawagi, Sunita	Wed-M-V-2-1	123	Shah, Sanket	Wed-E-V-1-7	161 212
Rothkrantz, Leon I.M.	Tue-M-SS-1-1 77	Salijoo, Seyyeu Saeeu	Fri-A-O-1-3	241	Shaik, M.A. Basha	Wed-M-SS-1-6	137
Rouas, Jean-Luc	Wed-E-V-2-1 163	Sarkar, Saurjya	Thu-A-V-3-8	206	Shamsi, Meysam	Tue-A-S&T-1-5	99
Rouditchenko, Andrew .	Wed-M-V-2-12 126	Sarma, Mousmita	Thu-M-O-3-6	179	Shandiz, Amin H	Wed-A-O-3-6	140
	Thu-M-V-2-8 184 Thu-A-V-6-1 211	Sarinan, Priyankoo	Wed-M-V-1-8	123	Shang, Shidong	Wed-M-SS-1-2 Wed-M-V-3-7	127
Rouhe, Aku	Wed-M-O-1-3 119	Sarni, Salvatore	Fri-A-V-4-3	250	- 0, 01 0	Fri-M-V-4-9	231
Routray, Aurobinda	Thu-A-V-1-9 201	Saruwatari, Hiroshi	Tue-M-V-2-3	65	Shangguan, Yuan	Wed-M-V-6-3	133
Rowe, Hannan P	Fri-A-SS-2-5 258		Wed-M-V-3-6	127		Wed-A-V-3-14 Fri-A-V-3-6	147 248
Roy, Anwesha	Tue-E-V-3-1 107	Sato, Hiroshi	Tue-E-V-2-10	107	Shao, Jing	Wed-M-V-5-6	131
D (D())	Fri-A-S&T-1-8 261	Contra Data la	Wed-M-V-6-6	134	Shao, Qijie	Fri-M-V-6-11	236
Roze, Patricia	Wed-M-V-2-10 125	Sawata, Kyosuke	<u>Γμε-Ε-V-2-7</u> Fri-Δ-V-6-9	255	Shao, Zongru	Thu-A-SS-2-12 Fri-A-V-6-4	219
Rudnicky, Alexander	Wed-M-V-2-8 125	Schamoni, Shigehiko	Tue-E-V-5-7	113	Sharma, Dushyant	Fri-M-O-1-1	219
	Thu-A-V-1-1 200	Scharenborg, Ödette	Wed-M-V-2-2	124	Sharma, Manthan	Tue-A-V-2-3	85
Rumberg, Lars	Fri-M-O-1-3 219	Scheibler, Robin	Wed-A-V-4-3	148	Sharma, Mayank	Thu-M-V-5-6	189
Ryant, Neville	Thu-A-V-4-6 208	Schiller, Dominik	Tue-M-SS-1-12	79	Shechtman, Slava	Fri-A-V-5-11	253
Rybach, David	Wed-M-V-6-4 133	0.110	Fri-M-O-2-2	220	Shekhar, Shashank	Tue-A-V-1-5	82
	Wed-A-V-2-6 143	Schluter, Ralf	Wed-A-O-1-4	138 145	Shen, Guang	Fri-A-V-2-9	247 65
Rybicka, Magdalena	Tue- <u>A-O-1-2</u> 80		Thu-M-O-1-5	177		Tue-M-V-2-7 Tue-M-V-2-9	65 66
Rychlikowski, Paweł	Tue-E-O-1-1 100	İ	Thu-M-O-3-2	179	Shen, Yuan	Fri-A-V-5-6	252
B 1 · W 1	Tue-E-O-1-2 100		Thu-M-O-3-3	179	Shenoy, Ashish	Thu-M-V-7-8	193
Kyokai, Kimiko	Wed-E-V-2-2 163	Schmitt, Manuel	Wed-E-O-3-2	159^{-44}	Sherman, Alexander V	Fri-A-SS-2-2 Thu-M-SS-2-3	257 195
Kyuiiiiia, Liena	Thu-A-V-6-9 213	Schneider, Aaron	Tue-A-V-5-3	91	Shi, Guixin	Fri-M-SS-1-5	240
Ryzhova, Margarita	Fri-M-V-2-4 225	Schönherr, Lea	Tue-A-SS-1-5	95	Shi, Huiyu	Thu-M-SS-1-3	194
S		Schuller, Biörn W.	Tue-A-V-3-1 Tue-M-SS-1-1	80 77	Sni, Jiatong	Tue-E-V-3-9	109 244
Cashaa Anda			Tue-M-SS-1-9	78	JIII, KC	Fri-A-S&T-1-6	261
Sabzi Shahrebabaki. A	Tue-E-V-3-7 109		Tue-A-V-2-6	85 02	Shi, Liang	Tue-E-V-6-9	116
Saddler, Mark R	Wed-E-V-5-2 171		Tue-A-V-6-4	95 97	Shi, Ning	Wed-A-V-2-3	143 66
Sadhu, Samik	Tue-A-V-4-2 89		Tue-A-SS-2-7	97	Sill, Lallgyallg	Wed-M-V-2-8	133
Sadiadi, Seved Omid	Tue-E-O-1-4 100		Tue-A-SS-2-9	98 102	İ	Wed-A-V-3-1	145
Saebi, Mandana	Wed-A-V-2-12 144		Tue-E-O-3-3 Thu-M-V-2-9	102		Wed-A-V-3-14	147
Saeboe, Lilja	Wed-M-O-2-3 120		Thu-M-V-5-8	190		Fri-A-V-3-6 Fri-A-V-3-10	248 249
Sagnir, Hamidreza	Fri-A-V-6-8 255 Wed-M-V-1-2 122	Schuller Dagmar M	Fri-M-V-5-5	232	Shi, Yao	Wed-E-V-6-4	173
Sumuman, Mu	Wed-M-V-4-6 128	schuner, Dagmar M		102	Shi, Yi	Fri-M-V-4-5	230
	Thu-M-SS-1-2 194				Shigabeev. Ilva	Wed-A-SS-1-1	155
	Fri-M-V-7-9 238						

