National Conference on Man-Machine Speech Communication

2026 NCMMSC

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

目录 / Contents

大会主席寄语 / Message from the General Chairs 1

程序与出版工作前言 / Preface from the Program and Publication Chairs 2

组织者列表 / Organizer List 3

简明议程 / Program at a Glance 5

详细议程 / Detailed Program 9

嘉宾信息 / Guest Profiles 17

大会报告 / Plenary Talks 17

AACL 联合论坛 / AACL Joint Forum 19

教程报告 / Tutorials 21

青年论坛 / Young Forums 23

科普教育论坛 / Education and Outreach Forum 27

工业论坛 / Industry Forums 28

会议议程 / Session Program 31

Thursday, November 5

[14:00-17:00] Thu-R - 会议报到 31

Friday, November 6

[08:00-21:00] Fri-R - 会议报到 31

[09:00-17:00] Fri-EXH - 赞助商展览 31

[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 31

[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 31

[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 31

[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 32
[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 32
[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 32
[13:30-15:00] Fri-A-IND-1 - Industry Forum I 32
[15:30-17:00] Fri-A-IND-2 - Industry Forum II 33
[18:30-20:30] Fri-E-CCF - CCF 语音对话与听觉专委会工作会议 33
[20:30-22:00] Fri-E-CIPS - CIPS 语音信息专委会工作会议 33

Saturday, November 7

[08:00-18:00] Sat-R - 会议报到 34
[08:30-18:00] Sat-EXH - 赞助商展览 34
[08:30-09:00] Sat-OPEN - 大会开幕式（含大会合影） 34
[09:00-09:45] Sat-M-PLEN-1 - The Way of Multimodal Communication: From Emotion to Health 34
[09:45-10:30] Sat-M-PLEN-2 - Speaker Tracking: Segment-online Multi-talker Organization with a Varying Number of Speakers 34
[11:00-12:30] Sat-M-O-1 - Phonetics, Pronunciation Assessment and ASR Error Correction 34
[11:00-12:15] Sat-M-O-2 - Robust, Multimodal and Low-Resource Speech Recognition 35
[11:00-12:30] Sat-M-O-3 - Expressive and Controllable Speech Synthesis 36
[11:00-12:30] Sat-M-O-4 - Neural Speech Coding, Quantization and Compression 37
[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 38
[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 38
[11:00-12:30] Sat-M-SS-3 - 特殊议题 8: 共情计算与认知交互 I 39
[11:00-12:30] Sat-M-YF - Young Forum 1 39
[13:30-17:00] Sat-A-EDU - 科普教育论坛 39
[13:30-15:30] Sat-A-AACL - AACL 联合论坛：语音 × 语言 × 多模态：迈向智能交互新范式 40
[13:30-15:00] Sat-A-SF-1 - Student Forum I 40
[13:30-15:00] Sat-A-O-1 - Spatial Audio Encoding and Speech, Music and Sound Separation 40
[13:30-15:00] Sat-A-O-2 - Speaker Representations, Diarization and Dialect Identification 41
[13:30-15:00] Sat-A-O-3 - Speaker Verification, Voice Privacy and Watermarking 42
[13:30-15:00] Sat-A-P-1 - Speech Science, Cognitive Modeling and Healthcare 43
[13:30-15:00] Sat-A-SS-1 - 特殊议题 8: 共情计算与认知交互 II 47
[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 48
[15:30-17:00] Sat-A-SF-2 - Student Forum II 48
[15:30-17:15] Sat-A-O-4 - Audio Foundation Models and Sound Event Understanding 48
[15:30-17:00] Sat-A-O-5 - Target Speaker Extraction and Audio-Visual Speech Enhancement 49
[15:30-17:00] Sat-A-O-6 - Speech Enhancement, Restoration and Noise Control 50
[15:30-17:00] Sat-A-P-2 - Speech Recognition, Audio Understanding and Multimodal Interaction 51

[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 56
[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 57
[17:20-18:05] Sat-A-PLEN-1 - 四十年从理想到现实 57
[20:30-22:30] Sat-E-FUND - 基金分享沙龙 57

Sunday, November 8

[08:00-18:00] Sun-R - 会议报到 58
[09:00-18:00] Sun-EXH - 赞助商展览 58
[09:00-09:45] Sun-M-PLEN-1 - YuE2: Bringing the Human into the Loop of Symbiotic Creativity in Music 58
[09:45-10:30] Sun-M-PLEN-2 - 迈向全模态交互模型 58
[11:00-12:30] Sun-M-O-1 - Spotlight Paper Presentation I 58
[11:00-12:30] Sun-M-O-2 - Speech Spoofing and Content Manipulation Detection 59
[11:00-12:30] Sun-M-O-3 - Generative Speech Synthesis and Latent Representations 60
[11:00-12:30] Sun-M-SS-1 - 特殊议题 7: 语音与音乐的动态交互 61
[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 61
[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 62
[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 62
[11:00-12:30] Sun-M-YF - Young Forum 2 63
[12:30-14:00] Sun-N-COMM - NCMMSC 常设机构会议 63
[14:00-15:30] Sun-A-O-1 - Spotlight Paper Presentation II 63
[14:00-15:30] Sun-A-O-2 - Multimodal Emotion, Sentiment and Group Affect Recognition 64
[14:00-15:30] Sun-A-O-3 - Spoken Dialogue Systems, Turn-Taking and Interaction Modeling 65
[14:00-15:30] Sun-A-P-1 - Speech and Audio Generation, Coding and Restoration 66
[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 70
[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 70
[14:00-15:30] Sun-A-YF-1 - Young Forum 3 71
[15:45-17:15] Sun-A-O-4 - Speech and Multimodal Intelligence for Health and Cognitive Assessment 71
[15:45-17:00] Sun-A-O-5 - Speech and Audio-Visual Resources, Retrieval and Evaluation 72
[15:45-16:45] Sun-A-O-6 - Speech, Audio and Visual Speech Generation 73
[15:45-17:15] Sun-A-P-2 - Robust Speech Processing, Speaker Modeling and Audio Security 74
[15:45-17:15] Sun-A-SS-3 - 特殊议题 11: 2026 年 CCF 先进音频技术竞赛 78
[15:45-17:15] Sun-A-YF-2 - Young Forum 4 79
[17:35-18:00] Sun-CLOSE - 大会闭幕式 79

会务指南 / Conference Information 80

大会主席寄语

Message from the General Chairs

尊敬的各位嘉宾、各位同仁、各位朋友：

欢迎参加第二十一届全国人机语音通讯学术会议（NCMMSC 2026）！我们谨代表大会组委会，向莅临珠海横琴的各位专家学者、产业界同仁和青年学子表示热烈欢迎，向长期关心和支持 NCMMSC 发展的各界朋友致以诚挚感谢！

全国人机语音通讯学术会议自 1990 年创办以来，始终致力于推动语音科学与技术领域的学术交流和创新发展。从语音科学的基础研究，到语音识别、语音合成与人机交互技术的持续突破，NCMMSC 见证并参与了我国语音科学与技术的发展历程，逐步成为连接学术界与产业界、促进成果交流与科研合作的重要平台。2026 年，NCMMSC 正式入选**中国计算机学会（CCF）推荐学术会议目录（C 类）**，标志着会议发展迈上新的台阶，也激励我们持续提升学术质量与国际影响力，更好地服务学术共同体。

本届会议于 2026 年 11 月 5 日至 8 日在珠海横琴天沐琴台国际会展中心举行，由中国计算机学会（CCF）与中国中文信息学会（CIPS）联合主办，CCF 语音对话与听觉专业委员会、CIPS 语音信息专业委员会、清华大学深圳国际研究生院、香港中文大学（深圳）及北京理工大学（珠海）联合承办，并同期举行两个专业委员会的学术年会。时隔二十余年，NCMMSC 再次来到粤港澳大湾区。我们期待依托横琴毗邻澳门、联通国际的区位优势，进一步深化学术交流与国际合作。

本届会议以“**语音 × 语言 × 多模态：迈向智能交互新范式**”为主题。随着人工智能大模型技术的快速发展，语音、语言与多模态智能正加速融合，为人机交互带来新的机遇与挑战。大会在持续关注语音科学与技术核心问题的基础上，进一步拓展至语言智能、音频与音乐、多模态交互、情感与认知计算、智能语音医疗及信息无障碍等交叉领域，共同探索更加自然、智能、以人为中心的人机交互新范式。

本届会议的征稿工作得到了广大研究人员的积极响应，**共收到论文投稿 465 篇，最终录用 353 篇，投稿量与录用量均创新高**。投稿论文覆盖语音、语言、音频及多模态智能等多个研究方向，充分展现了相关领域的研究活力，也反映了 NCMMSC 学术吸引力与影响力的持续提升。广大青年学者和学生的积极参与，更为会议注入了新的活力。

本届会议还与自然语言处理与中文计算会议（NLPCC 2026）、亚洲计算语言学会议（AACL 2026）同期协同举办，进一步促进语音、语言与多模态智能领域不同学术社区之间的交流合作。大会设置了大会特邀报告、NCMMSC–AACL 联合论坛、教程报告、特殊议题、青年论坛、学生论坛、工业论坛、科普教育论坛、技术竞赛以及论文口头报告与海报展示等丰富活动，为学术交流、跨领域合作与青年人才培养搭建多元化平台。

本届会议还得到了产业界的大力支持，**共有 15 家企业和机构参与赞助，其中白金赞助商 1 家、金牌赞助商 8 家、银牌赞助商 6 家**，涵盖互联网科技、智能语音、大模型、音频声学、人工智能数据及智能终端等多个领域。产业伙伴的积极参与，为大会的顺利举办提供了有力支持，也进一步促进了学术研究与产业实践的交流合作。

本届会议的顺利召开，凝聚了众多同仁的智慧与心血。在此，我们衷心感谢各主办和承办单位、全体组委会成员、报告嘉宾、论文作者与审稿专家，特别感谢 15 家产业伙伴的鼎力支持，以及会务团队和志愿者的辛勤付出！正是各方的共同努力，使本届大会得以汇聚丰富的学术成果与多元的交流活动，为学术交流、产业合作与人才培养搭建更加广阔的平台。

最后，衷心祝愿各位嘉宾和同仁在横琴度过充实而愉快的会议时光，在交流中收获启发，在研讨中深化合作，在相聚中增进友谊。

预祝 NCMMSC 2026 圆满成功！

大会主席：俞凯、贾珈、吴志勇

2026 年 11 月

程序与出版工作前言

Preface from the Program and Publication Chairs

尊敬的各位嘉宾、各位同仁、各位朋友：

第二十一届全国人机语音通讯学术会议（NCMMSC 2026）的征稿工作得到了语音、语言及多模态智能领域广大研究人员的积极响应。本届大会共收到 **465 篇投稿**，其中英文论文 417 篇、中文论文 48 篇；主会投稿 411 篇、特殊议题投稿 54 篇。投稿论文涵盖语音科学与语言学、语音识别与合成、音频与声学处理、语言理解与大模型、对话与多模态交互、医疗健康及信息无障碍等多个研究方向，体现了相关领域研究的广泛性与交叉融合趋势。值得一提的是，**第一作者明确为在校学生的投稿占 80.7%**，充分展现了青年科研人员参与学术研究与交流的热情与活力。

为保障论文评审质量，程序委员会坚持学术规范，重点关注论文的原创性、技术正确性、研究贡献与实验验证，组织开展了同行评审与综合评议。本届大会共向 **262 位审稿专家**分配评审任务，其中 252 位专家累计提交 **1,439 份评审意见**。在已有评审记录的 430 篇稿件中，420 篇获得至少三份评审意见，为论文的综合评价提供了较为充分的依据。程序委员会在综合同行评审意见、元评审建议及相关审议结果的基础上，最终**录用 353 篇论文**，其中主会论文 311 篇、特殊议题论文 42 篇，占全部投稿的 **75.9%**。录用的论文中，有 **120 篇论文被推荐并收录至 Springer CCIS**（Communications in Computer and Information Science）会议论文集。相关研究成果将在大会期间通过论文口头报告、海报展示及特殊议题等形式，与参会同仁进行深入交流。

在论文出版方面，大会结合论文评审结果、论文类型及作者发表意愿，开展了多轮论文推荐、出版协调与作者沟通工作。目前，120 篇英文论文正在准备 Springer CCIS 论文集出版材料。与此同时，部分优秀论文已推荐至《信号处理》《人工智能》《应用声学》《中国语音学学报》以及 *Journal of Shanghai Jiao Tong University (Science)*、*Intelligent Computing*、*Data Intelligence* 等中英文期刊。部分推荐论文已进入拟接收或作者修改阶段，其他论文的投稿、评审与出版协调工作仍在有序推进。大会将继续与论文作者、合作期刊编辑部及出版社保持密切沟通，做好后续论文出版与成果传播工作，并以规范的同行评审和多元化的成果出版机制为基础，持续 NCMMSC 的学术质量与影响力。

论文投稿、评审与出版工作的顺利开展，离不开广大作者、审稿专家、组委会成员以及合作期刊和出版社的共同努力。在此，我们谨向所有论文作者对 NCMMSC 2026 的信任与支持表示诚挚感谢，向认真负责、无私奉献的审稿专家和元评审专家致以衷心谢意，向参与论文组织、评审协调与出版工作的各位同仁表示由衷感谢。

我们期待以本届会议为契机，进一步促进语音、语言与多模态智能领域的高质量学术交流，推动优秀研究成果的传播与共享，为相关领域的持续创新与发展贡献力量。

程序委员会主席：董明会、刘瑞、万小军

出版主席：李雅、吴锡欣、蔡思祺

2026 年 11 月

组织者列表

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)

- 邹月娴（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）

简明议程

Program at a Glance

11 月 5 日		
时间	**内容**	**地点**
14:00–17:00	会议报到	会议室前厅

11 月 6 日		
时间	**内容**	**地点**
08:00–21:00	会议报到	会议室前厅
09:00–17:00	赞助商展览	会议室前厅
09:00–10:30	教程报告 1：连政（同济大学）	8M 会议室
09:00–10:30	教程报告 2：刘瑞（内蒙古大学）	8F 会议室
10:30–11:00	茶歇	会议室前厅
11:00–12:30	教程报告 3：陈丽萍（中国科学技术大学）	8M 会议室
11:00–12:30	教程报告 4：钱馨园（北京科技大学）	8F 会议室
12:30–13:30	午餐	天沐厅
13:30–15:00	教程报告 5：陈泽华（清华大学）	8M 会议室
13:30–15:00	工业论坛 I	8F 会议室
15:00–15:30	茶歇	会议室前厅
15:30–17:00	教程报告 6：江文斌（杭州电子科技大学）	8M 会议室
15:30–17:00	工业论坛 II	8F 会议室
17:00–18:30	欢迎晚宴	天沐厅
18:30–20:30	CCF 语音对话与听觉专委会工作会议	8M+8F 会议室
20:30–22:00	CIPS 语音信息专委会工作会议	8M+8F 会议室

