



第二十一届 全国人机语音通讯学术会议

2026年11月5-8日 中国·横琴

Thursday, November 5

[14:00-17:00] Thu-R - 会议报到 1

Friday, November 6

[08:00-17:00] Fri-R - 会议报到 1

[09:00-17:00] Fri-EXH - 赞助商展览 1

[09:00-10:30] Fri-M-T-1 - 语音交互 + 情感大模型：从判别式情感识别迈向生成式情感理解 1

[09:00-10:30] Fri-M-T-2 - 鲁棒对话智能：复杂场景下人机语音交互感知与表达 1

[11:00-12:30] Fri-M-T-3 - 面向深度伪造语音的话者隐私保护：语音匿名化与水印 1

[11:00-12:30] Fri-M-T-4 - 空间多模态听觉感知：面向真实复杂场景的说话人定位与提取 1

[13:30-15:00] Fri-A-T-1 - 操控生成：面向音频生成与编辑的条件扩散、桥、与引导方法 2

[15:30-17:00] Fri-A-T-2 - 语音增强前沿进展：从预测式估计迈向生成式建模 2

[13:30-15:00] Fri-A-IND-1 - Industry Forum I 2

[15:30-17:00] Fri-A-IND-2 - Industry Forum II 2

Saturday, November 7

[08:00-18:00] Sat-R - 会议报到 3

[08:30-18:00] Sat-EXH - 赞助商展览 3

[08:30-09:00] Sat-OPEN - 大会开幕式（含大会合影） 3

[09:00-09:45] Sat-M-PLN-1 - The Way of Multimodal Communication: From Emotion to Health 3

[09:45-10:30] Sat-M-PLN-2 - Speaker Tracking: Segment-online Multi-talker Organization with a Varying Number of Speakers	3
[11:00-12:30] Sat-M-O-1 - Phonetics, Pronunciation Assessment and ASR Error Correction	3
[11:00-12:30] Sat-M-O-2 - Robust, Multimodal and Low-Resource Speech Recognition	4
[11:00-12:30] Sat-M-O-3 - Expressive and Controllable Speech Synthesis	5
[11:00-12:30] Sat-M-O-4 - Neural Speech Coding, Quantization and Compression	6
[11:00-12:30] Sat-M-SS-1 - 特殊议题 1: ChinaVoices Challenge 2026 中文多方言语音识别挑战赛	7
[11:00-12:30] Sat-M-SS-2 - 特殊议题 2: 面向具身智能的空间听觉与人机交互 I	8
[11:00-12:30] Sat-M-SS-3 - 特殊议题 8: 共情计算与认知交互 I	8
[11:00-12:30] Sat-M-YF - Young Forum 1	8
[13:30-17:00] Sat-A-EDU - 教育科普论坛	9
[13:30-15:30] Sat-A-AAEL - AAEL 联合论坛: 语音 × 语言 × 多模态: 迈向智能交互新范式	9
[13:30-15:00] Sat-A-SF-1 - Student Forum I	9
[13:30-15:00] Sat-A-O-1 - Spatial Audio Encoding and Speech, Music and Sound Separation	10
[13:30-15:00] Sat-A-O-2 - Speaker Representations, Diarization and Dialect Identification	10
[13:30-15:00] Sat-A-O-3 - Speaker Verification, Voice Privacy and Watermarking	11
[13:30-15:00] Sat-A-P-1 - Speech Science, Cognitive Modeling and Healthcare	12
[13:30-15:00] Sat-A-SS-1 - 特殊议题 8: 共情计算与认知交互 II	18
[13:30-15:00] Sat-A-SS-2 - 特殊议题 2: 面向具身智能的空间听觉与人机交互 II	18
[15:30-17:00] Sat-A-SF-2 - Student Forum II	19
[15:30-17:15] Sat-A-O-4 - Audio Foundation Models and Sound Event Understanding	19
[15:30-17:00] Sat-A-O-5 - Target Speaker Extraction and Audio-Visual Speech Enhancement	20
[15:30-17:00] Sat-A-O-6 - Speech Enhancement, Restoration and Noise Control	21
[15:30-17:00] Sat-A-P-2 - Speech Recognition, Audio Understanding and Multimodal Interaction	21
[15:30-17:00] Sat-A-SS-3 - 特殊议题 5: 残障研究生视角下的无障碍交互与病理语音智能处理	28
[15:30-17:00] Sat-A-SS-4 - 特殊议题 6: 面向生成式音频的可信音频处理方法	28
[17:20-18:05] Sat-A-PLN-1 - 四十年从理想到现实	29
[20:30-22:30] Sat-E-FUND - 基金分享沙龙	29

Sunday, November 8

[08:00-18:00] Sun-R - 会议报到	30
[09:00-18:00] Sun-EXH - 赞助商展览	30
[09:00-09:45] Sun-M-PLEN-1 - YuE2: Bringing the Human into the Loop of Symbiotic Creativity in Music	30
[09:45-10:30] Sun-M-PLEN-2 - 迈向全模态交互模型	30
[11:00-12:30] Sun-M-O-1 - Spotlight Paper Presentation I	30
[11:00-12:30] Sun-M-O-2 - Speech Spoofing and Content Manipulation Detection	31
[11:00-12:30] Sun-M-O-3 - Generative Speech Synthesis and Latent Representations	32
[11:00-12:30] Sun-M-SS-1 - 特殊议题 7: 语音与音乐的动态交互	33
[11:00-12:30] Sun-M-SS-2 - 特殊议题 3: 第三届多语言智能语音与语言技术研讨会	34
[11:00-12:40] Sun-M-SS-3 - 特殊议题 12: 动态场景下基于麦克风阵列的声场建模与语音感知	34
[11:00-12:30] Sun-M-SS-4 - 特殊议题 10: 语音模型评测体系构建与开源生态创新	35
[11:00-12:30] Sun-M-YF - Young Forum 2	35
[14:00-15:30] Sun-A-O-1 - Spotlight Paper Presentation II	36
[14:00-15:30] Sun-A-O-2 - Multimodal Emotion, Sentiment and Group Affect Recognition	37
[14:00-15:30] Sun-A-O-3 - Spoken Dialogue Systems, Turn-Taking and Interaction Modeling	38
[14:00-15:30] Sun-A-P-1 - Speech and Audio Generation, Coding and Restoration	38
[14:00-15:30] Sun-A-SS-1 - 特殊议题 9: 中国汉语方言及民族语言音节韵尾的共时变异与历时演化: 跨学科的多维视角	44
[14:00-15:30] Sun-A-SS-2 - 特殊议题 4: 复杂声学场景下的目标音频提取	45
[14:00-15:30] Sun-A-YF-1 - Young Forum 3	45
[15:45-17:15] Sun-A-O-4 - Speech and Multimodal Intelligence for Health and Cognitive Assessment	46
[15:45-17:15] Sun-A-O-5 - Speech and Audio-Visual Resources, Retrieval and Evaluation	47
[15:45-17:15] Sun-A-O-6 - Speech, Audio and Visual Speech Generation	48
[15:45-17:15] Sun-A-P-2 - Robust Speech Processing, Speaker Modeling and Audio Security	48
[15:45-17:15] Sun-A-SS-3 - 特殊议题 11: 2026 年 CCF 先进音频技术竞赛	54
[15:45-17:15] Sun-A-YF - Young Forum 4	55
[17:35-18:00] Sun-CLOSE - 大会闭幕式	55
Organizer List	56

Thursday, November 5

Thu-R: 会议报到
Room: 会议室前厅
Type: Registration
November 5, 2026, 2:00 PM to 5:00 PM

Friday, November 6

Fri-R: 会议报到
Room: 会议室前厅
Type: Registration
November 6, 2026, 8:00 AM to 5:00 PM

Fri-EXH: 赞助商展览
Room: 会议室前厅
Type: Exhibition
November 6, 2026, 9:00 AM to 5:00 PM

Fri-M-T-1: 语音交互 + 情感大模型：从判别式情感识别迈向生成式情感理解
Room: 8M 会议室
Type: Tutorial
November 6, 2026, 9:00 AM to 10:30 AM
Presenter: 连政老师，同济大学副教授、博士生导师

Fri-M-T-2: 鲁棒对话智能：复杂场景下人机语音交互感知与表达
Room: 8F 会议室
Type: Tutorial
November 6, 2026, 9:00 AM to 10:30 AM
Presenter: 刘瑞老师，内蒙古大学教授、博士生导师

Fri-M-T-3: 面向深度伪造语音的话者隐私保护：语音匿名化与水印
Room: 8M 会议室
Type: Tutorial
November 6, 2026, 11:00 AM to 12:30 PM
Presenter: 陈丽萍老师，中国科学技术大学特任副研究员

Fri-M-T-4: 空间多模态听觉感知：面向真实复杂场景的说话人定位与提取
Room: 8F 会议室
Type: Tutorial
November 6, 2026, 11:00 AM to 12:30 PM
Presenter: 钱馨园老师，北京科技大学副教授

Fri-A-T-1: 操控生成：面向音频生成与编辑的条件扩散、桥、与引导方法

Room: 8M 会议室

Type: Tutorial

November 6, 2026, 1:30 PM to 3:00 PM

Presenter: 陈泽华, 清华大学博士后

Fri-A-T-2: 语音增强前沿进展：从预测式估计迈向生成式建模

Room: 8M 会议室

Type: Tutorial

November 6, 2026, 3:30 PM to 5:00 PM

Presenter: 江文斌老师, 杭州电子科技大学特聘副教授、硕士生导师

Fri-A-IND-1: Industry Forum I

Room: 8F 会议室

Type: Industry Forum

November 6, 1:30 PM to 3:00 PM

Fri-A-IND-2: Industry Forum II

Room: 8F 会议室

Type: Industry Forum

November 6, 3:30 PM to 5:00 PM

Saturday, November 7

Sat-R: 会议报到
Room: 会议室前厅
Type: Registration
November 7, 2026, 8:00 AM to 6:00 PM

Sat-EXH: 赞助商展览
Room: 会议室前厅
Type: Exhibition
November 7, 2026, 8:30 AM to 6:00 PM

Sat-OPEN: 大会开幕式 (含大会合影)
Room: 沐辉厅 A+B+C
Type: Opening Ceremony
November 7, 2026, 8:30 AM to 9:00 AM

Sat-M-PLEN-1: The Way of Multimodal Communication: From Emotion to Health
Room: 沐辉厅 A+B+C
Type: Plenary
November 7, 2026, 9:00 AM to 9:45 AM
Presenter: Björn W. Schuller 老师

Sat-M-PLEN-2: Speaker Tracking: Segment-online Multi-talker Organization with a Varying Number of Speakers
Room: 沐辉厅 A+B+C
Type: Plenary
November 7, 2026, 9:45 AM to 10:30 AM
Presenter: 汪德亮老师

Sat-M-O-1: Phonetics, Pronunciation Assessment and ASR Error Correction
Room: 7 号会议室
Type: Oral
November 7, 11:00 AM to 12:30 PM

11:00-11:15

Sat-M-O-1-1: Mandarin Native Speakers' Perception of Voicing and Aspiration in Hindi Consonants Based on the Perceptual Assimilation Model

Yusheng Mu (Shanghai International Studies University); Xingwen Han (The Chinese University of Hong Kong, Shenzhen); lei zhu (Shanghai International Studies University)*

11:15-11:30

Sat-M-O-1-2: Prosodic features of Chinese sports commentary discourse: a case study of basketball broadcasting

Zepeng Wang (Wuhan University; University of Cologne)*; Duo Xu (Harbin Engineering University); Fangfang Long (Guizhou Normal University); Yansheng Mao (Harbin Engineering University)

11:30-11:45

Sat-M-O-1-3: An Initial-Final-Tone Rubric-Constrained Diagnostic Scoring Framework for Mandarin Word-Level Pronunciation Assessment

Yuxi Jiang (Xinjiang University)*; Zaokere Kadeer (Xinjiang University); Aishan Wumaier (Xinjiang University); Xudong Pang (Xinjiang University); Mulati Kahaer (Xinjiang University); Liye Chen (Xinjiang University)

11:45-12:00

Sat-M-O-1-4: HuM2PA: Multi-Aspect, Multi-Granularity Automatic Pronunciation Assessment with HuPER-Derived Phoneme Posterior Features

Haifeng Wang (Xinjiang University)*; Zaokere Kadeer (Xinjiang University); Aishan Wumaier (Xinjiang University); Mulati Kahaer (Xinjiang University); Xudong Pang (Xinjiang University); Liye Chen (Xinjiang University); Chendi Li (Xinjiang University)

12:00-12:15

Sat-M-O-1-5: Self-training with Pseudo-label Alignment and Label-type Prompting Enhances Mispronunciation Detection and Diagnosis Ability in Multimodal Large Language Models

Minglin Wu (The Chinese University of Hong Kong)*; Jing Xu (The Chinese University of Hong Kong); Tianhua Zhang (The Chinese University of Hong Kong); Xueyuan Chen (The Chinese University of Hong Kong); Helen Meng (The Chinese University of Hong Kong)

12:15-12:30

Sat-M-O-1-6: PAFCorrect: Pronunciation-Guided Acoustic-Phonetic-Text Fusion for Accented ASR Correction

Jiahao Yan (Tianjin University)*; Nan Li (Tianjin Normal University); Yueying Chen (Tianjin University); Peiyuan Jiang (Tianjin University); Tianrui Wang (Tianjin University); Longbiao Wang (Tianjin University); Jianwu Dang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)

Sat-M-O-2: Robust, Multimodal and Low-Resource Speech Recognition

Room: 8M 会议室

Type: Oral

November 7, 11:00 AM to 12:30 PM

11:00-11:15

Sat-M-O-2-1: CTT-ED: A Unified Morpheme-Tail Framework for Low-Resource Speech Recognition in Morphologically Rich Languages

Yuhui Ming (Xinjiang University)*; Maihemuti Maimaiti (Xinjiang University); Yinfeng Yu (Xinjiang University); Zhifan Pan (Xinjiang University); Aishan Wumaier (Xinjiang University)

11:15-11:30

Sat-M-O-2-2: Heard but Not Transcribed: Diagnosing Chinese ASR Pipelines with Controlled Remixes of Real Drone Ego-Noise

Yanxiao Li (Macao Polytechnic University); Junhao Wei (Macao Polytechnic University); Yifu Zhao (Macao Polytechnic University); Xu Yang (Macao Polytechnic University); Yapeng Wang (Macao Polytechnic University)*

11:30-11:45

Sat-M-O-2-3: VSR-VLM: CTC-Grounded Vision-Language Decoding for Chinese Continuous Visual Speech Recognition

Bochao Hu (bupt)*; Zehua Liu (Beijing University of Posts and Telecommunications); Xiaolou Li (Beijing University of Posts and Telecommunications); Lantian Li (Beijing University of Posts and Telecommunications); Dong Wang (Tsinghua University)

11:45-12:00

Sat-M-O-2-4: Balancing Mandarin and Dialect Recognition in Chinese Multi-Dialect ASR via On-Policy Self-Distillation

shuiyuan wang (Northwestern Polytechnical University)*; bingshen mu (Northwestern Polytechnical University); pengshen zhang (WeNet Community); chengyou wang (Northwestern Polytechnical University); yujie liao (Northwestern Polytechnical University); chengdong liang (WeNet Community); binbin zhang (WeNet Community); qiangze feng (NEXDATA TECHNOLOGY INC.); lei xie (Northwestern Polytechnical University)

12:00-12:15

Sat-M-O-2-5: Low-Resource Cross-Regional Dialect Speech Recognition Based on Knowledge Distillation

yongduo zhang (Guizhou Provincial Laboratory of Big Data, State Key Laboratory of Public Big Data, Guizhou University, Guiyang, Guizhou 550025); juntian feng (Guizhou Provincial Laboratory of Big Data, State Key Laboratory of Public Big Data, Guizhou University, Guiyang, Guizhou 550025); lei zhang (Guizhou Provincial Laboratory of Big Data, State Key Laboratory of Public Big Data, Guizhou University, Guiyang, Guizhou 550025)*; hong peng (Sichuan Chuantou Zhisheng Digital Technology Co., Ltd.)

12:15-12:30

Sat-M-O-2-6: Balanced Multi-Stage Adaptation of Pretrained ASR for Low-Resource Chinese Dialects

Shenghui Lu (Malanshan Audio&Video Laboratory)*; Yongfeng Tu (Malanshan Audio&Video Laboratory); Yishuang Li (Malanshan Audio&Video Laboratory); Caixia Gong (Malanshan Audio&Video Laboratory); Wanli Dong (Malanshan Audio&Video Laboratory); Weihao Gan (Malanshan Audio&Video Laboratory)

Sat-M-O-3: Expressive and Controllable Speech Synthesis

Room: 8F 会议室

Type: Oral

November 7, 11:00 AM to 12:30 PM

11:00-11:15

Sat-M-O-3-1: Zero-Shot Emotional Voice Conversion with Flow-Step-Aware Emotion-Speaker Modulation

nianzu li (Beijing University of Posts and Telecommunications)*; Liu Gang (Beijing University of Posts and Telecommunications); Yang Shiqi (Beijing University of Posts and Telecommunications)

11:15-11:30

Sat-M-O-3-2: Efficient Emotion and Speaker Adaptation in LLM-Based TTS via Characteristic-Specific Partial Fine-Tuning

Tianrui Wang (Tianjin University)*; Meng Ge (Tianjin University); Cheng Gong (Tianjin University); Chunyu Qiang (Tianjin University); Haoyu Wang (Tianjin University); Zikang Huang (Tianjin University); Yu Jiang (Tianjin University); Ye Ni (Southeast University); Yuheng Lu (Tianjin University); Xiaobao Wang (Tianjin University); Engsiong Chng (Nanyang Technological University); Xie Chen (Shanghaijiaotong University); Longbiao Wang (Tianjin University); Jianwu Dang (Tianjin University)

11:30-11:45

Sat-M-O-3-3: Harness TTS: Towards Context-Aware Expressive Speech Synthesis with Harness Layer

Shengfan Shen (Nanjing University); Di Wu (MiLM Plus, Xiaomi Inc.); Xingchen Song (MiLM Plus, Xiaomi Inc.); Dinghao Zhou (MiLM Plus, Xiaomi Inc.); Pengyu Cheng (MiLM Plus, Xiaomi Inc.); Sixiang Lyu (The Chinese University of Hong Kong, Shenzhen); Jian Luan (MiLM Plus, Xiaomi Inc.); Shuai Wang (Nanjing University)*

11:45-12:00

Sat-M-O-3-4: CapDiT: A Caption-Guided Diffusion Transformer for Zero-Shot Text-to-Speech with Expressive Paralinguistic Vocalizations

Tingxiao Zhou (Shanghai Jiao Tong University)*; Leying Zhang (Shanghai Jiao Tong University); Ya Liu (Shanghai Jiao Tong University); Yijun Huang (Shanghai Jiao Tong University); Chenda Li (Shanghai Jiao Tong University); Yang Wang (Didichuxing Co. Ltd); Zhiqiang Lv (Didichuxing Co. Ltd); Liyin Liang (Didichuxing Co. Ltd); Yifeng Qian (Didichuxing Co. Ltd); Yanmin Qian (Shanghai Jiao Tong University)

12:00-12:15

Sat-M-O-3-5: Coarse-to-Fine Style Prompt Enrichment for User-Friendly Controllable Speech Generation

yubao Sun (Inner Mongolia University); Rui Liu (Inner Mongolia University)*; guanglai Gao (Inner Mongolia University)

12:15-12:30

Sat-M-O-3-6: Time-Layer Adaptive Alignment for Speaker Similarity in Flow-Matching Based Zero-Shot TTS

Haoyu Li (Shanghai Jiao Tong University)*; Mingyang Han (Taobao & Tmall Group of Alibaba); Yu Xi (Shanghai Jiao Tong University); Dongxiao Wang (Taobao & Tmall Group of Alibaba); Hankun Wang (Shanghai Jiao Tong University); Haoxiang Shi (Taobao & Tmall Group of Alibaba); Boyu Li (Taobao & Tmall Group of Alibaba); Jun Song (Taobao & Tmall Group of Alibaba); Bo Zheng (Taobao & Tmall Group of Alibaba); Shuai Wang (Nanjing University); Kai Yu (Shanghai Jiao Tong University)

Sat-M-O-4: Neural Speech Coding, Quantization and Compression

Room: 2R 会议室

Type: Oral

November 7, 11:00 AM to 12:30 PM

11:00-11:15

Sat-M-O-4-1: Robust Residual Finite Scalar Quantization for Neural Compression

Xiaoxu Zhu (XPeng Motors)*; Dongchuan Ran (Independent Researcher); Yiming Ren (Beijing Tongming Lake Information Technology Application Innovation Center (TLAIC),); Baoxiang Li (Shanghai Artificial Intelligence Laboratory)

11:15-11:30

Sat-M-O-4-2: TOSS-Codec: A Temporal-Offset-Supervised Streaming Audio Codec for Mobile Devices

Yingda Shen (The Chinese University of Hong Kong, Shenzhen); Yuxuan Chen (The Chinese University of Hong Kong, Shenzhen); Jiaqi Li (The Chinese University of Hong Kong, Shenzhen); Ruixiang Hang (Transsion Holdings Co. Ltd); Ruihai Jing (Transsion Holdings Co. Ltd); Junwen Qiu (The Chinese University of Hong Kong, Shenzhen); Yudong Li (The Chinese University of Hong Kong, Shenzhen); Zhizheng Wu (The Chinese University of Hong Kong, Shenzhen)*

11:30-11:45

Sat-M-O-4-3: MrCodec: A Lightweight Multiple Bitrate Scalar Quantization Neural Speech Codec for Real-Time Communication

Xusheng Yang (Peking University)*; Wei Xiao (Tencent); Yongkang Yin (Peking University); Zhengyu Chen (Zhejiang University); Shidong Shang (Tencent); Weibei Dou (Tsinghua University); Yuexian Zou (Peking University)

11:45-12:00

Sat-M-O-4-4: UniStream: Multi-Expert Residual Vector Quantization for 48 kHz Causal Streaming Audio Coding

Mingyu Zhao (Tsinghua Shenzhen International Graduate School)*; Zhiyong Wu (Tsinghua Shenzhen International Graduate School)