Sinn, Jong Won	Tue-A-V-3-4	87
Shin, Myeong Cheol	Fri-M-V-6-8	235
Shin, Seungmin	Wed-M-V-4-13	130
Shinozaki, Takahiro	Thu-A-V-2-1	202
Shintani, Taiken	Wed-A-V-1-8	142
Shirouzu, Hajime	Tue-M-V-6-4	73
Shon, Suwon	Thu-A-V-1-11	202
Shor, Joel	Tue-E-V-3-11	110
Chuoi Wong	Fri-A-SS-2-14	260
Shully, Wang	Tue-E-V-6-10	110
Shulman Michael D		110
Si Shijing	Tue-A-V-1-11	84
5i, 5iijiiig	Wed-M-V-3-9	127
Si, Yuke	Tue-M-V-4-7	70
Siddhartha, Nikhil	Tue-A-V-4-6	89
Siddi, Sara	Tue-A-V-2-6	85
Siegert, Ingo	Wed-M-V-4-9	129
Sigtia, Siddharth	Fri-M-V-6-4	234
Silpachai, Alif	Wed-A-V-1-4	141
Silva, Daniel P.P. da	Tue-M-SS-1-4	/8
Silva, Samuel	I ue-A-V-2-9	154
Silvestie-Cerua, JA	Thu-A-V-6-10	213
Sim Khe Chai	Tue-A-V-4-6	89
	Tue-E-V-5-4	113
Simantiraki, Olympia	Wed-M-V-5-4	131
Simblett, Sara	Tue-A-V-2-6	85
Simha, Pramod	Fri-M-V-6-4	234
Siminyu, Kathleen	Tue-M-V-5-2	70
	Thu-A-V-6-3	211
Simonchik, Konstantin .	Wed-A-SS-1-1	155
Singh, Abhayjeet	I UE-E-V-3-I	107
Singh Ankita	Fri-A-5&1-1-8	201
Singh Dinesh K	Wed-M-V-4-11	129
Singh, Maneesh	Tue-A-SS-2-11	98
Singh, Prachi	Thu-A-V-4-1	207
	Thu-A-V-4-6	208
Singh, Rita	Tue-A-V-1-12	84
	Fri-M-V-5-11	233
	Fri-A-V-4-11	251
Singh, Satwinder	Tue-M-V-3-11	68
Shirscalchi, Sabato M	Tue-A-55-1-0	100
Siohan. Olivier	Wed-M-V-6-10	134
- · · , - · ·		
	Thu-M-V-2-10	184
Sirojan, T	Thu-M-V-2-10 Thu-A-V-6-7	184 212
Sirojan, T Sisman, Berrak	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2	184 212 93
Sirojan, T Sisman, Berrak Sitaram Sunavana	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7	184 212 93 252 161
Sirojan, T Sisman, Berrak Sitaram, Sunayana Siyaprasad, Sarath	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3	184 212 93 252 161 252
Sirojan, T Sisman, Berrak Sitaram, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4	184 212 93 252 161 252 85
Sirojan, T Sisman, Berrak Sitaram, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1	184 212 93 252 161 252 85 170
Sirojan, T Sisman, Berrak Sitaram, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9	184 212 93 252 161 252 85 170 88
Sirojan, T Sisman, Berrak Sitaram, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1	184 212 93 252 161 252 85 170 88 120
Sirojan, T Sisman, Berrak Sitaram, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skerry-Ryan, R.J	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-2-7	184 212 93 252 161 252 85 170 88 120 65
Sirojan, T Sisman, Berrak Sivaprasad, Sarath Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skerry-Ryan, R.J Sklyar, Ilya	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7	184 212 93 252 161 252 85 170 88 120 65 134 210
Sirojan, T Sisman, Berrak Sivaprasad, Sarath Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Simon Singen, Simon	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4	184 212 93 252 161 252 85 170 88 120 65 134 210 242
Sirojan, T Sisman, Berrak Sivaprasad, Sarath Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skerry-Ryan, R.J Sklyar, Ilya Slangen, Simon Šmíld, Luboš	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10	$184 \\ 212 \\ 93 \\ 252 \\ 161 \\ 252 \\ 85 \\ 170 \\ 88 \\ 120 \\ 65 \\ 134 \\ 210 \\ 242 \\ 226 \\$
Sirojan, T Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skiyar, Ilya Slangen, Simon Smiljanic, Rajka Snowden, Julie S	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-SS-2-7	$184 \\ 212 \\ 93 \\ 252 \\ 161 \\ 252 \\ 85 \\ 170 \\ 88 \\ 120 \\ 65 \\ 134 \\ 210 \\ 242 \\ 226 \\ 218 \\$
Sirojan, T Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaramakrishnan, V Sivasankaran, Sunit Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skiyar, Ilya Slangen, Simon Smiljanic, Rajka Snowden, Julie S Soares, Anderson da S	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-5-3 Tue-A-V-3-9 Wed-M-O-2-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-SV-2-7 Thu-A-V-5-12	$184 \\ 212 \\ 93 \\ 252 \\ 161 \\ 252 \\ 85 \\ 170 \\ 88 \\ 120 \\ 65 \\ 134 \\ 210 \\ 242 \\ 226 \\ 218 \\ 211 \\$
Sirojan, T Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarn, Simon Singen, Simon Sindl, Luboš Smiljanic, Rajka Snowden, Julie S Soares, Anderson da S Södergren, Isabella	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-A-O-2-4 Fri-A-V-2-10 Thu-A-SS-2-7 Thu-A-SS-2-4	$184 \\ 212 \\ 93 \\ 252 \\ 161 \\ 252 \\ 85 \\ 170 \\ 88 \\ 120 \\ 65 \\ 134 \\ 210 \\ 242 \\ 226 \\ 218 \\ 211 \\ 97 \\$
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skiyar, Ilya Singen, Simon Smidl, Luboš Smidl, Luboš Smidl, Luboš Smidl, Luboš Soires, Anderson da S Soares, Anderson da S Sodergren, Isabella Solera-Ureña, Rubén	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-2-7 Fri-A-O-2-4 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-SS-2-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-A-SS-2-2	184 212 93 252 161 252 85 170 88 120 65 134 210 242 226 218 211 97 77
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skiyar, Ilya Slangen, Simon Sindl, Luboš Smidl, Luboš Smidl, Luboš Smidl, Luboš Soires, Anderson da S Sodres, Anderson da S Sodres, Anderson da S Sodergren, Isabella Solera-Ureña, Rubén Solera-Company, Juan	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-2-4 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-V-5-12 Tue-A-V-5-12 Tue-A-V-5-12 Wed-A-S&-1-15	184 212 93 252 161 252 85 170 88 120 65 134 210 242 226 218 211 97 77
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Solares, Anderson da S Soders, Anderson da S Soderser, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Soler-Company, Juan	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-2-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-SS-2-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-A-SS-1-2 Wed-A-S&T-1-5 Tue-M-O-2-11	184 212 93 252 161 252 85 170 88 120 65 134 210 242 226 218 211 97 77 157 60
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skaryar, Rya Skaryar, Rya Sklyar, Ilya Singi, Cajka Soares, Anderson da S Södergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Soler-Company, Juan Soltau, Hagen Song Charache	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-2-4 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-SS-2-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-A-SS-2-12 Wed-A-S&T-1-5 Tue-M-O-2-11 Fri-A-V-2-13 Fri-A-V-2-13 Fri-A-V-2-13 Fri-A-V-2-13 Fri-A-V-2-13	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 2218\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 252\\ 252\\ 252\\ 252\\ 252\\ 252\\ 25$
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skary-Ryan, R.J Sklyar, Ilya Sklyar, Ilya Sindl, Luboš Smidl, Luboš Smidl, Luboš Songeren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Soler-Company, Juan Solomon, Nancy Pearl Song, Changhe	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-SS-2-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-A-SS-1-2 Wed-A-S&T-1-5 Tue-M-O-2-1 Fri-A-V-1-3 Fri-A-V-1-3 Fri-A-V-1-3	184 212 93 252 161 252 85 170 88 120 65 134 210 242 226 2218 211 97 77 157 60 243 252
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skary.Ryan, R.J Sklyar, Ilya Sindl, Luboš Sindl, Luboš Somiljanic, Rajka Soares, Anderson da S Södergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Soler-Company, Juan Solomon, Nancy Pearl Song, Changhe Song, Dandan	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-2-4 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-1-2 Wed-A-S&T-1-5 Tue-M-SS-1-3 Wed-A-V-5-7 Thu-A-V-5-7 Thu-A-V-5-7	184 212 93 252 161 252 85 170 88 8120 65 134 210 242 226 218 211 97 77 157 60 243 252 194
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skary-Ryan, R.J Sklyar, Ilya Sidangen, Simon Smidl, Luboš Smidl, Luboš Smidl, Luboš Soing, Changhe Song, Changhe Song, Eunwoo	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-A-O-2-4 Fri-A-V-5-12 Tue-A-SS-2-7 Thu-A-V-5-12 Wed-A-S&T-1-5 Tue-M-SS-1-2 Wed-A-S&T-1-5 Tue-M-O-2-1 Fri-A-V-1-3 Fri-A-V-5-7 Thu-M-SS-13 Wed-A-V-5-11 Thu-A-V-5-12	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 218\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 229\end{array}$
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skiyar, Ilya Sidangen, Simon Simidl, Luboš Smidl, Luboš Smidl, Luboš Smidl, Luboš Soing, Changhe Song, Changhe Song, Eunwoo Song, Hongwei	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-A-O-2-4 Fri-A-V-5-12 Tue-A-SS-2-7 Thu-A-V-5-12 Tue-A-SS-1-2 Wed-A-S&T-1-5 Tue-M-O-2-1 Fri-A-V-1-3 Fri-A-V-5-7 Thu-M-SS-1-3 Wed-A-V-5-11 Thu-A-V-5-2 Tue-E-V-1-8	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 218\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 229\\ 104 \end{array}$
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skary-Ryan, R.J Sklyar, Ilya Sindil, Luboš Sindil, Luboš Sindil, Luboš Sindil, Luboš Somiljanic, Rajka Soniljanic, Rajka Soares, Anderson da S Södergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solor-Company, Juan Solor-Company, Juan Solotau, Hagen Song, Changhe Song, Dandan Song, Eunwoo Song, Hongwei Song, Hui	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-A-O-2-4 Fri-A-V-5-12 Tue-A-SS-2-7 Thu-A-V-5-12 Tue-A-SS-1-2 Wed-A-S&T-1-5 Tue-M-O-2-1 Fri-A-V-5-7 Thu-M-SS-1-3 Wed-A-V-5-11 Thu-A-V-5-11 Thu-A-V-5-8	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 218\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 229\\ 104\\ 206\\ \end{array}$
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sivasankaran, Sunit Skarnitzl, Rajek Sight, Rajka Similjanic, Rajka Soiniljanic, Rajka Soiniljanic, Rajka Soares, Anderson da S Södergren, Isabella Solera-Ureña, Rubén Soler-Company, Juan Solomon, Nancy Pearl Soltau, Hagen Song, Changhe Song, Eunwoo Song, Hongwei Song, Hui	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-V-5-12 Tue-M-SS-2-4 Tue-M-SS-2-4 Tue-M-SS-1-2 Wed-A-S&T-1-5 Tue-M-SS-1-3 Wed-A-V-5-11 Thu-A-V-5-2 Tue-E-V-1-8 Thu-A-V-3-11 Tue-A-V-3-4	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 218\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 209\\ 194\\ 152\\ 209\\ 104\\ 206\\ 87\\ \end{array}$
Sirojan, T Sisman, Berrak Sisman, Berrak Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sivasankaran, Sunit Skarnitzl, Rajek Sidangen, Simon Simidl, Luboš Simidl, Luboš Somiljanic, Rajka Soingen, Jsabella Solera-Ureña, Rubén Soler-Company, Juan Solotau, Hagen Soltau, Hagen Song, Changhe Song, Dandan Song, Eunwoo Song, Hungwei Song, Hui Song, Meishu	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-3-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-SS-2-7 Thu-A-V-5-12 Tue-M-SS-1-2 Wed-A-S&T-15 Tue-M-O2-11 Fri-A-V-1-3 Fri-A-V-5-7 Thu-M-SS-1-3 Wed-A-V-5-11 Thu-A-V-5-2 Tue-E-V-1-8 Thu-A-V-3-14 Thu-A-V-3-4 Fri-M-V-5-5	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 252\\ 134\\ 210\\ 242\\ 226\\ 218\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 209\\ 104\\ 206\\ 87\\ 232\\ 226\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 232\\ 206\\ 87\\ 206\\ 87\\ 232\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 206\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87$
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Solare, Simon Sonden, Jluie S Soares, Anderson da S Södergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solare, Nancy Pearl Solau, Hagen Song, Changhe Song, Changhe Song, Eunwoo Song, Hongwei Song, Hui Song, Hyungchan Song, Meishu	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-3-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-M-SS-1-2 Wed-A-SS-1-2 Wed-A-SS-1-2 Wed-A-SS-1-3 Fri-A-V-1-3 Fri-A-V-5-1 Thu-A-V-5-13 Fri-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-13 Thu-A-V-3-14 Thu-A-V-3-4 Fri-M-V-3-4 Fri-M-V-3-2 Tue-M-V-3-2	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 242\\ 226\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 209\\ 104\\ 206\\ 87\\ 232\\ 66\\ 61\\ 87\\ 232\\ 61\\ 206\\ 87\\ 232\\ 61\\ 206\\ 87\\ 232\\ 206\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87$
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sivasankaran, Sunit Skarnitzl, Radek Solargen, Simon Soindl, Luboš Soindl, Luboš Soinden, Julie S Sooares, Anderson da S Södergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solog, Changhe Song, Changhe Song, Changhe Song, Hui Song, Hui Song, Hui Song, Hyungchan Song, Qi Song, Qiong	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-3-3 Wed-M-O-2-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-M-V-2-10 Thu-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-1-2 Wed-A-S&T-1-5 Tue-M-SS-1-3 Wed-A-V-5-11 Thu-A-V-5-13 Fri-A-V-3-4 Thu-A-V-3-4 Fri-M-V-3-2 Thu-A-V-3-2 Thu-A-V-3-2 Thu-A-V-3-2 Thu-A-V-3-2 Thu-A-V-3-2 Thu-A-V-3-2 Thu-A-V-3-2	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 242\\ 226\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 209\\ 104\\ 226\\ 87\\ 232\\ 66\\ 87\\ 232\\ 66\\ 214\\ 9\\ 232\\ 66\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87\\ 87$
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sklyar, Ilya Slangen, Simon Similjanic, Rajka Smiljanic, Rajka Sonoden, Julie S Sooderg, Anderson da S Södergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solau, Hagen Song, Changhe Song, Dandan Song, Hui Song, Hui Song, Hui Song, Hui Song, Meishu Song, Qi Song, Yan	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Thu-A-V-5-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-M-SS-1-2 Wed-A-S&T-1-5 Tue-M-V-5-13 Fri-A-V-5-13 Fri-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-13 Wed-A-V-3-11 Thu-A-V-5-15 Tue-A-V-3-2 Thu-A-V-5-13 Thu-A-V-5-13 Thu-A-V-5-13 Thu-A-V-5-13 Thu-A-V-5-13 Thu-A-V-6-13 Tue-A-V-1-3 Wed-A-V-1-3	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 218\\ 211\\ 97\\ 77\\ 77\\ 77\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 209\\ 104\\ 252\\ 194\\ 152\\ 209\\ 104\\ 2262\\ 194\\ 152\\ 209\\ 104\\ 2262\\ 194\\ 152\\ 209\\ 104\\ 212\\ 209\\ 104\\ 212\\ 209\\ 104\\ 212\\ 209\\ 104\\ 212\\ 209\\ 104\\ 226\\ 214\\ 82\\ 209\\ 104\\ 212\\ 209\\ 212\\ 212\\ 209\\ 212\\ 212\\ 209\\ 212\\ 212\\ 209\\ 212\\ 212\\ 212\\ 209\\ 212\\ 212\\ 212\\ 212\\ 212\\ 212\\ 212\\ 21$
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sklyar, Ilya Slangen, Simon Sindl, Luboš Similjanic, Rajka Smiljanic, Rajka Sonwden, Julie S Sonder, Julie S Sodergren, Isabella Sodergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Soltau, Hagen Song, Changhe Song, Hongwei Song, Hongwei Song, Hui Song, Hyungchan Song, Qi Song, Qiong Song, Yan	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Thu-A-V-5-7 Thu-A-V-5-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-M-SS-1-2 Wed-A-S&T-1-5 Tue-M-V-2-10 Fri-A-V-5-12 Tue-A-V-5-12 Tue-A-V-5-13 Fri-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-2 Tue-E-V-1-8 Thu-A-V-5-3 Tue-A-V-3-2 Tue-A-V-3-2 Tue-A-V-3-2 Tue-A-V-3-2 Tue-A-V-3-2 Thu-A-V-6-13 Tue-A-V-1-3 Wed-M-V-1-3 Wed-M-V-1-3	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 218\\ 211\\ 97\\ 77\\ 77\\ 60\\ 243\\ 252\\ 194\\ 152\\ 209\\ 104\\ 252\\ 194\\ 152\\ 209\\ 104\\ 82\\ 262\\ 124\\ 206\\ 87\\ 232\\ 66\\ 214\\ 82\\ 222\\ 247\\ 247\\ 247\\ 247\\ 247\\ 247\\ 24$
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sklyar, Ilya Slangen, Simon Sindl, Luboš Smiljanic, Rajka Smiljanic, Rajka Sonden, Julie S Sodergren, Isabella Sodergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Soler-Company, Juan Soltau, Hagen Song, Changhe Song, Changhe Song, Hongwei Song, Hui Song, Hui Song, Hui Song, Qi Song, Qi Song, Yan	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Thu-A-V-5-7 Thu-A-V-5-7 Thu-A-V-5-10 Thu-A-V-5-12 Wed-A-S&T-1-5 Tue-M-V-2-10 Fri-A-V-5-13 Fri-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-2 Tue-E-V-1-8 Thu-A-V-5-13 Tue-A-V-3-2 Thu-A-V-5-3 Tue-A-V-3-2 Thu-A-V-5-13 Tue-A-V-3-2 Thu-A-V-5-13 Tue-A-V-3-2 Thu-A-V-5-13 Tue-A-V-3-2 Thu-A-V-6-13 Tue-A-V-1-3 Wed-M-V-1-3 Fri-A-V-2-10 Fri-A-V-2-10 Tue-M-V-7-10	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 61\\ 218\\ 211\\ 97\\ 77\\ 76\\ 20\\ 243\\ 252\\ 218\\ 211\\ 97\\ 77\\ 76\\ 200\\ 87\\ 232\\ 66\\ 214\\ 82\\ 212\\ 247\\ 76\end{array}$
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sklyar, Ilya Slangen, Simon Sindl, Luboš Smiljanic, Rajka Smiljanic, Rajka Sonden, Julie S Sodregren, Isabella Sodergren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Soler-Company, Juan Sologer, Changhe Song, Changhe Song, Dandan Song, Hongwei Song, Hui Song, Hui Song, Hui Song, Qi Song, Qi Song, Qi Song, Yan Soong, Yan	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-3-9 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-2-7 Wed-M-V-2-7 Wed-M-V-2-7 Thu-A-V-5-7 Thu-A-V-5-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-1-2 Wed-A-SS-1-2 Wed-A-SS-1-3 Wed-A-V-5-11 Thu-A-V-5-13 Tue-A-V-3-3 Tue-A-V-3-3 Tue-A-V-3-3 Tue-A-V-3-3 Tue-A-V-3-3 Tue-A-V-3-3 Tue-A-V-3-3 Tue-A-V-3-3 Tue-A-V-3-3 Tue-A-V-3-13 Tue-A-V-3-13 Tue-A-V-1-3 Tue-A-V-1-3 Fri-A-V-2-10 Tue-M-V-7-10 Fri-A-V-5-10	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 242\\ 226\\ 61\\ 210\\ 242\\ 226\\ 218\\ 211\\ 97\\ 77\\ 75\\ 60\\ 243\\ 252\\ 218\\ 212\\ 209\\ 104\\ 252\\ 209\\ 104\\ 87\\ 232\\ 66\\ 214\\ 82\\ 247\\ 76\\ 253\\ \end{array}$
Sirojan, T Sisman, Berrak Sisman, Berrak Sivarama, Sunayana Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sklyar, Ilya Sklyar, Ilya Similjanic, Rajka Smiljanic, Rajka Smiljanic, Rajka Songeren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Soler-Company, Juan Soler-Company, Juan Solare, Changhe Song, Changhe Song, Changhe Song, Hungen Song, Hungen Song, Hungen Song, Hui Song, Hui Song, Qi Song, Qi Song, Qi Song, Yan Song, Yan Soong, Yan Soong, Frank K Sootla, Sten	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-2-4 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-6-7 Thu-A-V-5-7 Fri-A-O-2-4 Fri-A-V-2-10 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-1-2 Wed-A-S&T-1-5 Tue-M-O-2-1 Fri-A-V-5-13 Fri-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-2 Tue-E-V-1-8 Thu-A-V-5-2 Tue-A-V-3-4 Fri-M-V-5-2 Tue-A-V-3-4 Fri-A-V-6-13 Tue-A-V-1-3 Wed-M-V-1-3 Fri-A-V-2-10 Fri-A-V-5-10 Fri-A-V-5-11	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 6218\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 209\\ 104\\ 232\\ 247\\ 76\\ 60\\ 214\\ 82\\ 247\\ 76\\ 253\\ 255\\ 255\\ \end{array}$
Sirojan, T Sisman, Berrak Sisman, Sunayana Sivaprasad, Sarath Sivaramakrishnan, V Sivaraman, Aswin Sivasankaran, Sunit Skarnitzl, Radek Skarnitzl, Radek Skarnitzl, Radek Sklyar, Ilya Sklyar, Ilya Sindl, Luboš Smiljanic, Rajka Smiljanic, Rajka Sonowden, Julie S Sodregren, Isabella Solera-Ureña, Rubén Solera-Ureña, Rubén Solera-Ureña, Rubén Soler-Company, Juan Solare, Company, Juan Solag, Changhe Song, Changhe Song, Changhe Song, Hui Song, Hui Song, Hui Song, Hui Song, Qi Song, Qi Song, Yan Song, Yan Soong, Yan Soong, Frank K Sootla, Sten Sorensen, Karsten	Thu-M-V-2-10 Thu-A-V-6-7 Tue-A-V-6-2 Fri-A-V-5-2 Wed-E-V-1-7 Fri-A-V-5-3 Tue-A-V-2-4 Wed-E-V-5-1 Tue-A-V-2-4 Wed-M-O-2-1 Tue-M-V-2-7 Wed-M-V-2-7 Wed-M-V-2-7 Wed-M-V-2-7 Thu-A-V-5-7 Thu-A-V-5-12 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-2-4 Tue-A-SS-1-2 Wed-A-S&T-1-5 Tue-M-O-2-1 Fri-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-13 Wed-A-V-5-11 Thu-A-V-5-2 Tue-E-V-1-8 Thu-A-V-3-11 Thu-A-V-3-13 Wed-A-V-3-3 Thu-A-V-3-13 Wed-M-V-1-3 Fri-A-V-2-10 Tue-M-V-7-10 Fri-A-V-5-11 Fri-A-V-5-11 Fri-A-V-5-11 Fri-A-V-5-11 Fri-A-V-5-11 Fri-A-V-5-11 Fri-A-V-5-11 Fri-A-V-5-11	$\begin{array}{c} 184\\ 212\\ 93\\ 252\\ 161\\ 252\\ 85\\ 170\\ 88\\ 120\\ 65\\ 134\\ 210\\ 242\\ 226\\ 61\\ 218\\ 211\\ 97\\ 77\\ 157\\ 60\\ 243\\ 252\\ 194\\ 152\\ 209\\ 104\\ 232\\ 219\\ 104\\ 206\\ 87\\ 232\\ 266\\ 214\\ 82\\ 122\\ 247\\ 75\\ 255\\ 255\\ 255\\ 255\\ 255\\ 255\\ 255$