11 月 7 日		
时间	**内容**	**地点**
08:00–18:00	会议报到	会议室前厅
08:30–18:00	赞助商展览	会议室前厅
08:30–09:00	大会开幕式（含大会合影）	沐辉厅 A+B+C
09:00–09:45	大会主旨报告 1：Björn W. Schuller（德国慕尼黑工业大学）- The Way of Multimodal Communication: From Emotion to Health	沐辉厅 A+B+C
09:45–10:30	大会主旨报告 2：汪德亮（香港中文大学 (深圳)）- Speaker Tracking: Segment-online Multi-talker Organization with a Varying Number of Speakers	沐辉厅 A+B+C
10:30–11:00	茶歇	会议室前厅
11:00–12:30	Sat-M-O-1: Phonetics, Pronunciation Assessment and ASR Error Correction	7 号会议室
11:00–12:30	Sat-M-O-2: Robust, Multimodal and Low-Resource Speech Recognition	8M 会议室
11:00–12:30	Sat-M-O-3: Expressive and Controllable Speech Synthesis	8F 会议室

11 月 7 日		
时间	**内容**	**地点**
11:00–12:30	青年论坛 1 青年报告 1：艾杨（中国科学技术大学） 青年报告 2：夏菁（中国科学院） 青年报告 3：郑达（小红书）、王翰坤/李波含（上海交通大学）	沐辉厅 A
11:00–12:30	特殊议题 1：China Voices Challenge 2026 中文多方言语音识别挑战赛	沐辉厅 B
11:00–12:30	特殊议题 2：面向具身智能的空间听觉与人机交互 I	沐辉厅 C
11:00–12:30	特殊议题 8：共情计算与认知交互 I	2D 会议室
11:00–12:30	Sat-M-O-4: Neural Speech Coding, Quantization and Compression	2R 会议室
12:30–13:30	午餐	天沐厅
13:30–15:30	Sat-A-AACL（NCMMSC-AACL 联合论坛） 特邀报告 1：王昊奋（同济大学） 特邀报告 2：Junichi Yamagishi（日本国立信息研究院） 特邀报告 3：Satoshi Nakamura（香港中文大学 (深圳)） 特邀报告 4：严睿（武汉大学）	沐和厅 B+C
13:30–17:00	Sat-A-EDU（科普教育论坛） 科普报告 1：马少平（清华大学） 科普报告 2：洪青阳（厦门大学） 科普报告 3：朱璇（三星电子中国研究院）	9 号会议室
13:30–15:00	Sat-A-O-1: Spatial Audio Encoding and Speech, Music and Sound Separation	7 号会议室
13:30–15:00	Sat-A-O-2: Speaker Representations, Diarization and Dialect Identification	8M 会议室
13:30–15:00	Sat-A-O-3: Speaker Verification, Voice Privacy and Watermarking	8F 会议室
13:30–15:00	学生论坛 I	沐辉厅 A
13:30–15:00	特殊议题 8：共情计算与认知交互 II	沐辉厅 B
13:30–15:00	特殊议题 2：面向具身智能的空间听觉与人机交互 II	沐辉厅 C
13:30–15:00	Sat-A-P-1: Speech Science, Cognitive Modeling and Healthcare	2D+2R 会议室 海报区
15:00–15:30	茶歇	会议室前厅
15:30–17:00	Sat-A-O-4: Audio Foundation Models and Sound Event Understanding	7 号会议室
15:30–17:00	Sat-A-O-5: Target Speaker Extraction and Audio-Visual Speech Enhancement	8M 会议室
15:30–17:00	Sat-A-O-6: Speech Enhancement, Restoration and Noise Control	8F 会议室
15:30–17:00	学生论坛 II	沐辉厅 A
15:30–17:00	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	沐辉厅 B
15:30–17:00	特殊议题 6：面向生成式音频的可信音频处理方法	沐辉厅 C
15:30–17:00	Sat-A-P-2: Speech Recognition, Audio Understanding and Multimodal Interaction	2D+2R 会议室 海报区
17:20–18:05	大会主旨报告 3：颜永红（中国科学院声学研究所）- 四十年从理想到现实	沐辉厅 A+B+C
18:30–20:30	大会晚宴	天沐厅
20:30–22:30	基金分享沙龙	沐辉厅 A+B+C

11 月 8 日		
时间	**内容**	**地点**
08:00–18:00	会议报到	会议室前厅
09:00–18:00	赞助商展览	会议室前厅
09:00–09:45	大会主旨报告 4：郭毅可（香港科技大学）- YuE2: Bringing the Human into the Loop of Symbiotic Creativity in Music	沐辉厅 A+B+C
09:45–10:30	大会主旨报告 5：邱锡鹏（上海创智学院/复旦大学）- 迈向全模态交互模型	沐辉厅 A+B+C
10:30–11:00	茶歇	会议室前厅
11:00–12:30	Sun-M-O-1: Spotlight Paper Presentation I	7 号会议室
11:00–12:30	Sun-M-O-2: Speech Spoofing and Content Manipulation Detection	8M 会议室
11:00–12:30	Sun-M-O-3: Generative Speech Synthesis and Latent Representations	8F 会议室
11:00–12:30	青年论坛 2 青年报告 4：李楠（天津师范大学） 青年报告 5：胡诗超（苏州大学） 青年报告 6：徐春阳（苏州大学）	沐辉厅 A
11:00–12:30	特殊议题 7：语音与音乐的动态交互	沐辉厅 B
11:00–12:30	特殊议题 3：第三届多语言智能语音与语言技术研讨会	沐辉厅 C
11:00–12:30	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	2D 会议室
11:00–12:30	特殊议题 10：语音模型评测体系构建与开源生态创新	2R 会议室
12:30–14:00	午餐	天沐厅
12:30–14:00	NCMMSC 常设机构会议	7 号会议室
14:00–15:30	Sun-A-O-1: Spotlight Paper Presentation II	7 号会议室
14:00–15:30	Sun-A-O-2: Multimodal Emotion, Sentiment and Group Affect Recognition	8M 会议室
14:00–15:30	Sun-A-O-3: Spoken Dialogue Systems, Turn-Taking and Interaction Modeling	8F 会议室
14:00–15:30	青年论坛 3 青年报告 7：潘怡霖（大连海事大学） 青年报告 8：薛浏蒙（南京大学） 青年报告 9：谢旭荣（中国科学院）	沐辉厅 A
14:00–15:30	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	沐辉厅 B
14:00–15:30	特殊议题 4：复杂声学场景下的目标音频提取	沐辉厅 C
14:00–15:30	Sun-A-P-1: Speech and Audio Generation, Coding and Restoration	2D+2R 会议室 海报区
15:30–15:45	茶歇	会议室前厅
15:45–17:15	Sun-A-O-4: Speech and Multimodal Intelligence for Health and Cognitive Assessment	7 号会议室
15:45–17:15	Sun-A-O-5: Speech and Audio-Visual Resources, Retrieval and Evaluation	8M 会议室
15:45–17:15	Sun-A-O-6: Speech, Audio and Visual Speech Generation	8F 会议室
15:45–17:15	青年论坛 4 青年报告 10：刘李（香港科技大学 (广州)） 青年报告 11：王潇蔓（上海人工智能实验室） 青年报告 12：拉巴顿珠（西藏大学）	沐辉厅 A
15:45–17:15	特殊议题 11：2026 年 CCF 先进音频技术竞赛	沐辉厅 C

11 月 8 日		
时间	**内容**	**地点**
15:45–17:15	Sun-A-P-2: Robust Speech Processing, Speaker Modeling and Audio Security	2D+2R 会议室 海报区
17:35–18:00	大会闭幕式	沐辉厅 A+B+C

详细议程

Detailed Program

11 月 5 日		
时间	**内容**	**地点**
14:00–17:00	**会议报到**	会议室前厅

11 月 6 日		
时间	**内容**	**地点**
08:00–21:00	**会议报到**	会议室前厅
09:00–17:00	**赞助商展览**	会议室前厅
09:00–10:30	**教程报告 1** **主持人：**陈丽萍（中国科学技术大学） **报告题目：**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 **报告嘉宾：**连政（同济大学）	8M 会议室
09:00–10:30	**教程报告 2** **主持人：**李琳（厦门大学） **报告题目：**鲁棒对话智能：复杂场景下人机语音交互感知与表达 **报告嘉宾：**刘瑞（内蒙古大学）	8F 会议室
10:30–11:00	**茶歇**	会议室前厅
11:00–12:30	**教程报告 3** **主持人：**陈丽萍（中国科学技术大学） **报告题目：**面向深度伪造语音的话者隐私保护：语音匿名化与水印 **报告嘉宾：**陈丽萍（中国科学技术大学）	8M 会议室
11:00–12:30	**教程报告 4** **主持人：**李琳（厦门大学） **报告题目：**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 **报告嘉宾：**钱馨园（北京科技大学）	8F 会议室
12:30–13:30	**午餐**	天沐厅
13:30–15:00	**教程报告 5** **主持人：**曹梦雪（中国社会科学院语言研究所） **报告题目：**操控生成：面向音频生成与编辑的条件扩散、桥、与引导方法 **报告嘉宾：**陈泽华（清华大学）	8M 会议室

11 月 6 日		
时间	**内容**	**地点**
13:30–15:00	**工业论坛 I** **主持人：**朱璇（北京三星通信技术研究有限公司） **大会主席致辞：**吴志勇（清华大学深圳国际研究生院） **宣讲主题 1：**语音 AI 全链路实践：从语音理解到智能交互 **专题 1：**听得懂，说得好，沟通无界——天籁语音大模型全新升级 **报告嘉宾：**商世东（腾讯） **专题 2：**Unified Streaming ASR and Word-Level Turn Detection with Interleaved Audio-Text LLMs **报告嘉宾：**杨栋（腾讯） **专题 3：**从语音对话到任务协作：语音交互基座模型的工业实践 **报告嘉宾：**Nova（腾讯） **宣讲主题 2：**从鸡尾酒会到热词检索——大模型 ASR 困难场景破局之路 **报告嘉宾：**范利春（小米集团） **宣讲主题 3：**近期语音数据热点洞察 **报告嘉宾：**胡婕（北京慧听科技有限公司）	8F 会议室
15:00–15:30	**茶歇**	会议室前厅
15:30–17:00	**教程报告 6** **主持人：**曹梦雪（中国社会科学院语言研究所） **报告题目：**语音增强前沿进展：从预测式估计迈向生成式建模 **报告嘉宾：**江文斌（杭州电子科技大学）	8M 会议室
15:30–17:00	**工业论坛 II** **主持人：**王磊（新加坡华为维纳研究所） **宣讲主题 4：**科大讯飞 Spark-Audio 序列大模型进展 **报告嘉宾：**熊世富（科大讯飞） **宣讲主题 5：**基于连续潜变量自回归建模的 CuteTTS 模型 **报告嘉宾：**梁敏（OPPO） **宣讲主题 6：**VoxCPM：连续自回归的语音生成模型 **报告嘉宾：**曾国洋（面壁智能联合创始人、首席技术官） **宣讲主题 7：**SmoothConv 与 DuplexConv：面向语音交互的互补型普通话多人对话语料库 **报告嘉宾：**王成有（QualiaLabs × 西北工业大学 ASLP） **宣讲主题 8：**猫王妙播音响：AI 音频产品生态的构建与创新实践 **报告嘉宾：**曾德钧（猫王妙播音响） **宣讲主题 9：**端侧多模态交互硬件的产业化实践——AI 语音智能体鼠标的方案与合作路径 **报告嘉宾：**谢雨田（安徽声云智能科技有限公司）	8F 会议室
17:00–18:30	**欢迎晚宴**	天沐厅
18:30–20:30	**CCF 语音对话与听觉专委会工作会议**	8M+8F 会议室
20:30–22:00	**CIPS 语音信息专委会工作会议**	8M+8F 会议室

11 月 7 日		
时间	**内容**	**地点**
08:00–18:00	**会议报到**	会议室前厅
08:30–18:00	**赞助商展览**	会议室前厅

11 月 7 日		
时间	**内容**	**地点**
08:30–09:00	**大会开幕式（含大会合影）** **主持人：**吴志勇（清华大学深圳国际研究生院）	沐辉厅 A+B+C
09:00–09:45	**大会主旨报告 1** **主持人：**俞凯（上海交通大学） **报告题目：**The Way of Multimodal Communication: From Emotion to Health **报告嘉宾：**Björn W. Schuller（德国慕尼黑工业大学）	沐辉厅 A+B+C
09:45–10:30	**大会主旨报告 2** **主持人：**谢湘（北京理工大学） **报告题目：**Speaker Tracking: Segment-online Multi-talker Organization with a Varying Number of Speakers **报告嘉宾：**汪德亮（香港中文大学（深圳））	沐辉厅 A+B+C
10:30–11:00	**茶歇**	会议室前厅
11:00–12:30	**Sat-M-O-1:** Phonetics, Pronunciation Assessment and ASR Error Correction **Session Chair:** 钱馨园（北京科技大学）	7 号会议室
11:00–12:30	**Sat-M-O-2:** Robust, Multimodal and Low-Resource Speech Recognition **Session Chair:** 柯登峰（北京语言大学）	8M 会议室
11:00–12:30	**Sat-M-O-3:** Expressive and Controllable Speech Synthesis **Session Chair:** 高迎明（北京邮电大学）	8F 会议室
11:00–12:30	**青年论坛 1** **主持人：**王雪涵（广州大学） **青年报告 1（11:00-11:30）** **报告题目：**低码率神经网络语音编解码方法研究 **报告嘉宾：**艾杨（中国科学技术大学） **青年报告 2（11:30-12:00）** **报告题目：**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 **报告嘉宾：**夏菁（中国科学院） **青年报告 3（12:00-12:30）** **报告题目：**迈向更稳定的全连续自回归语音合成 **报告嘉宾：**郑达（小红书）、王翰坤/李波含（上海交通大学）	沐辉厅 A
11:00–12:30	**特殊议题 1: China Voices Challenge 2026 中文多方言语音识别挑战赛** **组织者：**谢磊（西北工业大学）、薛浏蒙（南京大学）、刘和鑫（南洋理工大学）、石宪（阿里巴巴通义实验室）、胡婕（北京慧听科技有限公司） **主持人：**薛浏蒙（南京大学）	沐辉厅 B
11:00–12:30	**特殊议题 2: 面向具身智能的空间听觉与人机交互 I** **组织者：**朱梦尧（苏州大学）、杜俊（中国科学技术大学）、卜辉（希尔贝壳）、薛浏蒙（南京大学） **主持人：**朱梦尧（苏州大学）	沐辉厅 C
11:00–12:30	**特殊议题 8: 共情计算与认知交互 I** **组织者：**黄东延（深圳市智拓欣扬科技有限公司）、赵子平（天津师范大学）、张子兴（湖南大学）、李雅（北京邮电大学）、刘斌（中国科学院自动化研究所） **主持人：**李雅（北京邮电大学）	2D 会议室
11:00–12:30	**Sat-M-O-4:** Neural Speech Coding, Quantization and Compression **Session Chair:** 易江燕（清华大学）	2R 会议室