12:00-12:15

Sat-M-O-4-5: Deep Hierarchical Speech Coding

Junteng Zhang (Nanjing University)*; Zhaoxuan Liu (Nanjing University); Lizhong Wang (Samsung Research China-Beijing); Yuxing Zheng (Samsung Research China-Beijing); Liwen Shi (Huaqiao University); Ming Lu (Nanjing University); Zhan Ma (Nanjing University)

12:15-12:30

Sat-M-O-4-6: UniSRCCodec: Ultra Low-Bitrate Single-Codebook Audio Codec based on Decoupled Sub-band Reconstruction

Zhisheng Zhang (Tsinghua University)*; Xiang Li (Tsinghua University); Yixuan Zhou (Tsinghua University); Jing Peng (Tsinghua University); Shengbo Cai (Tsinghua University); Guoyang Zeng (ModelBest); Zhiyuan Zhao (International Digital Economy Academy); Jingyi Li (International Digital Economy Academy); Zhiyong Wu (Tsinghua University)

Sat-M-SS-1: 特殊议题 1: ChinaVoices Challenge 2026 中文多方言语音识别挑战赛

Room: 沐辉厅 B

Type: Special Session

November 7, 2026, 11:00 AM to 12:30 PM

Organizer: 谢磊老师, 薛浏蒙老师, 刘和鑫老师, 石宪老师, 胡婕老师

11:00-11:15

Sat-M-SS-1-1: 闽南话智能语音系统

Presenter: 洪青阳老师, 厦门大学教授

11:15-11:30

Sat-M-SS-1-2: 东南亚及"一带一路"国家多语言语音数据资源建设

Presenter: 陈谐老师, 上海交通大学、上海创智学院副教授

11:30-11:45

Sat-M-SS-1-3: G-STAR: 面向长时多说话人语音的端到端识别框架

Presenter: 王帅老师, 南京大学副教授

11:45-12:00

Sat-M-SS-1-4: WenetSpeech 方言语料家族

Presenter: 朱鹏程, WeNet 开源社区

12:00-12:15

Sat-M-SS-1-5: 方言数据构建与语音技术赋能

Presenter: 胡婕老师, 北京慧听科技有限公司总监

12:15-12:30

Sat-M-SS-1-6: ChinaVoices Challenge 2026 竞赛总结与分析

Presenter: 廖育杰博士, 西北工业大学

Sat-M-SS-2: 特殊议题 2: 面向具身智能的空间听觉与人机交互 I
Room: 沐辉厅 C
Type: Special Session
November 7, 2026, 11:00 AM to 12:30 PM
Organizer: 朱梦尧老师, 杜俊老师, 卜辉老师, 薛浏蒙老师

11:00-11:30
Sat-M-SS-2-1: Egocentric Audio-Visual Understanding: From Multimodal Signal Processing to Wearable Spatial Intelligence
Presenter: 王文武老师, 英国萨里大学教授

11:30-12:00
Sat-M-SS-2-2: 具身听觉: 智能体探索物理世界的感知新范式
Presenter: 陈凯老师, 深度机智 & 北京中关村学院具身智能方向负责人

12:00-12:30
Sat-M-SS-2-3: 空间音频智能化生成
Presenter: 赵洲老师, 浙江大学教授

Sat-M-SS-3: 特殊议题 8: 共情计算与认知交互 I
Room: 2D 会议室
Type: Special Session
November 7, 2026, 11:00 AM to 12:30 PM
Organizer: 黄东延老师, 赵子平老师, 张子兴老师, 李雅老师, 刘斌老师

11:00-11:30
Sat-M-SS-3-1: Emotion Recognition in Speech via Learning on Self-Estimated Labels Using a Dual-Branch MoE-Based Network
Presenter: 赵子平老师, 天津师范大学教授

11:30-12:00
Sat-M-SS-3-2: RA-MERG: A Reasoning-Augmented Multi-modal Empathetic Response Generation Benchmark
Presenter: 刘斌老师, 中国科学院自动化研究所副教授

12:00-12:30
Sat-M-SS-3-3: 面向共情与认知交互的非典型语音理解: 可解释与包容性视角
Presenter: 岳政君老师, 深圳河套学院语言模型与人机交互中心助理教授

Sat-M-YF: Young Forum 1
Room: 沐辉厅 A
Type: Young Forum
November 7, 2026, 11:00 AM to 12:30 PM

11:00-11:30
Sat-M-YF-1: 低码率神经网络语音编解码方法研究
Presenter: 艾杨老师, 中国科学技术大学语音及语言信息处理国家工程研究中心副研究员

11:30-12:00
Sat-M-YF-2: 面向自然人机交互的语音副语言能力评测: 从指令跟随到非语言发声
Presenter: 薛浏蒙老师, 南京大学智能科学与技术学院准聘助理教授

12:00–12:30

Sat-M-YF-3: 迈向更稳定的全连续自回归语音合成

Presenter: 郑达老师, 小红书 dots 语音团队; 王翰坤, 上海交通大学 X-LANCE 博士生; 李波含, 上海交通大学博士生

Sat-A-EDU: 教育科普论坛

Room: 9 号会议室

Type: Education and Outreach Forum

November 7, 2026, 1:30 PM to 5:00 PM

Sat-A-EDU-1: 人工智能的奥秘

Presenter: 马少平老师, 清华大学长聘教授

Sat-A-EDU-2: 人工智能-产业变革与青年未来

Presenter: 洪青阳老师, 厦门大学教授

Sat-A-EDU-3: Design thinking: 以人为本的创新范式, 如何守住 AI 时代的创新起点

Presenter: 朱璇, 三星电子中国研究院研发副总裁

Sat-A-AAACL: AAACL 联合论坛: 语音 × 语言 × 多模态: 迈向智能交互新范式

Room: 沐和厅 B+C

Type: AAACL Joint Forum

November 7, 2026, 1:30 PM to 3:30 PM

Sat-A-AAACLJF-1: Toward Multi-Layer Protection of Speech Against Real-World Misuse: What Can the Speech Community Do?

Presenter: Junichi Yamagishi 老师, 日本国立信息研究院

Sat-A-AAACLJF-2: Challenges and Progress in Automatic Speech-to-Speech Translation: Bridging the Gap to Human Simultaneous Interpretation

Presenter: Satoshi Nakamura 老师, 香港中文大学 (深圳)

Sat-A-AAACLJF-3:

Presenter: 嘉宾待定

Sat-A-AAACLJF-4:

Presenter: 嘉宾待定

Sat-A-SF-1: Student Forum I

Room: 沐辉厅 A

Type: Student Forum

November 7, 2026, 1:30 PM to 3:00 PM

Organizer: 洪青阳 (厦门大学), 连政 (同济大学), 孔秋强 (香港中文大学)

Sat-A-O-1: Spatial Audio Encoding and Speech, Music and Sound Separation**Room: 7 号会议室****Type: Oral****November 7, 1:30 PM to 3:00 PM**

13:30-13:45

Sat-A-O-1-1: A Microphone Padding Approach for Speech Separation with Varying Number of Microphones

Fulin Wu (Southern University of Science and Technology)*; Linglin He (Southern University of Science and Technology); Shulin He (Southern University of Science and Technology); Yuankun Huang (OPPO, Dongguan, China); Mou Wang (OPPO, Dongguan, China); Zhong-Qiu Wang (Southern University of Science and Technology)

13:45-14:00

Sat-A-O-1-2: Breaking the Quality–Intelligibility Trade-off in Streaming Target Speaker Extraction via Deep-Feature-Anchored Preference Optimization

ShuHai Peng (Tsinghua University)*; Jinjiang Liu (Tsinghua University); Hui Lu (The Chinese University of Hong Kong); Liyang Chen (Tsinghua University); Guiping Zhong (SenseTime); Jiakui Li (SenseTime); Shiyin Kang (SenseTime); Zhiyong Wu (Tsinghua University)

14:00-14:15

Sat-A-O-1-3: Hierarchical Streaming Memory with Mamba State Compression for Continuous Speech Separation

Yanlin Wang (Xinjiang University)*; Ying Hu (Xinjiang University)

14:15-14:30

Sat-A-O-1-4: Band-SCNet-RT: Lightweight Causal Real-time Stereo Music Separation with Vocal Refinement and Complex Multi-channel Masking

Wenxin Bao (Tsinghua University)*; Jinjiang Liu (Tsinghua University); Zhanjia Zhang (Tsinghua University); Jiehui Luo (Tsinghua University); Fengjie Zhu (Transsion); Yunfeng Xu (Transsion); Zhiyong Wu (Tsinghua University)

14:30-14:45

Sat-A-O-1-5: USS-TDA: Universal Sound Separation with Transformer Decoder-Based Attractor

Xiaoxue Luo (Beijing University of Posts and Telecommunications)*; Li Lantian (Beijing University of Posts and Telecommunications); Wang Dong (Tsinghua University)

14:45-15:00

Sat-A-O-1-6: Physics-Informed Ambisonics Encoding for Circular Microphone Arrays via Reliability-Weighted Spatial Cues and Gated Neural Modulation

Congyu Huang (School of Information and Electronics, Beijing Institute of Technology)*; Zhiyu Li (School of Information and Electronics, Beijing Institute of Technology); Jing Wang (School of Information and Electronics, Beijing Institute of Technology)

Sat-A-O-2: Speaker Representations, Diarization and Dialect Identification**Room: 8M 会议室****Type: Oral****November 7, 1:30 PM to 3:00 PM**

13:30-13:45

Sat-A-O-2-1: Probing-Guided Feature-Space Adaptation for Frozen Chinese HuBERT Embeddings

Haobo Wang (Nanjing University of Posts and Telecommunications)*; Yiqiang Cai (Xi'an Jiaotong-Liverpool University); Yang Xu (Nanjing University of Posts and Telecommunications); Shengchen Li (Xi'an Jiaotong-Liverpool University); Shao Xi (Nanjing University of Posts and Telecommunication)

13:45-14:00

Sat-A-O-2-2: MoE Architectures for Transformer Encoder-Based EEND: Towards Scalable Multi-Speaker Diarization

Ziyi Chen (Nanjing University)*; Haoyu Li (Nanjing University); Jing Peng (Shanghai Jiao Tong University); Mengtian Li (Central Media Technology Institute, Huawei); Kai Yu (Shanghai Jiao Tong University); Shuai Wang (Nanjing University)

14:00-14:15

Sat-A-O-2-3: SURE: Reliability-Aware Enrollment for Open-Set Speaker Identification in Mandarin Deaf and Hard-of-Hearing Speech

Tongxi Chen (Lancaster University)*; Yue Ma (Southeast University); Weizhi Meng (Lancaster University); Yu Chen (Southeast University)

14:15-14:30

Sat-A-O-2-4: PULSE: Personalized Understanding with Large Audio-Language Model for Speaker-Attributed Audio Evaluation

Yuwen Wang (University of Science and Technology Beijing); Ziyang Jiang (University of Science and Technology Beijing); Yilin Ren (University of Science and Technology Beijing); Fuqiang Luo (University of Science and Technology Beijing); Qingyu Pan (ZhengZhou University); Xinyuan Qian (University of Science and Technology Beijing)*

14:30-14:45

Sat-A-O-2-5: Fusion of Complementary Layer-Selected Acoustic Embeddings from Multiple ASR Models for Dialect Classification in ChinaVoices Challenge 2026

Sicheng Yu (Chongqing University)*

14:45-15:00

Sat-A-O-2-6: Summary of the ChinaVoices Challenge 2026: Data, Tasks, Baseline, and Methods

Yujie Liao (Northwestern Polytechnical University)*; Bingshen Mu (Northwestern Polytechnical University); Shuiyuan Wang (Northwestern Polytechnical University); Liumeng Xue (School of Intelligence Science and Technology, Nanjing University); Hexin Liu (Nanyang Technological University); Xian Shi (Independent Researcher); Jie Hu (Beijing Huiting Technology Co., Ltd.); Lei Xie (Northwestern Polytechnical University)

Sat-A-O-3: Speaker Verification, Voice Privacy and Watermarking

Room: 8F 会议室

Type: Oral

November 7, 1:30 PM to 3:00 PM

13:30-13:45

Sat-A-O-3-1: SRPM: Analytic Class-Incremental Speaker Recognition with a Precision Matrix

Wenjia Du (National University of Defense Technology)*; Qisheng Xu (National University of Defense Technology); Cheng Yang (National University of Defense Technology); kele Xu (National University of Defense Technology)

13:45-14:00

Sat-A-O-3-2: Sub-Phonetic State Residual Disentanglement for Ultra-Short-Utterance Speaker Verification

Yi Zhou (Shanghai Acoustics Laboratory, Chinese Academy of Sciences); Feng Hong (Shanghai Acoustics Laboratory, Chinese Academy of Sciences)*

14:00-14:15

Sat-A-O-3-3: Proactive Voice Privacy Protection for Chinese Zero-Shot Voice Cloning with Perceptually Constrained Perturbations

Yiwei Hu (Qinhai Minzu University); Kuntharrrgyal Khysru (Qinghai Minzu University)*; Minghu Tang (Qinghai Minzu University); Quecuo De (Qinghai Minzu University); Ethan Deci (Qinhai Minzu University); Di Jin (Tianjin University); Jianguo Wei (Tianjin University)

14:15-14:30

Sat-A-O-3-4: SEAM: A Sparse and Energy-Aware Masking Framework for Imperceptible and Robust Speech Watermarking

Xin Fang (University of Science and Technology of China); Jin Li (University of Science and Technology of China)*; Tian Gao (Anhui SparkShield Intelligent Technology Co., Ltd); Junhua Liu (Anhui SparkShield Intelligent Technology Co., Ltd); Liang Zou (China University of Mining and Technology)

14:30-14:45

Sat-A-O-3-5: Asynchronous Voice Anonymization by Combining Adversarial Perturbations on Speaker Embedding and Waveform

胡中山 (中国科学技术大学)*; 陈丽萍 (中国科学技术大学); Kong Aik Lee (香港理工大学); 凌震华 (中国科学技术大学)

14:45-15:00

Sat-A-O-3-6: OpenSVBench: Diagnosing Speaker Verification Generalization Across Deployment Scenarios

Qituan Shangguan (School of Intelligence Science and Technology, Nanjing University, Suzhou, China)*; Bosen Xu (School of Intelligence Science and Technology, Nanjing University, Suzhou, China); Shuai Wang (School of Intelligence Science and Technology, Nanjing University, Suzhou, China)

Sat-A-P-1: Speech Science, Cognitive Modeling and Healthcare

Room: 2D 会议室 + 2R 会议室 (海报区)

Type: Poster

November 7, 1:30 PM to 3:00 PM

1. Phonetics, Prosody and Speech Production (10 papers)

Sat-A-P-1-1: 对连续话语中音节时长分析

xiaoxia chen (institute of linguistics,cass,china)*; aijun li (institute of linguistics,cass.china)

Sat-A-P-1-2: Acoustic Correlates of Politeness in Japanese Speech: A Controlled Production Study of Spectral and Voice Quality Features

Yanlong Zhang (Xi'an International Studies University)*; Tianxiang Cao (Xi'an International Studies University); Meiyi Shi (Xi'an International Studies University); Yuxin Wu (Xi'an International Studies University)

Sat-A-P-1-3: WavFry: A Noise-Robust Data-Driven Approach for Creaky Voice Detection in Conversational Speech

宣涛 (北京外国语大学)*; 龙帅宇 (北京外国语大学)

Sat-A-P-1-4: Design of the ToBI-plus Universal Phonetic Annotation System

Zhigang Yin (Institute of Linguistics, Chinese Academy of Social Sciences)*; Lijun Gao (Department of Linguistics, University of Chinese Academy of Social Sciences)

Sat-A-P-1-5: 基于机器学习的汉语强调焦点句声学层级研究

Yiheng Sun (Nankai University)*; Jianghua Guo (College of Chinese Language and Culture,Nankai University); Xudong Shen (School of Data Information, Tsinghua University); Baoying Wen (College of Chinese Language and Culture,Nankai University)

Sat-A-P-1-6: 基于空气动力学特性的普通话擦音生理格局

Xiujuan Shi (shandong university)*

Sat-A-P-1-7: 电磁发音仪传感器对普通话单元音的声学影响

Zhiyao Jing (University of Chinese Academy of Social Sciences)*; Qiang Fang (Institute of Linguistics, Chinese Academy of Social Sciences)

Sat-A-P-1-8: 《基于自监督表征的元音归一化及其生理验证》

shuo shi (Hebei University)*; Qiang Jun Wang (Hebei University)

Sat-A-P-1-9: 汉语普通话语音鉴定中长时基频特征的时长阈限

Honglin Cao (China University of Political Science and Law)*; Chuyi Pan (China University of Political Science and Law); Yijing Lei (China University of Political Science and Law); Danlin Wang (Beijing Police College)

Sat-A-P-1-10: 基于共振峰参数与唇形特征的普通话韵母四呼无监督聚类分析

Liwen Jing (中山大学)*; Bei Yang (中山大学)

2. Dialects, Bilingualism and Second-Language Speech (14 papers)

Sat-A-P-1-11: 高州粤语长短元音的焦点效应研究

任昕 (College of Chinese Language and Culture, Jinan University); Maolin Wang (Zhanjiang University of Science and Technology College of Culture and Communication); Han Yu (College of Chinese Language and Culture, Jinan University)*

Sat-A-P-1-12: Interactional Strategies and Intonational Variation: Evidence from Kazakh-Mandarin Bilinguals' Natural Conversations

Yuxuan Wu (College of Language and Culture, Northwest A&F University)*

Sat-A-P-1-13: 语音变异程度对汉语二语者声调感知的影响

于闯 (新疆大学)*; 王玮 (新疆大学)

Sat-A-P-1-14: An Experimental Study on the Perception and Production of Mandarin Fricatives by Cantonese Native Speakers

JING YANG (Dongguan University of Technology)*

Sat-A-P-1-15: A Comparative Study on Tones in Hebei Dialects

Jinguang Zhang (Hebei Normal University)*

Sat-A-P-1-16: An Error Analysis of Chinese Disyllabic Sentences by Bangladeshi Learners

Jia Jing (Institute of Applied Linguistics, Ministry of Education)*; Ruiyuan Wang (Beijing Language and Culture University)

Sat-A-P-1-17: Cross-Generational Accommodation of Cantonese Sound Change Features Toward Human and AI Interlocutors

Yitian Hong (Department of Linguistics and Modern Languages, The Chinese University of Hong Kong)*; Manson Chun Man Wong (The Chinese University of Hong Kong); Grace Wenling Cao (University College Dublin); Pik Ki Peggy Mok (The Chinese University of Hong Kong)

Sat-A-P-1-18: 藏式英语塞音塞擦音偏误的合成语音复刻

Sihong Zhang (Guangdong University of Technology); Yu Zhang (Hefei University of Technology)*

Sat-A-P-1-19: 阳江闸坡疍家话单字调实验研究

Fan Jingyi (Jinan University); Gao Yingming (Beijing University of Posts and Telecommunications); Liu Min (Jinan University)*

Sat-A-P-1-20: Prosodic Realization of Nouns in Canonical vs. Non-Canonical Word Order in Second Language Mandarin by Learners of Tone vs. Non-Tone Languages

Yuxin Feng (Nanjing University of Science and Technology); Xinyue Wang (Nanjing University of Science and Technology); Ying Chen (Nanjing University of Science and Technology)*

Sat-A-P-1-21: Evaluating Montreal Forced Aligner for L2 English Stop Analysis: Pretrained and Speaker-Adapted Models on Chinese EFL Learners' Speech

邓斌 (Yangzhou University)*

Sat-A-P-1-22: Effects of Noise on the Recognition of Cantonese Tones: A Speaker-Independent XGBoost Classification Study

Wenqi DONG (University of Macau)*; Jingwei ZHANG (University of Macau)

Sat-A-P-1-23: 粤语韵尾功能负载研究——基于拼音-文本互信息模型的量化分析

Jinsong Zhang (Beijing Language and Culture University)*

Sat-A-P-1-24: 基于元音声学参数的阿美语方言分类研究

TIAN XiangXiang (Minzu University of China)*

3. Pronunciation Assessment and Child Speech (4 papers)

Sat-A-P-1-25: BiSP-RPR: End-to-End Canonical Phoneme-Tone Prediction for Reference Construction in Mandarin Pronunciation Assessment

韩肖肖 (新疆大学)*; Aishan Wumaier (XinJiang University); Zaokere Kadeer (Xinjiang University); Song Zhengping (Xinjiang University); Mulati Kahaer (XinJiang University); Gao Yingxiong (Xinjiang University)

Sat-A-P-1-26: ATR-MDD: An Attribute-Level, Tone- and Role-Aware Framework for Mandarin L2 Mispronunciation Detection and Diagnosis

Simiao Zhang (Shanghai Jiaotong University)*; Mengyue Wu (Shanghai Jiaotong University); Juqiang Chen (Shanghai Jiaotong University)

Sat-A-P-1-27: Multi-Factor Feedback and Scoring compensation for Automatic Pronunciation Assessment

ren ling (cvte); li liang (cvte)*; lei yan (cvte)

Sat-A-P-1-28: A Bi-LSTM Model to Detect Voicing Transition Errors in the Production of Single Words Spoken by Australian Pre-schoolers

Ying Li (Curtin University)*; Duc-Son Pham (Curtin University); Roslyn Ward (Curtin University); Neville Hennessey (Curtin University); Tele Tan (Curtin University)

4. Speech Perception, Hearing and Neural Modeling (10 papers)

Sat-A-P-1-29: 老年人群对 AI 合成普通话音节的识别研究

Wencui Liu (Tongji University)*; Ziyang Zhou (Tongji University); Zhaoyue Wang (Tongji University); Chenqi Wang (Tongji University); Xiaohu Yang (Tongji University)

Sat-A-P-1-30: Word Surprisal Modulates Information Redundancy Across Human Single Neurons, with a Similar Trend in LLMs

Xuanda Chen (The Chinese University of Hong Kong, Shenzhen)*; Mi Tang (Harbin Institute of Technology, Shenzhen); Siqi Cai (Harbin Institute of Technology, Shenzhen); Haizhou Li (The Chinese University of Hong Kong, Shenzhen)