Sorokin, Daniil	Tue-E-V-4-7	111
	Fri-A-V-6-3	254
Spanias, Andreas	Wed-A-V-4-12	150
Spathis, Dimitris	Tue-M-SS-1-1	77
Spong, Paul	Wed-E-O-3-2	159
Sprenkamp, Kilian	Thu-M-S&T-1-4	197
Squartini, Stefano	Fri-M-O-1-4	220
Sridhar, Prashant	Wed-M-V-6-14	135
Srinivasan, Sriram	Wed-E-SS-1-1	175
	Fri-A-SS-1-1	255
Srinivasan, S.	Tue-A-V-5-1	90
,	Tue-E-V-5-10	114
Sriram, Anuroop	Tue-A-V-4-4	89
Stack. Sarah	Tue-A-V-2-7	86
Stafvlakis. Themos	Tue-A-O-1-3	80
Stahl, Daniel	Tue-A-V-2-6	85
Staib. Marlene	Fri-M-O-2-3	221
	Fri-M-O-2-4	221
Stappen, Lukas	Tue-M-SS-1-1	77
, , , , , , , , , , , , , , , , , , ,	Thu-M-V-5-8	190
Stefansdottir, B.	Wed-F-V-4-13	170
Steinert, Lars	Tue-F-O-3-2	102
Stemmer, Georg	Wed-F-V-2-10	164
Stephenson, Brooke	Fri-M-O-2-1	220
Stern Richard M	Thu-A-V-1-1	200
Stering Identification Internet	Fri-A-O-1-2	241
Stoakes, Hywel	Fri-M-V-2-14	227
Stoidis Dimitrios	Wed-M-V-4-14	130
Stolcke Andreas	Tue-A-V-4-2	89
Storence, I murcus IIIII	Thu-M-V-4-1	187
	Thu-A-V-2-7	203
	Fri-A-V-4-1	249
	Fri-A-V-4-2	250
Stone, Michael	Thu-A-O-2-3	200
Stone. Simon	Thu-M-S&T-1-2	196
Strake, Maximilian	Wed-E-SS-1-3	175
	Fri-A-SS-1-4	256
Strauss. Martin	Fri-M-O-3-3	222
Strik. Helmer	Thu-M-V-1-1	180
Strimel. Grant P	Tue-A-V-5-2	91
	Wed-E-V-3-2	165
	Fri-A-V-3-4	248
	Fri-A-V-3-7	248
Strohman, Trevor	Wed-A-V-2-6	143
Stüker, Sebastian	Wed-M-V-6-1	133
	Wed-E-V-1-2	160
Šturm. Pavel	Wed-M-O-2-1	120
,	Wed-E-V-4-2	168
Styles, Suzy J	Wed-M-V-1-1	122
Stylianou, Yannis	Thu-A-V-5-8	210
Stypułkowski, Michał	Tue-E-O-1-1	100
, , ,	Tue-E-O-1-2	100
Su, Bo-Hao	Tue-E-V-3-6	108
Su, Dan	Tue-M-V-5-11	72
	Tue-E-V-2-2	106
	Wed-A-V-3-8	146
	Wed-A-V-5-2	150
	Thu-A-V-6-5	212
	Fri-A-V-5-4	252
Su, Hang	Wed-A-V-3-14	147
	Fri-A-V-3-6	248
Su, Jinru	Tue-A-V-5-2	91
	Fri-A-V-3-7	248
Su, Ruolin	Tue-M-V-4-3	69
	Tue-E-V-4-5	111
SubpalaKSnml, K.P.	1nu-A-SS-2-12	219
Sudoh Votenhitz	Tue-M-V-6-3	150
Suuon, Katsunito	Wed-A-V-6-7	155
Sugndormann Oaft D	wed-A-V-6-9	134
Suendermann-Oert, D	Wed-A-V-I-7	142
Sup Eric		122
Juli, LIIC	Wed-E-V-2-1-4	133 167
	Thu-A-V-2-10	204
Sun Huiming	Tuo-A-V-2-10	207
Sun Jianwoi		112
Sun Jinowa	Wed-M-V-6-9	13/
Sun Lei		188
Sun Lifa	Fri-A-SS-2-9	258
Sun Ming	Τμε-Δ-V-1-6	230
Sun Oinghua	Eri-Δ-V-2-4	246
Sun. Sining	Τμε-Δ-V-4-1	- 10
	Fri-Δ-V-3-11	240
Sun. Xiao	Wed-E-V-3-12	167
Sun. Xingwei	Tue-F-V-2-3	106
Sun. Yandan	Fri-A-V-6-5	254
Sun. Yuhang	Tue-M-V-3-1	66
~ ····, · ······		00