11 月 7 日		
时间	**内容**	**地点**
12:30–13:30	**午餐**	天沐厅
13:30–15:30	**NCMMSC-AACL 联合论坛** **主持人：**俞凯（上海交通大学）、张岳（西湖大学） **特邀报告 1（13:30-14:00）** **报告题目：**Token Economics in the Agent Era: From Token Consumption to Intelligent Scheduling **报告嘉宾：**王昊奋（同济大学） **特邀报告 2（14:00-14:30）** **报告题目：**Toward Multi-Layer Protection of Speech Against Real-World Misuse: What Can the Speech Community Do? **报告嘉宾：**Junichi Yamagishi（日本国立信息研究院） **特邀报告 3（14:30-15:00）** **报告题目：**Challenges and Progress in Automatic Speech-to-Speech Translation: Bridging the Gap to Human Simultaneous Interpretation **报告嘉宾：**Satoshi Nakamura（香港中文大学（深圳）） **特邀报告 4（15:00-15:30）** **报告题目：**Multi-Agent Collaboration for Complex Tasks **报告嘉宾：**严睿（武汉大学）	沐和厅 B+C
13:30–17:00	**科普教育论坛** **主持人：**关键（哈尔滨工程大学） **科普报告 1** **报告题目：**人工智能的奥秘 **报告嘉宾：**马少平（清华大学） **科普报告 2** **报告题目：**人工智能-产业变革与青年未来 **报告嘉宾：**洪青阳（厦门大学） **科普报告 3** **报告题目：**Design thinking：以人为本的创新范式，如何守住 AI 时代的创新起点 **报告嘉宾：**朱璇（三星电子中国研究院）	9 号会议室
13:30–15:00	**Sat-A-O-1:** Spatial Audio Encoding and Speech, Music and Sound Separation **Session Chair:** 黄公平（武汉大学）	7 号会议室
13:30–15:00	**Sat-A-O-2:** Speaker Representations, Diarization and Dialect Identification **Session Chair:** 雪巍（香港科技大学）	8M 会议室
13:30–15:00	**Sat-A-O-3:** Speaker Verification, Voice Privacy and Watermarking **Session Chair:** 张晓雷（西北工业大学）	8F 会议室
13:30–15:00	**学生论坛 I** **组织者：**洪青阳（厦门大学）、连政（同济大学）、孔秋强（香港中文大学） **特邀嘉宾：**俞凯（上海交通大学）、武执政（香港中文大学（深圳））、王帅（南京大学）、韩静（湖南大学）、孔秋强（香港中文大学）、张晴晴（晴数智慧创始人）、白烨（腾讯）、李哲（香港大学） **主持人：**连政（同济大学）	沐辉厅 A

11 月 7 日		
时间	**内容**	**地点**
13:30–15:00	**特殊议题 8: 共情计算与认知交互 II** **组织者：**黄东延（深圳市智拓欣扬科技有限公司）、赵子平（天津师范大学）、张子兴（湖南大学）、李雅（北京邮电大学）、刘斌（中国科学院自动化研究所） **主持人：**刘斌（中国科学院自动化研究所）	沐辉厅 B
13:30–15:00	**特殊议题 2: 面向具身智能的空间听觉与人机交互 II** **组织者：**朱梦尧（苏州大学）、杜俊（中国科学技术大学）、卜辉（希尔贝壳）、薛浏蒙（南京大学） **主持人：**杜俊（中国科学技术大学）	沐辉厅 C
13:30–15:00	**Sat-A-P-1:** Speech Science, Cognitive Modeling and Healthcare **Session Chair:** 吴锡欣（香港中文大学）	2D+2R 会议室 海报区
15:00–15:30	**茶歇**	会议室前厅
15:30–17:00	**Sat-A-O-4:** Audio Foundation Models and Sound Event Understanding **Session Chair:** 康世胤（深圳市商汤科技有限公司）	7 号会议室
15:30–17:00	**Sat-A-O-5:** Target Speaker Extraction and Audio-Visual Speech Enhancement **Session Chair:** 潘佳慧（内蒙古大学）	8M 会议室
15:30–17:00	**Sat-A-O-6:** Speech Enhancement, Restoration and Noise Control **Session Chair:** 王中秋（南方科技大学）	8F 会议室
15:30–17:00	**学生论坛 II** **组织者：**洪青阳（厦门大学）、连政（同济大学）、孔秋强（香港中文大学） **特邀嘉宾：**俞凯（上海交通大学）、武执政（香港中文大学（深圳））、王帅（南京大学）、韩静（湖南大学）、孔秋强（香港中文大学）、张晴晴（晴数智慧创始人）、白烨（腾讯）、李哲（香港大学） **主持人：**连政（同济大学）	沐辉厅 A
15:30–17:00	**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** **组织者：**王甦菁（中国科学院心理研究所）、吴锡欣（香港中文大学）、袁鑫（武汉科技大学）、燕楠（中国科学院深圳先进技术研究院）、于仕琪（南方科技大学）、陈莹（南京理工大学）、姚登峰（北京联合大学）、胡雯（福建师范大学） **主持人：**王甦菁（中国科学院心理研究所）	沐辉厅 B
15:30–17:00	**特殊议题 6: 面向生成式音频的可信音频处理方法** **组织者：**刘学琛（西交利物浦大学）、李圣辰（西交利物浦大学）、朱美能（国际关系学院） **主持人：**李圣辰（西交利物浦大学）	沐辉厅 C
15:30–17:00	**Sat-A-P-2:** Speech Recognition, Audio Understanding and Multimodal Interaction **Session Chair:** 武执政（香港中文大学（深圳））	2D+2R 会议室 海报区
17:20–18:05	**大会主旨报告 3** **主持人：**谢磊（西北工业大学） **报告题目：**四十年从理想到现实 **报告嘉宾：**颜永红（中国科学院声学研究所）	沐辉厅 A+B+C
18:30–20:30	**大会晚宴**	天沐厅
20:30–22:30	基金分享沙龙 **主持人：**陈谐（上海交通大学） **分享嘉宾：**邹月娴（北京大学）、王岚（中国科学院深圳先进技术研究院）、钱彦旻（上海交通大学）、韩静（湖南大学）、张子兴（湖南大学）、连政（同济大学）、陈谐（上海交通大学）	沐辉厅 A+B+C

11 月 8 日		
时间	**内容**	**地点**
08:00–18:00	**会议报到**	会议室前厅
09:00–18:00	**赞助商展览**	会议室前厅
09:00–09:45	**大会主旨报告 4** **主持人：**凌震华（中国科学技术大学） **报告题目：**YuE2: Bringing the Human into the Loop of Symbiotic Creativity in Music **报告嘉宾：**郭毅可（香港科技大学）	沐辉厅 A+B+C
09:45–10:30	**大会主旨报告 5** **主持人：**刘瑞（内蒙古大学） **报告题目：**迈向全模态交互模型 **报告嘉宾：**邱锡鹏（上海创智学院/复旦大学）	沐辉厅 A+B+C
10:30–11:00	**茶歇**	会议室前厅
11:00–12:30	**Sun-M-O-1:** Spotlight Paper Presentation I **Session Chair:** 凌震华（中国科学技术大学）	7 号会议室
11:00–12:30	**Sun-M-O-2:** Speech Spoofing and Content Manipulation Detection **Session Chair:** 赵子平（天津师范大学）	8M 会议室
11:00–12:30	**Sun-M-O-3:** Generative Speech Synthesis and Latent Representations **Session Chair:** 何佳毅（中国科学院自动化研究所）	8F 会议室
11:00–12:30	**青年论坛 2** **主持人：**徐珑婷（东华大学） **青年报告 4（11:00-11:30）** **报告题目：**融入听觉感知机理的语音增强方法研究 **报告嘉宾：**李楠（天津师范大学） **青年报告 5（11:30-12:00）** **报告题目：**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 **报告嘉宾：**胡诗超（苏州大学） **青年报告 6（12:00-12:30）** **报告题目：**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 **报告嘉宾：**徐春阳（苏州大学）	沐辉厅 A
11:00–12:30	**特殊议题 7: 语音与音乐的动态交互** **组织者：**杜忆（中国科学院心理研究所）、丁鼐（浙江大学） **主持人：**杜忆（中国科学院心理研究所）	沐辉厅 B
11:00–12:30	**特殊议题 3: 第三届多语言智能语音与语言技术研讨会** **组织者：**欧智坚（清华大学）、洪青阳（厦门大学） **主持人：**欧智坚（清华大学）、洪青阳（厦门大学）	沐辉厅 C
11:00–12:30	**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** **组织者：**钱彦旻（上海交通大学）、黄公平（武汉大学）、张王优（上海交通大学）、白仲鑫（武汉大学） **主持人：**黄公平（武汉大学）	2D 会议室
11:00–12:30	**特殊议题 10: 语音模型评测体系构建与开源生态创新** **组织者：**俞凯（上海交通大学）、陈谐（上海交通大学）、卜辉（希尔贝壳）、谢磊（西北工业大学）、王帅（南京大学）、朱梦尧（苏州大学） **主持人：**王帅（南京大学）	2R 会议室
12:30–14:00	**午餐**	天沐厅

11 月 8 日		
时间	**内容**	**地点**
12:30–14:00	**NCMMSC 常设机构会议**	7 号会议室
14:00–15:30	**Sun-A-O-1:** Spotlight Paper Presentation II **Session Chair:** 凌震华（中国科学技术大学）	7 号会议室
14:00–15:30	**Sun-A-O-2:** Multimodal Emotion, Sentiment and Group Affect Recognition **Session Chair:** 左昊麟（内蒙古大学）	8M 会议室
14:00–15:30	**Sun-A-O-3:** Spoken Dialogue Systems, Turn-Taking and Interaction Modeling **Session Chair:** 潘宁宁（西南财经大学）	8F 会议室
14:00–15:30	**青年论坛 3** **主持人：**燕楠（中国科学院深圳先进技术研究院） **青年报告 7（14:00-14:30）** **报告题目：**基于医患对话的老年认知健康智能诊断方法研究 **报告嘉宾：**潘怡霖（大连海事大学） **青年报告 8（14:30-15:00）** **报告题目：**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 **报告嘉宾：**薛浏蒙（南京大学） **青年报告 9（15:00-15:30）** **报告题目：**结合语音基础模型的障碍语音识别建模方法 **报告嘉宾：**谢旭荣（中国科学院）	沐辉厅 A
14:00–15:30	**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** **组织者：**罗颖艺（中国社会科学院语言研究所）、李蕙心（香港城市大学翻译及语言学系） **主持人：**李蕙心（香港城市大学）	沐辉厅 B
14:00–15:30	**特殊议题 4: 复杂声学场景下的目标音频提取** **组织者：**王帅（南京大学）、李明（香港中文大学（深圳））、钱馨园（北京科技大学）、关键（哈尔滨工程大学） **主持人：**李明（香港中文大学（深圳））	沐辉厅 C
14:00–15:30	**Sun-A-P-1:** Speech and Audio Generation, Coding and Restoration **Session Chair:** 张结（中国科学技术大学）	2D+2R 会议室 海报区
15:30–15:45	**茶歇**	会议室前厅
15:45–17:15	**Sun-A-O-4:** Speech and Multimodal Intelligence for Health and Cognitive Assessment **Session Chair:** 杨继臣（广东技术师范大学）	7 号会议室
15:45–17:15	**Sun-A-O-5:** Speech and Audio-Visual Resources, Retrieval and Evaluation **Session Chair:** 罗颖艺（中国社会科学院语言研究所）	8M 会议室
15:45–17:15	**Sun-A-O-6:** Speech, Audio and Visual Speech Generation **Session Chair:** 秦霄羽（清华大学）	8F 会议室

11 月 8 日		
时间	**内容**	**地点**
15:45–17:15	**青年论坛 4** **主持人：**燕楠（中国科学院深圳先进技术研究院） **青年报告 10（15:45-16:15）** **报告题目：**音频推理与拟人情感化语音生成技术进展 **报告嘉宾：**刘李（香港科技大学（广州）） **青年报告 11（16:15-16:45）** **报告题目：**“万卷 · 丝路”多模态小语种语料库的创新与实践 **报告嘉宾：**王潇蔓（上海人工智能实验室） **青年报告 12（16:45-17:15）** **报告题目：**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 **报告嘉宾：**拉巴顿珠（西藏大学）	沐辉厅 A
15:45–17:15	**特殊议题 11: 2026 年 CCF 先进音频技术竞赛** **组织者：**卜辉（希尔贝壳）、凌震华（中国科学技术大学）、黎椿键（华为终端） **主持人：**卜辉（希尔贝壳）	沐辉厅 C
15:45–17:15	**Sun-A-P-2:** Robust Speech Processing, Speaker Modeling and Audio Security **Session Chair:** 解焱陆（北京语言大学）	2D+2R 会议室 海报区
17:35–18:00	**大会闭幕式** **主持人：**吴志勇（清华大学深圳国际研究生院）	沐辉厅 A+B+C

嘉宾信息

Guest Profiles

大会报告 / Plenary Talks

Sat-M-PLEN-1: The Way of Multimodal Communication: From Emotion to Health

报告人 / Presenter: Björn W. Schuller（德国慕尼黑工业大学）

场次时间 / Session: November 7, 2026, 9:00 AM to 9:45 AM

地点 / Room: 沐辉厅 A+B+C

嘉宾简介 / Biography

Professor Björn W. Schuller is a leading force in speech intelligence, affective computing, and multimodal health AI. His research has helped define how machines analyse human voice, recognise emotion, and interpret behaviour for health and intelligent interaction. He holds the Chair of Health Informatics at the Technical University of Munich and is Professor of Artificial Intelligence at Imperial College London, where he leads the Group on Language, Audio & Music. As co-founder and Chief Scientific Officer of audEERING, he brings this research into commercial audio intelligence. His record includes 1,500+ publications, 90,000+ citations, an h-index above 130, and more than 50 awards. He holds Fellowships of ACM, IEEE, ISCA, ELLIS, BCS, and AAAC, recognising contributions across computer audition, affective computing, and health informatics. An ERC Starting Grant recipient, he was selected by the World Economic Forum among 40 extraordinary scientists under 40. His leadership extends across the institutions that shape the discipline: former Editor-in-Chief of IEEE Transactions on Affective Computing, Editor-in-Chief of AI Open, Field Chief Editor of Frontiers in Digital Health, and President-Emeritus of the Association for the Advancement of Affective Computing. An ACM Distinguished Speaker and IEEE Signal Processing Society Distinguished Lecturer, Schuller’s work spans the full arc of the field—from foundational methods in speech and emotion recognition to generative, multimodal systems for human interaction and health. His expertise is sought by companies including Huawei, Samsung, Barclays, and GN, while 300+ media appearances and contributions—spanning The Guardian, The Times, Scientific American, Newsweek, and MIT Technology Review—bring his work into the global conversation on AI.

