National Conference on
Man-Machine Speech
Communication

2023
NCMMSC

第十八届
全国人机语音通讯学术会议

2023年12月8-10日　江苏·苏州

Friday, December 8

Saturday, December 9

Sunday, December 10

Friday, December 8

Fri-M-T-1: Speaker Representation Learning: Theories, Applications and Practice
Room: 107
Type: Tutorial
December 8, 9:00 AM to 10:30 PM
Presenter: 王帅老师，韩冰博士
Chair: 钱彦旻老师

Fri-M-T-2: 语音预训练
Room: 107
Type: Tutorial
December 8, 10:45 AM to 12:15 PM
Presenter: 刘树杰老师
Chair: 钱彦旻老师

Fri-A-T-1: 语音可懂度的客观评估技术
Room: 107
Type: Tutorial
December 8, 1:00 PM to 2:30 PM
Presenter: 陈霏老师
Chair: 武执政老师

Fri-A-T-2: Building the Singing Voice Foundation Model
Room: 107
Type: Tutorial
December 8, 2:45 PM to 4:15 PM
Presenter: 雪巍老师
Chair: 武执政老师

Fri-A-IND: Industry Forum 1
Room: 107
Type: Industry Forum
December 8, 4:30 PM to 6:00 PM
Chair: 张超老师

16:30-17:00
Fri-A-IND-1: 对话式人工智能技术及其产业应用
Presenter: 樊帅 (思必驰)

17:00-17:25
Fri-A-IND-2: OPPO 小布助手的语音技术应用实践
Presenter: 田垚 (OPPO)

17:25-17:40
Fri-A-IND-3: 魔搭语音和音频多模态大模型研究和开源进展
Presenter: 张仕良 (阿里云)

17:40-17:55
Fri-A-IND-4: 智能语音技术赋能语言学习
Presenter: 王龙标 (智言科技)

Saturday, December 9

Sat-M-PLEN-1: ChatGLM：从千亿模型到 ChatGPT 的一点思考
Room: 206
Type: Plenary
December 9, 9:15 AM to 10:00 AM
Presenter: 唐杰老师
Chair: 贾珈老师

Sat-M-PLEN-2: Harnessing Machine Learning with Science in Speech and Language Technologies
Room: 206
Type: Plenary
December 9, 10:00 AM to 10:45 AM
Presenter: Helen Meng 老师
Chair: 贾珈老师

Sat-M-O-1: Best Paper Presentation
Room: 108
Type: Oral
December 9, 11:00 AM to 12:30 PM
Chair: 陈景东老师

Sat-M-O-1-1: Adaptive Deep Graph Convolutional Network For Dialogical Speech Emotion Recognition
Jiaxing Liu (Tianjin University); Sheng Wu (Tianjin University)*; Longbiao Wang (Tianjin University); Jianwu Dang (Tianjin University)

Sat-M-O-1-2: CAM-GUI: A Conversational Assistant on Mobile GUI
Zichen Zhu (Shanghai Jiao Tong University)*; Liangtai Sun (Shanghai Jiao Tong University); Jingkai Yang (Shanghai Jiao Tong University); Yifan Peng (Shanghai Jiao Tong University); Weilin Zou (Shanghai Jiao Tong University); Ziyuan Li (Shanghai Jiao Tong University); Wutao Li (Shanghai Jiao Tong University); Lu Chen (Shanghai Jiao Tong University); Yingzi Ma (Sichuan University); Danyang Zhang (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sat-M-O-1-3: Parkinson's-Specific Featureset under Read Speech Experiment
Pingyue Zhang (Shanghai Jiao Tong University); Mengyue Wu (Shanghai Jiao Tong University)*; Kai Yu (Shanghai Jiao Tong University)

Sat-M-O-1-4: APNet2: High-quality and High-efficiency Neural Vocoder with Direct Prediction of Amplitude and Phase Spectra
Hui-Peng Du (University of Science and Technology of China)*; Ye-Xin Lu (University of Science and Technology of China); Yang Ai (University of Science and Technology of China); Zhen-Hua Ling (University of Science and Technology of China)

Sat-M-O-1-5: 基于高可迁移性的通用扰动的快速对抗样本生成方法
Wang Zhen (Harbin Institute of Technology)*; Jiqing Han (Harbin Institute of Technology); 何勇军 (哈尔滨工业大学); 郑铁然 (哈尔滨工业大学); 郑贵滨 (哈尔滨工业大学)

Sat-M-O-1-6: 基于无注意力和分层模型的阿尔茨海默氏症检测方法

Zhongren Dong (Hunan University)*; Zixing Zhang (Hunan University)

Sat-M-O-2: Speech Recognition and Keyword Spotting
Room: 104
Type: Oral
December 9, 11:00 AM to 12:30 PM
Chair: 李明老师

Sat-M-O-2-1: An MFCCA and Speaker Diarization Based SA-ASR System for the Multi-channel Multi-party Meeting Transcription

Luzhen Xu (University of Science and Technology of China)*; Haoyin Yan (University of Science and Technology of China); Mao-Kui He (University of Science and Technology of China); Zixian Guo (University of Science and Technology of China); Zhou Yeping (China Merchants Bank); Peiqi Liu (CMB China); Jie Zhang (University of Science and Technology of China (USTC)); Lirong Dai (University of Science and Technology of China)

Sat-M-O-2-2: Research on language-assisted Mandarin-English Code-switching ASR

Hukai Huang (Xiamen University)*; Jiayan Lin (Xiamen University); Qingyang Hong (Xiamen University); Lin Li (Xiamen University)

Sat-M-O-2-3: Wav2vec-AD: Acoustic Unit Discovery Module-Integrated, Self-Supervised Contrastive Pre-training Approach for Speech Recognition

Nurmemet Yolwas (Xinjiang Univerisity); Lixu Sun (Xinjiang Univerisity)*; Xin Li (Xinjiang Univerisity); Qichao Liu (Xinjiang Univerisity); Zhixiang Wang (Xinjiang Univerisity)

Sat-M-O-2-4: EC-BERT: A BERT language model with Error Correction for Mandarin Chinese Speech Recognition

Sujie Xiao (University of Chinese Academy of Sciences;Institute of Acoustics, Chinese Academy of Sciences)*; Ruipeng Hao (Institute of Acoustics, Chinese Academy of Sciences); Gaofeng Cheng (Institute of Acoustics, Chinese Academy of Sciences); Xiaoyan Xu (Institute of Acoustics, Chinese Academy of Sciences); Ta Li (Institute of Acoustics, Chinese Academy of Sciences)

Sat-M-O-2-5: Advancing VAD Systems Based on Multi-Task Learning with Improved Model Structures

Lingyun Zuo (Speech Lab, Alibaba Group)*; Keyu An (Speech Lab, Alibaba Group, China); Shiliang Zhang (Alibaba Group); Zhijie Yan (Alibaba Inc.)

Sat-M-O-2-6: End-to-end Streaming Customizable Keyword Spotting based on text-adaptive neural search

Baochen Yang (Shanghai Jiao Tong University)*; Jiaqi Guo (AISpeech Ltd, Suzhou China); Yu Xi (Shanghai Jiao Tong University); Haoyu Li (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sat-M-O-2-7: 基于有限训练数据和开放集学习的鲁棒小型关键词检测系统

Zijun Huang (Northwestern Polytechnical University)*; Zhang Xiaolei (Northwestern Polytechnical University)

Sat-M-SS: 半监督智能语音语言技术
Room: 206
Type: Special Session
December 9, 11:00 AM to 12:30 PM
Organizer: 欧智坚老师，张鹏远老师
Chair: 欧智坚老师

Sat-M-SS-1: 语音识别中的半监督学习范式思考
Presenter: 颜永红老师，中国科学院声学研究所首席科学家

Sat-M-SS-2: Large-Scale Foundation Model for Speech Generation
Presenter: Wei-Ning Hsu (徐煒甯) 老师，Meta Fundamental AI Research (FAIR) Research Scientist / Audio Generation Team Lead

Sat-M-SS-3: 大规模语言模型的领域探索——以法律场景为例
Presenter: 冯岩松老师，北京大学王选计算机研究所副教授

Sat-M-SS-4: 零样本跨语言语音合成与翻译
Presenter: 刘树杰老师，微软亚洲研究院首席研究员

Sat-M-P-1: Speech Coding and Enhancement
Room: 105
Type: Poster
December 9, 11:00 AM to 12:30 PM

