### TABLE OF CONTENTS

- Sponsors List  4
- Exhibitors and Institutional Supporters  5
- Welcome Message from the ISCA President  6
- Welcome Message from the Interspeech 2022 General Chairs  8
- Welcome Message from the Technical Program Chairs  9
- Welcome to Incheon, Korea  11
- INTERSPEECH 2022 General Information  12
- ISCA Ethics for Authors and Attendees  13
- Programs  17
- Social Programs  22
- Organizers  23
- INTERSPEECH 2022 Organizing Committee  24
- ISCA Best Student Paper Award  25
- Travel Grants  26
- Keynote Speakers  27
- Survey Talk Speakers  30
- Tutorial Sessions  32
- Special Sessions and Challenges  40
- ISCA Student Advisory Committee  49
- Area Chairs  51
- Satellite Events  53
- Scientific Review Committee  54
- Future INTERSPEECH Conferences  64
### SPONSORS

#### FOUNDING SPONSOR

- **LOVO**

#### PLATINUM SPONSOR

- Amazon
- science
- Apple
- Hyundai
- kakao enterprise
- LG AI Research
- Samsung Research
- supertone

#### DIAMOND SPONSOR

- DATATANG
- Google
- kt
- NAVER
- typecast

#### GOLD SPONSOR

- Meta
- speech ocean
- DATUMO

#### SILVER SPONSOR

- appen
- HITACHI
- Inspire the Next
- MAGIC DATA

#### BRONZE SPONSOR

- ASAPP
- NC
Dear INTERSPEECH 2022 attendees, Dear friends of ISCA,

INTERSPEECH 2022 has opened its doors in Incheon, Korea, marking the 23rd event in the INTERSPEECH series of conferences dedicated to spoken language processing and technology. After visiting beautiful Jeju Island in 2004, this is the second time that our flagship conference is kindly hosted in Korea, but this time on the Korean mainland, close to its capital Seoul, with plenty of historic as well as modern charm promising not only an excellent scientific event, but also adding to our personal experience.

Despite this long tradition, a persistent virus has told us to re-think our established ways of scientific exchange involving communications. After the first fully virtual INTERSPEECH conference organized by our Chinese colleagues in 2020, and a first hybrid conference in Brno in 2021, INTERSPEECH 2022 marks a short way towards new hybrid conference formats. It comes with a very steep learning curve for organizers and participants alike: Whereas Brno was organized in a fashion to make the conference online-accessible from all over the world, catering for vastly different time zones, INTERSPEECH 2022 puts the focus again on in-person participation, and assumes a more-or-less equal share between in-person and virtual participants. This transition also signals ISCA's commitment to supporting outreach to those not able to attend in-person conferences (e.g., health, visa constraints, under-resourced regions, etc.) to expand remote participation from all worldwide. You will see that most of the event, including session timing, is more similar to “traditional” conference scheduling. However, the hybrid part of the conference, mainly reflected by live poster sessions every evening (time-wise accessible from Europe as well as from parts of the Americas) as well as by online streaming and recording, makes this a very inclusive event worldwide for those who prefer or cannot travel. The hybrid format is here to stay: the requirement of climate-preserving behavior, as well as our wish to be inclusive to participants who prefer or cannot travel, will make such formats indispensable. INTERSPEECH 2022 will mark an important milestone on this way, experimenting with new formats and schedules, introducing new collaborative tools, which sometimes will challenge not only our well-established routines, but sometimes also our nightly sleep habits – all in the name of supporting speech communication research and engagement with our colleagues worldwide.

Because of these specific requirements, organizing INTERSPEECH 2022 is a challenge: The number of submitted papers skyrocketed (more than 2100 submissions), making reviewers and meta-reviewers a scarce and overloaded resource. Plagiarism detection tools were also applied, installing the ethics control loop as an essential part of the review process to ensure all honor and respect the ethical commitment ISCA holds for our community.

For welcoming us in this setting, I want to express my deep gratitude to Hanseok Ko (General Chair) and John D.L. Hansen (General Co-Chair), as well as to the TPC Co-Chairs Kyogu Lee, Lori Lamel, Mark Hasegawa-Johnson, Karen Livescu, and Okim Kang. With the help of their entire team, they managed to set up a very strong conference program, “sponsored” by lots of dedication, and (luckily) by many commercial sponsors (for which ISCA and our community are sincerely grateful). A part of this program consists of world-class Keynote Speakers, expert overview and industry talks, a wide range of special sessions, tutorials and Show&Tell demos.

A particularly noteworthy part of the program will be the presentation by this year's recipient of the ISCA 2022 Medal for Scientific Achievement: Lin-shan Lee who has been awarded the Medal for his sustained innovation in all aspects of Chinese and English spoken language technologies. As many of you will know, Lin-shan Lee has pioneered a variety of spoken language technologies, including a voice-driven Chinese typewriter, and he has set bases for retrieving, understanding and organizing speech information. Since long, ISCA has recognized the importance of language diversity in speech research, and has been making efforts to connect linguistics, science and technologies. Lin-shan Lee has made significant, original contributions in all of these dimensions; thus, ISCA is very proud to award the 2022 Medal for Scientific achievements to him during the opening ceremony.
In addition, we will have six new ISCA Fellows from our scientific community:

- David House “for outstanding multidisciplinary work at the crossroads of speech science and technology”;
- Hideki Kawahara “for contributions to speech signal processing and speech science by introducing flexible speech analysis, modification and synthesis systems”;
- Hema Murthy “for contributions to signal processing for speech technology applications and for leadership in speech processing for Indian languages”;
- Tara Sainath “for contributions to deep learning for automatic speech recognition”;
- Alex Waibel “for pioneering contributions in multilingual and multimodal spoken language processing and translation”;
- Heiga Zen “for pioneering contributions to model-based speech synthesis”.

ISCA is very proud to have such world-class researchers amongst its members!

This technical part of the program will be blended with social and networking options for all. I am particularly grateful for our ISCA-SAC and ISCA-PECRAC for taking care of the needs of students and young researchers in this hybrid setting, as well as our Diversity Committee for their time and effort to make INTERSPEECH an even more inclusive event. For those travelling to Korea, the conference setting will allow to explore Korean culture and food, and the meeting venue will be an excellent starting point for discovering the Korean capital and nearby sights.

In my role as ISCA President, I would like to highlight some changes which our organization is experiencing, and which will have some notable effect on you – our scientific members! Most visibly, the ISCA Board has decided to hire a new conference coordinator who will support the organization of the technical program – from paper submission to session organization for future Interspeech conferences. We are very glad to welcome Antoine Serrurier who has agreed to take up this new duty. Second, the ISCA Board has slightly increased in size with Ioana Vasilescu who joined the ISCA Board and kindly accepted to oversee ISCA’s expanding workshop activities. Not yet visible, but to come is an update of ISCA’s outdated software tools: A task force has agreed on new tools to optimally support you with services around our common aim, and we are confident that you will be able to experience these tools well before the next INTERSPEECH. More updates on all of our programs, and a possibility to contribute with your own ideas and suggestions, will be given during the ISCA General Assembly on Tuesday, September 20; please join us and support your scientific association so that we can best serve all of you!

On behalf of the ISCA Membership and the speech communication community as a whole, I would like to express my sincere gratitude to the entire Organizing Committee of INTERSPEECH 2022. Together with the ISCA Board, I know that this team is at the core of having an excellent INTERSPEECH conference, which I wish to all of you, be it virtually or on site in Incheon.

Welcome to INTERSPEECH 2022 in Incheon, Korea!

Sebastian Möller
ISCA President
Dear Colleagues and attendees of INTERSPEECH 2022,

Welcome to the 23rd INTERSPEECH Conference taking place here in Incheon, Korea, under the theme Human and Humanizing Speech Technology. INTERSPEECH is the world’s largest and most comprehensive conference on the science and technology of spoken language processing. INTERSPEECH conferences emphasize interdisciplinary approaches, addressing all aspects of speech science and technology, ranging from basic theory to advanced applications.

The conference theme “Human and Humanizing Speech Technology” formulates the vision of the academic, research, and industrial community to commit endeavors to continue the effort in speech science toward humanizing spoken language technology. Recent advancements in AI have made intelligent machines more available and readily accessible to assist us in our daily lives. However, these advancements have also highlighted the need for improved natural interaction/engagement with such technology assist or human interface portals. With the rapid progress in AI on speech and language applications over the 5G network services provided worldwide, we are at the onset of building our vision of creating a full ecosystem of natural speech and language technology applications. We hope the impact becomes a transformative paradigm shift that goes beyond the current state-of-the-art we serve, and ultimately society as a whole experience the benefits.

Truly a city of the future, Songdo sits adjacent to Seoul, regarded as one of the technology capitals of the world. The city’s underground rail/subway already offers high-speed WiFi, with electronic panels at exits, providing waiting times for connecting to buses or trains, while companies such as Samsung Electronics are already working on linking household devices to mobile phones. On the technological front, Songdo is a new city that offers the chance to truly integrate innovation into daily life.

interspeech 2022 this week will serve as a showcase, virtually and physically, for demonstrating new and experimental speech technologies, theories, and scientific advancements. As organizers, we hope you will enjoy the Technical Program that features 4 keynote speakers (Lin-shan Lee, Yejin Choi, Rupal Patel, Daniel Lee), 6 Survey Talks, and 12 Industry Session speakers, as well as over 1100 presentations of highly impactful scientific/technical papers. In addition to the technical program, social events will feature Korean culture and social platforms for networking with all attendees to foster the advancement of speech science and technology.

The conference venue, CONVENSIA, sits in the ultra-modern city of Songdo and has several options for transport including an underground metro line, accommodations, food, shopping, and entertainment all in close vicinity to the venue. There are many historic places to visit in the greater Incheon and Seoul City, all reachable with a convenient underground metro line. We encourage all attendees to take some time and explore the people, culture, and history of Korea!

We are thrilled to see each other in person to share new research findings and enjoy plenty of time for networking with delegates from academia, research institutes, governments, and industries. The conference program will be full of rewarding and exciting experiences. We hope you will have a great time here at INTERSPEECH-2022.

Sincerely,

Hanseok Ko                                  John Hansen
General Chair                                            General Co-Chair
WELCOME MESSAGE FROM THE TECHNICAL PROGRAM CHAIRS

We are thrilled to have the privilege to serve as technical program chairs and put together the technical program for INTERSPEECH 2022 in Incheon, Korea. This is the 2nd INTERSPEECH coming to Korea after INTERSPEECH 2004 in Jeju Island. During the past 18 years, the speech community centered around ISCA has witnessed significant growth in many aspects, including not only the size and the technical advances, but also the outreach and diversity of the community. When we were given an opportunity to work as technical program chairs, we knew it would be very challenging not only because INTERSPEECH is the largest academic event on speech science and technology, but also due to the complicated conference format caused by the uncertain nature of the Covid-19 pandemic.

Nevertheless, the speech community keeps growing and we were very grateful to have record breaking numbers in INTERSPEECH 2022 again, following the previous years. We received 2,490 submissions at the initial submission stage and 2,140 papers were sent out for review after initial checks (9.4% and 7.5% increase compared with last year, respectively). Each paper was carefully evaluated by the technical committee, receiving at least three independent reviews. In total, 1,637 reviewers provided 6,540 reviews and there was a discussion period among the reviewers, following last year’s practice. In case of split reviews, the area chairs provided additional meta reviews or recruited another reviewer to make a fair and justified decision. Finally, the TPC meeting was held at the conference venue, Songdo Convensia in Incheon during May 30-31 in a hybrid format, where General Chairs, TPC chairs, area chairs and ISCA Technical Committee decided to accept 1,121 papers for presentation at INTERSPEECH 2022 Incheon.

After the accept/reject decisions were made, all accepted papers underwent a plagiarism check. In addition, several papers were flagged by area chairs or reviewers due to possible plagiarism during the review process. More than 100 papers were suspected to have ethical issues and were forwarded to the ISCA Ethics Committee. The ISCA Ethics Committee put a significant amount of effort and time to carefully assess these papers and decided that 19 papers rose to Level-2 violation. These 19 papers were thus rejected, resulting in 1,102 papers that will be presented at INTERSPEECH 2022 Incheon. Therefore, the overall acceptance rate for INTERSPEECH 2022 is 51.5%.

The papers accepted for presentation then went through the process of session making, where all area chairs and TPC chairs put significant amounts of time and effort to assign all accepted papers into individual sessions with coherent themes throughout the conference dates. Putting these sessions together was extremely challenging again this year due to constant adjustments in policies/regulations regarding the Covid-19 pandemic. We aimed to have as many on-site attendees and presentations as possible using the hybrid format of both on-site and virtual programs. Finally, we ended up having a total of 124 technical sessions where 77 sessions are presented on-site and 47 sessions virtually. Out of 77 on-site sessions, 51 sessions are oral presentations, and 26 sessions are poster presentations while 47 poster sessions are presented virtually.

In the process of paper review and session making, the reviewers and area chairs were asked to recommend the student papers for the best student paper award. The recommended papers were forwarded to the Best Student Paper award committee led by Joakim Gustafsson and 12 papers were finally nominated as Best Student Paper award candidates. The nominated students will be asked to present their work in a technical session. After reading the paper and the reviews, and watching the talk, the committee will decide the Best Student Paper awards during the conference.

One of the integral parts in every INTERSPEECH is keynote presentations. The INTERSPEECH 2022 plenary session chairs Julia Hirschberg, Richard Stern, and David Han invited 4 amazing keynote speakers – Lin-shan Lee, Yejin Choi, Rupal Patel and Daniel Lee. Their talks will cover a wide range of topics related to speech science and technology. Following the initiative of INTERSPEECH 2018 in Hyderabad, we will also have 6 survey talks (overseen by John Hansen) where a diverse group of experts from both academia and industry will share their insights accumulated from experience in the speech community.

In addition, the program contains 8 tutorials covering highly interesting and diverse topics (overseen by Hyung-Min Park and Mounya Elhilali), 12 Special sessions of a wide range of themes (overseen by Jon Barker, Odette Scharenborg, Jingdong
Chen, and Joon Son Chung), and 7 Show and Tell sessions (overseen by Dongsuk Yook and Ji-Hwan Kim) which will feature 21 demonstrations throughout the conference period. Finally, we also have 6 ISCA-supported Satellite Events and Workshops (overseen by Jong Won Shin, Yunjung Kim and Wenwu Wang) that will be held on-site or virtually surrounding the main INTERSPEECH conference, enriching the technical program.

Organizing the technical program for INTERSPEECH is a collaborative effort by a number of devoted colleagues without whom it may not be possible. We are therefore very grateful for those who helped and contributed in any way to make INTERSPEECH 2022 a great success, including (in random order): area chairs, reviewers, ISCA board, ISCA Ethics Committee, chairs for special sessions and challenges, plenary talks, survey talks, tutorials, satellite events, and show & tell, the organizers of special sessions and satellite events, authors, and the local organizers from Hyundai Asan.

The main theme of INTERSPEECH 2022 is “Human and Humanizing Speech Technology”, which is well reflected in a wide range of topics and areas in the technical program. We hope you will truly enjoy and be inspired by many social programs and industrial events as well as the diverse technical programs, and take home with you the spirit of Incheon and Korea!
WELCOME TO INCHEON, KOREA

The host city, Incheon, is the fastest gateway to enter Korea. As both sky and sea routes are widely open, the city has positioned itself as a professional hub. Incheon International Airport, as the 7th largest airport in passenger transfer, connects 186 cities in the world. Incheon is a beautiful port city in Korea that best exemplifies the attractive balance between a long, rich heritage of dynasties and a recently developed modern culture. Incheon is rapidly developing into a city of modern culture. It already completed the construction of various industrial and logistics complexes and urbanized the Songdo Business District.

Eco-Friendly Environment

There are a total of 168 islands in Incheon, ranging from the largest, Ganghwa Island (293 km²) to the smallest, Modo Island (810 m²). The islands are popular with tourists who enjoy cycling, camping, and backpacking along with white sandy beaches surrounded by forests. One of the islands, Daejikdo, is the only island designated as a marine ecosystem conservation area in Incheon. It is a mysterious sandy island that is submerged during high tide and exposed during low tide. It is called ‘whale shoal’ since the sand island rising above the wide sea and disappearing like a submarine resembles a whale.