Sat-M-PLEN-2: Speaker Tracking: Segment-online Multi-talker Organization with a Varying Number of Speakers

报告人 / Presenter: 汪德亮（香港中文大学（深圳））

场次时间 / Session: November 7, 2026, 9:45 AM to 10:30 AM

地点 / Room: 沐辉厅 A+B+C

嘉宾简介 / Biography

DeLiang Wang received the B.S. degree in 1983 and the M.S. degree in 1986 from Peking (Beijing) University, and the Ph.D. degree in 1991 from the University of Southern California, all in computer science. From 1991 to 2025, he was with the Department of Computer Science and Engineering and the Center for Cognitive and Brain Sciences, Ohio State University, where he was a Professor and University Distinguished Scholar. Since July 2025, he has been with the School of Data Science, the Chinese University of Hong Kong, Shenzhen, where he is a Presidential Chair Professor. His research interests include machine perception and deep learning. He was the recipient of the 1996 Office of Naval Research Young Investigator Award, and the 2008 Helmholtz Award from the International Neural Network Society, and the 2025 Neural Networks Pioneer Award from the IEEE Computational Intelligence Society. He also received the 2007 Outstanding Paper Award from the IEEE Computational Intelligence Society, the 2019 Best Paper Award and 2025 Donald G. Fink Overview Paper Award from the IEEE Signal Processing Society. He is a Fellow of IEEE, ISCA and AAIA, and served as the Editor-in-Chief of Neural Networks from 2011 to 2025.

Sat-A-PLEN-1: 四十年从理想到现实

报告人 / Presenter: 颜永红（中国科学院声学研究所）

场次时间 / Session: November 7, 2026, 5:20 PM to 6:05 PM

地点 / Room: 沐辉厅 A+B+C

嘉宾简介 / Biography

颜永红，中国科学院声学研究所首席科学家，1990 年毕业于清华大学电子工程系获学士学位，1995 年于美国俄勒冈研究院获计算机科学博士。曾任美国 OGI 副教授（1998）、英特尔人机交互学术委员会主席、英特尔中国研究中心主任（1998-2001）。从事语音信号处理和识别的研究四十年，牵头承担国家自然基金委重大项目、863、科技支撑、重点领域专项等国家级科研项目十余项，是国家自然基金委国家杰出青年基金获得者、全国优秀科技工作者称号获得者。发表论文 500 余篇，持有发明专利 100 余项，2006 年带领团队率先打破了外企对中国语音识别市场的垄断，为中国移动、中国电信、中国联通、华为、腾讯、阿里、百度等企业提供了语音识别底层技术，推动了国内语音识别产业从 0 到 1 的跨越和发展。

Sun-M-PLEN-1: YuE2: Bringing the Human into the Loop of Symbiotic Creativity in Music

报告人 / Presenter: 郭毅可（香港科技大学）

场次时间 / Session: November 8, 2026, 9:00 AM to 9:45 AM

地点 / Room: 沐辉厅 A+B+C

嘉宾简介 / Biography

Professor Yike Guo is a world-renowned computer scientist and a foreign member of the Chinese Academy of Engineering. He currently serves as Provost of the Hong Kong University of Science and Technology, Chair Professor in the Department of Computer Science and Engineering, and Founding Director of the Hong Kong Generative AI Research and Development Center (HKGAI). With his main research interests in machine learning and large-scale data management, he has been leading numerous large-scale AI research projects in Hong Kong, Mainland China, the UK, and other European countries. Previously, he was the Founding Director of the Data Science Institute at Imperial College London, and the Vice President (Research and Development) at Hong Kong Baptist University. In recognition of his work, he holds numerous prestigious titles, including Fellow of the Royal Academy of Engineering (FREng), Member of Academia Europaea (MAE), Fellow of the Hong Kong Academy of Engineering Sciences (FHKEng), Fellow of the Institute of Electrical and Electronics Engineers (FIEEE), Fellow of the British Computer Society (FBCS), and Fellow of the Chinese Association for Artificial Intelligence (FCAAI). He is also the recipient of 2023 Wu Wenjun Outstanding Contribution Award for Artificial Intelligence.

Sun-M-PLEN-2: 迈向全模态交互模型

报告人 / Presenter: 邱锡鹏（上海创智学院、复旦大学）

场次时间 / Session: November 8, 2026, 9:45 AM to 10:30 AM

地点 / Room: 沐辉厅 A+B+C

嘉宾简介 / Biography

邱锡鹏，上海创智学院副院长、复旦大学教授、模思智能创始人，CAAI Fellow，国家杰青获得者，主要研究方向为大模型、智能体以及科学工程等领域中的应用，发表学术论文被引用 4 万余次，5 次获得 AI 领域高水平会议的最佳/杰出等论文奖，主持研发了 MOSS 等系列高影响力开源大模型。获得 2022 年度钱伟长中文信息处理科学技术奖一等奖（第一完成人）、2022-2023 年度教育部 “高校计算机专业优秀教师奖励计划”，2024 年度 CCF-ACM 青年科技奖（每年仅 1 人）、2025 年度吴文俊人工智能自然科学奖一等奖（第一完成人）。著作《神经网络与深度学习》累计印刷 10 万余册，被上百家高校作为教材使用。

AACL 联合论坛 / AACL Joint Forum

Sat-A-AACL-1: Token Economics in the Agent Era: From Token Consumption to Intelligent Scheduling

报告人 / Presenter: 王昊奋（同济大学）

场次时间 / Session: November 7, 2026, 1:30 PM to 3:30 PM

报告时间 / Talk: 13:30-14:00

地点 / Room: 沐和厅 B+C

嘉宾简介 / Biography

Haofen Wang is a tenured professor and doctoral supervisor at Tongji University. His research focuses on knowledge graphs, knowledge-enhanced large language models, and industrial AI applications. His key contributions include KAG-Thinker, the world's first knowledge-augmented reasoning model; MemOS, a memory operating system; and AI-Ceping, a large language model evaluation platform, with applications in airport operations and financial risk management. He has authored six books and published over 100 papers, receiving more than 16,000 citations on Google Scholar. He has led more than ten national AI research projects, including key projects funded by the National Natural Science Foundation of China and research tasks under the New Generation Artificial Intelligence Major Project. Previously, he served as CTO of two AI unicorns, whose intelligent customer service systems have collectively handled over 10 billion service interactions. He currently serves as the inaugural rotating chair of OpenKG, the world's largest open knowledge graph community; deputy chair of the Terminology Committee and secretary-general of the Technical Committee on Natural Language Processing of the China Computer Federation (CCF); a council member of the Chinese Information Processing Society of China (CIPS), an executive committee member of its Technical Committee on Large Language Models, and deputy secretary-general of its Technical Committee on Language and Knowledge Computing; and deputy chair of the Technical Committee on Natural Language Processing of the Shanghai Computer Society.

Sat-A-AACL-2: Toward Multi-Layer Protection of Speech Against Real-World Misuse: What Can the Speech Community Do?

报告人 / Presenter: Junichi Yamagishi（日本国立信息研究院）

场次时间 / Session: November 7, 2026, 1:30 PM to 3:30 PM

报告时间 / Talk: 14:00-14:30

地点 / Room: 沐和厅 B+C

嘉宾简介 / Biography

Junichi Yamagishi received a Ph.D. from the Tokyo Institute of Technology in 2006 for a thesis that pioneered speaker-adaptive speech synthesis. He is a professor at the National Institute of Informatics, Tokyo, Japan. Previously, he held an EPSRC Career Acceleration Fellowship in the Centre for Speech Technology Research (CSTR) at the University of Edinburgh, U.K., from 2011 to 2016. He has authored and co-authored more than 400 refereed papers in international journals and conferences. He served as an elected member of the IEEE Speech and Language Technical Committee from 2013 to 2019, as an Associate Editor for the IEEE/ACM Transactions on Audio, Speech, and Language Processing from 2014 to 2017, as a chairperson of ISCA SynSIG from 2017 to 2021, as a Senior Associate Editor for the IEEE/ACM Transaction on Audio, Speech and Language Processing from 2019 to 2023, and as a technical committee of the Asia-Pacific Signal and Information Processing Association (APSIPA) Multimedia Security and Forensics from 2018 to 2023. He has been a member of the IEEE Signal Processing Society Education Board since 2021. He received the Itakura Prize from the Acoustic Society of Japan, the Kiyasu Special Industrial Achievement Award from the Information Processing Society of Japan, the Young Scientists' Prize from the Minister of Education, Science and Technology, the JSPS prize from the Japan Society for the Promotion of Science, the best paper award at the IEEE International Workshop on Information Forensics and Security (WIFS), the Docmo mobile science award, and IEEE Biometrics Council's BTAS/IJCB 5-Year Highest Impact Award in 2010, 2013, 2014, 2016, 2017, 2018, and 2023, respectively.

Sat-A-AACL-3: Challenges and Progress in Automatic Speech-to-Speech Translation: Bridging the Gap to Human Simultaneous Interpretation

报告人 / Presenter: Satoshi Nakamura（香港中文大学（深圳））

场次时间 / Session: November 7, 2026, 1:30 PM to 3:30 PM

报告时间 / Talk: 14:30-15:00

地点 / Room: 沐和厅 B+C

嘉宾简介 / Biography

Dr. Satoshi Nakamura is a Presidential Chair Professor at the School of AI, the Chinese University of Hong Kong, Shenzhen. He is also a professor emeritus at Nara Institute of Science and Technology (NAIST) and Honorarprofessor of Karlsruhe Institute of Technology, Germany. He received his B.S. from Kyoto Institute of Technology in 1981 and Ph.D. from Kyoto University in 1992. He was an Associate Professor in the Graduate School of Information Science at NAIST from 1994-2000. He was Department head and Director of ATR Spoken Language Communication Research Laboratories in 2000-2004, and 2005-2008, respectively, and Vice president of ATR in 2007-2008. He was Director General of Keihanna Research Laboratories and the Executive Director of Knowledge Creating Communication Research Center, National Institute of Information and Communications Technology, Japan 2009-2010. He moved to Nara Institute of Science and Technology as a full professor in 2011. He is a full professor at The Chinese University of Hong Kong, Shenzhen, China. His research interests include modeling and systems of spoken language processing, speech processing, spoken language translation, spoken dialog systems, natural language processing, and data science. He is one of the world leaders in speech-to-speech translation research. He has been serving various speech-to-speech translation research projects, including C-Star, A-Star, and the International Workshop on Spoken Language Translation IWSLT. He is currently the chairperson of ISCA SIG SLT (Spoken Language Translation). He was a committee member of IEEE SLTC 2016-2018. He was an Elected Board Member of the International Speech Communication Association, ISCA, from 2012 to 2019. He received the Antonio Zampolli Prize in 2012 and retained the titles of IEEE Fellow, ISCA Fellow, IPSJ Fellow, and ATR Fellow.

Sat-A-AACL-4: Multi-Agent Collaboration for Complex Tasks

报告人 / Presenter: Rui Yan（严睿）（武汉大学）

场次时间 / Session: November 7, 2026, 1:30 PM to 3:30 PM

报告时间 / Talk: 15:00-15:30

地点 / Room: 沐和厅 B+C

嘉宾简介 / Biography

Rui Yan is a full professor at Wuhan University. He was also recognized as a StarTrack scholar in Microsoft Research Asia and Beijing Academy of Artificial Intelligence (BAAI) scholar. He has published more than 100 research articles with more than 23,000 citations to date. His research interests include Natural Language Processing, Large Language Model and Agents. He regularly serves as a senior area chair/senior PC for top-tier venues and has been invited to give tutorial talks in a series of well-known conferences.