Sat-A-P-1-31: 双言经验对南京老年人声调感知的影响

Yuqing Wang (1. School of Foreign Studies, Tongji University; 2. Center for Speech and Language Processing, Tongji University)*; Qianwen Jin (1. School of Foreign Studies, Tongji University; 2. Center for Speech and Language Processing, Tongji University); Jinxin Ji (1. School of Foreign Studies, Tongji University; 2. Center for Speech and Language Processing, Tongji University); Xiaohu Yang (1. School of Foreign Studies, Tongji University; 2. Center for Speech and Language Processing, Tongji University)

Sat-A-P-1-32: EviARC: Reliable Single-Epoch EEG Decoding of Speaker-Change Responses via Evidential Adaptive Risk Calibration

Xingui Wang (Tsinghua University); Puyang Geng (Academy of Forensic Science); Yuxuan Jiang (Tsinghua University); Weibei Dou (Tsinghua University)*; Hong Guo (Academy of Forensic Science)

Sat-A-P-1-33: Auditory spatial attention detection based on ear EEG signals using a ResNet with channel attention

Sijia Song (Tianjin University); Zhuang Xie (Tianjin University); Mengze Du (Tianjin University); Gaoyan Zhang (Tianjin University)*

Sat-A-P-1-34: MDFNet: A Multi-Domain Fusion Network for EEG-Based Cross-Session Auditory Attention Decoding

Mengyuan Gao (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, School of Computer Science and Technology, Anhui University); Xingguang Dong (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, School of Computer Science and Technology, Anhui University); Jianyu Chen (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, School of Computer Science and Technology, Anhui University); Yu Jiang (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, School of Computer Science and Technology, Anhui University); Zhao Lv (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, School of Computer Science and Technology, Anhui University); Cunhang Fan (Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, School of Computer Science and Technology, Anhui University)*

Sat-A-P-1-35: Perceptual Characteristics of Emotional Text-to-Speech in Deaf and Hard-of-Hearing Individuals

Yinuo Wang (Southeast University); Tongxi Chen (Lancaster University)*; Weizhi Meng (Lancaster University); Yu Chen (Southeast University)

Sat-A-P-1-36: Segmental and Tonal Production in Mandarin-speaking Adults with Cochlear Implants: A Comparison of Prelingually and Postlingually Deafened Users

Ruwei Wu (Nanjing University of Science and Technology); Ping Tang (Nanjing University of Science and Technology)*

Sat-A-P-1-37: Speech Perception in Mandarin-Speaking Adults with Cochlear Implants During the Early Stage After Activation

Wenyu Xiang (Nanjing University of Science and Technology); Ping Tang (Nanjing University of Science and Technology)*

Sat-A-P-1-38: BRAE: A Brain-Region-Aware Evidence Framework for LLM-Generated Explanation of EEG-Based Emotion Recognition

Chenyang Li (Beijing Institute of Technology, Zhuhai); Xinfang Kang (Beijing Institute of Technology, Zhuhai); Lin Shen (Beijing Institute of Technology); Rui Wang (Beijing Institute of Technology, Zhuhai); Minghao Yan (Beijing Institute of Technology, Zhuhai); Kun Qian (Beijing Institute of Technology)*; Bin Hu (Beijing Institute of Technology); Björn Schuller (Technische Universität München)

5. Clinical Speech Recognition and Rehabilitation (9 papers)

Sat-A-P-1-39: Advancing Sentence-Level Dysarthric Audio-Visual Speech Recognition via Test-Time Adaptation

shiyao wang (Nankai University)*; Hui Wang (Nankai University); Jiaming Zhou (Nankai University); Shiwan Zhao (Nankai University); Yong Qin (Nankai University)

Sat-A-P-1-40: TTS-PRISM: A Perceptual Reasoning and Interpretable Speech Model for Fine-Grained Diagnosis

Xi Wang (Tsinghua University)*; Jie Wang (Xiaomi); Xingchen Song (Xiaomi); Jingran Xie (Tsinghua University); Baijun Song (Tsinghua University); Jiahe Shao (The University of Tokyo); Zijian Lin (Tsinghua University); Di Wu (Xiaomi); Meng Meng (Xiaomi); Jian Luan (Xiaomi); Zhiyong Wu (Tsinghua University)

Sat-A-P-1-41: Macro-Descriptors Derived from Large Language Models for Low-Resource Cantonese Aphasia Detection and Subtyping

YuFeng Nie (Visual Intelligence +X International Cooperation Joint Laboratory of MOE, Beijing Jiaotong University)*; Ying Qin (Visual Intelligence +X International Cooperation Joint Laboratory of MOE, Beijing Jiaotong University); Anthony Pak Hin Kong (Unit of Human Communication, Development, and Information Sciences, The University of Hong Kong, Hong Kong, China); shikui wei (Visual Intelligence +X International Cooperation Joint Laboratory of MOE, Beijing Jiaotong University)

Sat-A-P-1-42: 面向重测信度提升的普通话构音障碍严重度自动评估方法

李琪伦 (深圳大学)*; 李娟娟 (深圳市龙岗区耳鼻咽喉医院); 郑慧丽 (深圳市龙岗区耳鼻咽喉医院); 魏朝刚 (深圳市龙岗区耳鼻咽喉医院); 孟庆林 (华南理工大学); 郑能恒 (深圳大学)

Sat-A-P-1-43: 环境噪声下中文构音障碍严重程度分类鲁棒性研究

Lin Ma (Institute of Software Chinese Academy of Sciences); Zexia Fan (University of Science and Technology Beijing); Yicong Jiang (Institute of Software Chinese Academy of Sciences); Xurong Xie (Institute of Software Chinese Academy of Sciences)*; Xiaolan Peng (Institute of Software Chinese Academy of Sciences); Jin Huang (Institute of Software Chinese Academy of Sciences); Feng Tian (Institute of Software Chinese Academy of Sciences)

Sat-A-P-1-44: DuBA: Parameter-Efficient Adaptation of Self-Supervised Speech Models for Low-Resource Dysarthric ASR

Zongyu Han (Swinburne University of Technology); Zhenwei Cao (Swinburne University of Technology); Yan Zhang (Hebei University of Technology)*; Qing Hu (Hebei University of Technology); Xiaoqing Zhang (Hebei University of Technology); Zetong Cai (Hebei University of Technology)

Sat-A-P-1-45: Rethinking End-to-End ASR Architectures for Dysarthric Speech: A Comparative Study of Conformer and Paraformer

Xinran Zhao (Jiangsu University of Science and Technology); Jingting Li (Institute of Psychology, Chinese Academy of Sciences); Su-Jing Wang (Chinese Academy of Sciences)*

Sat-A-P-1-46: Dual-Branch Encoding with Spectral Band Debiasing and Globally Guided Attention Fusion for Pathological Voice Detection

Yan Zhou (Beijing Institute of Technology, Zhuhai); Kenan Wang (Beijing Institute of Technology, Zhuhai); Enze Li (Beijing Institute of Technology); Kun Qian (Beijing Institute of Technology)*; Hengxin Liu (Beijing Children's Hospital); Bin Hu (Beijing Institute of Technology); Björn Schuller (Technische Universität München)

Sat-A-P-1-47: Pathology-Knowledge-Guided Structured Annotation for Clinical Speech Audio

Yangqing Wu (Shanghai Jiao Tong University)*; Mengyue Wu (Shanghai Jiao Tong University); Juqiang Chen (Shanghai Jiao Tong University)

6. Mental Health and Multimodal Medical Assessment (7 papers)

Sat-A-P-1-48: Multimodal Fusion with a Language-Guided Heatmap for Mild Cognitive Impairment Detection from Mandarin Speech

Jun Huang (Southern University of Science and Technology; Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences; Shenzhen University of Advanced Technology)*; Zhixing Guo (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Rongfeng Su (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Lan Wang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Nan Yan (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)

Sat-A-P-1-49: SAMR-Net: A Structure-Aware and Prompt-Guided Multimodal Representation Network for EEG-Speech Depression Recognition

Ru Gao (Taiyuan University of Technology)*; Zhe Li (University of Hong Kong); JiaNing Song (Taiyuan University of Technology); Meirong Song (Taiyuan University of Technology); Xin Li (Taiyuan University of Technology); Rong Yao (Taiyuan University of Technology); HaiFang Li (Taiyuan University of Technology)

Sat-A-P-1-50: WavLM-Guided Dual-Gated Multi-Level Acoustic Feature Fusion for Speech-Based Depression Detection

徐庆龙 (新疆大学)*; 黄志华 (新疆大学)

Sat-A-P-1-51: SoulEcho: An ISTDP-Driven Multi-Modal Conversational Agent System for Mental Health Assessment

Yixin Zhang (Beijing university of Posts and Telecommunications)*; Runnan Li (Beijing university of Posts and Telecommunications); Yaohua Bu (Beijing university of Posts and Telecommunications)

Sat-A-P-1-52: A Hybrid Acoustic Attention Framework with Periodic Decoupling and Wavelet Fusion for Cross-Lingual Alzheimer's Detection

Xian-Xian Liu (university of macau)*; Simon Fong (University of Macau)

Sat-A-P-1-53: Folk Expressions Mask Red Flags: Response Safety Evaluation for Chinese LLM Health Consultation

Xiao-Hang Jiang (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); Lei Jiang (iFLYTEK Co., Ltd.); Zhi-Yang He (iFLYTEK Co., Ltd.); Cong Liu (iFLYTEK Co., Ltd.)

Sat-A-P-1-54: DermClin-CVQA: An Auditable Multi-Task Benchmark for Dermatology MLLMs

Zi Yu Wang (College of Information Science and Engineering, Hohai University)*; Xinyu Li (Jiangsu Second Chinese Medicine Hospital, the Second Affiliated Hospital of Nanjing University of Chinese Medicine); Guang Zhi Wang (Nanjing Fuzhiduo AI Technology Co., Ltd.); Xue Wu Zhang (College of Information Science and Engineering, Hohai University); Yanchun Ma (Wuhan Vocational College of Software and Engineering)

7. Biomedical Acoustic Sensing (4 papers)

Sat-A-P-1-55: EdgeHS: Noise-Robust Heart Sound Abnormality Detection via Attention-Guided Denoising and an Ultra-Lightweight Embedded Convolutional Neural Network

Xinyuan He He (Beijing Institute of Technology, Zhuhai); Enze Li (Beijing Institute of Technology); Kun Qian (Beijing Institute of Technology)*; Jiancheng Han (Beijing Anzhen Hospital); Bin Hu (Beijing Institute of Technology); Björn Schuller (Technische Universität München)

Sat-A-P-1-56: Cycle-Consistent Neural Collapse for Long-Tailed Pediatric Heart Sound Classification

Li Gu (Beijing Institute of Technology); Enze Li (Beijing Institute of Technology); Kun Qian (Beijing Institute of Technology)*; Fan Lai (Chengdu Women's and Children's Central Hospital); Bin Hu (Beijing Institute of Technology); Björn Schuller (Technische Universität München)

Sat-A-P-1-57: Structure-Enhanced Representation Learning for Respiratory Sound Classification

Xiaoping Hou (Beijing Institute of Technology); Mengkai Sun (Beijing Institute of Technology); Kun Qian (Beijing Institute of Technology)*; Yan Liu (Peking University International Hospital); Bin Hu (Beijing Institute of Technology); Björn Schuller (Technische Universität München)

Sat-A-P-1-58: Cross-Device Fourier Domain Augmentation for Robust Respiratory Sound Classification under Stethoscope Recording Variation

Zhou He (Beijing Institute Of Technology)*; MengYu Zhu (Beijing Institute of Technology)

Sat-A-SS-1: 特殊议题 8: 共情计算与认知交互 II

Room: 沐辉厅 B

Type: Special Session

November 7, 2026, 1:30 PM to 3:00 PM

Organizer: 黄东延老师, 赵子平老师, 张子兴老师, 李雅老师, 刘斌老师

13:30-14:00

Sat-A-SS-1-1: Investigation of Empathy Modeling in Active Inference for Robotics

Dong-Yan Huang (深圳市智拓欣扬科技有限公司)*; Ding Wan (深圳市智拓欣扬科技有限公司); Claire Ma (深圳市智拓欣扬科技有限公司); Ruihua Ma (深圳市智拓欣扬科技有限公司)

Presenter: 黄东延老师, 深圳市智拓欣扬科技有限公司首席科学家

14:00-14:30

Sat-A-SS-1-2: Mitigating Modal Semantic Inconsistency via Uncertainty-Aware Soft Labels for Missing-Modality Emotion Recognition

Presenter: 张子兴老师, 湖南大学教授

14:30-15:00

Sat-A-SS-1-3: Swapped-Query Cascaded Fusion for Multimodal Emotion and Intent Joint Recognition

Presenter: 李雅老师, 北京邮电大学教授

Sat-A-SS-2: 特殊议题 2: 面向具身智能的空间听觉与人机交互 II

Room: 沐辉厅 C

Type: Special Session

November 7, 2026, 1:30 PM to 3:00 PM

Organizer: 朱梦尧老师, 杜俊老师, 卜辉老师, 薛浏蒙老师

13:30-13:45

Sat-A-SS-2-1: Latent-Based Diffusion Transformer for Ambisonic Sound Field Reconstruction

Yang Xiao (Northwestern Polytechnical University)*; Bing Zhu (Northwestern Polytechnical University); Yichen Yang (Northwestern Polytechnical University); Wen Zhang (Northwestern Polytechnical University)

Presenter: Yang Xiao (Northwestern Polytechnical University)

13:45-14:00

Sat-A-SS-2-2: No-Reference Evaluation of Simulation-Based Noisy Speech Restoration: An Empirical Study of Metric-Human Correlations

Chunyang Xu (Soochow University)*; Shichao Hu (Soochow University); Mengyao Zhu (Soochow University); Xin Xu (Beijing AISHELL Technology Co., Ltd.); Hui Bu (Beijing AISHELL Technology Co., Ltd.); Zhenhua Ling (University of Science and Technology of China)

Presenter: Chunyang Xu (Soochow University)

14:00-14:15

Sat-A-SS-2-3: ATF-Conditioned Neural Ambisonic Matrix Encoding for Arbitrary Microphone Arrays with Variable Microphone Count

Shichao Hu (Soochow University)*; Chunyang Xu (Soochow University); Mengyao Zhu (Soochow University)

Presenter: Shichao Hu (Soochow University)

Sat-A-SF-2: Student Forum II

Room: 沐辉厅 A

Type: Student Forum

November 7, 2026, 3:30 PM to 5:00 PM

Organizer: 洪青阳 (厦门大学), 连政 (同济大学), 孔秋强 (香港中文大学)

Sat-A-O-4: Audio Foundation Models and Sound Event Understanding

Room: 7 号会议室

Type: Oral

November 7, 3:30 PM to 5:15 PM

15:30-15:45

Sat-A-O-4-1: UniAE-MoE: A Unified Audio Encoder via Mixture of Experts

Shengbo Cai (Shenzhen International Graduate School, Tsinghua University)*; Zhisheng Zhang (Shenzhen International Graduate School, Tsinghua University); Zichao Nie (Shenzhen International Graduate School, Tsinghua University); Jing Peng (Shenzhen International Graduate School, Tsinghua University); Jingran Xie (Shenzhen International Graduate School, Tsinghua University); Zhiyong Wu (Shenzhen International Graduate School, Tsinghua University)

15:45-16:00

Sat-A-O-4-2: Understanding Semantic-Acoustic Complementarity in Audio Encoder Fusion under the Frozen Adaptation Paradigm

Zijian Lin (Tsinghua University); Dapeng Wu (Tsinghua University); Kaifeng Yun (Tsinghua University); Yikai Huang (Tsinghua University); Xiang Li (Tsinghua University); Jinjiang Liu (Tsinghua University); Zhiyong Wu (Tsinghua University)*

16:00-16:15

Sat-A-O-4-3: SKAD: Sketch-Guided Audio-Contrastive Decoding for Reasoning in Large Audio-Language Models

Junyu Wang (Tianjin University)*; Peiyuan Jiang (Tianjin University); Siyuan Zhang (Tianjin University); Jian Zong (Tianjin University); Tianrui Wang (Tianjin University); Zikang Huang (Tianjin University); Meng Ge (Tianjin University); Longbiao Wang (Tianjin University); Jianwu Dang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)

16:15-16:30

Sat-A-O-4-4: SURE-EVAL: A Systematic and Unified Agentic Framework for Reproducible Evaluation

Jing Peng (Shanghai Jiaotong University)*; Junhao Du (AISpeech Co., Ltd.); Yixuan Wang (AISpeech Co., Ltd.); Bowenn Wang (University of Bristol); hanqi li (Shanghai Jiao Tong University); Chaolei Liu (Shanghai Jiao Tong University); Weihai Chen (Shanghai Jiao Tong University); Haohui Xie (Shanghai Jiao Tong University); Ruichen Sun (Shanghai Jiao Tong University); Chenghao Wang (Shanghai Jiao Tong University); Wen Wen (Shanghai Jiao Tong University); Guanyu Chen (Shanghai Jiao Tong University); Xiaoyu Gu (Shanghai Jiao Tong University); Haoyu Li (Shanghai Jiao Tong University); Yiwei Guo (Shanghai Jiao Tong University); Bohan Li (Shanghai Jiao Tong University); Tao Liu (Shanghai Jiao Tong University); Yucheng Wang (ETH Zurich); Yu Xi (Shanghai Jiao Tong University); Yihua Zhou (AISpeech Co., Ltd.); Qiang Zhou (AISpeech Co., Ltd.); Feng Lu (AISpeech Co., Ltd.); Shuai Wang (Nan Jing University); Kai Yu (Shanghai Jiao Tong University)

16:30-16:45

Sat-A-O-4-5: RADSA: Range-Aware Dual-Specialist Adaptation for Pre-trained Sound Event Localization and Detection

Lele Si (Dalian Minzu University); Rong Wang (Dalian Minzu University)*; Bingbing Zhang (Dalian Minzu University); Weiwei Zhang (Dalian Maritime University); Ruifang Wang (Suzhou City University)

16:45-17:00

Sat-A-O-4-6: MCCL-SED: Multi-Constraint Consistency Learning for Polyphonic Sound Event Detection

Ying Hu (Xinjiang University)*; Jiulong Guo (Xinjiang University); Lurong Liang (Xinjiang University)

17:00-17:15

Sat-A-O-4-7: Diffusion-Based Data Augmentation with Noise-Trajectory Mixing for Low-Resource Bioacoustic Event Classification

Biao Liu (Guangxi University); Wei Xie (Guangxi University)*; Wang Dong (Guangxi University); Xilu Wang (University of Surrey)

Sat-A-O-5: Target Speaker Extraction and Audio-Visual Speech Enhancement

Room: 8M 会议室

Type: Oral

November 7, 3:30 PM to 5:00 PM

15:30-15:45

Sat-A-O-5-1: Visually Tracked Dynamic Spatial Queries for Multichannel Target Speech Extraction of Moving Talkers

Zhengqing Niu (Tianjin University)*; Longbiao Wang (Tianjin University)

15:45-16:00

Sat-A-O-5-2: Spatially Aware Audio-Visual Target Speaker Extraction via Multi-Cue Neural Fusion

Ze Zhang (Inner Mongolia University)*; Jinglin Bai (RDG Group, iFlytek Research); Feng Cheng (RDG Group, iFlytek Research); Jiahui Pan (Inner Mongolia University); Xueliang Zhang (Inner Mongolia University)

16:00-16:15

Sat-A-O-5-3: A Lightweight Dual-Branch Visual Front-End for Efficient Audio-Visual Target Speaker Extraction

Zhan Jin (Wuhan University)*; Peijun Yang (Wuhan University); Juan Liu (Wuhan University); Ming Li (The Chinese University of Hong Kong, Shenzhen)

16:15-16:30

Sat-A-O-5-4: Robust Audio-Visual Target Speaker Extraction with Multiple Enrollment Fusion

Zhan Jin (Wuhan University)*; Bang Zeng (Wuhan University); Peijun Yang (Wuhan University); Jiarong Du (Wuhan University); Juan Liu (Wuhan University); Ming Li (The Chinese University of Hong Kong, Shenzhen)

16:30-16:45

Sat-A-O-5-5: Audio-Visual Speech Enhancement in Complex Scenarios with Separation and Dereverberation Joint Modeling

Zhan Jin (Wuhan University)*; Jiarong Du (Wuhan University); Peijun Yang (Wuhan University); Juan Liu (Wuhan University); Ming Li (The Chinese University of Hong Kong, Shenzhen)

16:45-17:00

Sat-A-O-5-6: Neural-Guided Target Speaker Extraction from Mixed Speech via Multi-Level Multi-Domain Fusion

Mengze Du (Tianjin University); Zhuang Xie (Tianjin University); Sijia Song (Tianjin University); Gaoyan Zhang (Tianjin University)*

Sat-A-O-6: Speech Enhancement, Restoration and Noise Control**Room: 8F 会议室****Type: Oral****November 7, 3:30 PM to 5:00 PM**

15:30-15:45

Sat-A-O-6-1: Loud-loss: A Perceptually Motivated Loss Function for Speech Enhancement Based on Equal-Loudness Contours

Zixuan Li (Inner Mongolia University)*; Xueliang Zhang (Inner Mongolia University); Changjiang Zhao (Inner Mongolia University); Shuai Gao (Inner Mongolia University); Lei Miao (Lenovo); Zhipeng Yan (Lenovo); Ying Sun (Lenovo); Chong Zhu (Lenovo)

15:45-16:00

Sat-A-O-6-2: SCSB: Sampler-Calibrated Schrödinger Bridge for Generative Speech Enhancement

Huate Zhu (Jiangsu University); Qing Yao (Jiangsu University); Lijian Gao (Jiangsu University); Xun Zhang (Jiangsu University); Qirong Mao (Jiangsu University)*

16:00-16:15

Sat-A-O-6-3: LGDK-Net: Linear-Gated Dynamic Kernel Compensation for Low-SNR Speech Enhancement

Bixin Wu (Xinjiang University)*; Zhihua Huang (Xinjiang University); Wenjie Su (Xinjiang University); Bowen Li (Xinjiang University)

16:15-16:30

Sat-A-O-6-4: A Cross-Gated 2D Auxiliary Branch for Lightweight Dual-Channel Speech Enhancement

