



第十九届 全国人机语音通讯学术会议

2024年08月15-18日 新疆·乌鲁木齐

Friday, August 16

[Tutorial][10:00-11:30][多功能厅] Fri-M-T-1 - 基于离散编码的语音合成	1
[Special Session][10:00-11:30][人文韬厅] Fri-M-SS-1 - 中文连续视觉语音识别挑战赛 CNVSRC 2024	1
[Special Session][10:00-11:30][文博厅] Fri-M-SS-2 - 共享语音平等：构音障碍语音识别的创新解决方案	1
[Tutorial][11:45-13:15][多功能厅] Fri-M-T-2 - 少样本低资源下语音识别及翻译技术研究	1
[Special Session][11:45-13:15][人文韬厅] Fri-M-SS-3 - 民族地区儿童国家通用语言的语音研究	1
[Special Session][11:45-13:15][文博厅] Fri-M-SS-4 - 基于大模型提示工程的线索语视频生成与识别研究	2
[Industry Forum][13:15-14:00][多功能厅] Fri-M-IND - Industry Forum 1	3
[Tutorial][15:30-17:00][多功能厅] Fri-A-T-1 - 结合抗卷绕相位预测的语音生成技术	3
[Special Session][15:30-17:00][人文韬厅] Fri-A-SS-1 - 基于脑与类脑听觉的语音信息处理及其应用	3
[Special Session][15:30-17:00][文博厅] Fri-A-SS-2 - 第二届多语言智能语音语言技术研讨会	3
[Tutorial][17:15-18:45][多功能厅] Fri-A-T-2 - 无监督工业异常声音检测技术	4
[Industry Forum][18:45-19:30][多功能厅] Fri-A-IND - Industry Forum 2	4

Saturday, August 17

[Plenary][10:45-11:30][多功能厅] Sat-M-PLEN-1 - 跨语言自然语言理解研究进展	5
[Plenary][11:30-12:15][多功能厅] Sat-M-PLEN-2 - 面向中华有声语言和文化展示的生理几何模型研究	5
[Oral][12:30-14:00][人文韬厅] Sat-M-O-1 - Best Paper Presentation	5

[Oral][12:30-14:00][文博厅] Sat-M-O-2 - Speech Recognition and Security	6
[Poster][12:30-14:00][多功能厅] Sat-M-P-1 - Speech Production, Perception, and Acquisition	6
[Poster][12:30-14:00][多功能厅] Sat-M-P-2 - Automatic Speech Recognition	8
[Poster][12:30-14:00][多功能厅] Sat-M-P-3 - Phonetics, Phonology, and Prosody	9
[Poster][12:30-14:00][多功能厅] Sat-M-P-4 - Speaker and Language Identification and Language Technology	9
[Young Forum][15:30-17:30][多功能厅] Sat-A-YF - Young Forum 1	10
[Oral][17:45-19:15][人文韬厅] Sat-A-O-1 - Speaker Recognition and Speech Signal Analysis	11
[Oral][17:45-19:15][文博厅] Sat-A-O-2 - Speech production, Perception and Acquisition	11
[Poster][17:45-19:15][多功能厅] Sat-A-P-1 - Speech Synthesis and Conversion	12
[Poster][17:45-19:15][多功能厅] Sat-A-P-2 - Speech Coding, Speech Enhancement, and Audio Signal Analysis	14
[Poster][17:45-19:15][多功能厅] Sat-A-P-3 - Multimodal and Others	14
[Poster][17:45-19:15][多功能厅] Sat-A-P-4 - Speech Science	16
Sunday, August 18	
[Plenary][10:00-10:45][多功能厅] Sun-M-PLN-1 - 大模型与语音技术	18
[Plenary][10:45-11:30][多功能厅] Sun-M-PLN-2 - SALMONN: 认知导向的视听觉大语言模型	18
[Oral][11:45-13:15][人文韬厅] Sun-M-O-1 - Miscellaneous: Multimodal System, Language Model, and Other	18
[Oral][11:45-13:15][文博厅] Sun-M-O-2 - Human-Machine Interaction and Sound Event Detection	19
[Student Forum][11:45-13:15][多功能厅] Sun-M-SF - Student Forum 1	19
[Industry Forum][13:15-13:55][多功能厅] Sun-M-IND - Industry Forum 3	20
[Oral][15:30-17:00][人文韬厅] Sun-A-O-1 - Speech Synthesis, Sound Synthesis and Pectorophony	21
[Oral][15:30-17:00][文博厅] Sun-A-O-2 - Phonetics, Phonology and Prosody	21
[Student Forum][15:30-16:50][多功能厅] Sun-A-SF - Student Forum 2	22
[Young Forum][17:15-18:45][多功能厅] Sun-A-YF - Young Forum 2	23
Organizer List	24

Friday, August 16

Fri-M-T-1: 基于离散编码的语音合成

Room: 多功能厅

Type: Tutorial

August 16, 10:00 AM to 11:30 AM

Presenter: 俞凯, 刘树杰, 陈诺, 杜晨鹏, 郭奕玮

Chair: 韩纪庆

Fri-M-SS-1: 中文连续视觉语音识别挑战赛 CNVSRC 2024

Room: 人文韬厅

Type: Special Session

August 16, 10:00 AM to 11:30 AM

Presenter: 王东, 李蓝天

Chair: 王东, 李蓝天

10:00-11:30

Fri-M-SS-1-1: 中文连续视觉语音识别挑战赛 CNVSRC 2024

Presenter: 王东 (清华大学), 李蓝天 (北京邮电大学)

Fri-M-SS-2: 共享语音平等: 构音障碍语音识别的创新解决方案

Room: 文博厅

Type: Special Session

August 16, 10:00 AM to 11:30 AM

Presenter: 王甦菁, 杜俊

Chair: 王甦菁, 杜俊

10:00-11:30

Fri-M-SS-2-1: 共享语音平等: 构音障碍语音识别的创新解决方案

Presenter: 王甦菁 (中国科学院心理研究所), 杜俊 (中国科学技术大学)

Fri-M-T-2: 少样本低资源下语音识别及翻译技术研究

Room: 多功能厅

Type: Tutorial

August 16, 11:45AM to 1:15 PM

Presenter: 屈丹

Chair: 韩纪庆

Fri-M-SS-3: 民族地区儿童国家通用语言的语音研究

Room: 人文韬厅

Type: Special Session

August 16, 11:45 AM to 1:15 PM

Presenter: 李爱军

Chair: 李爱军

Fri-M-SS-3-1: 白语学前儿童汉语声调习得研究

Presenter: 刘增会 (云南大学)

Fri-M-SS-3-2: 4-11 岁汉语和苗语儿童声调习得

Presenter: 王蓓, 陈绍艳 (北京理工大学)

Fri-M-SS-3-4: 布朗语学龄儿童国家通用语声调习得研究

Presenter: 郭江华 (云南大学)

Fri-M-SS-3-6: 新疆学前儿童普通话数据资源建设 (A Speech Corpus of Putonghua-Learning Preschoolers From the Uygur Ethnic Group in South Xinjiang Uygur Autonomous Region)

Presenter: 高军 (中国社会科学院语言所)

Fri-M-SS-3-7: 新疆学前儿童普通话发音评测 UY/CH-CHILD——一个公开的维吾尔族儿童汉语第二语言语音数据库

Presenter: 毛丽旦·尼加提 (新疆大学)

Fri-M-SS-3-8: 南疆学前儿童通用语声调感知 (Tone Perception by Putonghua-Learning Preschool Children in South Xinjiang Uyghur Autonomous Region)

Presenter: 李爱军, 王志伟, 周鑫, 高军 (中国社会科学院语言所)

Fri-M-SS-3-9: 不同语言经验对儿童语音意识发展的影响——一项来自维吾尔族双语儿童与汉族单语儿童的对比研究

Presenter: 韦晓保 (华南理工大学)

Fri-M-SS-4: 基于大模型提示工程的线索语视频生成与识别研究

Room: 文博厅

Type: Special Session

August 16, 11:45 AM to 1:15 PM

Presenter: 刘李

Chair: 刘李

11:45-12:15

Fri-M-SS-4-1: 手部残障人士使用键盘习惯的研究

Presenter: 王甦菁 (中国科学院心理研究所)

