Closure Report

On

ISCA Supported

Summer School on Speech Signal Processing

Theme:
Automatic Speech Recognition (ASR)

6-10 July 2024

Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT)

Gandhinagar, India.

URL: https://sites.google.com/view/s4p2024/

Message from Organizing Chair

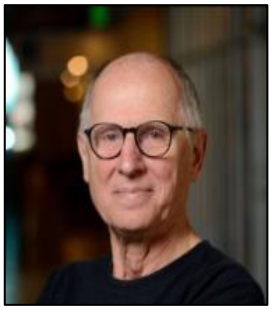
On behalf of the Organizing Committee, I record our appreciation for the valuable contribution made by eminent invited speakers (from academics and industry), participants, international program committee, DA-IICT faculty colleagues, administration, staff, and student volunteers towards conducting the 6th edition of summer school with the theme 'Automatic Speech Recognition (ASR)' during July 6-10, 2024 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 Automatic Speech Recognition (ASR). Furthermore, to encourage young talent, the school presented 5th edition of 5 Minute Ph.D. Thesis (5MPT) contest with four **ISCA endorsed** cash prizes.

We were honoured that we had eminent world-class expert, namely, Prof. (Dr.) Hynek Hermansky (Department of Electrical and Computer Engineering, Johns Hopkins University, USA), Dr. Bhuvana Ramabhadran (Google Research, USA), Dr. Mathew Magimai Doss (IDIAP Research Institute, Martigny, Switzerland), Prof. (Dr.) Chng Eng Siong (Nanyang Technological University (NTU), Singapore.), Prof.(Dr.) Srikanth Madikeri (University of Zurich, Switzerland). In addition, we have Prof. (Dr.) Bayya Yegnanarayana (IIIT, Hyderabad), Prof.(Dr.) C. V. Jawahar (IIIT, Hyderabad), Prof. (Dr.) Sriram Ganapathy (IISc, Bengaluru), Prof.(Dr.) Preethi Jyothi (IIT Bombay), Dr. Aparna Walanj (Kokilaben Dhirubhai Ambani Hospitals and Research Center, Mumbai), Prof. (Dr.) Samudravijaya (KL university,), and Prof. (Dr.) Hemant A. Patil (DA-IICT, Gandhinagar). At the Summer School, motivated from INTERSPEECH 2018, we have organized Industry Perspective Talks in which senior industry personnel, namely, Dr. Tara N. Sainath (Google Research, USA), Dr. Sunayana Sitaram (Microsoft Research Laboratory, Bangalore), Dr. Harish Arsikere (Amazon, Bangalore), Dr. Hardik B. Sailor (Institute for Infocomm Research (I2R), A*STAR, Singapore), Dr. Vikram C. Mathad (Samsung Research Institute, Bengaluru), Dr. Nirmesh J. Shah (Sony Research, India), Dr. Sunil Kumar Kopparapu (TCS Innovation Lab, Mumbai), Mr. Amitabh Nag (Ministry of Electronics & Information Technology (MeitY), New Delhi), Mr. Ajay Rajawat (Ministry of Electronics & Information Technology (MeitY), New Delhi), Mr. Dipesh K. Singh (Augnito, Mumbai), Ms. Gauri Prajapati (Microsoft Research, Bengaluru). Events of this kind cannot happen without generous financial support from potential sponsors. In this regard, we express our deep gratitude and appreciation to the sponsors, namely, DA-IICT Gandhinagar, Google, International Speech Communications Association (ISCA), Indian Speech Communication Association (IndSCA), The Ministry of Electronics and Information Technology (MeitY), and Digital India Bhashini Division (BHASHINI), without which organizing this event would not have been possible. In addition, we would like to Prof. (Dr.) Phil Green (University of Sheffield, UK) for their valuable feedback on our proposal for ISCA support to S4P 2024.

This summer school witnessed 95 attendees including researchers, industry personnel, faculty members, and students from all over world. We would like to sincerely acknowledge kind support from DA-IICT administration, including Prof. (Dr.) Tathagata Bandyopadhyay, Director, Shri. Siddharth Swaminarayan, Executive Registrar, Ms. Krutika Raval, Head, HR and all admin staff. Further, we thank Dr. Vikram Vij, who recommended Dr. Vikram C. M. as invited speaker from Voice Intelligence Group of Samsung R&D Institute, Bangalore. Furthermore, we acknowledge excellent support from KL University at Vijaywada and Hyderabad campus 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.

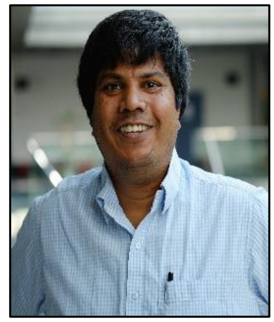
Prof. (Dr.) Hemant A. Patil (ISCA Member)
DA-IICT Gandhinagar
Associate Editor, IEEE Signal Processing Magazine for 2021-2023
ISCA Distinguished Lecturer for 2020-2022
APSIPA Distinguished Lecturer for 2018-2019

Invited Speakers



Hynek Hermansky
Johns Hopkins University, USA

Bhavana Ramabhadran
Google Research, USA



Mathew Magimai Doss
Idiap Research Institute, Martigny, Switzerland

Chng Eng Soing
Nanyang Technological University (NTU), Singapore

Srikanth Madikeri
University of Zurich, Switzerland

Bayya Yegnanarayana
IIIT, Hyderabad

C. V. Jawahar
IIIT, Hyderabad

Sriram Ganapathy
IISc, Bengaluru

Preethi Jyothi
Indian Institute of Technology, Bombay

Aparna Walanj
Kokilaben Dhirubhai Ambani Hospitals and Research Center, Mumbai

Hemant A. Patil
DA-IICT, Gandhinagar

Samudravijaya K
KL University, Vijayawada

Industry Perspective Talks

Tara N. Sainath
Google Research,
USA

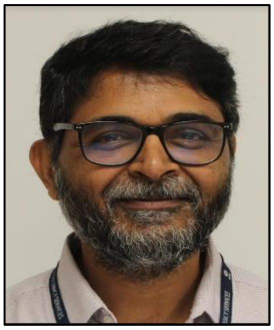
Sunayana Sitaram
Microsoft Research,
Bengaluru

Harish Arsikere
Amazon, Bengaluru

Hardik B. Sailor
A*STAR,
Singapore

Vikram C. Mathad
Samsung Research Institute,
Bengaluru

Nirmesh J. Shah
Sony Research, Mumbai



Sunil Kumar Kopparapu
TCS, Mumbai

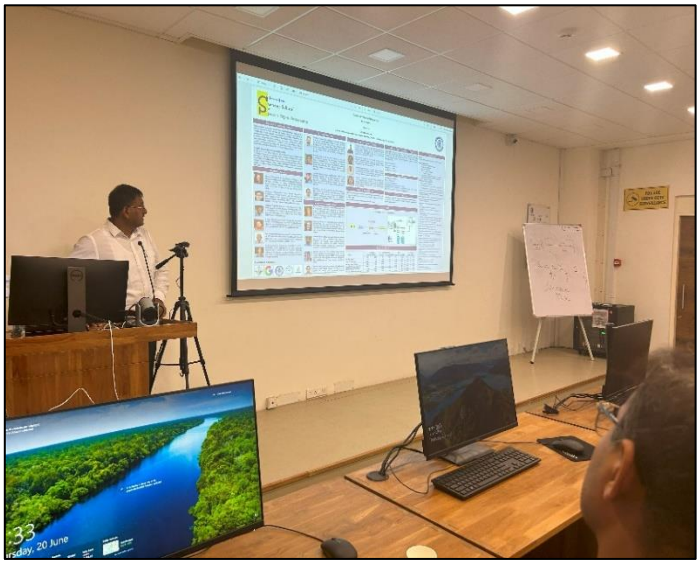
Amitabh Nag
CEO of Digital India Bhashini Division (BHASHINI), MeitY
New Delhi

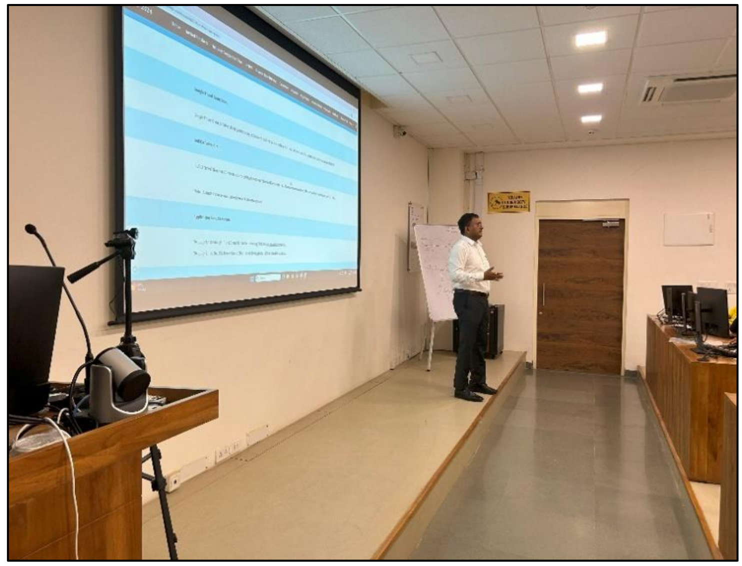
Dipesh K. Singh
Augnito, Mumbai

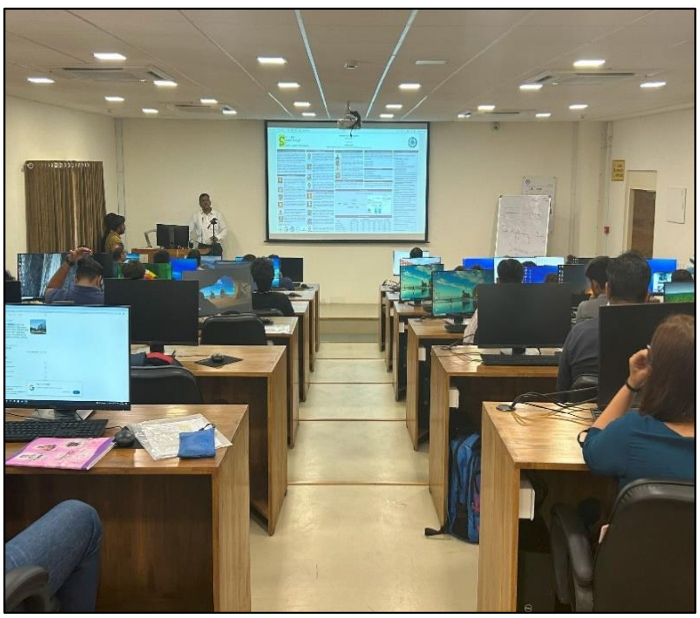
Gauri Prajapati
Microsoft Research, Bengaluru

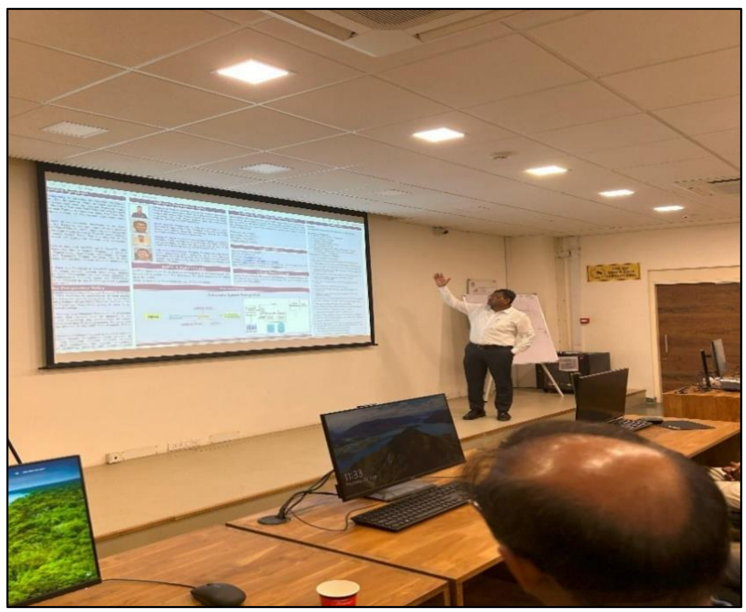
Ajay Rajawat
Manager (Support & Onboarding), Digital India Bhashini Division (BHASHINI), MeitY,
New Delhi

Promotion of Summer School (S4P 2024)









ISCA Supported Summer School on Speech Signal Processing (S4P 2024)

The *6th* edition of Summer School on Speech Signal Processing was held at Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar, India. S4P 2024 was a 5-day event during 6-10 July 2024, which was 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.

