Closure Report

On

ISCA Supported

Summer School
on
Speech Signal Processing

Speaker Recognition and Diarization

6-10 July 2019

Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT)
Gandhinagar, India.

URL: https://sites.google.com/site/s4p2019/
Message from Organizing Chair

On behalf of the Organizing Committee, I record our appreciation for the valuable contribution made by eminent invited speakers, participants, international program committee, DA-IICT faculty colleagues, staff, and student volunteers towards conducting the 5th edition of summer school with the theme ‘Speaker Recognition and Diarization’ during July 6-10, 2019 at DA-IICT Gandhinagar, India. This summer school gave a platform to interact with distinguished invited speakers, to discover novel methods and broaden our knowledge in the broad area of voice biometrics. Furthermore, to encourage young talent, the school presented fourth edition of 5 Minute Ph.D. Thesis (5MPT) contest with four ISCA endorsed cash prizes and four DE GRUYTER sponsored book prizes.

We were honoured that we had Prof. (Dr.) J. F. Bonastre (Ex. ISCA President, University of Avignon, France), Prof. (Dr.) Nicholas Evans (EURECOM, France), Prof. (Dr.) Thomas. F. Zheng (Tsinghua University, China), Prof. (Dr.) Dong Wang (Tsinghua University, China). In addition, we have Prof. (Dr.) B. Yegnanarayana (IEEE Fellow, ISCA Fellow, Retd. IIT Madras), Prof. (Dr.) Hema A. Murthy (ISCA Board Member, IIT Madras), Prof. (Dr.) T. K. Basu (Retd. IIT Kharagpur), Prof. (Dr.) S. R. M. Prasanna (IIT Dharwad), Prof. (Dr.) K. Samudravijaya (IIT Guwahati), Prof. (Dr.) Suryakant Gangashetty (IIT Hyderabad), and Prof. (Dr.) K. S. R. Murty (IIT Hyderabad). At the Summer School, motivated from INTERSPEECH 2018, we have organized Industry Perspective Talks in which senior industry personnel namely Dr. Vikram Vij (Samsung R&D Institute, Bangalore), Dr. Sunil Kumar Kopparapu (TCS Innovation Lab, Mumbai), Dr. Sunayana Sitaram (Microsoft Research, Bangalore), Mrs. Swaran Lata (Retd. MeitY, Delhi), and Mr. Pranaw Kumar (C-DAC, Mumbai).

Events of this kind cannot happen without generous financial support from sponsors. In this regard, we express our deep gratitude and appreciation to the sponsors, namely, DA-IICT Gandhinagar, Reliance Communications, International Speech Communications Association (ISCA), and DE GRUYTER. We also thank our technical co-sponsors, namely, IEEE SPS Gujarat Section Chapter, and APSIPA (in particular APSIPA DL by Prof. Dong Wang and Prof. Hemant A. Patil) without which organizing this event would not have been possible. In addition, we would like to thank Prof. Gerard Bailly (GIPSA Lab, CNRS, France) and Prof. Phil Green (University of Sheffield, UK) for their valuable feedback on our proposal for possible ISCA support to S4P 2019.

This summer school witnessed 65 attendees including researchers, industry personnel, faculty members, and students from all over world. We would like to sincerely acknowledge kind support by Dr. Vikram Vij, who recommended 07 speech scientist of ASR team from Voice Intelligence Group of Samsung R&D Institute, Bangalore. Furthermore, we acknowledge excellent support from ISRO Ahmedabad, TCS Innovation Lab and few startup companies who sponsored their employees to attend this event. The members of Organizing Committee hope that the participants and invited speakers had memorable experience and pleasant stay at Gandhinagar and hope that you will continue to visit DA-IICT in future and participate in such ISCA supported events.

Prof. (Dr.) Hemant A. Patil (ISCA Member)  
DA-IICT Gandhinagar  
ISCA Distinguished Lecturer for 2020-2021  
APSIPA Distinguished Lecturer for 2018-2019  

Prof. (Dr.) Hemant A. Patil (ISCA Member)  
DA-IICT Gandhinagar  
ISCA Distinguished Lecturer for 2020-2021  
APSIPA Distinguished Lecturer for 2018-2019
(a) **Thomas Fang Zheng**, Tsinghua University, China.

(b) **Jean-Francois Bonastre**, University of Avignon, France.

(c) **Nicholas Evans**, EURECOM, France.

(d) **Dong Wang**, Tsinghua University, China.

(e) **Bayya Yegnanarayana**, Retd., Indian Institute of Technology (IIT), Madras, India.

(f) **Tapan Kumar Basu**, Retd. Professor, Indian Institute of Technology (IIT) Kharagpur, India.

(g) **Hema A. Murthy**, Indian Institute of Technology (IIT), Madras, India.

(h) **S. R. M Prasanna**, Indian Institute of Technology (IIT), Dharwad, India.

(i) **K. Samudravijaya**, Indian Institute of Technology (IIT), Guwahati, India.

(j) **K. S. R. Murty**, Indian Institute of Technology (IIT), Hyderabad, India.

(k) **Suryakant Gangashetty**, International Institute of Information Technology (IIIT) Hyderabad, India.

(l) **Hemant A. Patil**, DA-IICT Gandhinagar, India.
Industry Perspective Talks

(a) Vikram Vij, Senior Vice President, Samsung Electronics Co. Ltd., Bangalore, India.
(b) Swaran Lata, Ex-Senior Director and Head, TDIL Programme, MeitY, Govt. of India.
(c) Sunil Kumar Kopparapu, Principal Scientist, TCS Research and Innovation Lab, Mumbai, India.
(d) Sunayana Sitaram, Senior Applied Scientist, Microsoft Research Laboratory, Bangalore, India.
(e) Pranaw Kumar, Senior Technical Officer, C-DAC, Mumbai, India.
The 5th edition of Summer School on Speech Signal Processing was held at Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar, India. S4P 2019 was a 5-day event during 6-10 July 2019. S4P 2019 targeted mainly the postgraduate students, college teachers, faculty in educational institutions, and scientists/researchers in research laboratories/industry who were interested to update their knowledge in this area from some of the best researchers in speech signal processing across India and abroad.
FOCUS Area: Speaker Recognition and Diarization

