IEEE SLT 2024 Program Details

Day 1, Dec 2, Monday

08:30-09:00 Opening Session (Venue: Lecture Hall)

09:00-10:00 Keynote Speech 1 (Venue: Lecture Hall)

Title: Towards Robust Audio Deepfake Detection and Attribution
Speaker: *Prof Jianhua Tao, Tsinghua University*
Chair: *Dr Minghui Dong*

10:00-10:30 Coffee Break

10:30-12:30 Poster Session 1: Speech Recognition (Venue: Poster Area)

Chair: Prof Yanmin Qian

Poster ID (Paper ID)	Title and Authors
P1-01-ASR (#28)	**PromptKWS: A Novel Prompt-Guided Open-Vocabulary Keyword Spotting Framework** *Gaopeng Xu (NIO)* *Chengfei Li (Qilu Normal University)* *Xianliang Wang (NIO)* *Li Zhu (NIO)* *Juan Wei (NIO)* *Wenpeng Li (NIO)* *Jianwei Niu (NIO)* *Jie Gao (NIO)*
P1-02-ASR (#43)	**Personalizing Large Sequence-to-Sequence Speech Foundation Models with Speaker Representations** *Dominik Wagner (Technische Hochschule Nürnberg Georg Simon Ohm)* *Ilja Baumann (Technische Hochschule Nürnberg Georg Simon Ohm)* *Thomas Ranzenberger (Technische Hochschule Nürnberg Georg Simon Ohm)* *Korbinian Riedhammer (Technische Hochschule Nürnberg Georg Simon Ohm)* *Tobias Bocklet (TH Nürnberg)*

P1-03-ASR (#97)	**Label-Looping: Highly Efficient Decoding for Transducers** *Vladimir Bataev (NVIDIA, University of London)* *Hainan Xu (NVIDIA)* *Daniel Galvez (NVIDIA)* *Vitaly Lavrukhin (NVIDIA)* *Boris Ginsburg (NVIDIA)*
P1-04-ASR (#102)	**Advancing Multi-Talker ASR Performance with Large Language Models** *Mohan Shi (University of California, Los Angeles)* *Zengrui Jin (The Chinese University of Hong Kong)* *Yaoxun Xu (Tsinghua University)* *Yong Xu (Tencent)* *Shi-Xiong Zhang (Capital One)* *Kun Wei (School of Computer Science, Northwestern Polytechnical University)* *Yiwen Shao (Johns Hopkins University)* *Chunlei Zhang (Bytedance)* *Dong Yu (Tencent AI Lab)*
P1-05-ASR (#114)	**Token-Weighted RNN-T for Learning from Flawed Data** *Gil Keren (Meta)* *Wei Zhou (Meta)* *Ozlem Kalinli (Meta)*
P1-06-ASR (#137)	**Enhancing Code-Switching Speech Recognition with LID-Based Collaborative Mixture of Experts Model** *Hukai Huang (Xiamen University)* *Jiayan Lin (Xiamen University)* *Kaidi Wang (Xiamen University)* *Yishuang Li (Xiamen University)* *Wenhao Guan (Xiamen University)* *Lin Li (Xiamen University)* *Qingyang Hong (Xiamen University)*
P1-07-ASR (#149)	**Language Bias in Self-Supervised Learning for Automatic Speech Recognition** *Ed Storey (Trinity College Dublin)* *Naomi Harte (Trinity College Dublin)* *Peter Bell (University of Edinburgh)*
P1-08-ASR (#150)	**Robust Audiovisual Speech Recognition Models with Mixture-of-Experts**

	Yihan Wu (Renmin University of China) *Yifan Peng (Carnegie Mellon University)* *Yichen Lu (Carnegie Mellon University)* *Xuankai Chang (Carnegie Mellon University)* *Ruihua Song (Renmin University of China)* *Shinji Watanabe (Carnegie Mellon University)*
P1-09-ASR (#162)	**Hybrid Attention-Based Encoder-Decoder Model for Efficient Language Model Adaptation** *Shaoshi Ling (Microsoft)* *Guoli Ye (Microsoft)* *Rui Zhao (Microsoft)* *Yifan Gong (Microsoft)*
P1-10-ASR (#165)	**SpatialEmb: Extract and Encode Spatial Information for 1-Stage Multi-Channel Multi-Speaker ASR on Arbitrary Microphone Arrays** *Yiwen Shao (Johns Hopkins University)* *Yong Xu (Tencent)* *Sanjeev Khudanpur (Johns Hopkins University)* *Dong Yu (Tencent AI Lab)*
P1-11-ASR (#171)	**Effective Text Adaptation for LLM-Based ASR through Soft Prompt Fine-Tuning** *Yingyi Ma (Meta)* *Zhe Liu (Meta)* *Ozlem Kalinli (Meta)*
P1-12-ASR (#173)	**Temporal Order Preserved Optimal Transport-Based Cross-Modal Knowledge Transfer Learning for ASR** *Xugang Lu (NICT)* *Peng Shen (NICT)* *Yu Tsao (Academia Sinica)* *Hisashi Kawai (NICT)*
P1-13-ASR (#187)	**Contextualized Automatic Speech Recognition with Dynamic Vocabulary** *Yui Sudo (Honda Research Institute Japan)* *Yosuke Fukumoto (Honda Research Institute Japan)* *Muhammad Shakeel (Honda Research Institute Japan)* *Yifan Peng (Carnegie Mellon University)* *Shinji Watanabe (Carnegie Mellon University)*
P1-14-ASR	**Do Prompts Really Prompt? Exploring the Prompt Understanding**

(#209)	**Capability of Whisper** *Chih-Kai Yang (National Taiwan University)* *Kuan-Po Huang (National Taiwan University)* *Hung-yi Lee (National Taiwan University)*
P1-15-ASR (#221)	**An Effective Context-Balanced Adaptation Approach for Long-Tailed Speech Recognition** *Yi-Cheng Wang (National Taiwan Normal University)* *Li-Ting Pai (National Taiwan Normal University)* *Bi-Cheng Yan (National Taiwan Normal University)* *Hsin-Wei Wang (NTNU)* *Chi-Han Lin (E.SUN Financial Holding Co., Ltd.)* *Berlin Chen (National Taiwan Normal University)*
P1-16-ASR (#300)	**Training Large ASR Encoders with Differential Privacy** *Geeticka Chauhan (Google DeepMind)* *Steve Chien (Google)* *Om Thakkar (Google)* *Abhradeep Thakurta (Google)* *Arun Narayanan (Google Inc.)*
P1-17-ASR (#309)	**Transducer Consistency Regularization for Speech-to-Text Applications** *Cindy S Tseng (Samsung Research America)* *Yun Tang (Samsung Research America)* *Vijendra Raj Apsingekar (Samsung Research America)*
P1-18-ASR (#324)	**Leave No Knowledge Behind During Knowledge Distillation: Practical and Effective Knowledge Distillation for Code-Switching ASR Using Realistic Data** *Liang-Hsuan Tseng (National Taiwan University)* *Zih-Ching Chen (National Taiwan University)* *Weishun Chang (National Taiwan University)* *Cheng-Kuang Lee (NVIDIA Corporation)* *Tsung-Ren Huang (National Taiwan University)* *Hung-yi Lee (National Taiwan University)*
P1-19-ASR (#389)	**CTC-Assisted LLM-Based Contextual ASR** *Guanrou Yang (Shanghai Jiao Tong University)* *Ziyang Ma (Shanghai Jiao Tong University)* *Zhifu Gao (Alibaba)* *Shiliang Zhang (Alibaba Group)* *Xie Chen (Shanghai Jiao Tong University)*

P1-20-ASR (#53)	**Automatic Time Alignment Generation for End-to-End ASR Using Acoustic Probability Modeling** *Dongcheng Jiang (University of Cambridge)* *Chao Zhang (Tsinghua University)* *Phil Woodland (Machine Intelligence Laboratory, Cambridge University Department of Engineering)*
P1-21-ASR (#73)	**Continual Learning with Embedding Layer Surgery and Task-Wise Beam** *Chin Yuen Kwok (Nanyang Technological University)* *Jia Qi Yip (Alibaba Group / Nanyang Technological University)* *Eng Siong Chng (Nanyang Technological University)*
P1-22-ASR (#93)	**BESTOW: Efficient and Streamable Speech Language Model with the Best of GPT and T5** *He Huang (NVIDIA)* *Zhehuai Chen (NVIDIA)* *Krishna C Puvvada (NVIDIA)* *Piotr Żelasko (NVIDIA)* *Jagadeesh Balam (NVIDIA)* *Boris Ginsburg (NVIDIA)* *Nithin Rao Koluguri (NVIDIA)* *Oleksii Hrinchuk (NVIDIA)*
P1-23-ASR (#116)	**Combining TF-GridNet and Mixture Encoder for Continuous Speech Separation for Meeting Transcription** *Peter Vieting (RWTH Aachen University)* *Simon Berger (RWTH Aachen University)* *Thilo von Neumann (Paderborn University)* *Christoph Boeddeker (Paderborn University)* *Ralf Schlüter (RWTH Aachen University)* *Reinhold Haeb-Umbach (Paderborn University)*
P1-24-ASR (#122)	**Mamba-Based Decoder-Only Approach with Bidirectional Speech Modeling for Speech Recognition** *Yoshiki Masuyama (Tokyo Metropolitan University)* *Koichi Miyazaki (CyberAgent)* *Masato Murata (CyberAgent)*
P1-25-ASR (#201)	**An Analysis of Linear Complexity Attention Substitutes with BEST-RQ** *Ryan Whetten (LIA - Avignon University)*

	Titouan Parcollet (Samsung AI Cambridge / University of Cambridge) *Adel Moumen (Avignon University)* *Marco Dinarelli (CNRS)* *Yannick Estève (LIA - Avignon University)*
P1-26-ASR (#214)	**Speech-Mamba: Long-Context Speech Recognition with Selective State Spaces Models** *Xiaoxue Gao (ASTAR)* *Nancy Chen (Institute for Infocomm Research)*
P1-27-ASR (#357)	**Lite ASR Transformer: A Lightweight Transformer Architecture for Automatic Speech Recognition** *Metilda Sagaya Mary N J (Indian Institute of Technology Madras)* *S Umesh (IIT Chennai)*

10:30-12:30 Challenge Session 1: Stutter Speech ASR/SED and Dysarthria WWS (Venue: Lecture Hall)

