

杨钊



Tel: 188-2921-1941 | Email: zhaoyang9425@gmail.com |
出生年月: 1994.02 | 籍贯: 河北 邢台 | 政治面貌: 中共党员

教育经历

- 西安交通大学 | 计算机科学与技术 电子与信息学部 | 学术型博士研究生 2019.03—2025.03 (预计)
新加坡南洋理工大学 | 国家留学基金委 访问学生 2021.10—2023.10
研究兴趣: 语音识别、语音翻译; 语音多模态学习 (语音 + 文本/视觉); 预训练声学模型和语言模型的结合
导师: 赵季中教授、惠维教授 访问期间外导: Chng Eng Siong
- 西安交通大学 | 软件工程 软件学院 | 硕士研究生 2017.09—2018.12
西安电子科技大学 | 计算机科学与技术 计算机学院 | 工学学士 2013.09—2017.06
GPA: 3.7/4.0 排名: 5/159, 曾获国家励志奖学金、校一等奖学金

学术论文

- **Zhao Yang**, Yue Heng Yeo, Rui Jiang, Xiao Fu, Weiguang Chen, Wei Xi, Jizhong Zhao. Injecting Visual Features into Whisper for Parameter-Efficient Noise-Robust Audio-Visual Speech Recognition. *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2025)* (语音领域顶级会议)
- **Zhao Yang**, Rui Jiang, Xiao Fu, Wei Xi, Jizhong Zhao. Open-Modality Latent Modality Interaction Maximization for Audio-Visual Learning. *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2025)* (语音领域顶级会议)
- **Zhao Yang**, Dianwen Ng, Chong Zhang, Xiao Fu, Rui Jiang, Wei Xi, Yukun Ma, Chongjia Ni, Eng Siong Chng, Bin Ma, Jizhong Zhao. Dual Acoustic Linguistic Self-supervised Representation Learning for Cross-Domain Speech Recognition. *INTERSPEECH 2023 (oral)* (语音领域顶级会议)
- **Zhao Yang**, Dianwen Ng, Chong Zhang, Rui Jiang, Wei Xi, Yukun Ma, Chongjia Ni, Jizhong Zhao, Bin Ma, Eng Siong Chng. A Unified Recognition and Correction Model under Noisy and Accent Speech Conditions. *INTERSPEECH 2023* (语音领域顶级会议)
- **Zhao Yang**, Dianwen Ng, Xizhe Li, Chong Zhang, Rui Jiang, Wei Xi, Yukun Ma, Chongjia Ni, Jizhong Zhao, Bin Ma, Eng Siong Chng. Dual-Memory Multi-Modal Learning for Continual Spoken Keyword Spotting with Confidence Selection and Diversity Enhancement. *INTERSPEECH 2023* (语音领域顶级会议)
- Rui Jiang, **Zhao Yang**, Xiao Fu, Wei Xi, Jizhong Zhao. Chinese Speech Processing via Chinese Character Feature. *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2025)*
- Xiao Fu, Wei Xi, **Zhao Yang**, Rui Jiang, Dianwen Ng, Jia Yang, Jizhong Zhao. MRFER: Multi-channel Robust Feature Enhanced Fusion For Multi-modal Emotion Recognition. *2024 IEEE International Conference on Multimedia and Expo (ICME 2024)*
- Dianwen Ng, Ruixi Zhang, Jia Qi Yip, **Zhao Yang**, Jinjie Ni, Chong Zhang, Yukun Ma, Chongjia Ni, Eng Siong Chng, Bin Ma. De'hubert: Disentangling Noise in a Self-Supervised Model for Robust Speech Recognition. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2023)*
- Dianwen Ng, Yang Xiao, Jia Qi Yip, **Zhao Yang**, Biao Tian, Qiang Fu, Eng Siong Chng, Bin Ma. Small Footprint Multi-channel Network for Keyword Spotting with Centroid Based Awareness. *INTERSPEECH 2023*
- Dianwen Ng, Jia Qi Yip, Tanmay Surana, **Zhao Yang**, Chong Zhang, Yukun Ma, Chongjia Ni, Eng Siong Chng, Bin Ma. I2CR: Improving Noise Robustness on Keyword Spotting using Inter-Intra Contrastive Regularization. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2022)*
- Fan Yu, Wei Xi, **Zhao Yang**, Ziye Tong, Jingtong Sun. LRTD: A Low-rank Transformer with Dynamic Depth and Width for Speech Recognition. *International Joint Conference on Neural Networks (IJCNN 2022)*