12:15-12:45

Fri-M-SS-4-2: Empowering Speech and Healing with AI: From Sound to Soul

Presenter: 雪巍 (香港科技大学)

12:45-13:15

Fri-M-SS-4-3: 基于大模型提示工程的线索语视频生成与识别研究

Presenter: 刘李 (香港科技大学-广州)

13:15-13:45

Fri-M-SS-4-4: 发音反演研究及听障儿童语言学习初探

Presenter: 高迎明 (北京邮电大学)

Fri-M-IND: Industry Forum 1

Room: 多功能厅

Type: Industry Forum

August 16, 1:15 PM to 2:00 PM

Chair: 谢磊

13:15-13:45

Fri-M-IND-1: 基于多模态 Agent 智能体的增强人机交互产业化

Presenter: 邢启洲 (捷通华声)

13:45-14:00

Fri-M-IND-2: 对话式人工智能技术及其产业应用

Presenter: 樊帅 (思必驰)

Fri-A-T-1: 结合抗卷绕相位预测的语音生成技术

Room: 多功能厅

Type: Tutorial

August 16, 3:30 PM to 5:00 PM

Presenter: 艾杨

Chair: 张卫强

Fri-A-SS-1: 基于脑与类脑听觉的语音信息处理及其应用

Room: 人文韬厅

Type: Special Session

August 16, 3:30 PM to 5:00 PM

Presenter: 范存航, 张结

Chair: 范存航, 张结

15:30-17:00

Fri-A-SS-1-1: 基于脑与类脑听觉的语音信息处理及其应用

Presenter: 范存航 (安徽大学), 张结 (中国科学技术大学)

Fri-A-SS-2: 第二届多语言智能语音语言技术研讨会**Room:** 文博厅**Type:** Special Session**August 16, 3:30 PM to 5:00 PM****Presenter:** 黄浩, 欧智坚**Chair:** 黄浩, 欧智坚

15:30-17:00

Fri-A-SS-2-1: 第二届多语言智能语音语言技术研讨会**Presenter:** 黄浩 (新疆大学), 欧智坚 (清华大学)**Fri-A-T-2: 无监督工业异常声音检测技术****Room:** 多功能厅**Type:** Tutorial**August 16, 5:15 PM to 6:45 PM****Presenter:** 关键**Chair:** 张卫强**Fri-A-IND: Industry Forum 2****Room:** 多功能厅**Type:** Industry Forum**August 16, 6:45 PM to 7:30 PM****Chair:** 谢磊

18:45-19:15

Fri-A-IND-1: 讯飞语音技术最新进展**Presenter:** 熊世富 (科大讯飞)

19:15-19:30

Fri-A-IND-2: 小布助手语音交互中的个性化技术**Presenter:** 单煜翔 (OPPO)

Saturday, August 17

Sat-M-PLN-1: 跨语言自然语言理解研究进展

Room: 多功能厅

Type: Plenary

August 17, 10:45 AM to 11:30 AM

Presenter: 吾守尔·斯拉木

Chair: 郑方

Sat-M-PLN-2: 面向中华有声语言和文化展示的生理几何模型研究

Room: 多功能厅

Type: Plenary

August 17, 11:30 AM to 12:15 AM

Presenter: 孔江平

Chair: 张劲松

Sat-M-O-1: Best Paper Presentation

Room: 人文韬厅

Type: Oral

August 17, 12:30 AM to 2:00 PM

Chair: 张劲松 (北京语言大学)

Sat-M-O-1-1: Investigation into the Impact of Speaker Adversarial Perturbation on Speech Recognition

Chenyang Guo (University of Science and Technology of China)*; Liping chen (University of Science and Technology of China); Kong Aik Lee (The Hong Kong Polytechnic University); Zhen-Hua Ling (University of Science and Technology of China); Wu Guo (University of Science and Technology of China)

Sat-M-O-1-2: Paraformer-v2: An improved single step non-autoregressive transformer for noise-robust speech recognition

Keyu An (alibaba)*; Zerui Li (alibaba); Shiliang Zhang (Alibaba Group)

Sat-M-O-1-3: Pruning and Quantization Enhanced Densely Connected Neural Network for Efficient Acoustic Echo Cancellation

Chen Chen (Institute of Acoustics, Chinese Academy of Sciences; University of Chinese Academy of Sciences); Sheng Yan (Institute of Acoustics, Chinese Academy of Sciences; University of Chinese Academy of Sciences)*; Chengpeng Hao (Institute of Acoustics, Chinese Academy of Sciences; University of Chinese Academy of Sciences)

Sat-M-O-1-4: 基于双对齐和对比学习的多模态情感识别

Yuhua Wen (Beijing University of Posts and Telecommunications)*; Qifei Li (Beijing University of Posts and Telecommunications); Yingying Zhou (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications); Ya Li (Beijing University of Posts and Telecommunications)

Sat-M-O-1-5: A Speech Corpus of Putonghua-Learning Preschoolers From the Uygur Ethnic Group in South Xinjiang Uygur Autonomous Region

Jun Gao (Institute of Linguistics, Chinese Academy of Social Sciences)*; Ziyu Xiong (CASS); Aijun LI ("Institute of Linguistics, CASS")

Sat-M-O-1-6: 基于注意力机制和数据过采样的酒瓶裂纹敲击异常声音检测系统

Mou Wang (OPPO)*; Jisheng Bai (Northwestern Polytechnical University); Siwei Huang (The 34th Research Institute of China Electronics Technology Group Corporation); Zhuo Li (OPPO); Xin Liu (OPPO); Feiren Yang (Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences); Ziteng Wang (OPPO)

Sat-M-O-2: Speech Recognition and Security

Room: 文博厅

Type: Oral

August 17, 12:30 AM to 2:00 PM

Chair: 杜俊 (中国科学技术大学)

Sat-M-O-2-1: Multilingual Speech Recognition Using Discrete Tokens with a Two-step Training Strategy

Zehan Li (Institute of Artificial Intelligence (TeleAI), China Telecom, China)*; Zhang XiaoLei (Northwestern Polytechnical University); Jie Li (Institute of Artificial Intelligence (TeleAI), China Telecom, China); Jian Kang (Institute of Artificial Intelligence (TeleAI), China Telecom, China); Yan Yang (Institute of Artificial Intelligence (TeleAI), China Telecom, China); Xueqing Li (Northwestern Polytechnical University)

Sat-M-O-2-2: Amphion: An Open-Source Audio, Music and Speech Generation Toolkit

Xueyao Zhang (The Chinese University of Hong Kong, Shenzhen); Liumeng Xue (The Chinese University of Hong Kong, Shenzhen); Yicheng Gu (The Chinese University of Hong Kong, Shenzhen); Yuancheng Wang (The Chinese University of Hong Kong, Shenzhen); Jiaqi Li (The Chinese University of Hong Kong, Shenzhen); Haorui He (The Chinese University of Hong Kong, Shenzhen); Chaoren Wang (The Chinese University of Hong Kong, Shenzhen); Xi Chen (Chinese University of Hong Kong (Shenzhen)); Junan Zhang (Fudan University); Zihao Fang (The Chinese University of Hong Kong, Shenzhen); Haopeng Chen (The Chinese University of Hong Kong, Shenzhen); Yze Ying Tang (The Chinese University of Hong Kong, Shenzhen); Lexiao Zou (Harbin Institute of Technology (Shenzhen)); Mingxuan Wang (The Chinese University of Hong Kong, Shenzhen); Jun Han (The Chinese University of Hong Kong, Shenzhen); Kai Chen (Shanghai AI Laboratory); Haizhou Li (The Chinese University of Hong Kong, Shenzhen); Zhizheng Wu (Chinese University of Hong Kong, Shenzhen)*

Sat-M-O-2-3: DA-KWFormer: A Domain Adaptation Network with K-Weight Transformer for Speech Emotion Recognition

fujie xu (Xinjiang University)*; Huamin Yang (Xinjiang University); Ying Hu (Xinjiang University)

Sat-M-O-2-4: Phoneme Semantic Backdoor Attacks with Multiple Task Learning for Speech Classification Task

Ye Xiao (XiangTan University)*; Wenhan Yao (Xiangtan University); Zexin Li (XiangTan University); Jiangkun Yang (XiangTan University); Weiping Wen (Peking University)

Sat-M-O-2-5: Are Transformers in pretrained LM A Good ASR Encoder? An Empirical Study

Keyu An (alibaba)*; Shiliang Zhang (Alibaba Group); Zhijie Yan (Alibaba Inc.)