Speaker recognition (voice biometrics) has gained a lot of interests in a large number of e-commerce related applications, such as speaker forensics, banking transactions and of late in smartphones. Discussion related to the development of speaker recognition systems which are robust to spoofing, noise, channel variability, intrinsic variability, etc., is the major goal of S4P 2019. In addition, the recent research work related to the development of speaker diarization and countermeasures against spoofing and tampering attacks will also be covered in this summer school. This summer school will be highly benefited by invited talks from eminent and world-class researchers from academia, industry, and research laboratories from India and abroad. In addition, S4P 2019 is highly benefited by world renowned program committee (PC) members. Furthermore, to encourage young talent, the school presents a fourth edition of 5 Minutes Ph.D Thesis Contest (5MPT) with four ISCA endorsed cash prizes. Finally, for the first time, motivated from INTERSPEECH 2018, S4P 2019 introduced industry perspective talks from major speech technology industries in India. The aim of S4P 2019 was to further advance our understanding of these technological challenges (and open) research issues.
Announcement and Publicity of S4P 2019

Announcement for S4P 2019 was made worldwide through the following:

1) **S4P 2019 Poster Distribution:**
   - ISCA Speech Labs across the world
   - ICASSP 2019, Brighton, UK
   - WiSSAP 2019, Thiruvananathapuram, Kerala, India
   - Speech Research Lab in China (Thanks to Prof. Thomas F. Zheng in this great help)
   - 1000 potential colleges, institutes (IISc, IIT’s, NIT’s, etc.), R&D Labs, industry across India.

2) **S4P 2019 website:**
   - URL: [https://sites.google.com/site/s4p2019/](https://sites.google.com/site/s4p2019/)

**Participants**

Around 65 participants from India have attended S4P 2019. The brief detail of the participants is shown in Figure 1.

![Figure 1](image1.png)

*Figure 1: The distribution of the participants that attended S4P 2019 at DA-IICT Gandhinagar.*

The participants represented 31 institutes/colleges/universities across India. The distribution of the participants is given in Figure 2.

![Figure 2](image2.png)

*Figure 2: The institute-wise distribution of the participants who attended S4P 2019 at DA-IICT Gandhinagar.*
Patron

- K. S. Dasgupta, Director, DA-IICT Gandhinagar.

Program Committee (PC) Members

S4P 2019 was organized with the guidance and support of various speech researchers across the country and abroad. The Program Committee (PC) members of S4P 2019 who guided and supported in planning and designing the technical program and other related tasks are:

- Douglas O'Shaughnessy, McGill University, Canada
- Sadaoki Furui, Toyota Technological Institute (TTI), Chicago, USA
- Douglas Reylonds, MIT Lincoln Lab, USA
- Pedro Torres-Carrasquillo, MIT Lincoln Lab, USA
- Hideki Kawahara, Wakayama University, Japan
- Odette Scharenborg, Delft University of Technology, The Netherlands
- Torbjorn Svendsen, Norwegian University of Science and Technology, Norway
- Jiri Navratil, IBM Thomas J. Watson Research Center, NY, USA
- Yannis Stylianou, University of Crete, Greece
- Tomi H. Kinnunen, University of Eastern Finland, Finland
- Junichi Yamagishi, National Institute of Informatics, Japan
- Eduardo Lleida, University of Zaragoza, Spain
- Luis Javier Rodriguez-Fuentes, University of Basque Country, Euskal Herriko Unibertsitatea, Spain
- Lori Lamel, LIMSI, France
- Saikat Chatterjee, KTH Royal Institute of Technology, Stockholm
- Julia Hirschberg, Columbia University, New York, USA
- Kong Aik Lee, NEC Corporation, Japan
- Amy Neustein, Linguistic Technology Systems, New Jersey, USA
- Tanja Schultz, University of Bremen, Germany
- Dong Ming Hui, Institute for Infocomm Research (I2R), A*STAR, Singapore
- Eliathamby Ambikairajah, University of New South Wales, Sydney, Australia
- Dong Wang, Tsinghua University, China
- Tara Sainath, AI Google, USA
- Tatsuya Kawahara, Kyoto University, Japan
- Jianwu Wang, JAIST, Japan
- Csapó Tamás Gábor, Budapest University of Technology and Economics, Hungary
- Bernd Möbius, Saarland University, Germany
- Pascal Perrier, GIPSA-Lab, Grenoble-INP, France
- Daryush Mehta, Center of Laryngeal and Voice Rehabilitation, Harvard University, Boston

Organizing Committee, DA-IICT Gandhinagar

- Hemant A. Patil, Chair
- Suman K. Mitra, Convener
- Yash Vasavada
- Sanjeev Gupta
- Srimanta Mandal
- Soman Nair
Sponsors
The Organizing Committee would like to thank the following sponsors for the generous support extended to conduct S4P 2019:
- Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar
- Reliance Communications Limited
- International Speech Communication Association (ISCA)
- DE GRUYTER

Technical Co-Sponsors
- IEEE Signal Processing Society (SPS) Gujarat Section Chapter
- Asia-Pacific Signal and Information Processing Association (APSIPA)

Banner of S4P 2019

Summer School on Speech Signal Processing (S4P)
Speech Recognition and Diarization
06-10 July 2019

Sponsors
- DA-IICT
- RELIANCE
- ISCA
- DE GRUYTER

Technical Co-Sponsors
- SIGNAL PROCESSING SOCIETY
- APSIPA
Inauguration Ceremony:

The summer school began with an inauguration ceremony, which included prayer to the Almighty, followed by lighting of the lamps by Prof. (Dr.) K. S. Dasgupta (Director, DA-IICT Gandhinagar), Prof. (Dr.) Nicholas Evans, Prof. (Dr.) J. F. Bonastre, Prof. (Dr.) B. Yegnanarayana, Prof. (Dr.) Suryakant V. Gangashetty, and Mrs. Swaran Lata.
Inaugural Address By Prof. K. S. Dasgupta, Director, DA-IICT Gandhinagar

Welcome Address by Prof. Hemant A. Patil, Organizing Chair, S4P 2019

Prof. Patil appreciated all the sponsors of S4P 2019. In addition, he also appreciated DE GRUYTER sponsored book prizes for the winner of 5MPT contest.
# Design of Technical Program and Schedule