Historical and Cultural Heritages

Incheon is located at the mouth of the Han River, which links Seoul with the West Sea. Because of its location, it was a site of great military and geographic importance. So, Ganghwa used to be located at the center of Korean history, and it has still many historical and cultural heritages, thereby being called as the roofless museum. It can be said that the history of Ganghwa is just history of Korean people, starting from the dolmen of prehistoric times designated as the UNESCO World Heritage. It feels like time travel where past, present and future coexist.

Urban Parks and Cityscape

The cityscape is modern and beautiful. The eye-catching architecture creates a sophisticated urban atmosphere, and the parks give the city a fresh color. It is good for a walk along the trail surrounding the waterway, and at night you can enjoy the wonderful night view of Songdo International City surrounded by skyscrapers. In the water park, which is the first in Korea created by pumping water from the sea, water taxis are operated and various water leisure activities such as boats, kayaks, and canoes can be experienced. Songdo international city park shows an “exotic” landscape due to the harmony of traditional Korean-style houses and skyscrapers.
INTERSPEECH 2022 GENERAL INFORMATION

VENUE
Songdo Convensia, 123 Central-ro, Yeonsu-gu, Incheon 21998 Korea

MOBILE APP
The INTERSPEECH 2022 mobile app is a native application for tablet and smartphone devices (iPhone and Android). The mobile app provides easy-to-use interactive capabilities to enhance your experience as an attendee. Features include:
• Agenda: View schedules, explore sessions and find social events. Create your own personal schedule for easy tracking.
• Update: Receive latest news about the conference program.
• Activity Feed: Push notifications, chatting on sessions and presentations.
• Sponsors/Exhibitors: A complete list of sponsors/exhibitors.
To download the application, visit your app store and search for “INTERSPEECH 2022”.

REGISTRATION
Location: Grand Ballroom Foyer, 2F, Songdo Convensia
• Sunday, September 18 ....................... 08:00–20:30
• Monday, September 19 ...................... 08:00–17:00
• Tuesday, September 20 ...................... 08:00–17:00
• Wednesday, September 21 .................. 08:00–17:00
• Thursday, September 22 .................... 08:00–17:00

BADGES & MATERIALS
As a registered attendee, you will be issued an INTERSPEECH name badge when you pick up your registration materials on-site. You will be required to display your name badge for admission to all official functions. In the event of a lost badge, you may purchase a replacement badge for a EUR 20 fee.
• Materials: The conference proceedings (USB stick)
  A printed abstract book (only if you have requested upon your online registration)
  A conference bag
  A mobile Android or iOS app
  Access to all official conference sessions
  Coffee breaks throughout the conference
  One admission to the Welcome Reception on Monday, September 19
  One admission to the Gala Banquet on Wednesday, September 21

GENERAL INFORMATION
• Time Zone: The Republic of Korea observes Korea Standard Time (UTC+9, KST) all year.
• Local Currency: The Republic of Korea has the Korean won (KRW) as its sole currency. One euro represents approximately 1,370 KRW.
• Electricity: In the Republic of Korea, the electrical voltage used is 220 volts at 60 Hertz, and the outlet has two round holes.
• Wifi: Free WiFi will be provided at the conference venue.

EMERGENCY
Please notify any of the INTERSPEECH 2022 staff members if you need medical assistance. The general emergency call number in the Republic of Korea is 119.

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 2022 organizers will not accept any responsibility for any delegate failing to take out adequate insurance.

DISCLAIMER
The organizers are not liable for any loss or damage incurred by the conference delegates or by any other individuals accompanying them, both during the official activities as well as going to/from the conference. The organizers also cannot accept liability for injuries arising from accidents or other situations during or as a consequence of the conference attendance. Delegates are responsible for their own safety and belongings.
ISCA ETHICS FOR AUTHORS AND ATTENDEES

ISCA 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. Authors should meet the following standards:

1. The work does NOT include fabrication, falsification, or any kind of data breach. Authors should retain their code and maintain a log of the data that produced the results in their paper. Authors are also encouraged to make their code and dataset freely available.

2. The work does NOT include plagiarism or significant self-plagiarism. The work must be original, and any paper which significantly overlaps with previous work is not allowed. Proper reference to previous work is also required. Verbatim copying of work that has been distributed but not refereed, such as technical reports and arXiv articles, is permitted only if the authors are the same. ISCA (and conference organizers or journal editors) may use tools to detect (self-)plagiarism and reject papers without review. The work MAY NOT BE submitted to any other conference, workshop or journal during the review process.

3. 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.

4. 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 involved in the work.

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

Ethical Standard checking is not limited to these 5 points.

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, which may include rejection/removal of the paper (and other papers in the same conference/workshop by the same authors) 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).

ISCA 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 against 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.
<table>
<thead>
<tr>
<th>Time (KST)</th>
<th>Sep. 18 (Sun)</th>
<th>Sep. 19 (Mon)</th>
<th>Sep. 20 (Tue)</th>
<th>Sep. 21 (Wed)</th>
<th>Sep. 22 (Thu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 – 08:30</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>08:30 – 09:00</td>
<td>Opening Session</td>
<td>Keynote 2</td>
<td>Keynote 3</td>
<td>Keynote 4</td>
<td></td>
</tr>
<tr>
<td>09:00 – 09:30</td>
<td>Morning Tutorial I</td>
<td>Morning Tutorial II</td>
<td>Morning Tutorial III</td>
<td>Morning Tutorial IV</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>09:30 – 10:00</td>
<td>Keynote 1</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>10:00 – 10:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>10:30 – 11:00</td>
<td>Lunch Break</td>
<td>Lunch Break</td>
<td>Lunch Break</td>
<td>Lunch Break</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>11:00 – 11:30</td>
<td>Lunch Break</td>
<td>Lunch Break</td>
<td>Lunch Break</td>
<td>Lunch Break</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>11:30 – 12:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>12:00 – 12:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>12:30 – 13:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>13:00 – 13:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>13:30 – 14:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>14:00 – 14:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>14:30 – 15:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>15:00 – 15:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Closing Session</td>
</tr>
<tr>
<td>16:00 – 16:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td></td>
</tr>
<tr>
<td>16:30 – 17:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td></td>
</tr>
<tr>
<td>17:00 – 17:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td></td>
</tr>
<tr>
<td>17:30 – 18:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td></td>
</tr>
<tr>
<td>18:00 – 18:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td></td>
</tr>
<tr>
<td>18:30 – 20:30</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Gala Banquet</td>
</tr>
<tr>
<td>20:30 – 22:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Gala Banquet</td>
</tr>
<tr>
<td>21:00 – 23:00</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Oral Sessions</td>
<td>Poster, Show&amp;Tell</td>
<td>Gala Banquet</td>
</tr>
<tr>
<td>17:00 – 17:30</td>
<td>ISCA General Assembly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:30 – 18:00</td>
<td>Registration Desk Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18:00 – 18:30</td>
<td>Welcome Reception</td>
<td>Student Reception</td>
<td>Reviewers Reception</td>
<td>Gala Banquet</td>
<td></td>
</tr>
<tr>
<td>18:30 – 20:30</td>
<td>Welcome Reception</td>
<td>Student Reception</td>
<td>Reviewers Reception</td>
<td>Gala Banquet</td>
<td></td>
</tr>
<tr>
<td>20:30 – 22:00</td>
<td>Welcome Reception</td>
<td>Student Reception</td>
<td>Reviewers Reception</td>
<td>Gala Banquet</td>
<td></td>
</tr>
</tbody>
</table>

**SESSION CODE DESCRIPTION**

(Ex. Tue-O-VR-1-1)

<table>
<thead>
<tr>
<th>Tue</th>
<th>O / P / SS / ST</th>
<th>OS / VR</th>
<th>#</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>O: Oral / P: Poster / SS: Special Session / ST: Show &amp; Tell</td>
<td>OS: On-Site / VR: Virtual</td>
<td>Session #</td>
<td>Sub-Session ##</td>
</tr>
</tbody>
</table>

- **Hybrid:** Green indicates programs that will be on-site and live-streamed on the virtual platform.
- **On-Site:** Yellow indicates programs that will take place on-site only.
- **Virtual:** Blue indicates programs that will take place on the virtual platform only.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:30 - 09:00</td>
<td>Keynote 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:30 - 10:00</td>
<td>Coffee Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00 – 12:30</td>
<td>Lunch Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20:00 – 23:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

September 22, Thursday

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:30 - 09:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:30 - 10:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00 – 12:30</td>
<td>Lunch Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:00 – 17:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21:00 – 23:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Live Q&A for Virtual Poster Session

**TECHNICAL AREAS**

<table>
<thead>
<tr>
<th>Area Number</th>
<th>Area Name and Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Speech Perception, Production and Interaction by Human Listeners</td>
</tr>
<tr>
<td>2</td>
<td>Phonetics, Phonology and Prosody</td>
</tr>
<tr>
<td>3</td>
<td>Paralinguistics in Speech and Language: Human and Automatic Analysis and Processing</td>
</tr>
<tr>
<td>4</td>
<td>Speaker and Language Identification</td>
</tr>
<tr>
<td>5</td>
<td>Analysis of Speech and Audio Signals</td>
</tr>
<tr>
<td>6</td>
<td>Speech Coding and Enhancement</td>
</tr>
<tr>
<td>7</td>
<td>Speech Synthesis and Spoken Language Generation</td>
</tr>
<tr>
<td>8</td>
<td>Speech Recognition Signal Processing, Acoustic Modeling, Robustness, Adaptation</td>
</tr>
<tr>
<td>9</td>
<td>Speech Recognition: Architecture, Search, and Linguistic Components</td>
</tr>
<tr>
<td>10</td>
<td>Speech Recognition: Technologies and Systems for New Applications</td>
</tr>
<tr>
<td>11</td>
<td>Spoken Dialog Systems and Conversational Analysis</td>
</tr>
<tr>
<td>12</td>
<td>Spoken Language Processing: Translation, Information Retrieval, Summarization, and Evaluation</td>
</tr>
<tr>
<td>13</td>
<td>Speech, Voice, and Hearing Disorders</td>
</tr>
<tr>
<td>14</td>
<td>Special Sessions and Challenges</td>
</tr>
<tr>
<td>15</td>
<td>Show and Tell</td>
</tr>
</tbody>
</table>

**INTERSPeECH 2022**

**HUMAN AND HUMANIZING SPEECH TECHNOLOGY**

**INTERSPeECH 2022**

**HUMAN AND HUMANIZING SPEECH TECHNOLOGY**
Welcome Reception (For All!)

- Monday, 19 September, 18:30 to 21:00
- Grand Ballroom, Songdo Convensia, Conference Venue

We welcome everyone to join the welcome reception at the Grand Ballroom of the conference venue right after the ISCA General Assembly has finished on Monday at 18:30. Please come and enjoy some finger food and drinks to begin the conference with lively networking and performances.

Student Reception (Students Only!)

- Tuesday, 20 September, 18:30 to 21:00
- Multi Square, Songdo Convensia, Conference Venue (Exit 7)

All students are invited to the famous ‘Chi-Mek Party’: the fantastic pairing of Korean fried chicken and Mekju (beer in Korean). With an EDM DJ, everyone is welcome to dance the night away! Don’t worry if you have dietary restrictions – we have you covered with vegetarian and halal options.

Reviewers & Fellows Reception (For Reviewers and Fellows Only!)

- Tuesday, 20 September, 18:30 to 21:00
- Arirang Hall (B1F), Gyeongwon-Jae

Experience the true Korean way of living at Gyeongwonjae with traditional Korean architecture. We would like to take this opportunity to convey our deepest gratitude to our large body of reviewers for delivering meticulous and timely reviews which are crucial for building an excellent scientific program. With the record high number of submitted papers, INTERSPEECH 2022 wouldn’t have been possible without your contribution. Enjoy the beautiful evening of Songdo!

Gala Banquet (For All!)

- Wednesday, 21 September, 18:30 to 21:00
- Grand Ballroom, Songdo Convensia, Conference Venue

Let’s enjoy the last day of the conference! Although we would have loved to take you to beautiful spots of Korea to celebrate the successful conclusion of the conference, due to large number of expected conference participants we will be transforming the Grand Ballroom of Songdo Convensia for the banquet. Enjoy the ‘Fun Zone’ with the photo wall where you can try on the traditional Korean costumes to take instant photos, and learn to play Korean traditional games. Spectacular performances will be ready for the Gala Banquet, so please come and enjoy!
International Speech Communication Association (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 Rene CARRE 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 it has put its financial and administrative functions on a firm professional footing. 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. ISCA Objectives are to stimulate scientific research and education, to organize conferences, courses and workshops, to publish, and to promote publication of scientific works.

www.isca-speech.org

The Acoustical Society of Korea (ASK) is a non-profit professional organization founded in 1981. The purpose of the society is to promote the academic and technical development of acoustics by fostering exchanges between academic professors and industrial experts in the acoustical field. ASK has set up and administers 12 academic committees, covering all aspects of acoustic research. ASK has over 3,000 regular members of whom 300 are actively engaged in events organized by ASK. The society members do aggressively lots of academic activities relating to acoustical fields such as Speech and Hearing, Signal Processing, Acoustic Communication, Electroacoustics, Psychology and Physiology, Musical Acoustics, Broadcasting Acoustics and New media, Underwater Acoustics, Physical Acoustics, Photoacoustics, Ultrasonics, Architectural Acoustics, Structural Acoustics and Vibration, Linguistics and Voice Synthesis, etc. The ASK has published annually six journal issues titled ‘The Journal of the Acoustical Society of Korea’ since 1982.

www.ask.or.kr
INTERSPEECH 2022 ORGANIZING COMMITTEE

General Chair
Hanseok Ko (Korea University, Korea)

General Co-Chair
John H. L. Hansen (University of Texas at Dallas, USA)

Technical Program
Kyogu Lee (Seoul National University, Korea)
Lori Lamel (CNRS-LIMSI, France)
Mark Hasegawa-Johnson (University of Illinois, USA)
Okim Kang (Northern Arizona University, USA)

Special Sessions/Challenges
Jon Barker (University of Sheffield, UK)
Odette Scharenborg (Delft University of Technology, Netherlands)
Jingdong Chen (Northwestern Polytechnical University, China)
Joon Son Chung (KAIST, Korea)

Plenary Sessions
Julia Hirschberg (Columbia University, USA)
Richard Stern (Carnegie Mellon University, USA)
David K. Han (Drexel University, USA)

Tutorial Sessions
Hyung-Min Park (Sogang University, Korea)
Mounya Elhilali (Johns Hopkins University, USA)

Satellite Workshop
Jong Won Shin (GIST, Korea)
Yunjung Kim (Florida State University, USA)
Wenwu Wang (University of Surrey, UK)

Show and Tell
Dongsuk Yook (Korea University, Korea)
Ji-Hwan Kim (Sogang University, Korea)

Industry Liaison
Chanwoo Kim (Samsung Research, Korea)
Sunhee Kim (Seoul National University, Korea)
Bhuvana Ramabhadran (Google, USA)
Akihiko K. Sugiyama (Yahoo Japan Corporation, Japan)
Anil Alexander (Oxford Wave Research, UK)
Jing Huang (Amazon Alexa AI, USA)
Zhengchen Zhang (JD Technology, China)

Exhibits
Jeff Holliday (Korea University, Korea)
Eun Jong Kong (Korea Aerospace University, Korea)

Awards (Grants)
Young-cheol Park (Yonsei University, Korea)
Jiyoung Shin (Korea University, Korea)

Publications
Ha-Jin Yu (University of Seoul, Korea)
Sunmee Kang (Seokyeong University, Korea)
Suyoun Kim (Meta, USA)

Publicity
Joon-Hyuk Chang (Hanyang University, Korea)
Joo-Kyeong Lee (University of Seoul, Korea)
Jaeha Kim (Hanyang University, Korea)

Finance
Hongsub Choi (Daejin University, Korea)

Volunteers
Tae-Jin Yoon (Sungshin Women’s University, Korea)

Local Arrangements
Wooil Kim (Incheon National University, Korea)
Eon-Suk Ko (Chosun University, Korea)

Local Operations
Bowon Lee (Inha University, Korea)

Operation Secretariat
Hyojeong Bang (Korea University, Korea)
The following 12 papers are shortlisted for the ISCA Best Student Paper Award 2022. Note that there may be changes to some of the paper sessions in the final program.