Sung Dooyong	Eri-A-V-3-3	247
Sung Iongmo	Wed-M-V-4-13	130
Sung Juno Sig		126
Sunlara Manica		114
Sulikara, Mollica	Tue-E-V-5-10	114
c . II lt		195
Suo, Hongbin	I nu-A-V-4-2	207
Suresh, Ananda I	1ue-A-SS-1-4	95
Suter, Benjamin	Tue-E-O-1-3	100
Svec, Jan	Tue-M-O-1-2	60
	Wed-A-S&T-1-2	156
	Fri-A-O-2-4	242
Svendsen, Torbjørn	Tue-E-V-3-7	109
Swaminathan, Rupak V.	Fri-A-V-3-4	248
Swapna, N	Tue-A-V-2-4	85
Swart, Albert	Tue-E-V-1-5	103
Swearengen, Michael	Tue-A-V-2-7	86
Syed, Muhammad S.S	Thu-A-SS-2-8	218
Sved, Zafi Sherhan	Thu-A-SS-2-8	218
Symonds, Helena	Wed-E-O-3-2	159
Synnaeve. Gabriel	Tue-M-V-5-10	72
-,,	Tue-A-V-4-4	89
	Tue-A-V-4-8	90
Szakav. Anita	Tue-E-SS-1-1	117
	Tue-E-SS-1-2	117
Szöke. Igor	Thu-M-SS-2-2	195
	Thu-M-SS-2-4	195
	Thu-M-SS-2-5	196
Sztahó Dávid	Tue-A-S&T-1-1	98
Szurley, Joseph	Tue-A-SS-1-8	96
Szuncy, Joseph	146 - 755 1 0	50
Т		
-		
1., Kıran Praveen	Wed-E-V-1-9	162
T., Mani Kumar	Thu-A-V-1-5	201
Tachibana, Kentaro	Thu-M-V-5-1	188
Tada, Kohei	Thu-M-V-6-1	190
Tahon, Marie	Tue-A-S&T-1-5	99
Tai, Wenxin	Thu-A-V-3-10	206
Tairum Cruz, Miguel	Fri-M-V-6-12	236
Tak, Hemlata	Tue-A-SS-1-1	94
	Wed-E-O-1-2	158
Takahashi, Shun	Wed-M-V-2-7	125
Takahashi, Shusuke	Tue-E-V-2-7	107
Takamichi, Shinnosuke	Wed-M-V-3-6	127
Takashima, Akihiko	Tue-M-V-6-6	74
	Tue-A-V-5-5	91
	Tue-A-V-5-5 Wed-E-V-3-13	91 167
	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6	91 167 228
	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11	91 167 228 245
Takashima, Yuki	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4	91 167 228 245 187
Takashima, Yuki	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8	91 167 228 245 187 188
Takashima, Yuki Takeda, Ryu	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7	91 167 228 245 187 188 232
Takashima, Yuki Takeda, Ryu Talkar, Tanya	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-O-2-1	91 167 228 245 187 188 232 60
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-O-2-1 Wed-E-V-5-9	91 167 228 245 187 188 232 60 172
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-O-2-1 Wed-E-V-5-9 Tue-M-V-7-6	91 167 228 245 187 188 232 60 172 76
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha .	Tue-A-V-5-5 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-O-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6	91 167 228 245 187 188 232 60 172 76 163
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-A-O-1-5	91 167 228 245 187 188 232 60 172 76 163 138
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-A-0-1-5 Thu-M-V-5-7	91 167 228 245 187 188 232 60 172 76 163 138 190
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-A-0-1-5 Thu-M-V-5-7 Fri-A-V-5-9	91 167 228 245 187 188 232 60 172 76 163 138 190 253
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin Tan, Kye Min	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Fri-M-V-5-7 Tue-M-V-5-7 Ued-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-E-V-2-5 Thu-M-V-5-9 Fri-A-S&T-1-6	91 167 228 245 187 188 232 60 172 76 163 138 190 253 261
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-A-0-1-5 Thu-M-V-5-7 Fri-A-V-5-9 Fri-A-S&T-1-6 Thu-A-V-2-1	91 167 228 245 187 188 232 60 172 76 163 138 190 253 261 202
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-A-0-1-5 Thu-M-V-5-7 Fri-A-V-5-9 Fri-A-S&T-1-6 Thu-A-V-2-1 Fri-A-V-5-6	91 167 228 245 187 188 232 60 172 76 163 138 190 253 261 202 252
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-O-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-A-O-1-5 Thu-M-V-5-7 Fri-A-V-5-9 Fri-A-S&T-1-6 Thu-A-V-2-1 Fri-A-V-5-6 Tue-E-V-2-7	91 167 228 245 187 188 232 60 172 76 163 138 190 253 261 202 252 252
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-A-0-1-5 Thu-M-V-5-7 Fri-A-V-5-9 Fri-A-S&T-1-6 Thu-A-V-2-1 Fri-A-V-5-7 Tue-E-V-2-7 Tue-M-V-6-6	91 167 228 245 187 188 232 60 172 76 163 138 190 253 261 202 252 252 107 74 4
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Thu-A-V-2-10 Fri-A-V-5-5 Tue-M-V-6-6 Tue-A-V-5-5	91 167 228 245 187 188 232 60 172 76 163 138 190 253 261 202 252 252 107 74 91
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Thu-A-V-2-1 Fri-A-V-5-6 Tue-E-V-2-7 Tue-M-V-6-6 Tue-A-V-5-5 Wed-M-V-6-6	91 167 228 245 187 188 232 60 0172 76 163 138 190 253 261 202 252 252 107 74 91 134
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-E-V-2-7 Fri-A-V-5-9 Fri-A-V-5-6 Tue-A-V-5-5 Wed-M-V-6-6 Wed-E-V-3-13	91 167 228 245 187 188 232 76 163 138 190 253 261 202 252 107 74 91 134 167 262 262 252 252 252 252 252 252
Takashima, YukiTakeda, RyuTalkar, TanyaTambwekar, AnujTamim, NourTammewar, AniruddhaTamminen, SamuTan, DaxinTan, Kye MinTan, XuTanaka, KeitaroTanaka, Tomohiro	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-E-V-2-7 Fri-A-V-5-9 Fri-A-S&T-1-6 Thu-A-V-2-1 Fri-A-V-5-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6	91 167 228 2457 188 232 260 172 76 163 138 190 253 262 252 252 252 252 252 2107 74 91 134 167 228
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha . Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-3-6 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-V-5-9 Fri-A-V-5-7 Thu-M-V-5-7 Thu-A-V-2-1 Fri-A-V-5-6 Tue-E-V-2-7 Tue-M-V-6-6 Tue-A-V-5-5 Wed-M-V-6-6 Wed-E-V-3-13 Fri-A-V-3-13 Fri-A-V-3-6 Fri-A-V-1-11	911167 2288 2455 18771888 232260 1722766 1633138 19002533 2611202252 2522252252 252210774 911344 1672288 2455
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tammewar, Aniruddha . Tammewar, Aniruddha . Tammewar, Aniruddha . Tammewar, Aniruddha . Tann, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-A-0-1-5 Thu-M-V-5-7 Fri-A-S&T-1-6 Thu-A-V-5-9 Fri-A-S&T-1-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V-5-3	91 167 2288 245 187 188 232 60 172 76 163 31 38 190 253 261 202 252 107 741 91 134 167 228 245 209 252 252 252 252 252 252 252 25
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin Tang, Fengyi	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-5 Wed-E-V-2-7 Tue-M-V-6-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-A-V-1-11 Thu-A-V-5-3 Thu-A-V-5-3	911 1677 2288 2455 1877 1888 2322 600 1726 766 1633 1388 1900 2533 2611 2022 2522 2522 2522 2107 744 911 1344 1677 2288 2099 2188
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin Tang, H. Lilian	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Thu-M-V-5-7 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-7 Tue-M-V-6-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-3 Fri-A-V-5-3 Fri-A-V-5-3 Thu-A-V-5-3 Thu-A-V-5-3 Thu-A-V-5-3	911 1677 2288 2455 1877 1888 232 600 1722 766 163 138 190 253 261 138 190 253 261 138 261 138 262 107 74 91 134 167 208 265 209 2188 245 209 2188 245 209 2188 245 209 2188 245 209 2188 245 209 2188 245 209 2187 200 200 200 200 200 200 200 200 200 20
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin Tang, H. Lilian	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-E-V-2-7 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-5 Wed-M-V-6-6 Fri-A-V-1-11 Thu-A-V-5-2 Thu-A-V-5-2 Thu-A-V-2-28 Thu-A-V-2-8 Tue-M-V-4-7	91 167 2288 245 187 188 232 60 172 76 163 138 190 253 261 202 252 252 252 2107 74 91 134 167 2285 245 245 205 215 205 205 205 205 205 205 205 20
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tamimewar, Aniruddha Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Fengyi Tang, H. Lilian Tang, Min	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-E-V-2-7 Tri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-6 Thu-A-V-2-1 Fri-A-V-5-5 Wed-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V-3-3 Thu-A-V-2-8 Thu-A-V-2-8 Thu-A-V-2-8 Tue-M-V-4-7 Wed-E-V-5-3	911 167722882452 187718822452187718822452187761822452177618219221222222221077744911134416772288245520992188144470012111
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin Tang, Fengyi Tang, H. Lilian Tang, Min Tang, Ping	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-3-6 Fri-A-V-1-11 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-E-V-2-7 Tri-A-V-5-9 Fri-A-S&T-1-6 Thu-A-V-2-1 Fri-A-V-5-5 Wed-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V-5-3 Thu-A-SS-2-11 Wed-E-V-5-3 Tue-M-V-4-7 Wed-E-V-5-3 Fri-M-V-1-4	91 167 2288 245 187 188 232 600 1722 76 163 138 190 2533 261 202 2522 107 74 91 134 167 2288 245 209 9218 144 70 171 228
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tann, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Fengyi Tang, H. Lilian Tang, Din Tang, Ping Tang, Xiao	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-3-6 Fri-A-V-1-11 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-A-0-1-5 Thu-M-V-5-7 Fri-A-V-5-5 Thu-M-V-5-7 Fri-A-V-5-5 Thu-A-V-2-1 Fri-A-V-5-5 Wed-M-V-6-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V-5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-6-9	911 167722882455 187718822455 187823261 1722767 1633138190 2533261 202252210774 9111344 167722824520992188 144470711223374
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Fengyi Tang, Lilian Tang, Nin Tang, Xiao Tang, Yufeng	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-5 Wed-A-0-1-5 Thu-M-V-5-7 Fri-A-V-5-9 Fri-A-V-5-9 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-A-V-1-11 Thu-A-V-5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-2-8 Tue-M-V-2-3 Fri-M-V-1-3 Fri-M-V-3-3 Fri-M-V-3-3 Fri-M-V-3-3 Fri-M-V-3-3 Fri-M-V-3-3 Fri-M-V-3-3 Fri-M-V-3-3	91 167 2288 245 187 188 232 600 172 76 163 138 190 253 261 107 74 91 134 167 228 209 218 144 700 171 223 245 209 218 215 209 218 218 209 218 218 219 218 219 218 219 218 219 219 219 219 219 219 219 219
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Nye Min Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Fengyi Tang, H. Lilian Tang, Min Tang, Xiao Tang, Yufeng Tang, Zhenyu	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Thu-M-V-5-7 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-3 Shu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-4-7 Wed-E-V-3-3 Fri-M-V-1-4 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-5-5	911 1677 2288 2455 1877 1888 232 600 1722 766 1633 138 1900 2533 2611 2022 2522 2522 2522 2522 2522 2522 25
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin Tang, Fengyi Tang, H. Lilian Tang, Nin Tang, Yufeng Tang, Zhenyu Tang, Zhiyuan	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-5 Wed-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-6-9 Wed-E-V-5-3 Fri-M-V-1-4 Tue-M-V-6-9 Wed-M-V-5-5 Tue-E-V-5-1	911 167 2288 245 187 188 232 60 172 76 60 163 138 190 253 261 202 252 252 252 252 252 2107 74 91 134 167 228 245 209 218 144 700 171 223 245 138 190 253 261 120 2252 202 210 74 91 134 167 228 245 202 210 74 107 245 202 202 202 202 200 218 245 245 245 245 245 245 245 245 245 245
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tamimewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Fengyi Tang, H. Lilian Tang, Nin Tang, Yufeng Tang, Yufeng Tang, Zhiyuan	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-7-6 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-7 Tri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-5 Wed-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V-5-3 Thu-A-V-2-8 Tue-M-V-6-9 Wed-E-V-5-3 Fri-M-V-1-4 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-5-5 Tue-M-V-6-8 Tue-M-V-5-5	911 167722882455 187718822455 187822455 1888232260 172276163 138819002533 261122252107744 91132622522107744 91132622522107744 9114167722882455 209921881444 70017112233744 1344711122183
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tambwekar, Aniuj Tammewar, Aniruddha Tam, Daxin Tan, Daxin Tan, Daxin Tan, Daxin Tan, Xu Tana, Xu Tanaka, Keitaro Tanaka, Tomohiro Tanag, Fengyi Tang, H. Lilian Tang, Min Tang, Min Tang, Xiao Tang, Yufeng Tang, Zhenyu Tang, Zhenyu Tang, Ziyue Tang, Ziyue	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-0-2-1 Wed-E-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-A-0-1-5 Thu-M-V-5-7 Fri-A-V-5-5 Thu-M-V-5-7 Fri-A-V-5-5 Thu-A-V-2-1 Fri-A-V-5-5 Wed-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V-5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-6-9 Wed-E-V-5-3 Fri-M-V-1-4 Tue-M-V-6-9 Wed-M-V-6-8 Tue-M-V-6-9 Wed-M-V-6-8 Tue-M-V-5-5 Tue-E-V-5-1 Thu-M-V-2-4 Wed-E-V-5-1	911 16772288 2455187 1882245 1877188 232261 1633261 202252 10774 911 13442 209228 2184 707228 245229 2184 7071223 74471 11223 744711 1124711 11
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tann, Daxin Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Fengyi Tang, H. Lilian Tang, Min Tang, Xiao Tang, Yufeng Tang, Zhenyu Tang, Zhiyuan Tang, Ziyue Tanguchi, Akira	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-3-6 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Wed-E-V-5-9 Tue-M-V-5-7 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-7 Tue-M-V-6-6 Tue-A-V-5-5 Wed-H-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-5-3 Fri-A-V-5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-6-9 Wed-M-V-6-8 Tue-M-V-6-9 Wed-M-V-6-8 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-5-5 Tue-M-V-6-8 Tue-N-6-8 Tue-N-6-8 Tue-V-6-8 Tue-V-6-8	91 167 2288 245 187 188 232 600 172 76 163 138 190 253 261 202 252 107 74 91 134 167 228 245 209 218 245 209 218 245 209 218 245 209 218 218 201 201 201 201 201 201 201 201
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tan, Nour Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin Tang, Fengyi Tang, H. Lilian Tang, Ding Tang, Yufeng Tang, Yufeng Tang, Zhiyuan Tang, Ziyue Tang, Kira Tang, Xiyao	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-7 Tue-M-V-6-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-6-9 Wed-M-V-6-8 Tue-M-V-6-8 Tue-E-V-5-1 Thu-M-V-6-8 Tue-E-V-6-8 Tue-E-V-6-8 Tue-E-V-6-8 Tue-E-V-6-8	911 1677 2288 2455 1877 1888 232 600 1726 766 163 1388 1900 2533 2611 2022 2522 2522 2522 2522 2522 209 2188 167 74 91 1344 700 1711 2285 2099 2188 1447 711 2285 2099 2188 1447 711 2285 2099 2188 1477 74 911 1577 2187 2187 209 2187 100 2187 2187 209 2187 2197 2197 2197 2197 2197 2197 2197 219
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tam, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Fengyi Tang, H. Lilian Tang, Min Tang, Min Tang, Min Tang, Xiao Tang, Xiao Tang, Zhenyu Tang, Zhenyu Tang, Ziyue Taniguchi, Akira Taniguchi, Tadahiro Taniguchi, Tadahiro Tanj, Iianhua	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Thu-M-V-5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-9 Fri-A-V5-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-3 Thu-A-S-2-11 Wed-A-V-2-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-E-V-5-1 Thu-M-V-2-4 Tue-E-V-6-8 Tue-E-V-6-8 Thu-M-V-6-3 Tue-M-V-6-3 Tue-M-V-6-3 Tue-M-V-5-3	911 1677 2288 2455 1877 1888 232 600 1722 766 1633 138 1900 2533 2611 2022 2522 2522 2532 2612 2022 2532 2612 2022 2532 2612 2022 2107 74 911 1344 1677 2288 2455 2090 2188 1444 700 1711 2233 2455 1077 74 911 1344 1677 2288 2455 2090 2188 1457 2090 2188 1457 2090 2188 1457 2090 2187 2090 2187 2090 2187 2090 2187 2090 2187 2090 2187 2090 2187 2090 2187 2090 2187 2090 2187 2090 2187 2090 2187 2090 2188 2455 2090 2188 2457 2090 2188 2457 2090 2188 2457 2090 2188 2457 2090 2188 2457 2090 2188 2457 2090 2188 2457 2090 2188 2457 2090 2188 2197 2090 2188 2197 2000 2000 2000 2000 2000 2000 2000 20
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin Tang, Fengyi Tang, H. Lilian Tang, Nin Tang, Yufeng Tang, Zhenyu Tang, Ziyue Tang, Ziyue Tang, Ziyue Tang, Ziyue Tang, Jianhua	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-1-11 Thu-M-V-4-4 Thu-M-V-4-4 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-9 Fri-A-V-5-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-5 Wed-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-5 Wed-M-V-6-6 Wed-E-V-3-13 Fri-A-V-5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-6-9 Wed-M-V-6-9 Wed-M-V-6-8 Tue-K-S-5 Tue-E-V-6-8 Tue-M-V-6-3 Tue-M-V-6-3 Tue-M-V-6-3 Tue-M-V-6-3 Tue-M-V-5-3	911 167722882452 187718822452452 1882232262767 1633138819002533261122767 163313826122022252107744 91341672288245209921881444 70112233744 71112233744 7111221833744 711112218331155190095
Takashima, Yuki Takeda, Ryu Talkar, Tanya Talkar, Tanya Tambwekar, Anuj Tamim, Nour Tamim, Nour Tammewar, Aniruddha Tammewar, Aniruddha Tamminen, Samu Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Chuanxin Tang, Fengyi Tang, H. Lilian Tang, Min Tang, Niao Tang, Yufeng Tang, Zhiyuan Tang, Zhiyuan Tang, Ziyue Taniguchi, Akira Tanji, Ryo Tanjianhua	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-3-6 Fri-A-V-1-11 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-7-6 Wed-E-V-2-6 Wed-E-V-2-6 Wed-A-O-1-5 Thu-M-V-5-7 Fri-A-V-5-9 Fri-A-S&T-1-6 Tue-A-V-5-9 Fri-A-S&T-1-6 Tue-A-V-5-5 Wed-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V-5-3 Thu-A-SS-2-11 Wed-E-V-5-3 Fri-M-V-3-6 Fri-M-V-3-6 Fri-M-V-3-5 Tue-M-V-6-9 Wed-E-V-5-3 Fri-M-V-1-4 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-K-5-5 Tue-A-SS-17 Tue-A-SS-17 Tue-A-SS-17 Tue-A-SS-17	91 167 2288 245 187 188 232 600 1722 76 163 138 190 253 261 202 2522 107 74 91 134 167 228 245 209 218 245 209 218 245 209 218 245 209 218 245 209 218 245 209 218 201 200 200 200 200 200 200 200 200 200
Takashima, Yuki Takeda, Ryu Talkar, Tanya Tambwekar, Anuj Tambwekar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tammewar, Aniruddha Tan, Daxin Tan, Kye Min Tan, Xu Tanaka, Keitaro Tanaka, Tomohiro Tang, Fengyi Tang, H. Lilian Tang, Nin Tang, Yufeng Tang, Xiao Tang, Zhenyu Tang, Ziyue Taniguchi, Akira Tanjuchi, Tadahiro Tanji, Ryo Tao, Jianhua	Tue-A-V-5-5 Wed-E-V-3-13 Fri-A-V-3-6 Fri-A-V-1-11 Thu-M-V-4-8 Fri-M-V-5-7 Tue-M-V-5-7 Tue-M-V-5-9 Tue-M-V-5-9 Tue-M-V-5-9 Fri-A-S&T-1-6 Thu-A-V-5-7 Fri-A-V-5-9 Fri-A-S&T-1-6 Thu-A-V-5-7 Tue-M-V-6-6 Tue-E-V-2-7 Tue-M-V-6-6 Tue-E-V-2-7 Tue-M-V-6-6 Wed-E-V-3-13 Fri-M-V-3-6 Fri-A-V-1-11 Thu-A-V-5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-6-9 Wed-E-V-5-3 Thu-A-SS-2-11 Wed-A-V-2-8 Tue-M-V-6-9 Wed-E-V-5-3 True-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-M-V-6-8 Tue-E-V-5-1 Thu-M-V2-4 Tue-E-V-6-8 Tue-E-V-6-8 Tue-M-V-6-3 Tue-A-V-5-1 Thu-A-V-5-1 Tue-A-SS-1-7 Wed-E-V-2-8 Wed-E-V-2-8	91 167 2288 245 187 188 232 600 1722 76 163 138 190 2533 2611 2022 2022 107 74 91 134 167 2282 245 209 2182 245 209 2187 102 2522 107 74 91 1134 163 2612 202 2152 107 74 91 1134 163 2612 202 2152 107 74 91 1134 167 202 202 2152 107 74 91 1134 167 202 202 2152 107 74 91 1134 167 202 202 2152 2152 107 74 91 1134 167 202 202 2152 2152 209 2184 70 1721 203 2152 2155 215

Tawara, Naohiro	Thu-A-V-4-5	208
Taylor, Jason	Wed-E-V-6-11	175
Taylor, Sarah	Thu-A-O-1-4	199
Ten, Kan Kuan	Fri-M-V-5-8	233
Ten, Han Huey	Fri-M-0-2-3	221
Teixeira, António	Tue-A-V-2-9	86
Teixeira, Francisco	Tue-M-SS-1-2	77
ten Bosch, Louis	Tue-A-O-2-4	81
	Tue-A-O-2-5	81
Terblanche Camryn	1 ue-A-O-2-6	132
Tevtaut. Yann	Tue-M-O-3-4	62
Thelapurath, Shrinath	Fri-A-SS-2-15	260
The RADAR-CNS C/m	Tue-A-V-2-6	85
Thiagarajan, J.J.	Wed-A-V-4-12	150
Thieberger, Nick	Fri-M-V-2-14	227
Thienpondt Jenthe	I hu-M-V-7-2	192
Thenpoliat, Jenthe	Thu-A-V-4-4	207
Thomas, Samuel	Tue-E-V-4-8	111
	Wed-M-V-2-12	126
	Thu-M-V-2-8	184
	Thu-A-V-2-6	203
Thompson Jennifer C	<u>FΓΙ-Α-V-6-6</u> Thu-Δ-SS-2-7	218
Thulke. David	Wed-A-O-1-4	138
Thyagachandran, Anand	Wed-E-V-1-6	161
Tian, Biao	Tue-E-V-2-9	107
Tian, Xiaohai	Tue-E-V-6-11	116
Tian, Znengkun	<u>IUE-M-V-5-1</u>	20
	Wed-M-V-4-5	128
	Fri-M-V-3-1	227
Tiede, Mark	Thu-A-O-1-4	199
Tihelka, Daniel	Tue-M-O-1-2	60
Tilson Sam	Fri-A-S&I-I-2	260
Tiscareno, Ingrid	Wed-M-O-2-4	120
Tiwari, Upasana	Tue-A-SS-2-10	98
Tjandra, Andros	Fri-M-V-4-1	229
Tobin, Jimmy	Fri-A-SS-2-1	257
	Fri-A-SS-2-12	259
Toda. Tomoki	Tue-E-V-6-2	114
	Wed-A-V-5-3	150
	Wed-A-V-5-9	152
Todisco Massimiliano	Thu-A-V-5-9	210
	Tue-A-SS-1-1	94
	Tue-E-V-1-13	105
	Wed-E-O-1-2	158
	Fri-M-V-7-9	238
Togami Masahito	Fri-M-V-7-13 Wed-A-V-4-3	239
	Thu-M-V-3-6	185
Tokuyama, Hirotaka	Wed-A-V-6-7	153
Tolias, Andreas Savas	Thu-A-V-3-7	206
Toma, Andrea	Wed-A-V-4-8	149
Tomanek, Katrin	Fri-A-SS-2-1	257
	Fri-A-SS-2-14	260
Tomasello, Paden	Tue-M-V-5-10	72
Tomashenko, Natalia	Tue-E-V-1-13	105
	Wed-M-O-1-5	119
Tong Audrey	Inu-A-V-4-8	200
Tong. Fuchuan	Thu-A-SS-1-8	215
8, 11	Fri-A-V-4-9	251
Tong, Ziyi	Wed-M-O-1-5	119
Tooley, Lauren	Fri-A-SS-2-15	260
Topar, Topias	Wed-A-S&T-1-3	157
Torgashov. Nikita	Wed-A-SS-1-1	155
Torralba, Antonio	Wed-M-V-2-12	126
Torresquintero, A	Fri-M-O-2-3	221
Táth Lác-lá	Fri-M-O-2-4	221
10111, Laszio	Tue-M-SS-1-10	79 140
Tran, Dung N.	Wed-E-V-5-5	171
Tran, Huy Dat	Fri-M-V-5-8	233
Tran, Khoa Viet	Wed-A-O-1-4	138
Tran, Trang	Wed-E-V-4-5	168
Trançoso Isabel	Tue-A-V-6-10	94 77
11ancoso, 15abel	Wed-E-V-2-9	164

