

EDUCATION	<div><div><div>The University of Texas at Austin, Austin, USA</div><div>Ph.D. Student, Computer Science</div><div>Advisor: Prof. David Harwath</div></div><div><div>Shanghai Jiao Tong University, Shanghai, China</div><div>B.E. in Information Engineering</div><div>Advisor: Prof. Xie Chen</div></div><div><div>Shanghai Jiao Tong University, Shanghai, China</div><div>Zhiyuan Honors Program of Engineering</div></div></div> <div><div>August 2024 - Present</div><div>September 2020 - June 2024</div><div>September 2020 - June 2024</div></div>
EMPLOYMENT	<div><div><div>Amazon</div><div>Applied Scientist Intern</div></div><div><div>Microsoft Research Asia</div><div>Research Intern (Supervisor: Lei He, Xu Tan)</div></div><div><div>The University of Texas at Austin, SALT Lab</div><div>Visiting Scholar (Supervisor: David Harwath, Eunsol Choi)</div></div><div><div>Shanghai Jiao Tong University, X-Lance Lab</div><div>Research Intern (Supervisor: Xie Chen)</div></div></div> <div><div>May 2025 - Present, Seattle, USA</div><div>Apr 2024 - Aug 2024, Beijing, China</div><div>July 2023 - Jan 2024, Austin, USA</div><div>Dec 2021 - Jun 2024, Shanghai, China</div></div>
SELECTED AWARDS & HONORS	<div><div><div>Shanghai Outstanding Graduates</div><div>2024</div></div><div><div>SenseTime Scholarship for Undergraduate AI Researchers (30 winners nationwide each year), SenseTime</div><div>2023</div></div><div><div>Rongchang Science and Technology Innovation Scholarship (<0.1%)</div><div>Shanghai Rongchang Public Welfare Foundation</div><div>2023</div></div><div><div>Best Student Paper Shortlist, INTERSPEECH</div><div>2023</div></div><div><div>Zhiyuan College Honors Scholarship (TOP 10%), Zhiyuan College</div><div>2020-2024</div></div><div><div>Tencent Scholarship (TOP 2%), Tencent Technology (Shenzhen) Co., Ltd.</div><div>2021</div></div></div>
PUBLICATIONS	<div><div><div>[1] DRCap: Decoding CLAP Latents with Retrieval-Augmented Generation for Zero-shot Audio Captioning.</div><div>Xiquan Li, Wenxi Chen, Ziyang Ma, Xuenan Xu, Yuzhe Liang, Zhisheng Zheng, Qiuqiang Kong, Xie Chen</div><div>International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2025.</div></div><div><div>[2] SLAM-AAC: Enhancing Audio Captioning with Paraphrasing Augmentation and CLAP-Refine through LLMs.</div><div>Wenxi Chen, Ziyang Ma, Xiquan Li, Xuenan Xu, Yuzhe Liang, Zhisheng Zheng, Kai Yu, Xie Chen</div><div>International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2025.</div></div></div>
BEFORE PH.D.	<div><div><div>[3] EmoBox: Multilingual Multi-corpus Speech Emotion Recognition Toolkit and Benchmark.</div><div>Ziyang Ma, Mingjie Chen, Hezhao Zhang, Zhisheng Zheng, Wenxi Chen, Xiquan Li, Jiaxin Ye, Xie Chen, Thomas Hain</div><div>INTERSPEECH 2024.</div></div><div><div>[4] Emotion2vec: Self-Supervised Pre-Training for Speech Emotion Representation.</div><div>Ziyang Ma, Zhisheng Zheng, Jiaxin Ye, Jinchao Li, Zhifu Gao, Shiliang Zhang, Xie Chen</div><div>Findings of ACL, 2024.</div></div></div>

- [5] BAT: Learning to Reason about Spatial Sounds with Large Language Models.
Zhisheng Zheng, Puyuan Peng, Ziyang Ma, Xie Chen, Eunsol Choi, David Harwath
International Conference on Machine Learning (ICML), 2024.
 - [6] EAT: Self-Supervised Pre-Training with Efficient Audio Transformer.
Wenxi Chen, Yuzhe Liang, Ziyang Ma, **Zhisheng Zheng**, Xie Chen
International Joint Conference on Artificial Intelligence (IJCAI), 2024.
 - [7] SJTU-THU AUTOMATED AUDIO CAPTIONING SYSTEM FOR DCASE 2024.
Wenxi Chen, Xiquan Li, Ziyang Ma, Yuzhe Liang, **Zhisheng Zheng**, Anbai Jiang, Yanmin Qian, Pingyi Fan, Wei-Qiang Zhang, Cheng Lu, Jia Liu, Xie Chen
Detection and Classification of Acoustic Scenes and Events (DCASE Challenge), 2024.
 - [8] Leveraging speech ptm, text llm, and emotional tts for speech emotion recognition.
Ziyang Ma, Wen Wu, **Zhisheng Zheng**, Yiwei Guo, Qian Chen, Shiliang Zhang, Xie Chen
International Conference on Acoustics, Speech, and Signal Processing (ICASSP; *Oral*), 2024.
 - [9] Fast-HuBERT: An Efficient Training Framework for Self-Supervised Speech Representation Learning.
Guanrou Yang, Ziyang Ma, **Zhisheng Zheng**, Yakun Song, Zhikang Niu, Xie Chen
IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2023.
 - [10] Exploring Effective Distillation of Self-Supervised Speech Models for Automatic Speech Recognition.
Yujin Wang, Changli Tang, Ziyang Ma, **Zhisheng Zheng**, Xie Chen, Wei-Qiang Zhang
IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2023.
 - [11] MT4SSL: Boosting Self-Supervised Speech Representation Learning by Integrating Multiple Targets.
Ziyang Ma, **Zhisheng Zheng**, Changli Tang, Yujin Wang, Xie Chen
INTERSPEECH 2023. *Best student paper shortlist.*
 - [12] Unsupervised Active Learning: Optimizing Labeling Cost-Effectiveness for Automatic Speech Recognition.
Zhisheng Zheng, Ziyang Ma, Yu Wang, Xie Chen
INTERSPEECH 2023.
 - [13] Pushing the Limits of Unsupervised Unit Discovery for SSL Speech Representation.
Ziyang Ma, **Zhisheng Zheng**, Guanrou Yang, Yu Wang, Chao Zhang, Xie Chen
INTERSPEECH 2023.
 - [14] Front-End Adapter: Adapting Front-End Input of Speech based Self-Supervised Learning for Speech Recognition.
Xie Chen, Ziyang Ma, Changli Tang, Yujin Wang, **Zhisheng Zheng**
International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2023.
- PREPRINTS
- [15] VoiceCraft-X: Unifying Multilingual, Voice-Cloning Speech Synthesis and Speech Editing.
Zhisheng Zheng, Puyuan Peng, Anuj Diwan, Cong Phuoc Huynh, Xiaohang Sun, Zhu Liu, Vimal Bhat, David Harwath
Submitted to EMNLP 2025 (Overall Assessment: 4/4/4), Under Review.
 - [16] Scaling rich style-prompted text-to-speech datasets.
Anuj Diwan, **Zhisheng Zheng**, David Harwath, Eunsol Choi
Submitted to EMNLP 2025, Under Review.

SERVICES

Conference Reviewer/Program Committee:
ACL Rolling Review 2024, ICML 2025