Sat-M-O-2-6: 基于尖峰特征的口音识别和语音识别多任务学习方法

Sujie Xiao (University of Chinese Academy of Sciences; Institute of Acoustics, Chinese Academy of Sciences)*; ta li (Institute of Acoustics, Chinese Academy of Sciences); Ruipeng Hao (Institute of Acoustics, Chinese Academy of Sciences); Duojie Renzeng (Tibet University, School of Information Science and Technology); Qingwei Zhao (Institute of Acoustics, Chinese Academy of Sciences)

Sat-M-P-1: Speech Production, Perception, and Acquisition**Room:** 多功能厅**Type:** Poster**August 17, 12:30 AM to 2:00 PM****Sat-M-P-1-1: 基于音高线索的普通话对比焦点感知研究——以阳平、去声为例**

璐妍李 (南开大学)*; 宝莹温 (南开大学)

Sat-M-P-1-2: A Comparative Analysis of Diphthong Acquisition in Standard Chinese by Learners from ‘the Belt and Road’

Yuan Jia (Institute of Linguistics, Chinese Academy of Social Sciences); Linjiao Pan (Peking University)*

Sat-M-P-1-3: 16 年跨度对普通话青年女声声学特征的影响

Honglin Cao (China University of Political Science and Law)*; Yuping Xiao (中国政法大学)

Sat-M-P-1-4: 中级俄语母语者汉语阳平双字调声学分析

Peining Song (Xinjiang University)*; Wei Wang (Xinjiang University)

Sat-M-P-1-5: 中级水平俄语母语者汉语停延习得研究

Qiao Xv (新疆大学)*

Sat-M-P-1-6: “去 1+VP+ 去 2” 中 “去 2” 的轻化程度及语义虚实分析

Weiqing Li (College of Chinese Language and Culture, Nankai University)*; Baoying Wen (College of Chinese Language and Culture, Nankai University)

Sat-M-P-1-7: A longitudinal study on the acquisition of standard Chinese monophthongs by intermediate- and advanced- level Korean Learners

DAN YUAN (East China Normal University)*; Feng LING (East China Normal University)

Sat-M-P-1-8: Contrastive focus perception in Mandarin from pitch – Take Tone 1 and Tone 3 as an example(Agree to publish/participate in the Best Paper evaluation in the journal)

wenya liu (Nankai University)*

Sat-M-P-1-9: A pilot study on the perception of "dearing" emotional speech

Kun Wang (Linguistic Lab of Peking University)*; Yao Lu (Peking University)

Sat-M-P-1-10: 俄语母语者普通话舌尖元音 [ɿ]、[ʅ] 产出训练研究

凌波姬 (北京语言大学)*; Jinsong Zhang (Beijing Language and Culture University); 玉蓉吴 (北京语言大学)

Sat-M-P-1-11: 儿化感知影响因素研究

芷欣黄 (深圳大学); Li Dong (Shenzhen University)*

Sat-M-P-1-12: 俄罗斯学生普通话辅音 /t/、/tɕ/ 的感知同化与区分

huang heting (Beijing Language and Culture University)*; Jinsong Zhang (Beijing Language and Culture University)

Sat-M-P-1-13: A New Parameter to Indicate the Syllable Stress Levels in Mandarin

Zhigang Yin (Institute of Linguistics, Chinese Academy of Social Sciences)*

Sat-M-P-1-14: 维吾尔语清塞音 k 习得的声学特征研究

宇刘 (新疆大学)*

Sat-M-P-1-15: 俄语母语者汉语同声调句中的声调产出研究

Li Weihao (Xinjiang University)*

Sat-M-P-1-16: 哈萨克斯坦汉语学习者辅音声母习得分析

Yuan Jia (Institute of Linguistics, Chinese Academy of Social Sciences); Zili Wang (Beijing Normal University)*; Jinsong Zhang (Beijing Language and Culture University)

Sat-M-P-1-17: An acoustic study of Mandarin Chinese vowels produced by Uyghur-speaking learners

hou yunhong (Nanjing university)*

Sat-M-P-1-18: 复杂动态系统背景下邹平方言人声调变异速率实证研究

Ying Liu (Qufu Normal University)*; Wei Zhang (Qufu Normal University)

Sat-M-P-1-19: 基于语音病理特征的不流畅语音片段标注系统

Huang Jinfeng (University of Tsukuba)*; Chen Yuan (University of Tsukuba); He Chengqi (University of Tsukuba); Xingfeng LI (Hainan University); Miyamoto Shoko (University of Tsukuba)

Sat-M-P-1-20: Sound Zone Control Based On A Kronecker Second-order Tensor Decomposition

Mao zhien (Northwestern Polytechnical University)*; Wen Zhang (Northwestern Polytechnical University)

Sat-M-P-2: Automatic Speech Recognition

Room: 多功能厅

Type: Poster

August 17, 12:30 AM to 2:00 PM

Sat-M-P-2-1: TeleSpeechPT: Large-Scale Chinese Multi-Dialect And Multi-Accent Speech Pre-Training**Sat-M-P-2-2: AESR: Speech Recognition With Speech Emotion Recognizing Learning**

RongQi Han (Inner Mongolia University)*

Sat-M-P-2-3: CTC-Assisted LLM-Based Contextual ASR

guanrou yang (Shanghai Jiao Tong University)*; Ziyang Ma (Shanghai Jiao Tong University); zhifu gao (alibaba); Shiliang Zhang (Alibaba Group); Xie Chen (Shanghai Jiaotong University)

Sat-M-P-2-4: Enhancing Mispronunciation Detection with Multi-Speaker Text-to-Speech and Mixture-of-Experts Network

wenqian Bao (Beijing Language and Culture University)*; Yan Yuchen (beijing language and culture university); Jinsong Zhang (Beijing Language and Culture University)

Sat-M-P-2-5: 融合语种信息的端到端多语种语音识别方法

yuning wang (Harbin Institute of Technology); jiqing Han (Harbin Institute of Technology)*; 铁然郑 (哈尔滨工业大学); 贵滨郑 (哈尔滨工业大学); Yongjun He (Harbin Institute of Technology)

Sat-M-P-2-6: XCB: an effective contextual biasing approach to bias cross-lingual phrases in speech recognition

Xucheng Wan (Huawei Technologies Co., Ltd.); Naijun Zheng (Huawei Technologies Co., Ltd.); KAI LIU (Huawei Technologies Co., Ltd.); zhou huan (AARC, Huawei Technologies Co., Ltd.)*

Sat-M-P-2-7: 基于注意力机制软切分的发音偏误检测探究

Junyao Fan (Beijing Language and Culture University); Xiaolin Wu (Beijing Language and Culture University); Jinsong Zhang (Beijing Language and Culture University)*

Sat-M-P-2-8: Adaptive Context Biasing in Transformer-based ASR Systems

努尔麦麦提尤鲁瓦斯 (新疆大学); Yineng Cai (Xinjiang University)*; Zhiwu Sun (Xinjiang university); Qiqi Du (Xinjiang University); 吾守尔斯拉木 (新疆多语种信息技术实验室新疆大学信息科学与工程学院)

Sat-M-P-2-9: Mongolian Speech Recognition Based on Semi-Supervised Learning and Syllable Subword Modeling Units**Sat-M-P-2-10: Enhancing Transducers for Robust Keyword Spotting by Duration Modeling**

Xiaoyu Gu (aispeech)*; Haoyu Li (Shanghai Jiao Tong University); Yu Xi (Shanghai Jiao Tong University); Hao Li (AISpeech Ltd, Suzhou China); Kai Yu (Shanghai Jiao Tong University)