<table>
<thead>
<tr>
<th>Time Slot</th>
<th>6th July (Saturday)</th>
<th>7th July (Sunday)</th>
<th>8th July (Monday)</th>
<th>9th July (Tuesday)</th>
<th>10th July (Wednesday)</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00-08:30</td>
<td>Registration and Inauguration</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>08:30-10:00</td>
<td>B. Yegnanarayana (L1)</td>
<td>J. F. Bonastre (L6)</td>
<td>Hema A. Murthy (L11)</td>
<td>T. F. Zheng (L15)</td>
<td>J. F. Bonastre (L20)</td>
</tr>
<tr>
<td>10:00-10:30</td>
<td>Tea Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:30–12:00</td>
<td>T. K. Basu (L2)</td>
<td>K. S. R. Murty (L7)</td>
<td>Hemant A. Patil (L12)</td>
<td>Vikram Vij (L16)</td>
<td>N. Evans (L21)</td>
</tr>
<tr>
<td>12:00–13:30</td>
<td>Lunch Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:30–15:00</td>
<td>J. F. Bonastre (L3)</td>
<td>N. Evans (L8)</td>
<td>T. F. Zheng (L13)</td>
<td>Sponsors Presentation (S1)</td>
<td>Swaran Lata (L17)</td>
</tr>
<tr>
<td>15:00-15:30</td>
<td>Tea Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:30–17:00</td>
<td>N. Evans (L4)</td>
<td>S. R. M. Prasanna (L9)</td>
<td>Pranaw Kumar (L14)</td>
<td>S. V. Gangashetty (L18)</td>
<td>Award Ceremony &amp; Valedictory Function</td>
</tr>
<tr>
<td>17:00–17:30</td>
<td>Tea Break</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:30-19:00</td>
<td>K. Samudravijaya (L5)</td>
<td>Dong Wang (L10)</td>
<td>5 Min Ph. D. Thesis’ (5 MPT)</td>
<td>S. Sitaram (L19)</td>
<td>-</td>
</tr>
</tbody>
</table>

*Selected Ph.D. scholars will present their doctoral research work within 5 minutes.*
## Technical Content of S4P 2019

<table>
<thead>
<tr>
<th>SPEAKERS</th>
<th>LECTURE TOPICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Extraction of Hidden Features in Speech Signals</td>
</tr>
<tr>
<td>L2</td>
<td>Signal Processing Aspects of Speech</td>
</tr>
<tr>
<td>L3</td>
<td>Speaker Recognition: Advances and Limits</td>
</tr>
<tr>
<td>L4</td>
<td>Speaker Recognition: A Reliable, Trustworthy Biometric?</td>
</tr>
<tr>
<td>L5</td>
<td>Speaker Recognition using Gaussian Mixture Model</td>
</tr>
<tr>
<td>L6</td>
<td>Forensic Voice Characterization</td>
</tr>
<tr>
<td>L7</td>
<td>Speaker Embedding Extraction from Virtual Phonetic Information</td>
</tr>
<tr>
<td>L8</td>
<td>ASVSPoof: The Automatic Speaker Verification Spoofing and Countermeasures Challenge Initiative</td>
</tr>
<tr>
<td>L9</td>
<td>Spoof Detection using Excitation Source Information</td>
</tr>
<tr>
<td>L10</td>
<td>Deep Feature Learning and Normalization in Speaker Recognition</td>
</tr>
<tr>
<td>L11</td>
<td>A) Variable Top-C and Feature Switching - Speaker Recognition, Anti Spoofing B) Variable IB and Speaker Diarization</td>
</tr>
<tr>
<td>L12</td>
<td>Generative Adversarial Networks (GANs) for Speaker Recognition</td>
</tr>
<tr>
<td>L13</td>
<td>The On-Going Plan of Trusted Identity Authentication in China and the Research Challenges of Speaker Recognition (Part I)</td>
</tr>
<tr>
<td>L14</td>
<td>Development Journey of Indian Language ASR, TTS and Related Applications</td>
</tr>
<tr>
<td>L15</td>
<td>The On-Going Plan of Trusted Identity Authentication in China and the Research Challenges of Speaker Recognition (Part II)</td>
</tr>
<tr>
<td>L16</td>
<td>Large Scale ASR for Multi Device Ecosystem</td>
</tr>
<tr>
<td>L17</td>
<td>The Research Challenges of Speaker &amp; Speech Recognition for Punjabi Language</td>
</tr>
<tr>
<td>L18</td>
<td>Voice Conversion Approaches for Spoofing Attacks</td>
</tr>
<tr>
<td>L19</td>
<td>Speech Recognition and Speaker Diarization for Meeting Transcription</td>
</tr>
<tr>
<td>L20</td>
<td>Representation Learning for Voice Characterization</td>
</tr>
<tr>
<td>L21</td>
<td>Automatic Speaker Verification and Spoofing Countermeasures: Integration and Metrics</td>
</tr>
<tr>
<td>L22</td>
<td>Pathological Speech Processing</td>
</tr>
</tbody>
</table>

### Sponsors Presentation

<table>
<thead>
<tr>
<th>S1</th>
<th>Prof. Sanjeev Gupta, Dean (R&amp;D), DA-IICT Gandhinagar, India.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2</td>
<td>Prof. Thomas F. Zheng, Tsinghua University, China. (Vice President, Conferences APSIPA Board)</td>
</tr>
</tbody>
</table>
Invited Talks

Prof. (Dr.) B. Yegnanarayana

*Extraction of Hidden Features in Speech Signals*

The objective of this talk was to highlight features of speech production which can be perceived, but cannot be extracted easily using the standard short-time analysis methods, due to dynamic nature of these features. Speaker call these features as hidden features in the speech signal. Some new signal processing methods which can bring out these dynamic features were discussed. In particular, the features reflecting the continuously changing dynamic vocal tract system due to glottal vibration and vowel nasalization were addressed in detail. There are many hidden features which human perception uses in various tasks, as in speaker recognition and emotion recognition. But it is a challenge to identify features specific for these tasks. That is why nonlinear models like artificial neural networks are usually explored for these tasks by posing them as classification problems, rather than as feature extraction problem. But representation of the speech signal information is a major challenge for developing these applications. (Most of this material is based on the PhD thesis of speaker’s recently graduated student, Dr. RaviShankar Prasad.)