12:30-14:00 Lunch

14:00-15:00 Invited Talk 1 (Venue: Lecture Hall)

Title: Challenges and Progress in Automatic Speech-to-Speech Translation: Bridging the Gap to Real-Time Interpretation
Speaker: *Prof Satoshi Nakamura, The Chinese University of Hong Kong, Shenzhen*
Chair: *Prof Xie Chen*

15:00-15:30 Coffee Break

15:30-17:30 Poster Session 2: Speech Recognition and Enhancement (Venue: Poster Area)

Chair: Prof Jun Du

Poster ID (Paper ID)	Title and Authors
P2-01-ASR (#22)	**Serialized Speech Information Guidance with Overlapped Encoding Separation for Multi-Speaker Automatic Speech Recognition** *Hao Shi (Kyoto University)* *Yuan Gao (Kyoto University)* *Zhaoheng Ni (Meta AI)* *Tatsuya Kawahara (Kyoto University)*
P2-02-ASR	**Efficient Extraction of Noise-Robust Discrete Units from Self-**

(#161)	**Supervised Speech Models** *Jakob Poncelet (KU Leuven)* *Yujun Wang (Xiaomi)* *Hugo Van Hamme (KU Leuven)*
P2-03-ASR (#170)	**Controlling Whisper: Universal Acoustic Adversarial Attacks to Control Multi-Task Automatic Speech Recognition Models** *Vyas Raina (University of Cambridge)* *Mark Gales (University of Cambridge)*
P2-04-ASR (#283)	**Improving Rare-Word Recognition of Whisper in Zero-Shot Settings** *Yash Jogi (Sprinklr)* *Vaibhav Aggarwal (Sprinklr)* *Shabari S Nair (Sprinklr)* *Yash Verma (Sprinklr)* *Aayush Kubba (Sprinklr)*
P2-05-ASR (#343)	**Augmenting Automatic Speech Recognition Models with Disfluency Detection** *Robin Amann (Karlsruher Institut für Technologie)* *Zhaolin Li (Karlsruhe Institute of Technology)* *Barbara Bruno (Karlsruhe Institute of Technology)* *Jan Niehues (Karlsruhe Institute of Technology)*
P2-06-ASR (#246)	**Enhancing Unified Streaming and Non-Streaming ASR through Curriculum Learning with Easy-to-Hard Tasks** *Yuting Yang (NetEase Yidun AI Lab)* *Yuke Li (NetEase Yidun AI Lab)* *Lifeng Zhou (NetEase Yidun AI Lab)* *Binbin Du (NetEase Yidun AI Lab)* *Haoqi Zhu (NetEase Yidun AI Lab)*
P2-07-ASR (#78)	**DQ-Whisper: Joint Distillation and Quantization for Efficient Multilingual Speech Recognition** *Hang Shao (Shanghai Jiao Tong University)* *Bei Liu (Shanghai Jiao Tong University)* *Wei Wang (Shanghai Jiao Tong University)* *Xun Gong (Shanghai Jiao Tong University)* *Yanmin Qian (Shanghai Jiao Tong University)*
P2-08-ASR (#129)	**Fusion of Discrete Representations and Self-Augmented Representations for Multilingual Automatic Speech Recognition**

	Shih-Heng Wang (National Taiwan University) *Jiatong Shi (Carnegie Mellon University)* *Chien-Yu Huang (National Taiwan University)* *Shinji Watanabe (Carnegie Mellon University)* *Hung-Yi Lee (National Taiwan University)*
P2-09-ASR (#157)	**Longer is (Not Necessarily)** *Nithin Rao Koluguri (NVIDIA)* *Travis M Bartley (NVIDIA* *CUNY)* *Hainan Xu (NVIDIA)* *Oleksii Hrinchuk (NVIDIA)* *Jagadeesh Balam (NVIDIA)* *Boris Ginsburg (NVIDIA)* *Georg Kucsko (Suno Inc.)*
P2-10-ASR (#178)	**Semi-Supervised Learning for Code-Switching ASR with Large Language Model Filter** *Yu Xi (NVIDIA)* *Wen Ding (NVIDIA)* *Kai Yu (Shanghai Jiao Tong University)* *Junjie Lai (NVIDIA)*
P2-11-ASR (#301)	**Parameter Averaging is All You Need to Prevent Forgetting** *Peter W Plantinga (JPMorgan Chase & Co.)* *Jaekwon Yoo (JPMorgan Chase & Co.)* *Abenezer G Girma (JP Morgan Chase)* *Chandra Dhir (JPMorgan Chase)*
P2-12-ASR (#304)	**Advancing CTC Models for Better Speech Alignment: A Topological Approach** *Zeyu Zhao (University of Edinburgh)* *Peter Bell (University of Edinburgh)*
P2-13-SES (#37)	**DualSep: A Lightweight Dual-Encoder Convolutional Recurrent Network for Real-Time In-Car Speech Separation** *Ziqian Wang (Northwestern Polytechnical University)* *Jiayao Sun (Northwestern Polytechnical University)* *Zihan Zhang (Northwestern Polytechnical University)* *Xingchen Li (Northwestern Polytechnical University)* *Jie Liu (Huawei Cloud)* *Lei Xie (NWPU)*

P2-14-SES (#39)	**DDTSE: Discriminative Diffusion Model for Target Speech Extraction** *Leying Zhang (Shanghai Jiao Tong University)* *Yao Qian (Microsoft)* *Linfeng Yu (Shanghai Jiao Tong University)* *Heming Wang (The Ohio State University)* *Hemin Yang (Microsoft)* *Shujie Liu (Microsoft Research Asia)* *Long Zhou (Microsoft Research Asia)* *Yanmin Qian (Shanghai Jiao Tong University)*
P2-15-SES (#89)	**An Investigation of Incorporating Mamba for Speech Enhancement** *Rong Chao (National Taiwan University)* *Wen-Huang Cheng (National Taiwan University)* *Moreno La Quatra (Kore University of Enna)* *Sabato M Siniscalchi (University of Palermo)* *Chao-Han Huck Yang (NVIDIA Research)* *Szu-Wei Fu (NVIDIA)* *Yu Tsao (Academia Sinica)*
P2-16-SES (#117)	**Effective Noise-Aware Data Simulation for Domain-Adaptive Speech Enhancement Leveraging Dynamic Stochastic Perturbation** *Chien-Chun Wang (National Taiwan Normal University)* *Li-Wei Chen (United Link Co., Ltd.)* *Hung-Shin Lee (United Link Co., Ltd.)* *Berlin Chen (National Taiwan Normal University)* *Hsin-Min Wang (Academia Sinica)*
P2-17-SES (#120)	**SMRU: Split-and-Merge Recurrent-Based UNet for Acoustic Echo Cancellation and Noise Suppression** *Zhihang Sun (Tencent AI Lab)* *Andong Li (Tencent AI Lab)* *Rilin Chen (Tencent)* *Hao Zhang (Tencent AI Lab)* *Meng Yu (Tencent)* *Yi Zhou (CQUPT)* *Dong Yu (Tencent AI Lab)*
P2-18-SES (#139)	**On the Effectiveness of Enrollment Speech Augmentation for Target Speaker Extraction** *Junjie Li (The Hong Kong Polytechnic University)* *Ke Zhang (Northeastern University)* *Shuai Wang (Shenzhen Research Institute of Big Data, Chinese University*

	of Hong Kong (Shenzhen)) *Haizhou Li (The Chinese University of Hong Kong, Shenzhen)* *M W Mak (HK PolyU)* *Kong Aik Lee (The Hong Kong Polytechnic University)*
P2-19-SES (#142)	**Diffusion-Based Generative Modeling with Discriminative Guidance for Streamable Speech Enhancement** *Chenda Li (Shanghai Jiao Tong University)* *Samuele Cornell (Carnegie Mellon University)* *Shinji Watanabe (Carnegie Mellon University)* *Yanmin Qian (Shanghai Jiao Tong University)*
P2-20-SES (#216)	**NeuroSpex: Neuro-Guided Speaker Extraction with Cross-Modal Fusion** *Dashanka D N De Silva (University of Bremen)* *Siqi Cai (National University of Singapore)* *Saurav Pahuja (University of Bremen)* *Tanja Schultz (University of Bremen)* *Haizhou Li (The Chinese University of Hong Kong, Shenzhen)*
P2-21-SES (#325)	**Enhancing Speaker Extraction through Rectifying Target Confusion** *Jiahe Wang (Shanghai Jiao Tong University)* *Shuai Wang (Shenzhen Research Institute of Big Data, Chinese University of Hong Kong (Shenzhen))* *Junjie Li (The Hong Kong Polytechnic University)* *Ke Zhang (Northeastern University)* *Yanmin Qian (Shanghai Jiao Tong University)* *Haizhou Li (The Chinese University of Hong Kong, Shenzhen)*
P2-22-SES (#368)	**Diff-PLC: A Diffusion-Based Approach for Effective Packet Loss Concealment** *Da-Hee Yang (Hanyang University)* *Joon-Hyuk Chang (Hanyang University)*
P2-23-SES (#369)	**Improving Curriculum Learning for Target Speaker Extraction with Synthetic Speakers** *Yun Liu (National Institute of Informatics)* *Xuechen Liu (National Institute of Informatics)* *Junichi Yamagishi (National Institute of Informatics)*
P2-24-SS02 (#415)	**Language Model Based Generative Error Correction: A Challenge and Baselines for Speech Recognition, Speaker Tagging, and Emotion Recognition**