Sat-M-P-3: Phonetics, Phonology, and Prosody

Room: 多功能厅

Type: Poster

August 17, 12:30 AM to 2:00 PM

Sat-M-P-3-1: 普通话学习者与母语者的嗓音质量对比分析

Xinyi Zhao (Beijing language and culture university)*; Jinsong Zhang (Beijing Language and Culture University)

Sat-M-P-3-2: 普通话与重庆话长时基频特征的时长阈值研究

艺雷 (中国政法大学); 覃晓婧 (中国政法大学); Honglin Cao (China University of Political Science and Law)*

Sat-M-P-3-3: A Study of Phonetic Differences in Hankou Dialect in the late Qing Dynasty – Based on the records of A Chinese-English Dictionary and The Hankow Syllabary

Tu guoxiang Tu guoxiang (Tu guoxiang)*

Sat-M-P-3-4: 鄂温克语阴阳元音舌根位置的声学表现

Yan Liu (Minzu University of China)*; Qingqing Zhang (Minzu University of China); Hua Chen (Nanjing University)

Sat-M-P-3-5: 俄语母语者汉语朗读流利度评分模型探究

亦宸孙 (新疆大学)*; Wei Wang (International Cultural Exchange College Xinjiang University)

Sat-M-P-3-6: the analysis for vowels in neutral tone syllable

xiaoxia chen (institute of linguistics,cass,china)*

Sat-M-P-4: Speaker and Language Identification and Language Technology**Room:** 多功能厅**Type:** Poster**August 17, 12:30 AM to 2:00 PM****Sat-M-P-4-1: BER: Balanced Error Rate For Speaker Diarization**

Tao Liu (Shanghai Jiao Tong University)*; Shuai Fan (AISpeech Ltd); Kai Yu (Shanghai Jiao Tong University)

Sat-M-P-4-2: Improving Speaker Verification Back-end with Graph Neural Networks

Jinfeng Chen (Xinjiang University); Zhihua Fang (Xinjiang University); Liang HE (Tsinghua University)*

Sat-M-P-4-3: Dual-Path Spectrogram Refinement Network for Robust Speaker Verification

Zonghui Wang (Xinjiang University); Zhihua Fang (Xinjiang University); Liang HE (Tsinghua University)*

Sat-M-P-4-4: 基于解耦学习的鲁棒说话人验证研究**Sat-M-P-4-5: An local-aggregation based method for robust speaker verification**

Wei-Lin Xie (University of Science and Technology of China)*; Yan Song (USTC); Wu Guo (University of Science and Technology of China); Zhongfu Ye (University of Science and Technology of China); Lirong Dai (University of Science and Technology of China)

Sat-M-P-4-6: Demystifying the Robustness of Deep Speaker Recognition Against Non-Speech Segments

Pengqi Li (Xinjiang University)*; Huan Liu (Xinjiang University); Jinfu Li (Xinjiang University); Xueying Liao (Xinjiang University); Askar Rozi (Xinjiang University); Askar Hamdulla (Xinjiang University)

Sat-M-P-4-7: 抑郁症情感表达的多模态生理数据库构建及分析**Sat-M-P-4-8: 面向心音识别的自监督联邦学习****Sat-M-P-4-9: 融合多源知识的文物描述自动生成方法研究**

Chenggang Mi (Xi'an International Studies University)*; Ying (颖) Dong (董) (Xi'an International Studies University); Shaoliang (韶亮) Xie (谢) (Xi'an International Studies University); Yuan (媛) Ma (马) (Xi'an International Studies University)

Sat-M-P-4-10: Analysis and Construction of Corpus Based on Kazakh Text Character Encoding

Jiayina Jianaer (Xinjiang police college)*

Sat-A-YF: Young Forum 1**Room:** 多功能厅**Type:** Young Forum**August 17, 3:30 PM to 5:30 PM****Chair:** 欧智坚**Sat-A-YF-1: 体系化 AI 下的语音技术研究**

Presenter: 张世磊, 中国移动研究院人工智能与智慧运营中心语音技术负责人, 多媒体信息处理全国重点实验室听觉内容分析技术负责人, 中关村试听产业技术创新联盟第三届理事会理事

Sat-A-YF-2: 基于数据的空间音频处理

Presenter: 张雯, 西北工业大学航海学院教授/博士生导师, 国家级青年人才

Sat-A-YF-3: 从数据驱动到意图导向的语音生成方法研究

Presenter: 温正棋, 启元实验室副研究员

Sat-A-YF-4: 面向通用任务的基于大规模数据自监督预训练的声音基础模型

Presenter: 张俊博, 小米声学语音技术专家

Sat-A-O-1: Speaker Recognition and Speech Signal Analysis

Room: 人文韬厅

Type: Oral

August 17, 5:45 PM to 7:15 PM

Chair: 张晓雷 (西北工业大学)

Sat-A-O-1-1: Cross-Model Knowledge Distillation and Metadata Fusion for Respiratory Sound Classification

Zhengyang Sun (Xinjiang university)*; Zhihua Huang (Xinjiang university); xueyuan xu (新疆大学); binyu li (新疆大学)

Sat-A-O-1-2: PadAug: Robust Speaker Verification with Simple Waveform-Level Silence Padding

Zijun Huang (Northwestern Polytechnical University)*; Chengdong Liang (General User Agent); Jiadi Yao (Northwestern Polytechnical University); Zhang XiaoLei (Northwestern Polytechnical University)

Sat-A-O-1-3: 刻意伪装场景下的说话人确认

Xiaoyi Qin (DukeKunshan University)*; Ze Li (Wuhan University); Dong Liu (Duke Kunshan University); Ming Li (Duke Kunshan University)

Sat-A-O-1-4: 面向说话人识别的最近邻惩罚损失函数

Zhuo Li (OPPO)*; Mou Wang (OPPO); Ziteng Wang (OPPO); Xin Liu (OPPO); Feiran Yang (Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences)

Sat-A-O-1-5: Speaker extraction with verification of present and absent target speakers

Ke Zhang (Northeastern University)*; Marvin Borsdorf (University of Bremen); Tianchi Liu (National University of Singapore); Shuai Wang (Shenzhen Research Institute of Big Data, Chinese University of Hong Kong (Shenzhen)); Yangjie Wei (Northeastern University); Haizhou Li (The Chinese University of Hong Kong, Shenzhen)

Sat-A-O-1-6: An Unsupervised Domain Adaptation Method based on Distribution Alignment for Speaker Verification

Qing Gu (University of Science and Technology of China)*; Yan Song (USTC); Wu Guo (University of Science and Technology of China); Zhongfu Ye (University of Science and Technology of China); Lirong Dai (University of Science and Technology of China)

Sat-A-O-2: Speech production, Perception and Acquisition

Room: 文博厅

Type: Oral

August 17, 5:45 PM to 7:15 PM

Chair: 高迎明 (北京邮电大学)

Sat-A-O-2-1: 跨语言音系对比及错误分析研究

Gulnur Arkin (Institute of Public Administration Xinjiang University of Finance and Economics)*

Sat-A-O-2-2: 样本量大小对双元音/ei/话者区分能力评判的影响

Honglin Cao (China University of Political Science and Law)*; Xuehui Li (新兴发展集团有限公司); Danlin Wang (北京警察学院)

Sat-A-O-2-3: A study on the allocations of prosody boundaries in L2 Mandarin speech by Vietnamese learners based on dependency syntax

Yuqi Shi (Beijing Language and Culture University)*; Yukai Wan (Beijing Language & Culture University SAIT Lab); How Yang (Beijing Advanced Innovation Center for Language Resource); Xinyi Zhao (Beijing language and culture university); Jinsong Zhang (Beijing Language and Culture University)

Sat-A-O-2-4: 越南学生汉语-越南语声调的感知同化模式

Min Liu (Jinan University)*; 张 zap 艾平 (暨南大学)

Sat-A-O-2-5: Study on Prosodic Disambiguation of VP/NP syntactic structure by Chinese EFL Learners

Jia Guo (College of Foreign Languages, Nankai University)*; Qiuxiang Ang (NanKai University); Jiachun Liu (NanKai University); Yunyuan Zhang (NanKai University)

Sat-A-O-2-6: 新疆少数民族农牧民的国家通用语言文字声调感知研究

zulihuma tuohutimaiti (Xinjiang University)*

Sat-A-P-1: Speech Synthesis and Conversion

Room: 多功能厅

Type: Poster

August 17, 5:45 PM to 7:15 PM

Sat-A-P-1-1: A Comprehensive Data Processing Pipeline for LLM based Text-to-Speech Models

Runqiang Han (kuaishou); Chen Zhang (北京达佳互联信息技术有限公司)*

Sat-A-P-1-2: Expressive TTS Driven by Natural Language Prompts Using Few Human Annotations**Sat-A-P-1-3: Stage-Wise and Piror-Aware Neural Speech Phase Prediction****Sat-A-P-1-4: Language-independent Prosody-enhanced Speech Representations for Multilingual Speech Synthesis**

Chang Liu (University of Science and Technology of China); Zhen-Hua Ling (University of Science and Technology of China)*; Ya-Jun Hu (iFLYTEK Co., Ltd.)