- Fan Yu, Jiawei Guo, Wei Xi, **Zhao Yang**, Rui Jiang, Chao Zhang. Audio DistilBERT: a distilled audio BERT for speech representation learning. *International Joint Conference on Neural Networks (IJCNN 2021)*
- Yinhui Zhang, Wei Xi, **Zhao Yang**, Sitao Men, Rui Jiang, Yuxin Yang, Jizhong Zhao. Speech2Stroke: Generate Chinese Character Strokes Directly from Speech. *International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2020)*
- 刘志林, 杨颜瑜, **杨钊**, 赵鲲, 惠维. 基于卷积神经网络的单一特征及多特征融合阿尔茨海默症音频分类方法. 第十六届人机语音通讯会议 (*NCMMSC 2021*)
- **Zhao Yang**, Dianwen Ng, Xiao Fu, Liping Han, Wei Xi, Rui Wang, Rui Jiang, Jizhong Zhao. On the Effectiveness of Pinyin-Character Dual-Decoding for End-to-End Mandarin Chinese ASR. *arXiv 2022*

工程项目经历

小词汇量实时语音指令识别系统研发与部署 | 中国联通智能城市研究院 (远程) 2020.09—2020.12

- 搭建语音指令识别系统, 包含唤醒语音控制、打开商业体征、打开视频监控、返回主页等近 20 条语音指令。当用户对设备 (麦克风) 说出固定语音操作指令后, 将语音指令识别为对应的文字
- 负责包括训练数据采集、模型搭建与训练、接口集成 (同微单元 IOC 系统对接、测试) 等算法及工程化工作
- 数据集采集与处理: 利用语音合成 API 合成总计约 1 小时语音数据, 包含 1600 多条语音指令; 人工录制约 1 小时语音数据; 使用数据增强策略实现数据增广
- 指令识别模型结构: VGG-19、TC-ResNet 等
- 识别结果后处理: 根据模型识别概率对结果后处理, 高概率则输出结果, 低概率则忽略
- 端点检测: 采用基于阈值的 VAD 算法实时监测语音流, 通过设置合理的阈值, 用以区分语音和非语音

基于 LDA 主题模型的汉典古籍建模与分析 | 西安交通大学哲学系计算哲学实验室 2018.09 — 至今

- 利用潜在狄利克雷分配 (LDA, Latent Dirichlet Allocation) 主题模型对中国古籍语义、词源进行分析, 梳理传统文化传承与中国古代哲学的脉络
- 负责汉典古籍数据集爬取、TopicExplorer 工具更新、模型可视化与分析、网站维护等各种技术开发工作
- 参加清华大学计算哲学工作坊、华北电力大学马克思主义理论数字化研究技能研修班, 并作 LDA 主题建模技术宣讲报告

科研项目经历

基于数据空间的归因认知与智能决策 | 国家重点研发计划 | 参与人员 2020.11—2023.10

跨域协同的多模态高效感知与增强智能 | 国家自然科学基金重点项目 | 参与人员 2018.01—2020.12

荣誉奖励

- 2021.10 在第十六届人机语音通讯会议举办的“阿尔兹海默症识别竞赛”短音频赛道中荣获二等奖
- 2014.11 西安电子科技大学校级优秀学生

个人总结

- 具备扎实的计算机基础和语音领域的专业知识, 能进行创新性的学术研究和工程探索, 始终保持对科学研究的热情
- 具有较强的学习能力和接受新事物的能力。从博士期间孤独踏入语音领域时的“瞎折腾”, 到深入了解并确定自己的研究兴趣, 离不开自己独立探索和深入学习的努力; 从最初跨领域接触 LDA 主题模型时的懵懂, 到如今熟练掌握数字人文涉及的主要技术工作, 离不开日以继夜的理论学习和代码钻研
- 具备良好的沟通能力、协作精神和合作意识。连续两年参与阿里巴巴新加坡语音组的周会, 定期汇报自己的研究进展、遇到的问题及拟解决方法, 认真听取团队成员的意见和建议。同时, 不定期参与团队聚餐和公司活动, 始终与团队保持和谐融洽的合作关系
- 工作认真勤奋, 做人踏实稳重, 责任感强。在南洋理工大学访问交流期间, 在外导鼓励参与组会 (但没有强制要求) 的前提下, 其他访问交流学生纷纷放弃参加组会, 本人两年来坚持参加外导实验室的组会, 彼此分享前沿的研究成果, 从中获益匪浅