Tsiky Rakotomalala, Pierre Baraduc and Pascal Perrier: Trajectories predicted by optimal speech motor control using LSTM networks
Mon-O-OS-2-2 Monday, September 19, 14:30 – 16:30
IN-PERSON: Speech Production

Philipp Buech, Rachid Ridouane and Anne Hermes: Pharyngealization in Amazigh: Acoustic and articulatory marking over time
Wed-O-OS-7-3 Wednesday, September 21, 13:30 – 15:30
IN-PERSON: Phonetics and Phonology

Ambika Kirkland, Harm Lameris, Éva Székely and Joakim Gustafson: Where’s the uh, hesitation? The interplay between filled pause location, speech rate and fundamental frequency in perception of confidence
Thu-O-OS-10-3 Thursday, September 22, 13:30 – 15:30
IN-PERSON: Emotional Speech Production and Perception

Bei Liu, Zhengyang Chen and Yanmin Qian: Attentive Feature Fusion for Robust Speaker Verification
Mon-P-VR-1-4 Monday September 19, 11:00 – 13:00
VIRTUAL: Embedding and Network Architecture for Speaker Recognition

Katharine Patterson, Kevin Wilson, Scott Wisdom and John R. Hershey: Distance-Based Sound Separation
Mon-P-VR-2-4 Monday, September 19, 12:30 – 16:30
VIRTUAL: Spatial Audio

Vinay Kothapally and John H.L. Hansen: Complex-Valued Time-Frequency Self-Attention for Speech Dereverberation
Tue-P-VR-5-4 Tuesday, September 20, 16:00 – 18:00
VIRTUAL: Dereverberation and Echo Cancellation

Minchan Kim, Myeonghun Jeong, Byoung Jin Choi, Sungwan Ahn, Joun Yeop Lee and Nam Soo Kim: Transfer Learning Framework for Low-Resource Text-to-Speech using a Large-Scale Unlabeled Speech Corpus
Mon-P-OS-2-2 Monday, September 19, 14:30 – 16:30
IN-PERSON: Speech Synthesis: Acoustic Modeling and Neural Waveform Generation I

Guangzhi Sun, Chao Zhang and Phil Woodland: Tree-constrained Pointer Generator with Graph Neural Network Encodings for Contextual Speech Recognition
Tue-P-VR-4-4 Tuesday, September 20, 13:00 – 15:00
VIRTUAL: Neural Transducers, Streaming ASR and Novel ASR Models

Qu Yang, Qi Liu and Haizhou Li: Deep Residual Spiking Neural Network for Keyword Spotting in Low-Resource Settings
Wed-P-OS-6-3 Wednesday, September 21, 10:00 – 12:00
IN-PERSON: Resource-constrained ASR

Bowen Shi, Wei-Ning Hsu and Abdelrahman Mohamed: Robust Self-Supervised Audio-Visual Speech Recognition
Tue-P-VR-4-5 Tuesday, September 20, 13:00-15:30
VIRTUAL: Zero, Low-resource and Multi-Modal Speech Recognition II

Sarenne Carrol Wallbridge, Catherine Lai and Peter Bell: Investigating perception of spoken dialogue acceptability through surprisal
Thu-O-OS-9-4 Thursday, September 22, 10:00-12:00
IN-PERSON: Spoken Dialogue Systems

Hyeon-Kyeong Shin, Hyewon Han, Doyeon Kim, Soo-Whan Chung and Hong-Goo Kang: Learning Audio-Text Agreement for Open-vocabulary Keyword Spotting
Tue-O-OS-4-4 Tuesday, September 20, 13:00-15:30
IN-PERSON: Spoken Term Detection and Voice Search
TRAVEL GRANTS

A total of 60 ISCA Travel Grants for virtual or in-person participation in Interspeech 2022 have been awarded. Congratulations to all recipients!

Kechun Li
A Arunkumar
Sondes Abderrazek
Binu Abeysinghe
Andrea Alicehajic
Franklin Alvarez Cardinale
Ananya Ayasi
Abdul Hameed Azeemi
Alan Baade
Matthew Baas
Shahaf Bassan
Debarpan Bhattacharya
Han Bing
Andrei Birladeanu
Catarina Botelho
Siddarth Chandrasekar
Gerasimos Chatzoudis
Jaejin Cho
Trung Duc Anh Dang
Helen Gent
Chou Huang-Cheng
Li Jingyu
Georgios Karakasisidis
Piotr Kawa
Joseph Konan
Kunnar Kukk
Boram Lee
Yuna Lee
Yi Lei
Guan-Ting Lin
Bei Liu
Luong Manh
Debasish Ray Mohapatra
Jonathan Mukiibi
Takayuki Nagamine
Bernard Opoku
Zexu Pan
Darshana Priyasad
Fareeha Rana
Anwesha Roy
Matsui Sanae
Nicolas Schmidt
Yael Segal
Naoaki Suzuki
Tuende Szalay
Dehua Tao
Francisco Teixeira
Nguyen Luong Tran
Ioannis Tsiamas
Werner Van Der Merwe
Lester Phillip Violeta
Sarenne Wallbridge
Haibin Wu
Peter Wu
Ting-Wei Wu
Liumeng Xue
He Yuhang
Jing Zhao
Running Zhao
Da-Wei Zhou
Lin-shan Lee: From Semantics to Self-Supervised Learning for Speech and Beyond

Monday, September 19, 09:30 KST - Grand Ballroom

Lin-shan Lee has been teaching in Electrical Engineering and Computer Science at National Taiwan University since 1979.

He invented, published and demonstrated the earliest but very complete set of fundamental technologies and systems for analyzing Chinese sentences (1986-91) and synthesizing (1984-89) and recognizing (1987-97) Chinese speech with unlimited text, considering the structural features of Chinese language (monosyllable per character, limited number of distinct monosyllables, tones, sentence structures, etc.) and the extremely limited resources.

He then focused his work on retrieval of speech information from the Internet, proposing a whole set of approaches offering better retrieval accuracy and efficiency. This part of work applies equally to all different languages, and was described as the stepping stones towards "a spoken version of Google" when Nature selected him in 2018 as one of the 10 "Science Stars of East Asia" in a special issue on scientific research in East Asia.

He has also pioneered a wide variety of other new frontiers in spoken language technologies, with examples including structured learning for ASR, code-switched speech recognition, unsupervised discovery of acoustic tokens, audio signal embedding and unsupervised ASR. So his work was not only across different languages with very different linguistic structures, not only from TTS to ASR to retrieval to understanding, but from acoustics, linguistics to semantics, from utterances to spoken documents, and from supervised to unsupervised.

He served as a member of the Board (2001-2009) of International Speech Communication Association (ISCA), the Distinguished Lecturer of IEEE Signal Processing Society (SPS) (2007-2008), and the general chair of International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2009 held in Taipei. He was elected IEEE Fellow in 1993 and ISCA Fellow in 2010. He also received the Meritorious Service Award from IEEE SPS in 2011. He received Presidential Science Prize of Taiwan in 2015 and was elected as an academician of Academia Sinica in 2016, both of which are the highest honor for scientists in Taiwan.

Abstract: Although spoken language understanding has been extensively investigated for many years, the gap between signals and semantics in speech remains to be a challenging problem. But the semantics in speech may offer a bridge towards the realization of "a spoken version of Google". On the other hand, with the fast advances in deep learning in recent years, self-supervised learning seems to be very attractive. In this latter approach, a model is first pre-trained with a big unlabeled dataset without considering any specific application, based on which many different downstream tasks can be achieved much more easier with simpler model structure and smaller labeled dataset.

In this talk, the speaker will present his thoughts and experiences on these two research areas, and show how the research paths for them seem to merge in some way.
Yejin Choi:
David V.S. Goliath: the Art of Leaderboarding in the Era of Extreme-Scale Neural Models

Tuesday, September 20, 08:30 KST - Grand Ballroom

Yejin Choi is Brett Helsel Professor at the Paul G. Allen School of Computer Science & Engineering at the University of Washington and a senior research manager at AI2 overseeing the project Mosaic. Her research investigates a wide range of problems including commonsense knowledge and reasoning, neuro-symbolic integration, multimodal representation learning, and AI for social good. She is a co-recipient of the ACL Test of Time award in 2021, the CVPR Longuet-Higgins Prize in 2021, a NeurIPS Outstanding Paper Award in 2021, the AAAI Outstanding Paper Award in 2020, the Borg Early Career Award in 2018, the inaugural Alexa Prize Challenge in 2017, IEEE AI's 10 to Watch in 2016, and the ICCV Marr Prize in 2013.

Abstract: Scale appears to be the winning recipe in today’s leaderboards. And yet, extreme-scale neural models are still brittle to make errors that are often nonsensical and even counterintuitive. In this talk, I will argue for the importance of knowledge, especially commonsense knowledge, as well as inference-time algorithms, and demonstrate how smaller models developed in academia can still have an edge over larger industry-scale models, if powered with knowledge or algorithms.

First, I will introduce “symbolic knowledge distillation”, a new framework to distill larger neural language models into smaller commonsense models, which leads to a machine-authored KB that wins, for the first time, over a human-authored KB in all criteria: scale, accuracy, and diversity.

Next, I will highlight how we can make better lemonade out of neural language models by shifting our focus to unsupervised, inference-time algorithms. I will demonstrate how unsupervised models powered with algorithms can match or even outperform supervised approaches on hard reasoning tasks such as nonmonotonic reasoning (such as counterfactual and abductive reasoning), or complex language generation tasks that require logical constraints.

Finally, I will introduce a new (and experimental) conceptual framework, Delphi, toward machine norms and morality, so that the machine can learn to reason that “helping a friend” is generally a good thing to do, but “helping a friend spread fake news” is not.

Rupal Patel:
Blurring the Line Between Human and Computer-Generated Speech: Opportunities and Challenges

Wednesday, September 21, 08:30 KST - Grand Ballroom

Dr. Rupal Patel is a Professor of Communication and Computer Sciences at Northeastern University and Director of the Center for Speech Science and Technology, an interdisciplinary research group that leverages AI and machine learning to identify acoustic signals in speech for diagnostic, therapeutic and prosthetic applications. She is also the founding Chief Executive Officer of VocaliD, an AI Voice company that creates synthetic voices with personality for brands that understand the power of relatable voice and individuals living with speechlessness who want to be heard in their own unique voice. Rupal is also the Founder of STEM4SocialChange, a youth leadership organization that engages youth in STEM-based initiatives that impact those with health and social challenges and AITHOS, a consortium of synthetic media organizations that promote the ethical and fair use of computer-generated media.
Abstract: Computer-generated speech is becoming nearly indistinguishable from human recorded speech, especially for short-form audio content. The neural models that generate life-like speech require a fraction of the training data compared to previous concatenative and parametric methods. Moreover, combining text-to-speech and speech conversion methodologies can further blur the line between real and artificial. This talk will shed light on the shift in scientific focus from striving for human parity to protecting identity, safeguarding usage, and sometimes even introducing artifacts to convey provenance. Through a series of use cases, we will explore the opportunities and challenges in this new dawn of humanized speech technologies.

Daniel Dongyuel Lee:
Representations and Geometry for Multimodal Learning

Thursday, September 22, 08:30 KST - Grand Ballroom

Dr. Daniel Dongyuel Lee is the Tisch University Professor in Electrical and Computer Engineering at Cornell Tech and Executive Vice President and Head of the Global AI Center for Samsung Research. He received his B.A. summa cum laude in Physics from Harvard University and his Ph.D. in Condensed Matter Physics from the Massachusetts Institute of Technology. He was also a researcher at Bell Labs in the Theoretical Physics and Biological Computation departments. He is a Fellow of the IEEE and AAAI and has received the NSF CAREER award and the Lindback award for distinguished teaching. He was also a fellow of the Hebrew University Institute of Advanced Studies in Jerusalem, an affiliate of the Korea Advanced Institute of Science and Technology, and organized the US-Japan National Academy of Engineering Frontiers of Engineering symposium and Neural Information Processing Systems (NeurIPS) conference. His group focuses on understanding general computational principles in biological systems and on applying that knowledge to build autonomous systems.

Abstract: The advent of deep neural networks has brought significant advancements in the development and deployment of speech and multimodal technologies. Compared with conventional signal processing approaches, end-to-end neural network architectures have shown significantly better performance for natural language understanding, machine translation, automatic speech recognition, and text-to-speech generation. Samsung Research has focused on incorporating these neural network models across Samsung’s billions of devices and users. But how can we understand how the representations of sensor input signals are transformed by deep neural networks? I show how insights can be gained by analyzing the high-dimensional geometrical structure of these representations as they are reformatted in neural network hierarchies.
SURVEY TALK SPEAKERS

Nancy F. Chen, A*STAR, Singapore:
Summarizing Conversations: From Meetings to Social Media Chats

Monday, September 19, 13:10–13:40

Abstract: Human-to-human conversation is a dynamic process of information exchange between multiple parties that unfolds and evolves over time. It remains one of the most natural means of how humans pass down knowledge, increase mutual understanding, convey emotions, and collaborate with one another. However, it is nontrivial to automatically distill such unstructured information into summaries useful for downstream tasks. In this talk we will examine recent approaches and applications in addition to discussing future trends.

Esther Klabbers, ReadSpeaker, USA:
Speech Synthesis in Conversational AI

Monday, September 19, 13:50–14:20

Abstract: Conversational AI is a type of artificial intelligence that enables consumers to interact with computer applications the way they would with other humans. State-of-the-art neural network speech synthesis is capable of generating expressive speech that sounds intelligible and natural, but there are several issues to consider when developing speech synthesis for conversational AI applications. In this survey talk we will give an overview of some of these considerations pertaining to training data selection, acoustic model architectures, and evaluation of the generated speech in the context of conversational AI.

Theodora Chaspari, Texas A&M University:
Emotion and Affective Speech Processing/Technology

Wednesday, September 21, 12:10–12:40

Abstract: In the era of Internet of Things (IoT) and artificial intelligence (AI), voice-enabled technologies promise a unique opportunity for bettering our lives via human-centered services. Affective speech processing is central to such technologies, since speech can be a valuable indicator of emotion, communicative intent, and social behavior. However, the implementation and broad adoption of speech-enabled affective technologies comes with various computational and societal challenges, that yield from the complex and heterogeneous data spaces, limited availability of reliably labelled data, inherently high inter-individual variability, and the risk of leaking of personally identifiable information. This talk will overview the major challenges in designing robust speech emotion recognition models for supporting human-centered affective technologies, present recent advances on privacy preservation, fairness, and explainability, and discuss thoughts on rendering these technologies trustworthy to the stakeholders and the general public.
Nicolas Cummins, King’s College London, Thymia:
Machine Learning for Speech-Based Health Analysis:
State-of-the-art and Future Challenges

Wednesday, September 21, 12:50–13:20

Abstract: Speech is a unique and rich health signal: no other signal contains its singular combination of cognitive, neuromuscular and physiological information. This presentation will first describe how our voice is a tacit communicator of our health. Then, it will present an overview of current state-of-the-art in the prediction of health state using speech. It will end by highlighting future challenges in relation to the translation of speech analysis into clinic practise.

Leibny Paola Garcia Perera, Johns Hopkins University, Center for Language and Speech Processing, USA:
Current Trends in Diarization

Thursday, September 22, 12:10–12:40

Abstract: The primary goal of speaker diarization is to answer the question “who spoke when” in a recording, identifying regions containing speech and assigning speaker identity labels to each utterance. In this talk, we will explore the state-of-the-art approaches to solve speaker diarization. We will highlight the evolution from cascaded systems to end-to-end diarization and the current improvements. Finally, we will point out which directions, such as multimodal diarization, are evolving rapidly.

Tom Bäckström, Aalto University, Department of Signal Processing and Acoustics:
Privacy in Speech Technology

Thursday, September 22, 12:50–13:20

Abstract: Speech signals contain a host of information beyond the intended message, including speaker identity, state of health, affiliations and gender identity, much of which is private information. We have recently come to understand that this exposes users of speech technology to exploitation such as price gauging, discrimination, stalking and extortion. Consequently, research in the area has picked up speed and aims to improve the quality of experience and usability by preserving privacy. This talk introduces the main concepts and progress in the main research areas.
Learning from Weak Labels

One of the key bottlenecks in training diverse accurate audio classifiers is the need for “strongly-labeled” training data, that provide precisely demarcated instances of the audio events to be recognized. Such data are, however, difficult to obtain, particularly in bulk. The alternate, more popular approach is to train models using “weakly” labelled data, comprising recordings in which only the presence or absence of sound classes is tagged, without additional details of the number of occurrences of the sounds or their locations in the recordings. Weakly labelled data are much easier to obtain than strongly labelled data; however training with such data comes with many challenges. In this tutorial we will discuss the problem of training audio (and other) classifiers from weakly labelled data, including several state-of-art formalisms, their restrictions and limitations, and areas of future research.