Sat-A-P-1-5: Analyzing the Improvement of Supplementary Features on Voice Conversion Using Disentangled Representation Learning

Qi Luo (Beijing Language and Cultural University)*; Zimeng Han (Zimeng Han); dengfeng ke (blcu.edu.cn)

Sat-A-P-1-6: MIXDIFF: Mixture Diffusion Model for Efficient Text-to-Speech

Xuankai Wang (Xinjiang university)*; Liejun Wang (Xinjiang University); Yinfeng Yu (Xinjiang University)

Sat-A-P-1-7: LightAC: Lightweight Accent Conversion via Module-Wise Distillation

Mingyang Zhang (Chinese University of Hong Kong, ShenZhen)*; Yi Zhou (National University of Singapore); Haizhou Li (The Chinese University of Hong Kong, Shenzhen)

Sat-A-P-1-8: Tibetan Speech Synthesis Based on Pre-trained Mixture Alignment FastSpeech2

Qing Zhou (Minzu University of China); Xiaona Xu (Minzu University of China)*; Yue Zhao (Minzu University of China)

Sat-A-P-1-9: 基于互信息特征解耦的情感一致性语音转换

Dudu Li (Beijing Forestry University); Shaorong Wang (Beijing Forestry University); Yining Zhen (Beijing Forestry University); Chunling Wang (Beijing Forestry University)*

Sat-A-P-1-10: Burmese Speech Synthesis Based on Diffusion Model

Xiaoyan Zhang (Guangxi Minzu University); Donghong Qin (.); Yu Chen (Guangxi Minzu University); Fengbo BAI (Guangxi Minzu University)*; Yan Liu (Tsinghua University); Zhijian Ou (Tsinghua University)

Sat-A-P-1-11: IUMSS-CETL: Low-Resource Iu Mien Speech Synthesis based on Transfer Learning**Sat-A-P-1-12: Zero-Shot Personalized Voice Synthesis with Cross-Attention Speaker Embeddings**

Kexin Lu (Xinjiang University)*; Zihua Huang (Xinjiang university); Mingming Yin (Xinjiang University); Ke Chen (Xinjiang university)

Sat-A-P-1-13: Domain Adaptation for Front-End Module in TTS with LoRAs

Jie WANG (Xiaomi)*; Binbin Shen (Xiaomi Inc.); Meng Meng (Xiaomi); Yujun Wang (xiaomi)

Sat-A-P-1-14: A Neural Denoising Vocoder for Clean Waveform Generation from Noisy Mel-Spectrogram based on Amplitude and Phase Predictions

Hui-Peng Du (University of Science and Technology of China)*; Ye-Xin Lu (University of Science and Technology of China); Yang Ai (University of Science and Technology of China); Zhen-Hua Ling (University of Science and Technology of China)

Sat-A-P-1-15: 基于神经网络的重音提取及重音描述提示生成

Kehan Li (Tsinghua University); Zeyu Jin (Tsinghua University); Yixuan Zhou (Tsinghua University); Xiaoyu Qin (Tsinghua University); Zhiyong Wu (Tsinghua University)*

Sat-A-P-1-16: Task-Specific Dementia Risk Detection Based on Psycholinguistic Knowledge and Linguistic Features

Mengyue Wu (Shanghai Jiao Tong University)*; Dengdeng Huang (Shanghai Jiao Tong University); 居强陈 (上海交通大学)

Sat-A-P-1-17: 基于计划梯度反转的说话人无关韵律表征研究

Leyuan Qu (Zhejiang Lab)*; Taihao Li (Zhejiang Lab)

Sat-A-P-1-18: Dementia Risk Detection via Text Augmentation-based Multi-Task Learning

Mengyue Wu (Shanghai Jiao Tong University)*; Dengdeng Huang (Shanghai Jiao Tong University)

Sat-A-P-1-19: StyleSVC: Singing Voice Conversion with Multi-Scale Style Transfer**Sat-A-P-1-20: Emotion-Aware Prosodic Phrasing for Expressive Text-to-Speech**

bin liu (Inner Mongolia University, Hohhot, China); Rui Liu (Inner Mongolia University)*; Haizhou Li (The Chinese University of Hong Kong, Shenzhen)

Sat-A-P-2: Speech Coding, Speech Enhancement, and Audio Signal Analysis**Room:** 多功能厅**Type:** Poster**August 17, 5:45 PM to 7:15 PM****Sat-A-P-2-1: MDCTCodec: A Lightweight MDCT-based Neural Audio Codec towards High Sampling Rate and Low Bitrate Scenarios****Sat-A-P-2-2: On the effectiveness of enrollment speech augmentation for Target Speaker Extraction****Sat-A-P-2-3: M-CMGAN: Attempting to Use Mamba on Speech Enhancement**

Yujie Xiong (Xinjiang University); Zhihua Huang (Xinjiang university)*

Sat-A-P-2-4: SESNet: A Speech Enhancement and Separation Network in Noisy Reverberant Environments

Liusong Wang (Xinjiang University)*; Yuan Gao (Xinjiang University); Kaimin Cao (Xinjiang University); Ying Hu (Xinjiang University)

Sat-A-P-2-5: Exploring Discrete Tokens Suitable for Speech Synthesis**Sat-A-P-2-6: ExARN: Target Speaker Extraction with Attentive Recurrent Networks**

Pengjie Shen (Inner Mongolia University)*; Shulin He (College of Computer Science, Inner Mongolia University); Xueliang zhang (Inner Mongolia University)

Sat-A-P-2-7: Data Augmentation and Progressive Learning in Acoustic Echo Cancellation for Duplex conversations on Mobile Devices

Yiheng Jiang (Alibaba Group)*; Biao Tian (-)

Sat-A-P-2-8: D-AGNet: A Dual-branch Network with Attention Guidance for Speech Emotion Recognition

Huamin Yang (Xinjiang University); Ying Hu (Xinjiang University)*

Sat-A-P-2-9: Speech emotion recognition based on multi acoustic feature fusion

shanshan xiang (Xinjiang University)*; Hankiz Yilahun (Xinjiang University); Askar Hamdulla (Xinjiang University)

Sat-A-P-2-10: 基于声学参数的蒙古呼麦情感表现研究

雅馨席 (西北民族大学)*; 惠鹏任 (西北民族大学); 阿旭胡 (西北民族大学)

Sat-A-P-2-11: A Brief Survey on the Explainability for Deep Speech Processing Models

Pengqi Li (Xinjiang University)*; Lantian Li (Beijing University of Posts and Telecommunications); Askar Hamdulla (Xinjiang University); Dong Wang (Tsinghua University)

Sat-A-P-2-12: Efficient Singular Spectrum Mode Ensembler Capable of Extracting Wide-band Components in Spectrum Overlapping Scenarios

zhang hongcheng (天津大学)*; Wenhuan Lu (Tianjin University); Jianguo Wei (School of Computer Software, Tianjin University, Tianjin, China)

Sat-A-P-2-13: 视频信息辅助的歌声旋律提取

Xiaozhen PENG (Dalian Maritime University); Weiwei Zhang (Dalian Maritime University)*

Sat-A-P-3: Multimodal and Others**Room:** 多功能厅**Type:** Poster**August 17, 5:45 PM to 7:15 PM****Sat-A-P-3-1: Baseline Systems for Chinese Continuous Visual Speech Recognition Challenge 2024**