Announcement and Publicity of S4P 2024

Announcement for S4P 2024 was made worldwide through the following:

1) S4P 2024 Poster Distribution:

- ISCA Speech Labs across the world
- SPCOM 2024, IISc Bengaluru (July 1-4, 2024)
- 1000 potential Institutes of National Importance (IISc, IIT's, NIT's, etc.), Universities, Colleges, R&D Labs, and industries across India.

2) S4P 2024 website:

URL: https://sites.google.com/view/s4p2024/

Participants

Around 72 participants from India have attended S4P 2024. The brief detail of the participants is shown in Figure 1.

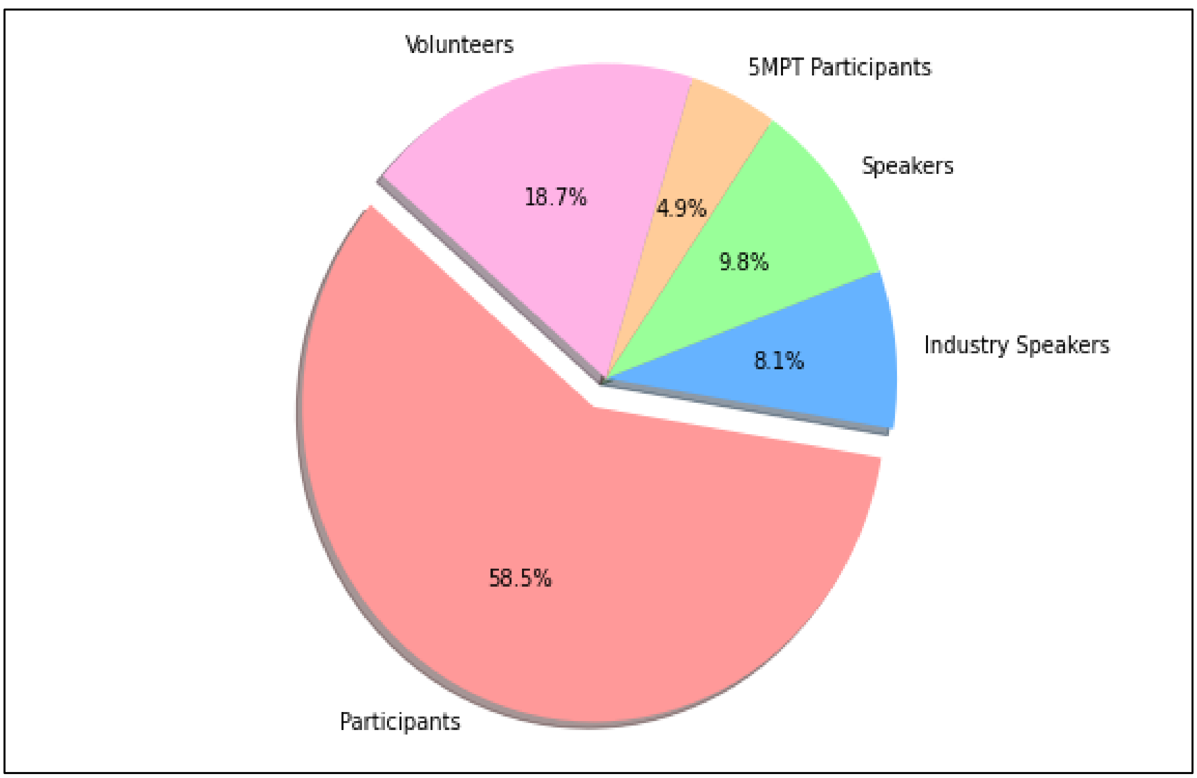


Figure 1: The distribution of the participants that attended S4P 2024 at DA-IICT Gandhinagar.

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

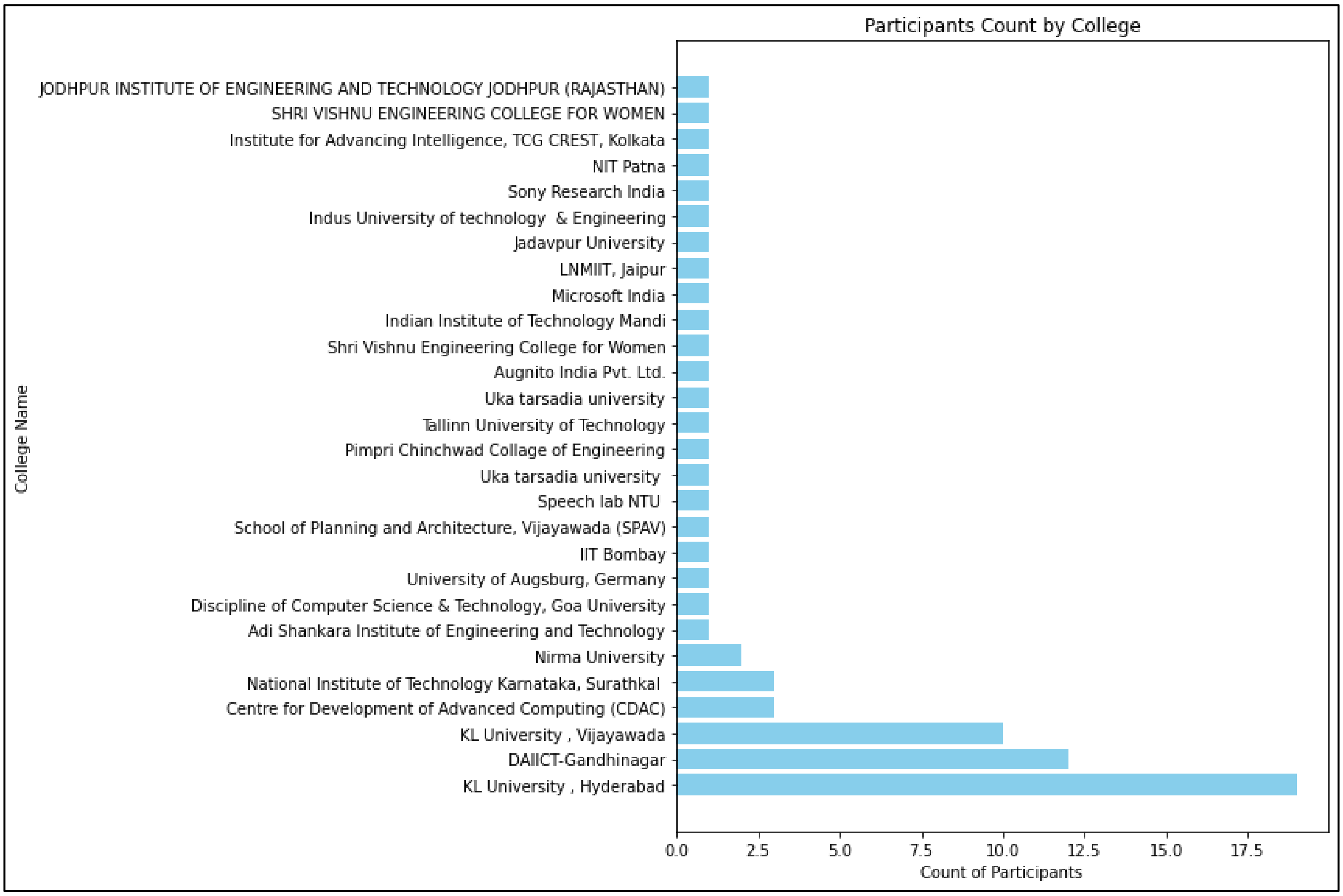


Figure *2*: The institute-wise distribution of the participants who attended S4P 2024 at DA-IICT Gandhinagar.

Patron

- Tathagata Bandyopadhyay, Director, DA-IICT Gandhinagar.

Program Committee (PC) Members

S4P 2024 was organized with the guidance and support of various speech researchers across the country and abroad. The Program Committee (PC) members of S4P 2024, who guided and supported in planning and designing the technical program and other related tasks are:

- Hermann Ney, RWTH Aachen, Germany
- Yu Tsao, Academia Sinica, Taiwan
- Shrikanth (Shri) Narayanan, University of Southern California, USA
- Tatsuya Kawahara, School of Informatics, Kyoto University, Japan
- Petros Maragos, National Technical University of Athens, Greece
- Douglas O'Shaughnessy, Énergie Matériaux Telecommunications Research Centre, Canada
- Dilek Hakkani-Tur, University of Illinois Urbana-Champaign, USA
- Mark Allan Hasegawa-Johnson, University of Illinois Urbana-Champaign, USA
- Lori Lamel, French National Centre for Scientific Research, France
- Odette Scharenborg, Delft University of Technology, The Netherlands
- Yannis Stylianou, University of Crete, Greece
- Torbjørn Svendsen, Norwegian University of Science and Technology, Norway
- Kong Aik Lee, The Hong Kong Polytechnic University, Hong Kong
- Saikat Chatterjee, KTH Royal Institute of Technology, Sweden
- Thomas Fang Zheng, Tsinghua University, China
- Nancy Zlatintsi, National Technical University of Athens, Greece
- Heidiki Kawahara, Wakayama University, Japan
- Hynek Hermansky, John Hopkins University, USA
- Nancy F. Chen, Institute of Infocom Research (I2R), Singapore
- Hardik B. Sailor, Institute of Infocom Research (I2R), Singapore
- Dong Yu, Tencent AI Lab, USA
- Rodrigo Capobianco Guido, São Paulo State University, Brazil
- Monojit Chaudhary, Mohamed bin Zayed University of Artificial Intelligence, UAE
- Rita Singh, Carnegie Mellon University, USA
- Tiago H. Falk, University of Quebec, Canada
- Kai Yu, Shanghai Jiao Tong University, China
- Jan (Honza) Cernocky, Brno University of Technology, Czech Republic
- Eliathamby Ambikairajah, University of New South Wales (UNSW), Australia
- Isabel Trancoso, University of Lisbon, Portugal
- Elmar Nöth, Friedrich-Alexander Universität Erlangen-Nürnberg, Germany
- Rohit Prabhavalkar, Google Research, USA

- Jean-François Bonastre, INRIA, France
- Hervé Bourlard, IDIAP Research Institute, Switzerland
- Ma Bin, Alibaba Inc., Singapore
- Nicholas Evans, EURECOM, France

Organizing Committee

- Hemant A. Patil, DA-IICT Gandhinagar (Chair)
- Suryakanth V. Gangashetty, KL University, Vijayawada (Convener)
- Preethi Jyothi, IIT Bombay
- Vipul Arora, IIT Kanpur
- Priyankoo Sarmah, IIT Guwahati
- A.D. Dileep, IIT Dharwad
- Sunil Kumar Kopparapu, TCS Innovation Labs, Mumbai
- Pranaw Kumar, C-DAC Mumbai

Sponsors

The Organizing Committee would like to thank the following sponsors for the generous support extended to conduct S4P 2024:

International Speech Communication Association (ISCA)

Google

Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT)

Indian Speech Communication Association (IndSCA)

Digital Indian Bhashini Division (BHASHINI)

The Ministry of Electronics and Information Technology (MeitY)

