



Publication Partner



Supported by



(Centre of Excellence – AI,  
IGDTUW)

## **3<sup>rd</sup> INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND SPEECH TECHNOLOGY (AIST-2021)**

**12<sup>th</sup>-13<sup>th</sup> November, 2021**

### **Conference Report**

Indira Gandhi Delhi Technical University for Women organized the third version of the International Conference on Artificial intelligence and Speech Technology (AIST-2021). The event was held on 12th and 13th November 2021 which witnessed a great success and guided by many eminent guests and speakers from International levels such as Japan, Germany, France, Australia and others. Total 55 papers were presented by researchers. Event was attended by more than 200 students of UG, PG and Research Scholars from IGDTUW.

The inauguration of the event was held in the prestigious presence of Honorable Vice Chancellor Dr. (Mrs.) Amita Dev and the Chief Guest Shri. Parimal Kumar, Scientific Advisor to Chief of Integrated Defence Staff, Ministry of Defence, GOI. The honorable Vice-Chancellor of IGDTUW, being an active researcher in Speech technology herself, addressed the crowd and spoke about the importance of AI and Speech technology in subsequent times, and emphasized the reliability of AI and ST in our day-to-day lives. The Chief Guest of the event, Shri. Parimal Kumar, Scientific Advisor to Chief of Integrated Defence Staff, Ministry of Defence, GOI conveyed ample applause to all the esteemed researchers and congratulated them for their generous efforts in the area of Artificial Intelligence and Speech Technology. He concluded by encouraging future graduates and students of IGDTUW to become capable enough to generate occupation and undertake the enterprise.

After the Inauguration session, renowned academicians and researchers delivered their Invited talks starting with Dr. Rajkumar Buyya, Professor, University of Melbourne, Australia. He discussed the latest developments and trends in Cloud computing and IoT embedded with Artificial Intelligence and Machine Learning.

Afterwards for the next talk Dr. Satoshi Nakamura, Honorary Chair AIST 2021, from NAIST, Japan, discussed his research on “Recent Advances in Speech-to-Speech Translation”. Sir discussed “Speech translation” technology to translate a source language speech to a target language speech, which includes speech recognition, machine translation, and speech synthesis. The research and development of this technology was launched in the 1980's, and the great effort and hard work of many researchers brought the “speech translation” to the current stage of practical use.

However, what people expect from the real-time speech translation is the “interpretation”. The human interpreters simultaneously and incrementally listen, understand, and produce interpreted utterances considering word order differences between source and target languages. In this talk, he introduced the recent advances in speech-to-speech translation towards automatic speech interpretation at their team including incremental speech recognition, machine translation, and speech synthesis for distant language pairs like English and Japanese.

For the next talk we had Dr. Tanja Schultz, Professor, University of Bremen, Germany. Her talk revolves around “Beyond Acoustic Speech Communication”, where she explained the relation between Speech and biosignals. These biosignals – stemming from the articulators, the articulator muscle activities, the neural pathways, and the brain itself – can be used to circumvent limitations of conventional speech processing in particular, and to gain insights into the process of speech production in general. She also presented her ongoing research at the Cognitive Systems Lab (CSL), where she investigated a range of speech-related biosignals beyond acoustics, such as muscle and brain activities to establish alternative communication interfaces. Based on machine learning methods her team implement biosignal-based speech processing systems and devices for communication applications in everyday situations and for speech rehabilitation. Several applications described such as Silent Speech Interfaces that rely on articulatory muscle movement captured by electromyography to recognize and synthesize silently produced speech, Brain-to-text interfaces that use brain activity captured by electrocorticography to recognize speech, and Brain-to-Speech interfaces that synthesize speech with low latency to close the interaction loop.

Day 1 of the conference ended with the talk of Dr. Laurent Besacier, Principal Scientist, Naver Labs Europe, France. He discussed “Self-Supervised Learning for Low Resource Speech Tasks”. Self-supervised learning using huge unlabeled data has been successfully explored for image processing and natural language processing. Since 2019, recent works also investigated self-supervised representation learning from speech. They were notably successful to improve

performance on downstream tasks such as speech recognition. These recent works suggest that it is possible to reduce dependence on labeled data for building speech systems through acoustic representation learning. In this talk he gives an overview of these recent approaches to self-supervised learning from speech and shows his own investigations to use them in spoken language processing tasks for which size of training data is limited.

Day 2 started with the Keynote address of Dr. S Sakti, Professor, from JAIST/NAIST, Japan. The title of the talk is “Listening while Speaking and Visualizing: A Semi-supervised Approach with Multimodal Machine Speech Chain” where she introduces a semi-supervised learning mechanism based on a machine speech chain framework. First, it describes the primary machine speech chain architecture that learns not only to listen or speak but also to listen while speaking. The framework enables ASR and TTS to teach each other given unpaired data. Then, the recent multimodal machine chain framework that mimics overall human communication to listen while speaking and visualizing is discussed. With the support of image captioning and production models, the framework further reduces the need for a large amount of unpaired data. It enables ASR and TTS to improve their performance using an image-only dataset.

And the last keynote address of the conference is headed by Prof. S. Umesh, Professor, from IIT Madras. In his talk sir discussed about their team efforts to build speech recognition systems in Indian languages. In particular, they have recently developed the idea of a common label set for Indian languages which helps pool speech data and address the problems of resource scarcity. He also talked about recent efforts to make technical lectures available in Indian languages by automatic transcription and translation as a part of the National Language translation Mission.

Along with these keynote address authors presented their research papers in between 7 defined and specialised tracks.

The event came to its periphery with a valedictory address by Dr Anupam Shukla, Director, IIIT Pune, who exhorted the researchers to develop low cost technological solutions with enthusiasm for research and high impulse. Dr. Shyam Sundar Agrawal, Emeritus Scientist and Chair of the Technical Program Committee and Professor Arun Sharma, Conference Secretary and Head of the Department of Artificial Intelligence and Data Science, IGDTUW contributed significantly to make this event a success.

A handwritten signature in blue ink, appearing to read "Arun Sharma", is located in the bottom right corner of the page. The signature is written in a cursive style with a horizontal line underneath.