Treep, Jelle	Tue-M-SS-1-1	77
1,5	Tue-M-SS-1-8	78
Trella, Paweł	Fri-A-O-2-1	242
Tripathi, Anshuman	Thu-M-V-2-10	184
-	Thu-A-V-2-4	203
Tripiana, Antonio	Thu-M-S&T-1-4	197
Trmal, Jan	Thu-M-O-3-6	179
	Thu-A-V-6-5	212
Trouvain, Jürgen	Thu-M-V-6-5	191
	Thu-M-V-6-7	191
	Fri-M-V-2-4	225
Truong, Thinh Hung	Fri-A-V-6-1	254
Tsao, Yu	Tue-M-V-3-8	68
	Tue-M-V-3-9	68
	Tue-E-V-6-2	114
	Wed-M-V-4-2	128
	Thu-A-V-5-9	210
Tschoepe, Constanze	Tue-E-O-2-3	101
Tseng, Shu-Chuan	Tue-A-V-2-1	84
Tseng, Wei-Cheng	Tue-E-V-3-9	109
	Wed-E-V-6-9	174
Tsiakoulis, Pirros	Wed-M-V-3-2	126
Tsukada, Kimiko	Fri-M-V-1-1	222
Tsunoo, Emiru	Tue-M-V-5-8	72
Tsuzaki, Minoru	Thu-M-V-6-9	192
Tu, Wei-Wei	Fri-M-V-6-11	236
Tu, Youzhi	Tue-M-V-1-6	63
Tu, Zehai	Tue-A-V-3-8	88
Tulics, Miklós Gábriel	Tue-A-S&T-1-1	98
Tur, Gokhan	Wed-A-V-2-13	145
Turchi, Marco	Wed-A-V-6-11	154
	Thu-A-V-6-2	211
Turrisi, Rosanna	Tue-M-O-2-5	61
Tuske, Zoltan	Tue-E-V-4-8	111
	Wed-M-V-6-9	134
	Wed-A-V-2-11	144
	Wed-A-V-3-5	145
T	Wed-E-V-3-12	167
Tzanetakis, George	Wed-E-O-3-2	159
izimiropoulos, G	Thu-A-V-1-5	201
I ZIRAKIS, Panaglotis	Tue-M-SS-1-1	122
I zudir, Moakala	Wed-M-V-1-8	123
TI		

U

Udupa, Sathvik Tue-E	-V-3-1 1	07
Fri-A-S&	xT-1-8 20	61
Ukkonen, Antti Wed-A	-0-1-5 1	38
ul Haq, Ehsan Tue-M	-V-7-5	75
Um, Seyun Thu-A	-V-5-2 20	90
Umesh, Tejas Tue-A	-V-2-3	85
Unni, Vinit Wed-E	-V-1-7 1	61
Unoki, Masashi Thu-M	-V-3-4 18	85
Urooj, Saba Tue-M	-V-7-5	75
Utsuro, Takehito Tue-M	-V-6-7	74
Wed-E	-V-1-4 1	61

V

Vaaras, Einari	Thu-A-V-1-3	200
Vaideeswaran, Rakesh	Wed-E-V-1-7	161
Valentini-Botinhao, C	Wed-E-V-6-2	173
Vali, Mohammad Hassan	Thu-A-O-2-1	199
Valin, Jean-Marc	Tue-E-V-2-5	106
	Wed-M-V-4-8	129
Vallée, Nathalie	Tue-E-O-2-5	101
Vallés-Pérez, Iván	Thu-M-V-5-2	189
Valstar, Michel	Thu-A-V-1-5	201
Vamvoukakis, Georgios	Wed-M-V-3-2	126
Van de Velde, Hans	Fri-M-V-2-13	227
Van Gysel, Christophe	Wed-A-V-2-12	144
Van hamme, Hugo	Tue-M-O-2-4	61
	Tue-A-O-2-3	81
	Wed-A-O-3-3	140
	Wed-E-O-3-3	160
van Hout, Roeland	Thu-M-V-1-1	180
Van keirsbilck, Matthijs	Fri-A-V-3-1	247
van Niekerk, Benjamin .	Wed-M-V-2-3	124
	Wed-M-V-2-6	124
van Niekerk, Daniel	Thu-M-V-6-6	191
van Os, Marjolein	Fri-M-V-2-4	225
van Rijn, Pol	Tue-M-SS-1-12	79
	Fri-M-O-2-2	220
van Son, Rob J.J.H	Tue-E-O-3-1	102
Variani, Ehsan	Wed-M-V-6-4	133
	Fri-M-V-3-3	228
Varma, Rajat	Thu-A-V-4-1	207
	Thu-A-V-4-6	208
Varol, Huseyin Atakan .	Wed-E-V-6-10	174

Vasilescu. Ioana	Tue-E-O-2-2	101
Vásquez-Correa, J.C	Tue-M-O-2-2	61
_	Tue-M-SS-1-3	77
	Thu-A-SS-2-2	217
Vaysse, Robin	Wed-A-O-3-2	139
Vega Rodriguez, Jenifer	Tue-E-O-2-5	101
Velichko, Alena	Tue-M-SS-1-11	79
Venkataramani, Sh	Tue-E-V-2-5	106
Venkataramani, Sw	Wed-E-V-3-12	167
Venkatesh, Ganesh	Fri-A-V-3-10	249
Venue, Kalph	Tue-E-O-3-4	103
Venugopalan, S	Fri-A-55-2-14	200
vepa, jimenura	Wed-A-S&I-I-6	157
	wea-A-S&I-I-7	102
Verkholvak Ovana	1 nu-M-V-7-6	195
Verkholyak, Oxalia	Thu-A-V-6-0	213
Vocolý Karol		105
vesely, Karer		195
		106
Vetráb Mercedes	Tuo-M-SS-1-10	70
Vidal Jazmín	Eri-A-V-1-4	2/3
Vijavasenan Deenu	FII-A-V-1-4	08
Villalba Jesús	Tue-A-33-2-0	74
villaiba, jesus	Tuo-A-O-1-2	80
	Tue-F-V-1-9	104
	Thu-Δ-SS-1-12	216
	Thu-Δ-SS-2-10	218
	Fri-M-V-7-10	238
Villatoro-Tello, Esaú	Wed-A-O-3-5	140
Vincent, Emmanuel	Tue-A-V-3-9	88
Vinciarelli, Alessandro .	Wed-E-V-2-4	163
· · · · · , · · · · · ·	Wed-E-V-2-5	163
Vinogradova, Alisa	Wed-A-SS-1-3	155
Virkar, Yogesh	Thu-M-V-5-6	189
Viswanathan, Vijay	Tue-E-V-4-10	112
Vít, Jakub	Fri-A-S&T-1-2	260
Viveros Muñoz, Rhoddy	Tue-A-V-3-7	88
Volná, Veronika	Tue-A-S&T-1-2	99
von Neumann, Thilo	Thu-A-V-3-3	205
Vozila, Paul	Wed-A-V-2-7	143
Vuissoz, Pierre-André	Thu-A-O-1-1	198
Vuong, Tyler	Fri-M-V-5-11	233
	Fri-A-O-1-2	241
Vyas, Apoorv	Thu-M-O-1-6	178
Vyas, Piyush	Wed-A-V-6-12	154
Vyas, Saurabh	Wed-E-V-1-7	161
Vydana, Hari Krishna	Thu-M-O-3-5	179
W		
Wagner Michael	Wod EV 44	160
wagner, michael	Wed-E-V-4-4	100
Wagner Detra	wed-E-V-4-11	1/0
wagner, Petra	Tue-E-O-2-4	101
Waihal Alay	1hu-A-O-1-3	198
waldel, Alex	Wed-M-V-6-1	133

, , , , , , , , , , , , , , , , , , , ,		
	Thu-A-O-1-3	198
Waibel, Alex	Wed-M-V-6-1	133
	Wed-E-V-1-2	160
Walker, Daniel D	Tue-A-V-5-3	91
Wallbridge, Sarenne	Wed-E-O-2-4	159
Wallington, Electra	Tue-A-V-4-3	89
	Thu-M-O-3-1	178
Wallis, Christopher G.R.	Fri-M-O-2-3	221
	Fri-M-O-2-4	221
Wan, Moquan	Tue-M-V-2-6	65
Wan, Xucheng	Thu-M-V-4-3	187
Wang, Bin	Fri-M-V-4-5	230
Wang, Binling	Thu-M-SS-1-1	193
Wang, Boxin	Wed-A-V-2-3	143
Wang, Bruce Xiao	Tue-M-V-7-2	75
Wang, Changhan	Wed-A-V-6-3	153
	Wed-A-V-6-4	153
Wang, Chao	Wed-A-SS-1-5	156
Wang, Chao	Tue-A-V-1-6	83
Wang, Chenglong	Tue-A-SS-1-7	95
0. 0 0	Wed-M-V-4-5	128
Wang, Chengyu	Fri-A-V-3-2	247
Wang, Chunfeng	Thu-M-V-5-5	189
Wang, Congyi	Fri-M-V-4-5	230
Wang, Dan	Tue-E-V-1-4	103
Wang, DeLiang	Tue-A-V-3-6	88
0. 0	Tue-E-V-2-8	107
Wang, Ding	Thu-M-SS-1-4	194
Wang, Disong	Tue-E-V-6-5	115
0. 0	Thu-M-V-1-10	182
	Fri-A-SS-2-8	258
Wang, Dong	Thu-M-SS-1-1	193
<u> </u>	Fri-M-V-7-8	238
Wang Fan-Lin	Thu-M-V-3-8	186

INTERSPEECH 2021 276 BRNO | CZECHIA

Wang Fei	Thu-A-V-2-11	206	Wang We
		200	wang, we
wang, Feng	I ue-A-V-2-2	84	
Wang, Gary	Tue-A-V-4-7	90	Wang, We
Wang, Guan-Bo	Thu-A-V-6-5	212	
8,	Fri-M-SS-1-5	240	Wang Wi
Warra Correction		100	Wang Vie
wang, Guangsen	wed-E-V-I-8	162	wang, Ala
Wang, Haoxiang	Thu-A-V-1-8	201	Wang, Xia
	Fri-A-V-2-2	245	
Wang Haoyu	Tuo.M.V.4.8	70	
Wang, Habyu		00	Wong Vi
wang, Helin	I ue-A-V-I-2	82	wang, Ala
	Tue-E-V-2-2	106	Wang, Xia
	Tue-E-V-3-2	108	Wang, Xii
Wang Hsin-Min	Tue-F-V-6-2	114	Wang Xi
wang, man min		1 - 2	Wang, An
	wea-A-v-6-5	135	wang, An
	Thu-M-V-3-8	186	
	Thu-A-V-5-9	210	
Wang Huaming	Wed-E-V-5-3	171	Wang Xie
Wong Hui		100	wung, m
	wea-m-v-1-5	122	T 17 T 7
Wang, Jianping	Fri-M-V-6-7	235	Wang, Xu
Wang, Jianrong	Tue-A-V-2-5	85	
8,5 8	Thu.M.V.2.4	183	Wang Ya
Wang Jianzong		100	Wong Vo
wallg, Jializolig	Tue-A-O-T-5	00	wallg, Ia
	Tue-A-V-1-11	84	Wang, Yil
	Tue-E-V-3-4	108	Wang, Yo
	Wed-M-V-3-9	127	0,
	Wed FO14	158	Wong Vu
	weu-E-0-1-4	100	wang, Tu
T17 T:	Thu-A-V-5-1	209	wang, Yu
Wang, Jie	Tue-A-V-6-9	94	Wang, Yu
Wang, lie	Fri-A-V-4-9	251	Wang Yu
Wang Jin	Tue-E-V-1-2	103	
Wang, Jin		144	T 17 T 7
wang, Jinchao	wed-A-V-2-9	144	Wang, Yu
Wang, Jindong	Thu-A-V-2-1	202	Wang, Yu
Wang, Jing	Tue-E-V-2-1	105	Wang, Zh
Wang Jingsong	Eri.M.V.6.11	236	Wang 7h
Wong Jinhon		110	Wang, Zh
wang, jinnan	Tue-E-V-5-3	112	wang, Zit
Wang, Junjie	Fri-M-V-6-10	235	Wani, Yas
Wang, Jun	Tue-A-V-2-10	86	Wanner, l
-	Fri-A-SS-2-2	257	Warden, l
Wang Jun	Tue-M-V-5-11	72	Watanahe
Wong Voi		10	watanabi
wang, Kal	I hu-M-V-3-5	185	
Wang, Kang	Wed-E-V-3-9	166	
Wang, L	Fri-M-O-1-2	219	
Wang Ian	Wed-M-V-5-6	131	
Wallg, Lall		750	
T17 T .	FII-A-55-2-7	230	
Wang, Li	Fri-M-V-6-6	235	
Wang, Liyuan	Tue-A-V-5-6	91	
0. 1	Wed-A-V-3-10	146	
		222	
		223	
	Fri-M-V-1-7	223	
	Fri-A-V-1-7	244	
Wang, Lizhong	Wed-M-V-4-3	128	
Wang Longhiao	Tue-M-V-2-1	64	
Wallg, Longblao		65	
	Tue-M-V-2-5	20	
	Iue-M-V-4-7	70	
	Tue-E-V-1-11	105	
	Fri-A-V-2-6	246	
	Fri-A-V-2-10	247	
Wang, Lu	Fri-A-V-6-5	254	
Wong Minggin	F.:: A V 1 2	242	Watore A
wang, Mingqiu	Fri-A-V-1-5	245	Waters, A
wang, Mingxuan	Wed-A-V-6-8	154	watson, (
Wang, Naigang	Wed-E-V-3-12	167	Wei, Bo .
Wang, Nick I.C.	Fri-A-V-6-5	254	Wei, Wen
Wang Ning	Thu A \$\$ 2.12	210	Weingart
Wang Deidang	Med EX 2-0	160	Weice Do
wally, reluoily	wea-E-V-3-8	100	,, c135, KU
Wang, Pinyi	Wed-E-V-1-5	161	T. 7 1 1 TT
Wang, Pu	Tue-M-O-2-4	61	weld, Hei
Wang, Oing	Wed-F-SS-1-8	176	Wen, Liar
Wang Qing	Tuo E V 1 12	105	Wen, Xue
	1ue-E-V-1-12	105	Won 7ho
Wang, Quan	Tue-E-V-1-6	104	wen, zne
	Fri-M-V-6-3	234	
ĺ	Fri-A-V-4-4	250	Wen, Zuji
Wang, Rui	Wed-M-V-5-11	132	Weng, Ch
Wong Duili		60	3, 55
wang, Kulli	Tue-M-V-3-TT	100	
wang, snaojun	Wed-M-V-3-5	127	TA7
	Wed-E-V-5-12	172	werner, F
	Fri-M-V-4-4	230	
Wang Shenghei	Thu-M-V-2-4	185	
Mang Sifer	THU-M-V-5-4	140	Westerm
wang, sifan	Wed-A-V-4-6	149	Wester
Wang, Taihui	Wed-M-SS-1-3	136	weston, I
Wang, Tao	Wed-M-V-4-5	128	Whang, T
Wang Tianzi		015	White, Ha
παμε. Παμλι	Thu-A-SS-1-6	215	.,
8,	Thu-A-SS-1-6	215	White Ka
	Thu-A-SS-1-6 Thu-A-SS-1-9	215	White, Ka
Wang, Wei	Thu-A-SS-1-6 Thu-A-SS-1-9 Tue-E-V-5-1	215 215 112	White, Ka Wiechma
Wang, Wei Wang, Weiging	Thu-A-SS-1-6 Thu-A-SS-1-9 Tue-E-V-5-1 Tue-E-V-1-2	215 215 112 103	White, Ka Wiechma Wiesler, S
Wang, Wei Wang, Weiqing Wang, Wei	Thu-A-SS-1-6 Thu-A-SS-1-9 Tue-E-V-5-1 Tue-E-V-1-2 Wed-A-V-2-2	215 215 112 103 143	White, Ka Wiechma Wiesler, S Wiesner,
Wang, Wei Wang, Weiqing Wang, Wei	Thu-A-SS-1-6 Thu-A-SS-1-9 Tue-E-V-5-1 Tue-E-V-1-2 Wed-A-V-2-3	215 215 112 103 143	White, Ka Wiechma Wiesler, S Wiesner,
Wang, Wei Wang, Weiqing Wang, Wei Wang, Wen	Thu-A-SS-1-6 Thu-A-SS-1-9 Tue-E-V-5-1 Tue-E-V-1-2 Wed-A-V-2-3 Tue-A-V-5-4	215 215 112 103 143 91	White, Ka Wiechma Wiesler, S Wiesner, Wiest Th