Prof. (Dr.) Tapan Kumar Basu

*Signal Processing Aspects of the Speech*

The subject of Digital Signal Processing (DSP) has evolved out of the research in various areas of Mathematics and Electrical Engineering. At the end of the nineteenth century, with the development of electrical generators and power transmission systems, the subject of network theory, analysis, and synthesis techniques became the frontier of electrical science. New mathematical concepts were developed to explain the behaviour of electrical networks (positive real functions, lattice, and ladder network synthesis, minimum functions, synthesis from parts of network functions, etc.). Similar relations were later on developed for magnitude and phase relations following Hilbert transform. The uses of Laplace transform and Fourier series in solving problems of networks under various conditions were successfully exploited. Bode plot to study the frequency response of linear systems and using them for prediction of performance of a network was a major breakthrough in electrical engineering. The behaviour of travelling waves following a disturbance in a transmission system at a junction of two different lines or at a termination has been a subject of study for a long time. It has its analogous counterpart in speech wave transmission in a vocal tract. Filter design techniques with passive elements were developed for improving the performance of radio communication systems. It was also used by power engineers for improving power systems’ performance. Modern filter design is based on realising some approximate mathematical functions (Butterworth, Chebyshev, etc.). The design of Infinite Impulse Response (IIR) filters is based on such analog filters.
Subsequently, the subject of control systems and study of behaviour of different systems along the lines of network theories were successfully exploited. Study of behaviour of nonlinear systems very often required tedious long hand calculation using various numerical methods and need for development of some hardware devices was strongly felt. Initially, analog computers were used to carry out such studies of simple set of equations. The development of active elements like transistors, op-amps, Gyrator/NIC (negative impedance converter), and FDNR (Frequency Dependant Negative Resistance) paved the way for developing any network function with active elements.

While numerical methods and discrete mathematics attracted researchers for centuries, the steady development of VLSI technology in the post- WW-II period has been the mainstay for development of very powerful computers. Such a development required also a simultaneous growth of theory and the attendant softwares. The theory of discrete-time systems were now developed at a much faster pace to develop softwares for use in computers. The limitations of speed and memory were gradually circumvented by continuous progress in VLSI technology. DSP is the Brain of present day civilization and its two most important wings are Speech Processing and Image Processing. The branches of each of them are many. Today, we are living in a global village, where we would like to break the barriers of languages to communicate with each other with the tools of Speech Technology. Currently, we have started a small venture towards developing some tribal languages of India. This talk enriched all the attendees with different dimensions of Speech Processing field.

Prof. (Dr.) Jean-Francois Bonastre

*Speaker Recognition: Advances and Limits*

The progresses made in the field of speaker characterization during the past decades are impressive. In the 90's, speaker recognition was mainly limited to close set identification task, with a few dozen of speakers and read, clean speech. Nowadays, we are talking about text-independent, language-independent, environmental noise robust, smartphone-based speaker authentication practical solutions. Based on this practical observation, it is interesting to try to first understand what allowed these progresses and, second, if we really know more now on speaker recognition than 30 years ago. This talk by Prof. Bonastre first proposed a short view of speaker recognition over last 30 years history. Then, it questioned a little bit more deeply the noted progresses and showed the limits of current approaches in speaker recognition, including neural network (NN) - based approaches.
Forensic Voice Characterization (FVC)

Forensic Voice Characterization (FVC) is more often viewed as a subpart of speaker recognition field. But... Even if the links between the two topics are straightforward, FVC is a specific area, which tries to answer to different scientific questions than the classical speaker recognition. This second talk by Prof. Bonastre presented FVC and showed the differences with speaker recognition. Then, it questioned the latest speaker recognition technical improvements, like $i$-vectors and $x$-vectors, in this specific FVC context.

Representation Learning for Voice Characterization

The recent development of neural network (NN) - based approaches opens new avenue for new tasks in voice characterization area. In this talk, speaker presented the AI-assisted voice casting in the context of multimedia documents dubbing. This area is a good representative of a class of problems, where there is a strong human know-how demonstrated by a lot of examples of problem solutions but with few or no explicit information on the process. This third talk by Prof. Bonastre presented the voice casting problem. It helped to define the in-interest class of problems. Then, a neural network approach is first used to see if the problem resolution is feasible. Finally, a representation learning approach was presented.
Prof. (Dr.) Nicholas Evans

**Speaker Recognition: A Reliable, Trustworthy Biometric?**
First talk by Prof. Evans started by outlining the requirements for successful biometric characteristics and explores whether these can be met with the use of voice recordings for speaker recognition. The remainder of the talk focused upon one requirement, in particular, relating to the robustness of biometric systems to circumvention. Speaker discussed early work that explored security aspects of speaker recognition systems and more recent work which has begun to investigate the potential of the same technology to intrude on personal privacy. The final part of the talk touched upon recent regulation which places restrictions upon the extent to which voice recordings can be lawfully collected, stored, and processed.

**ASVspoof: The Automatic Speaker Verification Spoofing and Countermeasures Challenge Initiative**
Second talk by Prof. Evans expanded upon the issue of circumvention. It focuses exclusively on the threat of presentation attacks, better known in our field as spoofing. The talk described the making of the community-led, competitive challenges known as ASVspoof. It then presented the first edition of the ASVspoof challenge series held in 2015. After describing the RedDots Replayed dataset, the talk presented the second edition of ASVspoof held in 2017. A selection of the most successful anti-spoofing solutions and general trends were identified. The talk concluded with a discussion of the progress achieved through the first two editions of ASVspoof and the remaining challenges that they highlighted.
**Automatic Speaker Verification and Spoofing Countermeasures: Integration and Metrics**

Third talk by Prof. Evans focused on the integration of biometric systems with spoofing countermeasures. The first two editions of ASVspoof focused upon the assessment of countermeasures in isolation from automatic speaker verification systems; the talk explained the reasons behind this strategy. With spoofing countermeasures always operating alongside a biometric recogniser, however, a more meaningful approach to assessment should reflect the intertwined operation and impact of each separate component on the single, combined system as a whole. The talk discussed different solutions to the integration of automatic speaker verification and spoofing countermeasures and shows how the detection error cost function used for the assessment of automatic speaker verification systems can be extended to assess their performance when they are combined with spoofing countermeasures.