Sat-M-P-1-1: 用于听觉注意检测的音视频脑电数据集
Hongyu Zhang (Anhui University School of Computer Science and Technology); Jingjing Zhang (Anhui University); Xingguang Dong (Anhui University); Zhao Lv (anhui university); Jianhua Tao (Department of Automation, Tsinghua University); Jian Zhou (Anhui University); Xiaopei Wu (Anhui University School of Computer Science and Technology); Cunhang Fan (Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, School of Computer Science and Technology, Anhui University)*

Sat-M-P-1-2: 基于梅尔谱与压缩激励加权量化的语音神经编解码方法
Junzuo Zhou (Institute of Automation, Chinese Academy of Sciences)*; Jiangyan Yi (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences); Jianhua Tao (Department of Automation, Tsinghua University); Yong Ren (Institute of Automation, Chinese Academy of Sciences); Tao Wang (Institute of Automation, Chinese Academy of Sciences)

Sat-M-P-1-3: 用于目标语音增强的去噪扩散条件期望模型
Leying Zhang (Shanghai Jiao Tong University)*; Yanmin Qian (Shanghai Jiao Tong University)

Sat-M-P-1-4: 融合 MVDR 和多目标 GAN 网络的多通道语音增强
Tuo Liu (Shenzhen University)*; Zhijian Jiang (shenzhen university); Yunda Chen (Shenzhen University); Yupeng Shi (Shenzhen University); Nengheng Zheng (Shenzhen University)

Sat-M-P-1-5: Joint speech and noise estimation using SNR-adaptive target learning for deep-learning-based speech enhancement
Xiaoran Li (USTC)*; Zilu Guo (USTC); Jun Du (University of Science and Technology of China); Chin-Hui Lee (Georgia Institute of Technology); Yu Gao (Midea); Wenbin Zhang (Midea)

Sat-M-P-1-6: Improving Speech Intelligibility through Sub-band temporal envelope Characteristics
Ruilin Wu (Xinjiang University); Zhihua Huang (Xinjiang university)*; Jingyi Song (Xinjiang University); Xiaoming Liang (Xinjiang University)

Sat-M-P-1-7: 基于两阶段神经网络的低复杂度实时声学回声消除系统

Haoming Li (Shenzhen University)*; Weitong Zhao (Shenzhen University); Zhijian Jiang (shenzhen university); Yuyong Kang (Shenzhen Jingxinhao Technology Co., Ltd); Nengheng Zheng (Shenzhen University)

Sat-M-P-1-8: A Bone Conducted Speech Based Loss Function For Speech Enhancement

Qiu Keying (Shenzhen University)*; Yunda Chen (Shenzhen University); Nengheng Zheng (Shenzhen University)

Sat-M-P-1-9: Implementation of Multichannel Speech Coding Method Based on the Opus Codec and Spatial Parameters

Xue Yang (Beijing University of Technology); Changchun Bao (Beijing University of Technology)*; Jing Zhou (Beijing University of Technology); Xu Zhang (Beijing University of Technology); Haiwei Duan (Beijing University of Technology); Yunhao Zhao (Beijing University of Technology); Wenwen Li (Beijing University of Technology)

Sat-M-P-1-10: Target Speech Separation with Steering Vector Beamforming in Multi-Channel Microphone

Xiaokang Yang (Tianjin University)*; Wenjun Ke (Tianjin University); 臣白国 (天津大学); Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China)

Sat-M-P-1-11: 基于差和比和质心修正的风噪抑制算法

张瑜聪 (Shenzhen university)*; 孙伟 (华勤技术股份有限公司); 陈运达 (Shenzhen university); Nengheng Zheng (Shenzhen University); Zhijian Jiang (shenzhen university)

Sat-M-P-2: Analysis of Speech and Audio Signals
Room: 105
Type: Poster
December 9, 11:00 AM to 12:30 PM

Sat-M-P-2-1: Multi-source DOA Estimation Method Based on the Signal Subspace within the RSP Domain

Haiwei Duan (Beijing University of Technology); Changchun Bao (Beijing University of Technology)*

Sat-M-P-2-2: Multi-Branch Network with Cross-domain Feature Fusion for Anomalous Sound Detection

Wenjie Fang (Xinjiang University)*; Ying Hu (Xinjiang University); Xin Fan (Xinjiang University)

Sat-M-P-2-3: 基于声矢量传感器网络的声源定位算法

Wangzhe Jin (Inner Mongolia University); De Hu (Inner Mongolia UNiversity)*; Xu Wang (Inner Mongolia University); Rui Liu (Inner Mongolia University)

Sat-M-P-2-4: ADA-SID：基于无监督对抗域自适应解决歌手识别中的专辑效应问题

Yu Xiao (Tianjin University)*; Wenhuan Lu (Tianjin University); Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China); Ruiteng Zhang (Tianjin University); Yuting Zhao (Tianjin University)

Sat-M-P-2-5: Singular Spectrum Harmonic Decomposition for the non-stationary analysis of vibration signals

Zhang Hongcheng (天津大学)*

Sat-M-P-2-6: Integration of Short-Term and Long-Term Harmonic Peaks in a Two-Level Discriminative Weight Training Framework for Voice Activity Detection

Yingwei Tan (Volkswagen-Mobvoi (Beijing) Information Technology Co., Ltd)*

Sat-M-P-2-7: Respiratory Sound Classification based on Self-Supervised Model
Shiyu Miao (Xiamen University)*; Qi Su (Xiamen University); Yishuang Li (Xiamen University); Lin Li (Xiamen University); Qingyang Hong (Xiamen University)

Sat-M-P-2-8: Iteratively Refined multi-channel Speech Separation Based on Beam-TFDPRNN
Xu Zhang (Beijing University of Technology); Changchun Bao (Beijing University of Technology)*

Sat-M-P-2-9: Improving the Performance of Speech SSL Student Model by Fusing FBANK in Knowledge Distillation
Qiuyi Lu (Xinjiang University)*; Liang He (Tsinghua University); Yuanyuan Leng (Xinjiang University)

Sat-M-P-2-10: Noise-Aware Masked Autoencoders for Voice Activity Detection
Yingwei Tan (Volkswagen-Mobvoi (Beijing) Information Technology Co., Ltd)*; Xuefeng Ding (Volkswagen-Mobvoi (Beijing) Information Technology Co., Ltd)

Sat-M-P-2-11: Fine-Grained Heart Sound Classification for Congenital Heart Defects Using Multiple Deep Learning Techniques
Yiyang Jiang (Shanghai Artificial Intelligence Laboratory); Zihan Zhao (Shanghai Jiao Tong University)*; Yuxin Duan (Shanghai Jiao Tong University); Yan-Feng Wang (Cooperative medianet innovation center of Shanghai Jiao Tong University); Xin Sun (Xinhua Hospital Affiliated To Shanghai Jiaotong University School Of Medicine); Yu Wang (Shanghai Jiao Tong University)

Sat-M-P-3: Human-Machine Interaction
Room: 105
Type: Poster
December 9, 11:00 AM to 12:30 PM

Sat-M-P-3-1: A Study of Mispronunciation Detection and Diagnosis Based on Meta-learning
Yukai Wan (Beijing Language & Culture University SAIT Lab)*; Yuqi Shi (Beijing Language and Culture University); Binghuai Lin (Tencent Technology Co.); Yanlu Xie (Beijing Language and Culture University)

Sat-M-P-3-2: Time-Frequency Generative Adversarial Network with Transformer Module for Packet Loss Concealment
Yunhao Zhao (Beijing University of Technology); Changchun Bao (Beijing University of Technology)*

Sat-M-P-3-3: 回放样本增强的持续学习
Rui Niu (Tsinghua University)*; Zhiyong Wu (Tsinghua University); Jia Jia (Tsinghua University)

Sat-M-P-3-4: A System Self-Noise Suppression Method based on Multi-modal and Perceptual Loss Function
Yan Zhou (Northwestern Polytechnical University)*

Sat-M-P-3-5: 聋人健美操视频学习注意力机制研究
Yu Chen (Tianjin University of Technology)*; 王王明华 (天津理工大学); Ju Zhang (Tianjin University); 李红艳 (天津理工大学)