Wang, Wenchao	Fri-M-V-7-5	237
Wong Wonwa	Fri-A-V-4-7	250
wallg, wellwu	Thu-M-V-3-4	185
Wang, Wupeng	Thu-A-V-3-12	206
Wang, Xiaobo	Fri-A-V-3-2	247
Wang, Xiaofei	Wed-E-V-5-3	171
	Inu-A-V-2-2 Fri-Δ-V-1-2	202
Wang, Xiaoqiang	Wed-A-V-2-2	143
Wang, Xiaorui	Thu-M-SS-1-3	194
Wang, Xingming	Fri-M-V-6-10	235
Wang, Xin	Thu-M-SS-1-3	194
walig, Alli	Fri-M-V-7-1	230
	Fri-M-V-7-9	238
Wang, Xiong	Fri-M-V-3-5	228
Wang Yuwang	Fri-A-V-3-11	249
wallg, Auyallg	Fri-M-V-6-10	235
Wang, Yang	Wed-E-V-5-4	171
Wang, Yannan	Wed-M-SS-1-4	136
Wang, Yih-Wen		63
wallg, rollgqillg	Thu-A-V-6-13	212
Wang, Yuguang	Tue-M-V-2-1	64
Wang, Yuhua	Fri-A-V-2-9	247
Wang, Yujun	Thu-A-V-6-13	214
wallg, rull	1ue-E-V-3-13 Wed-M-V-4-12	130
Wang, Yu-Xuan	Thu-M-V-4-6	188
Wang, Yuxuan	Tue-M-V-3-7	67
Wang, Zhichao	Tue-A-V-6-6	93
Wang, Zhiming	Tue-E-V-1-/	104
Wani, Yash R	Tue-A-SS-2-6	97
Wanner, Leo	Wed-A-S&T-1-5	157
Warden, Pete	Fri-M-V-6-5	234
watanabe, Shinji	Tue-A-V-1-13	7 Z 84
	Tue-E-V-3-9	109
	Tue-E-V-4-10	112
	Wed-M-O-1-4	119
	Wed-M-V-2-8	121
	Wed-M-V-6-14	135
	Wed-E-V-1-12	162
	<u> </u>	185
	Thu-M-V-4-8	188
	Thu-A-V-1-11	202
	Thu-A-V-3-9	206
	Thu-A-V-4-3 Thu-A-V-6-5	217
	Thu-A-SS-1-2	214
	Thu-A-SS-1-6	215
	Thu-A-SS-1-7	215
Waters. Austin	Thu-A-33-1-9	182
Watson, Catherine I	Wed-E-V-6-5	174
Wei, Bo	Tue-M-V-6-9	74
Wei, wengi Weingartová Lenka	wea-M-V-3-9 Tue-A-S&T-1-2	127 90
Weiss, Ron J.	Wed-E-V-3-8	166
	Thu-A-SS-1-11	216
Weld, Henry	Fri-A-V-6-10	255
Wen, Liang	Wed-M-V-4-3	120
Wen, Zhengqi	Tue-M-V-5-1	70
Mar. 7	Fri-M-V-3-1	227
Wen, Zujie	<u>Fri-A-S&I-I-/</u>	261
weng, endo	Tue-E-V-2-2	106
	Thu-A-V-6-5	212
Werner, Raphael	Thu-M-V-6-5	191
	<u> </u>	225
Westerman, Ryan	Thu-A-V-2-9	204
Weston, Heather	Thu-A-O-1-6	199
Whang, Taesun	Fri-M-V-6-8	235
white, Hannan White, Katie M	іue-E-SS-I-I Тие-А-V-2-6	117
Wiechmann, Daniel	Thu-A- <u>SS-2-6</u>	217
Wiesler, Simon	Wed-M-V-6-7	134
wiesner, Matthew		179
Wiest, Thomas	Thu-M-V-5-8	190
Wilbrandt Alexander	Thu.M.S&T.1.2	106

Willett Daniel	Wed-E-V-1-1	160
Williams Jason		111
Williamaan Danald C		111
williamson, Donald S	1ue-M-V-3-12	68
	Wed-A-V-6-12	154
Winata. Genta Indra	Wed-E-V-1-8	162
Wintrode Ionathan	Tue-Δ-V-5-7	92
Withrock Michael		202
WILDIOCK, MICHAEI	1 nu-A-V-1-10	202
Witkowski, Marcin	Fri-M-O-3-2	221
Witzlack-Makarevich, A.	Wed-M-O-2-3	120
Włodarczak, Marcin	Tue-A-V-5-10	92
modulezady marchi mi		101
Western Destate	Tue-E-0-2-4	101
Wochner, Daniela	Fri-M-V-1-3	223
Woisard, Virginie	Tue-M-O-2-3	61
Wolf. Lior	Fri-M-O-3-1	221
Wolff Matthias	Tue-F-0-2-3	101
Woltron Thomas		60
Wolffold, Illollias	Tue-M-0-1-1	157
Wood, Chris	Wed-A-S&T-1-4	157
Woodland, Philip C	Wed-E-V-3-15	168
	Fri-M-V-3-8	229
Wren Vyonne	Thu-M-V-1-8	181
Wright Ionathan		120
Wilgilt, Jollatliali	wed-A-O-2-2	150
Wright, Richard A	Wed-A-V-2-5	143
Wu, Anne	Wed-A-V-6-3	153
	Wed-A-V-6-4	153
Wu Rin	Wed-M-V-5-7	131
Wu Bo		106
	Tue-E-V-Z-2	144
wu, Bo	Wed-A-V-2-8	144
Wu, Chunyang	Tue-M-V-2-8	66
	Wed-A-V-3-1	145
	Wed-A-V-3-14	147
		240
W D	Fri-A-V-5-0	240
Wu, Di	Fri-M-V-3-5	228
Wu, Felix	Wed-M-V-6-14	135
Wu, Haibin	Fri-M-V-7-8	238
Wu Iianhan	Тие-А-V-1-11	84
Wu Jian		18/
wu, jiali		104
	I hu-M-V-3-9	180
	Thu-A-V-6-4	212
Wu, Jian	Fri-M-V-3-11	229
Wil. Mengvile	Tue-M-V-6-11	75
Wu Ming-hui	Eri.M.SS.1.3	230
	FIFWF33-1-3	233
wu, Minglin	Fri-M-V-1-10	224
Wu, Minhua	Tue-A-V-4-2	89
	Thu-A-V-2-7	203
Wu. Sarah	Fri-A-SS-2-15	260
Wii Ting-Wei	Tue-M-V-4-3	69
wa, mig wei		111
147 37	Tue-E-V-4-5	111
wu, Xixin	I ue-M-V-2-6	65
	Thu-A-SS-1-13	216
	Fri-M-V-7-12	238
	Fri-A-SS-2-8	258
Wii Vanfeng	Tue-M-V-1-4	63
Wii Vomi		101
	Tue-E-O-2-2	101
WU, YIDO	lue-E-V-I-II	105
Wu, Yi-Chiao	Wed-A-V-5-3	150
	Thu-A-V-5-9	210
Wu. Yifei	Thu-M-V-2-11	184
Wu Vonghui	Tue-M-V-2-7	65
wa, rongnar		66
T17 T7	Tue-M-V-2-9	100
wu, ru	Wed-E-V-3-14	10/
	Thu-M-V-3-1	184
	Thu-M-V-3-9	186
	Thu-A-V-2-2	202
	Thu-A-V-2-10	204
Way Zolin		145
wu, Zeiiii	wed-A-V-3-4	145
Wu, Zhiyong	Tue-A-V-6-9	94
	Thu-A-SS-1-13	216
	Fri-M-V-7-8	238
	Fri-A-V-5-7	252
Wu Zhizheng	Tuo-E-V-6-11	116
Wu, Zhizheng		110
wu, zhongqin	Tue-E-V-5-6	115
	Thu-M-V-2-11	184
	Thu-M-SS-1-5	194
	Thu-A-V-1-2	200
Wykes, Til	Tue-A-V-2-6	85
		00
X		
4 b		
Xia, Rui	Tue-M <u>-V-3-7</u>	67
Xia, Rui Xia, Tong	Tue-M-V-3-7 Tue-M-SS-1-1	67 77
Xia, Rui Xia, Tong	Tue-M-V-3-7 Tue-M-SS-1-1	67 77 181
Xia, Rui Xia, Tong	Tue-M-V-3-7 Tue-M-SS-1-1 Thu-M-V-1-9	67 77 181
Xia, Rui Xia, Tong Xia, Yangyang	Tue-M-V-3-7 Tue-M-SS-1-1 Thu-M-V-1-9 Thu-A-V-1-1	67 77 181 200
Xia, Rui Xia, Tong Xia, Yangyang	Tue-M-V-3-7 Tue-M-SS-1-1 Thu-M-V-1-9 Thu-A-V-1-1 Fri-A-O-1-2	67 77 181 200 241
Xia, Rui Xia, Tong Xia, Yangyang Xiang, Lu	Tue-M-V-3-7 Tue-M-SS-1-1 Thu-M-V-1-9 Thu-A-V-1-1 Fri-A-O-1-2 Fri-A-V-6-2	67 77 181 200 241 254
Xia, Rui Xia, Tong Xia, Yangyang Xiang, Lu Xiao, Alex	Tue-M-V-3-7 Tue-M-SS-1-1 Thu-M-V-1-9 Thu-A-V-1-1 Fri-A-O-1-2 Fri-A-V-6-2 Wed-A-V-3-1	67 77 181 200 241 254 145
Xia, Rui Xia, Tong Xia, Yangyang Xiang, Lu Xiao, Alex	Tue-M-V-3-7 Tue-M-SS-1-1 Thu-M-V-1-9 Thu-A-V-1-1 Fri-A-O-1-2 Fri-A-V-6-2 Wed-A-V-3-14	67 77 181 200 241 254 145 147
Xia, Rui Xia, Tong Xia, Yangyang Xiang, Lu Xiao, Alex Xiao, Jachyn	Tue-M-V-3-7 Tue-M-SS-1-1 Thu-M-V-1-9 Thu-A-V-1-1 Fri-A-O-1-2 Fri-A-V-6-2 Wed-A-V-3-1 Wed-A-V-3-14 Eri: M-V-5-2	67 77 181 200 241 254 145 147 231