	Chao-Han Huck Yang (NVIDIA Research) *Tae Jin Park (NVIDIA)* *Yuan Gong (Massachusetts Institute of Technology)* *Yuanchao Li (University of Edinburgh)* *Yen-Ting Lin (National Taiwan University)* *Zhehuai Chen (NVIDIA)* *Yuchen Hu (Nanyang Technological University)* *Chen Chen (Nanyang Technological University)* *Kunal Dhawan (NVIDIA)* *Piotr Żelasko (NVIDIA)* *Chao Zhang (Tsinghua University)* *Yun-Nung Chen (National Taiwan University)* *Yu Tsao (Academia Sinica)* *Jagadeesh Balam (NVIDIA)* *Boris Ginsburg (NVIDIA)* *Sabato M Siniscalchi (University of Palermo)* *Eng Siong Chng (Nanyang Technological University)* *Peter Bell (University of Edinburgh)* *Catherine Lai (University of Edinburgh)* *Shinji Watanabe (Carnegie Mellon University)* *Andreas Stolcke (Uniphore)*
P2-25-SS06 (#400)	**FGCL: Fine-Grained Contrastive Learning for Mandarin Stuttering Event Detection** *Han Jiang (Xi'an Jiaotong University)* *Wenyu Wang (Xi'an Jiaotong University)* *Yiquan Zhou (XJTU)* *Hongwu Ding (Happy Elements)* *Xu Jiacheng (Happy Elements)* *Jihua Zhu (Xi'an Jiaotong University)*
P2-26-SS06 (#402)	**Data Augmentation Techniques for Improved Performance in the SLT 2024 Mandarin Stuttering Event Detection and ASR Challenge** *Weiwei Wang (Chery HuiYin Motor Finance Service Co., Ltd.)* *Zhijin Feng (Chery HuiYin Motor Finance Service Co., Ltd.)* *Qingyuan Song (Chery HuiYin Motor Finance Service Co., Ltd.)* *Wenyang Wei (Chery HuiYin Motor Finance Service Co., Ltd.)* *Yansong Wang (Chery HuiYin Motor Finance Service Co., Ltd.)*
P2-27-SS06 (#403)	**Findings of the 2024 Mandarin Stuttering Event Detection and Automatic Speech Recognition Challenge** *Hongfei Xue (NWPU)* *Rong Gong (StammerTalk)* *Mingchen Shao (NWPU)* *Xin Xu (AISHELL)* *Lezhi Wang (StammerTalk)*

	Lei Xie (NWPU) *Hui Bu (AISHELL)* *Jiaming Zhou (Nankai University)* *Yong Qin (Nankai University)* *Jun Du (University of Science and Technology of China)* *Ming Li (Wuhan University)* *Binbin Zhang (WeNet Open Source Community)* *Bin Jia (StammerTalk)*
P2-28-SS06 (#410)	**Enhanced ASR for Stuttering Speech: Combining Adversarial and Signal-Based Data Augmentation** *Shangkun Huang (Beijing Fosafer Information Technology Co., Ltd.)* *Dejun Zhang (Beijing Fosafer Information Technology Co., Ltd.)* *Jing Deng (Beijing Fosafer Information Technology Co., Ltd.)* *Rong Zheng (Beijing Fosafer Information Technology Co., Ltd.)*

15:30-17:30 Challenge Session 2: Singing Voice Deepfake Detection (SVDD) (Venue: Lecture Hall)

17:30-18:30 Recent Breakthrough (Venue: Poster Area)

Chair: Prof Eng Siong Chng

Poster ID (Paper ID)	Title and Authors
RB-1 (#1)	**Reverb: Open-Source ASR and Diarization from Rev** *Nishchal Bhandari, Danny Chen, Miguel Ángel del Río Fernández, Natalie Delworth, Jennifer Drexler Fox, Migüel Jetté, Quinten McNamara, Corey Miller, Ondřej Novotný, Ján Profant, Nan Qin, Martin Ratajczak, Jean-Philippe Robichaud*
RB-2 (#2)	**GAMA: A Large Audio-Language Model with Advanced Audio Understanding and Complex Reasoning Abilities** *Sreyan Ghosh, Sonal Kumar, Ashish Seth, Chandra Kiran Reddy Evuru, Utkarsh Tyagi, S Sakshi, Oriol Nieto, Ramani Duraiswami, Dinesh Manocha*
RB-3 (#3)	**MMAU: A Massive Multi-Task Audio Understanding and Reasoning Benchmark** *S Sakshi, Utkarsh Tyagi, Sonal Kumar, Ashish Seth, Ramaneswaran Selvakumar, Oriol Nieto, Ramani Duraiswami, Sreyan Ghosh, Dinesh Manocha*
RB-4 (#4)	**Vevo: Controllable Zero-Shot Voice Imitation with Self-Supervised Disentanglement**

	Xueyao Zhang, Xiaohui Zhang, Kainan Peng, Zhenyu Tang, Vimal Manohar, Yingru Liu, Jeff Hwang, Dangna Li, Yuhao Wang, Julian Chan, Yuan Huang, Zhizheng Wu, Mingbo Ma
RB-5 (#5)	**SD-Eval: A Benchmark Dataset for Spoken Dialogue Understanding Beyond Words** *Junyi Ao, Yuancheng Wang, Xiaohai Tian, Dekun Chen, Jun Zhang, Lu Lu, Yuxuan Wang, Haizhou Li, Zhizheng Wu*
RB-6 (#6)	**SuperM2M: Supervised and Mixture-to-Mixture Co-Learning for Speech Enhancement and Robust ASR** *Zhong-Qiu Wang*
RB-7 (#7)	**Speech Recognition Corpus of the Khinalug Language for Documenting Endangered Languages** *Zhaolin Li, Monika Rind-Pawlowski, Jan Niehues*
RB-8 (#8)	**MaskGCT: Zero-Shot Text-to-Speech with Masked Generative Codec Transformer** *Yuancheng Wang, Haoyue Zhan, Liwei Liu, Ruihong Zeng, Haotian Guo, Jiachen Zheng, Qiang Zhang, Xueyao Zhang, Shunsi Zhang, Zhizheng Wu*
RB-9 (#9)	**Cacophony: An Improved Contrastive Audio-Text Model** *Ge Zhu, Jordan Darefsky, Zhiyao Duan*
RB-10 (#10)	**LlamaPartialSpoof: An LLM-Driven Fake Speech Dataset Simulating Disinformation Generation** *Hieu-Thi Luong, Haoyang Li, Lin Zhang, Kong Aik Lee, Eng Siong Chng*
RB-11 (#11)	**Self-Supervised ASR Models and Features for Dysarthric and Elderly Speech Recognition** *Shujie Hu, Xurong Xie, Mengzhe Geng, Zengrui Jin, Jiajun Deng, Guinan Li, Yi Wang, Mingyu Cui, Tianzi Wang, Helen Meng, Xunying Liu*
RB-12 (#12)	**PAT: Parameter-Free Audio-Text Aligner to Boost Zero-Shot Audio Classification** *Ashish Seth, Ramaneswaran Selvakumar, Sonal Kumar, Sreyan Ghosh, Dinesh Manocha*
RB-13 (#13)	**A Transformer Framework for Simultaneous Segmentation,**

	Classification, and Caller Identification of Marmoset Vocalization *Bin Wu, Shinnosuke Takamichi, Sakriani Sakti, and Satoshi Nakamura*
RB-14 (#14)	**UTDUSS: UTokyo-SaruLab System for Interspeech2024 Speech Processing Using Discrete Speech Unit Challenge** *Wataru Nakata, Kazuki Yamauchi, Dong Yang, Hiroaki Hyodo, Yuki Saito*
RB-15 (#15)	**Optimizing Contextual Speech Recognition Using Vector Quantization for Efficient Retrieval** *Nikolaos Flemotomos, Roger Hsiao, Pawel Swietojanski, Takaaki Hori, Dogan Can, Xiaodan Zhuang*
RB-16 (#16)	**Streaming Speech-to-speech Speech-LLM for Simultaneous Translation and Multi-turn Conversation** *Elena Rastorgueva, Zhehuai Chen, He Huang, Edresson Casanova, Jason Li, Krishna Puvvada, Ryan Langman, Ante Jukić, Kunal Dhawan, Nithin Rao Koluguri, Subhankar Ghosh, Piotr Żelasko, Oleksii Hrinchuk, Andrei Andrusenko, Vitaly Lavrukhin, Jagadeesh Balam, Boris Ginsburg*

19:00-20:30 Welcome Reception

Day 2, Dec 3, Tuesday

09:00-10:00 Keynote Speech 2 (Venue: Lecture Hall)

Title: End-to-End Audio Processing: From On-Device Models to LLMs
Speaker: *Dr Tara N. Sainath, Google DeepMind*
Chair: *Prof Hung-yi Lee*

10:00-10:30 Coffee Break

10:30-12:30 Poster Session 3: Speech Processing (Venue: Poster Area)

Chair: Prof Junichi Yamagishi

Poster ID (Paper ID)	Title and Authors
P3-01-ANA (#66)	**Property Neurons in Self-Supervised Speech Transformers** *Tzu-Quan Lin (National Taiwan University)* *Guan-Ting Lin (National Taiwan University)* *Hung-Yi Lee (National Taiwan University)* *Hao Tang (The University of Edinburgh)*

P3-02-ANA (#145)	**Privacy vs Emotion Preservation Trade-Offs in Emotion-Preserving Speaker Anonymization** *Zexin Cai (Johns Hopkins University)* *Henry Li Xinyuan (Johns Hopkins University)* *Ashi Garg (Bharti Vidyapeeth College of Engineering)* *Nicholas O Andrews (Johns Hopkins University)* *Paola Garcia (Johns Hopkins University)* *Matthew S Wiesner (Johns Hopkins University)* *Kevin Duh (Johns Hopkins University)* *Sanjeev Khudanpur (Johns Hopkins University)*
P3-03-ANA (#217)	**Estimating the Completeness of Discrete Speech Units** *Sung-Lin Yeh (University of Edinburgh)* *Hao Tang (The University of Edinburgh)*
P3-04-ANA (#314)	**Investigation of Speaker Representation for Target-Speaker Speech Processing** *Takanori Ashihara (NTT Corp.)* *Takafumi Moriya (NTT Corporation)* *Shota Horiguchi (NTT Corporation)* *Junyi Peng (Brno University of Technology)* *Tsubasa Ochiai (NTT)* *Marc Delcroix (NTT)* *Kohei Matsuura (NTT)* *Hiroshi Sato (NTT Corporation)*
P3-05-MMP (#226)	**Crossmodal ASR Error Correction with Discrete Speech Units** *Yuanchao Li (University of Edinburgh)* *Pinzhen Chen (University of Edinburgh)* *Peter Bell (University of Edinburgh)* *Catherine Lai (University of Edinburgh)*
P3-06-MMP (#241)	**Listen and Speak Fairly: A Study on Semantic Gender Bias in Speech Integrated Large Language Models** *Yi-Cheng Lin (National Taiwan University)* *Tzu-Quan Lin (National Taiwan University)* *Chih-Kai Yang (National Taiwan University)* *Ke-Han Lu (National Taiwan University)* *Wei-Chih Chen (National Taiwan University)* *Chun-Yi Kuan (National Taiwan University)* *Hung-Yi Lee (National Taiwan University)*
P3-07-MMP (#265)	**Learning Video Temporal Dynamics with Cross-Modal Attention for Robust Audio-Visual Speech Recognition**