Bhiksha Raj is a Professor in the School of Computer Science at Carnegie Mellon University. His areas of research include speech and audio processing and acoustic scene analysis. He was one of the pioneers of the field of learning audio classifiers from weak labels. Raj has previously conducted several tutorials at ICASSP, Interspeech and various other conferences. He is a fellow of the IEEE.

Anurag Kumar is a research scientist at Meta Research. Anurag completed his PhD from Carnegie Mellon University in 2019, and has been at Meta Research since then. Kumar is the recipient of the Samsung Innovation Award and has been a finalist for the Qualcomm Fellowship for his work on audio analysis. Along with Professor Raj, he has been one of the pioneers in the field of learning audio classifiers from weak labels.

Ankit Shah is a PhD student in the Language Technologies institute in the School of Computer Science at Carnegie Mellon University. His areas of interest are audio understanding, machine learning, and deep learning. His thesis focuses on learning in the presence of weak, uncertain, and incomplete labels, where he has made several key contributions, including the setting up of DCASE challenges on the topic. He has won the Gandhian Young Technological Innovator (GYTI) award in India for his contribution to building a never-ending learner of sound systems.

Reinforcement Learning and Bandits for Speech and Language Processing

In recent years, reinforcement learning and bandits have transformed a wide range of real-world applications including healthcare, finance, recommendation systems, robotics and computer vision, and last but not least, the speech and language processing. While most speech and language applications of reinforcement learning algorithms are centered around improving deep network training with its flexible optimization properties, there are still many grounds to explore to utilize the benefits of reinforcement learning, such as its reward-driven adaptability, state representations, temporal structures, and generalizability. In this one-session tutorial, we will overview the recent advancements of reinforcement learning and bandits and discuss how they can be employed to solve various speech and natural language processing problems with models that are interpretable and scalable.
**Baihan Lin** is a machine learning and neuroscience researcher at Columbia University. His current theoretical research interest lies in the intersection between reinforcement learning, dynamical systems, geometric topology and complex networks, with extensive application interest in multiscale systems, especially in understanding the neural systems and the theory of neural networks, as well as developing neuroscience-inspired algorithms in signal processing, speech recognition, and computer vision domains that are efficient, scalable, interpretable and interactive. Before his PhD at Columbia University, he held a Masters Degree in Applied Mathematics from the University of Washington, Seattle. In the past few years, he maintains close collaborations with IBM, Google, Microsoft and Amazon on various application domains. According to the Google Scholar, he has authored 30+ publications with an H-index of 11+ and served on program committees or as reviewers for over 15 conferences including IJCAI, AAMAS, INTERSPEECH, AISTATS, NeurIPS, ICML, CVPR, ICCV, KDD, ICLR, AAAI, EMNLP and MICCAI, etc., as well as over 15 journals including Nature Scientific Reports, PLOS ONE, JACS, JInf, Entropy, Adv Complex Syst, IEEE Trans Knowl Data Eng, Comput Commun, Mathematics, Electronics, Signals, Algorithms, Behav Res Methods, Appl Sci, Front AI, Front Comp Neuro, Front Psychol, Front Robot AI and Comput Commun.

**Self-supervised Representation Learning for Speech Processing**

Although Deep Learning models have revolutionized the speech and audio processing field, they build specialist models for individual tasks and application scenarios. Deep neural models also bottlenecked dialects and languages with limited labelled data. Self-supervised representation learning methods promise a single universal model to benefit a collection of tasks and domains. They recently succeeded in NLP and computer vision domains, reaching new performance levels while reducing required labels for many downstream scenarios. Speech representation learning is experiencing similar progress with three main categories: generative, contrastive, and predictive. Other approaches relied on multi-modal data for pre-training, mixing text or visual data streams with speech. Although self-supervised speech representation is still a developing research area, it is closely related to acoustic word embedding and learning with zero lexical resources. This tutorial session will present self-supervised speech representation learning approaches and their connection to related research areas. Since many current methods focus solely on automatic speech recognition as a downstream task, we will review recent efforts on benchmarking learned representations to extend the application of such representations beyond speech recognition. A hands-on component of this tutorial will provide practical guidance on building and evaluating speech representation models.

**Hung-yi Lee** received PhD degree from National Taiwan University (NTU). He was a visiting scientist at the Spoken Language Systems Group of MIT CSAIL. He is an associate professor at National Taiwan University. He is the co-organizer of the special session on "New Trends in self-supervised speech processing" at Interspeech (2020), and the workshop on "Self-Supervised Learning for Speech and Audio Processing" at NeurIPS (2020).

**Abdelrahman Mohamed** is a research scientist at Meta AI research. He received his PhD from the University of Toronto where he was part of the team that started the Deep Learning revolution in Spoken Language Processing in 2009. He has been focusing lately on improving, using, and benchmarking learned speech representations, e.g. HuBERT, Wav2vec 2.0, TextlessNLP, and SUPERB.

**Shinji Watanabe** is an Associate Professor at CMU. He was a research scientist at NTT, Japan, a visiting scholar in Georgia Tech, a senior principal research scientist at MERL, and an associate research professor at JHU. He has published more than 200 peer-reviewed papers. He served as an Associate Editor of the IEEE TASLP. He was/has been a member of several technical committees, including the APSIPA SLA, IEEE SPS SLTC, and MLSP.
**Tara Sainath** is a Principal Research scientist at Google. She received her PhD from MIT in the Spoken Language Systems Group. She is an IEEE and ISCA Fellow and the recipient of the 2021 IEEE SPS Industrial Innovation Award. Her research involves applications of deep neural networks for automatic speech recognition, and has been very active in the community organizing workshops and special sessions on this topic.

**Karen Livescu** is a Professor at TTI-Chicago. She completed her PhD at MIT in the Spoken Language Systems group. She is an ISCA Fellow and an IEEE Distinguished Lecturer, and has served as a program chair for ICLR 2019 and Interspeech 2022. Her recent work includes multi-view representation learning, acoustic word embeddings, visually grounded speech models, spoken language understanding, and automatic sign language recognition.

**Shang-Wen Li** is a Research and Engineering Manager at Meta AI. He worked at Apple Siri, Amazon Alexa and AWS. He completed his PhD in 2016 from the Spoken Language Systems group of MIT CSAIL. He co-organized the workshop of “Self-Supervised Learning for Speech and Audio Processing” at NeurIPS (2020) and AAAI (2022), and the workshop of “Meta Learning and Its Applications to Natural Language Processing” at ACL (2021).

**Shu-wen Yang** is a PhD student at National Taiwan University. He co-created a benchmark for Self-Supervised Learning in speech, Speech processing Universal PERformance Benchmark (SUPERB). Before SUPERB, he created the S3PRL toolkit with Andy T. Liu, which supports numerous pretrained models and recipes for both pre-training and benchmarking. He gave a tutorial at the Machine Learning Summer School, Taiwan, 2021.

**Katrin Kirchhoff** is a Director of Applied Science at Amazon Web Services, where she heads several teams in speech and audio processing. She was a Research Professor at the UW, Seattle, for 17 years, where she co-founded the Signal, Speech and Language Interpretation Lab. She served on the editorial boards of Speech Communication and Computer, Speech, and Language, and was a member of the IEEE Speech Technical Committee.

---

**Speech Enhancement for Cochlear Implants: From Psychoacoustics to Machine Learning**

Cochlear implantation, as one of the most profound technological advances in modern medicine, is able to restore partial hearing and speech communication abilities to a large amount of profoundly deaf people. Cochlear implants (CIs) provide a valuable scientific model to investigate the impacts of psychoacoustic and linguistic cues on speech perception. Understanding speech in noisy environments remains challenging for most CI users; hence speech enhancement plays an extremely important role for CI speech processing and perception. Innovations from psychoacoustic knowledge to recent machine learning technology may provide novel solutions for the design of CI speech enhancement methods in challenging listening tasks.
Fei Chen received the B.Sc. and M.Phil. degrees from the Department of Electronic Science and Engineering, Nanjing University in 1998 and 2001, respectively, and the PhD degree from the Department of Electronic Engineering, The Chinese University of Hong Kong in 2005. He continued his research as post-doctor and senior research fellow in University of Texas at Dallas (supervised by Prof. Philipos Loizou) and The University of Hong Kong. He is now a full professor at Department of Electrical and Electronic Engineering, Southern University of Science and Technology (SUSTech), Shenzhen, China. Dr. Chen is leading the speech and physiological signal processing (SPSP) research group in SUSTech, with research focus on speech perception, speech intelligibility modeling, speech enhancement, and assistive hearing technology. He published over 100 journal papers and over 80 conference papers in IEEE journals/conferences, Interspeech, Journal of Acoustical Society of America, etc. He was tutorial speaker of "Intelligibility evaluation and speech enhancement based on deep learning" at Interspeech 2020, Shanghai, and organized special session "Signal processing for assistive hearing devices" at ICASSP 2015, Brisbane. He received the best presentation award in the 9th Asia Pacific Conference of Speech, Language and Hearing, and 2011 National Organization for Hearing Research Foundation Research Awards in States. Dr. Chen is an APSIPA distinguished lecturer, and is now serving as associate editor of <Biomedical Signal Processing and Control>, <Frontiers in Human Neuroscience>.

Yu Tsao received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2008. From 2009 to 2011, he was a Researcher with the National Institute of Information and Communications Technology, Tokyo, Japan, where he engaged in research and product development in automatic speech recognition for multilingual speech-to-speech translation. He is currently a Research Fellow (Professor) and Deputy Director with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. His research interests include speech and speaker recognition, acoustic and language modeling, audio coding, and bio-signal processing. He is currently an Associate Editor for the IEEE/ACM Transactions on Audio, Speech, and Language Processing and IEEE Signal Processing Letters and a Distinguished Lecturer of APSIPA. He was tutorial speakers of Interspeech 2020, ICASSP 2018, APSIPA 2021, APSIPA 2020, APSIPA 2019, APSIPA 2018 and ISCSLP 2018. He was the recipient of the Academia Sinica Career Development Award in 2017, the National Innovation Award in 2018, 2019, 2020, Future Tech Breakthrough Award 2019, and the Outstanding Elite Award, Chung Hwa Rotary Educational Foundation 2019–2020.

Sunday, September 18, 2022, 14:00 – 17:30 KST

Deep Spoken Keyword Spotting

Spoken keyword spotting (KWS) deals with the identification of keywords in audio streams and has become a fast-growing technology thanks to the paradigm shift introduced by deep learning a few years ago. This has allowed the rapid embedding of deep KWS in a myriad of small electronic devices with different purposes like the activation of voice assistants. In this tutorial, we will present a review into deep spoken KWS intended for practitioners and researchers who are interested in this technology. Particularly, we will cover an analysis of the main components of deep KWS systems, robustness methods, audio-visual KWS, applications and experimental considerations before concluding by identifying a number of directions for future research.

Iván López-Espejo received the M.Sc. degree in Telecommunications Engineering, the M.Sc. degree in Electronics Engineering and the Ph.D. degree in Information and Communications Technology, all from the University of Granada, Granada, Spain, in 2011, 2013 and 2017, respectively. In 2018, he was the leader of the speech technology team of Veridas, Pamplona, Spain. Since 2019, he is a post-doctoral researcher at the section for Artificial Intelligence and Sound at the Department of Electronic Systems of Aalborg University, Aalborg, Denmark. His research interests include speech enhancement, robust speech recognition and keyword spotting, multi-channel speech processing, and speaker verification.
Professor Zheng-Hua Tan received the B.Sc. and M.Sc. degrees in electrical engineering from Hunan University, Changsha, China, in 1990 and 1996, respectively, and the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 1999. He is a Professor in the Department of Electronic Systems and a Co-Head of the Centre for Acoustic Signal Processing Research at Aalborg University, Aalborg, Denmark. He is also a Co-Lead of the Pioneer Centre for AI, Denmark. He was a Visiting Scientist at the Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, USA, an Associate Professor at SJTU, Shanghai, China, and a postdoctoral fellow at KAIST, Daejeon, Korea. His research interests include machine learning, deep learning, pattern recognition, speech and speaker recognition, noise-robust speech processing, multimodal signal processing, and social robotics. He has authored/coauthored over 200 refereed publications. He is the Chair of the IEEE Signal Processing Society Machine Learning for Signal Processing Technical Committee (MLSP TC). He is an Associate Editor for the IEEE/ACM Transactions on Audio, Speech, and Language Processing. He has served as an Editorial Board Member for Computer Speech and Language and was a Guest Editor for the IEEE Journal of Selected Topics in Signal Processing and Neurocomputing. He was the General Chair for IEEE MLSP 2018 and a TPC Co-Chair for IEEE SLT 2016.

John H. L. Hansen (Fellow, IEEE) received the B.S.E.E. degree from the College of Engineering, Rutgers University, New Brunswick, NJ, USA, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA. In 2005, he joined the Erik Jonsson School of Engineering and Computer Science, the University of Texas at Dallas, Richardson, TX, USA, where he is currently an Associate Dean for research and a Professor of electrical and computer engineering. He also holds the Distinguished University Chair in telecommunications engineering and a joint appointment as a Professor of speech and hearing with the School of Behavioral and Brain Sciences. From 2005 to 2012, he was the Head of the Department of Electrical Engineering, the University of Texas at Dallas. At UT Dallas, he established the Center for Robust Speech Systems. From 1998 to 2005, he was the Department Chair and a Professor of speech, language, and hearing sciences, and a Professor of electrical and computer engineering with the University of Colorado Boulder, Boulder, CO, USA, where he co-founded and was an Associate Director of the Center for Spoken Language Research. In 1988, he established the Robust Speech Processing Laboratory. He has supervised 92 Ph.D. or M.S. thesis students, which include 51 Ph.D. and 41 M.S. or M.A. He has authored or coauthored 765 journal and conference papers including 13 textbooks in the field of speech processing and language technology, signal processing for vehicle systems, co-author of the textbook Discrete-Time Processing of Speech Signals (IEEE Press, 2000), Vehicles, Drivers and Safety: Intelligent Vehicles and Transportation (vol. 2 DeGruyter, 2020), Digital Signal Processing for In-Vehicle Systems and Safety (Springer, 2012), and the lead author of The Impact of Speech Under ‘Stress’ on Military Speech Technology (NATO RTO-TR-10, 2000). His research interests include machine learning for speech and language processing, speech processing, analysis, and modeling of speech and speaker traits, speech enhancement, signal processing for hearing impaired or cochlear implants, machine learning-based knowledge estimation and extraction of naturalistic audio, and in-vehicle driver modeling and distraction assessment for human–machine interaction. He is an IEEE Fellow for contributions to robust speech recognition in stress and noise, and ISCA Fellow for contributions to research for speech processing of signals under adverse conditions. He was the recipient of Acoustical Society of America’s 25 Year Award in 2010, and is currently serving as ISCA President (2017–2022). He is also a Member and the past Vice-Chair on U.S. Office of Scientific Advisory Committees (OSAC) for OSAC-Speaker in the voice forensics domain from 2015 to 2021. He was the IEEE Technical Committee (TC) Chair and a Member of the IEEE Signal Processing Society: Speech-Language Processing Technical Committee (SLTC) from 2005 to 2008 and from 2010 to 2014, elected the IEEE SLTC Chairman from 2011 to 2013, and elected an ISCA Distinguished Lecturer from 2011 to 2012. He was a Member of the IEEE Signal Processing Society Educational Technical Committee from 2005 to 2010, a Technical Advisor to the U.S. Delegate for NATO (IST/TG-01), an IEEE Signal Processing Society Distinguished Lecturer from 2005 to 2006, an Associate Editor for the IEEE Transactions on Audio, Speech, and Language Processing from 1992 to 1999 and the IEEE Signal Processing Letters from 1998 to 2000, Editorial Board Member for the IEEE Signal Processing Magazine from 2001 to 2003, and the Guest Editor in October 1994 for Special Issue on Robust Speech Recognition for the IEEE Transactions on Audio, Speech, and Language Processing. He is currently an Associate Editor for the JASA, and was on the Speech Communications Technical Committee for Acoustical Society of America from 2000 to 2003. In 2016, he was awarded the honorary degree Doctor Technices Honoris Causa from Aalborg University, Aalborg, Denmark in recognition of his contributions to the field of speech signal processing and speech or language or hearing sciences. He was the recipient of the 2020 Provost’s Award for Excellence in Graduate Student Supervision from the University of Texas at Dallas and the 2005 University of Colorado Teacher Recognition Award. He organized and was General Chair for ISCA Interspeech-2002, Co-Organizer and Technical Program Chair for the IEEE ICASSP-2010, Dallas, TX, and Co-Chair and Organizer for IEEE SLT-2014, Lake Tahoe, NV. He will be the Tech. Program Chair for the IEEE ICASSP-2024, and Co-Chair and Organizer for ISCA Interspeech-2022.
**Personalized Speech Enhancement: Data- and Resource-Efficient Machine Learning**

The tutorial introduces recent advancements in the emerging field of personalized speech enhancement. Personalizing a speech enhancement model leads to a compressed, thus efficient machine learning inference because the model focuses on a particular user's speech characteristics or their acoustic environment rather than trying to solve the general-purpose enhancement task. Since the test-time task can be seen as a smaller sub-problem of the generic speech enhancement problem, the model can also achieve better performance by solving a smaller and easier optimization problem. Moreover, since the goal is to adapt to the unseen test environment, personalization can improve the fairness of AI models for the users who are not well represented in the big training data. Meanwhile, personalized speech enhancement is challenging as it is difficult to utilize the personal information (e.g., clean speech) of the unknown test-time users ahead of time. In addition, acquiring such private data can increase concerns about privacy issues in speech applications. In this tutorial, we will explore various definitions of personalized speech enhancement in the literature and relevant machine learning concepts, such as zero- or few-shot learning approaches, data augmentation and purification, self-supervised learning, knowledge distillation, and domain adaptation. We will also see how these methods can improve data and resource efficiency in machine learning while achieving desired speech enhancement performance.