Sat-M-P-4: ASRU Paper Seminar: Showcasing and Discussing
Room: 105
Type: Poster
December 9, 11:00 AM to 12:30 PM

Sat-M-P-4-1: An Exploration of Task-decoupling on Two-stage Neural Post Filter for Real-time Personalized Acoustic Echo Cancellation
Presenter: 张子晗 (西北工业大学)

Sat-M-P-4-2: Bayesian Sample Selection for Speech-based In-context Learning
Presenter: 王嗣尹 (清华大学)

Sat-M-P-4-3: BiSinger: Bilingual Singing Voice Synthesis
Presenter: 周华丽 (武汉大学)

Sat-M-P-4-4: Efficient Text-Only Domain Adaptation for CTC-based ASR
Presenter: 龚勋 (上海交通大学)

Sat-M-P-4-5: Exploring Effective Distillation of Self-Supervised Speech Models for Automatic Speech Recognition
Presenter: 唐昌礼 (清华大学)

Sat-M-P-4-6: Fast-HuBERT: An Efficient Training Framework for Self-Supervised Speech Representation Learning
Presenter: 杨冠柔 (上海交通大学)

Sat-M-P-4-7: FAT-HuBERT: Front-end Adaptive Training of Hidden-unit BERT for Distortion-Invariant Robust Speech Recognition
Presenter: 杨冬宁 (上海交通大学)

Sat-M-P-4-8: Haha-Pod: An Attempt for Laughter-based Non-Verbal Speaker Verification
Presenter: Yuke Lin (杜克昆山大学)

Sat-M-P-4-9: HiGNN-TTS : Hierarchical Prosody Modeling with Graph Neural Networks for Expressive Long-form TTS
Presenter: 郭大可 (西北工业大学)

Sat-M-P-4-10: Improving Speech Enhancement Using Audio Tagging Knowledge from Pre-trained Representations and Multi-task Learning
Presenter: 林少雄 (上海交通大学)

Sat-M-P-4-11: SALT: Distinguishable Speaker Anonymization Through Latent Space Transformation
Presenter: 吕元骏 (西北工业大学)

Sat-M-P-4-12: Spike-Triggered Contextual Biasing for End-to-End Mandarin Speech Recognition
Presenter: 黄凯勋 (西北工业大学)

Sat-M-P-4-13: Transferring speech-generic and depression-specific knowledge for Alzheimer's disease detection
Presenter: Ziyun Cui (清华大学)

Sat-M-P-4-14: VITS-Based Singing Voice Conversion Leveraging Whisper and multi-scale F0 Modeling
Presenter: 宁子谦 (西北工业大学)

Sat-M-P-4-15: VoiceExtender: Short-utterance Text-independent Speaker Verification with Guided Diffusion Model
Presenter: 张旭龙 (平安科技)

Sat-M-P-4-16: ZERO-SHOT EMOTION TRANSFER FOR CROSS-LINGUAL SPEECH SYNTHESIS
Presenter: 李玉珂 (西北工业大学)

Sat-M-YF: Young Forum 1
Room: 206
Type: Young Forum
December 9, 1:30 PM to 3:00 PM
Chair: 张晓雷老师

Sat-M-YF-1: 大语言模型的原理、实现及应用
Presenter: 车万翔老师，哈尔滨工业大学计算学部长聘教授，人工智能研究院副院长

Sat-M-YF-2: AI for Cross-Modal Interaction: Speech-Driven Human Body Language Generation
Presenter: 刘李老师，香港科技大学（广州）人工智能学域助理教授

Sat-M-YF-3: 广域海洋水下多目标被动声学感知
Presenter: 朱晨阳老师，南京邮电大学通信与信息工程学院副研究员

Sat-A-O-1: Speaker Recognition: Verification and Diarization
Room: 108
Type: Oral
December 9, 3:15 PM to 5:15 PM
Chair: 张晓雷老师

Sat-A-O-1-1: A simple unsupervised linear multi-domain adaptation approach for deep speaker embedding
Wan Lin (CSLT, BNRist, Tsinghua University)*

Sat-A-O-1-2: BER: Balanced Error Rate For Speaker Diarization
Tao Liu (Shanghai Jiao Tong University)*; Shuai Fan (AiSpeech); Kai Yu (Shanghai Jiao Tong University)

Sat-A-O-1-3: Semi-supervised Speaker Verification System Based on Pre-trained Model
Yishuang Li (Xiamen University)*; Zhicong Chen (Xiamen University); Shiyu Miao (Xiamen University); Qi Su (Xiamen University); Lin Li (Xiamen University); Qingyang Hong (Xiamen University)

Sat-A-O-1-4: The DKU-MSXF Diarization System for the VoxCeleb Speaker Recognition Challenge 2023
Ming Cheng (Duke Kunshan University)*; Weiqing Wang (Data Science Research Center, Duke Kunshan University, China); Xiaoyi Qin (Dukekunshan University); Yuke Lin (Wuhan University); Ning Jiang (Mashang Consumer Finance Co., Ltd.); Guoqing Zhao (Mashang Consumer Finance Co., Ltd); Ming Li (Duke Kunshan University)

Sat-A-O-1-5: Adapting Universal Perturbation with Improved Perceptual Loss for High-quality Adversarial Examples Generation to Speaker Identification
Zhuhai Li (University of Science and Technology of China)*; Wu Guo (University of Science and Technology of China)

Sat-A-O-1-6: Within- and Between-Class Sample Interpolation Based Supervised Metric Learning for Speaker Verification

Jian-Tao Zhang (University of Science and Technology of China); Hao-Yu Song (The Australian National University); Wu Guo (University of Science and Technology of China); Yan Song (USTC)*; Lirong Dai (University of Science and Technology of China)

Sat-A-O-1-8: Joint Optimization of Speaker Extraction and Diarization

Junyi Ao (The Chinese University of Hong Kong (Shenzhen))*; Gui Chi (The Chinese university of HongKong (shenzhen)); Meng Ge (Tianjin University); Jingru Lin (National University of Singapore); Shuai Wang (Shenzhen Institute of Big Data); Haizhou Li (The Chinese University of Hong Kong (Shenzhen))

Sat-A-O-1-7: 通过多层次域适配器增强跨域说话人验证

Wen Huang (SJTU)*; Bing Han (Shanghai Jiao Tong University); Shuai Wang (Shenzhen Research Institute of Big Data); Zhengyang Chen (Shanghai Jiao Tong University); Yanmin Qian (Shanghai Jiao Tong University)

Sat-A-O-2: Source Separation, Localization and Speech Enhancement
Room: 104
Type: Oral
December 9, 3:15 PM to 5:15 PM
Chair: 付强老师

Sat-A-O-2-1: A Lightweight Music Source Separation Model with Graph Convolution Network

Mengying Zhu (Xinjiang University)*; Ying Hu (Xinjiang University); Liusong Wang (Xinjiang University)

Sat-A-O-2-2: Ultra-Low Complexity Residue Echo and Noise Suppression Based on Recurrent Neural Network

Jianquan Zhou (tencent)*; Jacky Gao (Tencent Chengdu Branch); Serrie Zhang (tencent)

Sat-A-O-2-3: Simultaneous Speech Extraction for Multiple Target Speakers under the Meeting Scenarios

Bang Zeng (Wuhan University)*; Hongbin Suo (OPPO); Yulong Wan (OPPO); Ming Li (Duke Kunshan University)

Sat-A-O-2-4: A Framework Combining Separate and Joint Training for Neural Vocoder-Based Monaural Speech Enhancement

Qiaoyi Pan (Shanghai Jiao Tong University)*; Wenbin Jiang (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sat-A-O-2-5: Iteration Noisy-target Approach: Speech Enhancement without Clean Speech

Yifan Zhang (Shanghai Jiao Tong University)*; Wenbin Jiang (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sat-A-O-2-6: Multi-Resolution Speech Representations for Speech Enhancement and Separation

Jingru Lin (National University of Singapore)*; Wupeng Wang (NUS); Meng Ge (Tianjin University); Haizhou Li (The Chinese University of Hong Kong (Shenzhen))

Sat-A-O-2-7: Joint time-domain and frequency-domain progressive learning for single-channel speech enhancement and recognition