Jiatao Yan (Hangzhou Dianzi University)*; Wenbin Jiang (Hangzhou Dianzi University); Xiang Ye (Hangzhou Dianzi University); Jingrui Qian (Hangzhou Dianzi University); Yang Zhang (Hangzhou Dianzi University)

16:30-16:45

Sat-A-O-6-5: SemMamba-GAN: Semantically Constrained Bone-to-Air Speech Restoration

Zhiqiang Duan (Anhui University)*

16:45-17:00

Sat-A-O-6-6: From Reference-Only to Primary-Path-Aware Neural Active Noise Control

Fei Zhao (School of Computer Science, Inner Mongolia University)*; Xueliang Zhang (School of Computer Science, Inner Mongolia University)

Sat-A-P-2: Speech Recognition, Audio Understanding and Multimodal Interaction**Room: 2D 会议室 + 2R 会议室 (海报区)****Type: Poster****November 7, 3:30 PM to 5:00 PM****1. Multilingual and Dialect Speech Recognition (10 papers)****Sat-A-P-2-1: A Language-Agnostic Hierarchical LoRA-MoE Architecture for CTC-based Multilingual ASR**

yuang zheng (Shanghai Normal University); dongxu chen (Shangrao Normal University); yuxiang mei (Shanghai Normal University); dongxing xu (Unisound AI Technology Co.); jie chen (Shanghai Normal University); yanhua long (Shanghai Normal University)*

Sat-A-P-2-2: Advancing LLM-based phoneme-to-grapheme for multilingual speech recognition

Lukuang Dong (TASI Technology); Ziwei Li (TASI Technology); Saierdaer Yusuyin (Xinjiang University); Xianyu Zhao (TASI Technology); Zhijian Ou (Tsinghua University)*

Sat-A-P-2-3: An Empirical Study of Parameter-Efficient Fine-Tuning for Low-Resource Cross-Lingual ASR: Capacity, Script Stability, and Evaluation Bias

李雨洁 (西南科技大学)*; 马立平 (西南科技大学)

Sat-A-P-2-4: Parameter-Efficient Fine-Tuning of Whisper for Low-Resource Amdo Tibetan Speech Recognition: A Capacity-Centric Study

杨热杨 (青海民族大学)*; 安见才让安见 (青海民族大学); 茸西娜姆茸西 (青海民族大学)

Sat-A-P-2-5: 面向低资源藏汉语音翻译的预训练模型迁移与数据增强方法研究

范佳雪 (中央民族大学)*; 朱丽平 (中央民族大学)

Sat-A-P-2-6: 中英闽多语种流式识别与低资源闽南语建模

魏志鹏 (厦门大学电子科学与技术学院)*; 洪青阳 (厦门大学信息学院); 李琳 (厦门大学电子科学与技术学院)

Sat-A-P-2-7: Analysis of acoustic feature differences among multiple Chinese dialects and their impacts on automatic speech recognition performance

Yuan Fang (Zhongnan University of Economics and Law); Jingzhe Li (Zhongnan University of Economics and Law)*; Wenxu Jiang (Zhongnan University of Economics and Law); Yuhao Zhang (Zhongnan University of Economics and Law); Yixiao Su (Zhongnan University of Economics and Law)

Sat-A-P-2-8: Decoding Tonal Mergers: High-Frequency Spectral Cues Drive Robust Tone Classification in the Laiwu Dialect

Haizheng Li (Beijing Normal University)*; Gaowu Wang (Beijing Normal University)

Sat-A-P-2-9: 基于 SenseVoice 的低资源方言辨识方法

詹国伟 (安徽大学)*; 曹宜策 (安徽大学); 徐金荣 (太原工业学院); 申娟娟 (沈阳建筑大学); 李东泽 (河北大学)

Sat-A-P-2-10: Seeing Dialect Beyond Labels: Coupled Generation for Chinese Dialect Identification and ASR

Shihong Li (University of Electronic Science and Technology of China); linrong pan (tencent); Juntao Xu (University of Electronic Science and Technology of China); Zhiyong Feng (Harbin Institute of Technology); Siqi Yang (University of Electronic Science and Technology of China)*

2. Robust Recognition, Adaptation and Language Processing (10 papers)

Sat-A-P-2-11: TSF-KWM: Structure-Aware and Reliability-Calibrated Embedding for Babble-Robust Keyword Mamba

HongZhen Li (Shenzhen University)*; Xueliang Li (Shenzhen University)

Sat-A-P-2-12: First-order Cross-Modal Meta-Transfer for Cross-Speaker Silent Speech Recognition

Minghao Guo (Tianjin University)*; Qiang Fang (Tianjin University); Han Liu (Tianjin University); Wenhuan Lu (Tianjin University); Jianguo Wei (Tianjin University)

Sat-A-P-2-13: From Implicit to Explicit: A Two-Stage Evidence-Driven Framework for LLM-based ASR Error Correction

Longwei Li (Xinjiang University)*; Liang He (Tsinghua University); Xiang He (Xinjiang University); Zhida Song (Xinjiang University); Jie Liu (Xinjiang University)

Sat-A-P-2-14: When Context Hurts: Selective Context Adaptation for On-device LLM Translation

Changqing Kong (USTC); Hang Chen (USTC)*; Ting Qi (iFlytek Research); Yanan Song (USTC); Henshun Zhou (iFlytek Research); Qing Wang (USTC); Jun Du (USTC); Jianqing Gao (iFlytek Research); Yu Hu (USTC)

Sat-A-P-2-15: 基于全模态大语言模型与群组相对策略优化的多模态幻灯片辅助自动语音识别

Boyi Yao (School of Software, Northwestern Polytechnical University)*; Jingxuan Zhang (School of Artificial Intelligence and Computer Science, Shaanxi Normal University); Zekun Zhang (iFLYTEK Research, iFLYTEK Co., Ltd.); Yeran Yang (School of Artificial Intelligence and Computer Science, Shaanxi Normal University); Genshun Wan (iFLYTEK Research, iFLYTEK Co., Ltd.); Jianqing Gao (iFLYTEK Research, iFLYTEK Co., Ltd.); Hongbo Ni (School of Computer Science, Northwestern Polytechnical University); Zhonghua Fu (School of Computer Science, Northwestern Polytechnical University; iFLYTEK Research, iFLYTEK Co., Ltd.)

Sat-A-P-2-16: TOASR: Breaking the Data Barrier in LLM-Based ASR via Text-Only Training

Qing Zhao (Megatronix(Beijing) technology Co., Ltd)*; Hongwei Cai (Megatronix(Beijing) technology Co., Ltd); Wei Liang (Megatronix(Beijing) technology Co., Ltd)

Sat-A-P-2-17: Tokenizer-Preserving Phonological Side Channels for Mandarin Error Recovery in Large Language Models

李亭皓 (中国科学技术大学)*

Sat-A-P-2-18: X-ASR: Optimizing Lightweight and Streaming Open-Source ASR Model on Million-Hour Speech Data

Jun Guo (Huazhong University of Science and Technology)*; Chonghao Cai (Shanghai Jiao Tong University); Yifan Yang (Shanghai Jiao Tong University); Zhikang Niu (Shanghai Jiao Tong University); Yujie Tu (University of Chinese Academy of Sciences); Xingjian Zhao (Fudan University); Qinyuan Cheng (Fudan University); Xipeng Qiu (Fudan University); Lianbo Guo (Huazhong University of Science and Technology); Zhipeng Tan (Huazhong University of Science and Technology); Xie Chen (Shanghai Jiao Tong University)

Sat-A-P-2-19: Marco-ASR: A Principled and Metric-Driven Framework for Fine-Tuning Large-Scale ASR Models for Domain Adaptation

Xuanfan Ni (Alibaba Group)*; Fei Yang (Alibaba Group); Fengping Tian (Alibaba Group); Qingjuan Li (Alibaba Group); Yuejiao Wang (Alibaba Group); Yiming Yao (Alibaba Group); Yichao Du (Alibaba Group); Chenyang Lyu (Alibaba Group); Longyue Wang (Alibaba Group); Weihua Luo (Alibaba Group)

Sat-A-P-2-20: CTC Blank-Guided Adaptive Temporal Compression for Efficient End-to-End Speech Recognition

Siyuan Wang (哈尔滨工程大学)*; Chen Xu (哈尔滨工程大学)

3. Speech Resources, Evaluation and Spoken Language Understanding (6 papers)

Sat-A-P-2-21: SCORE: Semantic Consequence-Oriented Recognition Evaluation for Context-Aware ASR

Xiaoyu Gu (Shanghai Jiao Tong University)*; Chencheng Wang (Shanghai Jiao Tong University); Zonghan Zhou (Shanghai Jiao Tong University); Wen Wen (Shanghai Jiao Tong University); FuYang Li (Beijing University of Posts and Telecommunications); Yu Xi (Shanghai Jiao Tong University); Jing Peng (Shanghai Jiao Tong University); Guanyu Chen (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sat-A-P-2-22: ATCCaps: A Call-Sign-Aware Speech Dataset for Air Traffic Control Recognition

Dongdong Li (East China University of Science and Technology)*; Jianwei Song (East China University of Science and Technology); Jianwei Wang (East China University of Science and Technology); Zhe Wang (East China University of Science and Technology)

Sat-A-P-2-23: When ASR Meets the Factory Floor: Ontology-Grounded Terminology Correction and Evidence-Tiered Abstention for Trustworthy Spoken Question Answering over Enterprise Data

Jinze YU (Waseda University)*

Sat-A-P-2-24: Agentic Data Sampling Framework for Multilingual ASR via Linguistic-Aware Error-Driven Optimization

Zhiheng Qian (Shanghai Jiao Tong University)*; Junhao Du (AISpeech Co., Ltd.); Jing Peng (Shanghai Jiao Tong University); Hanqi Li (Shanghai Jiao Tong University); Kunyang Peng (AISpeech Co., Ltd.); Feng Xue (AISpeech Co., Ltd.); Lu Chen (Shanghai Jiao Tong University); Shuai Wang (Nanjing University); Kai Yu (Shanghai Jiao Tong University)

Sat-A-P-2-25: Reference-Free ASR Hypothesis Evaluation with Acoustic Evidence and Linguistic Cues

Zhihan Li (Shanghai Jiao Tong University)*; Hankun Wang (Shanghai Jiao Tong University); Yiwei Guo (Shanghai Jiao Tong University); Bohan Li (Shanghai Jiao Tong University); Xie Chen (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sat-A-P-2-26: CM-OTD: Towards ASR-Robust Spoken Language Understanding via Contrastive Multiview and Optimal Transport Distance

Cao YiFan (Jiangnan University)*; Yu Zihao (Jiangnan University); Sun Jun (Jiangnan University)

4. Audio-Language Reasoning, Grounding and Retrieval (11 papers)

Sat-A-P-2-27: OMNI-CLST: Error-Aware Curriculum Learning With Guided Selective Chain-of-Thought For Audio Question Answering

Jinghua Zhao (Nankai University)*; Hang Su (MiLM Plus, Xiaomi Inc.); Lichun Fan (MiLM Plus, Xiaomi Inc.); Zhenbo Luo (MiLM Plus, Xiaomi Inc.); Hui Wang (Nankai University); Haoqin Sun (Nankai University); Yong Qin (Nankai University)

Sat-A-P-2-28: PGER-RAG: Progressive Generation-Oriented Evidence Reranking for Retrieval-Augmented Generation

jun zheng (School of Computer Science and Engineering Northeastern University)*; yubin bao (School of Computer Science and Engineering Northeastern University); fangling leng (School of Computer Science and Engineering Northeastern University); tiancheng zhang (School of Computer Science and Engineering Northeastern University); ge yu (School of Computer Science and Engineering Northeastern University)

Sat-A-P-2-29: Depth-Wise Reasoning Trajectory Alignment for Extending Large Language Models to Speech

Weidong Chen (The Chinese University of Hong Kong); Jiankun Zhao (The Chinese University of Hong Kong); Yuanyuan Wang (The Chinese University of Hong Kong); Dongrui Han (The Chinese University of Hong Kong); Rongjie Huang (The Chinese University of Hong Kong); Helen Meng (The Chinese University of Hong Kong); Xixin Wu (The Chinese University of Hong Kong)*

Sat-A-P-2-30: NVV-Locator: From Transcript Tags to Acoustic Boundaries for Fine-Grained Nonverbal Vocalization Grounding

Yuang Cao (Northwestern Polytechnical University)*; Bingshen Mu (Northwestern Polytechnical University); Zhennan Lin (Northwestern Polytechnical University); Guojian Li (Northwestern Polytechnical University); Haoyue Zhan (Shanghai Lingguang Zhaxian Technology); Jie Liu (Shanghai Lingguang Zhaxian Technology); Chuan Xie (Shanghai Lingguang Zhaxian Technology); Qiang Zhang (Shanghai Lingguang Zhaxian Technology); Liumeng Xue (Nanjing University); Lei Xie (Northwestern Polytechnical University)

Sat-A-P-2-31: SHARP: Headroom-Aware Post-SFT Refinement for Audio Question Answering

Tengfei Dong (Xinjiang University)*; Zhe Li (The University of Hong Kong); Zunwang Ke (Xinjiang University)

Sat-A-P-2-32: OSUM-Pangu: An Open-Source Multidimensional Speech Understanding Foundation Model Built on OpenPangu for Ascend NPUs

Yujie Liao (Northwestern Polytechnical University)*; Xuelong Geng (Northwestern Polytechnical University); Hongfei Xue (Northwestern Polytechnical University); Shuiyuan Wang (Northwestern Polytechnical University); Lei Xie (Northwestern Polytechnical University)

Sat-A-P-2-33: REAP: Reasoning-Oriented Audio Evidence Planning and Alignment for Complex Audio Reasoning
Peiyuan Jiang (Tianjin University)*; Siyuan Zhang (Tianjin University); Junyu Wang (Tianjin University); Tianrui Wang (Tianjin University); Jiahao Yan (Tianjin University); Jian Zong (Tianjin University); Longbiao Wang (Tianjin University); Jianwu Dang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)

Sat-A-P-2-34: Omni-AutoThink: Adaptive Reasoning for Omni-Modal Understanding
Dongchao Yang (The Chinese University of Hong Kong)*; Songxiang Liu (Meituan); Disong Wang (Meituan); Yuanyuan Wang (The Chinese University of Hongkong); Guanglu Wan (Meituan); Helen Meng (The Chinese University of Hongkong)

Sat-A-P-2-35: Generative-Discriminative Feature Complementarity for Enhancing Audio Understanding
Zeyu Xie (Peking University)*; Chenxing Li (Tencent); Xuenan Xu (SJTU); Mengyue Wu (SJTU); Wenfu Wang (Tencent); Ruibo Fu (Tencent); Meng Yu (Tencent); Dong Yu (Tencent); Yuexian Zou (Peking University)

Sat-A-P-2-36: Spatial Audio-Language Model with Structured Embeddings for Understanding and Editing
Jinbo Hu (Institute of Acoustics, Chinese Academy of Sciences)*; Yin Cao (Institute of Acoustics, Chinese Academy of Sciences); Ming Wu (Institute of Acoustics, Chinese Academy of Sciences); Zhenbo Luo (Xiaomi Inc.); Jun Yang (Institute of Acoustics, Chinese Academy of Sciences)

Sat-A-P-2-37: Boundary Quality Calibration for Candidate Ranking in Audio Moment Retrieval
Xiaokai Zhang (Xi'an Jiaotong-Liverpool University); Shengchen Li (Xi'an Jiaotong-Liverpool University)*

5. Audio and Audio-Visual Scene Understanding (4 papers)

Sat-A-P-2-38: DyFoST: Dynamic Foreground-aware Spectro-Temporal Learning for Environmental Sound Classification
Jiaqi Li (Zhengzhou University)*

Sat-A-P-2-39: Domain-specific Adapters and Routing Calibration for Domain-Incremental Audio Classification
李潜潜 (华南理工大学)*; Yanxiong Li (South China University of Technology); Guoqing Chen (South China University of Technology)

Sat-A-P-2-40: HyCE: Multi-Curvature Hyperbolic Alignment for Weakly-Supervised Audio-Visual Video Parsing
jiajun xu (University of Technology Sydney)*; Heyan Huang (Beijing Institute of Technology); xuefeng chen (Inkeverse group limited); Yanghao Zhou (Beijing Institute of Technology)

Sat-A-P-2-41: DetectCap: Neural Granger Temporal Dependency Modeling with MultiStage Narrative Feedback
tao ma (Inner Mongolia University); haolin zuo (Inner Mongolia University); zhenqi jia (Inner Mongolia University); rui liu (Inner Mongolia University)*

6. Emotion, Sentiment and Intent Understanding (12 papers)

Sat-A-P-2-42: RDGRN: A Reaction-Diffusion Graph Reasoning Network for Multimodal Emotion Recognition in Conversation

Xie Jiajun (Guangdong University of Technology); Huang Guoheng (Guangdong University of Technology)*; Xie Siyue (Guangdong University of Technology); Chen Zhuoheng (Guangdong University of Technology); Yuan Xiaochen (Macao Polytechnic University); Li Yan (Shenzhen Polytechnic University); Wang Shuqiang (Chinese Academy of Sciences); Tsang Kim-Fung (Chinese Academy of Sciences); Pun Chi-Man (University of Macau)

Sat-A-P-2-43: A Multi-Task Transfer Learning Framework for Efficient Speech Emotion Recognition

Junhui Li (University of Science and Technology of China)*; Yan Song (University of Science and Technology of China); Tinggan Huang (University of Science and Technology of China); Haoyu Song (Singapore Institute of Technology); Ian McLoughlin (Singapore Institute of Technology)

Sat-A-P-2-44: Adaptive Gating-Based Reliability-Aware Fusion for Multimodal Sentiment Analysis

yu hongxuan (Shandong University of Finance and Economics)*; zhang yunfeng (Shandong University of Finance and Economics); qiu han (Shandong University)

Sat-A-P-2-45: Semantic Reconstruction and Quality-Aware Fusion for Missing-Modality Multimodal Emotion Recognition in Conversations

Kaidong Wang (Shanghai University Of Engineering Science)*; Xin Yuan (Shanghai University of Engineering Science)

Sat-A-P-2-46: An Affective-Style Concept Bottleneck for Speaker-Independent Speech Emotion Recognition

Vladimir Vladimir (Beijing Institute of Technology); Lin Shen (Beijing Institute of Technology); Kun Qian (Beijing Institute of Technology)*; Xinyuan Qian (University of Science and Technology Beijing); Bin Hu (Beijing Institute of Technology); Björn Schuller (Technische Universität München)

Sat-A-P-2-47: Let Emotion Guide the Gate: Top-down Affective Modulation for Speech Emotion Recognition

Rui Wang (Beijing Institute of Technology, Zhuhai); Ziyang Zhang (Northeast Agricultural University); Chenyang Li (Beijing Institute of Technology, Zhuhai); Kun Qian (Beijing Institute of Technology)*; Bin Hu (Beijing Institute of Technology); Björn Schuller (Technische Universität München)

Sat-A-P-2-48: 中文大语言模型副语言能力评测体系构建

Bin Zhao (Institute of Linguistics, Chinese Academy of Social Sciences); Mengxue Cao (Institute of Linguistics, Chinese Academy of Social Sciences); Aijun Li (Institute of Linguistics, Chinese Academy of Social Science)*

Sat-A-P-2-49: AR Pole-Based Representations for Chinese Speech Emotion Recognition under Dual-Disjoint Evaluation

Bo CHAI (The Hong Kong Polytechnic University)*; Ruwen ZHANG (The Hong Kong Polytechnic University); Ran TAO (The Hong Kong Polytechnic University); Nizhuan WANG (The Hong Kong Polytechnic University); Wai Ting SIOK (The Hong Kong Polytechnic University)

Sat-A-P-2-50: Mitigating Modal Semantic Inconsistency via Uncertainty-Aware Soft Labels for Missing-Modality Emotion Recognition

zixing zhang (Hunan University)*; tengda yin (Hunan University); jing han (Hunan University)

Sat-A-P-2-51: Swapped-Query Cascaded Fusion for Multimodal Emotion and Intent Joint Recognition

Junxing Li (Beijing University of Posts and Telecommunications)*; Ya Li (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications)

Sat-A-P-2-52: TransDis-MMSA: A Mongolian Multimodal Sentiment Analysis Model Based on Modality Translation and Hierarchical Disentanglement

Qing-dao-er-ji Ren (College of Intelligent Science and Technology, Inner Mongolia University of Technology)*; Lulu Dai (College of Intelligent Science and Technology, Inner Mongolia University of Technology); Junzu Zhang (College of Intelligent Science and Technology, Inner Mongolia University of Technology)

Sat-A-P-2-53: Emotion Recognition in Speech via Learning on Self-Estimated Labels Using MoE-Based Complementary-Label Discriminance

ziping zhao (Tianjin Normal University)*; xinzhou xu (Nanjing University of Posts and Telecommunications); yandi zheng (Tianjin Normal University)

7. Spoken Dialogue, Robotics and Neural Semantic Decoding (5 papers)

Sat-A-P-2-54: Beyond Parallel Tracking: Interactive Multi-Feature Fusion Drives Isomorphic Semantic Reconstruction from Non-invasive Brain Recordings

Boda Xiao (Peking University); Xiran Xu (peking university); Songyi Li (peking university); Yujie Yan (peking university); Xihong Wu (peking university); Heping Cheng (peking university); Jing Chen (peking university)*

Sat-A-P-2-55: Audio large language models align with human neural responses during spoken-language comprehension

shiyao zhu (Beijing Language and Culture University)*; Yanlu Xie (Beijing Language and Culture University)

Sat-A-P-2-56: Multi-Speaker Intelligent Voice Dialogue Robot: Architecture Design and Engineering Implementation

kong shuoyu (city university of macau)*

Sat-A-P-2-57: Structured Multi-Dimensional Evaluation for Speech-to-Speech Spoken Dialogue with Specialized Large Audio Models

ruopan wang (Information Engineering University)*; Hao Zhang (Information Engineering University); Di Lu (Zhengzhou College of Finance and Economics); Dan Qu (Information Engineering University)

Sat-A-P-2-58: EEG Time Series Prediction and Its Application in Swarm Robot Cognition and Interaction