**Professor Minje Kim** is an Associate Professor in the Dept. of Intelligent Systems Engineering at Indiana University, where he leads his research group, Signals and AI Group in Engineering (SAIGE), and is affiliated with Data Science, Cognitive Science, Statistics, and Center for Machine Learning. He is also an Amazon Visiting Academic, working at Amazon Lab126. He earned his Ph.D. in the Dept. of Computer Science at the University of Illinois at Urbana-Champaign. Before joining IUUC, He worked as a researcher at ETRI, a national lab in Korea, from 2006 to 2011. Before then, he received his Master's and Bachelor's degrees in the Dept. of Computer Science and Engineering at POSTECH (Summa Cum Laude) and in the Division of Information and Computer Engineering at Ajou University (with honor) in 2006 and 2004, respectively. He is a recipient of various awards including NSF Career Award (2021), IU Trustees Teaching Award (2021), IEEE SPS Best Paper Award (2020), and Google and Starkey's grants for outstanding student papers in ICASSP 2013 and 2014, respectively. He is an IEEE Senior Member and also a member of the IEEE Audio and Acoustic Signal Processing Technical Committee (2018-2023). He is serving as an Associate Editor for EURASIP Journal of Audio, Speech, and Music Processing, and as a Consulting Associate Editor for IEEE Open Journal of Signal Processing. He is also a reviewer, program committee member, or area chair for the major machine learning and signal processing. He filed more than 50 patent applications as an inventor.

**A SpeechBrain for Everything: State of the PyTorch Ecosystem for Speech Technologies**

Neural models have become ubiquitous in speech technologies - almost all state-of-the-art speech technologies use deep learning as their foundation. This has made it possible for all applications to be built on top of a neural network library. This possibility is currently coming to fruition in the PyTorch ecosystem of speech technologies. This diverse landscape includes, for example, the automatic differentiation capable weighted finite state models in K2 and GTN, the powerful speech representations learned by pretrained wav2vec 2.0 models in Fairseq and Transformers, and the models, data loading and recipes in a large variety of tasks implemented by toolkits like ESPnet, NeMo, Asteroid and SpeechBrain, as well as a wealth of other tools such as the metric learning criteria of PyTorch Metric Learning. The first chapter of the tutorial covers PyTorch fundamentals and interfaces, and the subsequent chapters look at different practical challenges and tools that address them.

**Aku Rouhe** is currently a PhD student in Aalto University, Finland, working under the supervision of Prof. Mikko Kurimo. His research interests are novel speech recognition approaches, particularly end-to-end speech recognition, and how these new approaches should compare to more conventional hidden Markov model based approaches. Aku is also an experienced Python programmer, which he has been championing for close to 10 years. Aku is a core member of the original SpeechBrain development team.
Mirco Ravanelli is an Assistant Professor at Concordia University (Montréal, QC), an Adjunct Professor at the University of Montréal and an Associated Member of Mila. Previously, he was post-doc 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 co-author 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 (FR) 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 university of Montréal (Mila, QC, Canada) on the SpeechBrain project.

Peter Plantinga is an Applied AI/ML Associate at JPMorgan Chase & Co. He received his PhD in computer science from the Ohio State University (USA) under Prof. Eric Fosler-Lussier focusing on knowledge transfer for the tasks of speech enhancement, robust ASR, and reading verification. His current work involves adapting large-scale ASR models to work in the financial domain. Peter is one of the core members of the original SpeechBrain development team, and has continued to contribute since the original release.
Neural Speech Synthesis

Speech synthesis, which consists of several key tasks including text to speech (TTS) and voice conversion (VC), has been a hot research topic in the speech community and has broad applications in the industry. As the development of deep learning and artificial intelligence, neural network-based speech synthesis has significantly improved the quality of synthesized speech in recent years. In this tutorial, we give a comprehensive introduction to neural speech synthesis, which consists of four parts: 1) The history of speech synthesis technology and taxonomy of neural speech synthesis; 2) The key methods and applications of text to speech; 3) The key methods and applications of voice conversion; 4) Challenges in neural speech synthesis and future research directions.

Xu Tan is a Senior Researcher at Microsoft Research Asia. His research covers deep learning and its applications on language/speech/music processing, including text to speech, singing voice synthesis, automatic speech recognition, neural machine translation, music composition, etc. He has designed several popular text to speech (TTS) systems such as FastSpeech/NaturalSpeech, and transferred many technologies to Microsoft Azure TTS services. He has developed machine translation systems that achieved human parity on Chinese-English translation and won several champions on WMT machine translation competition, and developed language pre-training model MASS and AI music project Muzic. He has published more than 90 papers on top AI conferences and served as the action editor or area chair of some AI journals/conferences (e.g., TMLR, NeurIPS, AAAI). He is an executive member of the committee on Speech, Dialogue and Auditory Processing, and a member of the standing committee on Computational Art in China Computer Federation (CCF).

Hung-yi Lee is an Associate Professor of the Department of Electrical Engineering of National Taiwan University (NTU), with a joint appointment at the Department of Computer Science & Information Engineering of the university. His recent research focuses on developing technology that can reduce the requirement of annotated data for speech processing (including voice conversion and speech recognition) and natural language processing (including abstractive summarization and question answering). He won Salesforce Research Deep Learning Grant in 2019, AWS ML Research Award in 2020, Outstanding Young Engineer Award from The Chinese Institute of Electrical Engineering in 2018, Young Scholar Innovation Award from Foundation for the Advancement of Outstanding Scholarship in 2019, Ta-You Wu Memorial Award from Ministry of Science and Technology of Taiwan in 2019, and The 59th Ten Outstanding Young Person Award in Science and Technology Research & Development of Taiwan. He owns a YouTube channel teaching deep learning in Mandarin with about 100k Subscribers.
SPECIAL SESSIONS AND CHALLENGES

The Organizing Committee of INTERSPEECH 2022 is proudly announcing that the following special sessions and challenges for INTERSPEECH 2022 will be held.

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.

**APOLLO Fearless Steps**

The focus of this Special Session is to provide a forum for researchers working on the massive naturalistic audio collection stemming from the NASA Apollo Missions. UTDallas-CRSS under NSF support has led the Fearless Steps Initiative, a continued effort spanning eight years has resulted in the digitization, and recovery of over 50,000 hours of original analog audio data, as well as the development of algorithms to extract meaningful information from this naturalistic data resource, including an initial release of pipeline diarization meta-data for all 30 channels of APOLLO-11 and APOLLO-13 Missions. More than 500 sites worldwide have accessed the initial data. A current NSF Community Resource project is continuing this effort to recover the remaining Apollo missions (A7-A17; estimated to be 150,000hrs of data) in addition to motivating collaborative speech and language technology research through the Fearless Steps Challenge series.

**Organizers**

John H.L. Hansen, Univ. of Texas at Dallas  
Christopher Ceiri, Linguistic Data Consortium  
James Horan, NIST  
Aditya Joglekar, Univ. of Texas at Dallas  
Midia Yousefi, Univ. of Texas at Dallas  
Meena Chandra Shekar, Univ. of Texas at Dallas

**Audio Deep PLC Challenge**

The INTERSPEECH 2022 Audio Deep Packet Loss Concealment Challenge is intended to stimulate research in the area of Audio Packet Loss Concealment (PLC).

PLC is an important part of audio telecommunications technology and codec development, and methods for performing PLC using machine learning approaches are now becoming viable for practical use. Packet loss, either by missing packets or high packet jitter, is one of the top reasons for speech quality degradation in Voice over IP calls.

While there have been some groups publishing in this area, a lack of common datasets and evaluation procedures complicates the comparison of proposed methods and the establishment of clear baselines. With this challenge, we propose to address this situation: We will open source a dataset based on real-world (as opposed to the common synthetic) packet loss traces and bring the community together to, for the first time, compare approaches in this field on a unified test set.

As the gold standard for audio quality evaluation is human evaluator ratings, we will evaluate submissions using a crowd-source ITU-TP.808CCR approach. The top three approaches which achieve the highest average Mean Opinion Score on the blind set will be declared the winners of the INTERSPEECH 2022 Audio Deep Packet Loss Concealment Challenge. As an additional metric, to ensure that approaches are not degrading intelligibility, we will use the speech recognition rate, calculated using the Microsoft Cognitive Services Speech Recognition Service.

To help participants during the challenge, we will provide participants with access to our prototype "PLC-MOS" neural network model that provides estimates of human ratings of audio files with healed packet losses.

**Organizers**

Ross Cutler, Microsoft, USA  
Ando Saabas, Microsoft, Estonia
Challenges and Opportunities for Signal Processing and Machine Learning for Multiple Smart Devices

The increasing proliferation of smart devices in our lives offers tremendous opportunities to improve the customer experience by leveraging spatial diversity and distributed computational and memory capability. At the same time, multi sensor networks present unique challenges compared to single smart devices such as synchronization, arbitration, and privacy.

The purpose of this special session is to promote research in multiple device signal processing and machine learning by bringing together leading industry and academic experts to discuss the following topics including, but not limited to:
- Multiple device audio datasets
- Automatic speech recognition
- Keyword spotting
- Device arbitration (i.e. which device should respond to the user’s inquiry)
- Speech enhancement: de-reverberation, noise reduction, echo reduction
- Source separation
- Speaker localization and tracking
- Privacy sensitive signal processing and machine learning

The core motivation of this session is the recognition that "more is different". Robust speech recognition, enhancement, and analysis are foundational areas of speech signal processing with many publication outlets. The strength of the special session is to use the engineering specification of multiple devices as a backdrop against which creative solutions from these domains can be demonstrated. The session will co-locate top researchers working in the multi-sensor domain, and even though their specific applications may be different (e.g. enhancement vs acoustic event detection), the similarity of the problem space encourages cross pollination of techniques.

Organizers
- Jarred Barber, M.S., Amazon Alexa Speech
- Gregory Ciccarelli, Ph.D., Amazon Alexa Speech
- Israel Cohen, Ph.D., Amazon Alexa Speech, Technion-Israel Institute of Technology
- Tao Zhang, Ph.D., Amazon Alexa Speech

Inclusive and Fair Speech Technologies

Speech technologies have become increasingly used and now power a very large range of applications. Automatic speech recognition systems have indeed dramatically improved over the past decade thanks to the advances brought by deep learning and the effort on large-scale data collection. The speech technology community’s relentless focus on minimum word error rate has thus resulted in a productivity tool that works well for some categories of the population, namely for those of us whose speech patterns match its training data: typically, college-educated first-language speakers of a standardized dialect, with little or no speech disability.

For some groups of people, however, speech technology works less well, maybe because their speech patterns differ significantly from the standard dialect (e.g., because of regional accent), because of intra-group heterogeneity (e.g., speakers of regional African American dialects; second-language learners; and other demographic aspects such as age, gender, or race), or because the speech pattern of each individual in the group exhibits a large variability (e.g., people with severe disabilities).

The goal of this special session is (1) to discuss these biases and propose methods for making speech technologies more useful to heterogeneous populations and (2) to increase academic and industry collaborations to reach these goals.

Such methods include:
- analysis of performance biases among different social/linguistic groups in speech technology,
- new methods to mitigate these differences,
- new approaches for data collection, curation and coding,
- new algorithmic training criteria,
- new methods for envisioning speech technology task descriptions and design criteria.

Moreover, the special session aims to foster cross-disciplinary collaboration between fairness and personalization research, which has the potential to both improve customer experiences and algorithm fairness. The special session will bring experts from both fields to advance the cross-disciplinary study between fairness and personalization, e.g., fairness-aware personalization.

The session promotes collaboration between academia and industry to identify the key challenges and opportunities of fairness research and shed light on future research directions.

Organizers
Prof. Laurent Besacier, Naver Labs Europe, France, Principal Scientist,
Dr. Keith Burghardt, USC Information Sciences Institute, USA, Computer Scientist,
Dr. Alice Coucke, Sonos Inc., France, Head of Machine Learning Research,
Prof. Mark Allan Hasegawa-Johnson, University of Illinois, USA, Professor of Electrical and Computer Engineering,
Dr. Peng Liu, Amazon Alexa, USA, Senior Machine Learning Scientist,
Anirudh Mani, Amazon Alexa, USA, Applied Scientist,
Prof. Mahadeva Prasanna, IIT Dharwad, India, Professor, Dept of Electrical Engineering,
Prof. Priyankoo Sarmah, IIT Guwahati, India, Professor, Dept of Humanities and Social Sciences,
Dr. Odette Scharenborg, Delft University of Technology, the Netherlands, Associate professor,
Dr. Tao Zhang, Amazon Alexa, USA, Senior Manager.

Low-Resource ASR Development

The special session aims to bring together researchers from all sectors working on ASR (Automatic Speech Recognition) for low-resource languages and dialects to discuss the state of the art and future directions. It will allow for fruitful exchanges between participants in low-resource ASR challenges and evaluations and other researchers working on low-resource ASR development.

One such challenge is the Open ASR Challenge series conducted by NIST (National Institute of Standards and Technology) in coordination with IARPA’s (Intelligence Advanced Research Projects Activity) MATERIAL (Machine Translation for English Retrieval of Information in Any Language) program. The most recent challenge, OpenASR21, offered an ASR test of 15 low resource languages for conversational telephone speech, with additional data genres and case-sensitive scoring for some of the languages.

Another challenge is the Hindi ASR Challenge that was recently opened to evaluate regional variations of Hindi with the use of spontaneous telephone speech recordings made available by Gram Vaani, a social technology enterprise company. The regional variations of Hindi, together with spontaneity of speech, natural background, and transcriptions with varying degrees of accuracy due to crowd sourcing make it a unique corpus for automatic recognition of spontaneous telephone speech in low-resource regional variations of Hindi. A 1000 hours audio-only data (no transcription) is also released with this challenge to explore self-supervised training for such a low-resource framework.

We invite contributions from the OpenASR21 Challenge participants, the MATERIAL performers, the Hindi ASR Challenge participants, and any other researchers with relevant work in the low-resource ASR problem space.

Topics
Reports of results from tests of low-resource ASR, such as (but not limited to) the NIST/IARPA OpenASR21 Challenge, IARPA MATERIAL evaluations, and the Hindi ASR Challenge.