INTERSPEECH 2021 277 BRNO | CZECHIA

Xiao, Jing	Tue-A-O-1-5	80	Xue, Yawen	Thu-M-V-4-8	188
	Tue-A-V-1-11	84	V		
	Tue-E-V-3-4	108	1		
	Tue-E-V-5-11	114	Yadama, Tarun Reddy	Wed-A-S&T-1-7	157
	Wed-M-V-3-5	127	Yadav, Ravi	Thu-M-V-1-11	182
	wed-M-V-3-9	150	Yamada, Takeshi	Wed-A-V-1-3	141
		200	Yamagishi, Junichi	Fri-M-V-7-1	236
		209		Fri-M-V-7-2	237
	<u> </u>	214		Fri-M-V-7-9	238
Vino Dungiu		250	Yamamoto, Ayako	Tue-M-V-3-5	67
Viao Woj		230	Yamamoto, Ryuichi	Wed-A-V-5-11	152
Viao Viao		Q1		Thu-M-V-5-1	188
Via Eanglong		211	Yamazaki, Yoshihiro	Tue-M-V-4-5	69
Vio Loi		03	Yan, Brian	Wed-E-V-1-12	162
Ме, Lei	Tue-A-V-0-0	105	Yan, Chao	Wed-A-V-2-9	144
	Tue-E-V-1-12	116	Yan, Jie	Wed-A-SS-1-7	156
	Wed-A-V-5-2	150	Yan, Jinghao	Fri-M-SS-1-5	240
	Wed-F-SS-1-5	176	Yan, Tianhao	Fri-M-V-5-5	232
	Thu-A-V-6-4	212	Yan, Yonghong	Tue-E-V-2-3	106
i	Thu-A-SS-1-2	214		Wed-M-V-3-7	127
i	Fri-M-V-3-5	228		Fri-M-V-4-9	231
i	Fri-M-V-6-11	236	Yan, Yuzi	Fri-A-V-5-6	252
	Fri-A-V-3-11	249	Yan, Zhijie	Thu-A-V-4-2	207
	Fri-A-V-5-4	252	Yan, Zhiyong	Thu-A-V-6-5	212
	Fri-A-V-5-10	253		Thu-A-V-6-13	214
	Fri-A-SS-1-3	256	rang, Chao	Fri-M-V-3-5	228
Xie, Xurong	Fri-A-SS-2-4	257	rang, Chao-Han Huck	Tue-A-SS-1-6	- 95
	Fri-A-SS-2-6	258	rang, Dongchao	Tue-E-V-3-2	108
l	Fri-A-SS-2-7	258	rang, Eunho	Wed-E-V-3-1	165
Vie Ver	Fri-A-SS-2-9	259	rang, Feiran	Wed-M-SS-1-2	136
Ale, Yang	Fri-M-V-7-3	237	Vang For-	wed-M-SS-T-3	136
Ale, Yanlu	Wed-M-V-5-7	131	I dily, Fefig	Tue-M-V-3-3	- 07
Xie, Zongxia	Wed-E-V-3-9	100	Yang Cone Ding	Tue-A-V-6-6	100
Xin, Detai	Wed-M-V-3-6	127	Yang Homin		171
Xing, Chao	Wed-A-V-3-13	147	Vang Jinbuook	Wed-E-V-5-3	151
Vieng Coiming	Wed-M-V-3-4	160	Vang Jun	Wed-A-V-5-6	126
Viong Dovi	Wed-E-V-1-0	102	Vang Jun	Eri. A.V.2.5	2/8
Xiong Shengwii	Tue-F-V-1-4	103	Yang Lidong	Tue-F-V-2-1	105
Xiong Shifu	Tue-M-V-6-5	73	Yang, Liniu	Tue-M-V-3-1	66
Xiong, Tao	Tue-F-V-1-7	104	Yang, Meirong	Tue-M-V-6-9	74
Xiong, Yixuan	Fri-M-V-2-5	225	Yang, Minghui	Fri-A-S&T-1-7	261
Xiu, Zhiping	Tue-M-V-2-8	66	Yang, Shan	Wed-A-V-5-2	150
Xu, Angi	Thu-M-V-6-6	191		Fri-A-V-5-4	252
Xu, Bo	Tue-E-V-2-4	106	Yang, Shu-wen	Tue-E-V-3-9	109
	Wed-M-V-1-5	122	Yang, Song	Tue-E-V-5-6	113
Xu, Changliang	Wed-M-V-4-4	128		Thu-M-V-2-11	184
Xu, Chenglin	Wed-M-O-3-2	121	X 7	Thu-M-SS-1-5	194
	Thu-A-V-3-12	206	Yang, Wel	Tue-A-SS-T-9	96
V D	Fri-M-O-3-4	222	Yang, Xuerui	Wed-A-V-2-9	144
Xu, Doligxidilg	Wed-E-V-5-4	104	Vang Vuguang		105
Xu, Furong		112	Vang Zhonchuan		116
Yu Haibua	Wod-M-V-1-7	123	Tang, Zhenenuan	Wod-M-V-2-4	126
Xu Hainan	Wed-F-V-3-7	166	Yao Kaisheng	Tue-F-V-1-7	104
X_{11} Jiaming	Tue-F-V-2-4	106	Vao Shengvii	Wed-A-SS-1-7	156
Xu. Jing	Tue-M-V-1-4	63	Yao, Xuchen	Thu-A-V-6-5	212
Xu, Liang	Tue-E-V-2-1	105	Yao, Yiwu	Fri-A-V-3-5	248
Xu. Menglong	Wed-A-V-3-7	146	Yao, Zhuoyuan	Fri-M-V-3-5	228
Xu, Min	Wed-M-V-5-6	131	Yarra, Chiranjeevi	Tue-E-V-3-5	108
Xu, Mingxing	Fri-M-V-7-4	237		Wed-E-V-1-7	161
Xu, Qiantong	Tue-M-V-5-10	72	Yatabe, Kohei	Thu-M-V-6-9	192
-	Tue-A-V-4-4	89		Fri-A-S&T-1-1	260
	Tue-A-V-4-8	90	Ye, Guoli	Tue-E-V-5-9	114
Xu, Shugong	Fri-M-V-5-3	232		Wed-E-V-3-14	167
Xu, Weiyuan	Tue-M-V-4-6	70		Thu-A-V-2-2	202
Xu, Xiaoshuo	Tue-A-V-4-1	88	Vo Jioping	Fri-A-V-1-2	243
Vu Vin	Fri-A-V-T-5	243	Ve Pong		154
ли, лш	wed-E-V-6-4	212	Ye Shuaishuai	Thu-M-SS-1-4	104
Xu. Xinkang	Thu-M-SS-1-4	194	Ye. Zi	Fri-A-SS-2-4	257
Xu, Xinmeng	Wed-F-V-5-4	171	10, 21	Fri-A-SS-2-9	259
Xu. Xuenan	Tue-M-V-6-11	75	Yeh, Ching-Feng	Wed-A-V-2-1	142
Xu, Xuexin	Tue-E-V-6-9	116		Wed-A-V-3-1	145
Xu, Yi	Thu-M-V-6-6	191	Yemini, Yochai	Tue-E-V-2-6	106
i	Fri-M-V-2-2	225	Yeo, Eun Jung	Fri-A-SS-2-13	259
Xu, Yong	Tue-E-V-2-2	106	Yeung, Yu Ting	Tue-E-V-6-5	115
	Tue-E-V-2-4	106		Thu-M-V-1-10	182
	Wed-A-V-4-7	149	Vi lionarra	Fri-M-V-6-9	235
Vu Vuorbu	Thu-M-V-3-11	180	ri, jiangyan	Tue-M-V-5-1	- 70
AU, YUEZNU	Fri-A-V-2-9	247		Tue-A-SS-T-7	120
AU, ZIYI Yuo, Cheng	Wed-E-SS-T-3	127		wea-M-V-4-5	120
Aue, Cheng	Wed-M-SS-T-7	150	Yin, Bao-Cai	Thu-M-V-3-1	183
Xue liabin	Wed-F-V-3-4	165	Yin. Dacheng	Thu-A-V-5-3	209
Xue, Jian	Thu-A-V-2-10	204	Yin, Hengxin	Tue-E-V-5-1	112
Xue, Wei	Thu-M-V-1-1	180	Yin, Shouyi	Thu-M-SS-1-3	194

Yin, Xiang	Thu-M-V-5-5	189
Yin, Xuefeng	Thu-A-SS-2-6	217
Yokoyama, Hiroshi	Thu-M-V-6-4	191
Yoneyama, Reo	Wed-A-V-5-3	150
Yoon, Jaesam	Wed-A-V-5-7	151
Yoshinaga Taukaga		0/
TOSIIIIaga, TSukasa		101
Voshioka Takuva	Wed-E-V-5-3	171
rosinoka, rakuya	Thu-M-V-3-1	184
	Thu-M-V-3-9	186
	Thu-A-V-2-2	202
	Fri-A-V-1-2	243
You, Chenyu	Tue-M-V-4-2	69
	Tue-M-V-4-6	70
¥7	Thu-M-V-7-1	192
You, Jaeseong	Wed-A-V-5-1	150
You, Zhao	Wed-A-V-3-8	140
Vousefi Midia	1 nu-A-V-6-5	121
Yu Ring	Wed-M-SS-1-1	136
	Wed-F-SS-1-7	176
Yu. Cheng	Tue-M-V-3-8	68
,	Tue-M-V-3-9	68
Yu, Chengzhu	Tue-E-V-6-10	116
Yu, Dong	Tue-M-V-5-11	72
	Tue-E-V-2-2	106
	Tue-E-V-2-4	106
	Wed-A-V-3-8	146
	Wed-A-V-4-7	149
	Fri-A-V-4-5	250
Yu. Fan	Fri-M-V-3-5	228
Yu, Guochen	Wed-E-SS-1-2	175
Yu, Guoxing	Wed-M-V-5-11	132
Yu, Haibin	Thu-M-SS-1-5	194
Yu, Jiahui	Wed-M-V-6-4	133
Yu, Jianwei	Tue-E-V-2-2	106
	Fri-A-SS-2-4	257
		250
Vu Kai	Tue-M-V-6-11	233
1 u, Kui	Wed-A-V-2-10	144
	Thu-M-V-5-3	189
Yu, Mei	Tue-A-V-2-5	85
	Thu-M-V-2-4	183
Yu, Meng	Tue-E-V-2-2	106
	Tue-E-V-2-4	106
	Wed-A-V-4-7	149
Vu Duiguo		180
V_{11} Shi	Eri.M.\/.2.9	226
Yu Xiangzhan	Thu-M-V-2-3	184
Yu. Yan	Fri-M-SS-1-1	239
Yu, Zhesong	Tue-M-V-6-1	73
Yuan, Jiahong	Fri-A-V-2-7	246
Yuan, Weitao	Thu-M-V-3-4	185
Yue, Xianghu	Tue-A-V-4-9	90
Yue, Zhengjun	Thu-M-V-1-8	181
Yun, Jaewoong	Tue-A-V-5-9	92
Yurong	Fri-M-V-1-1	222
rurt, Metenan	Iue-M-O-3-2	- 62 242
i usui, bolaji Vyon Francois	Fri-A-0-2-2	242
1 von, mançois	Fn-A-0-2-5	2 4 2
Z		
7ahnan Dittan Vathani		122
Zanner-Ritter, Katharina	Wed-M-V-5-9	132

Lamer Miller, Nathanna	weu-wi-v-J-J	104
	Wed-E-V-4-9	169
	Fri-M-V-1-3	223
Zaiem, Salah	Thu-M-O-1-4	177
Zainkó, Csaba	Wed-A-V-5-8	151
Zaitova, Iuliia	Fri-M-V-6-1	234
Zandie, Rohola	Wed-E-V-6-3	173
Zanon Boito, Marcely	Wed-M-O-1-5	119
Zaporowski, Szymon	Fri-M-V-1-2	222
Zare, Ali	Tue-E-V-6-6	115
Zarrieß, Sina	Thu-A-O-1-3	198
Zayats, Vicky	Tue-A-V-5-3	91
Zdansky, Jindrich	Wed-M-O-3-3	121
Zeineldeen, Mohammad	Thu-M-O-1-5	177
	Thu-M-O-3-2	179