---

**Prof. (Dr.) K. Samudravijaya**

**Speaker Recognition using Gaussian Mixture Model**

An overview of text-independent speaker recognition using Gaussian Mixture Model (GMM) was presented. The talk began with an overview of speaker recognition, recapitulation of computation of popular feature vector, mel scale cepstral coefficients. Then, GMMs was introduced as model of speaker characteristics. Estimation of parameters of GMMs using Expectation Maximization (EM) algorithm was covered along with an illustrative computation. The usage of GMMs for speaker recognition, feature normalization, and model adaptation techniques, evaluation of speaker recognition systems was illustrated later.
Prof. (Dr.) K. Sri Rama Murty

**Speaker Embedding Extraction from Virtual Phonetic Information**

In the recent past, deep neural networks have been successfully employed to extract fixed-dimensional speaker embeddings from the speech signal. The commonly used x-vectors are extracted by projecting the magnitude spectral features extracted from the speech signal onto a speaker-discriminative space. As the x-vectors do not explicitly capture the speaker-specific phonological pronunciation variability, phonetic vectors extracted from an automatic speech recognition (ASR) engine were supplied as auxiliary information to improve the performance of the x-vector system. However, the development of ASR engine requires a huge amount of manually transcribed speech data. In this work, Prof. Murty proposed to transcribe the speech signal in an unsupervised manner with the cluster labels obtained from a mixture of autoencoders (MoA) trained on a large amount of speech data. The unsupervised labels, referred to as virtual phonetic transcriptions, are used to extract the phonetic vectors. The virtual phonetic vectors extracted using MoA are supplied as auxiliary information to the x-vector system.

Prof. (Dr.) S. R. M. Prasanna

**Spoof Detection using Excitation Source Information**

There are several attempts in the literature for spoof detection. Most of these have innovations at two levels, namely, features that rely on vocal tract information and modeling techniques based on latest trends in machine learning and deep learning. This talk by Prof. Prasanna explained some of the attempts made in exploiting the excitation source information. The hypothesis is that the replay attack condition modifies the speech signal characteristics and the same may be characterized using the excitation source information. The outcomes of these findings are encouraging and demonstrate their practical utility.
Prof. (Dr.) Dong Wang

Deep Feature Learning and Normalization in Speaker Recognition
(APSIPA Distinguished Lecture)

Deep learning has gained significant interest in speaker recognition. Two key concepts developed in DNN-based speaker recognition are deep feature learning and deep speaker embedding. While the former is frame-based representation with the aim of discovering the ‘essential features’ from compositional speech signals, the latter focuses on segment-based representations, targeting to the speaker recognition task directly. This APSIPA DL talk by Prof. Wang compared these two approaches and summarized some interesting applications of each. Moreover, both deep features and deep embeddings suffer from an unregularized distribution, attributed to the lack of prior in their distributions. Talk described a normalization technique based on the variational inference, which introduces a prior of diagonal Gaussian and so produces more regularized features and embeddings.

Prof. (Dr.) Hema A. Murthy

Variable Top-C and Feature Switching -- Speaker Recognition, Antispoofing & Variable IB and Speaker Diarization.

Speaker verification/spoof detection is a two-class problem. Conventionally, UBM GMM/i-vector systems are used. Scoring in the UBM-GMM framework uses top-C most contributing mixtures for each class. In this talk, Prof. Murthy focused on a variable top-C approach to scoring and then extended this approach to the i-vector framework for speaker verification. Combined with “Feature Switching”, this approach yields a significant improvement over state-of-the-art in both the tasks. Speaker Diarization is the task of identifying “who spoke when?” Bottom up
clustering where the utterance is split into fixed length segments is standard. Speaker showed that “information rate” is a better measure for initial clustering, in that it leads to much lesser diarization error rates.

Prof. (Dr.) Hemant A. Patil

Generative Adversarial Networks (GANs) for Speaker Recognition
(APSIPA Distinguished Lecture)

Adversarial training or Generative Adversarial Networks (GANs) is the most interesting and technologically challenging idea (pioneered by I. J. Goodfellow in 2014) in the field of machine learning. GAN is a recent framework for estimating generative models via the adversarial training mechanism in which we simultaneously train two models, namely, a generator G that captures the (true) data distribution and a discriminator model D that estimate the probability that a sample came from training data rather than G. The training procedure of GANs (which is challenging w.r.t. convergence, mode collapse, etc.) for G is to maximize the probability of D making mistake. This framework corresponds to a mini-max two-player game (such as thief-Police game!). In the function space of arbitrary differentiable functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to ½ everywhere (that D is fooled by generator). When G and D are defined by multilayer perceptron, the entire system can be trained with backpropagation.

GANs are widely used in various applications (first used in image processing and computer vision and recently in speech areas). In particular, image (sample) generation, single image superresolution, text-to-image synthesis, and several speech technology applications (mostly after 2017), such as voice conversion, Non-audible Murmur (NAM)-to-whisper conversion, whisper-to-normal conversion, voice imitation, speech enhancement, Text-to-Speech (TTS) synthesis, and a very recent application to speaker recognition, etc. The objective of this APSIPA DL talk by Prof. Patil was to first understand the fundamentals of GANs w.r.t. motivation, applications, various GAN architectures along with future research directions. Finally, the talk proposed several open research problems (relationship with variational autoencoders (VAEs and their asymptotic consistency, convergence of GANs), that needs immediate attention to fully realize the potential of GANs in several speech technological applications.
Prof. (Dr.) Thomas Fang Zheng

The On-Going Plan of Trusted Identity Authentication in China and the Research Challenges of Speaker Recognition (Part I & Part II)

With the rapid development of information technologies and telecommunication technologies, the mobile applications have been becoming so popular that the cyberspace which human beings depend on has been greatly extended. In this situation, the cyberspace security is very important and the trusted identity authentication can be regarded one of the most important security issues in cyberspace security. In this talk, by Prof. Zheng as a specific form of trusted identity authentication, the unsupervised identity authentication (USIA) was studied, covering two related scenarios, self-services in certain physical space and activities in whole cyberspace. Many researchers and developers have been making efforts based on biometric recognition technologies, including fingerprint, face, iris, and voice. However, two issues have brought great attention from people, one is the anti-spoofing problem, and another is the privacy protection problem.