Gongzhen Zou (University of Science and Technology of China)*; Jun Du (University of Science and Technology of China); Shutong Niu (University of Science and Technology of China); Hang Chen (USTC); Chin-Hui Lee (Georgia Institute of Technology); Yuling Ren (china mobile online services company limited); Qinglong Li (China Mobile Online Services Company Limited); Ruibo Liu (China Mobile Online Services Company Limited)

Sat-A-O-2-8: 基于卡尔曼滤波与频率聚焦的单声源到达方向实时估计与跟踪方法

Jing Zhou (Beijing University of Technology); Changchun Bao (Beijing University of Technology)*; Haiwei Duan (Beijing University of Technology)

Sat-A-SS: 语音交流中的神经生理信号解码及其应用
Room: 206
Type: Special Session
December 9, 3:15 PM to 5:15 PM
Organizer: 陈霏老师，张结老师
Chair: 陈霏老师

Sat-A-SS-1: EEG-based Multi-class Auditory Attention Decoding of Attended Speaker Direction
Presenter: 卢晶老师，南京大学教授

Sat-A-SS-2: EEG-based auditory attention decoding with audiovisual speech for hearing-impaired listeners
Presenter: 陈婧，北京大学研究员

Sat-A-SS-3: Sparse EEG Channel Selection for Brain-Assisted Speech Enhancement
Presenter: 张结老师，中国科学技术大学副研究员

Sat-A-SS-4: 解码汉语发音想象及其脑机接口应用
Presenter: 陈霏老师，南方科技大学教授

Sat-A-SS-5: Graph Self-Distillation for EEG-Based Auditory Spatial Attention Detection
Presenter: 范存航老师，安徽大学副教授

Sat-A-SS-6: 听声辨位——基于脑电的空间听觉注意检测
Presenter: 白艳茹老师，天津大学副教授

Sat-A-P-1: Speech Recognition and Keyword Spotting
Room: 105
Type: Poster
December 9, 3:15 PM to 5:15 PM

Sat-A-P-1-1: Exploring the Use of Large Language Models for Text-based Domain Adaptation in Automatic Speech Recognition
Junzhe Liu (Shanghai Jiaotong University)*; Xie Chen (Shanghai Jiaotong University); Jianwei Yu (Tencent AI lab)

Sat-A-P-1-2: Task-adaptive Generative Adversarial Network based Speech Dereverberation for Robust Speech Recognition
Liu Ji (TianJin University)*

Sat-A-P-1-3: Pseudo Label Refining for Semi-supervised Automatic Speech Recognition
Jin Zezhong (Polyu)*; Ruochen Zhang (Polyu)

Sat-A-P-1-4: Semi-supervised Learning for Multilingual ASR Using Articulatory Features
Xinmei Su (Beijing Institute of Technology)*; Xiang Xie (Beijing Institute of Technology); Chenguang Hu (Beijing Institute of Technology); Jing Wang (Beijing Institute of Technology)

Sat-A-P-1-5: Knowledge-driven Text Generation for Zero-shot ASR Domain Adaptation
Xizhuo Zhang (Shanghai Jiao Tong University)*; Baochen Yang (Shanghai Jiao Tong University); Sen Liu (Shanghai Jiao Tong University); Yiwei Guo (Shanghai Jiao Tong University); Zheng Liang (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sat-A-P-1-6: Self-Supervised Regularization for End-to-End Automatic Speech Recognition
Xianghu Yue (National University of Singapore)*; Jingru Lin (National University of Singapore); Xiaoxue Gao (National University of Singapore); Junyi Ao (The Chinese University of Hong Kong (Shenzhen)); Duo Ma (The Chinese University of Hong Kong，Shenzhen); Meng Ge (Tianjin University); Haizhou Li (The Chinese University of Hong Kong, Shenzhen)

Sat-A-P-1-7: Improving VAD Performance in Long-form Noisy Condition Through ASR Integration
Bingqing Zhu (AISpeech)*; Shaofei Xue (AISpeech)

Sat-A-P-1-8: 利用方言信息改进藏语多方言语音识别
Qinya Zhang(Northwest Minzu University)*

Sat-A-P-1-9: Leveraging Synthetic Speech for CIF-based Customized Keyword Spotting
Shuiyun Liu (Northwestern Polytechnical University)*; Ao Zhang (Northwestern Polytechnical University); Lei Xie (NWPU)

Sat-A-P-1-10: TA-KWS：基于测试时间自适应的开放语音指令词识别技术
Huang Jiayi (Shanghai Jiao Tong University)*; Meng Shan (Shanghai Jiao Tong University)

Sat-A-P-1-11: Adopting Multilingual ASR Foundation Models for Low-Resource Languages
Han Zhu (University of Chinese Academy of Sciences)*; Jie Zheng (Institute of Acoustics,Chinese Academy of Sciences); Gaofeng Cheng (Institute of Acoustics, Chinese Academy of Sciences); Liuping Luo (Guangdong Provincial Public Security Department); Qingwei Zhao (Institute of Acoustics, Chinese Academy of Sciences); Pengyuan Zhang (Institute of Acoustics, Chinese Academy of Sciences)

Sat-A-P-2: Speech Synthesis, Sound Synthesis and Voice Conversion
Room: 105
Type: Poster
December 9, 3:15 PM to 5:15 PM

Sat-A-P-2-1: 面向生成语音溯源的分布外检测
Xinrui Yan (中国科学院自动化研究所)*; Jiangyan Yi (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences); Hao Gu (Institute of Automation，Chinese Academy of Sciences); Jianhua Tao (Department of Automation, Tsinghua University); Ruibo Fu (National Laboratory of Pattern Recognition, Institute of Automation of the Chinese Academy of Sciences)

Sat-A-P-2-2: An Audio anti-spoofing model Using Whisper features and Improved spectro-temporal graph attention networks
王思远 (新疆大学); Zhihua Huang (Xinjiang university)*

Sat-A-P-2-3: 面向任意源话者的端到端声音转换方法
Yannian Chen (University of Science and Technology of China); Zhen-Hua Ling (University of Science and Technology of China)*

Sat-A-P-2-4: Enhancing Robustness of Cross-Domain Voice Conversion with Strong Conditioned Diffusion Models
Long Xiao (Tsinghua University)*; Xiang Li (Tsinghua University); Jing Yang (Huawei Inc); Fan Fan (Huawei Inc); Zhiyong Wu (Tsinghua University)

Sat-A-P-2-5: 基于特征解耦的语音转换系统设计与实现
Wei Chen (Shenzhen International Graduate School Tsinghua University)*; Long Xiao (Tsinghua University); Xintao Zhao (Tsinghua University); Zhiyong Wu (Tsinghua University)

Sat-A-P-2-6: 互信息和孪生结构增强的语音转换模型
Pengcheng Li (University of Science and Technology of China); Xulong Zhang (Ping An Technology (Shenzhen) Co., Ltd.); Jianzong Wang (Ping An Technology (Shenzhen) Co., Ltd)*; Ning Cheng (Ping An Technology (Shenzhen) Co., Ltd); Jing Xiao (Ping An Insurance (Group) Company Of China)

Sat-A-P-2-7: 基于互信息和文本引导的自动语音表征解耦
Ziqi Liang (University of Science and Technology of China); Xulong Zhang (Ping An Technology (Shenzhen) Co., Ltd.); Jianzong Wang (Ping An Technology (Shenzhen) Co., Ltd)*; Ning Cheng (Ping An Technology (Shenzhen) Co., Ltd); Jing Xiao (Ping An Insurance (Group) Company Of China)

Sat-A-P-2-8: Real-time Automotive Engine Sound Simulation with Deep Neural Network
Hao Li (Z-one Technology co. Ltd, Shanghai, China); Weiqing Wang (Data Science Research Center, Duke Kunshan University, China); Ming Li (Duke Kunshan University)*

Sat-A-P-2-9: 基于深度学习的多轨音乐自动生成与评测方法研究
Cheng Liu (Nankai University)*; Yong Qin (Nankai University)

Sat-A-P-2-10: Contextual Spectrum and Prosody Integration for Text-Based Speech Editing
Zheng Liang (Shanghai Jiao Tong University)*; Chenpeng Du (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University); Xie Chen (Shanghai Jiaotong University)

Sat-A-P-2-11: Explore Speaker Adaptation Text-to-Speech Method Based ProbSparse Self-attention
Yuanyuan Leng (Xinjiang University)*; Liang He (Tsinghua University)