甘玉玺 (清华大学)*; Zengkang Gan (Institute of Intelligent Manufacturing, Shenzhen Polytechnic University, Shenzhen, China); Xiao Chen (School of Artificial Intelligence, Shenzhen Polytechnic University, Shenzhen, China); Changyun Ge (School of Big Data and Artificial Intelligence, Fuyang University of Technology, Fuyang, China); Junyu Xie (R&D Department, Shenzhen Abby Bionic Robot Technology Co., Ltd., Shenzhen, China); Haoyang Fu (Dalian Neusoft University of Information, Dalian 116023, China); Gang Huang (R&D Department, Shenzhen Abby Bionic Robot Technology Co., Ltd., Shenzhen, China); Mingyu Xu (R&D Department, Weikang (Shenzhen) Intelligence Co., Ltd, Shenzhen, China); Zhihu Jin (R&D Department, Shenzhen Jinzhi Technology Co., Ltd., Shenzhen, China)

8. NLPCC Papers (11 papers)

Sat-A-P-2-59: How Little Poisoning Is Enough? Backdoor Vulnerabilities of Low-Resource Large Language Models

董时维 (西藏大学)

Sat-A-P-2-60: Zero-shot Cross-domain Slot Filling via Domain-aware Contrastive Prompting with Post-hoc Entity Refining

彭尧 (河北工程大学)

Sat-A-P-2-61: LEAP: LLM-Guided Experience Augmentation with MCTS Planning for Task-Oriented Dialogue Policy Learning

谢任超 (长沙理工大学)

Sat-A-P-2-62: Quantization-Robust LLM Unlearning via Module-Sensitive Updates

黄嘉兴 (武汉大学)

Sat-A-P-2-63: Listen Before You Look: Joint Visual-Audio Mixture of Retrieval Experts for Video Comprehension

薛海威 (香港科技大学)

Sat-A-P-2-64: BVPA: Branch-Verified Preference Augmentation for Long-Horizon LLM Agents

王越余 (国防科技大学)

Sat-A-P-2-65: VeriCite: Type-Aware Verification and Revision for Faithful Citations in Scientific Survey Generation
张崇发（国防科技大学）

Sat-A-P-2-66: SCORE: Towards Robust Situational Cognitive Reasoning in LLMs
王怡睿（南开大学）

Sat-A-P-2-67: LLM-RaF: Decision-Decoupled Rationale Fusion for Social-Context Fake News Detection
张嘉琦（国际关系学院）

Sat-A-P-2-68: SEA: A Structured Evidence Adapter for LLM-Augmented Dialogue Relation Extraction
傅湘婷（吉林大学）

Sat-A-P-2-69: A Fine-Grained Trustworthiness Assessment of LLM Safety Benchmark Results
曾锴秦（华东师范大学）

Sat-A-SS-3: 特殊议题 5: 残障研究生视角下的无障碍交互与病理语音智能处理
Room: 沐辉厅 B
Type: Special Session
November 7, 2026, 3:30 PM to 5:00 PM
Organizer: 王甦菁老师, 吴锡欣老师, 袁鑫老师, 燕楠老师, 于仕琪老师, 陈莹老师, 姚登峰老师, 胡雯老师

15:30-15:45
Sat-A-SS-3-1: 以结构化合约驱动的学术论文演讲视频自动生成
Presenter: 吴竞琦, 南方科技大学、中关村学院博士生

15:45-16:00
Sat-A-SS-3-2: 面向无障碍交互的构音障碍语音智能识别方法研究
Presenter: 康新晨, 南京理工大学计算机科学与工程博士生

16:00-16:15
Sat-A-SS-3-3: 面向肿瘤坏消息沟通训练的语音感知虚拟患者系统
Presenter: 王星淇, 清华大学博士生

16:15-16:30
Sat-A-SS-3-4: 面向听障群体的信息无障碍结构化字幕生成评测
Presenter: 施杰, 北京联合大学机器人学院硕士生

16:30-16:45
Sat-A-SS-3-5: 批评隐喻视角下 DeepSeek 主题英汉新闻话语的中国科技形象建构对比研究
Presenter: 黄清河, 福建师范大学硕士生

Sat-A-SS-4: 特殊议题 6: 面向生成式音频的可信音频处理方法
Room: 沐辉厅 C
Type: Special Session
November 7, 2026, 3:30 PM to 5:00 PM
Organizer: 刘学琛老师, 李圣辰老师, 朱美能老师

15:30-15:45
Sat-A-SS-4-1: 通过鲁棒说话人表征学习实现最大化说话人可分性
Presenter: 李哲, 香港大学博士后研究员

15:45-16:00

Sat-A-SS-4-2: 基于真伪语音差异建模的深伪语音检测研究

Presenter: 李梦璐，清华大学深圳国际研究生院博士后研究员

16:00-16:15

Sat-A-SS-4-3: 基于去噪扩散概率模型的匿名说话人韵律-情感修护方法研究

Presenter: 王婷婷老师，南京邮电大学讲师

16:15-16:30

Sat-A-SS-4-4: 面向音频生成模型的成员推断攻击：从扩散内部信号到结构匹配校准

Presenter: 刘宇轩，西交利物浦大学博士研究生

16:30-17:00

Sat-A-SS-4-5: 圆桌讨论

Sat-A-PLEN-1: 四十年从理想到现实

Room: 沐辉厅 A+B+C

Type: Plenary

November 7, 2026, 5:20 PM to 6:05 PM

Presenter: 颜永红老师

Sat-E-FUND: 基金分享沙龙

Room: 沐辉厅 A+B+C

Type: Fundings Session

November 7, 2026, 8:30 PM to 10:30 PM

Sunday, November 8

Sun-R: 会议报到
Room: 会议室前厅
Type: Registration
November 8, 2026, 8:00 AM to 6:00 PM

Sun-EXH: 赞助商展览
Room: 会议室前厅
Type: Exhibition
November 8, 2026, 9:00 AM to 6:00 PM

Sun-M-PLN-1: YuE2: Bringing the Human into the Loop of Symbiotic Creativity in Music
Room: 沐辉厅 A+B+C
Type: Plenary
November 8, 2026, 9:00 AM to 9:45 AM
Presenter: 郭毅可老师

Sun-M-PLN-2: 迈向全模态交互模型
Room: 沐辉厅 A+B+C
Type: Plenary
November 8, 2026, 9:45 AM to 10:30 AM
Presenter: 邱锡鹏老师

Sun-M-O-1: Spotlight Paper Presentation I
Room: 7 号会议室
Type: Oral
November 8, 11:00 AM to 12:30 PM

11:00-11:15

Sun-M-O-1-1: Listener-Judgment Evidence of Dialect Change: Urbanization of the Baoshan Variety in Shanghai Dialects

Tongxi Chen (Lancaster University)*; Siqi He (University of Massachusetts); Yinuo Wang (Southeast University); Yanyu Lu (Technical University of Denmark); Yu Chen (Southeast University); Weizhi Meng (Lancaster University)

11:15-11:30

Sun-M-O-1-2: Towards Fine-Grained Multi-Dimensional Speech Understanding: Data Pipeline, Benchmark, and Model

Guojian Li (Northwestern Polytechnical University)*; Zhixian Zhao (Northwestern Polytechnical University); Zhennan Lin (Northwestern Polytechnical University); Jingbin Hu (Northwestern Polytechnical University); Qirui Zhan (Northwestern Polytechnical University); Yang Cao (Northwestern Polytechnical University); Pengyuan Xie (Shanghai Lingguang Zhaxian Technology); Chuan Xie (Shanghai Lingguang Zhaxian Technology); Jie Liu (Shanghai Lingguang Zhaxian Technology); Qiang Zhang (Shanghai Lingguang Zhaxian Technology); Zhonghua Fu (Northwestern Polytechnical University); Lei Xie (Northwestern Polytechnical University)

11:30-11:45

Sun-M-O-1-3: Multilingual TASU: Extending Text-Only Alignment to Multilingual Speech Recognition via Simulated CTC Posteriors

Chenghao Wang (AISpeech Ltd)*; Yi Yang (X-LANCE Lab, School of Computer Science, Shanghai Jiao Tong University); Jing Peng (X-LANCE Lab, School of Computer Science, Shanghai Jiao Tong University); Yu Xi (X-LANCE Lab, School of Computer Science, Shanghai Jiao Tong University); Shuai Wang (Nanjing University); Kai Yu (X-LANCE Lab, School of Computer Science, Shanghai Jiao Tong University)

11:45-12:00

Sun-M-O-1-4: Towards Robust Chinese Dialect Speech Recognition via Systematic Data-Centric Learning

Qing Wang (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd)*; Boqing Liu (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Guangmin Xia (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Mengjie Du (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Xuechao Zhu (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Chao Wang (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Yiming Zhou (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Yan Yang (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Hang Lv (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Zehan Li (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Hongjie Chen (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Jian Kang (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Jie Li (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd); Yongxiang Li (Xingchen AGI Lab, China Telecom Artificial Intelligence Technology Co. Ltd)

12:00-12:15

Sun-M-O-1-5: Generalizable Feature Learning and Adaptive Prototype Expansion for Few-Shot Domain-Incremental Audio Deepfake Detection

Wang Dong (Guangxi University); Wei Xie (Guangxi University)*; Xilu Wang (University of Surrey); Wenwu Wang (University of Surrey)

12:15-12:30

Sun-M-O-1-6: PA-TSE: Presence-Aware Target Speaker Extraction for Complex Sparse-Overlap Conversations

Guanyu Chen (Shanghai Jiao Tong University)*; Haoyu Li (Shanghai Jiao Tong University); Yu Xi (Shanghai Jiao Tong University); Jing Peng (Shanghai Jiao Tong University); Wen Wen (Shanghai Jiao Tong University); Xiaoyu Gu (Shanghai Jiao Tong University); Shuai Wang (Nanjing University); Kai Yu (Shanghai Jiao Tong University)

Sun-M-O-2: Speech Spoofing and Content Manipulation Detection

Room: 8M 会议室

Type: Oral

November 8, 11:00 AM to 12:30 PM

11:00-11:15

Sun-M-O-2-1: WCAF-Net: Windowed Cross-Attention Fusion Network for Wild Singing Voice Deepfake Detection

Xinyi Wan (National University of Defense Technology); Ying Wang (National University of Defense Technology); Ke Wang (National University of Defense Technology); Nan Jia (National University of Defense Technology); Zhenyu Qiu (National University of Defense Technology); Lin Liu (National University of Defense Technology); Yuchuan Luo (National University of Defense Technology)*; Shaojing Fu (National University of Defense Technology)

11:15-11:30

Sun-M-O-2-2: Robust Speech Spoofing Detection via Progressive Hybrid Feature Aggregation and Hierarchical Gated Bidirectional Mamba

Yanping Li (Nanjing University of Posts and Telecommunications)*; Jiaqi Luan (Nanjing University of Posts and Telecommunications); Xiang Gao (Nanjing University of Posts and Telecommunications); Linhui Sun (Nanjing University of Posts and Telecommunications); Xi Shao (Nanjing University of Posts and Telecommunications)

11:30-11:45

Sun-M-O-2-3: A Static-Dynamic Difficulty Fusion Curriculum Sampling Method for Speech Deepfake Detection

Minjiao Yang (Beijing University of Posts and Telecommunications); Kangfeng Zheng (Beijing University of Posts and Telecommunications); Bin Wu (Beijing University of Posts and Telecommunications); Chunhua Wu (Beijing University of Posts and Telecommunications); Jvjie Wang (Beijing University of Posts and Telecommunications); Likai Jia (Beijing Institute of Computer Technology and Application); Yulin Yao (Beijing University of Posts and Telecommunications)*

11:45-12:00

Sun-M-O-2-4: ConSTMamba: Spectro-Temporal State Space Modeling in Conformer for Audio Deepfake Detection

Minjiao Yang (Beijing University of Posts and Telecommunications)*; Kangfeng Zheng (Beijing University of Posts and Telecommunications); Yulin Yao (Beijing University of Posts and Telecommunications); Jujie Wang (Beijing University of Posts and Telecommunications); Yaru Zhao (University of International Relations)

12:00-12:15

Sun-M-O-2-5: Wav2Vec-DMP-Net Fusion with Noise-Aware Geometric Bi-Directional Cross-Attention for Robust Speech Spoofing Detection

Kaian Cheang (Sun Yat-sen University); Nanli Zeng (China Mobile Internet Co., Ltd.); Yunchu Li (Sun Yat-sen University); Jianguo Hu (Sun Yat-sen University); Kun Zeng (Sun Yat-sen University)*

12:15-12:30

Sun-M-O-2-6: Lip-Focused Audio-Visual Consistency Detects Speech-Content Manipulation beyond Coarse Synchronization

Jian Zong (Tianjin University)*; Chenghan Lin (Tianjin University); Meng Liu (Tianjin University); Zhao Xu (Tianjin University); Junyu Wang (Tianjin University); Siyuan Zhang (Tianjin University); Peiyuan Jiang (Tianjin University); Longbiao Wang (Tianjin University); Jianwu Dang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)

Sun-M-O-3: Generative Speech Synthesis and Latent Representations

Room: 8F 会议室

Type: Oral

November 8, 11:00 AM to 12:30 PM

11:00-11:15

Sun-M-O-3-1: PhaseGAN: High-Fidelity Vocoder via Decoupled Amplitude and GAN-Driven Phase Reconstruction

Wenzheng Zhang (Inner Mongolia University)*; Xueliang Zhang (Inner Mongolia University); Shulin He (Inner Mongolia University); Fei Zhao (Inner Mongolia University); Xin Liu (Inner Mongolia University); Pengjie Shen (Inner Mongolia University); Zhenlong Guo (Inner Mongolia University); Zixuan Xue (Inner Mongolia University); Hongtao Bao (Inner Mongolia University); Zixuan Li (Inner Mongolia University); Jiahui Pan (Inner Mongolia University)

11:15-11:30

Sun-M-O-3-2: Towards Flow-Matching-based TTS without Classifier-Free Guidance

Yuzhe Liang (Shanghai Jiao Tong University)*; Chunyu Qiang (Kuaishou); Wenzhe Liu (Kuaishou); Zhikang Niu (Shanghai Jiao Tong University); Yushen Chen (Shanghai Jiao Tong University); Xiaopeng Wang (Kuaishou); Kang Yin (Kuaishou); Wenxi Chen (Shanghai Jiao Tong University); Nan Li (Kuaishou); Teng Ma (Kuaishou); Ziyu Zhang (Kuaishou); Chen Zhang (Kuaishou); Xie Chen (Shanghai Jiao Tong University)

11:30-11:45

Sun-M-O-3-3: ResiSpeech: Residual-Logit Steering of a Frozen Speech-Only Autoregressive Prior for Zero-Shot TTS

Haiyang Sun (Shanghai Jiao Tong University)*; Wei Wang (Shanghai Jiao Tong University); Chenyang Le (Shanghai Jiao Tong University); Tianteng Gu (Shanghai Jiao Tong University); Shuai Shi (Vui Lab); Mengxiao Bi (Vui Lab); Yanmin Qian (Shanghai Jiao Tong University)

11:45-12:00

Sun-M-O-3-4: SphereVAE: Hyperspherical Latent Autoencoders for Robust Autoregressive Speech Representation Modeling

Haoyu Zhang (Northwestern Polytechnical University)*; Jingbin Hu (Northwestern Polytechnical University); Hanke Xie (Northwestern Polytechnical University); Qirui Zhan (Northwestern Polytechnical University); Wenhao Li (Northwestern Polytechnical University); Ziyu Zhang (Northwestern Polytechnical University); Xiaming Ren (Northwestern Polytechnical University); Yue Li (NetEase Inc.); Xunyu Zhu (NetEase Inc.); Zhipeng Chen (NetEase Inc.); Lei Xie (Northwestern Polytechnical University)

12:00-12:15

Sun-M-O-3-5: DiTAR+: Dual Optimization for Robust Autoregressive Diffusion Speech Synthesis

Ziyu ZHANG (Northwestern Polytechnical University)*; Tianlun Zuo (Northwestern Polytechnical University); Hanzhao Li (Northwestern Polytechnical University); Haoyu Zhang (Northwestern Polytechnical University); Lei Xie (Northwestern Polytechnical University)

12:15-12:30

Sun-M-O-3-6: Speculative Decoding for Autoregressive Text-to-Speech in Continuous Latent Space

Xuetao Lin (Xiamen University)*; Yunchao He (XiaoAI Plus, Xiaomi Inc., Beijing, China); Di Wu (XiaoAI Plus, Xiaomi Inc., Beijing, China); Jian Luan (XiaoAI Plus, Xiaomi Inc., Beijing, China); Lin Li (Xiamen University); Qingyang Hong (Xiamen University)

Sun-M-SS-1: 特殊议题 7: 语音与音乐的动态交互

Room: 沐辉厅 B

Type: Special Session

November 8, 2026, 11:00 AM to 12:30 PM

Organizer: 杜忆老师, 丁鼐老师

11:00-11:20

Sun-M-SS-1-1: 语音、音乐与人脑中的节奏

Presenter: 丁鼐老师, 浙江大学教授

11:20-11:40

Sun-M-SS-1-2: Poetry and Music Jointly Shape Neural Tracking of Acoustic Envelopes and Phrase Boundaries in Songs

Presenter: 滕相斌老师, 香港中文大学心理学系助理教授

11:40-12:00

Sun-M-SS-1-3: 基础听觉特征自动整合依赖于意识的参与

Presenter: 田兴老师, 上海纽约大学副教授

12:00-12:20

Sun-M-SS-1-4: 音乐情绪对语音特征的动态调制神经机制

Presenter: 杜忆老师, 中国科学院心理研究所研究员

Sun-M-SS-2: 特殊议题 3: 第三届多语言智能语音与语言技术研讨会**Room:** 沐辉厅 C**Type:** Special Session**November 8, 2026, 11:00 AM to 12:30 PM****Organizer:** 欧智坚老师, 洪青阳老师

11:00-11:20

Sun-M-SS-2-1: 大模型范式下的语音识别与理解

Presenter: 谢磊老师, 西北工业大学教授

11:20-11:40

Sun-M-SS-2-2: 多语言语音识别与合成模型进展及展望

Presenter: 陈谐老师, 上海交通大学副教授

11:40-12:00

Sun-M-SS-2-3: 迈向全语种覆盖的音色克隆语音合成模型

Presenter: 朱涵老师, 小米高级算法工程师

12:00-12:20

Sun-M-SS-2-4: NIM4-ASR: 面向 Agent 时代的语音识别大模型

Presenter: 谢源, 蔚来汽车 Staff Engineer

Sun-M-SS-3: 特殊议题 12: 动态场景下基于麦克风阵列的声场建模与语音感知**Room:** 2D 会议室**Type:** Special Session**November 8, 2026, 11:00 AM to 12:40 PM****Organizer:** 钱彦旻老师, 黄公平老师, 张王优老师, 白仲鑫老师

11:00-11:20

Sun-M-SS-3-1: 复杂环境下多通道声信号感知与处理的挑战

Presenter: 陈景东老师, 武汉大学教授

11:20-11:40

Sun-M-SS-3-2: 视听融合的空间协同感知技术与应用

Presenter: 李军锋老师, 中国科学院声学研究所研究员

11:40-12:00

Sun-M-SS-3-3: 从声学感知到声场调控——声学阵列技术的进展与展望

Presenter: 卢晶老师, 南京大学教授

12:00-12:20

Sun-M-SS-3-4: 干扰环境下动态声源的提取方法

Presenter: 潘超老师, 西北工业大学副教授

12:20-12:40

Sun-M-SS-3-5: 通用声音分离中的混合专家模型——声学与语义引导的稀疏路由

Presenter: 李晨达老师, 上海交通大学助理研究员

Sun-M-SS-4: 特殊议题 10: 语音模型评测体系构建与开源生态创新
Room: 2R 会议室
Type: Special Session
November 8, 2026, 11:00 AM to 12:30 PM
Organizer: 俞凯老师, 陈谐老师, 卜辉老师, 谢磊老师, 王帅老师, 朱梦尧老师

11:00-11:20
Sun-M-SS-4-1: 迈向标准化可复现的开放语音模型评测体系
Presenter: 俞凯老师, 上海交通大学教授

11:20-11:35
Sun-M-SS-4-2: 语音评测数据集现状与标准化挑战
Presenter: 卜辉老师, 希尔贝壳创始人

11:35-11:50
Sun-M-SS-4-3: 语音评测标准体系的思考与实践
Presenter: 朱梦尧老师, 苏州大学教授

11:50-12:05
Sun-M-SS-4-4: 语音赛事相关情况汇总与思考
Presenter: 谢磊老师, 西北工业大学教授

12:05-12:30
Sun-M-SS-4-5: 圆桌论坛: 语音模型评测如何走向可信、开放与可复现
Presenter: 主持: 王帅老师, 南京大学副教授

Sun-M-YF: Young Forum 2
Room: 沐辉厅 A
Type: Young Forum
November 8, 2026, 11:00 AM to 12:30 PM

11:00-11:30
Sun-M-YF-1: 融入听觉感知机理的语音增强方法研究
Presenter: 李楠老师, 天津师范大学计算机与信息工程学院讲师

11:30-12:00
Sun-M-YF-2: 面向具身智能的空间听觉与语言交互: 从声场表征到跨设备泛化
Presenter: 胡诗超老师, 苏州大学未来科学与工程学院听觉及语言智能研究中心副教授、硕士生导师

12:00-12:30
Sun-M-YF-3: 从均方误差到听觉感知: 声音控制与重现的优化、实现与评价
Presenter: 徐春阳老师, 苏州大学听觉与语言智能研究中心讲师

Sun-A-O-1: Spotlight Paper Presentation II**Room: 7 号会议室****Type: Oral****November 8, 2:00 PM to 3:30 PM**

14:00-14:15

Sun-A-O-1-1: SecoustiCodec: Cross-Modal Aligned Streaming Single-Codebook Speech Codec

Chunyu Qiang (Tianjin University)*; Haoyu Wang (Tianjin University); Cheng Gong (Tianjin University); Tianrui Wang (Tianjin University); Ruibo Fu (Institute of Automation, Chinese Academy of Sciences); Tao Wang (Institute of Automation, Chinese Academy of Sciences); Ruilong Chen (Kuaishou Technology); Jiangyan Yi (Tsinghua University); Zhengqi Wen (Tsinghua University); Chen Zhang (Kuaishou Technology); Longbiao Wang (Tianjin University); Jiangwu Dang (Tianjin University); Jianhua Tao (Tsinghua University)