Prof. (Dr.) Suryakant V. Gangashetty

Voice Conversion Approaches for Spoofing Attacks

Voice conversion (VC) is an area of speech processing that deals with the conversion of the perceived speaker identity. In other words, the speech signal uttered by a first speaker, the source speaker, is modified to sound as if it was spoken by a second speaker, referred to as the target speaker. Voice conversion is closely related to speaker verification. The former analyze and synthesize the voice characteristics of speakers, while the latter distinguishes one from another. Thus, VC is a research challenge where the identity of a speaker from his/her speech has to be modified such that, it has to be perceived as if spoken by a desired (target) speaker. However, the information provided by the speaker has to be intact. The speaker identity lies at the acoustic-level and also at the linguistic-level. Nevertheless, much attention has been on modeling the spectral features and fundamental frequency from the speakers (source) speech. With the emerging technologies and computing abilities, once prescient ideas to emulate the human nature in machines are eventually achievable in the recent past.

We specifically discuss the artificial production of speech by leaning into Voice Conversion (VC) systems. VC technology enables the speaker variations through the artificial imitation of the desired target speaker. We have thus explored VC frameworks for speaker variations (conversion) with and without parallel corpus. VC systems mainly find their applications in emotion conversion, personalized TTS systems, normal-to-Lombard speech conversion, speech to singing conversion and carry out spoofing attack. Voice conversion has become one of the most easily accessible techniques to carry out spoofing attack, therefore, presents a threat to speaker verification systems. This talk discussed artificial speech generation for text-dependent voice (speaker) conversion. In addition, this talk briefly introduced various spoofing attack studies under different conditions with a focus on voice conversion spoofing attack. A traditional parallel VC approach was then demonstrated.
Industry Perspective Talks

Dr. Vikram Vij

Large Scale ASR for Multi Device Ecosystem

In this talk, Dr. Vikram Vij shared the journey that Samsung has undertaken in developing its Voice Assistant (Bixby) and particularly, Automatic Speech Recognition (ASR) system that powers it. Several independent components, such as pre-processors (acoustic echo cancellation, noise suppression, neural beamforming and so on), wake word detectors, end-point detectors, hybrid decoders, inverse text normalizers work together to make a complete ASR system. We are in an exciting period with tremendous advancements made in recent times. The development of End-to-End (E2E) ASR systems is one such advancement that has boosted recognition accuracy significantly and it has the potential to make speech recognition ubiquitous by fitting completely on-device thereby bringing down the latency and cost and addressing the privacy concerns of the users. Samsung envisions a huge value in bringing Bixby to a variety of existing devices and new devices, such as social robots, which throws many technical challenges particularly in making the ASR very robust. In his talk, Dr. Vikram presented the cutting-edge technologies that his team is developing - Far-Field Speech Recognition, E2E ASR, Whisper Detection, Contextual End-Point Detection (EPD), On-device ASR and so on. He also elaborated on the research activities of his team at Samsung R&D Institute, Bangalore.
Mrs. Swaran Lata

The Research Challenges of Speaker & Speech Recognition for Punjabi Language

Punjabi belongs to the Indo-Aryan family of languages but is unique due to its tonal characteristics. It is a less-resource language (LRL) on account of non-availability of electronic resources. Punjabi literature reveals that the suprasegmental phonemes, such as tone, nasalization, and stress are realized at the syllable-level. There is abundance of geminated words in which stress co-occurs on the geminated consonant. The disyllabic words have highest frequency of occurrence. There are very few quadrisyllabic/polysyllabic words. There is limited work available on Punjabi generative phonology. The tone and linguistic stress patterns of Punjabi language were discussed. The research challenges of speech and speaker recognition was presented based on $F_0$ contours of a small set of words in varying linguistic contexts for Punjabi disyllabic words. This talk by Mrs. Swaran Lata presented relevance of Punjabi language for voice biometrics.

Dr. Sunil Kumar Kopparapu

Pathological Speech Processing

Ability to process and possibly understand pathological speech has huge and deep implications on improving the quality of life for people with speech disorders and elderly. Speech technology can be exploited, in a smart way, to address problems in various areas, such as early assessment, therapy, and making commercial conversational interfaces inclusive. Beside visual and tactile, voice as a modality is set to enhance our interaction with machines; it is predicted that a decade from now we will be swamped by 10 times more speech interfaces than what we see today! Performance of speech recognition as a usable technology can be attributes to availability of huge amounts of speech and language data and an ability to deep machine learn. However, privacy and practical issues, such as patient fatigue in addition to the diversity of pathological conditions prevents pathological data collection at scale. In this talk, keeping these limitations in view, Dr. Sunil Kumar talked about some of his recently published work focusing around early diagnosis, severity assessment, enhancement of pathological speech from patients with dysarthria, dementia, and articulation problems.
Dr. Sunayana Sitaram

*Speech Recognition and Speaker Diarization for Meeting Transcription*

Meeting summarization is an application of speech and language technologies which can enable participants who were not present at a meeting receive a summary of important points discussed during the meeting. Automatic transcription of multi-party meetings requires high quality Automatic Speech Recognition (ASR) to determine what was said, as well as speaker diarization to attribute the speech to a participant. This talk looked at recent work at Microsoft on ASR in combination with speaker diarization for meeting transcription. The system uses virtual microphone arrays along to perform diarization using blind beamforming, followed by ROVER/confusion network-based ASR hypotheses combination. Dr. Sunayana concluded her talk with state-of-the-art results and challenges remaining in meeting transcription.