KLEF Vaddeswaram

Travel Grants

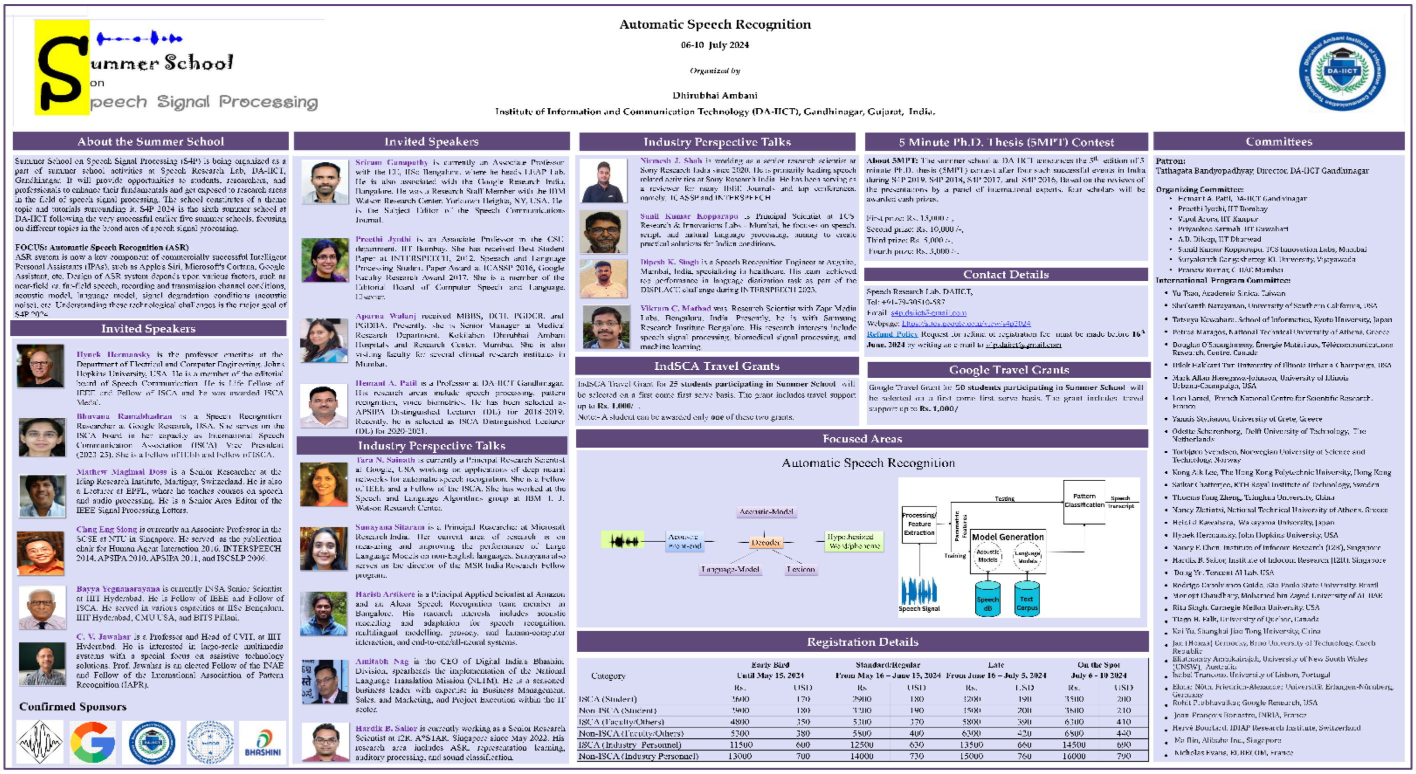
- Google Travel Grants for selected on a first serve basis in summer school:

Sr No.	Name	Program	Affiliation
1	Atish Shankar Ghone	Doctoral Degree (Ph.D. or higher)	Pimpri Chinchwad Collage of Engineering
2	Pratik Ranjan Roy Chowdhuri	Doctoral Degree (Ph.D. or higher)	National Institute of Technology Karnataka Surathkal
3	Garima Pandey	Doctoral Degree (Ph.D. or higher)	NITK Surathkal
4	Nikunj Dalsaniya	Doctoral Degree (Ph.D. or higher)	Institute of Technology, Nirma University
5	Ankita	Doctoral Degree (Ph.D. or higher)	NIT Patna
6	Baidyanath Mahato	Doctoral Degree (Ph.D. or higher)	Jadavpur University
7	Nikhil Raghav	Doctoral Degree (Ph.D. or higher)	Institute for Advancing Intelligence, TCG CREST, Kolkata
8	Darshan Prabhu	Doctoral Degree (Ph.D. or higher)	IIT Bombay
9	Akansha Tyagi	Doctoral Degree (Ph.D. or higher)	IIT Mandi
10	Pragya Gupta	Master's Degree (M.Tech, M.Sc, or other)	NIT Surathkal Karnataka

- IndSCA Travel Grants for selected on a first serve basis in summer school:

Sr No.	Name	Program	Affiliation
1	Atish Shankar Ghone	Doctoral Degree (Ph.D. or higher)	Pimpri Chinchwad Collage of Engineering
2	Nikunj Dalsaniya	Doctoral Degree (Ph.D. or higher)	Institute of Technology, Nirma University
3	Ankita	Doctoral Degree (Ph.D. or higher)	NIT Patna
4	Baidyanath Mahato	Doctoral Degree (Ph.D. or higher)	Jadavpur University
5	Nikhil Raghav	Doctoral Degree (Ph.D. or higher)	Institute for Advancing Intelligence, TCG CREST, Kolkata
6	Akansha Tyagi	Doctoral Degree (Ph.D. or higher)	IIT Mandi
7	Pragya Gupta	Master's Degree (M.Tech, M.Sc, or other)	NIT Surathkal Karnataka

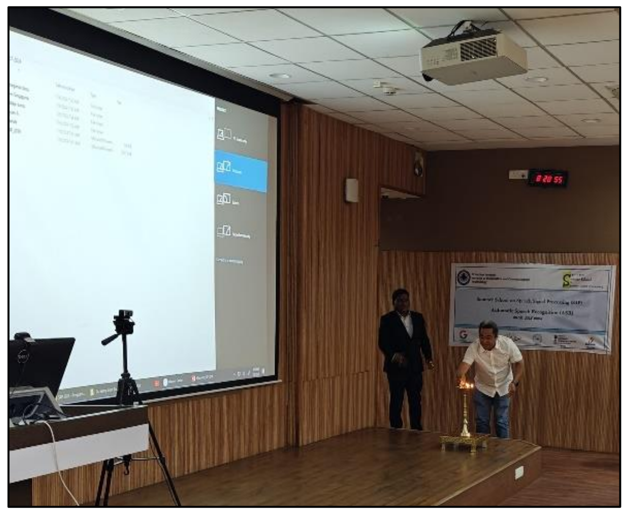
Poster of S4P 2024

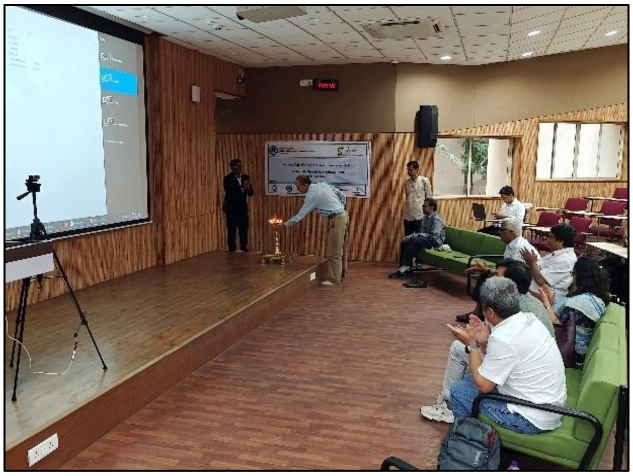


Banner of S4P 2024

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.) B. Yegnanarayana, Prof. (Dr.) Chng Eng Soing, Prof. (Dr.) Mathew Magimai Doss, Dr. Aparna Walanj, Prof, (Dr.) Yash Vasavada, Prof. (Dr.) Manik Lal Das (Dean, Faculty), Prof. (Dr.) Bhaskar Chaudhary (Dean of Academic Programs).





Inaugural and Welcome Address by Prof. Hemant A. Patil, Organizing Chair, S4P 2024

- Prof. Patil expressed appreciation to all the sponsors of S4P 2024. He also encouraged participants to consider attending the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2025), which will be held in Hyderabad, India, and urged them to take the opportunity to submit their research after gaining valuable insights from the event.

Design of Technical Program

Program Schedule

Time Slot	6th July (Saturday)	7th July (Sunday)	8th July (Monday)	9th July (Tuesday)	10th July (Wednesday)
08:00-08:30	Registration and Inauguration	-	-	-	
08:30– 09:30	B. Yegnanarayana (L1)	Srikanth Madikeri (L8)	Amitabh Nag, Ajay Rajawat (L15)	Srikanth Madikeri (L20)	Srikanth Madikeri (L27)
09:30– 10:30	Mathew M. Doss (L2)	Mathew M. Doss (L9)	Sriram Ganapathy (L16-Part-1)	Mathew M. Doss (L21)	Mathew M. Doss (L28)
10:30– 11.00	Tea Break				
11:00 12:00	Chng Eng Siong (L3)	Samudravijaya K. (L10)	Sriram Ganapathy (L16-Part2)	Vikram C. M. (L22)	Nirmesh J. Shah (L29)
12:00 13:00	B. Yegnanarayana (L4)	Suryakanth V Gangashetty (L11)	Chng Eng Siong (L17)	Hardik B. Sailor (L23)	Preethi Jyothi (L30)
13:00– 14:30	Lunch Break				
14:30– 15:30	Srikanth Madikeri (L5)	Aparna Walanj (L12)	Mathew M. Doss (L18)	C. V. Jawahar (L24)	Sunayana Sitaram (L31)
15:30-16:00	Tea Break				
16:00– 17:00	Harish Arsikere (L6)	Gauri Prajapati (L13)	Dipesh Kumar Singh (L19)	Hemant A. Patil (L25)	Sunil Kumar Kopparapu (L32)
17:00– 17:30	Tea Break				
17:30- 18:30	Samudravijaya K. (L7)	Tara Sainath (L14)	Sponsors' Presentation 5 Min Ph. D. Thesis Contest*	Hynek Hermansky (L26)	Bhuvana Ramabhadran (L33) Award Ceremony and Valedictory

Technical Content of S4P 2024

Lecture Topics	
L1	Processing Phase of Speech Signals
L2	Fundamentals of ASR– A Symbolic Perspective: Part-I Abstract Formulation of the ASR
L3	The NTU Speech Team's Experience in Adapting Whisper
L4	Fundamentals of Speech Processing: Signals and Systems Perspective
L5	Loss Functions for Training Automatic Speech Recognition Systems
L6	Bootstrapping ASR for New Languages
L7	An Overview of Traditional Approaches to ASR
L8	Publicly Available Open Source ASR Models and Their Applications
L9	Fundamentals of ASR– A Symbolic Perspective: Part-II Statistical (Bayesian) Approach to ASR
L10	ASR for Low Resource Indian Languages
L11	TBD
L12	Ethics in Research
L13	Voice Privacy in the Age of AI and Big Data
L14	TBD
L15	BHASHINI: MeitY's Vision of Bridging the Language Barriers
L16-1	Beyond the Frame: Multi-Scale Self-Supervised Speech Representation Learning: Part-1
L16-2	Beyond the Frame: Multi-Scale Self-Supervised Speech Representation Learning: Part-2
L17	Enabling Large Language Models (LLM) for ASR
L18	Fundamentals of ASR– A Symbolic Perspective: Part-III Posterior Based Approach to ASR
L19	Advancements in Multi-Accent and Noise Robust ASR Using Semi-Supervised Learning and Multimodal Approaches
L20	Training ASR with Limited Resources
L21	Fundamentals of ASR– A Symbolic Perspective: Part-IV
L22	Speech AI: From Command Recognition to Live Call Translation
L23	Representation Learning for Speech: From Unimodal to Multimodal
L24	Seeing is Listening
L25	Part-1: Voice Conversion Based Data Augmentation Using CycleGAN for Children's ASR Part-2: Dysarthric ASR: Assistive Speech Technology
L26	Why Should We Ask Why?
L27	TBD
L28	Pathological Speech Processing: Relevance to ASR
L29	Evolution of Speech Foundation Models and Its Applications in Speech AI
L30	Text-Only Adaptation of End-to-End Speech Recognition Models
L31	Evaluating LLMs on Languages Beyond English: Challenges and Opportunities
L32	Audio & Speech Processing
L33	TBD

Interaction of Prof. (Dr.) Tathagata Bandyopadhyay (Director, DA-IICT) and Prof. (Dr.) Hemant A. Patil (Chair, S4P) with Invited Speakers

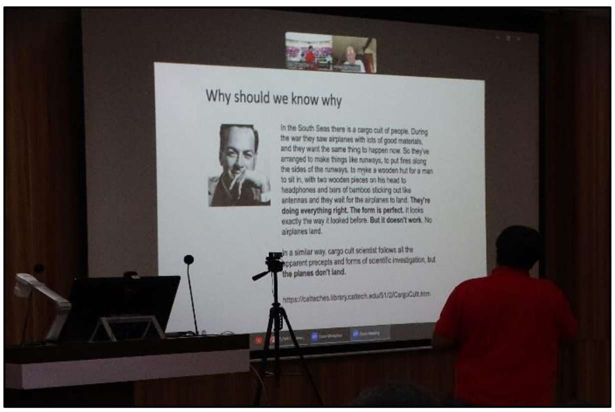
- During this interaction, the Director appreciated the contribution of invited speakers, organizing committee, student volunteers, faculty, and admin staff to the grand success of S4P 2024.

Invited Speaker's Talks

Prof. (Dr.) Hynek Hermansky

Why should we ask why?

We often present advances in automatic recognition of speech (ASR) by describing the most successful configuration of available open software processing modules, sometimes adding new elements, and reporting the accuracy of the obtained results. So, what is being reported to the community is HOW the work was done and WHAT has been the output. That is understandable since reviewers are evaluating our papers by checking if the work is replicable (the HOW element) and if the progress is demonstrated (the WHAT element). However, one can argue that more scientific progress could be made when the report also contains an explanation of WHY the processing was effective. Some attempts to follow this advice in our own work are discussed in the talk.