Sat-A-P-2-12: EEG-Based Speech Envelope Decoding: Structured State Space and U-Net Model Integration
Yueqian Lin (Duke Kunshan University); Ming Li (Duke Kunshan University)*

Sat-A-P-2-13: Cross-Speaker Emotion Transfer Through Disentangling Speaker and Emotion Attributes in Emotional Speech Synthesis
Mingming Yin (Xinjiang University); Zhihua Huang (Xinjiang university)*; Kexin Lu (Xinjiang University)

Sat-A-P-2-14: 文本预训练在 TTS 声学模型中的应用
Shuoyi Zhou (Tsinghua University)*; Yixuan Zhou (Tsinghua University); Shun Lei (Tsinghua University); Changhe Song (Tsinghua University); Zhiyong Wu (Tsinghua University); Jia Jia (Tsinghua University)

Sat-A-P-2-15: 完全端到端的藏语语音合成方法研究
Lhakpa Dondrub (Tibet University)*

Sat-E-IND: Industry Forum 2
Room: 107
Type: Industry Forum
December 9, 6:30 PM to 7:45 PM
Chair: 刘瑞老师

18:30-19:00
Sat-E-IND-1: 大规模语音应用中的落地思考
Presenter: 熊世富 (科大讯飞)

19:00-19:15
Sat-E-IND-2: 大模型时代的语音技术发展
Presenter: 郝玉峰 (海天瑞声)

19:15-19:30
Sat-E-IND-3: 基于数据大模型的数据标注平台
Presenter: 周彤 (标贝科技)

Sat-E-SS: 多模态识别竞赛
Room: 206
Type: Special Session
December 9, 6:30 PM to 7:45 PM
Organizer: 李科老师
Chair: 王东老师

Sunday, December 10

Sun-M-PLEN-1: 多体微观智能：快速算法与软件
Room: 206
Type: Plenary
December 10, 9:00 AM to 9:45 AM
Presenter: 金石老师
Chair: 凌震华老师

Sun-M-PLEN-2: Seeing to Hear Better
Room: 206
Type: Plenary
December 10, 9:45 AM to 10:30 AM
Presenter: 李海洲老师
Chair: 凌震华老师

Sun-M-O-1: Miscellaneous: Multimodal System, Language Model, and Other
Room: 108
Type: Oral
December 10, 11:00 AM to 12:30 PM
Chair: 吴志勇老师

Sun-M-O-1-1: DSNet: Disentangled Siamese Network with Neutral Calibration for Speech Emotion Recognition
Chengxin Chen (The Institute of Acoustics of the Chinese Academy of Sciences)*; Pengyuan Zhang (Institute of Acoustics, Chinese Academy of Sciences)

Sun-M-O-1-2: Emotional Support Dialog System Through Recursive Interactions Among Large Language Models
Keqi Chen (Beijing University of Posts and Telecommunications)*; Huijun Lian (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications); Ya Li (Beijing University of Posts and Telecommunications)

Sun-M-O-1-3: A Study on Domain Adaptation for Audio-visual Speech Enhancement
Chenxi Wang (University of Science and Technology of China)*; Hang Chen (USTC); Jun Du (University of Science and Technology of China); Chenyue Zhang (University of Science and Technology of China); Yuling Ren (china mobile online services company limited); Qinglong Li (china mobile online services company limited); Ruibo Liu (china mobile online services company limited); Chin-Hui Lee (Georgia Institute of Technology)

Sun-M-O-1-4: Adaptive Alignment for Audio-Visual Speech Separation
Xinyuan Wu (Shanghai Jiao Tong University); Yifei Wu (Shanghai Jiao Tong University); Chenda Li (Shanghai Jiao Tong University); Jiahong Li (Shanghai Jiao Tong University)*; Yanmin Qian (Shanghai Jiao Tong University)

Sun-M-O-1-5: Combing acoustic and behavioristic analysis of Dysarthria characteristics in subacute stroke patients
Yanbo Wang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)*

Sun-M-O-1-6: A Comparative Study of Pre-trained Audio and Speech Models for Heart Sound Detection

Yuxin Duan (Shanghai Jiao Tong University); Chenyu Yang (Shanghai Jiao Tong University); Zihan Zhao (Shanghai Jiao Tong University); Yiyang Jiang (Shanghai Artificial Intelligence Laboratory); Yan-Feng Wang (Cooperative medianet innovation center of Shanghai Jiao Tong University); Yu Wang (Shanghai Jiao Tong University)*

Sun-M-O-1-7: 面向知识图谱补全的自适应多教师蒸馏策略

Yujie Chen (Anhui Univesity)*; Jiangyan Yi (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences); Jianhua Tao (Tsinghua University); Zhao Lv (anhui university); Cunhang Fan (Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, School of Computer Science and Technology, Anhui University)

Sun-M-O-2: Human-Machine Interaction and Sound Event Detection
Room: 104
Type: Oral
December 10, 11:00 AM to 12:30 PM
Chair: 吴梦玥老师

Sun-M-O-2-1: Unraveling Predictive Mechanism in Speech Perception and Production: Insights from EEG analyses of Brain Network Dynamics

Bin Zhao (Institute of Linguistics, Chinese Academy of Social Sciences)*; Gaoyan Zhang (Tianjin University); Jianwu Dang (School of Computer Science and Technology, Tianjin University, Tianjin, China; School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan); Longbiao Wang (Tianjin University); Zhongjie Li (Tianjin University); Aijun Li ("Institute of Linguistics, CASS")

Sun-M-O-2-2: A Packet Loss Concealment Method Based on the Demucs Network Structure

Wenwen Li (Beijing University of Technology); Changchun Bao (Beijing University of Technology)*

Sun-M-O-2-3: Data Augmentation By Finite Element Analysis for Enhanced Machine Anomalous Sound Detection

Zhixian Zhang (Duke Kunshan University)*; Yucong Zhang (Duke Kunshan University); Ming Li (Duke Kunshan University)

Sun-M-O-2-4: DP-MAE: A Dual-Path Masked AutoEncoder Based Self-Supervised Learning Method for Anomalous Sound Detection

Wei-Lin Xie (University of Science and Technology of China); Haoyu Song (The Australian National University); Xiaomin Zeng (University of Science and Technology of China)*; Lirong Dai (University of Science and Technology of China); Yan Song (USTC)

Sun-M-O-2-5: 基于动态主代理损失的深度度量学习方法

Ruihao Jing (Northwestern Polytechnical University)*; Lei Zhao (Northwestern Polytechnical University); Zhang Xiaolei (Northwestern Polytechnical University)

Sun-M-O-2-6: "普通话七音"在线听力测试系统研发与应用效果测试

Yu Chen (Tianjin University of Technology)*; Yu Bai (天津理工大学); Ju Zhang (Tianjin University)

Sun-M-SF: Student Forum 1
Room: 206
Type: Student Forum
December 9, 11:00 AM to 12:30 PM
Chair: 王龙标老师

11:00-11:08
Sun-M-SF-1: Learning From Yourself: A Self-Distillation Method for Fake Speech Detection
Presenter: 薛军 (安徽大学)

11:08-11:16
Sun-M-SF-2: 基于并行分支循环注意力的时频域语音分离方法
Presenter: 杨雪 (北京工业大学)

11:16-11:24
Sun-M-SF-3: 一种低延时鲁棒 MVDR 波束形成器的设计方法
Presenter: 周静 (北京工业大学)

11:24-11:32
Sun-M-SF-4: 基于预训练网络的半监督声音事件检测
Presenter: 徐亮 (北京理工大学)

11:32-11:40
Sun-M-SF-5: 基于汉语中介语料库的二语韵律边界偏误检测
Presenter: 李瑞珊 (北京语言大学)

11:40-11:48
Sun-M-SF-6: 海口普通话"子"缀的轻声分布差异
Presenter: 钱薪屹 (北京语言大学)

11:48-11:56
Sun-M-SF-7: Anomalous Sound Detection Using Self-Attention-Based Frequency Pattern Analysis of Machine Sounds
Presenter: 张合静 (哈尔滨工程大学)

11:56-12:04
Sun-M-SF-8: 面向声音事件检测的特征及决策解耦初探
Presenter: 高利剑 (江苏大学)

12:04-12:12
Sun-M-SF-9: 跨年龄说话人确认：学习年龄不变的说话人表征
Presenter: 覃晓逸 (昆山杜克大学)