Żelasko, Piotr	Tue-M-V-6-10	74	Zhang, Shi-Xiong	Tue-E-V-2-2	106	Zh
	Tue-A-O-1-2	80		Tue-E-V-2-4	106	
	Tue-F-V-1-9	104		Wed-A-V-4-7	149	7h
		124			196	211
	wea-m-v-z-z	124		I nu-M-V-3-1 I	160	
	Thu-A-V-2-9	204		Fri-A-V-4-5	250	
	Thu-A-SS-1-12	216	Zhang, Shuai	Tue-M-V-5-1	70	Zh
	Thu-A-SS-2-10	218	0.	Fri-M-V-3-1	227	7h
	Eri-M-V-7-10	238	7hang Shucong	Wod E V 2 2	165	211
Zellene Mengenet		120		weu-E-v-5-5	105	
Zellers, Margaret	Wed-M-O-2-3	120	Zhang, Sijia	Wed-E-V-4-11	170	Zh
Zellou, Georgia	Tue-A-O-2-2	81	Zhang, Siyuan	Tue-A-V-3-3	87	Zh
Zen. Heiga	Tue-M-V-2-7	65	Zhang Tao	Tue-M-V-6-9	74	7h
		66	Zhang, Huo		167	76
	Tue-IVI-V-2-9	00	Zhang, wei	wed-E-V-3-12	107	ΖΠ
	Tue-A-V-4-7	90	Zhang, Weibin	Tue-E-V-6-10	116	
	Thu-A-SS-1-11	216	-	Wed-M-V-3-4	126	Zh
Zeng. Biao	Wed-M-V-5-11	132	7hang Weicheng	Tuo.F.V.4.4	111	7h
Zong Michael		110			104	21
		200	Zhang, wel-Qlang	1 nu-M-SS-1-5	194	Zn
zeng, wenjun	I hu-A-V-5-3	209		Thu-A-V-6-5	212	Zh
Zerr, Benoît	Thu-A-V-3-13	207		Fri-M-SS-1-5	240	
Zever. Albert	Wed-A-V-3-3	145		Eri. A.V. 5.6	252	7h
,,		177	7hang Via		250	211
		170		FIFA-33-2-11	233	-1
	Тпи-м-О-3-3	179	Zhang, Xianwei	Fri-A-V-6-7	255	Zh
Zhan, Haoyue	Wed-M-V-3-3	126	Zhang, Xiao-Lei	Tue-M-V-5-6	71	Zh
Zhan, Jinsong	Fri-A-V-1-9	244	8,	Wod-A-V-3-7	146	7h
7han Duming	Wod A V 2 7	1/13	7hong Vice Oi		171	71
	weu-A-v-2-7	173		Wed-E-V-5-6	1/1	Zn
Zhang, Baiyan	Fri-M-V-7-7	238	Zhang, Xinyi	Wed-M-V-1-1	122	Zh
Zhang, Binbin	Fri-M-V-3-5	228	Zhang, Xu	Wed-M-SS-1-1	136	Zh
Zhang, Chao	Wed-F-V-3-15	168	<u>,</u>	Wed-F-SS-1-7	176	7h
7hang Chan	Wod M SS 1-1	126	7hong Vudong		170	ъц 71
Linalig, Clicii	weu-M-SS-T-T	130	Zinang, Audong	wed-E-V-5-11	1/2	Zn
	Wed-E-SS-1-7	176	Zhang, Xueliang	Wed-M-SS-1-5	136	Ζh
Zhang, Chuanyao	Tue-A-V-1-11	84	-	Wed-A-SS-1-6	156	Zh
Zhang, Chunlei	Wed-A-V-4-7	149		Wed-E-SS-1-6	176	7h
Zhang Cong		171	Thong Vong	Weu-E-55-1-0	220	ப
	Wed-E-V-5-4	1/1	Zhang, Yang	FLI-M-V-7-8	238	
Zhang, Cong	Tue-E-V-3-10	109	Zhang, Yang	Tue-E-V-5-8	113	
Zhang, Dawei	Thu-M-SS-1-3	194		Wed-F-V-5-2	171	Zh
Zhang Deiun	Eri.A.V.6.5	254	7hang Vang	WodEV69	174	7h
	FIT-A-V-U-J	120		weu-E-V-0-8	1/4	71
Zhang, Didi	Wed-M-V-4-12	130		Fri-A-V-1-13	245	Zn
Zhang, Dong	Fri-M-V-6-10	235		Fri-A-S&T-1-3	260	Zh
Zhang, Guangyan	Thu-M-V-5-7	190	Zhang, Yike	Tue-A-V-4-1	88	Zh
Linang, Guang/an IIIII		252	7hang Vinning	Eri-M-V-6-10	235	7h
7hours Holtows	FIFA-V-3-0	232	Zhang, Imping		140	211
Znang, Halteng	Fri-M-V-4-3	230	Znang, 11	wed-A-V-3-6	146	
Zhang, Haitong	Wed-M-V-3-3	126	Zhang, Yi	Thu-A-V-3-11	206	Zh
Zhang, Hao	Fri-M-V-3-3	228	Zhang, Yi	Thu-M-V-7-7	193	Zh
Zhang Hao		88	Zhang Vou	Eri M \/ 7 11	238	Zh
	Tue-A-V-5-0	107			230	7h
_	Tue-E-V-2-8	107	Zhang, Yu	I ue-M-V-2-7	65	21
Zhang, Haoyu	Thu-M-V-5-5	189		Tue-M-V-2-9	66	Zn
Zhang, Haozhe	Wed-M-V-3-7	127		Tue-A-V-4-7	90	
		221		Thu-A-V-2-8	204	Zh
7hana Haraan	F11-IVI-V-4-9	231			204	7h
Znang, Huayun	Fri-A-V-1-8	244		I nu-A-SS-I-3	214	71
	Fri-A-S&T-1-6	261		Thu-A-SS-1-11	216	Zn
Zhang, Hui	Thu-A-V-3-11	206		Fri-M-V-3-8	229	Zh
7hang Jiaiun	Eri.A.V.6.2	254		Fri-M-V-3-9	229	Zh
Zhang, Jiajun	T N1/212	234			220	7h
Znang, Jianwei	Tue-M-V-3-13	68	7hana Vaalaai		110	71
Zhang, Jicheng	Wed-M-V-1-7	123	Znang, ruekai	Wed-M-O-T-4	119	
Zhang, Jie	Fri-M-SS-1-3	239	Zhang, Yu-Jia	Tue-M-V-1-3	63	Zh
Zhang Jinsong	Wod-M-V-5-7	131		Fri-M-SS-1-4	240	
Zitang, Jinsong		224	7hang Vuging	Wed-M-V-5-7	131	7h
_	Fri-M-V-1-9	224	Zhang, ruqing		131	7h
Zhang, Jisi	Thu-A-V-3-4	205		Fri-M-V-I-9	224	
Zhang, Ju	Tue-M-V-2-1	64	Zhang, Yuxiang	Fri-M-V-7-5	237	Ζh
7hang lun	Wod M V 6 8	13/	Zhang, Zhenchuan	Fri-M-V-7-3	237	Zh
Zhang herebe	weu-w-v-0-8	1010	7hang 7hihong	Τμο-Ε-V-6-0	116	7h
Zitalig, Jundo	Thu-A-V-6-5	212	Zhang Zhihai	T	100	71
	Thu-A-V-6-13	214		Tue-E-V-I-4	103	<u>ح</u> یا
Zhang, Kaili	Tue-M-V-2-5	65	Znang, Zhiwen	Thu-A-V-6-13	214	Ľ١,
Zhang, Kanghao	Wed-E-SS-1-6	176	Zhang, Zhivu	Fri-M-V-4-4	230	Zil
7hang Voyan	Mod May 4-4	120	Zhang, Zhuohuang	Thu-M-V-3-11	186	Zil
	wea-m-v-4-4	120	7hao Chandong		200	7n
Zhang, Leying	Wed-A-O-2-4	139	Zhao, Chenuong	Thu-A-V-5-1	209	211
Zhang, Li	Tue-E-V-1-12	105	znao, Chunyu	Fri-M-V-6-11	236	20
7hang Lin	Fri-M-V-7-2	237	Zhao, Ding	Fri-M-V-6-3	234	Zo
Zhang Dai		145	7hao Guanlong	Wed-A-V-1-4	141	Ζo
Linding, rel	wed-A-V-3-2	140	Linuo, Guainong		101	70
Zhang, Peng	Wed-A-SS-1-6	156		Thu-M-V-1-6	101	20
Zhang, Pengvuan	Tue-A-V-6-5	93	Znao, Hang	Tue-E-V-6-1	114	
	Wed-M-V-2-7	127	Zhao, Huan	Thu-M-V-2-3	183	
		170	Zhao, Jing	Thu-M-SS-1-5	194	Zo
	wed-E-V-5-10	1/2			240	
	Fri-M-V-4-9	231	7h and the state	FTI-M-55-1-5	240	
	Fri-M-V-7-5	237	znao, jinming	Fri-A-V-2-3	240	
	Fri-A-V-4-7	250	Zhao, Junan	Tue-M-V-1-4	63	
Zhang Qian	Thu-A-V-2-4	203	Zhao, Liang	Wed-E-V-5-11	172	
Zhang Qian		200	Zhao Rui		90	
Znang, Qianyutong	Fri-M-V-1-4	223	Ziiau, Kui	Tue-A-V-4-10	30	70
Zhang, Qinglin		91	znao, sneng	Wed-A-V-2-2	143	7
	Tue-A-V-5-4				050	7.11
	<u>Tue-A-V-5-4</u> Tue-E-V-4-6	111		Fri-A-V-5-6	252	20
7hang Qiquan	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2	111	Zhao, Shuaijiang	Fri-A-V-5-6 Tue-E-V-5-1	252 112	Zu
Zhang, Qiquan	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2	111 66	Zhao, Shuaijiang	Fri-A-V-5-6 Tue-E-V-5-1	252 112 212	Zu Zw
Zhang, Qiquan Zhang, Shaoji	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2 Wed-E-V-6-4	111 66 173	Zhao, Shuaijiang	Fri-A-V-5-6 Tue-E-V-5-1 Thu-A-V-6-5	252 112 212	Zu Zw
Zhang, Qiquan Zhang, Shaoji Zhang, Shilei	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2 Wed-E-V-6-4 Wed-A- <u>SS-1-5</u>	111 66 173 156	Zhao, Shuaijiang Zhao, Xi	Fri-A-V-5-6 Tue-E-V-5-1 Thu-A-V-6-5 Tue-E-V-5-1	252 112 212 112	Zu Zw
Zhang, Qiquan Zhang, Shaoji Zhang, Shilei Zhang, Shiliang	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2 Wed-E-V-6-4 Wed-A-SS-1-5 Thu-A-V-4-2	111 66 173 156 207	Zhao, Shuaijiang Zhao, Xi Zhao, Xintao	Fri-A-V-5-6 Tue-E-V-5-1 Thu-A-V-6-5 Tue-E-V-5-1 Tue-A-V-6-9	252 112 212 112 94	Zu Zw
Zhang, Qiquan Zhang, Shaoji Zhang, Shilei Zhang, Shiliang	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2 Wed-E-V-6-4 Wed-A-SS-1-5 Thu-A-V-4-2	111 66 173 156 207 248	Zhao, Shuaijiang Zhao, Xi Zhao, Xintao Zhao, Yucheng	Fri-A-V-5-6 Tue-E-V-5-1 Thu-A-V-6-5 Tue-E-V-5-1 Tue-A-V-6-9 Thu-A-V-5-3	252 112 212 112 94 209	Zu Zw
Zhang, Qiquan Zhang, Shaoji Zhang, Shilei Zhang, Shiliang	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2 Wed-E-V-6-4 Wed-A-SS-1-5 Thu-A-V-4-2 Fri-A-V-3-5	111 66 173 156 207 248	Zhao, Shuaijiang Zhao, Xi Zhao, Xintao Zhao, Yucheng Zhao, Yur	Fri-A-V-5-6 Tue-E-V-5-1 Thu-A-V-6-5 Tue-E-V-5-1 Tue-A-V-6-9 Thu-A-V-5-3	252 112 212 112 94 209	Zu Zw
Zhang, Qiquan Zhang, Shaoji Zhang, Shilei Zhang, Shiliang Zhang, Shimin	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2 Wed-A-SS-1-5 Thu-A-V-4-2 Fri-A-V-3-5 Wed-E-SS-1-5	111 66 173 156 207 248 176	Zhao, Shuaijiang Zhao, Xi Zhao, Xintao Zhao, Yucheng Zhao, Yun	FrI-A-V-5-6 Tue-E-V-5-1 Thu-A-V-6-5 Tue-E-V-5-1 Tue-A-V-6-9 Thu-A-V-5-3 Wed-A-V-2-9	252 112 212 112 94 209 144	Zu Zw
Zhang, Qiquan Zhang, Shaoji Zhang, Shilei Zhang, Shiliang Zhang, Shimin	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2 Wed-E-V-6-4 Wed-A-SS-1-5 Thu-A-V-4-2 Fri-A-V-3-5 Wed-E-SS-1-5 Fri-A-SS-1-3	111 66 173 156 207 248 176 256	Zhao, Shuaijiang Zhao, Xi Zhao, Xintao Zhao, Yucheng Zhao, Yun Zhao, Yun	Fr1-A-V-5-6 Tue-E-V-5-1 Thu-A-V-6-5 Tue-E-V-5-1 Tue-A-V-6-9 Thu-A-V-5-3 Wed-A-V-2-9 Wed-E-V-5-12	252 112 212 112 94 209 144 172	Zu Zw
Zhang, Qiquan Zhang, Shaoji Zhang, Shilei Zhang, Shiliang Zhang, Shimin Zhang, Shimin	Tue-A-V-5-4 Tue-E-V-4-6 Tue-M-V-3-2 Wed-E-V-6-4 Wed-A-SS-1-5 Thu-A-V-4-2 Fri-A-V-3-5 Wed-E-SS-1-5 Fri-A-SS-1-3 Wed-A-V-3-2	111 66 173 156 207 248 176 256 145	Zhao, Shuaijiang Zhao, Xi Zhao, Xintao Zhao, Yucheng Zhao, Yun Zhao, Yunxin Zhao, Zhengdao	FrI-A-V-5-6 Tue-E-V-5-1 Thu-A-V-6-5 Tue-E-V-5-1 Tue-A-V-6-9 Thu-A-V-5-3 Wed-A-V-2-9 Wed-E-V-5-12 Fri-A-V-2-9	252 112 212 112 94 209 144 172 247	Zu Zw

Zhao, Zhou	Wed-M-V-4-4	128
7hong Chongohi	Wed-E-V-6-6	174
Zneng, Cnengsni	I UE-M-V-3-6	175
	Fri-A-SS-1-5	256
Zheng, Nengheng	Wed-A-V-4-10	149
Zheng, Renjie	Wed-A-V-6-1	152
71 0	Fri-A-V-2-7	246
Zheng, Siqi	Thu-A-V-4-2	207
Zheng Tieran	Wed-E-V-3-4	165
Zheng, Xiguang	Wed-M-SS-1-1	136
	Wed-E-SS-1-7	176
Zheng, Xu	Tue-A-V-1-3	82
Zheng, Yuxing	Wed-M-V-4-3	128
Zheng, Zuoyun	Thu-M-O-3-2	1/9
2.111, 1.111111g	Fri-A-V-4-10	251
Zhong, Xionghu	Wed-A-V-4-13	150
	Fri-M-V-5-12	233
Zhou, Hang	Fri-M-V-6-9	235
Zhou, Hengsnun	Tue-M-V-6-5	107
Zhou, Huali Zhou liatong	Fri-A-V-3-6	248
Zhou. Jiavu	Thu-A-SS-2-11	218
Zhou, Kun	Tue-A-V-6-2	93
Zhou, Peilin	Tue-M-V-4-6	70
Zhou, Shiyu	Wed-M-V-1-5	122
znou, nafiyan Zhou Wei		102
Zhou, Wei	Thu-M-O-3-2	179
	Thu-M-O-3-3	179
	Fri-A-V-1-10	244
Zhou, Xiao	Fri-M-V-4-7	231
Znou, Ainyong	Tuo-E-V-6-11	93
Zhou, Yu	Fri-A-V-6-2	254
Zhou, Yuanfu	Wed-A-V-2-9	144
Zhou, Zhikai	Tue-E-V-5-2	112
71	Wed-A-SS-1-8	156
Zhou, Zimu Zhu Bilei	Thu-M-V-7-2	192
Zhu, bhenguang	Tue-E-V-4-1	110
Zhu, Dawei	Tue-A-S&T-1-4	99
Zhu, Ge	Tue-M-V-1-7	64
7hu Hongning	Fri-M-V-7-11	238
Ziiu, Holigiiiig Zhu Huifeng	Tue-M-V-1-9	66
Zhu, Huijia	Tue-E-V-1-7	104
Zhu, Jian	Tue-E-V-3-10	109
Zhu, Pai	Tue-M-V-6-3	73
Zhu, Qingying	Wed-M-V-3-5	127
Zhu, Qiu-Shi Zhu Rui	Wed-M-SS-1-2	136
Liiu, Kui	Wed-M-SS-1-3	136
Zhu, Tinglong	Tue-M-V-1-5	63
Zhu, Xinghua	Wed-M-V-3-9	127
ZIIU, YOUXIANG	Thu-A-SS-2-3	217
Zhu, Tuni Zhu, Yunzheng	Tue-F-V-5-3	112
Zhu, Zirun	Wed-E-V-5-3	171
Zi, Yunfei	Tue-E-V-1-4	103
Žilinec, Matúš	Wed-E-O-2-2	158
Zilka, Miri Zmolikova Vatarina	Wed-A-V-1-2	141
Zillolikova, Katerilla Zoehrer Matthias	Fri-A-SS-1-2	256
Zong, Chengqing	Fri-A-V-6-2	254
Zorilă, Cătălin	Thu-A-V-3-4	205
Zou, Wei	Tue-E-V-5-1	112
	Wed-M-V-2-4	124
Zou, Yuexian	Tue-M-V-6-5	69
,	Tue-M-V-4-6	70
	Tue-A-V-1-2	82
	Tue-E-V-3-2	108
	I hu-M-V-7-1	192 235
Zou, Yuxiang	Thu-M-V-5-5	189
Zuluaga, Maria A	Tue-A-SS-2-2	96
Zuluaga-Gomez, Juan	Thu-M-SS-2-4	195
Zwerts, Joeri A	Tue-M-SS-1-1	77
	- тue-м-55-1-8	10

INTERSPEECH 2021 (279) BRNO | CZECHIA

NOTES

NOTES

NOTES

Learn more about Amazon's unique approach to customer-obsessed science, and about our global science career opportunities at amazon.science.