Mr. Pranaw Kumar

*Development Journey of Indian Language ASR, TTS and Related Applications*

Devices and applications having speech interface is gaining popularity, and as a result speech technology industry is witnessing significant growth. New technologies have enabled: 1) speech recognition system to understand different languages and accents, and 2) Text-to-speech synthesis system to produce natural sounding speech output. Due to illiteracy, lack of computer knowledge, numerous languages, unfamiliarity with English language, etc. speech interface in Indian languages is boon for Indian population. Popularity of smartphones has provided a very convenient way to deploy and spread the speech technology. However, due to lack of language resources and tools, development of speech recognition and speech synthesis systems has always been challenging. This talk presented the development journey of Indian language ASR, TTS and related applications. With the time, different methods like phoneme and syllable-based unit selection, HMM-based parametric speech synthesis, deep learning etc. have been tried to get natural speech synthesis output, and various desktop, web and android applications have been developed. Similarly, development of Indian Language ASR system started with keyword and phrase recognition, and now large vocabulary continuous speech recognition (LVCSR) system is being developed. ASR system has also been deployed in various voice response and command systems. The talk by Mr. Pranaw Kumar helped participants to understand significance of ASR & TTS in Indian Context.
Felicitation of Invited Speakers

The invited speakers spared their valuable time and shared their rich and wide research experience and expertise with the participants during the Summer School. They were felicitated at a special function organized as a part of Summer School.
5 Minute Ph.D. Thesis (5 MPT) Contest

The, ‘5 Minutes Ph.D. Thesis’ contest was organized fourth time after last three such successful events in India during S4P 2018, S4P 2017, and S4P 2016, where doctoral students from India participated. The main objective of this event was to provide a unique opportunity to doctoral scholars to present their research work concisely in the broad areas of speech and audio signal processing strictly within 5 minutes time. By such initiative, doctoral students got an opportunity to interact with eminent researchers both from academia and industry. This helped them to get visibility of their work and improve their communication and presentation skills. 3 doctoral scholars presented their research work within 5 minutes time. The students who participated in ‘5 Minutes Ph.D. Thesis’ were:

(a) Ms. Nitya Tiwari, Samsung R&D Institute, Bangalore, India.
(b) Mr. Natraj K. S., Indian Institute of Technology (IIT), Bombay, India.
(c) Mr. Likhit Sai, International Institute of Information Technology (IIIT), Hyderabad, India.

The details are given at https://sites.google.com/site/s4p2019/5MPT
To select the best presentation for the **ISCA endorsed and DE GRUYTER sponsored book prizes**, an expert committee consisting of the following members was constituted by the Chair of Organizing Committee:

(a) **Thomas Fang Zheng**, Tsinghua University, China.
(b) **Jean-Francois Bonastre**, University of Avignon, France.
(c) **Nicholas Evans**, EURECOM, France.
(d) **Dong Wang**, Tsinghua University, China.
(e) **K. Samudravijaya**, Indian Institute of Technology (IIT), Guwahati, India.
(f) **Vikram Vij**, Samsung Electronics Co. Ltd., Bangalore, India.

**Note**: Dr. Vikram Vij was kindly requested to co-ordinate the overall evaluation process and announce the winners of 5MPT competition. Based on the reviews by the expert committee, two scholars were awarded **ISCA endorsed cash prizes** and **DE GRUYTER sponsored book prizes**.

![Expert Committee members to evaluate 5 Min Ph.D. Thesis contest.](image)

**Important**: To avoid conflicts of professional interest, Prof. (Dr.) Hemant A. Patil (Chair of Organizing Committee S4P-2019) was **not** included in the Expert Committee. In addition, no member of Organising Committee or faculty colleagues from DA-IICT was involved.
5 Minute Ph.D. Thesis Award Ceremony

Based on the assessment of presentations by the Expert Committee, Dr. Nitya Tiwari (Samsung R&D, Bangalore) and Mr. Natraj K. S. (IIT Bombay) were awarded ISCA endorsed First Prize of Rs. 15,000 /-, and Second Prize of Rs. 10,000 /-, respectively, and both the winners also awarded DE GRUYTER sponsored book prize.

1st Prize Winner Ms. Nitya Tiwari (Samsung R&D Institute, Bangalore)

2nd Prize Winner Mr. Natraj K. S. (IIT Bombay)
Sponsors Presentation

1. DA-IICT Gandhinagar ([https://www.daiict.ac.in/](https://www.daiict.ac.in/))

   Prof. Sanjeev Gupta presenting summary of R&D achievements on behalf of DA-IICT Gandhinagar


   Prof. Thomas F. Zheng presenting about APSIPA, and APSIPA DL Program
APSIPA Distinguished Lecture (DL)

Prof. Dong Wang presenting about APSIPA, APSIPA DL program, and delivered APSIPA Distinguished Lecture on, “Deep Feature Learning and Normalization in Speaker Recognition”.

Prof. Hemant A. Patil presenting about APSIPA, APSIPA DL program, and delivered APSIPA Distinguished Lecture on, “Generative Adversarial Networks (GANs) for Speaker Recognition”.

DE GRUYTER Sponsored Book Prizes for Winners of 5MPT Contest

DE GRUYTER sponsored books were displayed in the auditorium of S4P 2019 for all five days during the event.
Interaction between Participants and Speakers during Tea/Coffee Breaks and Lunch

In addition to the expert and invited talks and research sessions, there were extensive interaction sessions between the participants during the sessions and during tea, lunch breaks, and dinner time. Thus, participants of summer school got excellent opportunity to interact more closely with the invited speakers.
Volunteers of S4P 2019, DA-IICT Gandhinagar

Prof. Patil along with student volunteer for preparation of proceeding for S4P 2019

Thus, summer school activity strengthens a great bond of team spirit and interpersonal skills within members of Speech Group @ DA-IICT.

"Summer School on Speaker Recognition and Diarization
July 06 to July 10, 2019 at DA-IICT, Gandhinagar "
Valedictory Function: Prof. Hemant A. Patil appreciated highly independent and immense contribution of student volunteers during the valedictory function for the grand success of S4P 2019.

The Organizing Committee records its appreciation for the painstaking efforts and great team work of the volunteers and staff without which successful organizing of S4P 2019 at DA-IICT Gandhinagar would not have been possible. The students and staff members from DA-IICT Gandhinagar who have worked as volunteers are:

- Anshu Chittora
- Nirmesh J. Shah
- Madhu R. Kamble
- Ankur T. Patil
- Satyam Kumar
- Divyesh Rajpura
- Rajul Acharya
- Mirali Purohit
- Mihir Parmar
- Maitreya Patel
- Jalansh Munshi
- Hashit Malaviya
- Harsh Kota
- Akshar Joshi
- Kirit Pandya
- Santosh Pandit
- Bhavesh Shah
- Rajendra Shah
- Sudhir Dave
- Jitendra Parmar
- Sawan Kumar Sachaniya
- Nimesh Patel
- Rajesh Patel
- Gyanesh Pandya
- Prabhunath Sharma
- Shirish Varma
- Niketa Raval
- Geeta Nair
Jhanvi Chauhan
Kirtana Phatnani
Yesha Gosaliya
Krusha Zala

Ramesh Prajapati
Dinesh Prajapati
Abhilash Bhaskaran
Jainik Patel
Chaitanya Bhamare
Ashvin Chaudhari
Darshan Parajapati
Priyank Santola
Ankit Sharma
Hemangi Sharma
Call For Papers

INTERSPEECH 2019

"CROSSROADS OF SPEECH AND LANGUAGE"

www.interspeech2019.org

CALL FOR PAPERS AND PROPOSALS FOR TUTORIALS, SPECIAL SESSIONS/CHALLENGES, AND SHOW & TELL

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 theories to advanced applications. In addition to regular oral and poster sessions, INTERSPEECH 2019 will feature keynote talks by internationally renowned experts, tutorials, special sessions and challenges, show & tell sessions, and exhibits. A number of satellite events will also take place around INTERSPEECH 2019.