Dr. Mathew Magimai Doss

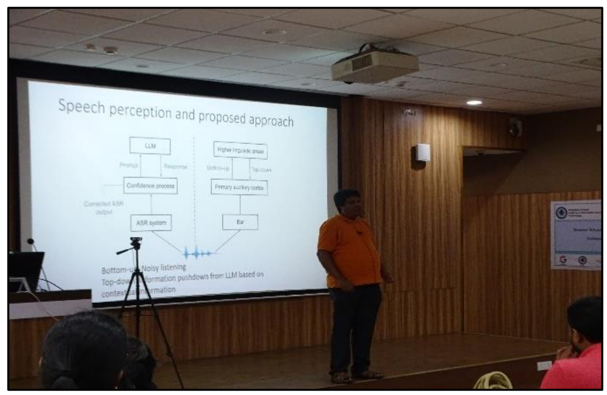
Fundamentals of Automatic Speech Recognition – A Symbolic Perspective

Over the decades, automatic speech recognition (ASR) approaches have evolved from more knowledge-driven to data-driven. A question that arises is: whether the approaches are so different from each other? In this series of presentations, I will present a symbolic perspective of the ASR problem through which, I will provide links between (a) knowledge-based approach, (b) instance-/template-based approach, and (c) statistical ASR approach, and show that these approaches are not all that different.

Talk 1: In this lecture, I will present an abstract formulation of the ASR problem, where ASR can be seen as combination of language generation (generation of word hypotheses) and matching of word hypothesis with the observed speech signal. Based on that formulation, I will elucidate knowledge-based approach and instance-based approach and discuss relevance of these approaches in the deep learning-based ASR era.

Talk 2: This lecture will extend the abstract formulation to statistical ASR approach. In that direction, I will focus on Bayesian formulation of the ASR problem and will then largely deal with likelihood-based ASR approach, more precisely, hidden Markov model-based approach. I will dwell into different aspects such as, (a) different types of statistical estimators, (b) pronunciation modeling, and (c) end-to-end learning.

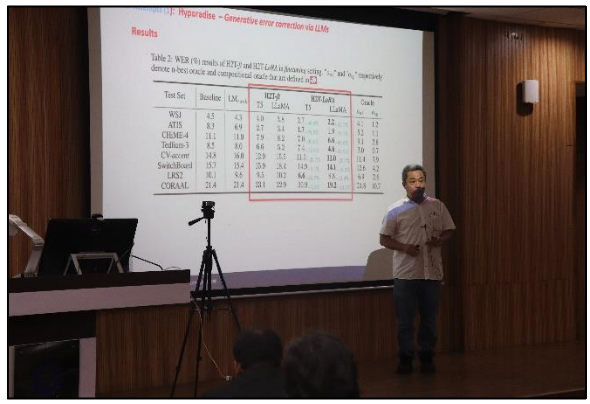
Talk 3: The third lecture will continue with the statistical ASR approach, where the focus will be on "posterior-based" approach. I will present an HMM-based approach where the HMM states are parameterized by categorical distributions. I will demonstrate how such an approach allows us (a) to overcome some of the limitations of conventional HMM-based approach (presented in Talk 2), (b) to unify instance-based approach and HMM-based approach, (c) to model different types of subword units and phonological representations, (d) to deal with data-scarcity issues, and (e) to holistically deal with speech recognition and speech assessment.



Prof. (Dr.) Chng Eng Siong

Talk1: The NTU Speech Teams Experience in Adapting Whisper

Whisper is a speech recognition model released by OpenAI at end of 2022. It is now one of the most impactful model based on transformer for the speech community. Whisper has been trained with 680K hours, is capable of transcribing speech from 96 language and converting these languages to English. Due to its open-sourced nature and SOTA (State of the art) performance, the speech-research community has widely adopted Whisper as a foundation model. Researchers have further enhanced it for various applications, including adaptation to accented, under resourced and code-switch speech, as well as for streaming and real time transcription. Advancing Automatic Speech Recognition with Whisper and Transformers In this work, we present our team's efforts from the Nanyang Technological University (NTU) to leverage the Whisper model and transformer architectures for enhancing automatic speech recognition (ASR) capabilities. Specifically, we focus on the following contribution: 1) Code-switch speech recognition – By fine-tuning and modifying the language prompts in Whisper, our team demonstrated the model's ability to perform code-switched transcription, achieving state-of-the-art results on the SEAME (South-East Asia Code-switch Mandarin English) corpus. Our experiment results show that Whisper can effectively handle code-switching between multiple languages within the same utterance. 2) Speaker aware decoding – The vanilla Whisper model is speaker-agnostic, designed to be robust against variations in speaker identity, accent, and noise. However, research indicates that recognition accuracy can be further improved with target speaker's information. For instance, using speaker adaptation data or speaker identity allows for fine-tuning or conditioning the model. In our previous work, we have demonstrated that incorporating a speaker-identity vector into the transformer encoder's key-value input makes the model speaker-aware. Experiments on the Libri Speech, Switchboard, and AISHELL-1 ASR tasks showed that our proposed model achieved relative word error rate (WER) reductions of 4.7% to 12.5%.



Talk2: Enabling LLM for ASR

The decoder only LLM such as ChatGPT was originally developed to only accept text as input. Recent advances have enabled it for other modalities: such as audio, video and images. Our focus in this talk is the integration of speech modality into LLM. For this task, the research community has proposed various innovative approaches: e.g, applying discrete representations, integrating pre-trained encoder to existing LLM decoder architectures (Qwen), multitask learning and multimodal pertaining. In the talk, I will a review the recent approaches of ASR task using LLM, and b) introduce 2 of our NTU's speech lab works for this task: i) "Hyporadise": Applying

LLM on N-best hypothesis generated by traditional ASR models to improve the top1 ASR transcription result. Our results show that LLM not only exceed the performance of traditional LM re-scoring, LLM can recover and generate correct words not found in the N-best hypothesis - we call such an ability GER (Generative Error Correction). ii) Leveraging LLMs for ASR and Noise-Robust ASR: In this work, we extend Hyporadise approach to include hypothesis (language) noise information into the LLM. Our insight is that under low SNR speech condition, there will be more diverse N-best hypothesis due to higher decoding uncertainty. This diversity can be captured and represented as an embedding vector called noisy language embedding. This embedding can then be exploited as a prompt. With fine-tuning on a training set, the LLM can be shown to have improve performance for the GER task.

Prof. (Dr.) Srikanth Madikeri

Talk 1: Loss functions for Training Automatic Speech Recognition Systems

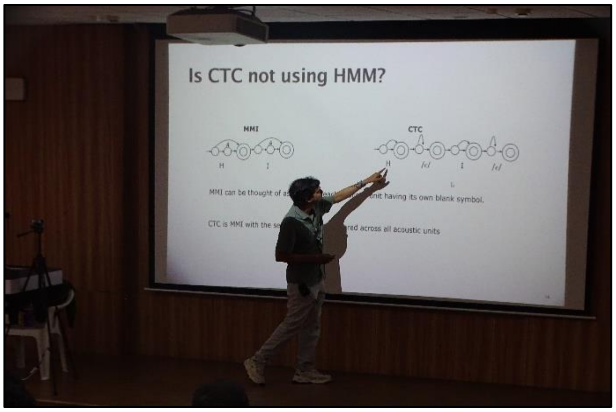
This talk will focus on commonly used loss functions to train neural-network based speech recognition systems in a supervised fashion. We will cover the fundamental cross-entropy loss, the popular connectionist temporal classification loss, and discriminative training with MMI. Finally, we will look at the most recent Transducer based approach to training ASR. The goal of this talk is to build on the content from other lectures, and understand the similarity and differences in the different approaches.

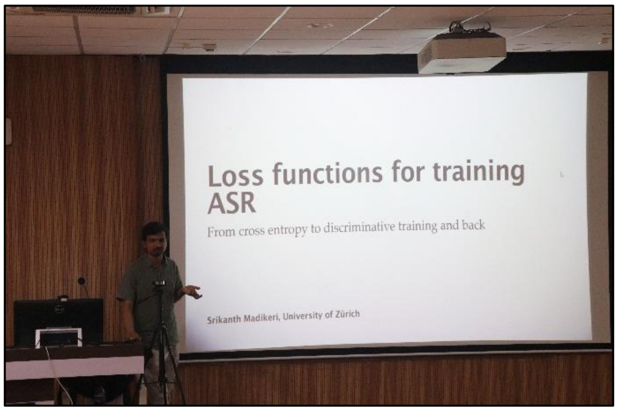
Talk 2: Publicly available open-source ASR models and their applications

Nowadays, the most common approach to train an ASR for a custom domain or a new language involves fine-tuning a bootstrap model trained from 1000s of hours to million hours of data in either supervised, weakly supervised, or self-supervised fashion. In this talk we will look at the different open-source options available for such bootstrapping. We will look at the commonly available options such as wav2vec 2.0, HuBERT, WavLM, etc., understand the architectural differences, advantages and limitations.

Talk 3: Training ASR with limited resources

In this talk we will look at different strategies, often complementary to each other, to leverage pre-trained acoustic models in conditions with limited resources. We will address situations that involve hardware and data constraints. First, we will look at semi-supervised learning for low-resource conditions. To address limited hardware constraints, we will introduce parameter efficient fine-tuning methods such as Low Rank Adapters and its variants.

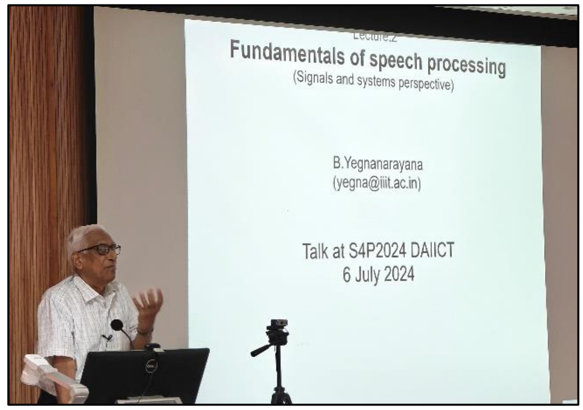




Prof. (Dr.) Bayya Yegnanarayana

Processing Phase of Speech Signals

In this somewhat provocative talk, I would like to discuss the need to process only the phase spectra of signals, in general, and of speech signals, in particular. I will show that phase spectrum has all the information, whereas the magnitude spectrum has limited information, which in principle can be derived from the phase spectrum. In order to understand and exploit the phase spectral information in signals, it is necessary to obtain the true phase without wrapping. Recently, I have proposed a method to obtain phase without the need for phase wrapping. I will show that many of the speech production features can be derived from the phase representation of the speech signal. It appears that speech information need to be represented only through the phase spectrum, rather than through the magnitude spectrum for most of the speech applications. I will give the necessary signal processing background to appreciate the points I will be making in this talk. Most of this work is not published yet, but I would like to take this opportunity to introduce these new ideas for the first time to the workshop audience.



Prof. (Dr.) C. V. Jawahar

Seeing is Listening

Understanding how humans perceive the signals around has been always fascinating. Traditionally computer vision and speech processing have identified themselves as areas with very less to share. Though there have been cognitive studies on the relationship between these two modalities of perception, computational approaches were very different. In recent years, we have been seeing more convergence in the computational methods. Far more, we are seeing the

emergence of audio-visual methods where one modality helps/catalyze the perception of other modality. In this talk, we discuss some of the recent works and directions.

Prof. (Dr.) Sriram Ganapathy

Beyond the Frame: Multi-Scale Self-Supervised Speech Representation Learning

This talk delves into the exciting field of self-supervised learning (SSL) for speech processing, specifically focusing on capturing the rich, multi-scale information embedded within speech signals. While conventional SSL approaches primarily target frame-level representations (20-30 ms), capturing semantic content, speech inherently encompasses information at various levels: utterance-level non-semantic cues and even recording session-specific channel/ambient characteristics. I will review key aspects of prior works on speech representation learning at frame and utterance levels that are prevalent in the field.