Topics focused on aspects of challenges and solutions in low-resource setting, such as:
- Zero- or few-shot learning methods
- Transfer learning techniques
- Cross-lingual training techniques
- Use of pretrained models
- Factors influencing ASR performance (such as dialect, gender, genre, variations in training data amount, or casing)
- Any 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
Prasanta Ghosh, Indian Institute of Science, Bangalore
William Hartmann, Raytheon BBN Technologies
Kay Peterson, National Institute of Standards and Technology
Aaditeshwar Seth, Indian Institute of Technology, Delhi

Low Resource Spoken Language Understanding

Progress in speech processing has been facilitated by shared datasets and benchmarks. Historically these have focused on automatic speech recognition (ASR), speaker identification, or other lower level tasks. Interest has been growing in higher-level spoken language understanding (SLU) tasks, including using end-to-end models, but there are fewer annotated datasets for such tasks, and the existing datasets tend to be relatively small. At the same time, recent work shows the possibility of pre-training generic representations and then fine-tuning for several tasks using relatively little labeled data. In this special session, we would like to foster a discussion and invite researchers in the field of SLU working on tasks such as named entity recognition (NER), sentiment analysis, intent classification, dialogue act tagging, or others, using either audio or ASR transcripts.

We invite contributions any relevant work in the low-resource SLU problem includes (but are not limited to):
- Training/fine-tuning approach using self/semi-supervised model for SLU tasks
- Comparison between pipeline and end-to-end SLU systems
- Self/semi-supervised learning approach focusing on SLU
- Multi-task/transfer/student-teacher learning focusing on SLU tasks
- Theoretical or empirical study on low-resource SLU problems

Organizers  Suwon Shon - ASAPP
Felix Wu - ASAPP
Pablo Brusco - ASAPP
Kyu J. Han - ASAPP
Karen Livescu - TTI at Chicago
Ankita Pasad - TTI at Chicago
Yoav Artzi - Cornell University
Katrin Kirchhoff- Amazon
Samuel R. Bowman - New York University
Zhou Yu - Columbia University

Non-Intrusive Objective Speech Quality Assessment (NISQA) Challenge for Online Conferencing Applications

ConferencingSpeech 2022 challenge is proposed to stimulate research in Non-intrusive speech quality assessment for online conferencing applications. For a long time, speech quality assessment of communication application was carried out by subjective experiments or obtained via computational model relying on the reference clean and degraded speech in an intrusive manner. However, for quality monitoring purpose non-intrusive speech quality model or so-called single-ended model which do not need reference speech is highly preferred and remains a difficult and challenging topic. The challenge aims to bring together researchers from all sectors working on speech quality to show the potential performance of different models, explore new ideas, and discuss the state of the art and future directions. We believe this could accelerate the research topic to make non-intrusive speech quality assessment more reliable and increase the possibility that those models being adopted by online conferencing applications in a near future.

This challenge will provide comprehensive training datasets, a comprehensive test dataset and a baseline system. The final
ranking of this challenge will be decided by the accuracy of the predicted MOS scores from the submitted model or algorithm on the test dataset. More details about the data and challenge can be found from the evaluation plan. Please let us know if you have questions or need clarification about any aspect of the challenge.

Organizers  Gaoxiong Yi, Tencent, China  
Wei Xiao, Tencent, China  
Yiming Xiao, Tencent, China  
Babak Naderi, Technical University of Berlin, Germany  
Sebastian Möller, Technical University of Berlin, Germany  
Gabriel Mittag, Machine Learning Scientist, Microsoft  
Ross Cutler, Partner Applied Scientist Manager, Microsoft  
Zhuohuang Zhang, Indiana University Bloomington, USA  
Donald S. Williamson, Assistant Professor, Indiana University Bloomington, USA  
Fei Chen, Professor, Southern University of Science and Technology, China  
Fuzheng Yang, Professor, XiDian University, China  
Shidong Shang, Senior Director, Tencent, China

Speaking Styles and Interaction Styles

Style is becoming more important, as we increasingly deploy variations of one basic dialog system across domains and genres, and as we aim to better customize and individualize our dialog systems.

Style has been a focus of much recent work in speech synthesis, with remarkable advances also in style transfer, style discovery, style recognition, and style modeling, both for utterance-level style properties and interaction-level and dialog-level properties. Nevertheless more work is needed in improving and simplifying our models, in generalizing and systematizing our understanding of style, and in translating research advances to value for users.

In this special session, we seek to promote interaction and collaboration between researchers working on different aspects of style and using different approaches. We encourage submissions that go beyond their technical or empirical contributions to also elaborate on how the work relates to the big picture of style in spoken dialog. We also welcome papers whose motivations, contributions, or implications highlight issues not commonly addressed at Interspeech.

Topics of interest include any aspects of speaking styles and interaction styles, including:
- style as it relates to expressiveness, pragmatic intents, genre, social role, social identity, stance, personality, entrainment, interpersonal dynamics, and so on
- universal and language-specific aspects of style
- style in monolog and dialog
- how styles are realized through phonetic, prosodic, lexical, and turn-taking means
- applications in dialog systems and beyond

Organizers  Nigel Ward, University of Texas at El Paso  
Kallirroi Georgila, University of Southern California  
Yang Gao, Carnegie-Mellon University  
Mark Hasegawa-Johnson, University of Illinois  
Koji Inoue, Kyoto University  
Simon King, University of Edinburgh  
Rivka Levitan, City University of New York  
Katherine Metcalf, Apple  
Eva Szekely, KTH Royal Institute of Technology  
Pol van Rijn, Max Planck Institute for Empirical Aesthetics  
Rafael Valle, NVIDIA
Speech and Language in Health: From Remote Monitoring to Medical Conversations

Technological advancements have been rapidly transforming healthcare in the last several years, with speech and language tools playing an integral role. However, this brings a multitude of unique challenges to consider to increase the generalisability, reliability, interpretability and utility of speech and language tools in healthcare and health research settings.

Many of these challenges are common to the two themes of this special session. The first theme, From Collection and Analysis to Clinical Translation, seeks to draw attention to all aspects of speech-health studies that affect the overall quality and reliability of any analysis undertaken on the data and thus affect user acceptance and clinical translation.

The second theme, Language Technology For Medical Conversations, covers a growing field of research in which automatic speech recognition and natural language processing tools are combined to automatically transcribe and interpret clinician-patient conversations and generate subsequent medical documentation.

By combining these themes, this session will bring the wider speech-health community together to discuss innovative ideas, challenges and opportunities for utilizing speech technologies within the scope of healthcare applications.

Suggested paper topics include, but are not limited to:
- Data collection protocols and speech elicitation strategies
- Device selection and related effects
- Acceptance of data collection in different health cohorts
- Longitudinal data collection and analysis
- Patient and Public Involvement in speech research
- User evaluation of speech technology in a healthcare setting
- Feature extraction and novel representations that provide clinical interpretability
- Advancements in analytics and machine learning methodologies that are clinically or biologically inspired
- Fusion of linguistic and paralinguistic information
- Health-related conversational analytics
- Speech recognition and natural language processing in healthcare settings
- Creation and annotation of medical conversation datasets
- Role of medical conversation understanding in reducing documentation burden
- Use of chatbots in healthcare
- Spoken language technologies in real-world health settings
- Utilising Electronic Health Records to personalise models in speech recognition or conversational analytics

Organizers

Nicholas Cummins (Kings's College London and Thymia)
Thomas Schaaf (3M)
Heidi Christensen (University of Sheffield)
Judith Dineley (King’s College London and University of Augsburg)
Julien Epps (University of New South Wales)
Matt Gormley (Carnegie Mellon University)
Sandeep Konam (Abridge.ai)
Emily Mower Provost (University of Michigan)
Chaitanya Shivade (Amazon.com)
Thomas Quatieri (MIT Lincoln Laboratory)

Speech Intelligibility Prediction for Hearing-Impaired Listeners

One of the greatest challenges for hearing-impaired listeners is understanding speech in the presence of background noise. Noise levels encountered in everyday social situations can have a devastating impact on speech intelligibility, and thus communication effectiveness, potentially leading to social withdrawal and isolation. Disabling hearing impairment affects 360 million people worldwide, with that number increasing because of the ageing population. Unfortunately, current hearing aid technology is often ineffective at restoring speech intelligibility in noisy situations.
To allow the development of better hearing aids, we need better ways to evaluate the speech intelligibility of audio signals. We need prediction models that can take audio signals and use knowledge of the listener’s characteristics (e.g., an audiogram) to estimate the signal’s intelligibility. Further, we need models that can estimate intelligibility not just of natural signals, but also of signals that have been processed using hearing aid algorithms - whether current or under development.

**The Clarity Prediction Challenge**

As a focus for the session, we have launched the ’Clarity Prediction Challenge’. The challenge provides you with noisy speech signals that have been processed with a number of hearing aid signal processing systems and corresponding intelligibility scores produced by a panel of hearing-impaired individuals. You are tasked with producing a model that can predict intelligibility scores given just the signals, their clean references and a characterisation of each listener’s specific hearing impairment. The challenge will remain open until the Interspeech submission deadline and all entrants are welcome. (Note, the Clarity Prediction Challenge is part of a 5-year programme with further prediction and enhancement challenges planned for the future.)

**Relevant Topics**

The session welcomes submission from entrants to the Clarity Prediction Challenge but is also inviting papers related to topics in hearing impairment and speech intelligibility, including, but not limited to:

- Statistical speech modelling for intelligibility prediction
- Modelling energetic and informational noise masking
- Individualising intelligibility models using audiometric data
- Intelligibility prediction in online and low latency settings
- Model-driven speech intelligibility enhancement
- New methodologies for intelligibility model evaluation
- Speech resources for intelligibility model evaluation
- Applications of intelligibility modelling in acoustic engineering
- Modelling interactions between hearing impairment and speaking style
- Papers using the data supplied with the Clarity Prediction Challenge

**Organizers**

- Trevor Cox - University of Salford, UK
- Fei Chen - Southern University of Science and Technology, China
- Jon Barker - University of Sheffield, UK
- Daniel Korzekwa - Amazon TTS
- Michael Akeroyd - University of Nottingham, UK
- John Culling - University of Cardiff, UK
- Graham Naylor - University of Nottingham, UK

**Spoofing-Aware Automatic Speaker Verification (SASV)**

While spoofing countermeasures, promoted within the sphere of the ASVspoof challenge series, can help to protect reliability in the face of spoofing, they have been developed as independent subsystems for a fixed ASV subsystem. Better performance can be expected when countermeasures and ASV subsystems are both optimised to operate in tandem. The first spoofing-aware speaker verification (SASV) challenge aims to encourage the development of original solutions involving, but not limited to:

back-end fusion of pre-trained automatic speaker verification and pre-trained audio spoofing countermeasure subsystems; integrated spoofing-aware automatic speaker verification systems that have the capacity to reject both non-target and spoofed trials.

While we invite the submission of general contributions in this direction, the Interspeech 2022 Spoofing-aware Automatic Speaker Verification special session incorporates a challenge – SASV 2022. Potential authors are encouraged to evaluate their solutions using the SASV benchmarking framework which comprises a common database, protocol and evaluation metric. Further details and resources can be found from the SASV challenge website.

**Organizers**

- Jee-weon Jung, Naver Corporation, South Korea
- Hemlata Tak, EURECOM, France
Trustworthy Speech Processing

Given the ubiquity of Machine Learning (ML) systems and their relevance in daily lives, it is important to ensure private and safe handling of data alongside equity in human experience. These considerations have gained considerable interest in recent times under the realm of Trustworthy ML. Speech processing in particular presents a unique set of challenges, given the rich information carried in linguistic and paralinguistic content including speaker trait, interaction and state characteristics. This special session on Trustworthy Speech Processing (TSP) was created to bring together new and experienced researchers working on trustworthy ML and speech processing. We invite novel and relevant submissions from both academic and industrial research groups, showcasing advancements in theoretical, empirical as well as real-world design of trustworthy speech applications.

Topics of interest cover a variety of papers centered on speech processing, including (but not limited to):
- Differential privacy
- Federated learning
- Ethics in speech processing
- Model interpretability
- Quantifying & mitigating bias in speech processing
- New datasets, frameworks and benchmarks for TSP
- Discovery and defense against emerging privacy attacks
- Trustworthy ML in applications of speech processing like ASR

Organizers
Anil Ramakrishna, Amazon Inc.
Shrikanth Narayanan, University of Southern California
Rahul Gupta, Amazon Inc.
Isabel Trancoso, University of Lisbon
Rita Singh, Carnegie Mellon University

The VoiceMOS Challenge

Human listening tests are the gold standard for evaluating synthesized speech. Objective measures of speech quality have low correlation with human ratings, and the generalization abilities of current data-driven quality prediction systems suffer significantly from domain mismatch. The VoiceMOS Challenge aims to encourage research in the area of automatic prediction of Mean Opinion Scores (MOS) for synthesized speech. This challenge has two tracks:

Main track: We recently collected a large-scale dataset of MOS ratings for a large variety of text-to-speech and voice conversion systems spanning many years, and this challenge releases this data to the public for the first time as the main track dataset.

Out-of-domain track: The data for this track comes from a different listening test from the main track. The purpose of this track is to study the generalization ability of proposed MOS prediction models to a different listening test context. A smaller amount of labeled data is made available to participants, and unlabeled audio samples from the same listening test are made available as well, to encourage exploration of unsupervised and semi-supervised approaches.

Participation is open to all. The main track is required for all participants, and the out-of-domain track is optional. Partici-
pants in the challenge are strongly encouraged to submit papers to the special session. The focus of the special session is on understanding and comparing MOS prediction techniques using a standardized dataset.

**Organizers**  
Wen-Chin Huang (Nagoya University, Japan)  
Erica Cooper (National Institute of Informatics, Japan)  
Yu Tsao (Academia Sinica, Taiwan)  
Hsin-Min Wang (Academia Sinica, Taiwan)  
Tomoki Toda (Nagoya University, Japan)  
Junichi Yamagishi (National Institute of Informatics, Japan)
ISCA STUDENT ADVISORY COMMITTEE

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 Catarina Botelho (General Coordinator), Francisco Teixeira (Event Coordinator), Thomas Rolland (Event Coordinator), and Omnia Ibrahim (Media Coordinator). More information about ISCA-SAC can be found on our website: http://www.isca-students.org/sacweb/index.php/about-us.

This year, we are pleased to announce that ISCA-SAC will host three student events at INTERSPEECH 2022: Doctoral Consortium, Students Meet Experts, as well two mentoring events, which are co-organized with ISCA’s Mentoring Committee. All information about these events is available at http://www.isca-students.org/sacweb/index.php and https://interspeech2022.org/students/student_events.php.

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.

8th Doctoral Consortium

Saturday, September 17th, 2022. Incheon, South Korea and online

The 8th Doctoral Consortium 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.

The Doctoral Consortium will be held in a hybrid format. Participants pre-submitted an extended abstract and were selected based on their submissions. During the event, students will be given 15 minutes to present their work, followed by 15 minutes of discussion with the panel of experts. After the event, the participants’ extended abstracts will also be published on ISCA-SAC’s website.

For those that missed this year’s submission deadline for the Doctoral Consortium please consider submitting an abstract next year! You can find extended abstracts published in previous years in the following link: http://www.isca-students.org/sacweb/index.php/resources

Mentoring Events at Interspeech 2022
– An Initiative of ISCA’s Mentoring Committee and ISCA-SAC –

Tuesday, September 20th, 2022. Incheon, South Korea and online

After three successful editions of the mentoring event, at Interspeech 2019 in Graz, Interspeech 2020 in Shanghai (online), and Interspeech 2021 in Brno (online), ISCA’s new Mentoring Committee is joining forces with the Student Advisory Committee, ISCA-SAC to organise mentoring events for the fourth time, at Interspeech 2022 in Incheon.

The same as last year, we will hold two events: round tables and one-on-one mentoring. We invite all researchers in any stage of their career to participate. Participants will be given the opportunity to engage in group and one-to-one discussions, respectively, with mid-career and senior researchers from academia and industry in a friendly, welcoming environment.

Round table discussions:

To cater for different needs and wishes, we will have separate round tables – tables for PhD students only and tables for a mix of people at different career stages. Each table will have an assigned topic, two mentors, and 6–8 participants. Mentors will be confirmed nearer the time.

Potential topics for discussion, subject to demand, are listed below.