Original papers are solicited in, but not limited to, the following areas:
1. Speech Perception, Production and Acquisition
2. Phonetics, Phonology, and Prosody
3. Analysis of Paralinguistics in Speech and Language
4. Speaker and Language Identification
5. Analysis of Speech and Audio Signals
6. Speech Coding and Enhancement
7. Speech Synthesis and Spoken Language Generation
8. Speech Recognition – Signal Processing, Acoustic Modelling, Robustness, and Adaptation
9. Speech Recognition – Architecture, Search, and Linguistic Components
10. Speech Recognition – Technologies and Systems for New Applications
11. Spoken Dialog Systems and Analysis of Conversations
12. Spoken Language Processing – Translation, Information Retrieval, Summarization, Resources, and Evaluation

A complete list of the scientific areas and topics including special sessions is available at www.interspeech2019.org

PAPER SUBMISSION

Papers intended for INTERSPEECH 2019 should be up to four pages of text. An optional fifth page could be used for references only. Paper submissions must conform to the format defined in the paper preparation guidelines and as detailed in the author’s kit on the conference website. Please be aware that INTERSPEECH 2019 will use new templates and submissions will be accepted only in the new format. Submissions may also be accompanied by additional files such as multimedia files, to be included on the proceedings USB drive. Authors must declare that their contributions are original and have not been submitted elsewhere for publication. Papers must be submitted via the online paper submission system. The working language of the conference is English, and papers must be written in English.

We look forward to receiving your submissions and to your participation in INTERSPEECH 2019.

IMPORTANT DATES

» November 30, 2018
» February 1, 2019
» February 15, 2019
» February 28, 2019
» March 29, 2019
» April 5, 2019
» April 20, 2019
» June 17, 2019
» June 24, 2019
» July 1, 2019

Special session/challenges proposals due
Tutorial proposals due
Submission portal opens
Satellite workshops/events proposals due
Paper submission deadline
Final paper submission deadline
Show & Tell proposals due
Acceptance/rejection notification
Registration opens
Camera-ready paper due

General Chairs:
Gamal Kishk, TU Graz, Austria
Zdenka Kadzig, University of Maribor, Slovenia

Technical Chairs:
Thomas Hirsch, U. Sheffield, UK
Bill Schroeder, University of Potsdam, Germany/UK

Organizing Committee Members:
Michele Broccardo, Google, USA
Gerhard Bohme, Sächsische Akademie der Wissenschaften, Germany
Janet Boll, TU Graz, Austria
Eugene Osterre, TU Graz, Austria

Contact: The technical program chairs: Thomas Hirsch and Björn Schauer at tpc-chairs@interspeech2019.org
CALL FOR PAPERS

Intelligent Signal & Information Processing

APSIPA ASC 2019 (www.apsipa2019.org) is the 11th annual conference organized by Asia-Pacific Signal and Information Processing Association (APSIPA), which will be held on November 18-21, 2019, in Lanzhou, China. Founded in 2009, APSIPA aims to promote research and education in signal processing, information technology, and communications. The annual conferences have been held previously in Sapporo, Japan (2009), Singapore (2010), Xi’an, China (2011), Los Angeles, USA (2012), Kaohsiung, Taiwan (2013), Siem Reap, Cambodia (2014), Hong Kong, China (2015), Jeju, Korea (2016), and Kuala Lumpur, Malaysia (2017) and Hawaii, USA (2018).

APSIPA is interested in all aspects of signal and information processing theories, algorithms, securities, implementations, and applications. All accepted papers will be indexed by EI compendex and archived by IEEE Xplore.

The technical program includes, but not limited to, the following areas:

- Biomedical Processing and Systems
- Image, Video, and Multimedia
- Signal and Information Processing Theories and Methods
- Signal Processing Systems: Design and Implementation
- Speech, Language, and Audio
- Wireless Communications and Networks
- Data Analytics and Machine Learning
- Deep Learning-based Signal Processing
- Multimedia Security and Forensics
- Signal and Information Processing in Energy and Sustainability
- Signal and Information Processing for the Internet of Things
- Signal and Information Processing Education

Paper Submission

Prospective authors are invited to submit papers of 4 to 10 pages in length.

<table>
<thead>
<tr>
<th>Important Dates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Submission of Proposals for Special Sessions</td>
<td>May 20, 2019</td>
</tr>
<tr>
<td>Submission of Proposals for Forum, Panel &amp; Tutorial</td>
<td>June 8, 2019</td>
</tr>
<tr>
<td>Submission of Regular Session Papers</td>
<td>June 20, 2019</td>
</tr>
<tr>
<td>Submission of Special Session Papers</td>
<td>June 30, 2019</td>
</tr>
<tr>
<td>Notification of Paper Acceptance</td>
<td>July 31, 2019</td>
</tr>
<tr>
<td>Submission of Camera-ready Papers</td>
<td>September 1, 2019</td>
</tr>
<tr>
<td>Author (early-bird) Registration Deadline</td>
<td>September 1, 2019</td>
</tr>
<tr>
<td>Tutorial Session Date</td>
<td>November 18, 2019</td>
</tr>
<tr>
<td>Summit and Conference Dates</td>
<td>November 18-21, 2019</td>
</tr>
</tbody>
</table>
Note: We have circulated CFP for INTERSPEECH 2019, APSIPA ASC 2019 in the participant’s kits of S4P 2019.