This talk will showcase our group's efforts in developing novel techniques for factorized representation learning across these multiple scales, leading to improved performance in various downstream speech processing tasks. The first part of the talk will introduce our approach to self-supervised representation learning directly from raw audio using a hidden unit clustering (HUC) framework. This computationally efficient method leverages convolutional neural networks (CNNs) for initial time-frequency representation extraction followed by processing with long short term memory (LSTM) layers. We'll delve into techniques employed to enhance speaker invariance in these learned representations. The efficacy of our approach will be demonstrated through its application in two distinct settings: completely unsupervised speech tasks within the ZeroSpeech 2021 challenge and semi-supervised automatic speech recognition (ASR) on the TIMIT and GramVaani challenge Hindi datasets. Notably, our method achieves state-of-the-art results for various Zero-speech tasks (as of 2023). The second part will shift focus to our recent "Learn2Diss" framework, designed for learning disentangled speech representations. We will discuss its architecture, comprising separate frame-level and utterance-level encoder modules, and detail the disentanglement process using a mutual information-based criterion. Through comprehensive evaluations on various downstream tasks, including those from the SUPERB challenge, we demonstrate the superior performance of Learn2Diss. Finally, we will touch upon related work in zero-shot emotion conversion and conclude by outlining future research avenues for these promising research streams.

Dr. Aparna Walanj

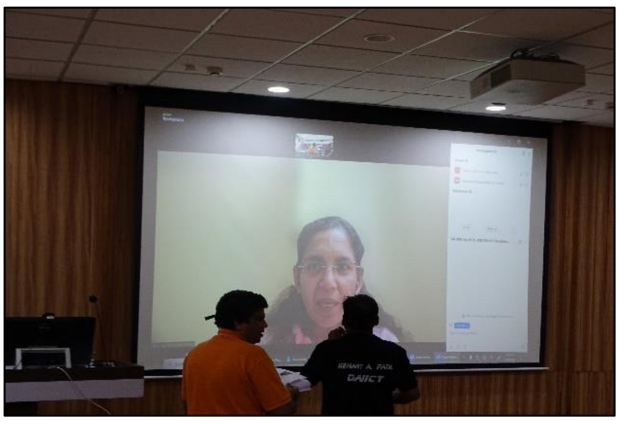
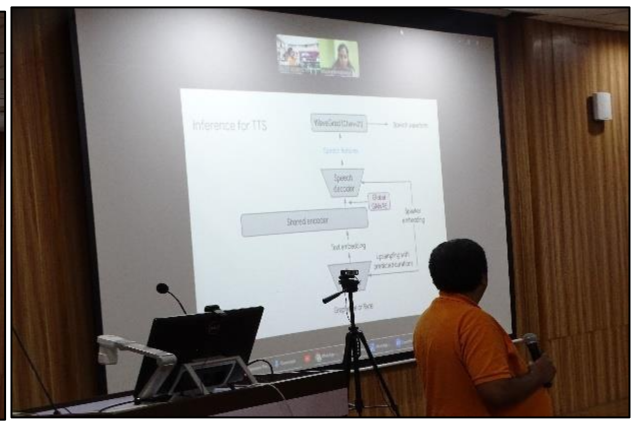
Ethics in Research

Ethics is an important component of any research, be it academic or clinical research. As the world today takes giant strides in science, technology and pioneering research, the credibility of the research community and the perception of the common man to accept new results firmly depends on the authenticity, accuracy and reliability of the results that have been published. It is important and crucial for researchers to be aligned and updated with the different guidelines and regulations to be followed when undertaking any research. This presentation will try to throw some light on the different guidelines in research, the role of Research Ethics Committees and will provide an insight on the process of submission of documents to the Ethics Committees.

Dr. Bhuvana Ramabhadran

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

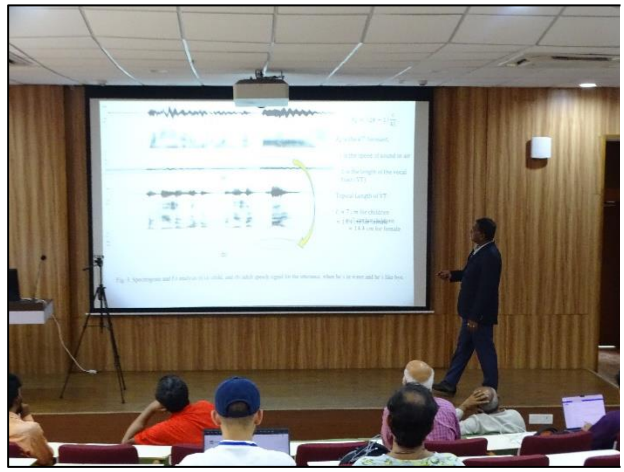
Speaker verification/spoof detection is a two-class problem. Conventionally, UBM GMM/*i*-vector systems are used. Scoring in the UBM-GMM framework uses top-C most contributing mixtures for each class. In this talk, Prof. Murthy focused on a variable top-C approach to scoring and then extended this approach 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

Dysarthric ASR: Assistive Speech Technology

Dysarthria is a speech disorder stemming from difficulties in controlling relevant muscles involved in natural speech production mechanism and thus, poses formidable challenges to dysarthric patients for effective communication. This disorder can happen due to various reasons, such as brain injury, brain tumour, stroke, and nervous system disorder including cerebral palsy, Parkinson's disease or Amyotrophic Lateral Sclerosis (ALS). To that effect, assistive technologies, such as dysarthric ASR can help to convert spoken words by the patients into text which can be easier for others to understand and thus, assist patients to communicate and participate in conversations. This talk will first present various challenges associated with processing of dysarthric speech, in particular, spectrographic *vs.* Linear Prediction (LP) analysis, shifts in formants and their – 3 dB bandwidths. Generally, formants are shifted to higher frequency region due to decrease in length of vocal tract system stemming from comprised contraction and relaxation of muscles for dysarthric patients. Further, as part ongoing efforts of the National Language Translation Mission (NLTM) consortium sponsored by MeitY, Govt. of India, the talk will review various dysarthric ASR systems reported in the literature including the recent work on noise robust whisper features using different classifier models, such as LSTM, BiLSTM, and BiGRU. Finally, talk will also discuss the significance of dysarthric severity-level classification system (as pre-processing) by invoking severity–specific ASR models to improve performance of dysarthric ASR system.



Industry Perspective Talks

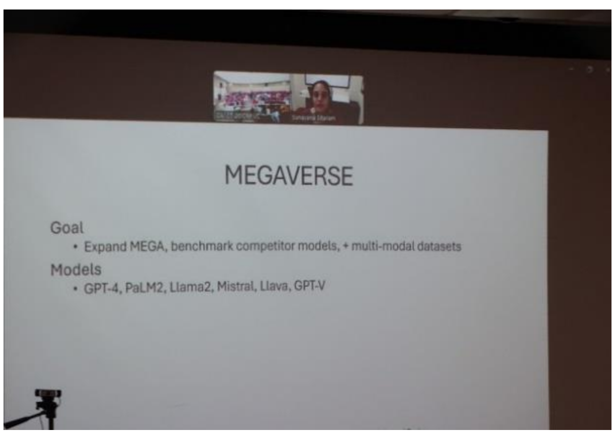
Dr. Tara N. Sainath

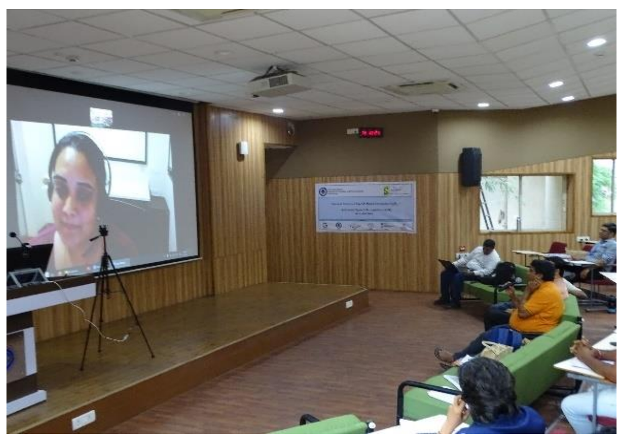
End-to-End Speech Recognition: The Journey from Research to Production

Dr. Sunayana Sitaram

Evaluating LLMs on Languages Beyond English: Challenges and Opportunities

The assessment of capabilities and limitations of Large Language Models (LLMs) through the lens of evaluation has emerged as a significant area of study. In this talk, I will discuss our research over the last 1.5 years on evaluating LLMs in a multilingual context, highlighting the lessons we learned and the general trends observed across various models. I will also discuss our recent efforts to evaluate Indic LLMs using a hybrid approach of human and LLM evaluators. Lastly, I will touch upon the challenges that remain in both advancing evaluation research and improving multilingual models.





Dr. Harish Arsikere

Bootstrapping ASR for new language

Dr. Vikram C. Mathad

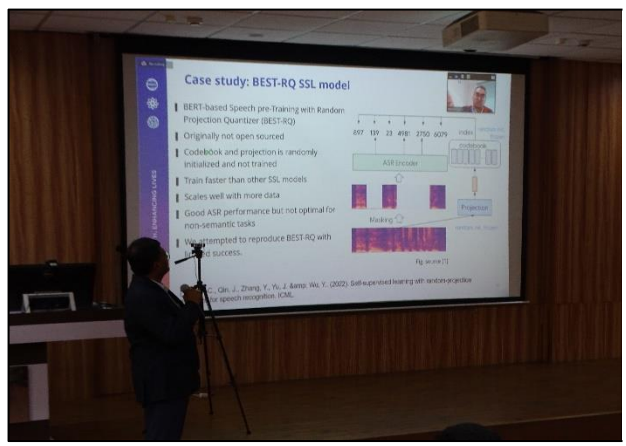
Speech AI: From Command Recognition to Live-Call Translation

ASR is most commonly used in smartphones voice-enabled applications such as Bixby, Alexa, Google's voice assistant. In these applications, ASR is used for command recognition purpose. Recent advances in ASR technology enables to extend the speech AI applications from the command-level to conversational speech recognition. Subsequently Live Call Translate, Interpreter, and Transcript Assist applications are introduced in the recent Samsung's Galaxy AI. To develop such applications several practical challenges like data preparation, background noise, multi-speaker, multi-lingual conditions, memory, and inference time need to be addressed. This talk gives an overview about incremental changes and the various challenges involved in the journey of command-based voice assistants to recent live call translation.

Dr. Hardik B. Sailor

Representation Learning for Speech: From Unimodal to Multimodal

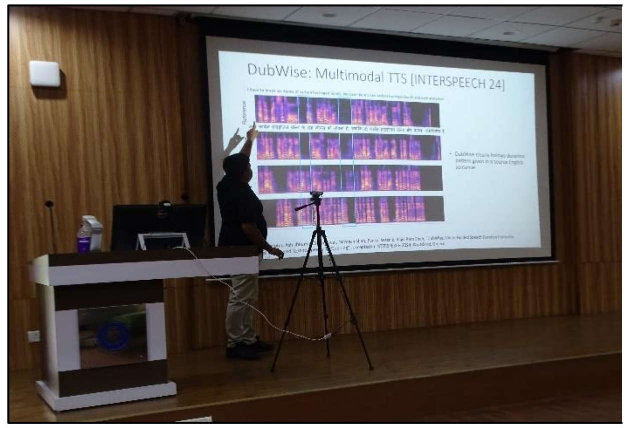
In this talk, Dr.Vikram Vij shared the journey that Samsung has undertaken in developing its Voice Assistant (Bixby) and particularly, Automatic Speech Recognition (ASR) system that powers it. Several independent components, such as pre-processors (acoustic echo cancellation, noise suppression, neural 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.



Dr. Nirmesh J. Shah

Evolution of Speech Foundation Models and Its Applications in Speech AI

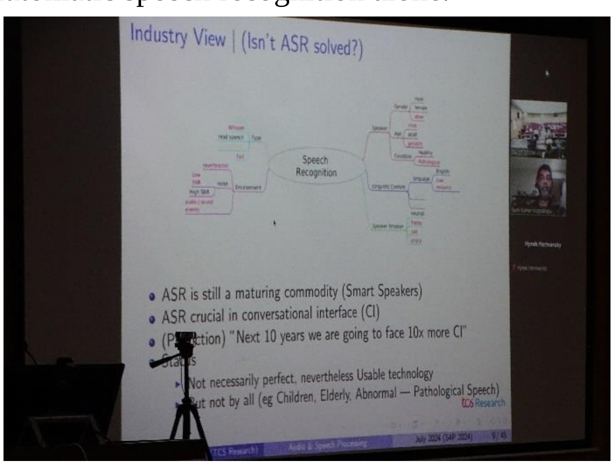
In the last couple of years, Large Language Models (LLM)s have brought remarkable advancements in the field of Generative AI, enabling machines to comprehend and generate human-like text. Hence, the research focus has shifted to developing foundation models for vision and speech modalities. In this talk, we will discuss various aspects of speech foundation models that are revolutionizing the landscape of speech recognition and synthesis in the future applications of Speech AI. We will specifically discuss the core principles and transformative potential of speech foundation models in Speech AI. We will explore how these models are pre-trained on vast amounts of speech data to learn contextually relevant speech representations and their potential applications, in the field of speech-to-text, text-to-speech synthesis, language translation, and more. Finally, we will examine the challenges and future directions in the development of Speech GPT, such as improving robustness to diverse accents and dialects, mitigating biases, and enhancing the ethical considerations surrounding AI-driven speech technologies.