	Sungnyun Kim (KAIST) *Kangwook Jang (KAIST)* *Sangmin Bae (KAIST)* *Hoirin Kim (KAIST)* *Se-Young Yun (KAIST)*
P3-08-MMP (#269)	**Data Efficient Reflow for Few Step Audio Generation** *Lemeng Wu (Meta)* *Zhaoheng Ni (Meta AI)* *Bowen Shi (Toyota Technological Institute at Chicago)* *Wei-Ning Hsu (Meta)* *Gael Le Lan (Meta)* *Varun Nagaraja (Meta)* *Anurag Kumar (Meta)* *Xinhao Mei (Meta)* *Yunyang Xiong (Meta)* *Bilge Soran (Meta)* *Raghuraman Krishnamoorthi (Facebook)* *Yangyang Shi (Facebook)* *Vikas Chandra (Meta)*
P3-09-MLS (#99)	**Optimizing Byte-Level Representation for End-to-End ASR** *Roger Hsiao (Apple)* *Liuhui Deng (Apple)* *Erik McDermott (Apple)* *Ruchir Travadi (Apple)* *Xiaodan Zhuang (Apple)*
P3-10-MLS (#179)	**Romanization Encoding for Multilingual ASR** *Wen Ding (NVIDIA)* *Fei Jia (NVIDIA Corporation)* *Hainan Xu (NVIDIA Corporation)* *Yu Xi (NVIDIA)* *Junjie Lai (NVIDIA)* *Boris Ginsburg (NVIDIA)*
P3-11-MLS (#254)	**Enhancing Code-Switching ASR Leveraging Non-Peaky CTC Loss and Deep Language Posterior Injection** *Tzu-Ting Yang (National Taiwan Normal University)* *Hsin-Wei Wang (NTNU)* *Yi-Cheng Wang (National Taiwan Normal University)* *Berlin Chen (National Taiwan Normal University)*
P3-12-MLS	**Language-Independent Prosody-Enhanced Speech**

(#263)	**Representations for Multilingual Speech Synthesis** *Chang Liu (University of Science and Technology of China)* *Zhen-Hua Ling (University of Science and Technology of China)* *Ya-Jun Hu (iFLYTEK Co., Ltd.)*
P3-13-MLS (#359)	**Classification of Spontaneous and Scripted Speech for Multilingual Audio** *Shahar Elisha (Spotify)* *Andrew J McDowell (Spotify)* *Mariano Beguerisse-Díaz (Spotify)* *Emmanouil Benetos (Queen Mary University of London)*
P3-14-EMR (#40)	**GMP-TL: Gender-Augmented Multi-Scale Pseudo-Label Enhanced Transfer Learning for Speech Emotion Recognition** *Yu Pan (Kyushu University)* *Yuguang Yang (Ximalaya Inc., Shanghai, China)* *Yuheng Huang (The University of Tokyo)* *Tiancheng Jin (Kyushu University)* *Jingjing Yin (Ximalaya)* *Yanni Hu (Ximalaya Inc., Shanghai, China)* *Heng Lu (Ximalaya Inc.)* *Lei Ma (The University of Tokyo / University of Alberta)* *Jianjun Zhao (Kyushu University)*
P3-15-EMR (#81)	**Embracing Ambiguity and Subjectivity Using the All-Inclusive Aggregation Rule for Evaluating Multi-Label Speech Emotion Recognition Systems** *Huang-Cheng Chou (Department of Electrical Engineering at National Tsing Hua University (NTHU))* *Haibin Wu (National Taiwan University)* *Lucas Goncalves (The University of Texas at Dallas)* *Seong-Gyun Leem (University of Texas at Dallas)* *Ali N Salman (University of Texas at Dallas)* *Carlos Busso (University of Texas at Dallas)* *Hung-Yi Lee (National Taiwan University)* *Chi-Chun Lee (National Tsing Hua University)*
P3-16-EMR (#83)	**Open-Emotion: A Reproducible EMO-SUPERB for Speech Emotion Recognition Systems** *Haibin Wu (National Taiwan University)* *Huang-Cheng Chou (Department of Electrical Engineering at National Tsing Hua University (NTHU))* *Kai-Wei Chang (National Taiwan University)* *Lucas Goncalves (The University of Texas at Dallas)* *Jiawei Du (National Taiwan University)*

	Jyh-Shing Roger Jang (National Taiwan University) *Chi-Chun Lee (National Tsing Hua University)* *Hung-Yi Lee (National Taiwan University)*
P3-17-EMR (#225)	**Speech Emotion Recognition with ASR Transcripts: A Comprehensive Study on Word Error Rate and Fusion Techniques** *Yuanchao Li (University of Edinburgh)* *Peter Bell (University of Edinburgh)* *Catherine Lai (University of Edinburgh)*
P3-18-EMR (#282)	**Beyond the Binary: Limitations and Possibilities of Gender-Related Speech Technology Research** *Ariadna Sanchez (The University of Edinburgh)* *Alice Ross (University of Edinburgh)* *Nina Markl (University of Essex)*
P3-19-EMR (#352)	**Enhancing Domain Generalization in Speech Emotion Recognition by Combining Domain-Variant Representations and Domain-Invariant Classifiers** *Shi-Wook Lee (National Institute of Advanced Industrial Science and Technology)*
P3-20-SS07 (#47)	**MDCTCodec: A Lightweight MDCT-Based Neural Audio Codec for High Sampling Rate and Low Bitrate Scenarios** *Xiao-Hang Jiang (University of Science and Technology of China)* *Yang Ai (University of Science and Technology of China)* *Rui-Chen Zheng (University of Science and Technology of China)* *Hui-Peng Du (University of Science and Technology of China)* *Ye-Xin Lu (University of Science and Technology of China)* *Zhen-Hua Ling (University of Science and Technology of China)*
P3-21-SS07 (#51)	**Addressing Index Collapse of Large-Codebook Speech Tokenizer with Dual-Decoding Product-Quantized Variational Auto-Encoder** *Haohan Guo (The Chinese University of Hong Kong)* *Fenglong Xie (Xiaohongshu)* *Dongchao Yang (The Chinese University of Hong Kong)* *Hui Lu (The Chinese University of Hong Kong)* *Xixin Wu (The Chinese University of Hong Kong)* *Helen Meng (The Chinese University of Hong Kong)*
P3-22-SS07 (#267)	**Investigating Neural Audio Codecs for Speech Language Model-Based Speech Generation** *Jiaqi Li (The Chinese University of Hong Kong, Shenzhen)*

	Dongmei Wang (Microsoft) *Xiaofei Wang (Microsoft)* *Yao Qian (Microsoft)* *Long Zhou (Microsoft Research Asia)* *Shujie Liu (Microsoft Research Asia)* *Midia Yousefi (Microsoft)* *Canrun Li (Microsoft)* *Chung-Hsien Tsai (Microsoft)* *Zhen Xiao (Microsoft)* *Yanqing Liu (Microsoft)* *Junkun Chen (Microsoft)* *Sheng Zhao (Microsoft)* *Jinyu Li (Microsoft)* *Zhizheng Wu (Chinese University of Hong Kong, Shenzhen)* *Michael Zeng (Microsoft)*
P3-23-SS07 (#280)	**ESPnet-Codec: Comprehensive Training and Evaluation of Neural Codecs for Audio, Music, and Speech** *Jiatong Shi (Carnegie Mellon University)* *Jinchuan Tian (Carnegie Mellon University)* *Yihan Wu (Renmin University of China)* *Jee-Weon Jung (Carnegie Mellon University)* *Jia Qi Yip (Alibaba Group / Nanyang Technological University)* *Yoshiki Masuyama (Tokyo Metropolitan University)* *William Chen (Carnegie Mellon University)* *Yuning Wu (Renmin University of China)* *Yuxun Tang (Renmin University of China)* *Massa Baali (CMU)* *Dareen Alharthi (Carnegie Mellon University)* *Dong Zhang (Fudan University)* *Ruifan Deng (Fudan University)* *Tejes Srivastava (University of Chicago)* *Haibin Wu (National Taiwan University)* *Alexander H Liu (MIT)* *Bhiksha Raj (Carnegie Mellon University)* *Qin Jin (Renmin University of China)* *Ruihua Song (Renmin University of China)* *Shinji Watanabe (Carnegie Mellon University)*
P3-24-SS07 (#336)	**Codec-SUPERB @ SLT 2024: A Lightweight Benchmark for Neural Codec Models** *Haibin Wu (National Taiwan University)* *Xuanjun Chen (National Taiwan University)* *Yi-Cheng Lin (National Taiwan University)* *Jiawei Du (National Taiwan University)* *Kai-Wei Chang (National Taiwan University)* *Ke-Han Lu (National Taiwan University)* *Alexander H Liu (MIT)*

	Ho Lam Chung (National Taiwan University) *Yuan-Kuei Wu (National Taiwan University)* *Dongchao Yang (The Chinese University of Hong Kong)* *Songxiang Liu (Tencent)* *Yi-Chiao Wu (Meta)* *Xu Tan (Microsoft Research Asia)* *James Glass (Massachusetts Institute of Technology)* *Shinji Watanabe (Carnegie Mellon University)* *Hung-Yi Lee (National Taiwan University)*
P3-25-SS08 (#61)	**Optimizing Dysarthria Wake-Up Word Spotting: An End-to-End Approach for SLT 2024 LRDWWS Challenge** *Shuiyun Liu (Northwestern Polytechnical University)* *Yuxiang Kong (Xiaomi Inc.)* *Pengcheng Guo (Northwestern Polytechnical University)* *Weiji Zhuang (Xiaomi Inc.)* *Peng Gao (Xiaomi Inc.)* *Yujun Wang (Xiaomi)* *Lei Xie (NWPU)*
P3-26-SS08 (#234)	**PB-LRDWWS System for the SLT 2024 Low-Resource Dysarthria Wake-Up Word Spotting Challenge** *Shiyao Wang (Nankai University)* *Jiaming Zhou (Nankai University)* *Shiwan Zhao (Nankai University)* *Yong Qin (Nankai University)*
P3-27-SS08 (#404)	**Summary of Low-Resource Dysarthria Wake-Up Word Spotting Challenge** *Ming Gao (University of Science and Technology of China)* *Hang Chen (USTC)* *Jun Du (University of Science and Technology of China)* *Xin Xu (Beijing AISHELL Technology Co., Ltd.)* *Hongxiao Guo (Beijing AISHELL Technology Co., Ltd.)* *Hui Bu (AISHELL)* *Ming Li (Wuhan University)* *Chin-Hui Lee (Georgia Institute of Technology)*
P3-28-SLP (#105)	**ProGRes: Prompted Generative Rescoring on ASR N-Best** *Ada D Tur (McGill University)* *Mirco Ravanelli (Concordia University, Université de Montréal, MILA)* *Adel Moumen (Avignon University)*
P3-29-SS02 (#212)	**FLANEC: Exploring Flan-T5 for Post-ASR Error Correction**