教程报告 / Tutorials

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

报告人 / Presenter: 连政（同济大学）

场次时间 / Session: November 6, 2026, 9:00 AM to 10:30 AM

地点 / Room: 8M 会议室

嘉宾简介 / Biography

连政（IEEE/CCF Senior Member），现任同济大学副教授，博士生导师。近年来已在情感计算、人机交互、大模型等领域开展一系列研究工作，累计发表论文和专利百余篇，包括 TPAMI, TNNLS, TASLP, TAFFC 等国际期刊和 ICML, NeurIPS, ICLR 等国际会议，Google Scholar 累计引用 5,400 余次（H-index: 38），入选全球前 2% 顶尖科学家榜单；作为" 多模态高鲁棒细微情感分析关键技术与系统" 的主要完成人之一，获得中国电子学会技术发明一等奖；发布了多模态情感识别数据库 MER，并围绕该数据库连续四年在 ACM Multimedia 和 IJCAI 上组织挑战赛和研讨会；担任 IEEE TAFFC/IEEE TASLP/PR Associate Editor，Information Fusion Area Editor，ACM Multimedia 和 ACL ARR Area Chair，ACM Multimedia 2026 Dataset Co-Chair。

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

报告人 / Presenter: 刘瑞（内蒙古大学）

场次时间 / Session: November 6, 2026, 9:00 AM to 10:30 AM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

刘瑞，内蒙古大学人工智能学院副院长，教授、博士生导师。入选第十届中国科协青年人才托举（青托）工程；主持国自然面上、国自然青年、内蒙古自治区杰出青年基金项目、中央引导地方科技发展计划项目、内蒙古自治区重点研发和成果转化计划项目、内蒙古自治区草原英才等 10 余项国家/省部级项目。主要研究方向为多语种多模态共情人机语音交互，相关成果以第一或通讯作者发表于 IEEE-TASLP、IEEE-TAFFC 和 ACL、ACM MM、AAAI 等领域顶级学术会议或期刊，其中 ESI 高被引 1 篇，IEEE 国际会议最佳论文奖 2 篇。获得内蒙古自治区科技进步一等奖、中国发明协会发明创业奖创新奖二等奖等。担任国际期刊 Information Fusion、IEEE-TAFFC、ACM-TALLIP 副编辑和 CCF 语音对话与听觉专委会执行委员、内蒙古计算机学会人工智能专委会主任、内蒙古自治区青年科技工作者协会人工智能与电子信息专委会秘书长等。

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

报告人 / Presenter: 陈丽萍（中国科学技术大学）

场次时间 / Session: November 6, 2026, 11:00 AM to 12:30 PM

地点 / Room: 8M 会议室

嘉宾简介 / Biography

陈丽萍于 2016 年在中国科学技术大学获得信号与信息处理专业博士学位。2016 年 7 月至 2022 年 12 月期间，她曾任职于微软公司，担任语音科学家。于 2023 年 1 月起，任中国科学技术大学特任副研究员。研究方向包括语音处理、语音隐私保护、语音生成以及说话人识别。

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

报告人 / Presenter: 钱馨园（北京科技大学）

场次时间 / Session: November 6, 2026, 11:00 AM to 12:30 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

钱馨园，北京科技大学计算机与通信工程学院副教授，主要研究方向为智能语音技术、多模态（视听觉）人机交互。博士毕业于英国伦敦玛丽女王大学计算机专业，2020 年至 2022 年于新加坡国立大学担任博士后研究员，曾在意大利 FBK 研究所、香港中文大学（深圳）从事研究工作。已在国际顶级会议及期刊上发表论文 80 余篇，担任多个国际会议/期刊的程序委员/客座编辑（如 IROS, ICASSP, ECAI, TCE 等）。主持或参与国家自然科学基金青年项目、北京市自然科学基金面上项目、科技创新 2030-新一代人工智能重大专项、北京市自然科学基金联合项目、CCF-腾讯犀牛鸟项目、深圳大数据研究院项目、新加坡人机交互项目及华为技术合作项目等。获评 ACM 新星奖（北京）、北京市" 高创计划" 青年托举人才、北京市图象图形学会最美女科技工作者。

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

报告人 / Presenter: 陈泽华（清华大学）

场次时间 / Session: November 6, 2026, 1:30 PM to 3:00 PM

地点 / Room: 8M 会议室

嘉宾简介 / Biography

陈泽华博士现为清华大学计算机系博士后，获" 水木学者" 计划资助，入选国家级博士后人才专项。此前博士毕业于英国帝国理工学院电气与电子工程系、任生数科技（Vidu）音频模态负责人。在概率生成模型（如扩散模型、桥类模型）及其在音-视频生成、健康监测等方面的应用持续开展研究工作，以第一或通讯作者（含共同）在 Nature Machine Intelligence, NeurIPS, ICML 等期刊与会议发表论文近 20 篇，其中多篇文章录取 CVPR, ACL, ACM MM, ICASSP 等会议 Highlight 或 Oral，文生音频工作 AudioLDM 单篇已获引用 1200 余次。

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

报告人 / Presenter: 江文斌（杭州电子科技大学）

场次时间 / Session: November 6, 2026, 3:30 PM to 5:00 PM

地点 / Room: 8M 会议室

嘉宾简介 / Biography

江文斌，杭州电子科技大学通信工程学院，特聘副教授，硕士生导师。上海交通大学电子工程系博士、计算机科学与工程系博士后。研究方向为智能语音信息处理，主要包括语音增强、语音唤醒、语音识别、语音理解大模型、低码率语音编解码与类脑听觉等。在 IEEE/ACM TASLP、IEEE SPL、ICASSP、Interspeech 等声学与语音领域顶级期刊与会议发表论文二十余篇，授权国家发明专利十余项；主持科技部长三角科技创新联合攻关课题、浙江省自然科学基金项目及多项企业横向项目；长期担任 IEEE TASLP、ICASSP、Interspeech、IEEE SLT、NCMMSC 等期刊与会议审稿人。

青年论坛 / Young Forums

Sat-M-YF-1: 低码率神经网络语音编解码方法研究

报告人 / Presenter: 艾杨（中国科学技术大学）

场次时间 / Session: November 7, 2026, 11:00 AM to 12:30 PM

报告时间 / Talk: 11:00–11:30

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

艾杨，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。主要研究方向包括语音编码、语音增强、语音合成、音频质量评价等，在语音技术领域知名期刊和会议上发表论文 70 余篇，其中第一作者/通讯作者论文近 60 篇。目前主持国家自然科学基金和安徽省自然科学基金等科研项目，参与战略性先导科技专项、国家重点研发计划等项目多项。在获奖方面，2024 年入选" 小米青年学者"，获得 Interspeech 2024 离散语音挑战赛（Discrete Speech Challenge）声码器赛道冠军（第一完成人）以及第十八届全国人机语音通讯学术会议最佳论文奖（通讯作者）等。

Sat-M-YF-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

报告人 / Presenter: 夏菁（中国科学院深圳先进技术研究院）

场次时间 / Session: November 7, 2026, 11:00 AM to 12:30 PM

报告时间 / Talk: 11:30–12:00

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

夏菁，中国科学院深圳先进技术研究院助理研究员、华南师范大学博士研究生，德国波鸿鲁尔大学博士联合培养。入选深圳市高层次人才、鹏城优才及南山区领航人才。长期聚焦视听通道认知与音乐治疗情绪障碍研究，致力于开发基于电生理信号的脑疾病智能诊断系统，开展抑郁等脑疾病相关研究。以主要作者在 Cell Reports、NeuroImage、Cortex 等国际期刊及 ICASSP、IEEE EMBC 等顶级会议发表认知神经科学与脑疾病相关论文 14 篇；申请脑电信号处理与脑疾病诊断相关专利 3 项、获授权 1 项。主持广东省自然科学基金、中国博士后科学基金、深圳市面上等项目 4 项，并参与国家重点研发计划、国自然面上项目、广东省重点及深圳市重点等项目 9 项。

Sat-M-YF-3: 迈向更稳定的全连续自回归语音合成

报告人 / Presenter: 郑达（小红书），王翰坤（上海交通大学），李波含（上海交通大学）

场次时间 / Session: November 7, 2026, 11:00 AM to 12:30 PM

报告时间 / Talk: 12:00–12:30

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

郑达硕士毕业于上海交通大学，导师为俞凯教授；曾搭建了业界领先的大规模中文语音识别系统 yitu-asr；在小红书 dots 语音团队进行文本和多模态的模型研发工作。目前在进行全双工语音方向的研究。

嘉宾简介 / Biography

王翰坤是上海交通大学跨媒体语言智能实验室 (X-LANCE) 博士生，导师为俞凯教授；本科毕业于哈尔滨工业大学。目前在小红书 dots 语音团队实习。他的主要方向为语音表征与合成，在 ICASSP, Interspeech, SLT, AAAI 等旗舰会议和期刊发表论文 10 余篇，并担任 ICASSP, Interspeech 等会议的审稿人。

嘉宾简介 / Biography

李波含于 2024 年获得上海交通大学计算机科学与技术专业（ACM 班）工学学士学位，目前在上海交通大学攻读博士学位，导师为俞凯教授。目前在小红书 dots 语音团队实习。他的研究方向主要为语音合成与理解。他已在语音与语言处理及机器学习领域的多个顶级会议上发表论文，包括 ICASSP、Interspeech、NeurIPS 和 AAAI 等。

Sun-M-YF-1: 融入听觉感知机理的语音增强方法研究

报告人 / Presenter: 李楠（天津师范大学）

场次时间 / Session: November 8, 2026, 11:00 AM to 12:30 PM

报告时间 / Talk: 11:00–11:30

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

李楠，天津师范大学计算机与信息工程学院讲师，天津大学计算机科学与技术专业工学博士。曾赴日本北陆先端科学技术大学院大学（JAIST）攻读双学位，博士期间获国家留学基金委资助赴新加坡国立大学开展联合培养。主要研究方向包括语音增强、语音端点检测、远场语音识别、语音去混响及听觉感知计算，重点关注听觉机理启发的低复杂度语音处理方法。已在 Expert Systems with Applications、Speech Communication、ICASSP、INTERSPEECH 等期刊和会议发表论文 16 篇，其中第一作者论文 7 篇、通讯作者论文 1 篇，并作为主要发明人申请发明专利 5 项。曾获 Intel 神经形态深度噪声抑制挑战赛 Top 2、音频深度合成检测挑战赛 Top 4 及方言识别挑战赛算法精英奖等荣誉。

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

报告人 / Presenter: 胡诗超（苏州大学）

场次时间 / Session: November 8, 2026, 11:00 AM to 12:30 PM

报告时间 / Talk: 11:30–12:00

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

胡诗超，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。本科、博士分别毕业于北京邮电大学和日本东北大学，主要研究方向包括空间听觉与智能声学等。曾在华为、腾讯音乐、Facebook 及日本理音从事音频与声学研究及产品研发。

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

报告人 / Presenter: 徐春阳（苏州大学）

场次时间 / Session: November 8, 2026, 11:00 AM to 12:30 PM

报告时间 / Talk: 12:00–12:30

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

徐春阳，苏州大学听觉与语言智能研究中心讲师，博士毕业于伦敦大学学院，曾任华为技术有限公司高级研究工程师。主要研究方向为心理声学与听觉感知，声场重现与控制，个性化空间音频，声品质设计与智能音频终端应用。

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

报告人 / Presenter: 潘怡霖（大连海事大学）

场次时间 / Session: November 8, 2026, 2:00 PM to 3:30 PM

报告时间 / Talk: 14:00–14:30

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

潘怡霖博士，大连海事大学人工智能学院讲师、硕士生导师，同济大学老龄语言与看护中心兼职研究员。于 2022 年博士毕业于英国谢菲尔德大学计算机科学专业，研究方向包括病理语音检测、多模态情感分析与智能语音对话等。自 2022 年入职大连海事大学以来，主持辽宁省自然科学基金一项，参与国家自然科学基金面上项目一项。近五年在 INTERSPEECH、ICASSP、IEEE TASLP、IEEE Signal Processing Letters、Computer Speech and Language 等语音领域权威会议及期刊发表论文十余篇，谷歌学术论文被引用量 400 余次。2018 年获得欧洲玛丽居里全额奖学金，并于 2022 年被认定为大连市青年才俊。

Sun-A-YF-1-2: 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

报告人 / Presenter: 薛浏蒙（南京大学）

场次时间 / Session: November 8, 2026, 2:00 PM to 3:30 PM

报告时间 / Talk: 14:30–15:00

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

薛浏蒙，南京大学智能科学与技术学院准聘助理教授。主要研究方向为智能语音，语音、音乐与音频理解与生成；情感语音生成和交互等。博士毕业于西北工业大学，曾在京东 AI Lab、腾讯 AI Lab、微软开展研究工作。此后，在香港中文大学（深圳）和香港科技大学从事博士后研究。作为联合发起人推出综合性音频生成与可视化平台，牵头主导开发统一音频理解与生成指令数据集 Audio-FLAN，一度位列 Hugging Face Dataset Trending 第二，下载量超过十万余次，下载单位包括 Google、Meta、Nvidia、字节跳动、腾讯、阿里巴巴、剑桥、牛津、清华、ETH Zurich AI Center、LAION 等国内外顶尖科技公司和知名高校研究组。参与 Llasa、Spark-TTS、YuE、AudioX 等语音、音乐与音频生成大模型研究。相关成果发表于 ACL、ICLR、ICASSP、INTERSPEECH、IEEE/ACM TASLP、Neural Networks 等高水平会议与期刊，相关工作受到 Apple、Amazon 等国际知名科技公司研究团队的关注与引用。参与组织 ISCSLP 2026 和 IEEE SLT 2024 国际会议；MLC-SLM@INTERSPEECH 2026、SmartGlasses@SLT2026、LLM4MA Workshop@ISMIR 2025、CoVoC@ISCSLP 2024 等学术挑战赛，长期担任 ACL、ACM MM、ICASSP、INTERSPEECH、IEEE/ACM TASLP、Speech Processing Letters、Speech Communication 等会议和期刊审稿人。

Sun-A-YF-1-3: 结合语音基础模型的障碍语音识别建模方法

报告人 / Presenter: 谢旭荣（中国科学院软件研究所）

场次时间 / Session: November 8, 2026, 2:00 PM to 3:30 PM

报告时间 / Talk: 15:00–15:30

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

谢旭荣现任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。他在香港中文大学电子工程系获得博士学位，并先后在中国科学院深圳先进技术研究院、香港中文大学系统工程与工程管理系工作学习。研究方向包括障碍语音处理、自适应技术、语音和语言建模、神经语音解码等，在相关领域发表期刊及会议论文 60 余篇，提出的神经网络声学模型贝叶斯自适应技术获 ICASSP 2019 最佳学生论文奖，语音基础模型障碍语音识别与检测技术论文入选 IEEE SPS 2024–2025 年 TASLP 期刊 Top 25 下载论文，以及获 ISCSLP 2024 多模态构音障碍严重程度评估挑战赛奖项、HHME 优秀论文等。承担国家自然科学基金青年项目、新一代人工智能重大项目子课题、国家重点研发计划项目子课题、中国科学院软件研究所重大项目子课题、中国科学院青年创新促进会会员项目等国家和中科院科研任务。