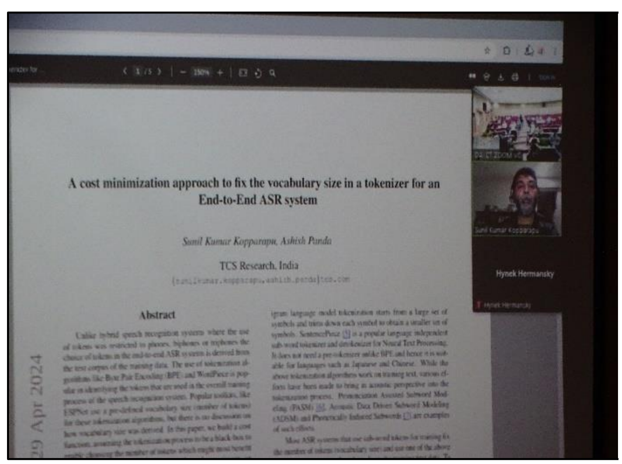


Dr. Sunil Kumar Kopparapu

Audio & Speech Processing | What have we been doing?

Availability of pre-trained acoustic models has narrowed the boundaries between a speech researcher & speech solution developer. Today it is not necessary to understand the fundamentals of speech production or speech perception to build meaningful speech solutions thanks to the wide availability of seemingly robust pre-trained model, be it for speech to text, text to speech or anything in between that might be required to build a solution to operate primarily on a speech signal. In this talk, we will dwell on some recent and a few ongoing activities in the Audio & Speech Processing team. To give a 360degree view of the things we do, we will not restrict to automatic speech recognition alone.





Mr. Amitabh Nag

Audio & Speech Processing | What have we been doing?

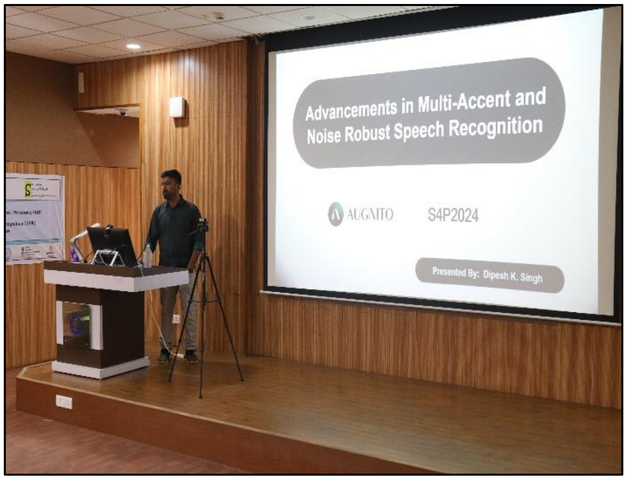
BHASHINI aims to transcend language barriers, ensuring that every citizen can effortlessly access digital services in their own language. Using voice as a medium, BHASHINI has the potential to bridge language as well as the digital divide. Launched by Honourable PM Shri Narendra Modi in July 2022 under the National Language Technology Mission, BHASHINI aims to provide technology translation services in 22 scheduled Indian languages. BHASHINI provides a full array of services to overcome language barriers and improve accessibility. It contains an easy-to-use web service portal, a mobile app in beta version, a dataset repository, a specialized services for Speech to Speech, Text to Text, Speech to Text, OCR, and Transliterations, a crowd-sourcing platform known as Bhasha Daan, which invites active data contributions for AI model training. This diverse strategy promotes inclusion, supports different languages, and encourages creativity in products and services.

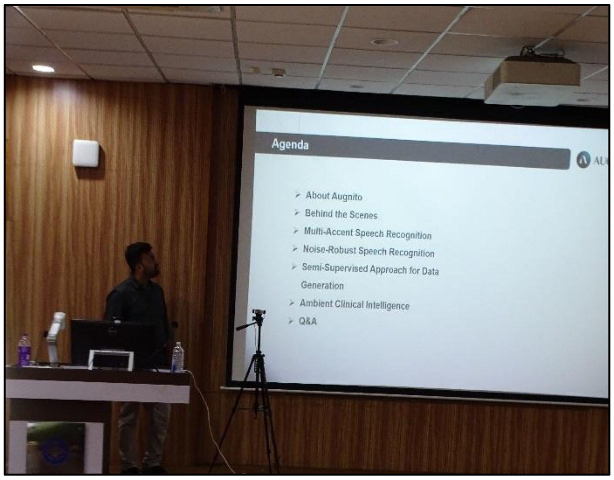
Mr. Dipesh K. Singh

Advancements in Multi-Accent and Noise Robust ASR Using Semi-Supervised Learning and Multimodal Approaches

In this talk, we explore cutting-edge techniques and advancements in the realm of ASR, focusing on multi-accent ASR, noise robustness, semi-supervised data generation, and multimodal integration. We begin by delving into the challenges posed by diverse accents and environmental noise in ASR. Leveraging recent developments in semi-supervised learning, we discuss novel approaches to efficiently generate labeled data and improve model performance across varying accent distributions. Next, we address the critical issue of noise robustness in ASR systems. Drawing on insights from recent research, we examine how multimodal approaches, integrating audio and contextual information, enhance the robustness of ASR models to noisy environments. We discuss methods such as noisy language embedding and multimodal pretraining, which enable

ASR systems to maintain accuracy even under challenging acoustic conditions. Furthermore, we explore the emerging field of ambient speech recognition, where the goal is to transcribe speech from everyday environments with high accuracy. We analyze recent advancements in this area, including the integration of large language models (LLMs) and the adaptation of transformer architectures for real-time ASR tasks. Finally, we reflect on the broader implications of these technological advancements, particularly in domains, such as ultrasound radiology report generation, where accurate and timely transcription of medical professionals' speech is crucial. By the end of the talk, attendees will gain a comprehensive understanding of the state-of-the-art techniques driving the evolution of ASR systems across diverse applications.



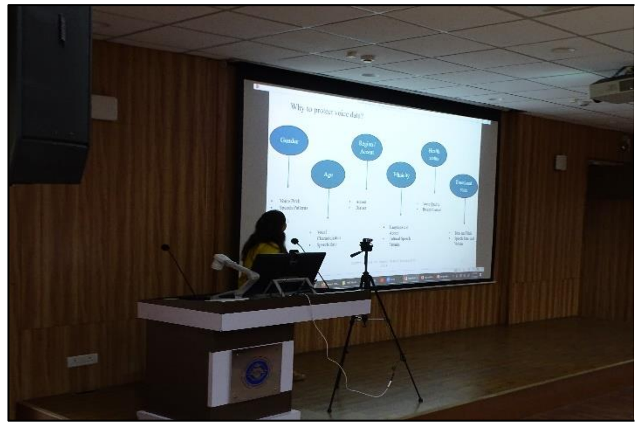


Ms. Gauri Prajapati

Voice Privacy in the Age of AI and Big Data

In the era of artificial intelligence (AI) and big data, the protection of voice privacy has become a critical issue. As voice-activated technologies and automatic speech recognition (ASR) systems proliferate, the risk of unauthorized access to and misuse of voice data has escalated. This talk will explore the multifaceted dimensions of voice privacy, addressing the unique challenges posed by voice data compared to other personal information. A significant focus will be on anonymization techniques for voice. We will delve into methods, such as voice synthesis, perturbation, and obfuscation that can effectively anonymize speech while maintaining its utility for applications, such as virtual assistants and automated transcription services. We will discuss the strengths and limitations of these techniques, and how they can be integrated into existing systems to protect individual privacy without compromising functionality.

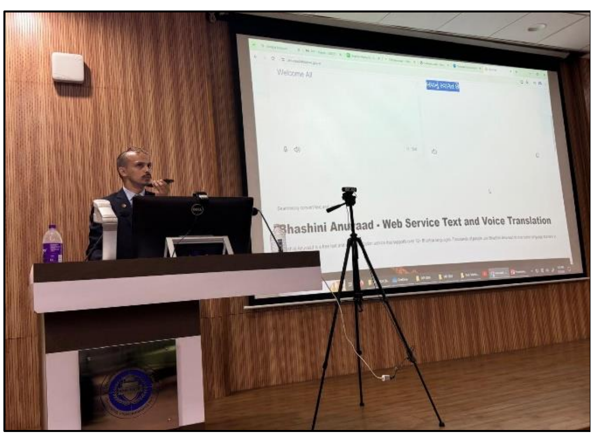
This talk will give insights into the balance between innovation and privacy, and the roles that individuals, organizations, and policymakers must play to protect voice privacy in an increasingly connected world.

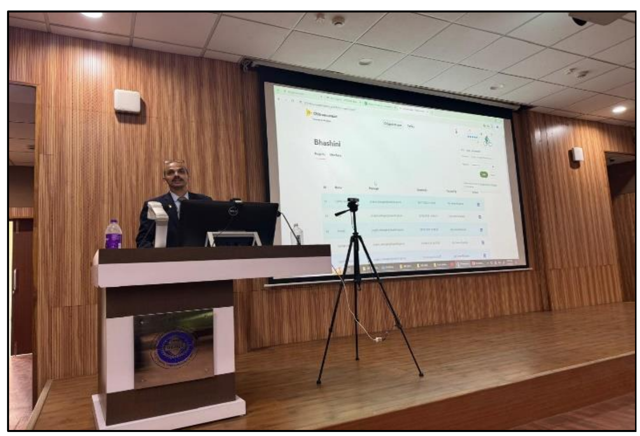


Mr. Ajay Rajawat

Audio & Speech Processing | What have we been doing?

BHASHINI aims to transcend language barriers, ensuring that every citizen can effortlessly access digital services in their own language. Using voice as a medium, BHASHINI has the potential to bridge language as well as the digital divide. Launched by Hon'ble PM Shri. Narendra Modi in July 2022 under the National Language Technology Mission, BHASHINI aims to provide technology translation services in 22 scheduled Indian languages. BHASHINI provides a full array of services to overcome language barriers and improve accessibility. It contains an easy-to-use web service portal, a mobile app in beta version, a dataset repository, a specialized services for Speech to Speech, Text to Text, Speech to Text, OCR, and Transliterations, a crowd-sourcing platform known as Bhasha Daan, which invites active data contributions for AI model training. This diverse strategy promotes inclusion, supports different languages, and encourages creativity in products and services.





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.

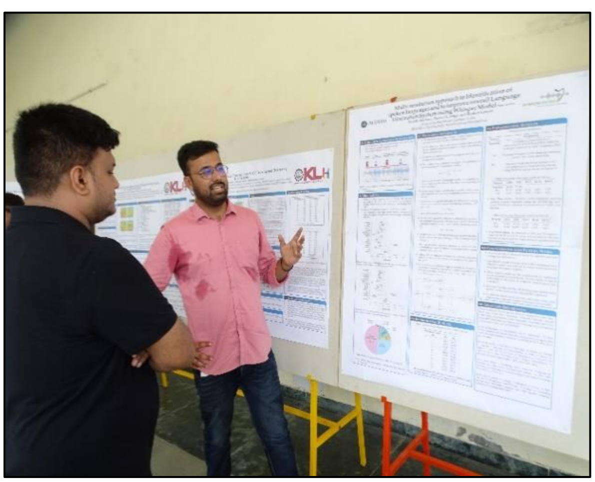
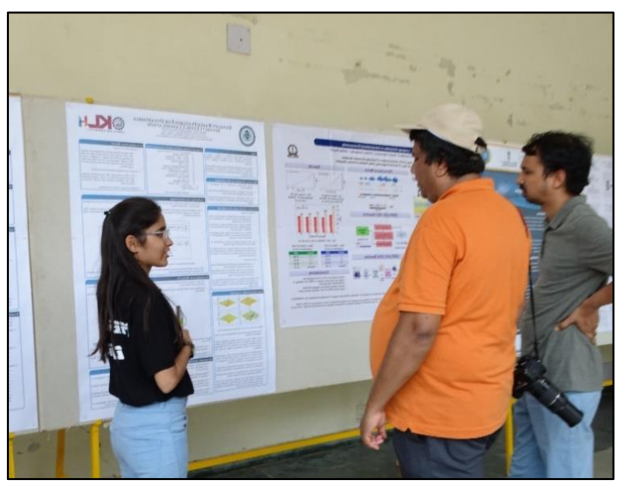