	Moreno La Quatra (Kore University of Enna) *Valerio Mario Salerno (Università degli Studi di Enna "Kore")* *Yu Tsao (Academia Sinica)* *Sabato Marco Siniscalchi (Kore University of Enna)*

10:30-12:30 Challenge Session 3: Source Speaker Tracing Challenge(SSTC) (Venue: Lecture Hall)

12:30-14:00 Lunch

14:00-15:00 Invited Talk 2 (Venue: Lecture Hall)

Title: Holistic Artificial Intelligence (HAI): From Big Models to Big Applications
Speaker: *Dr Junlan Feng, China Mobile Research Institute*
Chair: *Prof Lei Wang*

15:00-15:30 Coffee Break

15:30-17:30 Poster Session 4: Speech Synthesis (Venue: Poster Area)

Chair: Prof Xixin Wu

Poster ID (Paper ID)	Title and Authors
P4-01-TTS (#21)	**AS-Speech: Adaptive Style for Speech Synthesis** *Zhipeng Li (South China University of Technology)* *Xiaofen Xing (South China University of Technology)* *Jun Wang (Meituan)* *Shuaiqi Chen (South China University of Technology)* *Guoqiao Yu (Meituan)* *Guanglu Wan (Meituan)* *Xiangmin Xu (South China University of Technology)*
P4-02-TTS (#31)	**Room Impulse Responses Help Attackers Evade Deep Fake Detection** *Hieu-Thi Luong (Nanyang Technological University)* *Duc-Tuan Truong (Nanyang Technological University)* *Kong Aik Lee (The Hong Kong Polytechnic University)* *Eng Siong Chng (Nanyang Technological University)*
P4-03-TTS (#35)	**Attention-Constrained Inference for Robust Decoder-Only Text-to-Speech** *Hankun Wang (Shanghai Jiao Tong University)* *Chenpeng Du (Shanghai Jiao Tong University)* *Yiwei Guo (Shanghai Jiao Tong University)* *Shuai Wang (Shenzhen Research Institute of Big Data, Chinese University*

	of Hong Kong (Shenzhen)) *Xie Chen (Shanghai Jiao Tong University)* *Kai Yu (Shanghai Jiao Tong University)*
P4-04-TTS (#46)	**Stage-Wise and Prior-Aware Neural Speech Phase Prediction** *Fei Liu (University of Science and Technology of China)* *Yang Ai (University of Science and Technology of China)* *Hui-Peng Du (University of Science and Technology of China)* *Ye-Xin Lu (University of Science and Technology of China)* *Rui-Chen Zheng (University of Science and Technology of China)* *Zhen-Hua Ling (University of Science and Technology of China)*
P4-05-TTS (#52)	**SoCodec: A Semantic-Ordered Multi-Stream Speech Codec for Efficient Language Model-Based Text-to-Speech Synthesis** *Haohan Guo (The Chinese University of Hong Kong)* *Fenglong Xie (Xiaohongshu)* *Dongchao Yang (The Chinese University of Hong Kong)* *Xixin Wu (The Chinese University of Hong Kong)* *Helen Meng (The Chinese University of Hong Kong)* *Kun Xie (Xiaohongshu)* *Dake Guo (Northwestern Polytechnical University)*
P4-06-TTS (#56)	**Detecting the Undetectable: Assessing the Efficacy of Current Spoof Detection Methods Against Seamless Speech Edits** *Sung-Feng Huang (National Taiwan University)* *Heng-Cheng Kuo (National Taiwan University)* *Zhehuai Chen (NVIDIA)* *Xuesong Yang (NVIDIA Applied AI Research)* *Chao-Han Huck Yang (NVIDIA Research)* *Yu Tsao (Academia Sinica)* *Yu-Chiang Frank Wang (National Taiwan University)* *Hung-Yi Lee (National Taiwan University)* *Szu-Wei Fu (NVIDIA)*
P4-07-TTS (#62)	**DNN-Based Ensemble Singing Voice Synthesis with Interactions Between Singers** *Hiroaki Hyodo (The University of Tokyo)* *Shinnosuke Takamichi (Keio University)* *Tomohiko Nakamura (National Institute of Advanced Industrial Science and Technology (AIST))* *Junya Koguchi (Meiji University)* *Hiroshi Saruwatari (The University of Tokyo)*
P4-08-TTS (#87)	**Investigating Disentanglement in a Phoneme-Level Speech Codec for Prosody Modeling**

	Sotirios Karapiperis (Samsung) *Nikolaos Ellinas (Innoetics, Samsung Electronics)* *Alexandra Vioni (Innoetics, Samsung Electronics)* *Junkwang Oh (Mobile eXperience Business, Samsung Electronics)* *Gunu Jho (Mobile eXperience Business, Samsung Electronics)* *Inchul Hwang (Samsung Research)* *Spyros Raptis (Samsung Electronics Hellas / INNOETICS)*
P4-09-TTS (#128)	**InstructSing: High-Fidelity Singing Voice Generation via Instructing Yourself** *Chang Zeng (National Institute of Informatics)* *Chunhui Wang (Geely Automobile Research Institute)* *Xiaoxiao Miao (Singapore Institute of Technology)* *Jian Zhao (Geely)* *Zhonglin Jiang (Geely)* *Yong Chen (Geely Automobile Research Institute)*
P4-10-TTS (#166)	**E2 TTS: Embarrassingly Easy Fully Non-Autoregressive Zero-Shot TTS** *Sefik Emre Eskimez (Microsoft)* *Xiaofei Wang (Microsoft)* *Manthan Thakker (Microsoft)* *Canrun Li (Microsoft)* *Chung-Hsien Tsai (Microsoft)* *Zhen Xiao (Microsoft)* *Hemin Yang (Microsoft)* *Zirun Zhu (Microsoft)* *Min Tang (Microsoft)* *Xu Tan (Microsoft Research Asia)* *Yanqing Liu (Microsoft)* *Sheng Zhao (Microsoft)* *Naoyuki Kanda (Microsoft)*
P4-11-TTS (#167)	**Laugh Now Cry Later: Controlling Time-Varying Emotional States of Flow-Matching-Based Zero-Shot Text-to-Speech** *Haibin Wu (National Taiwan University)* *Xiaofei Wang (Microsoft)* *Sefik Emre Eskimez (Microsoft)* *Manthan Thakker (Microsoft)* *Daniel Tompkins (Microsoft)* *Chung-Hsien Tsai (Microsoft)* *Canrun Li (Microsoft)* *Zhen Xiao (Microsoft)* *Sheng Zhao (Microsoft)* *Jinyu Li (Microsoft)* *Naoyuki Kanda (Microsoft)*

P4-12-TTS (#184)	**Disentangling the Prosody and Semantic Information with Pre-Trained Model for In-Context Learning-Based Zero-Shot Voice Conversion** *Zhengyang Chen (Shanghai Jiao Tong University)* *Shuai Wang (Shenzhen Research Institute of Big Data, Chinese University of Hong Kong (Shenzhen))* *Mingyang Zhang (Chinese University of Hong Kong, Shenzhen)* *Xuechen Liu (National Institute of Informatics)* *Junichi Yamagishi (National Institute of Informatics)* *Yanmin Qian (Shanghai Jiao Tong University)*
P4-13-TTS (#195)	**NDVQ: Robust Neural Audio Codec with Distribution-Based Vector Quantization** *Zhikang Niu (Shanghai Jiao Tong University)* *Sanyuan Chen (Harbin Institute of Technology)* *Long Zhou (Microsoft Research Asia)* *Ziyang Ma (Shanghai Jiao Tong University)* *Xie Chen (Shanghai Jiao Tong University)* *Shujie Liu (Microsoft Research Asia)*
P4-14-TTS (#228)	**Fast, High-Quality, and Parameter-Efficient Articulatory Synthesis Using Differentiable DSP** *Yisi Liu (University of California, Berkeley)* *Bohan Yu (UC Berkeley)* *Drake Lin (UC Berkeley)* *Peter Wu (UC Berkeley)* *Cheol Jun Cho (UC Berkeley)* *Gopala Krishna Anumanchipalli (UC Berkeley)*
P4-15-TTS (#299)	**VISinger2+: End-to-End Singing Voice Synthesis Augmented by Self-Supervised Learning Representation** *Yifeng Yu (Georgia Institute of Technology)* *Jiatong Shi (Carnegie Mellon University)* *Yuning Wu (Renmin University of China)* *Yuxun Tang (Renmin University of China)* *Shinji Watanabe (Carnegie Mellon University)*
P4-16-TTS (#316)	**End-to-End Streaming Model for Low-Latency Speech Anonymization** *Waris Quamer (Texas A&M University)* *Ricardo Gutierrez-Osuna (Texas A&M University)*
P4-17-TTS	**Emotion-Coherent Speech Data Augmentation and Self-Supervised**