Sun-A-YF-1: 音频推理与拟人情感化语音生成技术进展

报告人 / Presenter: 刘李（香港科技大学（广州））

场次时间 / Session: November 8, 2026, 3:45 PM to 5:15 PM

报告时间 / Talk: 15:45–16:15

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

刘李目前是香港科技大学（广州）的副研究员、博导。她在法国格勒诺布尔阿尔卑斯大学 GIPSA-lab 获得博士学位。曾担任加拿大 Ryerson University 的博士后研究员。她的主要研究方向是语音处理、音视频理解与生成以及可信人工智能等。她目前以第一作者或通讯作者身份在这些领域已发表学术论文 75 篇，其中包括此领域顶级期刊和会议 TPAMI, TASLP, NeurIPS 等。她现担任国际 IEEE 机器学习信号处理委员会 MLSP Member Nominations & Election 分会主席。她曾担任 2022 年语音信号处理顶会 ICASSP 的 Local Chair (China site)。她获得了广东省青年拔尖人才以及深圳市海外高层次人才-孔雀人才称号。作为项目负责人，她主持国自然面上项目、国自然重点项目课题、国自然青年项目、广东省面上项目、广东省区域联合基金-青年基金项目、2024 CCF-腾讯犀牛鸟项目、2023/2025 腾讯 AI Lab 犀牛鸟专项计划、2025 年度 CCF-快手大模型探索者基金 2022 阿里巴巴创新研究计划项目、2022 以及 2024 腾讯公益创投计划等。她曾在 2017 年荣获法国 Sephora Berribi 数学与计算机领域女性科学家奖、IEEE Multimedia Signal Processing Rising Star Runner-up Award 以及 2024 年 CCF-腾讯犀牛鸟项目卓越奖。她团队文章获得 2024 年第 16 届国际社会机器人学会议 ICSR 最佳学生论文提名奖、CVPR 2025 CV4Animals Workshop 杰出论文以及 2022/2023 深圳市科协人工智能优秀论文奖。她团队在 2026 国际语音顶会 Interspeech 音频推理挑战赛中获得单模型赛道冠军和 Agent 赛道季军 (18 个国家和地区的 156 支队伍)。

Sun-A-YF-2: “万卷 · 丝路”多模态小语种语料库的创新与实践

报告人 / Presenter: 王潇蔓（上海人工智能实验室）

场次时间 / Session: November 8, 2026, 3:45 PM to 5:15 PM

报告时间 / Talk: 16:15–16:45

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

王潇蔓，上海人工智能实验室数据平台中心高级工程师，具有丰富的语料创新建设经验，长期从事多语种、多模态语料资源建设与数据治理相关工作，曾参与多项语料相关国家标准及团队标准的编写。研究方向主要包括多模态语料库建设、低资源语言数据构建、智能数据处理与高质量标注体系等。

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

报告人 / Presenter: 拉巴顿珠（西藏大学）

场次时间 / Session: November 8, 2026, 3:45 PM to 5:15 PM

报告时间 / Talk: 16:45–17:15

地点 / Room: 沐辉厅 A

嘉宾简介 / Biography

拉巴顿珠，藏族，博士，西藏大学信息科学技术学院副教授，博士生导师，主要研究方向为藏语智能信息处理、人工智能。长期从事计算机应用与藏文信息处理领域的教学科研工作。现任中国中文信息学会民族语言文字信息专业委员会（MLI）常务委员，CCF 专业会员、中国中文信息学会、中国民族语言学会、中国图象图形学学会等多家学术团体会员，西藏自治区" 成才杯" 大学生课外学术科技作品竞赛" 创业导师"。近年来，主持国家自然科学基金青年项目、教育部人文社会科学研究项目 (为 2021 年全区唯一获批项目，并获得" 免于鉴定" 结题)、西藏自治区自然科学基金青年项目、西藏自治区自然科学基金面上项目等 10 多项科研和人才项目；以主研身份参与国家重点研发项目等省部级及以上 10 余项；在国内外期刊和会议上发表学术论文近 30 篇；申请国家发明专利 5 项（其中，授权 2 项、3 项进入实质审查阶段）；先后研发了多项藏语自然语言处理应用软件，以及藏语语音识别、语音合成、机器翻译等智能交互系统，其研发的成果登记软件著作权 15 项；指导大学生专业竞赛荣获省部级及以上奖项近 20 项，指导大学生创新创业训练计划项目 10 余项；研究成果和技术在西藏自治区多家 IT 企业和高校中得到示范应用，得到良好的用户使用评价；积极参与学术前沿对话，先后应邀参加第七届北京国际藏学研讨会、MNLP 2026、NLPAI 2025、MLIP 2025、MLIP 2023、NCMMSC 2023、CNCC 2022、MNLP 2022、ICCPR 2021 等国内外学术交流 20 余场次；近两年，入选国家级青年人才候选人、2026 年全国新时代青年先锋、西藏自治区" 珠峰英才" 青年人才；教学成果荣获西藏自治区 2025 年研究生教学成果奖一等奖、西藏大学 2025 年研究生教学成果奖特等奖和本科教学成果奖三等奖；曾荣获两次博士研究生国家奖学金、西藏自治区优秀共青团干部、全国挑战杯优秀指导教师，以及西藏大学优秀硕士学位论文、优秀博士学位论文、优秀毕业研究生、研究生" 学术之星"、五四青年奖章、优秀团委书记、优秀教师等 30 余奖项及荣誉称号。

科普教育论坛 / Education and Outreach Forum

Sat-A-EDU-1: 人工智能的奥秘

报告人 / Presenter: 马少平（清华大学）

场次时间 / Session: November 7, 2026, 1:30 PM to 5:00 PM

地点 / Room: 9 号会议室

嘉宾简介 / Biography

马少平，清华大学计算机系长聘教授，知识表示与推理冠名教授。长期从事信息检索与推荐方面的研究工作，提出基于解释的信息推荐，获得 SIGIR 时间检验奖。多年从事人工智能教学工作，编写有《艾博士：深入浅出人工智能》《艾博士：深入浅出大语言模型》《图解人工智能》《人工智能通识（终身学习版）》，作为系列丛书主编组织编写《人工智能通识（小学版，初中版，高中版）》，3 到 8 年级、高一高二共 8 个年级、每个年级上下册共 16 本分年级丛书《人工智能》。

Sat-A-EDU-2: 人工智能-产业变革与青年未来

报告人 / Presenter: 洪青阳（厦门大学）

场次时间 / Session: November 7, 2026, 1:30 PM to 5:00 PM

地点 / Room: 9 号会议室

嘉宾简介 / Biography

洪青阳，厦门大学信息学院教授，主要研究方向为声纹识别、语音识别，先后主持国家自然科学基金三项，华为、海思、海信、滴滴等企业委托项目近二十项。牵头组建厦门大学智能语音实验室，带领 XMUSPEECH 团队获东方语种识别（OLR）、基于多模态信息的语音处理（MISP）等竞赛佳绩，研究成果落地华为手机、海信电视、声云录音笔等产品。成功研发闽南话识别与合成系统，填补区域应用空白。出版专著《语音识别：原理与应用》，已更新到第三版。获电子工业出版社" 优秀作者奖" 和华为" 优秀技术合作成果奖"。从事本科生、研究生的语音识别教学工作近二十年，多次受邀为中小学做 AI 科普报告。

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

报告人 / Presenter: 朱璇（三星电子中国研究院）

场次时间 / Session: November 7, 2026, 1:30 PM to 5:00 PM

地点 / Room: 9 号会议室

嘉宾简介 / Biography

朱璇，三星电子中国研究院研发副总裁，语言智能部门负责人。博士毕业于清华大学电子工程系，毕业后加入三星电子中国研究院，长期从事 AI 智能体、端侧大模型、口语对话系统、语音识别和合成、语音增强和分离、机器翻译和文本挖掘、音频分类和音乐检索等方面的研发工作，并且负责三星电子中国区 AI 产品相关技术的研究、高校和企业技术合作、以及产品商用化落地。自 2018 年以来，多次担任妇基会 Stem Girls 活动和科技馆青创大赛讲师。CCF 语音对话与听觉专委会办公室副主任、主任扩大会议成员，NCMMSC 委员会委员等。

工业论坛 / Industry Forums

Fri-A-IND-1-1: 听得懂，说得好，沟通无界——天籁语音大模型全新升级

报告人 / Presenter: 商世东（腾讯）

场次时间 / Session: November 6, 1:30 PM to 3:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

商世东，腾讯会议天籁实验室主任。拥有 20 余年音视频领域经验，曾任杜比中国和澳大利亚研发团队高级总监；2019 年加入腾讯，负责实时音频分析、处理与评估技术的研发和产品化。

Fri-A-IND-1-2: Unified Streaming ASR and Word-Level Turn Detection with Interleaved Audio-Text LLMs

报告人 / Presenter: 杨栋（腾讯）

场次时间 / Session: November 6, 1:30 PM to 3:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

杨栋，腾讯游戏语音算法负责人、专家研究员，研究领域包括实时语音增强、语音识别、语音合成及智能语音对话系统。

Fri-A-IND-1-3: 从语音对话到任务协作：语音交互基座模型的工业实践

报告人 / Presenter: Nova（腾讯）

场次时间 / Session: November 6, 1:30 PM to 3:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

Nova, 腾讯混元高级研究员, Hy-Realtime-3 核心贡献者, 深度参与强化学习及 Tool Call 能力建设。WavTokenizer、WavChat 等论文第一作者，Gander 论文通讯作者，在 CCF A 类会议发表论文 30 余篇，研究聚焦强化学习与语音 Agent。

Fri-A-IND-1-4: 从鸡尾酒会到热词检索——大模型 ASR 困难场景破局之路

报告人 / Presenter: 范利春（小米集团）

场次时间 / Session: November 6, 1:30 PM to 3:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

范利春，中国科学院自动化研究所博士，小米集团技术委员会语音交互方向负责人。深耕语音领域十余年，现负责小米智能座舱、手机、AIoT 等终端语音能力建设，涵盖拾音、唤醒、声纹、ASR、TTS 与全双工交互。

Fri-A-IND-1-5: 近期语音数据热点洞察

报告人 / Presenter: 胡婕（北京慧听科技有限公司）

场次时间 / Session: November 6, 1:30 PM to 3:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

胡婕，慧听科技商务总监，负责语音数据相关业务的市场推广及市场策略制定与执行，推动业务持续增长。

Fri-A-IND-2-1: 科大讯飞 Spark-Audio 序列大模型进展

报告人 / Presenter: 熊世富（科大讯飞）

场次时间 / Session: November 6, 3:30 PM to 5:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

熊世富，中国科学技术大学博士，科大讯飞 AI 研究院副院长。长期研究语音识别、同传、交互与语音基座大模型，参与讯飞语音输入法、听见会议系统、办公本转写系统、讯飞星火小星畅聊等产品研发。

Fri-A-IND-2-2: 基于连续潜变量自回归建模的 CuteTTS 模型

报告人 / Presenter: 梁敏（OPPO）

场次时间 / Session: November 6, 3:30 PM to 5:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

梁敏，OPPO AI 中心高级语音算法工程师，拥有超过 10 年语音合成研发经验，覆盖文本前端、声学模型与声码器等完整技术栈，关注自然度、表现力、个性化与实际交互体验。

Fri-A-IND-2-3: VoxCPM：连续自回归的语音生成模型

报告人 / Presenter: 曾国洋（面壁智能）

场次时间 / Session: November 6, 3:30 PM to 5:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

曾国洋，面壁智能联合创始人、首席技术官。曾参与 CPM 系列模型研发，开发 BMTrain、BMInf 等训练推理工具，曾任智源研究院语言大模型加速技术创新中心副主任，入选北京市科技新星计划及福布斯“2026 亚洲 30 位 30 岁以下精英榜”。

Fri-A-IND-2-4: SmoothConv 与 DuplexConv：面向语音交互的互补型普通话多人对话语料库

报告人 / Presenter: 王成有（QualiaLabs × 西北工业大学 ASLP）

场次时间 / Session: November 6, 3:30 PM to 5:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

王成有，西北工业大学 ASLP 实验室硕士研究生，师从谢磊教授，研究语音识别、多方言识别、语音理解大模型与全双工语音对话，参与 WenetSpeech-Wu、FastTurn 等研究。

Fri-A-IND-2-5: 猫王妙播音响：AI 音频产品生态的构建与创新实践

报告人 / Presenter: 曾德钧（猫王妙播音响）

场次时间 / Session: November 6, 3:30 PM to 5:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

曾德钧，猫王妙播音响创始人兼 CEO，中国电子音响行业协会副会长，资深电声专家与音响设计师，深耕行业近 50 年。曾获国家科技进步一等奖、9 次 CES 创新奖及中国音响行业终身成就奖，参与多项音频行业标准制定。

Fri-A-IND-2-6: 端侧多模态交互硬件的产业化实践——AI 语音智能体鼠标的方案与合作路径

报告人 / Presenter: 谢雨田（安徽声云智能科技有限公司）

场次时间 / Session: November 6, 3:30 PM to 5:00 PM

地点 / Room: 8F 会议室

嘉宾简介 / Biography

谢雨田，安徽声云智能科技有限公司副总经理，负责 AI 项目商务与产学研合作，具有多个人工智能项目从 0 到 1 的落地经验，现重点推进端侧多模态硬件成果转化。

会议议程

Session Program

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 9:00 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（同济大学）
Chair: 陈丽萍（中国科学技术大学）

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（内蒙古大学）
Chair: 李琳（厦门大学）

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（中国科学技术大学）
Chair: 陈丽萍（中国科学技术大学）

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（北京科技大学）
Chair: 李琳（厦门大学）

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）
Chair: 曹梦雪（中国社会科学院语言研究所）

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（杭州电子科技大学）
Chair: 曹梦雪（中国社会科学院语言研究所）

Fri-A-IND-1: Industry Forum I
Room: 8F 会议室
Type: Industry Forum
November 6, 1:30 PM to 3:00 PM
Chair: 朱璇（北京三星通信技术研究有限公司）

大会主席致辞：吴志勇（清华大学深圳国际研究生院）

Fri-A-IND-1-1: 听得懂，说得好，沟通无界——天籁语音大模型全新升级
Presenter: 商世东（腾讯）

Fri-A-IND-1-2: Unified Streaming ASR and Word-Level Turn Detection with Interleaved Audio-Text LLMs
Presenter: 杨栋（腾讯）

Fri-A-IND-1-3: 从语音对话到任务协作：语音交互基座模型的工业实践
Presenter: Nova（腾讯）

Fri-A-IND-1-4: 从鸡尾酒会到热词检索——大模型 ASR 困难场景破局之路
Presenter: 范利春（小米集团）

Fri-A-IND-1-5: 近期语音数据热点洞察
Presenter: 胡婕（北京慧听科技有限公司）

Fri-A-IND-2: Industry Forum II
Room: 8F 会议室
Type: Industry Forum
November 6, 3:30 PM to 5:00 PM
Chair: 王磊（新加坡华为维纳研究所）

Fri-A-IND-2-1: 科大讯飞 Spark-Audio 序列大模型进展
Presenter: 熊世富（科大讯飞）

Fri-A-IND-2-2: 基于连续潜变量自回归建模的 CuteTTS 模型
Presenter: 梁敏（OPPO）

Fri-A-IND-2-3: VoxCPM：连续自回归的语音生成模型
Presenter: 曾国洋（面壁智能）

Fri-A-IND-2-4: SmoothConv 与 DuplexConv：面向语音交互的互补型普通话多人对话语料库
Presenter: 王成有（QualiaLabs × 西北工业大学 ASLP）

Fri-A-IND-2-5: 猫王妙播音响：AI 音频产品生态的构建与创新实践
Presenter: 曾德钧（猫王妙播音响）

Fri-A-IND-2-6: 端侧多模态交互硬件的产业化实践——AI 语音智能体鼠标的方案与合作路径
Presenter: 谢雨田（安徽声云智能科技有限公司）

Fri-E-CCF: CCF 语音对话与听觉专委会工作会议
Room: 8M+8F 会议室
Type: Committee Meeting
November 6, 2026, 6:30 PM to 8:30 PM

Fri-E-CIPS: CIPS 语音信息专委会工作会议
Room: 8M+8F 会议室
Type: Committee Meeting
November 6, 2026, 8:30 PM to 10:00 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
Chair: 吴志勇（清华大学深圳国际研究生院）

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（德国慕尼黑工业大学）
Chair: 俞凯（上海交通大学）

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（香港中文大学（深圳））
Chair: 谢湘（北京理工大学）

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
Chair: 钱馨园（北京科技大学）

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 12:15 PM
Chair: 柯登峰（北京语言大学）

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: 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)

11:30-11:45
Sat-M-O-2-3: 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)

11:45-12:00
Sat-M-O-2-4: 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.)