12:12-12:20
Sun-M-SF-10: 基于神经同态合成的语音增强研究
Presenter: 何树林 (内蒙古大学)

Sun-M-P-1: Phonetics, Phonology and Prosody
Room: 105
Type: Poster
December 10, 11:00 AM to 12:30 PM

Sun-M-P-1-1: 基于依存句法的越南母语者汉语语音韵律结构考察
Yuqi Shi (Beijing Language and Culture University)*; Yukai Wan (Beijing Language & Culture University SAIT Lab); Hao Yang (Beijing Language and Culture University); Jinsong Zhang (Beijing Language and Culture University)

Sun-M-P-1-2: 基于大语言模型的语音合成前端训推一体化增强方法研究
刘旺 (中国科学院大学)*; Ruibo Fu (National Laboratory of Pattern Recognition, Institute of Automation of the Chinese Academy of Sciences); Wen Zhengqi (CASIA)

Sun-M-P-1-3: 汉语歧义句阅读加工的行为实验研究：语音意识与内隐韵律
Kexin Zhang (Tongji University)*

Sun-M-P-1-4: The Impact of Mandarin Tonal Coarticulation Mismatch on Auditory Perception
Yujia Liu (Beijing Language and Culture University)*; Jinsong Zhang (Beijing Language and Culture University); Xin Chen (Fudan University); Yukai Wan (Beijing Language & Culture University SAIT Lab); Chong Cao (Beijing Language and Culture University); Xiaoyue Zhang (Zhejiang Media Group)

Sun-M-P-1-5: 基于四种句型对比的汉语普通话"友好"态度韵律特征的研究
Jiawei Song (School of Foreign Languages, UESTC)*; Yutong Wang (School of Foreign Languages, UESTC); Yuzhen Zhang (School of Foreign Languages, UESTC)

Sun-M-P-1-6: The Effect of Language Experience on Cues for Tone Perception
Xiaoying Xu (Beijing Normal University)*; Yi Zhang (Beijing Normal University); Lishan Li (Beijing Normal University)

Sun-M-P-1-7: 基于虚词特征的藏语韵律单元选取方法
Lhakpa Dondrub (Tibet University)*

Sun-M-P-1-8: Chinese EFL Learners' Auditory and Visual Perception of English Statement and Question Intonation: The Effect of Stress
Qiunan Xu (南京理工大学)*

Sun-M-P-1-9: 俄语母语者普通话舌尖元音习得研究
姬凌波 (北京语言大学)*; Jinsong Zhang (Beijing Language and Culture University); Liangjie Huang (Beijing Language and Culture University); 姚文翰 (北京语言大学); 李紫叶 (北京语言大学)

Sun-M-P-1-10: The Production of Successive Addition Boundary Tone in Mandarin Preschoolers
Aijun Li ("Institute of Linguistics, CASS")*; Jun Gao (Institute of Linguistics, Chinese Academy of Social Sciences); Zhiwei Wang (BISU)

Sun-M-P-1-11: Music Experience Enhances Categorical Perception of English Vowels in Mandarin Speakers
Jiayu Liang (Shandong University)*; Hao Zhang (Shandong University)

Sun-M-P-1-12: 音高和发声类型在吴语常熟方言声调感知中的相互作用
姚程栋 (南京师范大学)*

Sun-M-P-1-13: 话剧演员和学员舞台音的声学特征分析
钱昭璇 (北京师范大学文学院)*; 郭人榕 (北京师范大学文学院); 汪高武 (北京师范大学文学院)

Sun-M-P-1-14: Acoustic Study on the Friction of the i Vowel in Qingpu Wu Chinese
Yan Zhao (East China Normal University)*; Junru Wu (East China Normal University)

Sun-M-P-1-15: The Effect of Fundamental Frequency Contour on The Intelligibility of Uyghur Sentences
Jingyi Song (Xinjiang University); Zhihua Huang (Xinjiang university)*; Ruilin Wu (Xinjiang University); Xiaoming Liang (Xinjiang University)

Sun-M-P-1-16: English Consonant Pronunciation Errors Research among Chinese K12 Learners of Different Ages
Yilin Tao (University of Chinese Academy of Sciences)*

Sun-M-P-1-17: 从感知角度看嘉兴方言阳上调和阳去调的合并
Qing Yao (Tongji University)*

Sun-M-P-2: Emotion Recognition, Spoken Language Processing and Digital Human Modeling
Room: 105
Type: Poster
December 10, 11:00 AM to 12:30 PM

Sun-M-P-2-1: Semi-End-to-End Nested Named Entity Recognition from Speech
Min Zhang (Huawei)*; Xiaosong Qiao (Huawei); Yanqing Zhao (Huawei); Chang Su (Huawei); Yuang Li (Huawei); Yinglu Li (HUAWEI TECHNOLOGIES CO., LTD.); Mengyao Piao (Huawei); Song Peng (Huawei); Shimin Tao (Huawei); Hao Yang (Huawei)

Sun-M-P-2-2: 动量对比学习实现跨模态嵌入对齐提升端到端口语理解
Zheng Bei Da (Xinjiang University)*

Sun-M-P-2-3: Beyond Classification: A Label-aware Prompt-based Framework for Joint Multiple Intent Detection and Slot Filling
Zhihong Zhu (Peking University)*; Yuexian Zou (Peking University)

Sun-M-P-2-4: A Fast Sampling Method in Diffusion-based Dance Generation Models
Puyuan Guo (Beijing University of Posts and Telecommunications)*; Yichen Han (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications); Ya Li (Beijing University of Posts and Telecommunications)

Sun-M-P-2-5: 一种基于扩散模型的情绪可定制的肖像动化方法
Bingyuan Zhang (Ping An Technology (Shenzhen) Co., Ltd., Shenzhen, China); Xulong Zhang (Ping An Technology (Shenzhen) Co., Ltd.); Jianzong Wang (Ping An Technology (Shenzhen) Co., Ltd)*; Ning Cheng (Ping An Technology (Shenzhen) Co., Ltd); Jing Xiao (Ping An Insurance (Group) Company Of China)

Sun-M-P-2-6: 基于单帧图像的语音驱动情感化数字人生成
Tianze Zhou (Tsinghua University)*; Liyang Chen (Tsinghua University); Zhiyong Wu (Tsinghua University); Jia Jia (Tsinghua University)

Sun-M-P-2-7: 基于情绪氛围软标签的对话情绪识别
Xiaolong Wu (Xinjiang University)*; Yiyang Zhao (Tsinghua University); Mingxing Xu (Tsinghua University); Thomas Fang Zheng ("CSLT, Tsinghua University"); Askar Hamdulla (Xinjiang University)

Sun-M-P-2-8: 基于文本模态引导注意力融合的多模态情感识别方法
苗阳 (东北林业大学)*

Sun-M-P-2-9: A Transformer Based Speech-Text Modality Experts for Multimodal Emotion Recognition
Weixiang Xu (Hunan University)*; Zixing Zhang (Hunan University)

Sun-A-O-1: Speech Synthesis, Sound Synthesis and Voice Conversion
Room: 108
Type: Oral
December 10, 1:30 PM to 3:30 PM
Chair: 凌震华老师

Sun-A-O-1-1: Zero-shot Singing Voice Conversion Based on Timbre Space Modeling and Excitation Signal Control
Yuan Jiang (University of Science and Technology of China); Yannian Chen (University of Science and Technology of China); Lijuan Liu (iflytek); Yajun Hu (iflytek); Xin Fang (iFlytek Research); Zhen-Hua Ling (University of Science and Technology of China)*

Sun-A-O-1-2: One-shot Voice Conversion based on Speaker Self-supervised Speech Representation and Dynamic Convolution
Na Guo (Tianjin university)*; Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China)

Sun-A-O-1-3: Leveraging Content-based Features from Multiple Acoustic Models for Singing Voice Conversion
Xueyao Zhang (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); Zihao Fang (The Chinese University of Hong Kong, Shenzhen); Lexiao Zou (Harbin Institute of Technology (Shenzhen)); Liumeng Xue (The Chinese University of Hong Kong, Shenzhen); Zhizheng Wu (Chinese University of Hong Kong, Shenzhen)