(#326)	**Contrastive Style Training for Enhancing Kids' Story Speech Synthesis** *Raymond Chung (LSCM)*
P4-18-TTS (#332)	**Discrete Unit-Based Masking for Improving Disentanglement in Voice Conversion** *Philip Lee (University of Texas at Dallas)* *İsmail Rasim Ülgen (University of Texas at Dallas)* *Berrak Sisman (University of Texas at Dallas)*
P4-19-TTS (#345)	**Cross-Dialect Text-To-Speech in Pitch-Accent Language Incorporating Multi-Dialect Phoneme-Level BERT** *Kazuki Yamauchi (The University of Tokyo)* *Yuki Saito (The University of Tokyo, Japan)* *Hiroshi Saruwatari (The University of Tokyo)*
P4-20-TTS (#353)	**Leveraging Diverse Semantic-Based Audio Pretrained Models for Singing Voice Conversion** *Xueyao Zhang (The Chinese University of Hong Kong, Shenzhen)* *Zihao Fang (The Chinese University of Hong Kong, Shenzhen)* *Yicheng Gu (The Chinese University of Hong Kong, Shenzhen)* *Haopeng Chen (The Chinese University of Hong Kong, Shenzhen)* *Lexiao Zou (Harbin Institute of Technology (Shenzhen))* *Junan Zhang (Fudan University)* *Liumeng Xue (The Chinese University of Hong Kong, Shenzhen)* *Zhizheng Wu (The Chinese University of Hong Kong, Shenzhen)*
P4-21-TTS (#394)	**TTSDS: Text-to-Speech Distribution Score** *Christoph D Minixhofer (The University of Edinburgh)* *Ondřej Klejch (University of Edinburgh)* *Peter Bell (University of Edinburgh)*
P4-22-SS03 (#261)	**Speech Foundation Model Ensembles for the Controlled Singing Voice Deepfake Detection (CtrSVDD)** *Anmol Guragain (Vellore Institute of Technology)* *Tianchi Liu (National University of Singapore)* *Zihan Pan (Institute for Infocomm Research (I2R), ASTAR, Singapore)* *Hardik B Sailor (I2R, ASTAR, Singapore)* *Qiongqiong Wang (ASTAR)*
P4-23-SS03 (#323)	**SVDD 2024: The Inaugural Singing Voice Deepfake Detection Challenge**

	You Zhang (University of Rochester) *Yongyi Zang (University of Rochester)* *Jiatong Shi (Carnegie Mellon University)* *Ryuichi Yamamoto (Nagoya University)* *Tomoki Toda (Nagoya University)* *Zhiyao Duan (University of Rochester)*
P4-24-SS03 (#348)	**XWSB: A Blend System Utilizing XLS-R and WavLM with SLS Classifier for the SVDD 2024 Challenge** *Zhang Qishan (Hubei Minzu University)* *Shuangbing Wen (Hubei Minzu University)* *Fangke Yan (Hubei Minzu University)* *Tao Hu (Hubei Minzu University)* *Jun Li (Hubei Minzu University)*
P4-25-SS03 (#416)	**Integrating Self-Supervised Pre-Training with Adversarial Learning for Synthesized Song Detection** *Yankai Wang (Beijing Fosafer Information Technology Co., Ltd.)* *Yuxuan Du (Beijing Fosafer Information Technology Co., Ltd.)* *Dejun Zhang (Beijing Fosafer Information Technology Co., Ltd.)* *Rong Zheng (Beijing Fosafer Information Technology Co., Ltd.)* *Jing Deng (Beijing Fosafer Information Technology Co., Ltd.)*
P4-26-SS05 (#396)	**The VoiceMOS Challenge 2024: Beyond Speech Quality Prediction** *Wen-Chin Huang (Nagoya University)* *Szu-Wei Fu (NVIDIA)* *Erica Cooper (National Institute of Information and Communications Technology)* *Ryandhimas E Zezario (Academia Sinica)* *Tomoki Toda (Nagoya University)* *Hsin-Min Wang (Academia Sinica)* *Junichi Yamagishi (National Institute of Informatics)* *Yu Tsao (Academia Sinica)*
P4-27-SS05 (#406)	**Pitch-and-Spectrum-Aware Singing Quality Assessment with Bias Correction and Model Fusion** *Yu-Fei Shi (University of Science and Technology of China)* *Yang Ai (University of Science and Technology of China)* *Ye-Xin Lu (University of Science and Technology of China)* *Hui-Peng Du (University of Science and Technology of China)* *Zhen-Hua Ling (University of Science and Technology of China)*
P4-28-SS05 (#407)	**The T05 System for The VoiceMOS Challenge 2024: Transfer Learning from Deep Image Classifier to Naturalness MOS Prediction of High-Quality Synthetic Speech**

	Kaito Baba (The University of Tokyo) *Wataru Nakata (The University of Tokyo)* *Yuki Saito (The University of Tokyo, Japan)* *Hiroshi Saruwatari (The University of Tokyo)*

15:15-17:30 Challenge Session 4: Codec SUPERB Challenge (Venue: Lecture Hall)

Day 3, Dec 4, Wednesday

09:00-10:00 Keynote Speech 3 (Venue: Lecture Hall)

Title: Large Language-Audio Models and Applications
Speaker: *Prof Wenwu Wang, University of Surrey*
Chair: *Dr Jinyu Li*

10:00-10:30 Coffee Break

10:30-12:30 Poster Session 5: Machine Learning & Resources (Venue: Poster Area)

Chair: Prof Ming Li

Poster ID (Paper ID)	Title and Authors
P5-01-TLP (#12)	**Automated Speaking Assessment of Conversation Tests with a Novel Graph-Based Modeling Method on Spoken Response Coherence** *Jiun-Ting Li (National Taiwan Normal University)* *Bi-Cheng Yan (National Taiwan Normal University)* *Tien-Hong Lo (National Taiwan Normal University)* *Yi-Cheng Wang (National Taiwan Normal University)* *Yung-Chang Hsu (EZ-AI)* *Berlin Chen (National Taiwan Normal University)*
P5-02-TLP (#115)	**Conditional Label Smoothing for LLM-Based Data Augmentation in Medical Text Classification** *Luca Manuel Becker (Institute of Communication Acoustics, Ruhr-Universität Bochum)* *Philip Pracht (Bochum Institute of Technology)* *Peter Sertdal (Fraunhofer Institute for High Frequency Physics and Radar Techniques FHR)* *Jil Uboreck (Bochum Institute of Technology)* *Alexander Bendel (Institute of Work and Qualification, University of*

	Duisburg-Essen) *Rainer Martin (Institute of Communication Acoustics, Ruhr-Universität Bochum)*
P5-03-TLP (#180)	**Plan, Generate and Optimize: Extending Large Language Models for Dialogue Systems via Prompt-Based Collaborative Method** *Mengfei Guo (China Mobile Research Institute (CMRI))* *Si Chen (China Mobile Research Institute (CMRI))* *Yi Huang (China Mobile Research)* *Junlan Feng (China Mobile Research)*
P5-04-TLP (#202)	**Taming NLU Noise: Student-Teacher Learning for Robust Dialogue Policy** *Mahdin Rohmatillah (Universitas Brawijaya)* *Jen-Tzung Chien (National Yang Ming Chiao Tung University)*
P5-05-RES (#15)	**HeightCeleb: An Enrichment of VoxCeleb Dataset with Speaker Height Information** *Stanisław Kacprzak (AGH University of Krakow)* *Konrad Kowalczyk (AGH University of Krakow)*
P5-06-RES (#57)	**ESPnet-EZ: Python-Only ESPnet for Easy Fine-Tuning and Integration** *Masao Someki (IBM)* *Kwanghee Choi (Carnegie Mellon University)* *Siddhant Arora (Carnegie Mellon University)* *William Chen (Carnegie Mellon University)* *Samuele Cornell (Carnegie Mellon University)* *Jionghao Han (Carnegie Mellon University)* *Yifan Peng (Carnegie Mellon University)* *Jiatong Shi (Carnegie Mellon University)* *Vaibhav Srivastav (Hugging Face, Inc.)* *Shinji Watanabe (Carnegie Mellon University)*
P5-07-RES (#106)	**Spoken Stereoset: On Evaluating Social Bias Toward Speaker in Speech Large Language Models** *Yi-Cheng Lin (National Taiwan University)* *Wei-Chih Chen (National Taiwan University)* *Hung-Yi Lee (National Taiwan University)*
P5-08-RES (#124)	**Amphion: An Open-Source Audio, Music, and Speech Generation Toolkit** *Xueyao Zhang (The Chinese University of Hong Kong, Shenzhen)*

	Liumeng Xue (The Chinese University of Hong Kong, Shenzhen) *Yicheng Gu (The Chinese University of Hong Kong, Shenzhen)* *Yuancheng Wang (The Chinese University of Hong Kong, Shenzhen)* *Jiaqi Li (The Chinese University of Hong Kong, Shenzhen)* *Haorui He (The Chinese University of Hong Kong, Shenzhen)* *Chaoren Wang (The Chinese University of Hong Kong, Shenzhen)* *Liu Songting (Nanyang Technological University)* *Xi Chen (Chinese University of Hong Kong (Shenzhen))* *Junan Zhang (Fudan University)* *Zihao Fang (The Chinese University of Hong Kong, Shenzhen)* *Haopeng Chen (The Chinese University of Hong Kong, Shenzhen)* *Tze Ying Tang (CUHK-Shenzhen)* *Lexiao Zou (Harbin Institute of Technology (Shenzhen))* *Mingxuan Wang (The Chinese University of Hong Kong, Shenzhen)* *Jun Han (The Chinese University of Hong Kong, Shenzhen)* *Kai Chen (Shanghai AI Laboratory)* *Haizhou Li (The Chinese University of Hong Kong, Shenzhen)* *Zhizheng Wu (The Chinese University of Hong Kong, Shenzhen)*
P5-09-RES (#126)	**Emilia: An Extensive, Multilingual, and Diverse Speech Dataset for Large-Scale Speech Generation** *Haorui He (The Chinese University of Hong Kong, Shenzhen)* *Zengqiang Shang (The Institute of Acoustics of the Chinese Academy of Sciences)* *Chaoren Wang (The Chinese University of Hong Kong, Shenzhen)* *Xuyuan Li (The Institute of Acoustics of the Chinese Academy of Sciences)* *Yicheng Gu (The Chinese University of Hong Kong, Shenzhen)* *Hua Hua (Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, China)* *Liwei Liu (The Chinese University of Hong Kong, Shenzhen)* *Chen Yang (The Institute of Acoustics of the Chinese Academy of Sciences)* *Jiaqi Li (The Chinese University of Hong Kong, Shenzhen)* *Peiyang Shi (The Institute of Acoustics of the Chinese Academy of Sciences)* *Yuancheng Wang (The Chinese University of Hong Kong, Shenzhen)* *Kai Chen (Shanghai AI Laboratory)* *Pengyuan Zhang (Institute of Acoustics, Chinese Academy of Sciences)* *Zhizheng Wu (The Chinese University of Hong Kong, Shenzhen)*
P5-10-RES (#191)	**FLORAS 50: A Massively Multilingual Multitask Benchmark for Long-Form Conversational Speech** *William Chen (Carnegie Mellon University)* *Brian Yan (Carnegie Mellon University)* *Chih-Chen Chen (TMU)* *Shinji Watanabe (Carnegie Mellon University)*
P5-11-RES	**Massively Multilingual Forced Aligner Leveraging Self-Supervised**