1. Handling disagreement and conflict in research
2. Why was my paper rejected? Insights from reviewers
3. Reaching out: Growing your network and creating new collaborations
4. Research integrity and being a responsible researcher
5. Writing your first grant proposal: tips for beginners
6. Contributing to your academic community: mentoring, supporting and being a good peer
7. Managing the supervisor - student relationship
8. Under pressure: how to navigate the competitive academic environment
9. The research-life juggle: time management, productivity and work-life balance
10. Planning your career in speech: strategies for success
11. Handling setbacks in your PhD
12. Research in academia versus industry: Which is right for me?
13. Defining research questions for your PhD
14. Future of Speech Domain Startup
15. Everyday tips for a PhD student

One-on-one mentoring:
You also have the chance to have one-on-one mentoring, for participants that feel more comfortable discussing certain topics in a one-on-one session with a mid- to senior career researcher.
The links for participating in the event and volunteering as a mentor can be found on the websites listed on the top.

9th Students Meet Experts

Wednesday, September 21th, Incheon. South Korea and online (tentative)

After successful editions in Lyon (2013), Singapore (2014), San Francisco (2016), Stockholm (2017), Hyderabad (2018), Graz (2019), virtually in both Shanghai (2020) and Brno (2021), we are excited to announce that the Students Meet Experts event is now coming to Interspeech 2022 in Incheon, South Korea. There will be a panel discussion with experts from academia and industry, where the experts respond to questions submitted by students. All students are welcome to participate!

You can already submit your questions using the following form:

https://forms.gle/4WryTGLrop1gepv79
| AREA CHAIRS |
|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| 1. Speech Perception, Production and Acquisition | 7. Speech Synthesis and Spoken Language Generation |
| • Jeesun Kim, Western Sydney University | • Berrak Sisman, National University of Singapore |
| • Fanny Meunier, University of Cote d’Azur | • Esther Klabbers-Judd, ReadSpeaker |
| • Prasanta Ghosh, Indian Institute of Science | • Heiga Zen, Google Inc. |
| 2. Phonetics, Phonology and Prosody | • Juhan Nam, Korea Advanced Institute of Science and Technology |
| • Priyankoo Sarmah, Indian Institute of Technology Guwahati | • Eva Navas, University of the Basque Country |
| • Keikichi Hirose, University of Tokyo | • Zhenhua Ling, University of Science and Technology of China |
| • Ioana Vasilescu, LISN-CNRS | • Jindrich Matousek, University of West Bohemia |
| 3. Analysis of Paralinguistics in Speech and Language | 8. Speech Recognition – Signal Processing, Acoustic Modeling Robustness and Adaptation |
| • Khiet Truong, University of Twente | • Marc Delcroix, NTT Communication Science Laboratories |
| • Shrikanth Narayanan, University of Southern California | • Thomas Hain, University of Sheffield |
| • Chloé Clavel, Télécom Paris | • Martin Karafiat, Brno University of Technology |
| • Carol Espy-Wilson, University of Maryland | • Penny Karanasou, Amazon |
| 4. Speaker and Language Identification | • Umesh Srinivasan, Indian Institute of Technology Madras |
| • Kong Aik Lee, Agency for Science, Technology and Research | • Kate Knill, University of Cambridge |
| • Alicia Lozano, Universidad Autonoma de Madrid | • Namsoo Kim, Seoul National University |
| • Srikanth Madikeri, Idiap Research Institute | 9. Speech Recognition – Architecture Search, and Linguistic Components |
| • Hong-Goo Kang, Yonsei University | • Preethi Jyothi, Indian Institute of Technology Bombay |
| • Joon Son Chung, KAIST | • Murat Saraclar, Boğaziçi University |
| • Zbyněk Koldovský, Technical University of Liberec | • Ngoc Thang Vu, University of Stuttgart |
| • K. Sri Rama Murty, Indian Institute of Technology Hyderabad | • Minhwa Chung, Seoul National University |
| • P. Vijayalakshmi, SSN College of Engineering | 10. Speech Recognition – Technologies and Systems for New Applications |
| • John Hershey, Google Research | • Florian Metze, Carnegie Mellon University |
| • Keunwoo Choi, ByteDance | • Jiangyan Yi, Chinese Academy of Sciences |
| • Zeyu Jin, Adobe Research | • Myoung-Wan Koo, Sogang University |
| • Yu Tsao, Academia Sinica | 11. Spoken Dialog Systems and Conversational Analysis |
| • Ina Kodrasi, Idiap Research Institute | • Alexandros Potamianos, National Technical University of Athens |
| • Vladimir Malenovsky, University of Sherbrooke | • Catharine Oertel, TU Delft |
| • Junfeng Li, Chinese Academy of Sciences | • Vivian Yun-Nung Chen, National Taiwan University |
| • Bastiaan Kleijn, Victoria University of Wellington | • Helen Meng, Chinese University of Hong Kong |
| • Rainer Martin, Ruhr-Universität Bochum |
12. Spoken Language Processing: Translation, Information Retrieval, Summarization, Resources and Evaluation

- Bin Ma, Alibaba Inc.
- Jan Trmal, Johns Hopkins University
- Sakriani Sakti, Japan Advanced Institute of Science and Technology
- Nancy Chen, Institute for Infocomm Research

13. Speech, Voice, and Hearing Disorders

- Mathew Magimai Doss, Idiap Research Institute
- Inyong Choi, University of Iowa
- Ning Ma, University of Sheffield

14. Special Sessions

- Jon Barker, University of Sheffield, UK
- Odette Sharenborg, Delft University of Technology, Netherlands
- Jingdong Chen, Northwestern Polytechnical University, China
- Joon Son Chung, KAIST
SATELLITE EVENTS

Speech, Music and Mind 2022

- Webpage: http://workshops.ifs.tuwien.ac.at/SMM22
- Date & Time: September 15, 2022
- Venue: Online Event (please see the event webpage for information)

Far Field Speaker Verification Challenge 2022 (FFSVC 2022)

- Webpage: https://ffsvc.github.io/
- Date & Time: September 17, 2022
- Venue: Online Event (please see the event webpage for information)

Young Female Researchers in Speech Workshop (YFRSW)

- Webpage: https://sites.google.com/view/yfrsw-2022/
- Date & Time: September 17, 2022, 9am-5pm
- Venue: Inha University, Korea

VoxCeleb Speaker Recognition Challenge and Workshop 2022 (VoxSRC-2022)

- Webpage: http://mm.kaist.ac.kr/voxsrc/
- Date & Time: September 22, 2022
- Venue: Songdo ConvensiA

2nd Symposium on Security and Privacy in Speech Communication Joined with 2nd VoicePrivacy Challenge

- Webpage: https://symposium2022.spsc-sig.org/
- Date & Time: September 23-24, 2022
- Venue: Hybrid (Online and Incheon National University, Korea)