12:00-12:15
Sat-M-O-2-5: 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
Chair: 高迎明（北京邮电大学）

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
Chair: 易江燕（清华大学）

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: UniSRCodec: 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（西北工业大学），薛浏蒙（南京大学），刘和鑫（南洋理工大学），石宪（阿里巴巴通义实验室），胡婕（北京慧听科技有限公司）
Chair: 薛浏蒙（南京大学）

11:00-11:15
Sat-M-SS-1-1: 闽南话智能语音系统
Presenter: 洪青阳（厦门大学）

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（上海交通大学、上海创智学院）

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（南京大学）

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（北京慧听科技有限公司）

12:15-12:30
Sat-M-SS-1-6: ChinaVoices Challenge 2026 竞赛总结与分析
Presenter: 廖育杰（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（苏州大学），杜俊（中国科学技术大学），卜辉（希尔贝壳），薛浏蒙（南京大学）
Chair: 朱梦尧（苏州大学）

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（英国萨里大学）

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（深度机智、北京中关村学院）

12:00-12:30
Sat-M-SS-2-3: 空间音频智能化生成
Presenter: 赵洲（浙江大学）

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（深圳市智拓欣扬科技有限公司），赵子平（天津师范大学），张子兴（湖南大学），李雅（北京邮电大学），刘斌（中国科学院自动化研究所）
Chair: 李雅（北京邮电大学）

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（天津师范大学）

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

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（深圳河套学院）

Sat-M-YF: Young Forum 1
Room: 沐辉厅 A
Type: Young Forum
November 7, 2026, 11:00 AM to 12:30 PM
Chair: 王雪涵（广州大学）

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（中国科学技术大学）

11:30–12:00
Sat-M-YF-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（中国科学院深圳先进技术研究院）

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（小红书），王翰坤（上海交通大学），李波含（上海交通大学）

Sat-A-EDU: 科普教育论坛
Room: 9 号会议室
Type: Education and Outreach Forum
November 7, 2026, 1:30 PM to 5:00 PM
Chair: 关键（哈尔滨工程大学）

Sat-A-EDU-1: 人工智能的奥秘
Presenter: 马少平（清华大学）

Sat-A-EDU-2: 人工智能-产业变革与青年未来
Presenter: 洪青阳（厦门大学）

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）

Sat-A-AACL: AACL 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: AACL Joint Forum
November 7, 2026, 1:30 PM to 3:30 PM
Chair: 俞凯（上海交通大学），张岳（西湖大学）

13:30-14:00
Sat-A-AACL-1: Token Economics in the Agent Era: From Token Consumption to Intelligent Scheduling
Presenter: 王昊奋（同济大学）

14:00-14:30
Sat-A-AACL-2: Toward Multi-Layer Protection of Speech Against Real-World Misuse: What Can the Speech Community Do?
Presenter: Junichi Yamagishi（日本国立信息研究院）

14:30-15:00
Sat-A-AACL-3: Challenges and Progress in Automatic Speech-to-Speech Translation: Bridging the Gap to Human Simultaneous Interpretation
Presenter: Satoshi Nakamura（香港中文大学（深圳））

15:00-15:30
Sat-A-AACL-4: Multi-Agent Collaboration for Complex Tasks
Presenter: Rui Yan（严睿）（武汉大学）

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）
Chair: 连政（同济大学）

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
Chair: 黄公平（武汉大学）

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
Chair: 雪巍（香港科技大学）

13:30-13:45
Sat-A-O-2-1: Probing-Guided Feature-Space Adaptation for Frozen Chinese HuBERT Embeddings
Haobo Wang (Nanjing Univesity of Posts and Telecommunictaions)*; Yiqiang Cai (Xi'an Jiaotong-Liverpool University); Yang Xu (Nanjing Univesity of Posts and Telecommunictaions); Shengchen Li (Xi'an Jiaotong-Liverpool University); Shao Xi (Nanjing Universuty 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
Chair: 张晓雷（西北工业大学）

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); Kuntharrgyal 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
Chair: 吴锡欣（香港中文大学）

1. Phonetics, Prosody and Speech Production (8 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: 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-1-3: 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-1-4: 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 Cinese Language and Culture,Nankai University); Xudong Shen (School of Data Information, Tsinghua University); Baoying Wen (College of Cinese Language and Culture,Nankai University)

Sat-A-P-1-5: 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-1-6: 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-1-7: 《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-1-8: 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 (10 papers)

Sat-A-P-1-9: 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-1-10: 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-1-11: 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-1-12: 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-1-13: 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 Unviersity of Hong Kong); Grace Wenling Cao (University College Dublin); Pik Ki Peggy Mok (The Chinese University of Hong Kong)

Sat-A-P-1-14: 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-1-15: 阳江闸坡蜑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-1-16: 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-1-17: Effects of Noise on the Automatic Recognition of Cantonese Tones: A Speaker-Independent XGBoost Classification Study
Wenqi DONG (University of Macau); Jingwei ZHANG (University of Macau)*

Sat-A-P-1-18: 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)*

3. Pronunciation Assessment and Child Speech (3 papers)

Sat-A-P-1-19: 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-1-20: 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-1-21: Multi-Factor Feedback and Scoring compensation for Automatic Pronunciation Assessment
ren ling (cvte); li liang (cvte)*; lei yan (cvte)

4. Speech Perception, Hearing and Neural Modeling (8 papers)

Sat-A-P-1-22: 老年人群对 AI 合成普通话音节的识别研究
Wencui Liu (Tongji University)*; Ziying Zhou (Tongji University); Zhaoyue Wang (Tongji University); Chenqi Wang (Tongji University); Xiaohu Yang (Tongji University)

Sat-A-P-1-23: 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-1-24: 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-1-25: 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-1-26: 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-1-27: 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-1-28: 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-1-29: 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-1-30: 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-1-31: 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-1-32: Macro-Descriptors Derived from Large Language Models for Low-Resource Cantonese Aphasia Detection and Subtyping
YuFeng Nie (Visual Intellgence +X International Cooperation Joint Laboratory of MOE，Beijing Jiaotong University)*; Ying Qin (Visual Intellgence +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 Intellgence +X International Cooperation Joint Laboratory of MOE，Beijing Jiaotong University)

Sat-A-P-1-33: 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-1-34: 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-1-35: 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-1-36: 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-1-37: 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-1-38: 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 (6 papers)

Sat-A-P-1-39: 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-1-40: 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-1-41: 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-1-42: 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-1-43: 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-1-44: 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 (3 papers)

Sat-A-P-1-45: 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-1-46: 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-1-47: 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-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（深圳市智拓欣扬科技有限公司），赵子平（天津师范大学），张子兴（湖南大学），李雅（北京邮电大学），刘斌（中国科学院自动化研究所）
Chair: 刘斌（中国科学院自动化研究所）

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（深圳市智拓欣扬科技有限公司）

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（湖南大学）

14:30-15:00
Sat-A-SS-1-3: Swapped-Query Cascaded Fusion for Multimodal Emotion and Intent Joint Recognition
Presenter: 李雅（北京邮电大学）

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（苏州大学），杜俊（中国科学技术大学），卜辉（希尔贝壳），薛浏蒙（南京大学）
Chair: 杜俊（中国科学技术大学）

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）
Chair: 连政（同济大学）

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
Chair: 康世胤（深圳市商汤科技有限公司）

15:30-15:45
Sat-A-O-4-1: 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)

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.); Bowewn Wang (University of Bristol); hanqi li (Shanghai Jiao Tong University); Chaolei Liu (Shanghai Jiao Tong University); Weihan Chen (Shanghai Jiao Tong Univerisity); 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 Univeristy)

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
Chair: 潘佳慧（内蒙古大学）

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
Chair: 王中秋（南方科技大学）

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
Chair: 武执政（香港中文大学（深圳））

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
Lukuan 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); Yuhan 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 (8 papers)

Sat-A-P-2-11: 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-2-12: 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-2-13: 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-2-14: 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-2-15: TOASR: Breaking the Data Barrier in LLM-Based ASR via Text-Only Training
Qing Zhao (Beijing Honor Terminal Co., Ltd.; Megatronix (Beijing) Technology Co., Ltd.)*; Hongwei Cai (Beijing Jiaotong University; Megatronix(Beijing) technology Co., Ltd); Wei Liang (Megatronix(Beijing) technology Co., Ltd)

Sat-A-P-2-16: 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-2-17: 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-2-18: 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 (7 papers)

Sat-A-P-2-19: 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-2-20: 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)*

Sat-A-P-2-21: 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)*

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: 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-2-24: 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-2-25: 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 (10 papers)

Sat-A-P-2-26: 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-2-27: 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-2-28: 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-2-29: 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-2-30: 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-2-31: 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-2-32: 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-2-33: 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 Unerversity)

Sat-A-P-2-34: 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-2-35: 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-2-36: 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 Techonology); Guoqing Chen (South China University of Techonology)

Sat-A-P-2-37: 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-2-38: 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)*

Sat-A-P-2-39: 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)

6. Emotion, Sentiment and Intent Understanding (9 papers)

Sat-A-P-2-40: 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)*

Sat-A-P-2-41: 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-2-42: 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-2-43: Let Emotion Guide the Gate: Top-down Affective Modulation for Speech Emotion Recognition
Rui Wang (Bejing institute of Tecnology, Zhuhai); Ziyang Zhang (Northeast Agricultural University); Chenyang Li (Beijing institute of Tecnology, Zhuhai); Kun Qian (Beijing Institute of Technology)*; Bin Hu (Beijing institute of Tecnology); Björn Schuller (Technische Universität München)

Sat-A-P-2-44: 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-2-45: 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-2-46: 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-2-47: 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-2-48: 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 (3 papers)

Sat-A-P-2-49: 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-2-50: 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-2-51: 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 (8 papers)

Sat-A-P-2-52: How Little Poisoning Is Enough? Backdoor Vulnerabilities of Low-Resource Large Language Models
董时维（西藏大学）

Sat-A-P-2-53: 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-2-54: 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-2-55: Quantization-Robust LLM Unlearning via Module-Sensitive Updates
黄嘉兴（武汉大学）

Sat-A-P-2-56: 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-2-57: BVPA: Branch-Verified Preference Augmentation for Long-Horizon LLM Agents
王越余（国防科技大学）

Sat-A-P-2-58: 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-2-59: SEA: A Structured Evidence Adapter for LLM-Augmented Dialogue Relation Extraction
傅湘婷（吉林大学）

> **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（中国科学院心理研究所），吴锡欣（香港中文大学），袁鑫（武汉科技大学），燕楠（中国科学院深圳先进技术研究院），于仕琪（南方科技大学），陈莹（南京理工大学），姚登峰（北京联合大学），胡雯（福建师范大学）**
> **Chair: 王甦菁（中国科学院心理研究所）**

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）

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）

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）

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）

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）

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（西交利物浦大学），李圣辰（西交利物浦大学），朱美能（国际关系学院）
Chair: 李圣辰（西交利物浦大学）

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）

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）

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（南京邮电大学）

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）

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（中国科学院声学研究所）
Chair: 谢磊（西北工业大学）

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
Chair: 陈谐（上海交通大学）

分享嘉宾 / Guests

邹月娴　北京大学教授

王岚　中国科学院深圳先进技术研究院研究员

钱彦旻　上海交通大学计算机学院特聘教授

韩静　湖南大学教授

张子兴　湖南大学教授

连政　同济大学副教授

陈谐　上海交通大学副教授

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-PLEN-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（香港科技大学）
Chair: 凌震华（中国科学技术大学）

Sun-M-PLEN-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（上海创智学院、复旦大学）
Chair: 刘瑞（内蒙古大学）

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
Chair: 凌震华（中国科学技术大学）

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); Yuang 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
Chair: 赵子平（天津师范大学）

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
Chair: 何佳毅（中国科学院自动化研究所）

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（中国科学院心理研究所），丁鼐（浙江大学）
Chair: 杜忆（中国科学院心理研究所）

11:00-11:20
Sun-M-SS-1-1: 语音、音乐与人脑中的节奏
Presenter: 丁鼐（浙江大学）

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（香港中文大学）

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（上海纽约大学）

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（中国科学院心理研究所）

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（清华大学），洪青阳（厦门大学）
Chair: 欧智坚（清华大学），洪青阳（厦门大学）

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（西北工业大学）

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（上海交通大学、上海创智学院）

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（小米集团）

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）

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（上海交通大学），黄公平（武汉大学），张王优（上海交通大学），白仲鑫（武汉大学）
Chair: 黄公平（武汉大学）

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（武汉大学）

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（中国科学院声学研究所）

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（南京大学）

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（西北工业大学）

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（上海交通大学）

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（上海交通大学），陈谐（上海交通大学、上海创智学院），卜辉（希尔贝壳），谢磊（西北工业大学），王帅（南京大学），朱梦尧（苏州大学）
Chair: 王帅（南京大学）

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（上海交通大学）

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（希尔贝壳）

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（苏州大学）

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（西北工业大学）

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（南京大学）

Sun-M-YF: Young Forum 2
Room: 沐辉厅 A
Type: Young Forum
November 8, 2026, 11:00 AM to 12:30 PM
Chair: 徐珑婷（东华大学）

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（天津师范大学）

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（苏州大学）

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（苏州大学）

Sun-N-COMM: NCMMSC 常设机构会议
Room: 7 号会议室
Type: Committee Meeting
November 8, 2026, 12:30 PM to 2:00 PM