Sun-A-O-1-4: CLAUDIO: Clustering, Augmentation, and Discriminator for Foley Sound Synthesis
Zeyu Xie (Shanghai Jiao Tong University)*; Xuenan Xu (Shanghai Jiao Tong University); Baihan Li (Shanghai Jiao Tong University); Mengyue Wu (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sun-A-O-1-5: Research on Prosody Prediction of Tibetan Language Based on Speech Synthesis Task
Zom Yang (Qinghai Minzu University)*; Peng Luo (Qinghai Normal University,); Yi Zhu (Tianjin University); Kuntharrgyal Khysru (Qinghai Minzu University); Daijicao Long (Qinghai Minzu University); Mengfei Hu (Tianjin University); Wenhuan Lu (Tianjin University); Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China)

Sun-A-O-1-6: Accent-VITS: accent transfer for end-to-end TTS
Linhan Ma (Northwestern Polytechnical University)*; Yongmao Zhang (Audio, Speech and Language Processing Group (ASLP@NPU), School Of Computer Science, Placeholder, Xi'an, China); Xinfa Zhu (Northwestern Polytechnical University); Yi Lei (Northwestern Polytechnical University); Ziqian Ning (northwestern Polytechnical University); Pengcheng Zhu (Northwestern Polytechnical University); Lei Xie (Fuxi AI Lab, NetEase Inc.)

Sun-A-O-1-7: 基于深度学习的蒙古语合成语音检测
Zhang Jinhua (Inner Mongolia University); Kailin Liang (Inner Mongolia University); Rui Liu (Inner Mongolia University)*

Sun-A-O-1-8: 基于强度重映射方法的情感迁移语音合成
Haoxiang Shi (USTC); Xulong Zhang (Ping An Technology (Shenzhen) Co., Ltd.); Jianzong Wang (Ping An Technology (Shenzhen) Co., Ltd)*; Ning Cheng (Ping An Technology (Shenzhen) Co., Ltd); Jing Xiao (Ping An Insurance (Group) Company Of China)

Sun-A-O-2: Phonetics, Phonology and Prosody
Room: 104
Type: Oral
December 10, 1:30 PM to 3:30 PM
Chair: 谢焱陆老师

Sun-A-O-2-1: Exploring the Generation of Pronunciation Lexicon for Low-Resource Language ASR Based on Generic Phone Recognizer

Jinpeng Li (Tsinghua University)*; Xie Chen (Shanghai Jiaotong University); Wei-Qiang Zhang (Tsinghua University)

Sun-A-O-2-2: An Experimental Study on Correction of Intonation Errors in Chinese Interlanguage

Rui Wang (Beijing Language and Culture University)*; Wen Cao (Beijing Language and Culture University)

Sun-A-O-2-3: Prosodic Cues in Polite Chengdu Dialect Speech Production and the Polite Speech Emotion Recognition of Chengdu Dialect

Yuxin Wu (Nanjing University)*; Jiayong He (University of Electronic Science and Technology of China)

Sun-A-O-2-4: Acoustic Features at Intonational Phrase Boundaries: Comparative Study of Native Speakers and L2 Learners of Chinese Mandarin

Ruishan Li (Beijing Language and Culture University)*; Lixia Hao (Dongguan University of Technology); Huang He Ting (Beijing language and culture university); Jinsong Zhang (Beijing Language and Culture University); Yanlu Xie (Beijing Language and Culture University)

Sun-A-O-2-5: 人工智能语音吸引力的韵律影响因素探究 A Pilot Study on the Prosodic Factors Influencing Voice Attractiveness of AI Speech

王毅慧 (北京师范大学)*; 卢浩程 (北京师范大学); Gaowu Wang (Beijing Normal University)

Sun-A-O-2-6: 普通话学龄前儿童塞音和塞擦音时长特征发展

Jun Gao (Institute of Linguistics, Chinese Academy of Social Sciences)*

Sun-A-O-2-7: 语言冗余和韵律结构对普通话音节时长的影响

Xiaowang Liu (Beijing Language and Culture University)*; Jinsong Zhang (Beijing Language and Culture University); Yun Hao (University of Groningen)

Sun-A-O-2-8: 北京话"吧"字句语气识别的性别差异

Qixin Li (BNU)*; Gaowu Wang (Beijing Normal University)

Sun-A-SF: Student Forum 2
Room: 206
Type: Student Forum
December 10, 1:30 PM to 3:30 PM
Chair: 薛少飞老师

13:30-13:38
Sun-A-SF-1: 面向声纹识别的隐蔽后门攻击
Presenter: 叶哲 (宁波大学)

13:38-13:46
Sun-A-SF-2: 迈向通用语音离散令牌：ASR 和 TTS 案例研究
Presenter: 杨亦凡 (上海交通大学)

13:46-13:54
Sun-A-SF-3: 基于簇感知 DINO 的自监督说话人验证
Presenter: 韩冰 (上海交通大学)

13:54-14:02
Sun-A-SF-4: Towards More Realistic，Powerful，and Efficient Speech-based Self-Supervised Learning
Presenter: 马子阳 (上海交通大学)

14:02-14:10
Sun-A-SF-5: 基于噪声解耦和度量学习的鲁棒说话人识别研究
Presenter: 孙瑶 (天津大学)

14:10-14:18
Sun-A-SF-6: 利用位置相关的局部-全局依赖进行合成语音检测
Presenter: 刘晓晖 (天津大学)

14:18-14:26
Sun-A-SF-7: 用于多通道语音分离的声源位置感知全神经波束模型
Presenter: 付燕杰 (天津大学)

14:26-14:34
Sun-A-SF-8: 用于对话行为识别的基于关系匹配的局部和全局上下文建模研究
Presenter: 司宇珂 (天津大学)

14:34-14:42
Sun-A-SF-9: 基于预训练语音表征学习的少监督语音合成
Presenter: 强春雨 (天津大学)

14:42-14:50
Sun-A-SF-10: 基于多任务学习的异常声音检测建模和多尺度评分
Presenter: 章羽聪 (武汉大学)

14:50-14:58
Sun-A-SF-11: Cross-Age Speaker Verification: Learning Age-Invariant Speaker Embeddings
Presenter: 覃晓逸 (武汉大学/昆山杜克大学)

14:58-15:06
Sun-A-SF-12: 基于 ad-hoc 麦克风阵列的多通道语音去混响融合模块
Presenter: 郭红梅 (西北工业大学)

15:06-15:14
Sun-A-SF-13: Maximal Speaker Separability via Robust Speaker Representation Learning
Presenter: 李哲 (香港理工大学)

15:14-15:22
Sun-A-SF-14: Complex Patterns of Tonal Realization in Taifeng Chinese
Presenter: 张小砚 (中国社会科学院语言研究所)

15:22-15:30
Sun-A-SF-15: 基于语音的端到端阻塞性睡眠呼吸暂停检测
Presenter: 张凯博 (清华大学)

Sun-A-P-1-1: Review of Explainability Studies in Voiceprint Recognition
Qi Su (Xiamen University)*; Shiyu Miao (Xiamen University); Yishuang Li (Xiamen University); Lin Li (Xiamen University); Qingyang Hong (Xiamen University)

Sun-A-P-1-2: Improving ECAPA-TDNN Performance with Coordinate Attention
Shuanghong Liu (Xinjiang University)*; Zhida Song (Xinjiang University); Liang He (Tsinghua University)

Sun-A-P-1-3: 一种用于说话人确认的局部和全局特征并行提取分支结构——Branch-ECAPA-TDNN
付先燕 (西北工业大学)*; Xiaolei Zhang (Northwestern Polytechnical University); Jiadi Yao(Northwestern Polytechnical University)

Sun-A-P-1-4: Gradient Frequency Attention Distillation for Speaker Verification
Wenhao W Yang (Tianjin Univerisity)*; Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China); Wenhuan Lu (Tianjin University)

Sun-A-P-1-5: Word-level Speaker Change Detection for CTC-based ASR
Naijun Zheng (Huawei Technologies Co., Ltd.)*; Ziqing Du (Huawei Technologies Co., Ltd.); Xucheng Wan (Huawei Technology); Kai Liu (Huawei Technologies Co., Ltd.); Huan Zhou (AARC, Huawei Technologies Co., Ltd.)

