MOHAN SHI

✉ shimohan@g.ucla.edu (smohan.speech@gmail.com) · ☎ (+1) 310-954-6825 ·

🎓 EDUCATION

University of California, Los Angeles (UCLA)   Los Angeles, CA, USA
Ph.D. student, Electrical and Computer Engineering   *Sep. 2024 - present*

- Advisor: Prof. **Abeer Alwan** (Fellow of IEEE / ISCA / ASA)

University of Science and Technology of China (USTC)   Hefei, Anhui, China
Master of Enginnering, Electronic Engineering and Information Science   *Sep. 2021 - Jun. 2024*

- Advisor: Prof. **Li-Rong Dai**

Dalian University of Technology (DUT)   Dalian, Liaoning, China
Bachelor of Engineering, Electronic Information Engineering   *Sep. 2017 - Jun. 2021*

- GPA: 89.5 / 100   Rank: **1** / 185 (top 1%)

👥 EXPERIENCE

Tencent AI Lab   Bellevue, WA, USA (Remote)
Research Intern   *Sep. 2023 - Aug. 2024*
Mentor(s): Dr. **Yong Xu**, Dr. **Shi-Xiong (Austin) Zhang**

- Conducted research on large language model (LLM)-based multi-talker ASR, achieving state-of-the-art results on the evaluation set of the AMI corpus.
- Conducted research on end-to-end all-in-one model integrating multi-talker ASR, speaker diarization, and speech separation.

Alibaba Damo Academy   Hangzhou, Zhejiang, China
Research Intern   *Jul. 2022 - May. 2023*
Mentor(s): Dr. **Shiliang Zhang**

- Conducted research on speaker-attributed ASR ("who spoke what" in multi-talker scenarios) and proposed a "Context-Aware Speaker-Attributed ASR" model that outperforms previous works.
- Conducted research on "Semantic VAD", reducing tail segmentation latency without significantly impact on back-end performance.
- Compared Multi-Channel speaker-attributed ASR approaches with single-channel methods.
- Participated in organizing "The Second Multi-Channel Multi-Party Meeting Transcription Challenge" (M2MeT2.0), where I developed the baseline system for the speaker-attributed ASR task.

✏ PUBLICATIONS

- **Mohan Shi**, Zengrui Jin, Yaoxun Xu, Yong Xu, Shi-Xiong Zhang, Kun Wei, Yiwen Shao, Chunlei Zhang, Dong Yu. Advancing Multi-talker ASR Performance with Large Language Models. (SLT 2024).
- Zengrui Jin*, Yifan Yang*, **Mohan Shi***, Wei Kang, Xiaoyu Yang, Zengwei Yao, Fangjun Kuang, Liyong Guo, Lingwei Meng, Long Lin, Yong Xu, Shi-Xiong Zhang, Daniel Povey. LibriheavyMix: A 20,000-Hour Dataset for Single-Channel Reverberant Multi-Talker Speech Separation, ASR and Speaker Diarization. (Interspeech 2024, **Oral**).
- **Mohan Shi**, Zhihao Du, Qian Chen, Fan Yu, Yangze Li, Shiliang Zhang, Jie Zhang, Li-Rong Dai. CASA-ASR: Context-Aware Speaker-Attributed ASR. (Interspeech 2023, Poster).
- **Mohan Shi**, Yuchun Shu, Lingyun Zuo, Qian Chen, Shiliang Zhang, Jie Zhang, Li-Rong Dai. Semantic VAD: Low-Latency Voice Activity Detection for Speech Interaction. (Interspeech 2023, **Oral**).

- **Mohan Shi**, Jie Zhang, Zhihao Du, Fan Yu, Qian Chen, Shiliang Zhang, Li-Rong Dai. A Comparative Study on Multichannel Speaker-Attributed Automatic Speech Recognition in Multi-party Meetings. (APSIPA ASC 2023).

- Yuhao Liang, **Mohan Shi**, Fan Yu, Yangze Li, Shiliang Zhang, Zhihao Du, Lei Xie, Yanmin Qian, Jian Wu, Zhuo Chen, Kong Aik Lee, Zhijie Yan, Hui Bu. The Second Multi-Channel Multi-Party Meeting Transcription Challenge (M2MeT 2.0): A Benchmark for Speaker-Attributed ASR. (ASRU 2023).

- Yangze Li, Fan Yu, Yuhao Liang, Pengcheng Guo, **Mohan Shi**, Zhihao Du, Shiliang Zhang, Lei Xie. SA-Paraformer: Non-autoregressive End-to-End Speaker-Attributed ASR. (ASRU 2023).

- Weitai Zhang, Zhongyi Ye, Haitao Tang, Xiaoxi Li, Xinyuan Zhou, Jing Yang, Jianwei Cui, Pan Deng, **Mohan Shi** et al. The USTC-NELSLIP Offline Speech Translation Systems for IWSLT 2022. (IWSLT 2022).

🏆 Honors and Awards

UCLA ECE Fellowship	*Sep. 2024*
National Scholarship	*Nov. 2018*
Outstanding graduate of Liaoning Province	*Mar. 2021*

⚙ Skills

- Toolkits: PyTorch   ESPnet   Nemo   FunASR (Contributor)
- Languages: Mandarin (Native speaker)   English (TOEFL Internet-based Test: 91)