Google


53 48.5 43.7 27.6
FT (30 epochs after T, 25% param
ng significantly improve

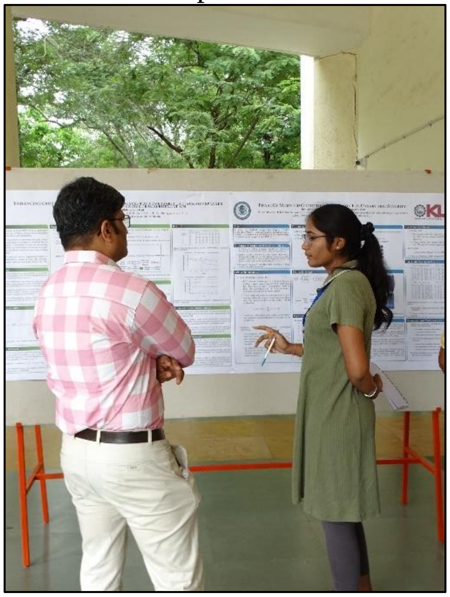
On-spot Poster Presentation:

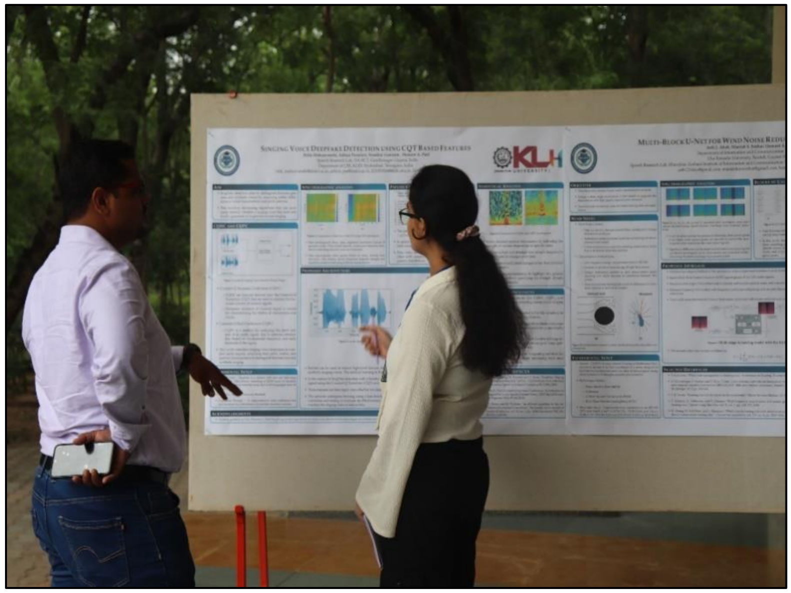


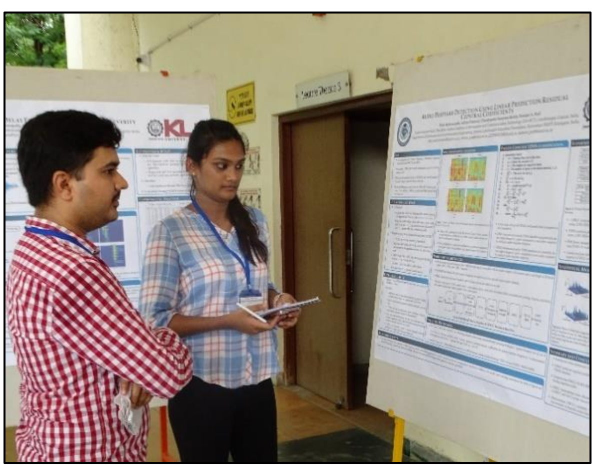


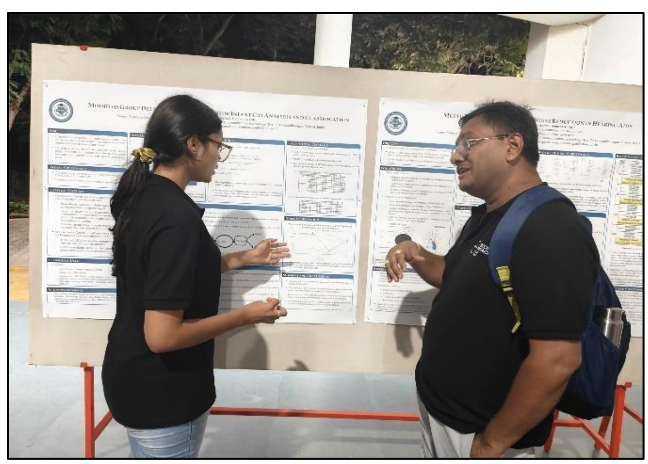




- We organized on-spot poster presentation session during Tea/Lunch break for all the five days of the S4P 2024. The key motivation of the session is to encourage the participants of summer school to present their ongoing research/published papers and get review feedback from experts and attendees of the event.









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, S4P 2016, and S4P 2019 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)

(b)

(c)

(d)

(a) **Ms. Gauri Deshpande**, University of Augsburg, Germany.
(b) **Ms. Ankita**, NIT Patna, India.
(c) **Ms. Akansha Tyagi**, IIT Mandi, India.
(d) **Mr. Swapnil Fadte**, Discipline of Computer Science & Technology, Goa University, Goa, India.

The details are given at https://sites.google.com/view/s4p2024/5-mpt

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) Amitabh Nag, Ministry of Electronics & Information Technology (MeitY), New Delhi, India.
(b) Mathew Magimai Doss, IDIAP Research Institute, Martigny, Switzerland.

(c) Vikram C. Mathad, Samsung Research Institute, Bengaluru, India.

(d) Dr. Samudravijaya K, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India.

Note: Dr. Ajay Rajawat (BHASHINI Division, Meity, New Delhi) 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, **four** scholars were awarded **ISCA endorsed** cash prizes.


(a) (b)

Expert Committee members to evaluate 5 Min Ph.D. Thesis contest.

Important: To avoid conflicts of professional interest, Prof. (Dr.) Hemant A. Patil (Chair of Organizing Committee S4P-2024) 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 were awarded **ISCA endorsed** First Prize of Rs. 15,000 /-, second prize of Rs. 10,000 /-, third Prize of Rs. 5,000 /-, and fourth Prize of Rs. 3,000 /-respectively,

1st Prize Winner Ms. Gauri Deshpande (University of Augsburg, Germany) and now at TCS, Mumbai

2nd Prize Winner Ms. Ankita. (NIT Patna)

3rd Prize Winner Ms. Akansha Tyagi (IIT Mandi)

4th Prize Winner Mr. Swapnil Fadte (Discipline of Computer Science & Technology, Goa University)

Sponsors Presentation

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

Prof. Hemant A. Patil presenting summary of R&D achievements on behalf of DA-IICT Gandhinagar.

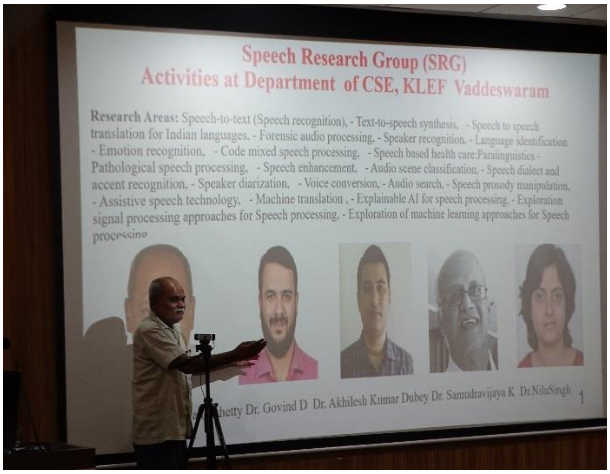
2. indSCA (http://www.apsipa.org/)

Prof. Hemant A. Patil presenting about indSCA, Behalf of Prof. Chandra Sekhar Seelamantula (IISc, Bengaluru).

3. BHASINI (http://www.apsipa.org/)

Mr. Amitabh Nag (CEO, BHASHINI Division) presenting about BHASINI achievement.

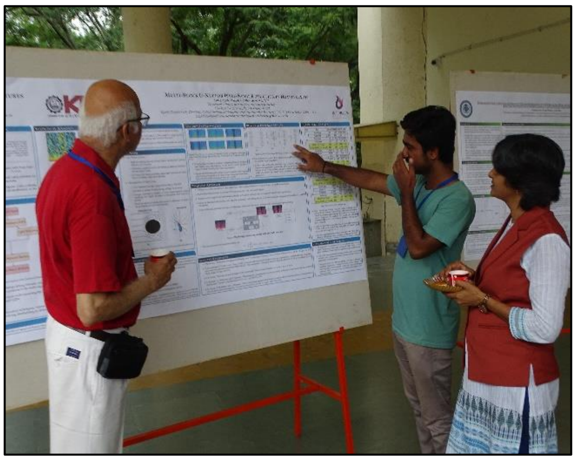
4. KLEF,Vaddeswaram
 (https://www.kluniversity.in/)



Prof. (Dr.) Suryakanth Gangashetty presenting about KLEF.

Interaction between Participants and Speakers during Tea/Coffee and Lunch Breaks

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 2024, DA-IICT Gandhinagar

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

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 2024.

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 2024 at DA-IICT Gandhinagar would not have been possible. The students and staff members from DA-IICT Gandhinagar who have worked as volunteers are

Speech Research Lab - DAIICT	**Admin & Help-Desk Member**
▪ Ravindrakumar Purohit ▪ Dharmendra H. Vaghera ▪ PSS Aditya ▪ Arushi Srivastava ▪ Arth J. Shah ▪ Hiya Chaudhari ▪ Kavya Kumar ▪ Ritik Mahyavanshi ▪ Dhrupal Narendra Kukadiya ▪ Dharmi Patel ▪ Nandini Mandaviya ▪ Manishkumar Suthar ▪ Savita Hiralal Yadav ▪ Atla Bhuvanika ▪ Samyana Reddy Chandupatla ▪ A. Meghana ▪ Nandini Gunnam ▪ Mahesh Reddy ▪ Polisetti Venkata Sarath Bhushan ▪ Alla Ujwal Sai ▪ K. Swapna ▪ I. Geeta Sai Sahasra ▪ Rohini Sri Mannepalli ▪ S. Sudrashan	▪ Kirit Pandya ▪ Santosh Pandit ▪ Jaydeep Panchal ▪ Anuradha Srivastava ▪ 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 ▪ Ramesh Prajapati ▪ Dinesh Prajapati ▪ Abhilash Bhaskaran ▪ Jainik Patel ▪ Chaitanya Bhamare ▪ Ashvin Chaudhari ▪ Darshan Parajapati ▪ Priyank Santola ▪ Ankit Sharma ▪ Hemangi Sharma ▪ Dhruti Joshi ▪ Saurabh Nayee ▪ Hoursekeeping Staff and Security Guards

Call For Papers:

INTERSPEECH 2024

INTERSPEECH 2024

Call for Papers

Important Dates
Paper Submission Portal Open: 20 January 2024
Paper Submission Deadline: 02 March 2024
Paper Update Deadline: 11 March 2024
Paper Acceptance Notification: 06 June 2024

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.

Interspeech 2024 will feature oral and poster sessions, plenary talks by internationally renowned experts, tutorials, special sessions and challenges, show & tell, exhibits, and satellite events.

The theme of Interspeech 2024 is ***Speech and Beyond***. For decades, our focus has been on enhancing speech technologies across various dimensions: spontaneous speech, large vocabulary, different conditions, robustness in adverse acoustic environments, multiple languages, speaker verification, and language identification. The organizing committee of Interspeech 2024 is committed to the advancement of speech technology in the wider sense while seeking new challenges. As technology evolves, we are ready to embrace the next set of challenges and explore new application domains. A non-exhaustive list of topics that goes beyond the traditional Interspeech topics includes: speech and health, animal voice recognition and understanding, speech for memory and heritage, voice communication across ages, and human-machine interaction, including gaming, virtual and augmented reality, and robot audition. We hope you will join our vision on "Speech and Beyond".

Paper Submission

Interspeech 2024 seeks original and innovative papers covering all aspects of speech science and technology. The language of the conference is English, so papers must be written in English. The paper length is up to four pages in two columns with an additional page for references and acknowledgments only. Submitted papers must conform to the format defined in the paper kit provided on the conference website (and as an overleaf template) and may optionally be accompanied by multimedia files. Interspeech 2024 will follow a double-blind review process so papers submitted for review should not reveal the identity or affiliation of the authors. Authors must declare that their contributions are original and that they have not submitted their papers elsewhere for publication. Papers must be submitted electronically and will be evaluated through rigorous peer review on the basis of novelty and originality, technical correctness, clarity of presentation, key strengths, and quality of references. The Technical Programme Committee will decide which papers to include in the conference programme using peer review as the primary criterion, with secondary criteria of addressing the conference theme, and diversity across the programme as a whole.