(#295)	**Discrete Units** *Hirofumi Inaguma (Meta)* *Ilia Kulikov (Meta)* *Zhaoheng Ni (Meta AI)* *Sravya Popuri (Meta)* *Paden P Tomasello (Meta)*
P5-12-RES (#298)	**Speech Recognition for Analysis of Police Radio Communication** *Tejes Srivastava (University of Chicago)* *Ju-Chieh Chou (TTIC)* *Priyank Shroff (University of Chicago)* *Karen Livescu (TTI-Chicago)* *Christopher Graziul (University of Chicago)*
P5-13-RES (#307)	**Large Language Models as User-Agents for Evaluating Task-Oriented Dialogue Systems** *Taaha Kazi (University of Illinois at Urbana-Champaign)* *Ruiliang Lyu (University of Illinois at Urbana-Champaign)* *Sizhe Zhou (University of Illinois at Urbana-Champaign)* *Dilek Hakkani-Tur (University of Illinois, Urbana-Champaign)* *Gokhan Tur (Amazon)*
P5-14-RES (#334)	**DFADD: The Diffusion and Flow-Matching Based Audio Deepfake Dataset** *Jiawei Du (National Taiwan University)* *I-Ming Lin (National Taiwan University)* *I-Hsiang Chiu (National Taiwan University)* *Xuanjun Chen (National Taiwan University)* *Haibin Wu (National Taiwan University)* *Wenze Ren (National Taiwan University)* *Yu Tsao (Academia Sinica)* *Hung-Yi Lee (National Taiwan University)* *Roger Jang*
P5-15-RES (#384)	**SpMis: An Investigation of Synthetic Spoken Misinformation Detection** *Peizhuo Liu (The Chinese University of Hong Kong, Shenzhen)* *Li Wang (The Chinese University of Hong Kong, Shenzhen)* *He Renqiang (The Chinese University of Hong Kong, Shenzhen)* *Haorui He (The Chinese University of Hong Kong, Shenzhen)* *Lei Wang (Huawei International)* *Huadi Zheng (Huawei Technology)* *Jie Shi (Huawei International)* *Tong Xiao (Northeastern University)* *Zhizheng Wu (The Chinese University of Hong Kong, Shenzhen)*

P5-16-MLS (#7)	**Self-Supervised Speech Models for Word-Level Stuttered Speech Detection** *Yi-Jen Shih (The University of Texas at Austin)* *Zoi Gkalitsiou (UT Austin)* *Alex Dimakis (UT Austin)* *David Harwath (The University of Texas at Austin)*
P5-17-MLS (#30)	**Enhancing Automatic Speech Assessment Leveraging Heterogeneous Features and Soft Labels for Ordinal Classification** *Wen Hsuan Peng (National Taiwan Normal University)* *Sally Chen (The Language Training & Testing Center)* *Berlin Chen (National Taiwan Normal University)*
P5-18-MLS (#92)	**Speech Recognition-Based Feature Extraction for Enhanced Automatic Severity Classification in Dysarthric Speech** *Jeehyun Lee (Sogang University)* *Yerin Choi (Sogang University)* *Myoung-Wan Koo (Sogang University)*
P5-19-MLS (#144)	**Efficient Training of Self-Supervised Speech Foundation Models on a Compute Budget** *Andy T. Liu (National Taiwan University)* *Yi-Cheng Lin (National Taiwan University)* *Haibin Wu (National Taiwan University)* *Stefan Winkler (National University of Singapore)* *Hung-Yi Lee (National Taiwan University)*
P5-20-MLS (#175)	**Improving Anomalous Sound Detection via Low-Rank Adaptation Fine-Tuning of Pre-Trained Audio Models** *Xinhu Zheng (Tsinghua University)* *Anbai Jiang (Tsinghua University)* *Bing Han (Shanghai Jiao Tong University)* *Yanmin Qian (Shanghai Jiao Tong University)* *Pingyi Fan (Tsinghua University)* *Jia Liu (Tsinghua University)* *Wei-Qiang Zhang (Tsinghua University)*
P5-21-MLS (#233)	**Exploring ASR-Based Wav2Vec2 for Automated Speech Disorder Assessment: Insights and Analysis** *Tuan Manh Nguyen (LIA, Avignon University)* *Corinne Fredouille (Avignon Université- LIA)* *Alain Ghio (Aix-Marseille University, LPL)*

	Mathieu Balaguer (IRIT) *Virginie Woisard (Hospitals of Toulouse)*
P5-22-MLS (#249)	**Hierarchical Multi-Path and Multi-Model Selection for Fake Speech Detection** *Chang Feng (Tsinghua University)* *Yiyang Zhao (Tsinghua University)* *Guangzhi Sun (University of Cambridge Department of Engineering)* *Zehua Chen (Tsinghua University)* *Shuai Wang (Shenzhen Research Institute of Big Data, Chinese University of Hong Kong (Shenzhen))* *Chao Zhang (Tsinghua University)* *Mingxing Xu (Tsinghua University)* *Thomas Fang Zheng (CSLT, Tsinghua University)*
P5-23-MLS (#251)	**Semi-Supervised Learning for Robust Speech Evaluation** *Huayun Zhang (ASTAR)* *Jeremy H. M. Wong (Institute for Infocomm Research)* *Geyu Lin (Agency of Science and Technology Research)* *Nancy Chen (Institute for Infocomm Research)*
P5-24-MLS (#264)	**GE2E-KWS: Generalized End-to-End Training and Evaluation for Zero-Shot Keyword Spotting** *Pai Zhu (Google)* *Jacob W Bartel (Google LLC)* *Dhruuv Agarwal (Google LLC)* *Kurt Partridge (Google)* *Hyun Jin Park (Google Inc.)* *Quan Wang (Google)*
P5-25-MLS (#312)	**A Simple HMM with Self-Supervised Representations for Phone Segmentation** *Gene-Ping Yang (The University of Edinburgh)* *Hao Tang (The University of Edinburgh)*
P5-26-MLS (#331)	**DASS: Distilled Audio State Space Models are Stronger and More Duration-Scalable Learners** *Saurabhchand Bhati (MIT)* *Yuan Gong (Massachusetts Institute of Technology)* *Leonid Karlinsky (MIT-IBM Watson AI Lab, IBM Research)* *Hilde Kuehne (University of Bonn)* *Rogerio Feris (MIT-IBM Watson AI Lab, IBM Research)* *James Glass (Massachusetts Institute of Technology)*

P5-27-MLS (#337)	**RAND: Robustness Aware Norm Decay for Quantized Neural Networks** *David Qiu (Google)* *David Rim (Google)* *Shaojin Ding (Google)* *Oleg Rybakov (Google)* *Yanzhang He (Google)*
P5-28-MLS (#377)	**SWIM: Short-Window CNN Integrated with Mamba for EEG-Based Auditory Spatial Attention Decoding** *Ziyang Zhang (Tsinghua University)* *Andrew Thwaites (University College London)* *Alexandra Woolgar (University of Cambridge)* *Brian C.J. Moore (University of Cambridge)* *Chao Zhang (Tsinghua University)*

10:30-12:30 Challenge Session 5: FutureDial RAG (Venue: Lecture Hall)

12:30-14:00 Lunch

14:00-15:00 Invited Talk 3 (Venue: Lecture Hall)

Title: Towards Safe, Truly Open, and Factual Large Language Models
Speaker: *Prof Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi*
Chair: *Dr Zhijian Ou*

15:00-15:30 Coffee Break

15:30-17:30 Panel Discussion

18:00-20:30 Gala Dinner

Day 4, Dec 5, Thursday

09:00-10:00 Keynote Speech 4 (Venue: Lecture Hall)

Title: A Theory of Unsupervised Speech Recognition
Speaker: *Prof. Mark Hasegawa-Johnson, University of Illinois at Urbana-Champaign*
Chair: *Prof Kong Aik Lee*

10:00-10:30 Coffee Break

10:30-12:30 Poster Session 6: Spoken Language Processing (Venue: Poster Area)

Chair: Dr Nan Yan

Poster ID (Paper ID)	Title and Authors
P3-28-SS04 (#411)	**The Database and Benchmark for the Source Speaker Tracing Challenge 2024** *Ze Li (Wuhan University)* *Yuke Lin (Duke Kunshan University)* *Yao Tian (AI Center, OPPO)* *Hongbin Suo (AI Center, OPPO)* *Pengyuan Zhang (Institute of Acoustics, Chinese Academy of Sciences)* *Yanzhen Ren (Computer School of Wuhan University)* *Zexin Cai (Johns Hopkins University)* *Hiromitsu Nishizaki (University of Yamanashi)* *Ming Li (Duke Kunshan University)*
P6-01-SLP (#65)	**Stutter-Solver: End-to-End Multi-Lingual Dysfluency Detection** *Xuanru Zhou (Berkeley Speech Group)* *Cheol Jun Cho (UC Berkeley)* *Ayati Sharma (University of California, Berkeley)* *Brittany Morin (UCSF)* *David Baquirin (UCSF)* *Jet Vonk (UCSF)* *Zoe Ezzes (UCSF)* *Zachary Miller (UCSF)* *Boon Lead Tee (UCSF)* *Maria Luisa Gorno Tempini (UCSF)* *Jiachen Lian (University of California, Berkeley)* *Gopala Krishna Anumanchipalli (UC Berkeley)*
P6-02-SS09 (#417)	**Domain Adaption and Unified Knowledge Base Motivate Better Retrieval Models in Dialog Systems with RAG** *Huadong Lin (South China University of Technology)* *Yirong Chen (South China University of Technology)* *Wenyu Tao (South China University of Technology)* *Mingyu Chen (South China University of Technology)* *Xiangmin Xu (South China University of Technology)* *Xiaofen Xing (South China University of Technology)*
P6-03-SLP (#153)	**SSAMBA: Self-Supervised Audio Representation Learning with Mamba State Space Model** *Siavash Shams (Columbia University)*