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
Chair: 凌震华（中国科学技术大学）

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
Shenghui Pan (XinJiang University)*; Lixu Sun (XinJiang University); Jian Peng (XinJiang University); Nurmemet 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
Chair: 左昊麟（内蒙古大学）

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: 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)

15:00-15:15
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
Chair: 潘宁宁（西南财经大学）

14:00-14:15
Sun-A-O-3-1: 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)

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)*; Kecan 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: 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); Kecan Mao (Beijing University of Posts and Telecommunications); Yingming Gao (Beijing University of Posts and Telecommunications); Ya Li (Beijing University of Posts and Telecommunications)

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
Chair: 张结（中国科学技术大学）

1. Expressive and Controllable Speech Synthesis (7 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: 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-1-6: MMSpeaker: A Multimodal Joint Framework for Expressive Controllable Speech Synthesis
Rui Niu (Tsinghua University); Zhiyong Wu (Tsinghua University)*

Sun-A-P-1-7: 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 (4 papers)

Sun-A-P-1-8: 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-1-9: 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-1-10: 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-1-11: 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)

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

Sun-A-P-1-12: MultiSwap: Cross-Modal Identity Conversion with Silent Video and Audio Inputs
Yifan Liu (Shanghai Jiao Tong University)*

Sun-A-P-1-13: 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-1-14: 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); Kuntharrgyal 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-1-15: 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-1-16: 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-1-17: 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-1-18: 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 Universit)

4. Neural Speech and Audio Coding (4 papers)

Sun-A-P-1-19: 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-1-20: 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-1-21: 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-1-22: 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 (7 papers)

Sun-A-P-1-23: 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-1-24: 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-1-25: 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-1-26: 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-1-27: 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-1-28: 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-1-29: 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)*

6. Music and Audio Generation, Understanding and Restoration (8 papers)

Sun-A-P-1-30: 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-1-31: 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-1-32: 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-1-33: 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-1-34: 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-1-35: 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-1-36: 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-1-37: LARA-Gen: Enabling Continuous Emotion Control for Music Generation Models via Latent Affective Representation Alignment
Jiahao Mei (Shanghai Jiaotong University)*; Xuenan Xu (Shanghai AI Lab, China); Zeyu Xie (Shanghai Jiaotong University); Zihao Zheng (Shanghai Jiaotong University); Ye Tao (Shanghai Jiaotong University); Yue Ding (Shanghai Mental Health Center); Mengyue Wu (Shanghai Jiaotong 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（中国社会科学院语言研究所），李蕙心（香港城市大学）
Chair: 李蕙心（香港城市大学）

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（香港城市大学）

14:15-14:30
Sun-A-SS-1-2: 粤方言韵尾功能负载研究
Presenter: 张劲松（北京语言大学）

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（中国社会科学院语言研究所）

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（中国社会科学院语言研究所）

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（中国社会科学院语言研究所）

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）

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（南京大学），李明（香港中文大学（深圳）），钱馨园（北京科技大学），关健（哈尔滨工程大学）
Chair: 李明（香港中文大学（深圳））

14:00-14:20
Sun-A-SS-2-1: 面向复杂声学场景的在线说话人日志与多说话人语音识别
Presenter: 李明（香港中文大学（深圳））

14:20-14:40
Sun-A-SS-2-2: Cross-Talk Speech Reduction, by Separation, for Separation
Presenter: 王中秋（南方科技大学）

14:40-15:00
Sun-A-SS-2-3: 真实场景下的目标语音提取：挑战与思考
Presenter: 王帅（南京大学）

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
Chair: 燕楠（中国科学院深圳先进技术研究院）

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（大连海事大学）

14:30–15:00
Sun-A-YF-1-2: 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（南京大学）

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（中国科学院软件研究所）

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
Chair: 杨继臣（广东技术师范大学）

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.); Xinchen 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 5:00 PM
Chair: 罗颖艺（中国社会科学院语言研究所）

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: Post-fusion Event-aware Time-Frequency Refinement for Audio Moment Retrieva
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 4:45 PM
Chair: 秦霄羽（清华大学）

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: 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)

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
Chair: 解焱陆（北京语言大学）

1. Predictive and Generative Speech Enhancement (10 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: 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 (4 papers)

Sun-A-P-2-11: 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-2-12: 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-2-13: Head Microphone Awaking for Geometry-Agnostic Speech Enhancement Using Distributed Microphone Arrays
chengqian Jiang (USTC)*; Jie zhang (USTC); Haowei Chen (USTC); Haoyin Yan (USTC)

Sun-A-P-2-14: Class-Balanced Dynamic Curriculum Learning for Sound Event Localization and Detection
Yufei Lu (Tianjin University)*; Jianguo Wei (Tianjin University); Wenhuan Lu (Tianjin University)

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

Sun-A-P-2-15: 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-2-16: 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-2-17: 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-2-18: 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-2-19: 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-2-20: 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-2-21: 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-2-22: 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-2-23: 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-2-24: 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-2-25: 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-2-26: 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-2-27: 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 (4 papers)

Sun-A-P-2-28: 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-2-29: 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-2-30: 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-2-31: 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)

6. Speaker Representations, Verification and Personalization (9 papers)

Sun-A-P-2-32: 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-2-33: 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-2-34: A Lightweight Overlapped Speech Detection Model for Edge Devices
Jacky Gao (Tencent Chengdu Branch)*

Sun-A-P-2-35: 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-2-36: 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-2-37: Auditing Textual-Attribute Explanations for Chinese Speaker Verification
Haobo Wang (Nanjing Univesity of Posts and Telecommunictaions)*; Zhuochen Hou (Nanjing Universuty of Posts and Telecommunication); Yang Xu (Nanjing Univesity of Posts and Telecommunictaions); Yiqiang Cai (Xi’an Jiaotong-Liverpool University); Xuechen Liu (Xi’an Jiaotong-Liverpool University); Xi Shao (Nanjing Universuty of Posts and Telecommunication)

Sun-A-P-2-38: 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-2-39: 声纹注册样本选择研究
Zhai Shusheng (厦门大学)*

Sun-A-P-2-40: 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 (7 papers)

Sun-A-P-2-41: Controllable Generative Speech Enhancement via ODE Truncation: A Case Study on Audio Deepfake Detection
Xinyu Yang (Beijing Institute of Technology, Zhuhai); Ziyu Wu (Beijing Institute of Technology, Zhuhai); Xiang Xie (Beijing Institute of Technology, Zhuhai)*

Sun-A-P-2-42: 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-2-43: 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-2-44: 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-2-45: 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-2-46: 多表征音频大模型重放检测
Yijie GAO (XMUspeech)*

Sun-A-P-2-47: 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); HOUJUN 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（希尔贝壳），凌震华（中国科学技术大学），黎椿键（华为终端）
Chair: 卜辉（希尔贝壳）

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: 声学与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-2: Young Forum 4
Room: 沐辉厅 A
Type: Young Forum
November 8, 2026, 3:45 PM to 5:15 PM
Chair: 燕楠（中国科学院深圳先进技术研究院）

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（香港科技大学（广州））

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（上海人工智能实验室）

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（西藏大学）

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
Chair: 吴志勇（清华大学深圳国际研究生院）

会务指南

Conference Information

1. 会议时间与地点

> **会议名称：**第二十一届全国人机语音通讯学术会议（NCMMSC 2026）
> **会议时间：**2026 年 11 月 5 日至 8 日
> **会议地点：**天沐琴台国际会展中心
> **会场地址：**中国珠海横琴天沐河人工岛
> **会议区域：**地下一层（B1）
> **用餐区域：**一层（F1）天沐厅

各项学术活动的具体时间与会场安排，请参阅本手册的简明议程和详细议程。

2. 会议报到

会议注册处位于**地下一层（B1）会议室前厅**，具体位置见 B1 会场平面图。报到时间安排如下：

日期	报到时间	地点
11 月 5 日（星期四）	14:00–17:00	B1 会议室前厅
11 月 6 日（星期五）	08:00–21:00	B1 会议室前厅
11 月 7 日（星期六）	08:00–18:00	B1 会议室前厅
11 月 8 日（星期日）	08:00–18:00	B1 会议室前厅

请于会议注册时领取餐券，并妥善保管。

3. 会场分布与平面图

3.1 B1 会场与赞助商展位

大会会议区域位于地下一层（B1），设有会议注册处、各会议室及赞助商展位。

B1 平面图中，**蓝色区域为 NCMMSC 2026 会场，红色区域为 AACL–IJCNLP 2026 会场**。图中同时标注了各会场入口、注册处、电梯间、卫生间及赞助商展位编号，供参会人员查阅。

功能或活动	地点
会议报到	B1 会议室前厅注册处
开幕式、大会主旨报告、闭幕式	B1 沐辉厅 A+B+C
教程报告	B1 的 8M、8F 会议室
其他论坛及论文交流	具体安排见会议日程
海报交流	B1 的 2D、2R 会议室海报区
赞助商展示	B1 前厅及通道展位区，具体位置见平面图
茶歇	B1 会议室前厅

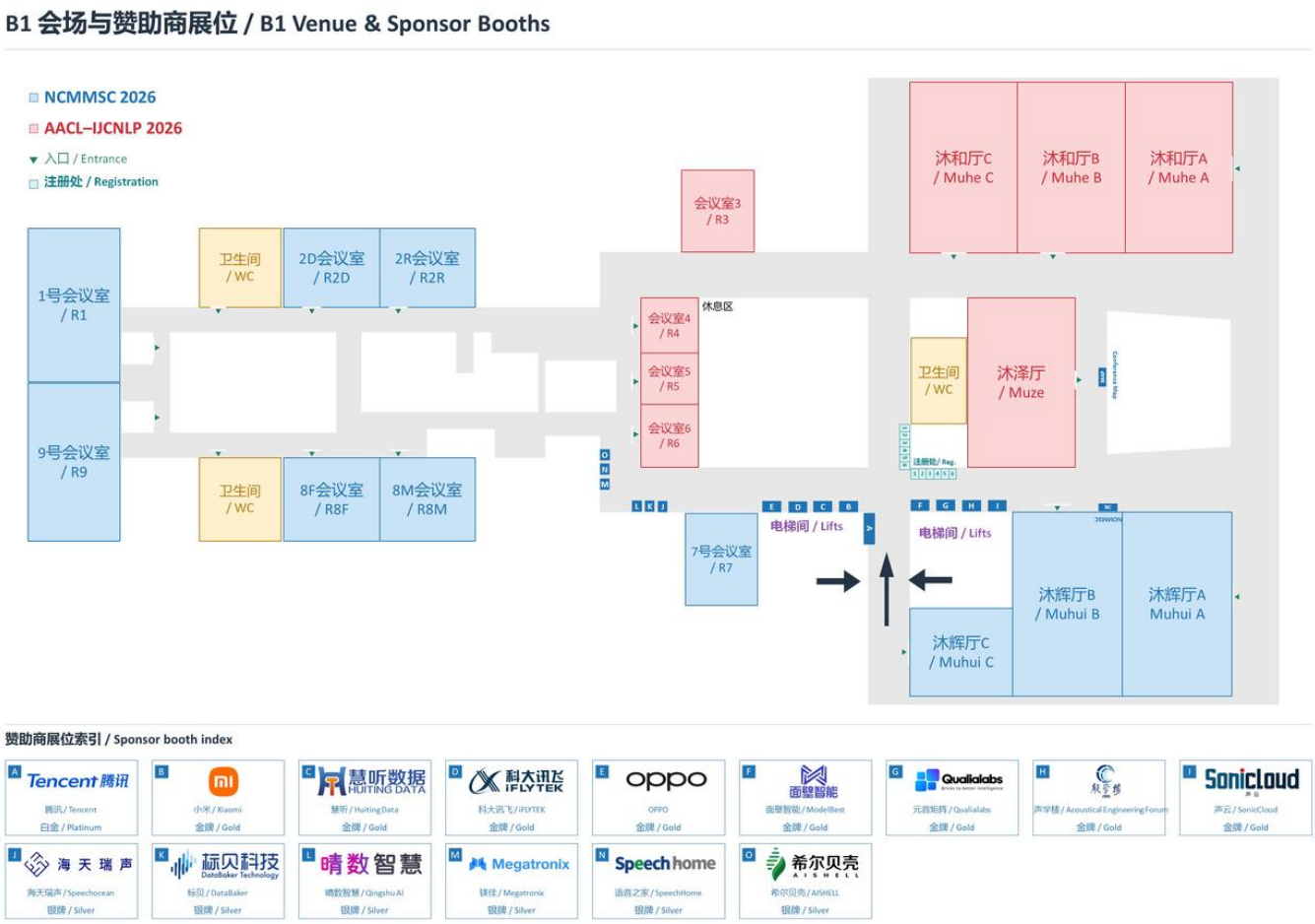


图 1　B1 会场与赞助商展位平面图

3.2 F1 用餐区域

会议用餐区域位于**一层（F1）天沐厅**。请按用餐时间前往，并凭餐券用餐。天沐厅位置见下图蓝色标注区域。



图 2　F1 用餐区域平面图

4. 会议用餐安排

参会人员**凭餐券用餐，餐券请于会议注册时领取**。具体安排如下：

日期	时间	用餐安排	地点	形式
11 月 6 日	12:30–13:30	午餐	F1 天沐厅	盒饭
11 月 6 日	17:00–18:30	欢迎晚宴	F1 天沐厅	自助餐
11 月 7 日	12:30–13:30	午餐	F1 天沐厅	盒饭
11 月 7 日	18:30–20:30	大会晚宴	F1 天沐厅	中式晚宴
11 月 8 日	12:30–14:00	午餐	F1 天沐厅	盒饭

5. 基本交通信息

5.1 会场导航

请在地图应用中搜索**“天沐琴台国际会展中心”**，并核对目的地位于珠海横琴天沐河人工岛。

抵达场馆后，请根据现场指引前往地下一层（B1）会议注册处。

5.2 主要抵达方式

出发地或抵达点	基本交通方式
珠海金湾机场	可乘出租车或网约车前往会场
珠海站	可乘出租车或网约车前往会场
横琴站、横琴口岸	可乘出租车或网约车前往会场
广州	可从广州南站乘铁路列车抵达珠海站，再换乘前往会场
香港	可选择经港珠澳大桥前往横琴的跨境巴士，抵达后换乘前往会场
澳门	可经横琴口岸通关后，换乘前往会场

具体班次、线路及预计用时，请以出行当日交通运营方和地图导航信息为准。