Zehua Liu (Beijing University of Posts and Telecommunications)*; Xiaolou Li (Beijing University of Posts and Telecommunications); Chen Chen (Tsinghua University); Yiwei Sun (Beijing University of Posts and Telecommunications); Yanzhi Shu (Beijing University of Posts and Telecommunications); Yonghao Sheng (Beijing University of Posts and Telecommunications); Shuaiming Meng (Beijing University of Posts and Telecommunications); Lantian Li (Beijing University of Posts and Telecommunications); Dong Wang (Tsinghua University)

Sat-A-P-3-2: MHAN: Bottleneck Fusion Model Based on Hybrid Attention Network for Multimodal Emotion Recognition

Kai Li (Xinjiang University)*; ya huang (Xinjiang University); Gang Zhong (Xinjiang University); 努尔麦麦提尤鲁瓦斯 (新疆大学); 吾守尔斯拉木 (新疆多语种信息技术实验室新疆大学信息科学与工程学院)

Sat-A-P-3-3: Investigation on Training Strategy for Cross-Modal Large Language Models with Speech and Text

Hang Zheng (Shanghai Jiao Tong University)*; Yuncong Liu (Shanghai Jiao Tong University); Hankun Wang (Shanghai Jiao Tong University); Kai Yu (Shanghai Jiao Tong University)

Sat-A-P-3-4: Evaluation of Data Inconsistency for Multi-modal Sentiment Analysis

Yufei Wang (Shanghai Jiao Tong University)*; Mengyue Wu (Shanghai Jiao Tong University)

Sat-A-P-3-5: 面向心理治疗的共情对话系统

Zekai Sun (Beijing University of Posts and Telecommunications)*; Keqi Chen (Beijing University of Posts and Telecommunications); Huijun Lian (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications); Ya Li (Beijing University of Posts and Telecommunications)

Sat-A-P-3-6: Towards Reliable and Empathetic Depression-Diagnosis-Oriented Chats

Mengyue Wu (Shanghai Jiao Tong University)*; Kunyao Lan (Shanghai Jiao Tong University); Lu Chen (Shanghai Jiao Tong University); Cong Ming (Shanghai Jiao Tong University); Binwei Yao (University of Wisconsin-Madison)

Sat-A-P-3-7: 多民族语言语音到语音翻译研究进展综述

Hetong Fan (Minzu University of China); Yue Zhao (Minzu University of China)*; Xiaona Xu (Minzu University of China)

Sat-A-P-3-8: Tibetan-Chinese Speech-to-Speech Translation Based on Large Speech Models

Hetong Fan (Minzu University of China); Yue Zhao (Minzu University of China)*; Zairan Gong (Minzu University of China); Xiaona Xu (Minzu University of China); Sijia Ma (Minzu University of China)

Sat-A-P-3-9: LDMME: Latent Diffusion Model for Music Editing

Runchuan Ye (Harbin Institute of Technology (Shenzhen)); Shiyin Kang (Skywork AI PTE. LTD.); Zhiyong Wu (Tsinghua University)*

Sat-A-P-3-10: 基于多流序列和对比学习的音乐生成研究

Zijian Lin (Tsinghua University); Shun Lei (Tsinghua University); Jiangnan Ye (Tsinghua University); Zhiyong Wu (Tsinghua University)*; Shiyin Kang (Skywork AI PTE. LTD.)

Sat-A-P-3-11: 基于去噪扩散概率模型的对抗攻击方法

Yan Zhao (Hebei University of Technology)*; Jiangyan Yi (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences); Jianhua Tao (Tsinghua University); Yongfeng Dong (Hebei University of Technology)

Sat-A-P-3-12: 基于跨语言数据迁移的端到端伪造语音检测方法

Xiaoke Qi (China University of Political Science and Law)*; Jiangyan Yi (National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences); Hao Gu (Institute of Automation, Chinese Academy of Sciences); Jianhua Tao (Tsinghua University)

Sat-A-P-3-13: NDVQ: Robust Neural Audio Codec with Distribution-Based Vector Quantization

Zhikang Niu (Shanghai Jiao Tong University)*; Xie Chen (Shanghai Jiaotong University); Ziyang Ma (Shanghai Jiao Tong University); Shujie Liu (Microsoft Research Asia)

Sat-A-P-3-14: 基于通道注意力特征融合的异常音检测方法

wang dong (Harbin Institute of Technology)*; jiqing Han (Harbin Institute of Technology); 铁然郑 (哈尔滨工业大学); 贵滨郑 (哈尔滨工业大学); yongjun he (Harbin Institute of Technology)

Sat-A-P-3-15: 基于脑控嵌入向量的语音分离网络

Haoran ZHOU (Anhui University); Jingjing Zhang (Anhui University); Cunhang Fan (Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, School of Computer Science and Technology, Anhui University)*; Jian Zhou (Anhui University); Zhao Lu (Anhui University)

Sat-A-P-4: Speech Science

Room: 多功能厅

Type: Poster

August 17, 5:45 PM to 7:15 PM

Sat-A-P-4-1: An electroencephalogram-based study of neural responses to imagined speech in Mandarin

Ran Zhao (Tianjin University)*; Hongxing Liu (Tianjin University); Shuming Zhang (Tianjin University); Qi Tang (Tianjin University); Xiaoli Yu (Tianjin University Children's Hospital); Yanru Bai (Tianjin University); Guangjian Ni (Tianjin University)

Sat-A-P-4-2: 基于脑电信号的汉语普通话语音分类研究

Shuming Zhang (Tianjin University)*; Ran Zhao (Tianjin University); Qi Tang (Tianjin University); Minghan Guo (Tianjin University); Xiaoli Yu (Tianjin University Children's Hospital); Guangjian Ni (Tianjin University); Yanru Bai (Tianjin University)

Sat-A-P-4-3: 削波语音声学特征研究

Liwen Jing (中山大学)*

Sat-A-P-4-4: 基于文化分析的跨语言语音情感识别

Xin Chen (Inner Mongolia University)*

Sat-A-P-4-5: 声调音高曲线的调头处理办法初探

Feng LING (East China Normal University)*; DAN YUAN (East China Normal University)

Sat-A-P-4-6: 潮州话入声调的韵律边界效应

Huanzhe Li (Nanjing Normal University School of Chinese Language and Literature)*; Han Hu (Nanjing Normal University School of Chinese Language and Literature); Wentao Gu (Nanjing Normal University; McGill University)

Sat-A-P-4-7: 藏语安多方言塞音的超声舌位和声学研究

Daijicao Long (Qinghai Minzu University)*

Sat-A-P-4-8: 藏语安多方言塞音发声的声学和电声门图研究

玛吉完 (青海民族大学)*

Sat-A-P-4-9: 宜人性与开放性人格特质对情绪语音特征的影响

YALEI LI (Nanjing Normal University)*; Wenyun Hai (Nanjing Normal University); Wentao Gu (Nanjing Normal University; McGill University); Han Hu (Nanjing Normal University; McGill University)

Sat-A-P-4-10: 外倾性和神经质人格特质对情绪语音特征的影响

Wenyun Hai (Nanjing Normal University)*; YALEI LI (Nanjing Normal University); Wentao Gu (Nanjing Normal University; McGill University); Han Hu (School of Chinese Language and Literature, Nanjing Normal University)

Sat-A-P-4-11: Emilia: An Extensive, Multilingual, and Diverse Speech Dataset for Large-Scale Speech Generation

Haorui He (The Chinese University of Hong Kong, Shenzhen)*; Zengqiang Shang (The Institute of Acoustics of the Chinese Academy of Sciences); Chaoren Wang (The Chinese University of Hong Kong, Shenzhen); xuyuan li (The Institute of Acoustics of the Chinese Academy of Sciences); Yicheng Gu (The Chinese University of Hong Kong, Shenzhen); Hua Hua (Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, China); Liwei Liu (The Chinese University of Hong Kong, Shenzhen); Chen Yang (The Institute of Acoustics of the Chinese Academy of Sciences); Jiaqi Li (The Chinese University of Hong Kong, Shenzhen); Peiyang Shi (The Institute of Acoustics of the Chinese Academy of Sciences); Yuancheng Wang (The Chinese University of Hong Kong, Shenzhen); Kai Chen (Shanghai AI Laboratory); pengyuan zhang (Institute of Acoustics, Chinese Academy of Sciences); Zhizheng Wu (Chinese University of Hong Kong, Shenzhen)