BLUE SKY TRACK

This year, we also encourage the authors to consider submitting to the new **BLUE SKY** track of highly innovative papers with strong theoretical or conceptual justification in fields or directions

EUSIPCO 2024

Call for Papers

On behalf of the European Association for Signal Processing (EURASIP), it is a great pleasure of the organizing committee to invite you to the 32nd European Signal Processing Conference, EUSIPCO 2024, to be held in Lyon, France, August, 26th-30th, 2024.

EUSIPCO is the flagship conference of EURASIP and offers a comprehensive technical program addressing all the latest developments in research and technology for signal processing.

EUSIPCO 2024 will feature world class speakers, oral and poster sessions, plenaries, exhibitions, demonstrations, tutorials, and satellite workshops, and is expected to attract many leading academic researchers and people from industry from all over the world.

EUSIPCO 2024 will also have a specific flavour with focus on interdisciplinary research geared toward major societal challenges. This will be highlighted by keynote speakers and special sessions.

Key Dates for the call for papers
Full Paper submission: March 3rd 2024
Notification of Acceptance: May 22nd 2024
Camera ready Paper submission: June 1st 2024

The technical scope of the conference is listed below.

Calls for special sessions, tutorial and 3-Minute Thesis,
with key dates are available in the next pages.

Accepted papers will be included in IEEE Xplore©.
EURASIP enforces a “no-show” policy.

Procedures to submit papers, proposals for special sessions, tutorials and satellite workshops can be found on the website.

TECHNICAL SCOPE
We invite the submission of original, unpublished technical papers on topics including but not limited to:

- Audio and acoustic signal processing
- Speech and language processing
- Image and video processing
- Multimedia signal processing
- Signal processing theory and methods
- Sensor array and multichannel signal processing
- Signal processing for communications
- Radar and sonar signal processing
- Signal processing over graphs and networks
- Nonlinear signal processing

ICPR 2024

ICPR 2024 INDIA

27TH International Conference on Pattern Recognition
December 01-05, 2024, Kolkata, India

CALL FOR PAPERS

General Chairs
Umapada Pal, India
Josef Kittler, UK
Anil Jain, USA

Program Chairs
Rama Chellappa, USA
Apostolos Antonacopoulos, UK
Cheng-Lin Liu, China
Subhasis Chaudhuri, India

Workshop Chairs
P. Shivakumara, UK
Stephanie Schuckers, USA
Jean-Marc Ogier, France
Prabir Bhattacharya, Canada

Tutorial Chairs
B. B. Chaudhuri, India
Guoying Zhao, Finland
Michael R. Jenkin, Canada

Competition Chairs
Richard Zanibbi, USA
Lianwen Jin, China
L. Likforman-Sulem, France

Doctoral Consortium Chairs
Daniel Lopresti, USA
Véronique Eglin, France
Mayank Vatsa, India

Publicity Chairs
Dipti Prasad Mukherjee, India
Bob Fisher, UK
Xiaojun Wu, China

Publications Chairs
Wataru Ohyama, Japan
Ananda S. Chowdhury, India

Awards Committee Chair
Arpan Pal, India

International Liaison / Visa Chairs
Balasubramanian Raman
Yue Lu, China

Finance Chairs
Kaushik Roy, India
Michael Blumenstein, Australia

Organizing Chairs
Saumik Bhattacharya, India
Palash Ghosal, India
Sk Md Obaidullah, India

The International Conference on Pattern Recognition (ICPR) is the flagship conference of the International Association of Pattern Recognition (IAPR) and the premier conference in pattern recognition, covering computer vision, image, speech and video processing, machine intelligence, and other related areas. It is a 5-day event that comprises the main conference, Workshops, Tutorials, different Competitions, Doctoral Consortium etc. ICPR-2024 is the 27th event of the series and it provides a great opportunity to nurture new ideas and collaborations for students, academics and industry researchers.

Main Topics of Interest

ICPR-2024 has 6 tracks as follows:

- Artificial Intelligence, Machine Learning for Pattern Analysis
- Computer and Robot Vision
- Image, Speech, Signal and Video Processing
- Biometrics and Human Computer Interaction
- Document Analysis and Recognition
- Biomedical Imaging and Bioinformatics

Important Dates

- Abstract Submission: April 10, 2024
- Full Paper submission: ~~March 20, 2024~~ April 17, 2024 (Extended)
- Reviews sent to authors: June 20, 2024
- Revision/ Author rebuttal deadline: July 10, 2024
- Acceptance notification: August 5, 2024
- Camera-ready submission: August 31, 2024
- Conference: December 1-5, 2024

Submission and Review

ICPR-2024 will follow a single-blind review process. Authors can include their names and affiliations in the manuscript.

Paper Format and Length

Springer LNCS format with maximum 15 pages (including references) during paper submission. To take care of reviewers' comments, one more page is allowed (without any charge) during revised/camera ready submission. Moreover, authors may purchase upto 2 extra pages. Extra page charges must be paid at the time of registration.

Contact: For any enquiry please contact the ICPR-2024 Secretariat via email at icpr2024@gmail.com and icpr2024@isical.ac.in

Track Chairs

Track 1: Artificial Intelligence, Machine Learning for Pattern Analysis
Larry O'Gorman, USA
Petia Radeva, Spain
Sushmita Mitra, India
Dacheng Tao, Australia
Jiliang Tang, USA

Track 2: Computer and Robot Vision
Maja Pantic, UK
C. V. Jawahar, India
João Paulo Papa, Brazil
Gang Hua, USA
Junwei Han, China

Track 3: Image, Speech, Signal and Video processing
P. K. Biswas, India
Shang-Hong Lai, Taiwan
Hugo Jair Escalante, Mexico
Sergio Escalera, Spain
Prem Natarajan, USA

Track 4 : Biometrics and Human Computer Interaction
Massimo Tistarelli, Italy
Wei-Shi Zheng, China
Richa Singh, India
Vishal Patel, USA
Jian Wang, USA

Track 5: Document Analysis and Recognition
Xiang Bai, China
Josep Llados, Spain
Mita Nasipuri, India
David Doermann, USA

Track 6: Biomedical Imaging and Bioinformatics
Xiaoyi Jiang, Germany
Seong-Whan Lee, Korea
J. Mukhopadhayaya, India

Women in ICPR Chairs
Ingela Nyström, Sweden
Alexandra B. Albu, Canada
Jing Dong, China
Sarbani Palit, India

Sponsorship Chairs
P. J. Narayanan, India
Yasushi Yagi, Japan
Venu Govindaraju, USA
Alberto Del Bimbo, Italy

ORGANIZING / TECHNICAL PARTNERS

www.icpr2024.org
icpr2024@gmail.com / icpr2024@isical.ac.in

APSIPA ASC 2024

ICASSP 2025

2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2025)

April 06 – 11, 2025

Hyderabad, India

IEEE Signal Processing Society

The 50th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) will be held in Hyderabad, India, from April 06 to 11, 2025, at the Hyderabad International Convention Centre. The flagship conference of the IEEE Signal Processing Society will offer a comprehensive technical program presenting the latest developments in signal processing research and applications. Featuring world-class oral and poster sessions, keynotes, exhibitions, demonstrations, tutorials, and short courses, the conference will attract leading researchers and global industry figures, providing great networking opportunities. Exceptional papers and contributors will be selected and recognized at ICASSP.

General Chairs
K.V.S. Hari (IISc, India)
V John Mathews (Oregon State Univ., USA)
Technical Program Chairs
Bhaskar D Rao (UCSD, USA)
Isabel Trancoso (Univ. of Lisbon, Portugal)
Gaurav Sharma, (Univ. of Rochester, USA)
Neelesh B. Mehta (IISc, India)
Plenary Chairs
P P Vaidyanathan (Caltech, USA)
T V Sreenivas (IISc, India)
Bhuvana Ramabhadran (Google, USA)
Special Session Chairs
A Lee Swindlehurst (UCI, USA)
Subhasis Chaudhuri (IIT Bombay, India)
Tutorial Chairs
Geert Leus (Delft Univ., Netherlands)
Chandra R Murthy (IISc, India)
Women in Signal Processing Chairs
Namrata Vaswani (Iowa State Univ., USA)
Educational Activities Chairs
Mrityunjoy Chakraborty (IIT Kgp., India)
Sumam David (NITK, India)
Industry Program & Exhibition Chairs
Ken Sugiyama (Yahoo, Japan)
N Venkatesh (Silicon Labs, India)
Local Organization Chairs
Chandra Sekhar Seelamantula (IISc, India)
Sanjeev Nimishakavi (Qualcomm, India)
Publication Chairs
R Venkatesh Babu (IISc, India)
Sethu Selvi (MSRIT, India)
Finance Chairs
Rajiv Soundararajan (IISc, India)
Soma Biswas (IISc, India)
Publicity Chairs
Angshul Majumdar (IIIT Delhi, India)
Supavadee Aramvith (Chulalongkorn Univ., Thailand)
Grand Challenge Chairs
Prasanta Ghosh (IISc, India)
Anubha Gupta (IIIT Delhi, India)
Sundeep Chepuri (IISc, India)
Ajit Bopardikar (U of Houston, USA)
International Liaisons
Homer Chen (National Taiwan Univ., Taiwan)
Björn Ottersten (Univ. of Luxembourg)
Vaughan Clarkson (Whipbird Signals, Australia)
José Bermudez (Fed. Univ. Santa Catarina, Brazil)
Raghuveer Rao (Army Research Lab, USA)

Technical Scope: We invite submissions of original, unpublished technical papers on topics including but not limited to:

- Applied Signal Processing Systems
- Audio & Acoustic Signal Processing
- Biomedical Imaging & Signal Processing
- Compressive Sensing, Sparse Modeling
- Computational Imaging
- Computer Vision
- Deep Learning/Machine Learning
- Image, Video & Multidimensional Signal Processing
- Industrial Signal Processing
- Information Forensics & Security
- Internet of Things
- Multimedia Signal Processing
- Quantum Signal Processing
- Remote Sensing & Signal Processing
- Sensor Array & Multichannel SP
- Signal Processing for Big Data
- Signal Processing for Communication
- Signal Processing for Cyber Security
- Signal Processing Education
- Signal Processing for Robotics
- Signal Processing Over Graphs
- Signal Processing Theory & Methods
- Speech & Language Processing

Location: Hyderabad is the capital of the Telangana state in southern India and has a unique and rich multicultural history. Its historic sites include the Golconda Fort, a former diamond-trading centre that was once the capital of the Qutb Shahi dynasty. The Charminar, a 16th-century monument whose 4 arches support towering minarets, is an old city landmark. A major center for the technology industry, Hyderabad is home to most of the global technology giants. The city is known for its uniquely delightful cuisine that includes the famous Hyderabadi biryani!

SP Society Journal Paper Presentations: Authors of papers published or accepted in IEEE SPS journals may present their work at ICASSP 2025 in appropriate tracks. These papers will not be re-reviewed or included in the proceedings. In addition, the IEEE Open Journal of Signal Processing (OJSP) will provide a special track for longer submissions with the same processing timeline as ICASSP. Accepted papers will be published in OJSP and presented at the conference but will not be included in the conference proceedings.

Open Preview: Conference proceedings will be available in IEEE Xplore, free of charge, to everyone, 30 days prior to the conference start date, through the conference end date.

Other Calls: Calls for proposals for Tutorials, Special Sessions, Workshops and Short Courses along with the Prospectus for Patrons and Exhibitors will be announced on 2025.ieeeicassp.org/calls.

Important Dates:

Special Session Proposals	July 8, 2024
Grand Challenge Proposals	July 8, 2024
Workshop Proposals	July 8, 2024
Tutorial Proposals	September 23, 2024
Short Course Proposals	September 23, 2024
Show and Tell Proposals	January 7, 2025
Paper submissions	September 9, 2024
Author response Period	November 15-26, 2024
SPS Journal Papers/Letters Deadline	December 11, 2024
Paper Acceptance	December 18, 2024
Camera Ready Paper Deadline	January 13, 2025
Author Registration	January 13, 2025
Open Preview Starts	March 7, 2025

Scan Me for More Details

2025.ieeeicassp.org

Note: We have circulated CFP for INTERSPEECH 2024, EUSIPCO 2024, ICPR 2024, APSIPA ASC 2024, ICASSP 2025 in the participant's kits of S4P 2024.