Speech for Social Good (S4SG) Workshop

- Webpage: http://s4sg-workshop.github.io
- Date of the event: September 24-25, 2022
- Venue: Online event (please see the event webpage for information)
SHAHINA A
ALBERTO ABAD
SOLOMON TEFFERA ABATE
OSSAMA ABDEL-HAMID
NASSIMA ABDELLI-BERUH
AJISH KURIAKOSE ABRAHAM
VINAYAK ABROL
ALEX ACERO
SIVANAND ACHANTA
M P ACTLIN JEEVA
ANDRE ADAMI
SHARATH ADAVANNE
MARTINE ADDA-DECKER
FARAH ADEEBA
YOSSI ADI
NAGARAJ ADIGA
AMBER AFSHAN
SACHIN AGARWAL
SHYAM AGRAWAL
CARLA AGURTO
REHAN AHMAD
MOHSIN AHMED
MASATO AKAGI
YUYA AKITA
MARYAM AL DABEL
OUAMYA AL DAKKAK
MD JAHANGIR ALAM
FIROJ ALAM
AARON ALBIN
FELIX ALBU
ZAKARIA ALDENEH
YAHYA ALDHLOMI
JAN ALEXANDERSSON
ANASTASIOS AXANDRIDIS
AHMED ALI
PAAVO ALKI
CYRIL ALLAUAUZEN
JESÚS B. ALONSO-HERNÁNDEZ
MOHAMMED SALAH AL-RADHI
SARAH ALSHAREEF
JALAL AL-TAMIMI
TANEL ALUMÄE
ABEEER ALWAN
ANGÉLIQUE AMELOT
NOAM AMIR
SHAHIN AMIRIPARIAN
GUOZHEN AN
TASOS ANASTASAKOS
BISTRA ANDREEVA
JESÚS ANDRÉS-FERRER
JORN ANEMULLER
PONTGEP ANKITITRAKUL
XAVIER ANGUERA
GOPALA KRISHNA ANUMANCHIPALLI
SALVATORE ANZALONE
VIJENDRA APSINGEKAR
TAKAYUKI ARAI
SHOKO ARAKI
JULIAN DAVID ARIAS LONDOÑO
DHANY ARIFIANTO
YASUO ARIKI
YOSHIKO ARIMOTO
EBRU ARISOY
HAGAI ARONOWITZ
DILEEP AROOR DINESH
ALI AROUDI
HARISH ARSIKERE
MICHAEL ASHBY
BHISHNU ATAL
BAGUS TRIS ATMARA
KARTIK AUDHKKASHI
CARLOS AVENDANO
MATTHEW AYLETT
DALIY AYOUB
BHARATHI B
MOLLY BABEL
MICHEL BACCHIANI
TOM BÄCKSTRÖM
SAMEER BADASKAR
PIERRE BADIN
DEBLIN BAGCHI
PAOLO BAGGIA
LADAN BAGHAI-RAVARY
YE BAI
LINXUE BAI
GÉRARD BAILLY
ANDREA BANDINI
YOSHIKI BANDO
CHANGCHUN BAO
JORGE BAPTISTA
PLINIO BARBOSA
NELLY BARBOT
DANTE BARONE
ROBERTO BARRA-CHICOTE
CLAUDE BARRAS
VINCENT BARRIAC
MURALI KARTHICK BASKAR
FERNANDO BATISTA
TIMO BAUMANN
ALI ORKAN BAYER
FREDERIC BECHET
TILMAN BECKER
MARY BECKMAN
STEVE BEEF
HOMAYOON BEIGI
PETER BELL
SOFIA BEN JEBARA
WAAD BEN KHEDER
ATEF BEN YOUSSEF
JOSE MIGUEL BENEID
ȘTEFAN BĂNNĂȘ
FAOUZI BENZEGBIBA
NICOLE BERINGER
FREDERIC BERTHOMMIER
LAURENT BESACIER
JONAS BESKOW
SIMON BETZ
JAYADEV BILLA
FRÉDÉRIC BIMBOT
PETER BIRKHOHL
JUDITH BISHOP
MARIA PAOLA BISSIRI
AJIRIJT BISWAS
JOSE LUIS BLANCO MURILLO
RONALD BÖCK
TOBIAS BOCKLET
BAJJABABA BOLLEPALI
ANTONIO BONAFONTE
JEAN-FRANCOIS BONASTRE
ZINNY BOND
ANNE BONNEAU
HÉLIÈNE BONNEAU
JONAS BORGSTROM
HYNEK BORIL
TOMÁŠ BOŘIL
MARVIN BORSFORD
MICHAL BORSKY
MARIEM BOUAFIF
GILLES BOULIANNE
RACHEL E. BOUSERHAL
PIERRE-MICHEL BOUSQUET
SUZANNE BOYCE
ANN BRADLOW
DANIEL BRAITHWAITE
DAVID BRAINTWAITE
ANGELIKA BRAUN
BEITINA BRAUN
SEBASTIAN BRAUN
HERVÉ BREDIN
Andrew Breun
JON BRIDLE
MARKUS BRÜCKL
Raymond BRUECKNER
Laurence Bruggeman
PABLO BRUSCO
ALESSIO BRUTTI
HARRY BUNT
Susanne Burger
Lukas Burget
Nam Soo Kim
Hoi Rin Kim
Jaebok Kim
Hyung Soon Kim
Sanghun Kim
Minje Kim
Byeongchang Kim
Kwangyoun Kim
Wool Kim
Jangwon Kim
Yelin Kim
Eugene Kim
Yeon-Jun Kim
Sunhee Kim
Jong-mi Kim
Simon King
Brian King
Brian Kingsbury
Tomi Kinnunen
Keisuke Kinosita
Irina Kipyatkova
Andreas Søeborg Kirkedal
Norihide Kitaoka
Esther Klabbers
Felicitas Kleber
Thomas Kleinbauer
Rebecca Kleinberger
Ondrej Klejch
Viacheslav Klimkov
Nicola Klingler
Kate Knill
Han-seok Ko
Satoshi Kobashikawa
Akio Kobayashi
Kazuhiko Kobayashi
Daniil Kocharov
Alexei Kochetov
Marcel Kockmann
Ina Kodrasi
Sri Rama Murty Kodukula
Laura Koenig
Tina Kohler
Daniel Kohlsdorf
Yuma Koizumi
Jachym Kolar
Zbyněk Koldovský
Thomas Kollar
Kazunori Komatani
Sunil Kumar Koppaparupu
Tomoki Koriyama
Srikanth Korse
Daniel Korzekwa
Takafumi Koshinaka
Hardik Kothare
Sandeep Reddy Kothinti
Maria Koutsogiannaki
György Kovács
Konrad Kowalczyk
Pavel Kral
Ivan Kraljevski
Kelsey Kraus
Sreenivasa Rao Krothapalli
Georg Kucsko
Roland Kuhn
Ivan Kukanov
Naveen Kumar
Manoj Kumar
Varun Kumar
Shankar Kumar
Kshitiz Kumar
Vivek Kumar
Anurag Kumar
R Kumaraswamy
Jimmy Kunzmann
Grace Chen-Hsiu Kuo
Gakuto Kurata
Mikko Kurimo
Frank Kurth
Imed Laaridh
Catherine Lai
Ying-Hui Lai
Divesh Lala
Pierre Lanchantin
Adrian Lancucki
Itshak Lapidot
Yves Laprie
Anthony Larcher
Staffan Larsson
Kornel Laskowski
Siddique Latif
Javier Latorre
Antoine Laurent
Galina Lavrentyeva
Aaron Lawson
Viet Bac Le
Duc Le
Gaël Le Lan
Sébastien Le Maguer
Jonathan Le Roux
Gwénolé Lecorvé
Chi-Chun Lee
Seunghun Lee
Yoonjeong Lee
Lin-shan Lee
Hung-yi Lee
Shi-wook Lee
Tan Lee
Donghyun Lee
Chin-Hui Lee
Jaewon Lee
Juheon Lee
Bowon Lee
Fabrice Lefèvre
Milan Legát
Zhenchun Lei
Yun Lei
Manon Lelandais
Cheung-Chi Leung
Gary Leung
Susannah Levi
Michael Levit
Rivka Levitan
Gina-Anne Levow
Sheng Li
Haoqi Li
Lantian Li
Ya Li
Yanxiong Li
Jinyu Li
Ruizhi Li
Bo Li
Ming Li
Xian Li
Kehuang Li
Yongwei Li
Xiang Li
Haifeng Li
Aijun Li
Lin Li
Pei-Chun Li
Peng Li
Dingzeyu Li
Ruirui Li
Zheng Lian
Shan Liang
Paul Pu Liang
Yuan-Fu Liao
Hank Liao
Tatiana Likhomanenko
Rasa Lileikyte
Jason Lilley
Carlos Lima
Amaro Lima
Kin Wah Edward Lin
Qiguang Lin
Baihan Lin
Ju Lin
Georges Linares
Yang Liu
Bin Liu
Wenju Liu
Rui Liu
Xunying Liu
Jia Liu
Yang Liu
Li Liu
Xiaoyu Liu
Xuefei Liu
Manoj Riyal
Matthew Roddy
Luis Javier
Rodriguez-Fuentes
Rosalia Rodriguez-Vazquez
Axel Roebel
Rungpat Roengpitya
Simon Rooesig
Johan Rohdin
Mikael Roll
Srikanth Ronanki
Richard Rose
Ralph Rose
Andrew Rosenberg
Sophie Rosset
Jean-Luc Rouas
Mickael Rouvier
Frank Rudzicz
Vesa Ruoppila
Michela Russo
Oleg Rybakov
Omid Sadjadi
Najmeh Sadoughi
Sylvie Saget
Lakshmi Saheer
Md Sahidullah
Saurabh Sahu
Hardik Sailor
Tara Sainath
Yuki Saito
Daisuke Saito
Shinsuke Sakai
Ken-Ichi Sakakibara
Elizabeth Salesky
Elliot Saltzman
Stan Salvador
Giampiero Salvi
Eugenia San Segundo
Fernández
Rubén San Segundo
Hernández
Victoria Sanchez
Jon Sanchez
Emilio Sanchis
Germán Sanchis-Trilles
Mufan Sang
Joao Felipe Santos
George Saon
Ashtosh Sapru
Murat Saraclar
Ibon Saratxaga
Seyyed Saeed Sarfjoo
Leda Sari
Rahi Sarikaya
Achintya kumar Sarkar
Mousmita Sarma
Milton Sarria-Paja
Akira Sasou
Johanan Sathyasingh
Antonio Satue Vilar
Michelina Savino
Oscar Saz
Thomas Schaaf
Felix Schaeffler
Robin Scheibler
Henning Schepek
David Schlangen
Ralf Schlüter
Sven Schmeier
Maximilian Schmitt
Jean Schoentgen
BJörn Schuller
Tanja Schultz
Barbara Schuppler
Petr Schwarz
James M Scobie
Jilt Sebastian
João Sedoc
Chandra Sekhar Seelamantula
Yael Segal
Encarna Segarra
Jose Segura
Ethan Selfridge
Gregory Sell
Michael Seltzer
Gregory Senay
Romain Serizel
Joan Serrà
Christophe Servan
Guruprasad Seshadri
Shreyas Seshadri
Vidhyasaharan Sethu
Anna Sfakianaki
Turaj Shabestary
Nirmesh Shah
Milind Shah
Mostafa Shahin
Syed Shahnawazuddin
Mahaboob Ali Basha Shaik
Ravi Shankar
Matt Shannon
Kalluri Shareef Babu
Stefanie Shattuck-Hufnagel
Slava Shechtman
Imran Sheikh
Peng Shen
Wade Shen
Sven Shepstone
Elizabeth Sherly
Chuang Shi
Hwan Shim
Tetsuya Shimamura
Hiroshi Shimodaira
Koichi Shinoda
Yasuaki Shinohara
Takahiro Shinozaki
Sayaka Shiota
Suwon Shon
Shijing Si
Ingo Siegert
Stephan Sigg
Siddharth Sigta
Vered Silber-Varod
Khe Chai Sim
Juraj Šimko
Flavio Simoes
Konstantin Simonchik
Simone Simonetti
Elliot Singer
Mittul Singh
Madhusudan Singh
Sabato Marco Siniscalchi
Olivier Siohan
Sunayana Sitaram
Dinkar Sitaram
Ganesh Sivaraman
Sunit Sivasankaran
Gabriel Skantze
Jan Skoglund
Pavel Skrelin
Malcolm Slaney
Kamel Smaili
Elena Sokolova
Maria Josep Solé
Rubén Solera-Ureña
Hagen Soltau
Peng Song
Eunwoo Song
Meet Soni
Mee Sonu
Alexander Sorin
Victor Sorokin
Laura Spinu
Richard Sproat
Stefano Squartini
Ajay Srinivasamurthy
Sundararajan Srinivasan
Sridhar Srinivasan
Anuroop Sriman
Siddharth Srivastava
Themis Stafylakis
Adriana Stan
tiomkin stas
Sabrina Stehwien
Peter Steiner
Georg Stemmer
Evgeny Stepanov
George Sterpu
Hywel Stoakes
INTERSPEECH 2022 HUMAN AND HUMANIZING SPEECH TECHNOLOGY
<table>
<thead>
<tr>
<th>Name</th>
<th>Name</th>
<th>Name</th>
<th>Name</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eugene Weinstein</td>
<td>Wei Xia</td>
<td>Hani Camille Yehia</td>
<td>Zixing Zhang</td>
<td></td>
</tr>
<tr>
<td>Melanie Weirich</td>
<td>Bo Xia</td>
<td>Serdar Yildirim</td>
<td>Chi (Leo) Zhang</td>
<td></td>
</tr>
<tr>
<td>Xue Wen</td>
<td>Xiong Xiao</td>
<td>Emre Yilmaz</td>
<td>Pengyuan Zhang</td>
<td></td>
</tr>
<tr>
<td>Liang Wen</td>
<td>Xurong Xie</td>
<td>Selma Yılmazıldız Kayaarma</td>
<td>Frank Zhang</td>
<td></td>
</tr>
<tr>
<td>Zhengqi Wen</td>
<td>Lei Xie</td>
<td>Sevinj Yolchuyeva</td>
<td>Hui Zhang</td>
<td></td>
</tr>
<tr>
<td>Chao Weng</td>
<td>Feng-Long Xie</td>
<td>Nestor Becerra Yoma</td>
<td>Shi-Xiong Zhang</td>
<td></td>
</tr>
<tr>
<td>Stefan Werner</td>
<td>Yanlu Xie</td>
<td>Minjoon Yoo</td>
<td>Tao Zhang</td>
<td></td>
</tr>
<tr>
<td>Allison Wetterlin</td>
<td>Feifei Xiong</td>
<td>Chang Yoo</td>
<td>Xiao-Lei Zhang</td>
<td></td>
</tr>
<tr>
<td>Gordon Wichern</td>
<td>Haihua Xu</td>
<td>Dongsuk Yook</td>
<td>Shiliang Zhang</td>
<td></td>
</tr>
<tr>
<td>Lorin Wilde</td>
<td>Yi Xu</td>
<td>Su-Youn Yoon</td>
<td>Wei Zhang</td>
<td></td>
</tr>
<tr>
<td>Daniel Willett</td>
<td>Minqiang Xu</td>
<td>Tsukasa Yoshinaga</td>
<td>Shuai Zhang</td>
<td></td>
</tr>
<tr>
<td>Donald Williamson</td>
<td>Jiaming Xu</td>
<td>Koichiro Yoshino</td>
<td>Xiaohui Zhang</td>
<td></td>
</tr>
<tr>
<td>Caroline Wiltshire</td>
<td>Hainan Xu</td>
<td>Midia Yousefi</td>
<td>Jinsong Zhang</td>
<td></td>
</tr>
<tr>
<td>Jonathan Wintrode</td>
<td>Mingxing Xu</td>
<td>Ha-Jin Yu</td>
<td>Guanlong Zhao</td>
<td></td>
</tr>
<tr>
<td>Marcin Wlodarczak</td>
<td>Chenglin Xu</td>
<td>Kai Yu</td>
<td>Liang Zhao</td>
<td></td>
</tr>
<tr>
<td>Matthias Wolff</td>
<td>Kele Xu</td>
<td>Dong Yu</td>
<td>Yuanjun Zhao</td>
<td></td>
</tr>
<tr>
<td>Krzysztof Wolk</td>
<td>Yong Xu</td>
<td>Yougen Yuan</td>
<td>Yunxin Zhao</td>
<td></td>
</tr>
<tr>
<td>Maria Wolters</td>
<td>Wei Xue</td>
<td>Jiahong Yuan</td>
<td>Kai Zhen</td>
<td></td>
</tr>
<tr>
<td>Minz Won</td>
<td>Shaofei Xue</td>
<td>Young-Sun Yun</td>
<td>Thomas Fang Zheng</td>
<td></td>
</tr>
<tr>
<td>Jeremy Heng Meng Wong</td>
<td>Nelson Yalta</td>
<td>François Yvon</td>
<td>Xinhui Zhou</td>
<td></td>
</tr>
<tr>
<td>Phil Woodland</td>
<td>Junichi Yamagishi</td>
<td>Zbynek Zajic</td>
<td>Yi Zhou</td>
<td></td>
</tr>
<tr>
<td>Jane Wottawa</td>
<td>Kazumasa Yamamoto</td>
<td>Alexander Zatvornitskiy</td>
<td>Nina Zhou</td>
<td></td>
</tr>
<tr>
<td>Szymon Woźniak</td>
<td>Hitoshi Yamamoto</td>
<td>Hossein Zeinali</td>
<td>Xuan Zhu</td>
<td></td>
</tr>
<tr>
<td>Zhiyong Wu</td>
<td>Nan Yan</td>
<td>Piotr Żelasko</td>
<td>Jian Zhu</td>
<td></td>
</tr>
<tr>
<td>Ji Wu</td>
<td>Xuesong Yang</td>
<td>Margaret Zellers</td>
<td>Xiaodan Zhuang</td>
<td></td>
</tr>
<tr>
<td>Chunyang Wu</td>
<td>Shan Yang</td>
<td>Biao Zeng</td>
<td>Bartosz Ziołko</td>
<td></td>
</tr>
<tr>
<td>Yaru Wu</td>
<td>Longfei Yang</td>
<td>Andrej Zgank</td>
<td>Imed Zitouni</td>
<td></td>
</tr>
<tr>
<td>Bo Wu</td>
<td>Yike Yang</td>
<td>Huayun Zhang</td>
<td>Hatice Zora</td>
<td></td>
</tr>
<tr>
<td>Zhizheng Wu</td>
<td>Zhaojun Yang</td>
<td>Chong Zhang</td>
<td>Catalin Zorila</td>
<td></td>
</tr>
<tr>
<td>Chung-Hsien Wu</td>
<td>Seung Yang</td>
<td>Mingyang Zhang</td>
<td>Yuexian Zou</td>
<td></td>
</tr>
<tr>
<td>Xixin Wu</td>
<td>Umit Yapanel</td>
<td>Chunlei Zhang</td>
<td>Enrico Zovato</td>
<td></td>
</tr>
<tr>
<td>Mengyue Wu</td>
<td>Chiranjeevi Yarra</td>
<td>Chao Zhang</td>
<td>Geoffrey Zweig</td>
<td></td>
</tr>
<tr>
<td>junru Wu</td>
<td>Keiichi Yasu</td>
<td>Zhengchen Zhang</td>
<td>Marzena Zygis</td>
<td></td>
</tr>
<tr>
<td>Chai Wutiwiwatchai</td>
<td>Guoli Ye</td>
<td>Wei-Qiang Zhang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hilary Wynne</td>
<td>Bayya Yegnanarayana</td>
<td>Xueliang Zhang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xianjun Xia</td>
<td>Raymond Yeh</td>
<td>Zhaoyan Zhang</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Save the Date and prepare to experience the Irish Céad Míle Fáilte – A Hundred Thousand Welcomes!

Sign up for regular e-updates about

INTER SPEECH 2023

20th-24th August 2023 - Dublin, Ireland

at www.interspeech2023.org
INTERSPEECH 2025
SPEECH AND THE INDIVIDUAL

Location: Rotterdam, The Netherlands

General Chair: Odette Scharenborg (TU Delft)

General Co-chairs: Khiet Truong (UTwente) & Catharine Oertel (TU Delft)

Technical Program Committee:
Esther Klabbers (chair; ReadSpeaker, US), Martin Cooke (IKERBASQUE, Spain) & Tomi Kinnunen (University of Eastern Finland)
Enabling Human Creativity

Explore a world beyond academic benchmarks.

LOVO believes technology truly shines when it creates real-life impact. Power millions of creators around the globe with unique Voice AI research.

Learn more at lovo.ai
Research Topics at LOVO

Controllable, Expressive TTS
Speaker Generation
Singing Synthesis
Voice Conversion
Speech Enhancement
Blockchain x AI

Backed by the best

#HASHED  kakao  kakao investment  LG CNS  NVIDIA
HUSTLE FUND  PRIMER SAZZE  Berkeley SKYDECK
PKO  YAT SIU  PATRICK LEE  HOLLY LIU  KUN GAO
Google for Startups Residency 2020  aws activate
At Amazon, we fundamentally believe that scientific innovation is essential to being the most customer-centric company in the world. It's the company's ability to have an impact at scale that allows us to attract some of the brightest minds in artificial intelligence, machine learning, and related fields.

Learn more at amazon.science
Hyundai Motor Company is taking a step forward in becoming a smart mobility solution provider company by utilizing various AI/SW technologies such as spoken language processing.

Learn more about HMC’s Technology and Recruiting in QR below
LG AI Research

Advancing AI for a better Life

Learn more about LG AI Research at www.lgresearch.ai
Kakao Enterprise brings AI into everyday work and life for a better future for all.

Kakao Enterprise AI Lab & Service is Kakao’s largest AI R&D organization. It provides technologies and services that innovate businesses based on its advanced research expertise. We have core AI technologies in a variety of areas including speech processing, computer vision, natural language processing, and time series prediction.

Speech Processing  Speech-to-Text  Text-to-Speech  Speaker Verification  Speaker Diarization
Speech Enhancement  Keyword Spotting  English Pronunciation Assessment
Computer Vision  NLP  Time Series Prediction

Winner of Global Challenges!
- MS MARCO Passage Ranking 1st
- FRVT 1:N Investigation Visa-Kiosk 1st
- FRVT 1:1 Verification Kiosk 1st
- WMT21 Machine Translation using Terminologies Task 1st
- KorQuAD 2.0 1st

Come grow with us
Shape the Future with Innovation and Intelligence

Samsung Research is the advanced R&D hub of Samsung’s Device eXperience (DX) Division to prepare the future of Samsung Electronics.

We lead the development of the future technologies with about 10,000 researchers and developers working in overseas R&D centers.

Under the vision of “Shape the Future with Innovation and Intelligence”, Samsung Research is actively conducting research and development to identify new future growth areas and secure advanced technologies to create new value and improve people’s lives.

Speech related research topics
- End-to-end speech recognition
- End-to-end speech synthesis
- Wake-up word detection & speaker recognition
- Speech/sound enhancement
- Sound Recognition
- Multi-modal speech recognition
- Spoken language understanding
- Speech translation

<table>
<thead>
<tr>
<th>Intelligence</th>
<th>Device Solution</th>
<th>Network &amp; SW</th>
</tr>
</thead>
<tbody>
<tr>
<td>🌐 Artificial Intelligence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>📈 Data Intelligence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>🧠 New Experience Innovation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>🤖 Robot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>📱 Next Generation Display &amp; Media</td>
<td></td>
<td></td>
</tr>
<tr>
<td>🛠 Intelligent Device &amp; Solution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>📡 Next Generation Communications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>🔒 Security &amp; Privacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>🌐 Intelligent SW Platform</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

https://research.samsung.com   |   @samsungresearch
Supertone is an AI audio tech startup whose mission is to break the limitations in content creation. We are passionate about innovating the content production landscape through our state-of-the-art technology such as expressive singing/speech synthesis, original voice design, and speech enhancement.

SINGING VOICE SYNTHESIS
TEXT-TO-SPEECH SYNTHESIS
GRAPHEME-TO-PHONEME
MELODY/LYRICS TRANSCRIPTION
STUDIO-QUALITY SPEECH ENHANCEMENT

VOICE CONVERSION
NATURAL LANGUAGE PROCESSING
SPEAKER VERIFICATION
AUTOMATIC SPEECH RECOGNITION

Learn more about Supertone at supertone.ai
Google is looking forward to connecting with the community at Interspeech 2022! We welcome you to stop by the Google booth during the conference to network with members of our research and staffing teams. You can also scan the below QR code to share your details with us for future Google events and career opportunities.
NAVER CLOVA

NAVER CLOVA, the artificial intelligence (AI) division of NAVER Corp., is introducing AI-enabled products and services to make our lives more convenient and comfortable. CLOVA is proving its technological competitiveness and outstanding capabilities in the field of AI research by making a new record every year for publishing academic papers.

Research Topics of CLOVA
- Computer vision
- NLP/Hyperscale LM
- Speech recognition
- Voice synthesis
- User behavior modeling
- Logistics optimization
- Auto ML
- HCI
- Healthcare AI
- Reinforcement learning

Learn more about NAVER Corp.
Full publication list of CLOVA
AI-powered virtual actors for everyone

Typecast uses artificial intelligence and deep learning technology to create characters with natural voices and various styles. Use Typecast to cast AI virtual actors for your video and audio contents.

Research topic

Speech synthesis
- Controllable / Expressive / Multi-lingual speech synthesis
- Singing voice synthesis
- Few-shot speech synthesis
- Voice conversion

Natural language processing
- Dialogue / text generation
- Sentence embedding

Talking head generation
- Emotion-controllable / One-shot / Few-shot talking head generation
- Talking head animation generation (2D/3D)