	Sukru Samet Dindar (Columbia University) *Xilin Jiang (Columbia University)* *Nima Mesgarani (Columbia University)*
P6-04-SLP (#215)	**Speech-Copilot: Leveraging Large Language Models for Speech Processing via Task Decomposition, Modularization, and Program Generation** *Chun-Yi Kuan (National Taiwan University)* *Chih-Kai Yang (National Taiwan University)* *Wei-Ping Huang (National Taiwan University)* *Ke-Han Lu (National Taiwan University)* *Hung-Yi Lee (National Taiwan University)*
P6-05-SLP (#174)	**CTC-GMM: CTC-Guided Modality Matching for Fast and Accurate Streaming Speech Translation** *Rui Zhao (Microsoft)* *Jinyu Li (Microsoft)* *Ruchao Fan (Microsoft)* *Matt Post (Microsoft)*
P6-06-SLP (#223)	**Long-Form End-to-End Speech Translation via Latent Alignment Segmentation** *Peter Polák (Charles University)* *Ondrej Bojar (Charles University)*
P6-07-SLP (#272)	**Confidence Estimation for LLM-Based Dialogue State Tracking** *Yijyun Sun (University of Illinois, Urbana-Champaign)* *Suvodip Dey (University of Illinois, Urbana-Champaign)* *Dilek Hakkani-Tur (University of Illinois, Urbana-Champaign)* *Gokhan Tur (Amazon)*
P6-08-SLP (#278)	**The 2nd FutureDial Challenge: Dialog Systems with Retrieval Augmented Generation (FutureDial-RAG)** *Yucheng Cai (Tsinghua University)* *Si Chen (China Mobile Research)* *Yuxuan Wu (Tsinghua University)* *Yi Huang (China Mobile Research)* *Junlan Feng (China Mobile Research)* *Zhijian Ou (Tsinghua University)*
P6-09-SLP (#130)	**Zero-Shot Audio Topic Reranking Using Large Language Models** *Mengjie Qian (Cambridge University)* *Rao Ma (University of Cambridge)*

	Aidan Liusie (University of Cambridge) *Erfan Loweimi (University of Cambridge)* *Katherine M Knill (University of Cambridge)* *Mark Gales (University of Cambridge)*
P6-10-SLP (#59)	**Clean Label Attacks Against SLU Systems** *Henry Li Xinyuan (Johns Hopkins University)* *Thomas Thebaud (Johns Hopkins University)* *Sonal Joshi (Johns Hopkins University)* *Jesus Antonio Villalba (Johns Hopkins University)* *Najim Dehak (Johns Hopkins University)* *Sanjeev Khudanpur (Johns Hopkins University)*
P6-11-SLP (#168)	**WHISMA: A Speech-LLM to Perform Zero-Shot Spoken Language Understanding** *Mohan Li (Toshiba Europe Ltd.)* *Cong-Thanh Do (Toshiba Research Europe Ltd.)* *Simon Keizer (Toshiba Europe Ltd.)* *Youmna Farag (Toshiba Europe Ltd.)* *Svetlana Stoyanchev (Toshiba Europe Ltd.)* *Rama S Doddipatla (Toshiba Europe Ltd.)*
P6-12-SLP (#169)	**Improving Transducer-Based Spoken Language Understanding with Self-Conditioned CTC and Knowledge Transfer** *Vishal Sunder (The Ohio State University)* *Eric Fosler-Lussier (The Ohio State University)*
P6-13-SLP (#208)	**Self-Supervised Syllable Discovery Based on Speaker-Disentangled HuBERT** *Ryota Komatsu (Independent Researcher)* *Takahiro Shinozaki (Tokyo Institute of Technology)*
P6-14-SLP (#355)	**Just ASR + LLM? A Study on Speech Large Language Models' Ability to Identify and Understand Speakers in Spoken Dialogue** *Junkai Wu (University of Washington)* *Xulin Fan (University of Illinois at Urbana-Champaign)* *Bo-Ru Lu (University of Washington)* *Xilin Jiang (Columbia University)* *Nima Mesgarani (Columbia University)* *Mark A Hasegawa-Johnson (University of Illinois)* *Mari Ostendorf (University of Washington)*
P6-15-SLR (#5)	**Enhancing Open-Set Speaker Identification through Rapid Tuning with Speaker Reciprocal Points and Negative Samples**

	Zhiyong Chen (Shanghai University) *Zhiqi Ai (Shanghai University)* *Xinnuo Li (Shanghai University)* *Shugong Xu (Shanghai University)*
P6-16-SLR (#10)	**Spoofing-Aware Speaker Verification Robust Against Domain and Channel Mismatches** *Chang Zeng (National Institute of Informatics)* *Xiaoxiao Miao (Singapore Institute of Technology)* *Xin Wang (National Institute of Informatics)* *Erica Cooper (National Institute of Information and Communications Technology)* *Junichi Yamagishi (National Institute of Informatics)*
P6-17-SLR (#49)	**Adversarial Purification for Speaker Verification by Two-Stage Diffusion Models** *Yibo Bai (The University of Hong Kong)* *Zhang Xiaolei (Northwestern Polytechnical University)* *Xuelong Li (Institute of Artificial Intelligence (TeleAI), China Telecom Corp. Ltd.)*
P6-18-SLR (#103)	**Measuring Sound Symbolism in Audio-Visual Models** *Wei-Cheng Tseng (The University of Texas at Austin)* *Yi-Jen Shih (The University of Texas at Austin)* *David Harwath (The University of Texas at Austin)* *Raymond Mooney (The University of Texas at Austin)*
P6-19-SLR (#111)	**Meta-Learning Approaches for Improving Detection of Unseen Speech Deepfakes** *Ivan Kukanov (KLASS Engineering and Solutions)* *Janne Laakkonen (UEF)* *Tomi H. Kinnunen (University of Eastern Finland)* *Ville Hautamäki (University of Eastern Finland)*
P6-20-SLR (#135)	**On the Generation and Removal of Speaker Adversarial Perturbation for Voice-Privacy Protection** *Chenyang Guo (University of Science and Technology of China)* *Liping Chen (University of Science and Technology of China)* *Zhuhai Li (University of Science and Technology of China)* *Kong Aik Lee (The Hong Kong Polytechnic University)* *Zhen-Hua Ling (University of Science and Technology of China)* *Wu Guo (University of Science and Technology of China)*

P6-21-SLR (#138)	**Towards Quantifying and Reducing Language Mismatch Effects in Cross-Lingual Speech Anti-Spoofing** *Tianchi Liu (National University of Singapore)* *Ivan Kukanov (KLASS Engineering and Solutions)* *Zihan Pan (Institute for Infocomm Research (I2R), ASTAR, Singapore)* *Qiongqiong Wang (ASTAR)* *Hardik B Sailor (I2R, ASTAR, Singapore)* *Kong Aik Lee (The Hong Kong Polytechnic University)*
P6-22-SLR (#141)	**Enhancing Low-Resource Spoken Language Identification via Cross-Modality Retrieval and Cross-Lingual Text-to-Speech Synthesis** *Min Ma (Google DeepMind)* *Yuan Wang (Google)* *Kyle Kastner (Google)* *Isaac Caswell (Google)* *Charles Yoon (Google)* *Andrew Rosenberg (Google LLC)*
P6-23-SLR (#148)	**Recursive Attentive Pooling for Extracting Speaker Embeddings from Multi-Speaker Recordings** *Shota Horiguchi (NTT Corporation)* *Atsushi Ando (NTT Corporation)* *Takafumi Moriya (NTT Corporation)* *Takanori Ashihara (NTT Corp.)* *Hiroshi Sato (NTT)* *Naohiro Tawara (NTT)* *Marc Delcroix (NTT)*
P6-24-SLR (#259)	**PDAF: A Phonetic Debiasing Attention Framework for Speaker Verification** *Massa Baali (CMU)* *Abdulhamid Aldoobi (Carnegie Mellon University)* *Hira Dhamyal (Carnegie Mellon University)* *Rita Singh (Carnegie Mellon University)* *Bhiksha Raj (Carnegie Mellon University)*
P6-25-SLR (#356)	**INX-SpeakerHub: A 2000-Hour Indian Multilingual Speaker Identification Corpus** *Metilda Sagaya Mary N J (Indian Institute of Technology Madras)* *S Umesh (IIT Chennai)*
P6-26-DIA (#100)	**Resource-Efficient Adaptation of Speech Foundation Models for Multi-Speaker ASR**

	Weiqing Wang (NVIDIA) *Kunal Dhawan (NVIDIA)* *Tae Jin Park (NVIDIA)* *Krishna C Puvvada (NVIDIA)* *Ivan Medennikov (NVIDIA)* *Somshubra Majumdar (NVIDIA)* *He Huang (NVIDIA)* *Jagadeesh Balam (NVIDIA)* *Boris Ginsburg (NVIDIA)*
P6-27-SS01 (#405)	**Exploring Self-Supervised Representations for Text-Dependent Speaker Verification** *Sankala Sreekanth (Indian Institute of Technology Hyderabad (IITH))*
P6-28-SS04 (#147)	**Distillation-Based Feature Extraction Algorithm for Source Speaker Verification** *Xinlei Ma (Tianjin University)* *Wenhuan Lu (Tianjin University)* *Ruiteng Zhang (Tianjin University)* *Junhai Xu (Tianjin Key Laboratory of Cognitive Computing and Application, College of Intelligence and Computing, Tianjin University)* *Xugang Lu (NICT)* *Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China)*
P6-29-SS04 (#224)	**Speaker Contrastive Learning for Source Speaker Tracing** *Qing Wang (Northwestern Polytechnical University)* *Hongmei Guo (Northwestern Polytechnical University)* *Jian Kang (Institute of Artificial Intelligence (TeleAI), China Telecom)* *Mengjie Du (China Telecom)* *Jie Li (Institute of Artificial Intelligence (TeleAI), China Telecom)* *Zhang Xiaolei (Northwestern Polytechnical University)* *Lei Xie (NWPU)*

10:30-12:30 Challenge Session 6: GenASR challenge (Venue: Lecture Hall)

12:30-13:00 Closing Ceremony (Venue: Lecture Hall)