Sat-A-P-4-12: 女性合成语音与真人语音的听感评价与分析

zhang jingying (xiaomi)*

Sunday, August 18

Sun-M-PLN-1: 大模型与语音技术

Room: 多功能厅

Type: Plenary

August 18, 10:00 AM to 10:45 AM

Presenter: 陶建华

Chair: 艾斯卡尔·艾木都拉

Sun-M-PLN-2: SALMONN: 认知导向的视听觉大语言模型

Room: 多功能厅

Type: Plenary

August 18, 10:45 AM to 11:30 AM

Presenter: 张超

Chair: 何亮

Sun-M-O-1: Miscellaneous: Multimodal System, Language Model, and Other

Room: 人文韬厅

Type: Oral

August 18, 11:45 AM to 1:15 PM

Chair: 李明 (昆山杜克大学)

Sun-M-O-1-1: A Backend-friendly On-device Multi-channel Speech Enhancement System with IPD and PHM

Wen Wen (shanghai Jiao Tong University)*; JingRui Qian (Hangzhou Dianzi University); Yifan Zhang (Shanghai Jiao Tong University); Yu Xi (Shanghai Jiao Tong University); Wenbin Jiang (Hangzhou Dianzi University); Qiang Zhou (AISpeech Ltd); Beiyi Liu (AISpeech Ltd); Yao Guo (AISpeech Ltd); Kai Yu (Shanghai Jiao Tong University)

Sun-M-O-1-2: Improved DOA Estimation of Sound Source of Small Amplitudes Using a Single Acoustic Vector Sensor

bing zhu (Center of Intelligent Acoustics and Immersive Communications, School of Marine Science and Technology, Northwestern Polytechnical University)*; Wen Zhang (Northwestern Polytechnical University)

Sun-M-O-1-3: Emotional Speech Synthesize via Visual Context Perception

Qixin Wang (Tsinghua University); Chenglin Guo (New York University); Songtao Zhou (Tsinghua University); Xiaoyu Qin (Tsinghua University)*

Sun-M-O-1-4: Improving Emotion Recognition with Pre-trained Models, Multimodality, and Contextual Information

Zhengshun Xia (Shanghai Jiao Tong University)*; Ziyang Ma (Shanghai Jiao Tong University); Zhisheng Zheng (Shanghai Jiao Tong University); Xie Chen (Shanghai Jiaotong University)

Sun-M-O-1-5: 基于前音节锚点音高规整的重音和间断特征考察

昊阳曹 (Beijing Language and Culture University)*; Jinsong Zhang (Beijing Language and Culture University)

Sun-M-O-1-6: A Study on the Effect of Focus on Vowel Duration and Formant in Cantonese

Maolin Wang (Jinan University)*; Han Yu (College of Chinese Language and Culture, Jinan University); Ziyu Xiong (CASS)

Sun-M-O-2: Human-Machine Interaction and Sound Event Detection

Room: 文博厅

Type: Oral

August 18, 11:45 AM to 1:15 PM

Chair: 张结 (中国科学技术大学)

Sun-M-O-2-1: Dynamic properties of diphthongs in Hefei Mandarin

Zhiheng Qian (Shanghai Jiao Tong University); Huifang Kong (Anhui Jianzhu University); Yifan Yang (Shanghai Jiao Tong University)*

Sun-M-O-2-2: ASD-Diff: Unsupervised Anomalous Sound Detection With Masked Diffusion Model

xin fan (xinjiang university); Wenjie Fang (Xinjiang University); Ying Hu (Xinjiang University)*

Sun-M-O-2-3: Emergence of Hemispheric Asymmetries and Predictive Coding in the Neural Mechanism of Speech Perception

Bin Zhao (Institute of Linguistics, Chinese Academy of Social Sciences)*; Jianwu Dang (Tianjin University); Aijun LI ("Institute of Linguistics, CASS")

Sun-M-O-2-4: Anomalous Sound Detection Using Time-Frequency Feature and Mixbatch

Shun Huang (Xinjiang University); Yunxiang Zhang (Xinjiang University); Zhihua Fang (Xinjiang University); MinRui Tang (Xinjiang University); Ruifeng Xu (Xinjiang University); Liang HE (Tsinghua University)*

Sun-M-O-2-5: Integrating Time-Frequency Domain Shallow and Deep Features for Speech-EEG Match-Mismatch of Auditory Attention Decoding

Yubang Zhang (University of Science and Technology of China); Qiushi Zhu (University of Science and Technology of China); QingTian Xu (University of Science and Technology of China); Jie Zhang (University of Science and Technology of China (USTC))*

Sun-M-O-2-6: Transformer-based Model for Auditory EEG Decoding

Jiixin Chen (University of Science and Technology of China)*; Jiahong Yuan (University of Science and Technology of China); Zhen-Hua Ling (University of Science and Technology of China)

Sun-M-SF: Student Forum 1

Room: 多功能厅

Type: Student Forum

August 18, 11:45 AM to 1:15 PM

Chair: 张晓雷, 艾斯卡尔·肉孜

11:45-11:55

Sun-M-SF-1: 基于 whisper 的语音识别的上下文学习

Presenter: 王嗣尹 (清华大学)

11:55-12:05

Sun-M-SF-2: 迈向实际应用的目标说话人提取

Presenter: 何树林 (内蒙古大学)

12:05-12:15

Sun-M-SF-3: IIANet: An Intra- and Inter-Modality Attention Network for Audio-Visual Speech Separation

Presenter: 李凯 (清华大学)

12:15-12:25

Sun-M-SF-4: A Prompt-Based Hierarchical Pipeline for Cross-Domain Slot Filling

Presenter: 魏笑 (天津大学)

12:25-12:35

Sun-M-SF-5: MBCFNet A Multimodal Brain-Computer Fusion Network for Human Intention Recognition

Presenter: 李忠杰 (天津大学)

12:35-12:45

Sun-M-SF-6: DBPNet: Dual-Branch Parallel Network with Temporal-Frequency Fusion for Auditory Attention Detection

Presenter: 张宏玉 (安徽大学)

12:45-12:55

Sun-M-SF-7: Ultimate Attainment in L2 Semantics: Evidence from Learning of Japanese Word Meaning by Chinese Native Speaker

Presenter: 赵翌 (Case Western Reserve University)

12:55-13:05

Sun-M-SF-8: 汉语单音节情绪语音刺激库的编制与评定

Presenter: 李梦媛 (北京大学护理学院)

13:05-13:15

Sun-M-SF-9: 汉语伪句中中性听觉语音库的编制

Presenter: 周安琪 (北京大学)

Sun-M-IND: Industry Forum 3

Room: 多功能厅

Type: Industry Forum

August 18, 1:15 PM to 1:55 PM

Chair: 谢磊

13:15-13:25

Sun-M-IND-1: 大模型场景下的数据标注

Presenter: 吴本谷 (标贝科技)

13:25-13:35

Sun-M-IND-2: 海天瑞声在语音大模型数据上的探索与实践

Presenter: 郝玉峰 (海天瑞声)

13:35-13:45

Sun-M-IND-3: 以数据 + 人才的社区生态建设

Presenter: 卜辉 (SpeechHome)

13:45-13:55

Sun-M-IND-4: 算力赋能，语音智联!