14:15-14:30

Sun-A-O-1-2: FAB-Diar: Activity Fusion and Assignment Backtracking for Online Speaker Diarization

Pan Shenghui (XinJiang University)*; Lixu Sun (XinJiang University); Jian Peng (XinJiang University); Nurmamet Yolwas (XinJiang University); Hongyu Lin (XinJiang University); Wushour Silamu (XinJiang University); Adinakiz Arken (Xinjiang Public Security Cybersecurity Management Center)

14:30-14:45

Sun-A-O-1-3: M4AVNet: A Mamba-Based Network with Multi-Level Fusion and Modality-Specific Modeling for Audio-Visual Speech Separation

Yu Jiang (Anhui University); Zhouxu Yang (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, (School of Computer Science and Technology), Anhui University, Hefei, Anhui, China); Mengyuan Gao (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, (School of Computer Science and Technology), Anhui University, Hefei, Anhui, China); Zhao Lv (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, (School of Computer Science and Technology), Anhui University, Hefei, Anhui, China); Cunhang Fan (State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, (School of Computer Science and Technology), Anhui University, Hefei, Anhui, China)*

14:45-15:00

Sun-A-O-1-4: StarTSE: Towards Streaming Target Speaker Extraction via Chunk-wise Interleaved Splicing of Autoregressive Language Model

ShuHai Peng (Tsinghua University)*; Hui Lu (The Chinese University of Hong Kong); Jinjiang Liu (Tsinghua University); Liyang Chen (Tsinghua University); Guiping Zhong (SenseTime); Jiakui Li (SenseTime); Huimeng Wang (The Chinese University of Hong Kong); Haiyun Li (Tsinghua University); Cao Liang (Tsinghua University); Shiyin Kang (SenseTime); Zhiyong Wu (Tsinghua University)

15:00-15:15

Sun-A-O-1-5: Speech-Based Dementia Detection Using Whisper and Log Neural Controlled Differential Equations

Sicheng Yu (College of Mathematics and Statistics, Chongqing University); Danyu Yang (College of Mathematics and Statistics, Chongqing University)*

15:15-15:30

Sun-A-O-1-6: Real-Validated UAV Audition Under Rotor Ego-Noise for Low-False-Alarm Human Detection

Junhao Wei (Macao Polytechnic University)*; Haochen Li (Macao Polytechnic University); Dexing Yao (Macao Polytechnic University); Yanxiao Li (Macao Polytechnic University); Yifu Zhao (Macao Polytechnic University); Baili Lu (Macao Polytechnic University); Zhenhong Peng (Zhongkai University of Agriculture and Engineering); Ngai Cheong (Macao Polytechnic University); Xu Yang (Macao Polytechnic University); Yapeng Wang (Macao Polytechnic University)

Sun-A-O-2: Multimodal Emotion, Sentiment and Group Affect Recognition**Room: 8M 会议室****Type: Oral****November 8, 2:00 PM to 3:30 PM**

14:00-14:15

Sun-A-O-2-1: Multimodal Embedding Model as Task-Oriented Affective Embedding Refiner for Multimodal Emotion Recognition

Wenpiao Su (Beijing University of Posts and Telecommunications)*; Runnan Li (Beijing University of Posts and Telecommunications); Xiaodan Zhao (Beijing University of Posts and Telecommunications); Peng Xu (Beijing University of Posts and Telecommunications); Peiyu Zheng (Beijing University of Posts and Telecommunications); Yaohua Bu (Beijing University of Posts and Telecommunications)

14:15-14:30

Sun-A-O-2-2: Evaluating Multimodal Large Language Models on Spoken Sarcasm Detection

Zhu Li (University of Groningen)*; Xiyuan Gao (University of Groningen); Yuqing Zhang (University of Groningen); Shekhar Nayak (University of Groningen); Matt Coler (University of Groningen)

14:30-14:45

Sun-A-O-2-3: AP-TFNet: An Amplitude-Phase Interaction and Adaptive Temporal Focusing Network for Speech Emotion Recognition

Huimin Lv (Xinjiang University); Gulnur Arkin (Xinjiang University of Finance and Economics); Saidiyaguli Ainiwaer (Xinjiang Agricultural Vocational and Technical University); Hankiz Yilahun (Xinjiang University)*; Askar Hamdulla (Xinjiang University)

14:45-15:00

Sun-A-O-2-4: Semantic-Anchored Multi-Prototype vMF Head for Speech Emotion Recognition

蔡易 (四川师范大学); 谢瑞 (四川师范大学); 杨军 (四川师范大学)*

15:00-15:15

Sun-A-O-2-5: MorphRA: Reliability-Aware Morphological Adaptation for Few-Shot Agglutinative Sentiment Analysis

茹诚 (XinJiang University); 彭坚 (Xinjiang University)*; 孙李旭 (XinJiang University); Yolwas Nurmemet (XinJiang University)

15:15-15:30

Sun-A-O-2-6: When Does Per-Person Modeling Matter? A Controlled Study of Multimodal Group Affect Recognition Across Acquisition Regimes

Godbless Mensah (School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China.); Qirong Mao (Jiangsu University)*; Nuku Atta Kordzo Abiew (School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China.); Lijian Gao (School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China.); Ernest Asimeng (School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China.); Qing Yao (School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China.)

Sun-A-O-3: Spoken Dialogue Systems, Turn-Taking and Interaction Modeling**Room: 8F 会议室****Type: Oral****November 8, 2:00 PM to 3:30 PM**

14:00-14:15

Sun-A-O-3-1: Structured Backchannel Control Prediction for Expressive Spoken Dialogue with PRISM

Zhengyao Ma (Beijing University of Posts and Telecommunications)*; Zhaoyang Meng (Beijing University of Posts and Telecommunications); Kekan Mao (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications); Ya Li (Beijing University of Posts and Telecommunications)

14:15-14:30

Sun-A-O-3-2: EmpathyBC: Enhancing Empathetic Spoken Dialogue with Semantic- and Emotion-Aware Backchannel Realization

Zhaoyang Meng (Beijing University of Posts and Telecommunications)*; Kekan Mao (Beijing University of Posts and Telecommunications); Zhengyao Ma (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications); Ya Li (Beijing University of Posts and Telecommunications)

14:30-14:45

Sun-A-O-3-3: Capability-Guided Iterative Online Optimization of Cascaded Spoken-Dialogue Modules

Yuhang Qiu (Shanghai Jiao Tong University)*; Bohan Li (Shanghai Jiao Tong University); Wenbin Huang (Shanghai Jiao Tong University); Zhihan Li (Shanghai Jiao Tong University); Yiwei Guo (Shanghai Jiao Tong University); Hankun Wang (Shanghai Jiao Tong University); Xie Chen (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

14:45-15:00

Sun-A-O-3-4: JAL-Turn: Joint Acoustic-Linguistic Modeling for Real-Time and Robust Turn-Taking Detection in Full-Duplex Spoken Dialogue Systems

yu pan (Recho Inc.)*; guangzhao yang (Recho Inc.); Muhammad Huzaifah (Recho Inc.); ningjie bai (Recho Inc.); shi qiu (Recho Inc.)

15:00-15:15

Sun-A-O-3-5: A Modular Full-Duplex Spoken Dialogue System with Multi-Class Turn Detection

Hui Lu (The Chinese University of Hong Kong)*; Xueyuan Chen (The Chinese University of Hong Kong); Huimeng Wang (The Chinese University of Hong Kong); Shuhai Peng (Tsinghua University); Zewen Zhan (The Chinese University of Hong Kong); Guiping Zhong (SenseTime Research); Shiyin Kang (SenseTime Research); Zhiyong Wu (Tsinghua University)

15:15-15:30

Sun-A-O-3-6: Auditing Weak Emotion Supervision for Title-Only Conversational Recommendation

Jiaxu Zhao (SUN YAT-SEN UNIVERSITY); Yonghe Lu (SUN YAT-SEN UNIVERSITY)*; Ngai Heng Cheong (META-ARCHIT SOFTWARE TECHNOLOGY LIMITED)

Sun-A-P-1: Speech and Audio Generation, Coding and Restoration**Room: 2D 会议室 + 2R 会议室 (海报区)****Type: Poster****November 8, 2:00 PM to 3:30 PM****1. Expressive and Controllable Speech Synthesis (9 papers)****Sun-A-P-1-1: emo-MB-iSTFT-VITS: A Lightweight Emotional TTS with Speech Emotion Embeddings**

Zijie Yin (Guilin University of Electronic TechnologyGuilin University of Electronic Technology)*

Sun-A-P-1-2: AuraRAG: Enabling Controllable and Emotion-Aware Speech Retrieval-Augmented Generation

xiaoyang luo (Beijing language and culture university)*; dengfeng ke (Beijing language and culture university); jinsong zhang (Beijing language and culture university); kaile su (Beijing language and culture university)

Sun-A-P-1-3: A Face-Conditioned Flow Matching Model for Text-to-Speech

Jichen Yang (GPNU)*

Sun-A-P-1-4: Graph-based Inter-Turn Emotional Dependency Modeling for Emotional Conversational Speech Synthesis

Xiaochun An (Xi'an University of Posts & Telecommunications)*; Chengye Lv (Xi'an University of Posts & Telecommunications); Cui Liu (Xi'an University of Posts & Telecommunications); Ercheng Pei (Xi'an University of Posts & Telecommunications); Xiaoge Li (Xi'an University of Posts & Telecommunications); Lang He (Xi'an University of Posts & Telecommunications)

Sun-A-P-1-5: Generated Chinese Affective Speech as Cross-Domain Cold-Start Data for Emotion Recognition: CER-Only Source Filtering and Few-Shot Target Calibration

Zhifeng Liang (University of Jinan)*; Hua Shao (University of Jinan); Xintang Wang (University of Jinan); Chuanming Yin (Institute of Information Engineering, Chinese Academy of Sciences)

Sun-A-P-1-6: 5D-Persona: A Five-Dimensional Interpretable Voice Persona Representation for Text-Driven Speech Generation

王旭 (北京邮电大学)*

Sun-A-P-1-7: Towards Continuous Emotional Voice Editing: A Unified Valence-Arousal Representation with VA-Guided Residual Diffusion Refinement

Dehui Luo (University of Electronic Science and Technology of China); Guobin Zhu (University of Electronic Science and Technology of China)*; Yang Zhao (University of Electronic Science and Technology of China); Yikun Liu (University of Electronic Science and Technology of China); Kai Ning (University of Electronic Science and Technology of China)

Sun-A-P-1-8: MMSpeaker: A Multimodal Joint Framework for Expressive Controllable Speech Synthesis

Rui Niu (Tsinghua University); Zhiyong Wu (Tsinghua University)*

Sun-A-P-1-9: Perturbation Self-Supervised Representations for Cross-Lingual Emotion TTS: Stage-Wise Modeling of Emotion and Speaker

Cheng Gong (Tianjin University)*; Chunyu Qiang (Tianjin University); Tianrui Wang (Tianjin University); Yu Jiang (Tianjin University); Yuheng Lu (Tianjin University); Ruihao Jing (Northwestern Polytechnical University); Xiaoxiao Miao (Duke Kunshan University); Xiaolei Zhang (Northwestern Polytechnical University); Longbiao Wang (Tianjin University); Jianwu Dang (Tianjin University)

2. TTS Modeling, Vocoding and Efficient Inference (8 papers)

Sun-A-P-1-10: Surpassing Per-Task BERT Baselines: A Unified Generative TTS Frontend via Jointly-Annotated Synthetic Data

rui song (USTC); Yuanjie Dong (iFLYTEK); Shichun Yan (iFLYTEK); Chang Liu (iFLYTEK); Yajun Hu (iFLYTEK); Lei Sun (iFLYTEK); Jianqing Gao (iFLYTEK); Zhen-Hua Ling (USTC); Yu Hu (iFLYTEK); Xin Fang (USTC)*

Sun-A-P-1-11: FAIR-TTS: Fast Alignment-Aware Inference and Alignment-Guided Post-Training via Attention Tracking for Robust Text-to-Speech Synthesis

Zhang Borui (Shanghai Jiaotong University)*; Hankun Wang (Shanghai Jiaotong University); Yiwei Guo (Shanghai Jiaotong University); Kai Yu (Shanghai Jiaotong University)

Sun-A-P-1-12: Posterior Representation Learning for Text-to-Speech

Chen Jiatai (Harbin University of Science and Technology); Chen Chen (Harbin University of Science and Technology)*; Feng Yuhongxu (Harbin University of Science and Technology); Li Ao (Harbin University of Science and Technology); Yang Hailu (Harbin University of Science and Technology)

Sun-A-P-1-13: Prompt-Distilled Speaker Injection for Transcript-Free Zero-Shot Text-to-Speech

Dekun Chen (The Chinese University of Hong Kong, Shenzhen)*; Yingda Shen (The Chinese University of Hong Kong, Shenzhen); Zhizheng Wu (The Chinese University of Hong Kong, Shenzhen)

Sun-A-P-1-14: CoStaP-Flow: Learning Conditional Start-State Priors for Low-Step Flow-Based Neural Vocoding

yonghe Wang (Inner Mongolia University)*; zi wang (Inner Mongolia University); long Fei (Inner Mongolia University); guanglai Gao (Inner Mongolia University)

Sun-A-P-1-15: Fast First Audio for Edge Text-to-Speech: Adaptive Chunking with Budgeted Decoding

Rui Guo (Beijing Polytechnic College); Ning Zhu (Shangyu Technology)*; Yuansu Zhao (Beijing Polytechnic College)

Sun-A-P-1-16: MoVoc: A Sparse Mixture-of-Experts Neural Vocoder

Rui Guo (Beijing Polytechnic College); Ning Zhu (Shangyu Technology)*; Yuansu Zhao (Beijing Polytechnic College)

Sun-A-P-1-17: SpeechLM-Inspired Hierarchical Speech Planning for Flow-Matching Text-to-Speech

Zeyu Qiu (Information Engineering University)*; Dan Qu (Information Engineering University); Nianwen Si (Information Engineering University); Qi Chen (Information Engineering University)

3. Multilingual Synthesis, Voice Conversion and Dubbing (7 papers)

Sun-A-P-1-18: MultiSwap: Cross-Modal Identity Conversion with Silent Video and Audio Inputs

Yifan Liu (Shanghai Jiao Tong University)*

Sun-A-P-1-19: FusionFlow-SVC: A Gated Fusion-Based Framework for Robust and High-Fidelity Singing Voice Conversion

胡玉叶 (Xinjiang University); 徐承烨 (Xinjiang University)*; 黄浩 (Xinjiang University)

Sun-A-P-1-20: 面向安多藏语复辅音特性的高频增强端到端语音合成研究

德却措 (青海民族大学); Kunthargyal Khysru (青海民族大学)*; 李美晨 (青海民族大学); 仁青华毛 (青海民族大学); 拉马杰 (青海民族大学); 魏建国 (天津大学); 金弟 (天津大学)

Sun-A-P-1-21: Paralinguistic-Aware Zero-Shot Voice Conversion via Gated Feature Fusion and Adversarial Speaker Disentanglement

Wei Chen (Tsinghua University)*; Zhipeng Chen (NetEase, Inc.); Dan Luo (Tsinghua University); Zhiyong Wu (Tsinghua University)

Sun-A-P-1-22: VoxDub: A Bilingual Lip-Synchronized Audio-Visual Dataset for Automatic Video Dubbing

kaidi wang (xiamen university)*; Wenhao Guan (xiamen university); Peijie Chen (xiamen university); Weijie Wu (xiamen university); Xuetao Lin (xiamen university); lin li (xiamen university); Qingyang Hong (xiamen university)

Sun-A-P-1-23: Lightweight Tibetan Speech Synthesis Based on Knowledge Distillation and Multi-Band Inverse Short-Time Fourier Transform

weizhao zhang (Northwest Normal University)*; qi shao (Northwest Normal University); hongwu yang (Northwest Normal University)

Sun-A-P-1-24: Robustness evaluation and failure diagnosis of complex text in low-resource multilingual speech synthesis

Tianlun Zuo (Northwestern Polytechnical University)*; Ziyu Zhang (Northwestern Polytechnical University); Tingzhi Mao (iFLYTEK Company Ltd., Xi'an); Zhonghua Fu (Northwestern Polytechnical University); Lei Xie (Northwestern Polytechnical University)

4. Neural Speech and Audio Coding (5 papers)

Sun-A-P-1-25: Comparative Study of Front-End Representations in Low-Bitrate Neural Speech Codec

Hui-Peng Du (University of Science and Technology of China)*; Xiao-Hang Jiang (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)

Sun-A-P-1-26: Mu-APCodec: A Dual-Domain STFT-CQT Neural Audio Codec for Music Compression

Chenyuan Ning (University of Science and Technology of China)*; Yang Ai (University of Science and Technology of China); Xiao-Hang Jiang (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)

Sun-A-P-1-27: AV-RVQ: Audio-Visual Adaptive Residual Vector Quantization for Robust Speech Compression

Tongzhen Chen Chen (辽宁大学)*; Yuan Jing Jing (LiaoNing University)

Sun-A-P-1-28: Sparse-Aware Structured Pruning for Lightweight MDCT Neural Speech Coding

En-Wei Zhang (University of Science and Technology of China)*; Hui-Peng Du (University of Science and Technology of China); Yang Ai (University of Science and Technology of China); Yuan Tian (University of Science and Technology of China); Zhen-Hua Ling (University of Science and Technology of China)

Sun-A-P-1-29: Usage-Aware Quantization and Horizon-Weighted Prediction for Enhanced RPQ-Based Speech Self-Supervised Learning

Yueying Chen (Tianjin University)*; Tianrui Wang (Tianjin University); Zhao Xu (Tianjin University); Xuanchen Li (Tianjin University); Zikang Huang (Tianjin University); Jiahao Yan (Tianjin University); Jintao Liang (Tianjin University); Xiao Wei (Tianjin University); Longbiao Wang (Tianjin University); Jianwu Dang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)

5. Generative Data Augmentation and Multimodal Content (10 papers)

Sun-A-P-1-30: Consistency Penalties and Fact-Density Control for Preference-Based RL in Story Generation: An Empirical Analysis

Xuhang Chen (University of Macau)*; Zhonghua Xie (Huizhou University); Lingjun Liu (Huizhou University); Jie Dong (Huizhou University); Chi-Man Pun (University of Macau)

Sun-A-P-1-31: Speech-Driven Structured Behavior Control for Affect-Aware MR Virtual Humans via Utility Behavior Trees

Zhenhao Zhang (Tianjin University); Baojie Yan (Tianjin University); Jinhao Dong (Tianjin University); Zhilei Liu (Tianjin University)*

Sun-A-P-1-32: Mitigating Distribution Shift in Data Augmentation via Text-Guided Synthesis for Dysarthric Speech Recognition

Siwen Guo (National University of Defense Technology)*; Yi Su (National University of Defense Technology); Hongyu Wen (National University of Defense Technology); Cheng Yang (National University of Defense Technology); Kele Xu (National University of Defense Technology)

Sun-A-P-1-33: 基于生成式数据增强的构音障碍语音重建方法

Heng Liu (Shenzhen university)*; Nengheng zheng (shenzhen university)

Sun-A-P-1-34: AlignToken-DSR: Zero-Shot Mandarin Dysarthric Speech Reconstruction with Alignment Supervision and Semantic Tokens

Yixiang Chen (University of Chinese Academy of Sciences); Gaoyuan Zhang (Institute of Software, Chinese Academy of Sciences); Shansong Liu (Independent Researcher); Xurong Xie (Institute of Software, Chinese Academy of Sciences)*; Hui Chen (Institute of Software, Chinese Academy of Sciences); Feng Tian (Institute of Software, Chinese Academy of Sciences)

Sun-A-P-1-35: ZEAL: Zipformer-Enhanced Articulation Learning with Flow-Regularized Dynamics for Speaker-Aware Lip Motion Synthesis

Jiawei Jiang (Beijing University of Posts and Telecommunications)*; Runnan Li (Beijing University of Posts and Telecommunications); Wenpiao Su (Beijing University of Posts and Telecommunications); Xiaodan Zhao (Beijing University of Posts and Telecommunications); Peng Xu (Beijing University of Posts and Telecommunications); Peiyu Zheng (Beijing University of Posts and Telecommunications)

Sun-A-P-1-36: Music as a Motion Planner: Prior-Preserving Shared-State Conditioning for Dance Video Generation

Meng Wang (Shenzhen International Graduate School. Tsinghua University)*; Liyang Chen (Shenzhen International Graduate School. Tsinghua University); Runnan Li (Beijing University of Posts and Telecommunications); Zhiyong Wu (Shenzhen International Graduate School. Tsinghua University)

Sun-A-P-1-37: PhyFlow: Multi-Modal Flow Matching and Physics-Driven Motion Synthesis for Interactive Virtual Humans

Keqin Tang (Beijing University of Posts and Telecommunications)*; Tuo Hao (Beijing University of Posts and Telecommunications); Wenxin Fu (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications); Ya Li (Beijing University of Posts and Telecommunications)

Sun-A-P-1-38: EIAFace: Explicit Identity-Aware Audio-Driven Facial Animation

Haifeng Chen (Shaanxi University of Science & Technology); Lina Mu (Shaanxi University of Science & Technology)*; Jian Li (Shaanxi University of Science & Technology); Dongmei Jiang (Peng Cheng Laboratory); Pei Nie (Hospital of Northwestern Polytechnical University)

Sun-A-P-1-39: IMAGE: Enriching Multi-Modal Dialogue Dataset with Generative Models

Dongning Rao (广东工业大学)*; Peng Chen (广东工业大学)

6. Music and Audio Generation, Understanding and Restoration (14 papers)

Sun-A-P-1-40: Dynamic Resolution Fusion Front-End for Time-Frequency Music Source Separation

Yue Li (东华大学); ZiJian Chai (东华大学)*; YangWu Li (东华大学); RuiMin Zhou (东华大学)

Sun-A-P-1-41: Expressive Background Music Generation with Dialogue Knowledge Aggregation: Dataset and Methodology

Shaoli Ge (Inner Mongolia University)*; Rui Liu (Inner Mongolia University)

