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**
- Research Topic: Speech-LLMs for Low-resource/Child Speech
- GPA: 4.0 / 4.0

University of Science and Technology of China (USTC)   Hefei, Anhui, China
Master of Engineering, 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 Rank: **1** / 185

👥 EXPERIENCE

UCLA - Speech Processing and Auditory Perception Lab   Los Angeles, CA, USA
*Graduate Student Researcher - Advisor: Prof. **Abeer Alwan***   *Sep. 2024 - Present*

- Conduct research on speech large language model (Speech-LLM)–based low-resource and child automatic speech recognition (ASR).
- Conduct research on discrete speech token representations for ASR and audio reconstruction. [2][3]

Microsoft Research   Redmond, WA, USA
*Research Intern - Manager: Dr. **Jinyu Li***   *Jun. 2025 - Sep. 2025*

- Conducted research on Speech-LLM–based joint ASR and speaker diarization, achieving state-of-the-art performance on both short and long audio while enabling streaming inference and training only on short audio. Submitted to *ICASSP 2026* [1] and was presented as part of an ASRU 2025 panel discussion.

Tencent AI Lab (US)   Bellevue, WA, USA (Remote)
*Research Intern - Manager: Dr. **Dong Yu***   *Sep. 2023 - Aug. 2024*

- Conducted research on LLM-based multi-talker ASR, achieving state-of-the-art results on a real-world meeting corpus. Paper accepted to *IEEE SLT 2024* [7].
- Conducted research on end-to-end all-in-one model integrating multi-talker ASR, speaker diarization, and speech separation.

Alibaba Speech Group   Hangzhou, Zhejiang, China
*Research Intern - Manager: Dr. **Shiliang Zhang***   *Jul. 2022 - May. 2023*

- Conducted research on speaker-attributed ASR ("who spoke what" in multi-talker scenarios) and developed a Context-Aware Speaker-Attributed ASR model that outperforms prior work. Paper accepted to *Interspeech 2023* [9].
- Conducted research on "Semantic VAD", reducing tail segmentation latency of voice activity detection without significantly impact on back-end performance. Paper accepted to *Interspeech 2023* [10].
- Compared Multi-Channel speaker-attributed ASR approaches with single-channel methods. Paper accepted to *APSIPA ASC 2023* [11].
- 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. Paper accepted to *ASRU 2023* [12].

✎ PUBLICATIONS

⧗ *Under Review*

[1] **Mohan Shi**, Xiong Xiao, Ruchao Fan, Shaoshi Ling, and Jinyu Li, *Train Short, Infer Long: Speech-LLM Enables Zero-Shot Streamable Joint ASR and Diarization on Long Audio*, submitted to *ICASSP 2026*.

[2] **Mohan Shi**, Natarajan Balaji Shankar, and Abeer Alwan, *Comparing Unsupervised and Supervised Semantic Speech Tokens: A Case Study of Child ASR*, accepted to *ASRU 2025 Satellite Workshop*.

[3] Kaiyuan Zhang*, **Mohan Shi***, Eray Eren, Natarajan Balaji Shankar, Zilai Wang, and Abeer Alwan, *STA-Codec: Semantic Token Assignment for Balancing Acoustic Fidelity and Semantic Information in Audio Codecs*, submitted to *ICASSP 2026*. (* equal contribution)

⊘ *Published*

[4] Natarajan Balaji Shankar, Zilai Wang, Kaiyuan Zhang, **Mohan Shi**, and Abeer Alwan, *CHSER: A Dataset and Case Study on Generative Speech Error Correction for Child ASR*, accepted to *Interspeech 2025*.

[5] Anyu Ying, Natarajan Balaji Shankar, Chyi-Jiunn Lin, **Mohan Shi**, Pu Wang, Hye-jin Shim, Siddhant Arora, Hugo Van Hamme, and Abeer Alwan, *Benchmarking Training Paradigms, Dataset Composition, and Model Scaling for Child ASR in ESPnet*, accepted to *WOCCI 2025*.

[6] Natarajan Balaji Shankar, Kaiyuan Zhang, Andre Mai, **Mohan Shi**, Alaria Long, Julie Washington, Robin Morris, and Abeer Alwan, *Leveraging ASR and LLMs for Automated Scoring and Feedback in Children's Spoken Language Assessments*, accepted to *SLaTE 2025*.

[7] **Mohan Shi**, Zengrui Jin, Yaoxun Xu, Yong Xu, Shi-Xiong Zhang, Kun Wei, Yiwen Shao, Chunlei Zhang, and Dong Yu, *Advancing Multi-talker ASR Performance with Large Language Models*, accepted to *SLT 2024*.

[8] 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, and Daniel Povey, *LibriheavyMix: A 20,000-Hour Dataset for Single-Channel Reverberant Multi-Talker Speech Separation, ASR and Speaker Diarization*, accepted to *Interspeech 2024*. (* equal contribution)

[9] **Mohan Shi**, Zhihao Du, Qian Chen, Fan Yu, Yangze Li, Shiliang Zhang, Jie Zhang, and Li-Rong Dai, *CASA-ASR: Context-Aware Speaker-Attributed ASR*, accepted to *Interspeech 2023*.

[10] **Mohan Shi**, Yuchun Shu, Lingyun Zuo, Qian Chen, Shiliang Zhang, Jie Zhang, and Li-Rong Dai, *Semantic VAD: Low-Latency Voice Activity Detection for Speech Interaction*, accepted to *Interspeech 2023*.

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

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

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

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

🏆 HONORS AND AWARDS

UCLA ECE Fellowship	*Sep. 2024*
Outstanding Graduate of USTC	*June. 2024*
Outstanding Graduate of Liaoning Province	*June. 2021*
National Scholarship	*Nov. 2018*

♜ ACADEMIC SERVICE

- Reviewer: ICASSP 2025/2026, Interspeech 2025, ASRU 2025, IEEE Signal Processing Letters (SPL)

⚙ SKILLS

- Toolkits: PyTorch   ESPnet   Nemo   FunASR (Contributor)
- Languages: Mandarin   English