Presenter: 唐华 (并行科技)

Sun-A-O-1: Speech Synthesis, Sound Synthesis and Pectorophony

Room: 人文韬厅

Type: Oral

August 18, 3:30 PM to 5:00 PM

Chair: 刘瑞 (内蒙古大学)

Sun-A-O-1-1: ESTVocoder: An Excitation-Spectral-Transformed Neural Vocoder Conditioned on Mel Spectrogram

Xiao-Hang Jiang (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); Ye-Xin Lu (University of Science and Technology of China); Zhen-Hua Ling (University of Science and Technology of China)

Sun-A-O-1-2: Attention-Constrained Inference for Robust Decoder-Only Text-to-Speech

Hankun Wang (Shanghai Jiao Tong University)*; Chenpeng Du (Shanghai Jiao Tong University); Yiwei Guo (Shanghai Jiao Tong University); Shuai Wang (Shenzhen Research Institute of Big Data, Chinese University of Hong Kong (Shenzhen)); Xie Chen (Shanghai Jiaotong University); Kai Yu (Shanghai Jiao Tong University)

Sun-A-O-1-3: 基于语音谐波结构的多通道语音增强网络

nu er ai li (天津大学)*; Jin Li (天津大学); Rui Cao (Tianjin University); Zhongjian Cui (天津大学); Haoyu Wang (Tianjin University); Tianrui Wang (Tianjin University); Meng Ge (National University of Singapore); Longbiao Wang (Tianjin University); Jianwu Dang (Tianjin University)

Sun-A-O-1-4: MCDubber: Multimodal Context-Aware Expressive Video Dubbing

Yuan Zhao (Inner Mongolia University)*; Zhenqi Jia (Inner Mongolia University); Rui Liu (Inner Mongolia University); De Hu (Inner Mongolia University); Feilong Bao (Inner Mongolia University); Guanglai Gao (Inner Mongolia University)

Sun-A-O-1-5: 基于 Transformer 编码改进 GCRN 网络的单通道语音增强方法

Xinyue Shen (Anhui University); lei luo (Anhui University); Enrui Liu (Anhui University); 彤朱 (安徽大学); Cunhang Fan (Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, School of Computer Science and Technology, Anhui University)*; zhao lv (anhui university)

Sun-A-O-1-6: The attention-based fusion of master-auxiliary network for speech enhancement

Hairong Jia (Taiyuan university of Technology)*; LI YaRong (Taiyuan University of Technology)

Sun-A-O-2: Phonetics, Phonology and Prosody

Room: 文博厅

Type: Oral

August 18, 3:30 PM to 5:00 PM

Chair: 曹洪林 (中国政法大学)

Sun-A-O-2-1: 单音节阳平声调拐点位置与音节结构关系考察

昕蕾王 (北京语言大学); Xinyi Zhao (Beijing language and culture university)*; Ruishan Li (Beijing Language and Culture University); Xiaoli Feng (Beijing Language and Culture University); Jinsong Zhang (Beijing Language and Culture University)

Sun-A-O-2-2: 普通话双音节词连上变调的信息机制研究

Liu Xiaowang (Beijing Language and Culture University)*; Jinsong Zhang (Beijing Language and Culture University)

Sun-A-O-2-3: 《广州话韵律边界时长模式研究》

Meier Wang (Jinan University); Maolin Wang (Jinan University)*

Sun-A-O-2-4: 基于统计分类的一级甲等普通话五度调值划分研究

Lu Xiyao (Beijing Language and Culture University)*; Xiaolin Wu (Beijing Language and Culture University); How Yang (Beijing Advanced Innovation Center for Language Resource); Jinsong Zhang (Beijing Language and Culture University)

Sun-A-O-2-5: 德昂语的焦点韵律实现

Zenghui Liu (Yunnan University)*; Jie Luo (Yunnan University)

Sun-A-O-2-6: 东部裕固语长短元音空间分布的声学统计分析

Min YUWEN (Northwest Minzu University)*; SHI-Liang LYU (Northwest Minzu University)

Sun-A-SF: Student Forum 2

Room: 多功能厅

Type: Student Forum

August 18, 3:30 PM to 4:50 PM

Chair: 李健, 黄志华

15:30-15:40

Sun-A-SF-1: 语音跨模态融合大模型基础: 从语音细粒度理解到精细可控合成

Presenter: 秦霄羽 (清华大学)

15:40-15:50

Sun-A-SF-2: Achieving Realistic and Controllable Emotional Voice Conversion: A Step Forward in Speech Synthesis

Presenter: 齐天铎 (东南大学)

15:50-16:00

Sun-A-SF-3: 基于自回归扩散 Transformer 的语音合成方法

Presenter: 刘知峻 (香港中文大学 (深圳))

16:00-16:10

Sun-A-SF-4: MM-NodeFormer: Node Transformer Multimodal Fusion for Emotion Recognition in Conversation

Presenter: 黄梓龙 (香港理工大学)

16:10-16:20

Sun-A-SF-5: Emotion Rendering for Conversational Speech Synthesis with Heterogeneous Graph-Based Context Modeling

Presenter: 胡一帆 (内蒙古大学)

16:20-16:30

Sun-A-SF-6: Contrastive Learning based Modality-Invariant Feature Acquisition for Robust Multimodal Emotion Recognition with Missing Modalities

Presenter: 左昊麟 (内蒙古大学)

16:30-16:40

Sun-A-SF-7: 基于自发语音的青少年自杀风险检测

Presenter: 崔紫云 (清华大学)

16:40-16:50

Sun-A-SF-8: A Path Signature Approach for Speech-based Dementia Detection

Presenter: 潘怡霖 (大连海事大学)

Sun-A-YF: Young Forum 2

Room: 多功能厅

Type: Young Forum

August 18, 5:15 PM to 6:45 PM

Chair: 努尔麦麦提·尤鲁瓦斯

Sun-A-YF-1: AI for Assistive Communication: Disordered Speech Reconstruction Using Speech Language Models

Presenter: Xixin Wu is currently an Assistant Professor at the Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong.

Sun-A-YF-2: 面向自然人机对话的语音合成技术

Presenter: 刘瑞, 内蒙古大学计算机学院研究员, 蒙古文智能信息处理国家地方联合工程研究中心成员

Sun-A-YF-3: 多模态情感识别 + 大模型

Presenter: 连政, 中国科学院自动化研究所多模态人工智能系统全国重点实验室助理研究员

Organizer List

名誉主席 (Honorary Chair)

- 吾守尔·斯拉木 - 中国工程院院士 - CCF 乌鲁木齐分部主席 - 新疆大学
- 郑方 - 清华大学

大会主席 (Conference Chairs)

- 张劲松 - 北京语言大学
- 俞凯 - 上海交通大学
- 艾斯卡尔·艾木都拉 - 新疆大学

程序委员会主席 (Program Committee Chairs)

- 凌震华 - 中国科学技术大学
- 陈谐 - 上海交通大学
- 何亮 - 清华大学/新疆大学

组织委员会主席 (Organizing Committee Chairs)

- 吴志勇 - 清华大学深圳研究生院
- 库尔班·吾布力 - CCF 乌鲁木齐分部副主席 - 新疆大学
- 吐尔地·托合提 - 新疆大学
- 伊克热木·买买提依明 - 新疆大学

工业联络主席 (Industrial Liaison Chairs)

- 贾珈 - 清华大学
- 谢磊 - 西北工业大学
- 于清 - 新疆大学
- 阿不都萨拉木·达吾提 - 新疆大学

宣传主席 (Publicity Chairs)

- 张超 - 清华大学
- 胡英 - 新疆大学
- 木合亚提·尼亚孜别克 - 新疆大学

出版主席 (Publication Chairs)

- 李雅 - 北京邮电大学
- 玛依热·依布拉音 - 新疆大学
- 古扎努尔·艾木肉拉 - 新疆大学

教程报告主席 (Tutorial Chairs)

- 韩纪庆 - 哈尔滨工业大学
- 张卫强 - 清华大学
- 米吉提·阿不里米提 - 常州分音塔科技有限公司

特殊议题主席 (Special Session Chairs)

- 王东 - 清华大学
- 李爱军 - 中国社会科学院语言研究所
- 袁家宏 - 中国科技大学
- 张华平 - 北京理工大学
- 黄浩 - 新疆大学

青年论坛主席 (Young Forum Chairs)

- 欧智坚 - 清华大学
- 努尔麦麦提·尤鲁瓦斯 - 新疆大学

学生论坛主席 (Student Forum Chairs)

- 张晓雷 - 西北工业大学
- 李健 - 清华海峡院
- 黄志华 - 新疆大学
- 艾斯卡尔·肉孜 - 新疆大学

财务主席 (Financial Chairs)

- 哈妮克孜·伊拉洪 - 新疆大学
- 古力努尔·艾尔肯 - 新疆财经大学