Sun-A-P-1-42: Balancing Alignment and Locality: Hybrid Attention for Lightweight Controllable Text-to-Music Generation

Li Zhang (Southwestern University of Finance and Economics)*; Ningning Pan (Southwestern University of Finance and Economics); Haoyu Wang (Southwestern University of Finance and Economics); Xiaoyi Shen (Institute of Acoustics, Chinese Academy of Sciences); Gongping Huang (Wuhan University)

Sun-A-P-1-43: SingVERSE: A Diverse, Real-World Benchmark for Singing Voice Enhancement

Junan Zhang (The Chinese University of Hong Kong, Shenzhen)*; Shaohan Jiang (The Chinese University of Hong Kong, Shenzhen); Yunjia Zhang (The Chinese University of Hong Kong, Shenzhen); Jing Yang (Central Media Technology Institute, Huawei); Fan Fan (Central Media Technology Institute, Huawei); Zhizheng Wu (The Chinese University of Hong Kong, Shenzhen)

Sun-A-P-1-44: 面向听障音乐的内容感知安全门控残差重混方法

Weiwen Huang (ShenZhen University)*; LiHuang Zhou (ShenZhen University); NengHeng Zheng (ShenZhen University)

Sun-A-P-1-45: LaMP-Cap: A Language-Guided Acoustic-Semantic Fusion Framework for Multilingual Music Captioning

Yi Lou (Inner Mongolia University)*; Zhenqi Jia (Inner Mongolia University); Yifan Hu (Inner Mongolia University); De Hu (Inner Mongolia University); Xiangyu Cui (Beijing Haitian Ruisheng Science Technology Ltd.); Rui Liu (Inner Mongolia University)

Sun-A-P-1-46: Audio Imitator: Controlling Timbre and Tempo in Video2Audio Synthesis with Audio Reference

Jiahui Zhao (Tianjin University)*; Tianrui Wang (Tianjin University); Chunyu Qiang (Tianjin University); Cheng Gong (Tianjin University); Longbiao Wang (Tianjin University)

Sun-A-P-1-47: DriftME: Restoring Full-Band Music from Mobile Recordings with Drifting-Inspired Latent Regularization

ZhanJia Zhang (TsingHua University)*; Wenxin Bao (TsingHua University); Jiehui Luo (TsingHua University); Jinjiang Liu (TsingHua University); Zhiyong Wu (TsingHua University)

Sun-A-P-1-48: STAR: End-to-End Speech-to-Audio Generation via Speech Semantic Representation Bridging

Zeyu Xie (Peking University)*; Xuenan Xu (SJTU); Yixuan Li (SJTU); Chenxing Li (Tencent); Mengyue Wu (SJTU); Yuexian Zou (Peking University)

Sun-A-P-1-49: SCALA: Recorded Music Restoration with Time-Frequency Modeling and Adversarial Supervision

Jinyi Lin (HIT); Jinping Ge (HIT)*; Jie Chen (HIT)

Sun-A-P-1-50: 声学内容与内容表征协同的生成式音乐修复

Chengzhong Wang (Institute of Acoustics, Chinese Academy of Sciences)*; Zhaoyang Wang (Institute of Acoustics, Chinese Academy of Sciences); Dingding Yao (Institute of Acoustics, Chinese Academy of Sciences)

Sun-A-P-1-51: 基于时间因果建模与多目标优化的轻量音乐修复

junzhe liu (苏州大学)*; Shichao Hu (苏州大学); Zeyi Shao (苏州大学); Mengyao Zhu (苏州大学)

Sun-A-P-1-52: The NPU-ASLP System for Music Restoration in the CCF-AATC 2026 Challenge

Zeyu Zhu (Northwestern Polytechnical University)*; Yike Zhu (Northwestern Polytechnical University); Xingchen Li (Northwestern Polytechnical University); Chaoren Yu (Northwestern Polytechnical University); Jingyi Wang (Northwestern Polytechnical University); Yurun Li (Northwestern Polytechnical University); Jierui Guo (Northwestern Polytechnical University); Chunjiang He (Northwestern Polytechnical University); Lei Xie (Northwestern Polytechnical University)

Sun-A-P-1-53: Band-MixNet-R4: Low-Complexity Real-World Music Restoration with Teacher-Student Distillation and Hierarchical Routing

Shuangrun Shao (CENTRAL CONSERVATORY OF MUSIC)*; Hongran An (CENTRAL CONSERVATORY OF MUSIC)

7. Active Noise Control and Competition Systems (5 papers)

Sun-A-P-1-54: RECR-ANC: Regional FIR Experts with Frozen Causal Routing for Active Noise Control

张虔 (大连理工大学)*; 陈喆 (大连理工大学); 殷福亮 (大连理工大学)

Sun-A-P-1-55: HybridHyperFIR: Response-Conditioned Dual-Rate Finite Impulse Response Filter Generation for Active Noise Control

Yi-Pei Zhang (Harbin Institute of Technology, Shenzhen); Zi-Han Hu (Harbin Institute of Technology, Shenzhen); Shun-Yu Zhou (Harbin Institute of Technology, Shenzhen); Wei Gao (Harbin Institute of Technology, Shenzhen); Hong-Wei Zhang (Harbin Institute of Technology, Shenzhen)*

Sun-A-P-1-56: Feedforward-Feedback Hybrid Dual-Branch Fusion For Active Noise Control

白璐 (南京大学)*; 南晓峰 (南京大学)

Sun-A-P-1-57: Multi-Expert Online Routing Active Noise Control System

Hucheng Wang (Center of Intelligent Acoustics and Immersive Communications, School of Marine Science and Technology, Northwestern Polytechnical University)*

Sun-A-P-1-58: FLASH-ANC: Feedback-Guided Latent Acoustic-State Hypernetwork for Secondary-Path-Robust Deep Active Noise Control

Jiaxing Han (Nanjing University)*

Sun-A-SS-1: 特殊议题 9: 中国汉语方言及民族语言音节韵尾的共时变异与历时演化: 跨学科的多维视角

Room: 沐辉厅 B

Type: Special Session

November 8, 2026, 2:00 PM to 3:30 PM

Organizer: 罗颖艺老师, 李蕙心老师

14:00-14:15

Sun-A-SS-1-1: 香港粤语韵尾塞音的发音和音变

Presenter: 李蕙心老师, 香港城市大学翻译及语言学系副教授

14:15-14:30

Sun-A-SS-1-2: 粤方言韵尾功能负载研究

Presenter: 张劲松老师, 北京语言大学教授

14:30-14:45

Sun-A-SS-1-3: 普通话儿童儿化韵尾习得的声学研究

Presenter: 陈树雯老师, 中国社会科学院语言研究所助理研究员

14:45-15:00

Sun-A-SS-1-4: 普通话鼻韵尾的言语感知机制与适应效应

Presenter: 李爱军老师, 中国社会科学院语言研究所研究员

15:00-15:15

Sun-A-SS-1-5: 方言经验对普通话鼻音韵尾合流的影响—基于甘肃地区的产出与认知证据

Presenter: 罗颖艺老师, 中国社会科学院语言研究所副研究员

15:15-15:30

Sun-A-SS-1-6: 10-13 世纪北方汉语方言韵尾综论

Presenter: 孙森, 中国社会科学院语言研究所博士后

Sun-A-SS-2: 特殊议题 4: 复杂声学场景下的目标音频提取

Room: 沐辉厅 C

Type: Special Session

November 8, 2026, 2:00 PM to 3:30 PM

Organizer: 王帅老师, 李明老师, 钱馨园老师, 关键老师

14:00-14:20

Sun-A-SS-2-1:

Presenter: 李明老师, 香港中文大学 (深圳) 人工智能学院教授

14:20-14:40

Sun-A-SS-2-2: Cross-Talk Speech Reduction, by Separation, for Separation

Presenter: 王中秋老师, 南方科技大学计算机科学与工程系副教授

14:40-15:00

Sun-A-SS-2-3:

Presenter: 王帅老师, 南京大学副教授

15:00-15:15

Sun-A-SS-2-4: PA-TSE: Presence-Aware Target Speaker Extraction for Complex Sparse-Overlap Conversations

Guanyu Chen (Shanghai Jiao Tong University)*; Haoyu Li (Shanghai Jiao Tong University); Yu Xi (Shanghai Jiao Tong University); Jing Peng (Shanghai Jiao Tong University); Wen Wen (Shanghai Jiao Tong University); Xiaoyu Gu (Shanghai Jiao Tong University); Shuai Wang (Nanjing University); Kai Yu (Shanghai Jiao Tong University)

Presenter: Guanyu Chen (Shanghai Jiao Tong University)

15:15-15:30

Sun-A-SS-2-5: Text-Guided Target Noise Extraction for Urban Acoustic Monitoring with Domain Adaptation and Activity-Focused ERB Regularization

Ganlong Wang (Hubei University of Economics)*; Longfei Felix Yan (Lincoln University); Xianglong Wang (Hubei University of Technology); Qiankun Zuo (Hubei University of Economics); Yuqi Lin (Huazhong University of Science and Technology)

Presenter: Ganlong Wang (Hubei University of Economics)

Sun-A-YF-1: Young Forum 3

Room: 沐辉厅 A

Type: Young Forum

November 8, 2026, 2:00 PM to 3:30 PM

14:00-14:30

Sun-A-YF-1-1: 基于医患对话的老年认知健康智能诊断方法研究

Presenter: 潘怡霖老师, 大连海事大学人工智能学院讲师、硕士生导师

14:30-15:00

Sun-A-YF-1-2: 音乐物理特性对多巴胺系统神经夹带效应量化表征的机制研究

Presenter: 夏菁老师, 中国科学院深圳先进技术研究院助理研究员

15:00–15:30

Sun-A-YF-1-3: 结合语音基础模型的障碍语音识别建模方法

Presenter: 谢旭荣老师, 中国科学院软件研究所人机交互技术与智能信息处理实验室副研究员

Sun-A-O-4: Speech and Multimodal Intelligence for Health and Cognitive Assessment

Room: 7 号会议室

Type: Oral

November 8, 3:45 PM to 5:15 PM

15:45-16:00

Sun-A-O-4-1: A Robust Automatic Speech Recognition System for Elderly Users: Dataset Construction, Model Optimization and Practical Evaluation

Xiaojun Zhang (Xi'an Jiaotong-Liverpool University); Hui Jiao (Xi'an Jiaotong-Liverpool University)*; Chunjiang He (Northwestern Polytechnical University); Chengyou Wang (Northwestern Polytechnical University); Zheng Zi (Northwestern Polytechnical University)

16:00-16:15

Sun-A-O-4-2: Continuous Emotion Dynamics in Heterogeneous Speech: A Neural Controlled Differential Equation Approach with Dual-Stream Fusion

Yihe Wang (Beijing Institute of Petrochemical Technology)*; Yongkang Wang (Beijing Institute of Petrochemical Technology); Zhiyuan Ouyang (Beijing Institute of Petrochemical Technology); Yu Meng (Beijing Institute of Petrochemical Technology); Ming Yang (Beijing Institute of Petrochemical Technology); Xinyi Yu (Beijing Institute of Petrochemical Technology); Yida Song (Beijing Institute of Petrochemical Technology); Qiang Liu (Beijing Institute of Petrochemical Technology); Qiang Yang (Beijing Institute of Petrochemical Technology)

16:15-16:30

Sun-A-O-4-3: MedMM-SafetyBench: A Benchmark for Evaluating Medical Safety in Multimodal Large Language Models

Huihui Wei (Shanghai Jiaotong university)*; Dengke Deng (Shanghai Jiaotong university); Weiyu You (Shanghai Jiaotong university); Pengcheng Luo (Shanghai Jiaotong university); Genke Yang (Shanghai Jiaotong university); Chengguang Wang (Ningbo Industrial Internet Research Institute)

16:30-16:45

Sun-A-O-4-4: Why Alzheimer's Speech Screening Fails to Generalize: Bridging the Deployment Gap via Cross-Corpus Evidence Anchoring

Zijian Lu (Nanjing University of Posts and Telecommunications); Sizhe Liu (University of Science and Technology of China); Yin Zhang (University of Science and Technology of China); Jixuan Deng (University of Science and Technology of China); Xinrong Lin (Nanjing Zhiwei Shouyi Technology Co., Ltd.); Xinchun Yuan (University of Science and Technology of China); Chicheng Jin (University of Science and Technology of China); Yiping Zuo (Nanjing University of Posts and Telecommunications); Yuanchao Li (University of Cambridge)*

16:45-17:00

Sun-A-O-4-5: VocalMind-Bridge: Acoustic Feature Coordination for Depression Detection from Participant Speech

Kang Yao (University of Science and Technology of China)*; Hao Li (University of Science and Technology of China); Congcong Tian (Faculty of Computing, Harbin Institute of Technology); Shengjiang Zhang (University of Science and Technology of China); Ting Dong (University of Science and Technology of China); JinJiang Cui (Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science); Weiwei Fu (Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science)

17:00-17:15

Sun-A-O-4-6: A Factorized Inverse Decision Model for Verbalized Cognitive Task Execution

Jiawen Kang (The Chinese University of Hong Kong)*; Dongrui Han (The Chinese University of Hong Kong); Xixin Wu (The Chinese University of Hong Kong); Helen Meng (The Chinese University of Hong Kong)

Sun-A-O-5: Speech and Audio-Visual Resources, Retrieval and Evaluation

Room: 8M 会议室

Type: Oral

November 8, 3:45 PM to 5:15 PM

15:45-16:00

Sun-A-O-5-1: ST-News: A Shantou News Dataset for Cross-Dialect Speech-to-Subtitle Generation and Non-Verbatim Mapping Analysis

Chen Jia (Sun Yat-sen University)*; Zhang Jinsong (Beijing Language and Culture University); Hu Jun (Sun Yat-sen University); Ke Dengfeng (Beijing Language and Culture University); Jian Chen (Shantou Media Convergence Group); Xiaoyun Chen (Shantou Media Convergence Group); Weiqing Huang (Shantou Media Convergence Group)

16:00-16:15

Sun-A-O-5-2: Frequency-Aware Attribute Prototype Learning for Fine-Grained Audio-Text Retrieval

Xun Zhang (Jiangsu University); Lijian Gao (Jiangsu University); Huate Zhu (Jiangsu University); Qirong Mao (Jiangsu University)*

16:15-16:30

Sun-A-O-5-3: CN-CVS3-P1: An Open-Source Large-Scale Chinese Audio-Visual Dataset for Visual Speech Recognition

Xiaolou Li (Beijing University of Posts and Telecommunications); Bochao Hu (bupt)*; Zehua Liu (Beijing University of Posts and Telecommunications); Chen Chen (Tsinghua University); Lantian Li (Beijing University of Posts and Telecommunications); Dong Wang (Tsinghua University)

16:30-16:45

Sun-A-O-5-4: SpeechAnnotator: A Context-Aware Multi-Agent Framework and Benchmark for Multidimensional Speech Annotation

qirui Zhan (Northwestern Polytechnical University)*; Shuiyuan Wang (Northwestern Polytechnical University); Jingbin Hu (Northwestern Polytechnical University); Liumeng Xue (Nanjing University); Haoyu Zhang (Northwestern Polytechnical University); Xiaming Ren (Northwestern Polytechnical University); Jinrui Liang (Northwestern Polytechnical University); Chaoren Yu (Northwestern Polytechnical University); Bengu Wu (yutuzhineng); Yunxiang Chen (Shenzhen Pimei Technology Co., Ltd., Guangdong, China); Houdun Liu (Shenzhen Pimei Technology Co., Ltd., Guangdong, China); Su Feng (Shenzhen Pimei Technology Co., Ltd., Guangdong, China); Lei Xie (Northwestern Polytechnical University)

16:45-17:00

Sun-A-O-5-5: Selection versus Perception: A Deterministic Diagnostic Framework for Fine-Grained Error Modes in Mandarin ASR

Yanxiao Li (Macao Polytechnic University); Junhao Wei (Macao Polytechnic University); Yifu Zhao (Macao Polytechnic University); Xu Yang (Macao Polytechnic University); Yapeng Wang (Macao Polytechnic University)*

17:00-17:15

Sun-A-O-5-6: Post-fusion Event-aware Time-Frequency Refinement for Audio Moment Retrieval

Feiyang Xiao (Harbin Engineering University); Li'ang Luo (Harbin Engineering University); Kejia Zhang (Harbin Engineering University); Qiaoxi Zhu (University of Technology Sydney); Guangjun He (State Key Laboratory of Space Information System and Integrated Application); Pengming Feng (State Key Laboratory of Space Information System and Integrated Application); Jian Guan (Harbin Engineering University)*

Sun-A-O-6: Speech, Audio and Visual Speech Generation**Room: 8F 会议室****Type: Oral****November 8, 3:45 PM to 5:15 PM**

15:45-16:00

Sun-A-O-6-1: BED: Boundary and Expansion Driven Framework for Accelerating Speech Generation

Dongfu Song (Tsinghua Shenzhen International Graduate School)*; Zhiyu Cui (Ant Group); Mingyu Zhao (Tsinghua Shenzhen International Graduate School); Yixuan Zhou (Tsinghua Shenzhen International Graduate School); Zijian Lin (Tsinghua Shenzhen International Graduate School); Zhiyong Wu (Tsinghua Shenzhen International Graduate School)

16:00-16:15

Sun-A-O-6-2: Pre-trained Semantic Enhancement and Multi-Scale Context Modeling for Low-Resource Agglutinative Language TTS

Bowen Li (Xinjiang University); Zhihua Huang (Xinjiang University)*; Wenjie Su (Xinjiang University); Bixin Wu (Xinjiang University)

16:15-16:30

Sun-A-O-6-3: A Training-Data-Free Timing Interface Between Frozen MT and TTS for Low-Resource Dubbing

张晖 (内蒙古大学)*; 王俊男 (内蒙古大学); 潘佳慧 (内蒙古大学); 飞龙 (内蒙古大学)

16:30-16:45

Sun-A-O-6-4: DeepASMR-NSpeech: A Fine-Grained Benchmark for Non-Speech ASMR Generation

ziyi yang (shanghai jiao tong university)*; leying zhang (shanghai jiao tong university); chenda li (shanghai jiao tong university); yanmin qian (shanghai jiao tong university)

16:45-17:00

Sun-A-O-6-5: OOD-VoiceBench: Benchmarking Prompt-Based Voice Design under Real-User Out-of-Distribution Prompts

Xiaosu Su (Institute of Information Engineering, Chinese Academy of Sciences)*; yun cao (Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China, School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100085, China); xiaowei yi (Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China, School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100085, China)

17:00-17:15

Sun-A-O-6-6: VSR-Gen: Controllable Multimodal Persona Synthesis for Scalable Visual Speech Recognition Data Generation

Chang Sun (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications)*

Sun-A-P-2: Robust Speech Processing, Speaker Modeling and Audio Security**Room: 2D 会议室 + 2R 会议室 (海报区)****Type: Poster****November 8, 3:45 PM to 5:15 PM****1. Predictive and Generative Speech Enhancement (11 papers)****Sun-A-P-2-1: URSE: VAD-Guided Unsupervised Robust Speech Enhancement**

Zhenlong Guo (Inner Mongolia University)*; shulin he (Inner Mongolia University); Jiahui Pan (Inner Mongolia University); Xueliang Zhang (Inner Mongolia University)

Sun-A-P-2-2: On the Effect of Training Targets in Speech Enhancement under Different Model Capacities and Computational Constraints

Zixuan Xue (Inner Mongolia University)*; Hongtao Bao (Inner Mongolia University); Wenzheng Zhang (Inner Mongolia University); Xueliang Zhang (Inner Mongolia University)

Sun-A-P-2-3: State Space and Attention Cross-Path Interaction Network for Monaural Speech Enhancement

Zhongping Huang (Shenzhen University)*; Nengheng Zheng (Shenzhen University)

Sun-A-P-2-4: SE-MoE: Dual-Path Joint Magnitude-Phase Speech Enhancement with Sample-Level Sparse Mixture-of-Experts

Wei Yangguang (Sichuan University); Zhou Leijiyu (Sichuan University); Lin Yi (Sichuan University); Yu Xincheng (Sichuan University)*

Sun-A-P-2-5: UniFlowSE: Unifying Data Prediction and Trajectory Consistency for One-Step Speech Enhancement

Yikai Huang (Tsinghua University)*; Xiang Li (Tsinghua University); Jinjiang Liu (Tsinghua University); Dongfu Song (Tsinghua University); Zhiyong Wu (Tsinghua University); Shiyin Kang (Tsinghua University)

Sun-A-P-2-6: Efficient Generative Speech Enhancement via Rectified Flow with Discriminative Prior

Kefan Lu (Hangzhou Dianzi University)*; Wen Zhang (Hangzhou Dianzi University); Zhen Wang (Hangzhou Dianzi University); Kaiying Wu (Hangzhou Dianzi University); Yang Zhang (Hangzhou Dianzi University); Wenbin Jiang (Hangzhou Dianzi University)

Sun-A-P-2-7: Dual-Branch Speech Enhancement Network via Mel-Spectrogram Generation and Phase Prediction

Yang Zhang (Hangzhou Dianzi University)*; Wenbin Jiang (Hangzhou Dianzi University); Jingrui Qian (Hangzhou Dianzi University); Kaiying Wu (Hangzhou Dianzi University); Wen Zhang (Hangzhou Dianzi University)

Sun-A-P-2-8: 面向远场语音识别的语音增强前端训练配置系统性研究

曾德辉 (shenzhen university)*

Sun-A-P-2-9: A Low-Complexity Speech Enhancement Front-End for Human-Machine Voice Interaction via Dual-Window Subband Hybrid Mamba

Hui Zeng (Shenzhen University); Nengheng Zheng (Shenzhen University)*

Sun-A-P-2-10: SS-FlowSE: Efficient Flow Matching for Speech Enhancement via Mam-ba-DiT Hybrid Architecture

戴炳宇 (陆军工程大学)*

Sun-A-P-2-11: Investigating Architectural Scaling and Learning Paradigms in Generative Speech Enhancement

吴开颖 (杭州电子科技大学); 张杨 (杭州电子科技大学); 张雯 (杭州电子科技大学); 江文斌 (杭州电子科技大学)*

2. Multichannel and Audio-Visual Processing and Localization (5 papers)

Sun-A-P-2-12: Adaptive Distortionless Beamforming with Directional Source Activity Detection for Multi-Speaker Speech Enhancement