Sun-A-P-1-6: 基于自适应类边界间隔的说话人识别方法
赵昊洋 (东北林业大学)*

Sun-A-P-1-7: UNSUPERVISED DOMAIN ADAPTATION WITH PSEUDO LABELING FOR ROBUST SPEAKER VERIFICATION
Bing Han (Shanghai Jiao Tong University)*; Zhengyang Chen (Shanghai Jiao Tong University); Yanmin Qian (Shanghai Jiao Tong University)

Sun-A-P-1-8: Voice Cloning-Based Data Augmentation for Speaker Recognition
Zezhong Jin (Polyu)*

Sun-A-P-1-9: 基于纹理特征的上下文感知掩蔽说话人验证模型
Chengli Sun (Guangzhou Maritime University)*

Sun-A-P-1-10: Magnitude and Phase-based Feature Fusion Using Co-attention Mechanism for Speaker recognition
Rongfeng Su (Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)*; Mengjie Du (China Telecom Corporation Ltd. Data & AI Technology Company); Xiaokang Liu (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Lan Wang (Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences); Nan Yan (Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)

Sun-A-P-1-11: Measurement of Rank-based Pruning on Lightweight Speaker Model
Baowei Zhao(Xinjiang University)*; Liang He (Tsinghua University)

Sun-A-P-1-12: An Improved System for Partially Fake Audio Detection Using Pre-trained Model
Jianqian Zhang (Beijing Institute of Technology)*; Hanyue Liu (Beijing Institute of Technology); Mengyuan Deng (Beijing Institute of Technology); Jing Wang (Beijing Institute of Technology); Liang Xu (Beijing Institute of Technology); Jiahao Li (Beijing Institute of Technology)

Sun-A-P-1-13: Replay Attack Detection Based on Raw Waveform and Sinc Filter
Jian Zhou (Anhui University)*

Sun-A-P-1-14: 基于发音基元字典的语音鉴伪方法
Feng Chen (Harbin Institute of Technology)*; Jiqing Han (Harbin Institute of Technology); 何勇军 (哈尔滨工业大学); 郑铁然 (哈尔滨工业大学); 郑贵滨 (哈尔滨工业大学)

Sun-A-P-2: Speech Analysis and Sound Event Detection
Room: 105
Type: Poster
December 10, 1:30 PM to 3:30 PM

Sun-A-P-2-1: "一级甲等"普通话基频特征值统计分析
Lu Xiyao (Beijing Language and Culture University)*; Yukai Wan (Beijing Language & Culture University SAIT Lab); Ruishan Li (Beijing Language and Culture University); Jinsong Zhang (Beijing Language and Culture University)

Sun-A-P-2-2: 海口普通话"子"缀的轻声分布差异
Xinyi Qian (BLCU)*; Wen Cao (BLCU)

Sun-A-P-2-3: 感知训练在泰语母语者汉语声调教学中的效果研究
Jingyan Shi (Xinjiang University)*; Zhu Sun (Xinjiang University); Wei Wang (Xinjiang University)

Sun-A-P-2-4: 声调对汉语言语可懂度的影响探究
孙亦宸 (新疆大学)*; 王玮 (新疆大学); Jingyan Shi (Xinjiang University)

Sun-A-P-2-5: 表示方式的"怎么 +VP"歧义结构韵律研究
Qian Shiqi (Beijing Normal University)*; Zhou Shihong (Beijing Normal University)

Sun-A-P-2-6: Sound Event Detection with Soft Labels by Aggregating Pre-trained Embeddings from Different Layers
Xuenan Xu (Shanghai Jiao Tong University); Ziyang Ma (Shanghai Jiao Tong University); Fei Yang (Shanghai Jiao Tong University)*; Guanrou Yang (Shanghai Jiao Tong University); Mengyue Wu (Shanghai Jiao Tong University); Xie Chen (Shanghai Jiaotong University)

Sun-A-P-2-7: A Comparative Study of Deep Learning Methods for Infant Cry Detection
Qingyu Liu (Shanghai Normal University)*; Li Li (Shanghai Normal University); Yifan Zhou (Shanghai Normal University); Dongxing Xu (Unisound AI Technology Co., Ltd.); Yanhua Long (Shanghai Normal University)

Sun-A-P-2-8: 基于改进 EfficientNet 的轻量化鸟鸣识别方法
Haolun He (Northeast Forestry University)*

Sun-A-P-3: Multimodal System and Machine Translation
Room: 105
Type: Poster
December 10, 1:30 PM to 3:30 PM

Sun-A-P-3-1: Label-Free Multi-Domain Machine Translation with Stage-wise Training
Fan Zhang (communication University of China)*

Sun-A-P-3-2: Cross-modal Domain generalization for Ultrasound Tongue Images-based Silent Speech Interfaces
Minghao Guo (Tianjin University)*; Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China); Ruiteng Zhang (Tianjin University); Yuting Zhao (Tianjin University); Qiang Fang (Chinese Academy of Social Sciences)

Sun-A-P-3-3: Joint Training or Not: An Exploration of Pre-trained Speech Models in Audio-Visual Speaker Diarization
Huan Zhao (Northwestern Polytechnical University)*; Li Zhang (Northwestern Polytechnical University); Yue Li (Northwestern Polytechnical University); Yannan Wang (Tencent); Hongji Wang (Tencent); Wei Rao (Tencent); Qing Wang (Northwestern Polytechnical University); Lei Xie (NWPU)

Sun-A-P-3-4: Learning from Parent Classes: a Multimodal Weighted Fusion Transformer with Parent Class Prediction
Yuxuan Wang (Shanghai Jiao Tong University); Mengyue Wu (Shanghai Jiao Tong University)*; Kai Yu (Shanghai Jiao Tong University)

Sun-A-P-3-5: 基于多任务学习和特征解耦技术的超声舌视频语音重建
Yuting Zhao (Tianjin University)*; Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China); Wenhuan Lu (Tianjin University); Yu Xiao (Tianjin University)

Sun-A-YF: Young Forum 2
Room: 206
Type: Young Forum
December 10, 3:45 PM to 5:15 PM
Chair: 杜俊老师

Sun-A-YF-1: 深度生成语音鉴别和取证溯源
Presenter: 易江燕老师，中国科学院自动化所多模态人工智能系统全国重点实验室副研究员

Sun-A-YF-2: 真实世界语音前端：从数据仿真到模型设计
Presenter: 罗艺老师，腾讯 AI Lab 高级研究员

Sun-A-YF-3: AI for Sound Understanding: Classification, Detection, and Separation
Presenter: 孔秋强老师，香港中文大学电气工程系助理教授

Organizer List

名誉主席 (Honorary Chair)
- 党建武 - 天津大学
- 郑方 - 清华大学

大会主席 (Conference Chairs)
- 鲍长春 - 北京工业大学
- 陶建华 - 清华大学
- 俞凯 - 上海交通大学

程序委员会主席 (Program Committee Chairs)
- 贾珈 - 清华大学
- 凌震华 - 中国科学技术大学
- 陈谐 - 上海交通大学

组织委员会主席 (Organizing Committee Chairs)
- 李明 - 昆山杜克大学
- 吴志勇 - 清华大学深圳研究生院
- 龚声蓉 - 常熟理工计算机学院
- 樊帅 - 思必驰科技股份有限公司

工业联络主席 (Industrial Liaison Chairs)
- 谢磊 - 西北工业大学
- 张超 - 清华大学
- 刘瑞 - 内蒙古大学

宣传主席 (Publicity Chairs)
- 王钰 - 上海交通大学
- 邹月娴 - 北京大学深圳研究生院
- 洪青阳 - 厦门大学

出版主席 (Publication Chairs)
- 李雅 - 北京邮电大学
- 张子兴 - 湖南大学

教程报告主席 (Tutorial Chairs)
- 钱彦旻 - 上海交通大学
- 武执正 - 香港中文大学（深圳）

特殊议题主席 (Special Session Chairs)
- 王东 - 清华大学
- 朱璇 - 三星电子中国研究院

青年论坛主席 (Young Forum Chairs)
- 张晓雷 - 西北工业大学
- 杜俊 - 中国科技大学
- 李圣辰 - 西交利物浦大学

学生论坛主席 (Student Forum Chairs)
- 王龙标 - 天津大学
- 易江燕 - 中国科学院自动化研究所
- 薛少飞 - 思必驰科技股份有限公司

财务主席 (Financial Chairs)
- 吴梦玥 - 上海交通大学