Wei Liu (Wuhan University)*; Gongping Huang (Wuhan University); Jinfu Wang (OPPO); Zhitao Liu (OPPO); Xin Liu (OPPO)

Sun-A-P-2-13: Confidence-aware Visual Distillation for Robust Audio-Visual Speech Enhancement

zewen Zhan (Tsinghua University)*; Jinjiang Liu (Tsinghua University); Zefeng Zhan (South China Agricultural University); Hui Lu (The Chinese University of Hong Kong); Huimeng Wang (The Chinese University of Hong Kong); Yikai Huang (Tsinghua University); Shiyin Kang (SenseTime Research); Zhiyong Wu (Tsinghua University)

Sun-A-P-2-14: Head Microphone Awakening for Geometry-Agnostic Speech Enhancement Using Distributed Microphone Arrays

chengqian Jiang (USTC)*; Jie zhang (USTC); Haowei Chen (USTC); Haoyin Yan (USTC)

Sun-A-P-2-15: Class-Balanced Dynamic Curriculum Learning for Sound Event Localization and Detection

Yufei Lu (Tianjin University)*; Jianguo Wei (Tianjin University); Wenhuan Lu (Tianjin University)

Sun-A-P-2-16: Rotated Orthogonal Dipole: Exploiting Robot Ego-Motion for Robust DOA Estimation

Feng Guo (Shanghai University)*; Luyuan Xie (Beijing University of Civil Engineering and Architecture); Yuhang Cao (Microsoft (China) Co., Ltd); Qin Zhao (Chinese Academy of Sciences)

3. Speech Separation and Target Speaker Extraction (8 papers)

Sun-A-P-2-17: Target Speaker Extraction Robust to Distance-Cue Errors: Tolerance-Conditioned Training for Co-Directional Talkers

Zhengqing Niu (Tianjin University)*; Longbiao Wang (Tianjin University)

Sun-A-P-2-18: DiS-TSE: Spatial-Cue-Preserving Multichannel-to-Multichannel Target Speech Extraction with Direction Clues

Wen Wen (shanghai Jiao Tong University)*; Changda Chen (waseda university); Qiang Zhou (AISpeech Ltd.); Bohan Li (shanghai jiaotong university); Yujie Xiong (Center of Auditory and Language Intelligence, Soochow University; AISpeech Ltd.); Guanyu Chen (Shanghai jiaotong university); Xiaoyu Gu (Shanghai jiaotong university); Jing Peng (Shanghai jiaotong university); Xin Guo (AISpeech Ltd.); Yu Xi (Shanghai jiaotong university); Shuai Wang (Nanjing University); Kai Yu (Shanghai jiaotong university)

Sun-A-P-2-19: Bridging Speech Separation and Target Speech Extraction: A Universal Model with Encoder-Decoder Attractors

Xin Zhou (Shanghai Jiao Tong University)*; Chenda Li (Shanghai Jiao Tong University); Hongyu Wang (Shanghai Jiao Tong University); Wangyou Zhang (Shanghai Jiao Tong University); Shuai Wang (Nanjing University); Qiang Zhou (AISpeech Ltd); Yanmin Qian (Shanghai Jiao Tong University)

Sun-A-P-2-20: Multi-Scale Local Temporal Enhancement for Separator-Aware Speech Separation

Linhui Sun (Nanjing University of Posts and Telecommunications)*; Weiqi Wang (Nanjing University of Posts and Telecommunications); Xiaolong Zhou (Nanjing University of Posts and Telecommunications); Yanping Li (Nanjing University of Posts and Telecommunications); WenMing Gui (Jinling Institute of Technology); Xi Shao (Nanjing University of Posts and Telecommunications)

Sun-A-P-2-21: A Dual-Path Efficient EEG Encoder for Brain-Assisted Target Speaker Extraction

wang xiang (wuhan university)*; Xue Zhang (Anhui University); Bang Zeng (Wuhan university); Cunhang Fan (Anhui University); Juan Liu (Wuhan university); Ming Li (The Chinese University of Hong Kong(Shenzhen))

Sun-A-P-2-22: MuseFlow-TSE: Music-Aware Target Speech Extraction via Cross-Source Band-Splitting and Data-to-Data Flow Matching

Lu Ma (JD.com, Inc.)*; Yongliang Wang (JD.com, Inc.); Zhewen Su (JD.com, Inc.); Xin Tian (JD.com, Inc.)

Sun-A-P-2-23: Ultrasound-Conditioned Target Speech Extraction via Articulatory Activity Ranking

Jiaao Zhao (Tianjin University)*; Hao Meng (Tianjin University); Guoning Ma (Tianjin University); Qin Wang (The First Affiliated Hospital of Anhui Medical University, 218 Jixi Road); Junhai Xu (Tianjin University)

Sun-A-P-2-24: Gender Composition Matters: A Mixture-of-Experts Approach to Target Speaker Extraction

Zihan Qian (Nanjing University)*; Haoyu Li (Nanjing University); Shuai Wang (Nanjing University)

4. Echo Cancellation, Dereverberation and Packet Restoration (5 papers)

Sun-A-P-2-25: Why Not Put a Microphone Near the Loudspeaker? A New Approach to Acoustic Echo Cancellation

Fei Zhao (School of Computer Science, Inner Mongolia University)*; Xueliang Zhang (School of Computer Science, Inner Mongolia University); Zhongqiu Wang (Southern University of Science and Technology)

Sun-A-P-2-26: 面向低资源场景的端到端回声消除网络

洪逸伟 (厦门大学)*; 詹宗毅 (厦门大学); 翟树盛 (厦门大学); 高艺杰 (厦门大学); 魏志鹏 (厦门大学); 潘泓辰 (厦门大学); 洪青阳 (厦门大学); 李琳 (厦门大学)

Sun-A-P-2-27: 全局时序动态融合的语音丢包恢复方法

马怡静 (陆军工程大学)*

Sun-A-P-2-28: Investigating Blind Multichannel Identification Algorithms and Their Application for Speech Dereverberation Using MINT

Qiang Zhou (Shanghai Jiao Tong University; AISpeech Ltd.); Yujie Xiong (Center of Auditory and Language Intelligence, Soochow University; AISpeech Ltd.)*; Xin Guo (AISpeech Ltd.); Wen Wen (Shanghai Jiao Tong University); Mengyao Zhu (Center of Auditory and Language Intelligence, Soochow University)

Sun-A-P-2-29: Hard-Partitioned Mixture-of-Experts with Adaptive Cross-Correlation for Robust Delay Estimation in Acoustic Echo Cancellation

Lu Ma (JD Content Group)*; Yongliang Wang (JD Content Group); Zhewen Su (JD Content Group); Xin Tian (JD Content Group)

5. Sound Event and Acoustic Anomaly Analysis (7 papers)

Sun-A-P-2-30: A Guided Residual Enhancement Network for Anomalous Sound Detection

Jiyu Lu (Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, CAS; Institute of Information Engineering, CAS; School of Cyber Security, University of CAS)*; Jiale Li (Hubei University); Yanhui Tong (Hubei University); Ming Zhang (Hubei University); Meilin Wan (Hubei University)

Sun-A-P-2-31: MG-DBTFNet: A Mask-Guided Dual-Branch Time-Frequency Fusion Network for Environmental Sound Classification

Xiang Sun (ZhengZhou University)*; Jiaqi Li (ZhengZhou University); Fangjie Wan (ZhengZhou University); Huawei Song (ZhengZhou University)

Sun-A-P-2-32: State Space Model with Frequency-Aware Representation Learning for Underwater Acoustic Target Recognition

Mingyue He (National University of Defense Technology)*; Qisheng Xu (National University of Defense Technology); Boda Zhou (National University of Defense Technology); Xiaoyi Tan (National University of Defense Technology); Yutao Dou (Hunan University); Kele Xu (National University of Defense Technology)

Sun-A-P-2-33: A Lightweight Wind-Aware Spectral Adapter for Sound Event Classification in Wind-Noise Environments

cui zq (beijing institute of technology)*; Zhiqi Song (China Tower Corporation Limited); Xiang Xie (Beijing Institute of Technology, Zhuhai); Chuizheng Kong (Beijing Institute of Technology, Zhuhai); Yan Liu (Beijing Institute of Technology, Zhuhai); Mingqiang Yuang (China Tower Corporation Limited); Donghui Wang (China Tower Corporation Limited); Kuo Zhang (China Tower Corporation Limited)

Sun-A-P-2-34: Efficient Heterogeneous Knowledge Distillation for Anomalous Sound Detection

Shangkai Zhao (Beijing Institute of Technology)*; Zhiqi song (China Tower Corporation Limited); Xiang Xie (Beijing Institute of Technology); Junjie Wang (Beijing Institute of Technology); Mingqiang Yuan (China Tower Corporation Limited); Donghui Wang (China Tower Corporation Limited); Kuo Zhang (China Tower Corporation Limited)

Sun-A-P-2-35: Domain-Gated Statistical Energy Difference Strategy for Anomalous Sound Detection

Shiheng Zhang (Harbin Engineering University); Feiyang Xiao (Harbin Engineering University); Wenbo Wang (Harbin Engineering University)*; Qiaoxi Zhu (University of Technology Sydney); Jian Guan (Harbin Engineering University)

Sun-A-P-2-36: MSTFA-SEUNet: A Sound Event Detection Network for Domestic Environments

chaofeng lan (Harbin University of Science and Technology)*; yuan gao (Harbin University of Science and Technology); lihua zhao (Harbin University of Science and Technology); yingqi chen (Harbin University of Science and Technology); rui guo (Harbin University of Science and Technology); xuemeng jin (Harbin University of Science and Technology)

6. Speaker Representations, Verification and Personalization (11 papers)

Sun-A-P-2-37: Interpretability-Based Speaker Disentanglement of Speech Pre-trained Models

Junhua Li (China National Aviation Fuel Supply Corporation Ltd.)*; Xiaoxu Zhu (XPENG Motors); Aaron J. Li (University of California, Berkeley)

Sun-A-P-2-38: StagnationAware Two Stage SelfDistillation for Speaker Representation Learning

Wenwen Li (Harbin University of Science and Technology); Chen Chen (Harbin University of Science and Technology)*; Dongyuan Li (Harbin University of Science and Technology); Hailu Yang (Harbin University of Science and Technology); Lili Wang (Harbin University of Science and Technology)

Sun-A-P-2-39: Voice-Keyed Personalized Agent Routing with Lightweight Dual-View Speaker-Phrase Embeddings

CHEN FU (China Energy Digital Intelligence Technology Development (Beijing) Co., Ltd.)*; HONGBIN LIANG (Chn Energy Digital Intelligence Technology Development (Beijing) Co., Ltd.); Haodian Wang (Chn Energy Digital Intelligence Technology Development (Beijing) Co., Ltd.); JING WANG (Chn Energy Digital Intelligence Technology Development (Beijing) Co., Ltd.); LILIN SUI (Chn Energy Digital Intelligence Technology Development (Beijing) Co., Ltd.)

Sun-A-P-2-40: Statistics-Aware Fusion Adapter Based on Combined Metric Learning for Cross-Device Speaker Verification

Ken Cheng (Nanjing University of Posts and Telecommunications); Xuechen Liu (Xi'an Jiaotong-Liverpool University); Xi Shao (Nanjing University of Posts and Telecommunications)*

Sun-A-P-2-41: A Lightweight Overlapped Speech Detection Model for Edge Devices

Jacky Gao (Tencent Chengdu Branch)*

Sun-A-P-2-42: SynthWake: LLM-Based Speech Synthesis for Data-Efficient Wake-Word Verification

Zhuojun Wu (Wuhan University); Dong Liu (Wuhan University); Juan Liu (Wuhan University); Lin Zheng (AI Center, OPPO); Xiangshi Chang (AI Center, OPPO); Ming Li (The Chinese University of Hong Kong, Shenzhen)*

Sun-A-P-2-43: Dynamic Teacher Sampling and Matching for Multi-teacher Knowledge Distillation

Jin Li (University of Science and Technology of China); Lirong Dai (University of Science and Technology of China); Tian Gao (Anhui SparkShield Intelligent Technology Co., Ltd); Junhua Liu (Anhui SparkShield Intelligent Technology Co., Ltd); Xin Fang (University of Science and Technology of China)*

Sun-A-P-2-44: Auditing Textual-Attribute Explanations for Chinese Speaker Verification

Haobo Wang (Nanjing University of Posts and Telecommunications)*; Zhuochen Hou (Nanjing University of Posts and Telecommunications); Yang Xu (Nanjing University of Posts and Telecommunications); Yiqiang Cai (Xi'an Jiaotong-Liverpool University); Xuechen Liu (Xi'an Jiaotong-Liverpool University); Xi Shao (Nanjing University of Posts and Telecommunications)

Sun-A-P-2-45: Synergizing Causally Independent Acoustic Features with Deep Embeddings for Physiology-Grounded Speaker Verification

Jining Xu (Xiamen University)*; Hongchen Pan (Xiamen University); Qingyang Hong (Xiamen University); Lin Li (Xiamen University)

Sun-A-P-2-46: 声纹注册样本选择研究

Zhai Shusheng (厦门大学)*

Sun-A-P-2-47: Listen to Detect: Onset-Prompted Personalized Voice Activity Detection

Zixuan Cui (Southern University of Science and Technology)*; Pengjie Shen (Inner Mongolia University); Jiacheng Li (Southern University of Science and Technology); Zhong-Qiu Wang (Southern University of Science and Technology)

7. Spoofing, Deepfake Detection and Voiceprint Security (11 papers)

Sun-A-P-2-48: Controllable Generative Speech Enhancement via ODE Truncation: A Case Study on Audio Deepfake Detection

杨昕宇 (Beijing Institute Of Technology)*

Sun-A-P-2-49: Anchor and Complement: CKA-Guided Layer Selection for Generalizable Audio Deepfake Detection

Rui Wang (Beijing University of Posts and Telecommunications)*; Gang Liu (Beijing University of Posts and Telecommunications)

Sun-A-P-2-50: Tree-Structured Spectral-Temporal Fusion for Spoofed Speech Detection

Yuhongxu Feng (Harbin University of Science and Technology); Chen Chen (Harbin University of Science and Technology)*; Yanjing Wang (Harbin University of Science and Technology); Fengwei Gu (Harbin University of Science and Technology); Lili Wang (Harbin University of Science and Technology)

Sun-A-P-2-51: Source Speaker Tracing for Forged Speech: A Chinese Voice Infringement Perspective

Puyang Geng (Academy of Forensic Science)*; Qimeng Lu (Academy of Forensic Science); Hong Guo (Academy of Forensic Science)

Sun-A-P-2-52: Audio Spoofing Detection Using Multi-Layer Feature Fusion and Attentive Statistics Pooling

Zhanhang Qiu (South China University of Technology); Yanxiong Li (South China University of Technology)*; Qianqian Li (South China University of Technology); Guoqing Chen (South China University of Technology)

Sun-A-P-2-53: CroMMAM-DF: Cross-modal Tri-domain Masked Modeling for Deepfake Detection

Jianbin Ye (National University of Defense Technology)*; Man Xiao (National University of Defense Technology); Bo Liu (Academy of Military Science); Huaping Hu (National University of Defense Technology); Kele Xu (National University of Defense Technology)

Sun-A-P-2-54: ERC-Conformer: Expanded Residual Connections for Robust Audio Deepfake Detection

Yaxuan Qiu (Xinjiang University)*; Zhe Li (The University of Hong Kong); Mieradilijiang Maimaiti (Xinjiang University); Wushour Silamu (Xinjiang University)

Sun-A-P-2-55: Spectral-Aware WavLM for Robust Audio Deepfake Detection under Noise and Propagation Degradations

laisheng kou (Shanghai Jiao Tong University)*

Sun-A-P-2-56: 多表征音频大模型重放检测

Yijie GAO (XMUSpeech)*

Sun-A-P-2-57: Voiceprint Recognition Based on Anti-Spoofing Feature AMGCC and Robust Transformer with Multi-Head Collaborative Attention

Xian-Xian Liu (university of Macau)*; Simon Fong (University of Macau)

Sun-A-P-2-58: DOSS-PERT: Data-Centric and Perturbation-Robust Training for Media-Transformed Speech Deepfake Detection

YUCHEN MAO (Shanghai Jiao Tong University)*; WEN HUANG (Shanghai Jiao Tong University); HOIJUN HUANG (Jiangsu University of Science and Technology); QIANG ZHOU (Shanghai Jiao Tong University, AISpeech); YANMIN QIAN (Shanghai Jiao Tong University)

Sun-A-SS-3: 特殊议题 11: 2026 年 CCF 先进音频技术竞赛

Room: 沐辉厅 C

Type: Special Session

November 8, 2026, 3:45 PM to 5:15 PM

Organizer: 卜辉老师, 凌震华老师, 黎椿键老师

Sun-A-SS-3-1: 基于时间因果建模与多目标优化的轻量音乐修复

Presenter: 刘俊哲 (苏州大学)

Sun-A-SS-3-2: 声学 with 内容表征协同的生成式音乐修复

Presenter: 丁殿哲 (中国科学院声学研究所)

Sun-A-SS-3-3: Band-MixNet-R4: Low-Complexity Real-World Music Restoration with Teacher-Student Distillation and Hierarchical Routing

Presenter: 邵双润 (中央音乐学院)

Sun-A-SS-3-4: The NPU-ASLP System for Music Restoration in the CCF-AATC 2026 Challenge

Presenter: 朱泽宇 (西北工业大学)

Sun-A-SS-3-5: SCALA: Recorded Music Restoration with Time-Frequency Modeling and Adversarial Supervision

Presenter: 林金逸 (哈尔滨工业大学)

Sun-A-SS-3-6: FLASH-ANC: Feedback-Guided Latent Acoustic-State Hypernetwork for Secondary-Path-Robust Deep Active Noise Control

Presenter: 韩佳星 (南京大学)

Sun-A-SS-3-7: Feedforward-Feedback Hybrid Dual-Branch Fusion For Active Noise Control

Presenter: 南晓峰 (南京大学)

Sun-A-SS-3-8: HybridHyperFIR: Response-Conditioned Dual-Rate Finite Impulse Response Filter Generation for Active Noise Control

Presenter: 张易培 (哈尔滨工业大学 (深圳))

Sun-A-SS-3-9: Multi-Expert Online Routing Active Noise Control System
Presenter: 王虎成（西北工业大学）

Sun-A-SS-3-10: RECR-ANC: Regional FIR Experts with Frozen Causal Routing for Active Noise Control
Presenter: 张虔（大连理工大学）

Sun-A-YF: Young Forum 4
Room: 沐辉厅 A
Type: Young Forum
November 8, 2026, 3:45 PM to 5:15 PM

15:45–16:15
Sun-A-YF-1: 音频推理与拟人情感化语音生成技术进展
Presenter: 刘李老师，香港科技大学（广州）副研究员、博导

16:15–16:45
Sun-A-YF-2: “万卷·丝路”多模态小语种语料库的创新与实践
Presenter: 王潇蔓老师，上海人工智能实验室数据平台中心高级工程师

16:45–17:15
Sun-A-YF-3: 融合语言学先验知识的低资源藏语语音合成优化方法研究
Presenter: 拉巴顿珠老师，西藏大学信息科学技术学院副教授、博士生导师

Sun-CLOSE: 大会闭幕式
Room: 沐辉厅 A+B+C
Type: Closing Ceremony
November 8, 2026, 5:35 PM to 6:00 PM

Organizer List

指导委员会 (Steering Committee)

- 李海洲 - 香港中文大学 (深圳)
- 党建武 - 中国科学院深圳先进技术研究院
- 蒙美玲 - 香港中文大学
- 中村哲 - 香港中文大学 (深圳)

名誉主席 (Honorary Chairs)

- 郑方 - 清华大学
- 张劲松 - 北京语言大学
- 毛启容 - 江苏大学

大会主席 (Conference Chairs)

- 俞凯 - 上海交通大学
- 贾珈 - 清华大学
- 吴志勇 - 清华大学深圳国际研究生院

程序委员会主席 (Program Committee Chairs)

- 董明会 - 深圳河套学院
- 刘瑞 - 内蒙古大学
- 万小军 - 北京大学

组织委员会主席 (Organizing Committee Chairs)

- 武执政 - 香港中文大学 (深圳)
- 钱昆 - 北京理工大学
- 秦霄羽 - 清华大学
- 周雁 - 北京理工大学 (珠海)

工业联络主席 (Industrial Liaison Chairs)

- 朱璇 - 三星电子中国研究院
- 王磊 - 华为新加坡研究所
- 谢磊 - 西北工业大学

宣传主席 (Publicity Chairs)

- 刘李 - 香港科技大学 (广州)
- 钱彦旻 - 上海交通大学
- 陈方玉 - 新加坡科技研究局

教育科普主席 (Education and Outreach Chairs)

- 王东 - 清华大学
- 卜辉 - 语音之家
- 李克之 - 伦敦大学学院

出版主席 (Publication Chairs)

- 李雅 - 北京邮电大学
- 吴锡欣 - 香港中文大学
- 蔡思祺 - 哈尔滨工业大学 (深圳)

教程报告主席 (Tutorial Chairs)

- 雪巍 - 香港科技大学
- 王帅 - 南京大学
- 曹梦雪 - 中国社会科学院语言研究所

特殊议题主席 (Special Session Chairs)

- 范存航 - 安徽大学
- 邵曦 - 南京邮电大学
- 王甦菁 - 中国科学院心理研究所

青年论坛主席 (Young Forum Chairs)

- 欧智坚 - 清华大学
- 李军锋 - 中国科学院声学研究所
- 燕楠 - 中国科学院深圳先进技术研究院

基金分享论坛主席 (Research Funding Forum Chairs)

- 邹月嫻 - 北京大学深圳研究生院
- 陈谐 - 上海交通大学

学生论坛主席 (Student Forum Chairs)

- 洪青阳 - 厦门大学
- 连政 - 同济大学
- 孔秋强 - 香港中文大学

网站主席 (Website Chair)

- 刘晋江 - 清华大学深圳国际研究生院

财务主席 (Finance Chairs)

- 张超 - 清华大学
- 丁惠君 - 深圳大学