Zhisheng Zheng

E-mail: zszheng@utexas.edu
Web: zhishengzheng.com

EDUCATION

The University of Texas at Austin, Austin, USA   August 2024 - Present
Ph.D. Student, Computer Science
Advisor: Prof. David Harwath

Shanghai Jiao Tong University, Shanghai, China   September 2020 - June 2024
B.E. in Information Engineering
Advisor: Prof. Xie Chen

Shanghai Jiao Tong University, Shanghai, China   September 2020 - June 2024
Zhiyuan Honors Program of Engineering

EMPLOYMENT

Microsoft Research Asia   Apr 2024 - Aug 2024, Beijing, China
Research Intern (Supervisor: Lei He, Xu Tan)

The University of Texas at Austin, SALT Lab   July 2023 - Jan 2024, Austin, USA
Visiting Scholar (Supervisor: David Harwath, Eunsol Choi)

Shanghai Jiao Tong University, X-Lance Lab   Dec 2021 - Jun 2024, Shanghai, China
Research Intern (Supervisor: Xie Chen)

SELECTED AWARDS & HONORS

Shanghai Outstanding Graduates   2024

**SenseTime Scholarship for Undergraduate AI Researchers
(30 winners nationwide each year)**, SenseTime   2023

Rongchang Science and Technology Innovation Scholarship (<0.1%)
Shanghai Rongchang Public Welfare Foundation   2023

Best Student Paper Shortlist, INTERSPEECH   2023

Zhiyuan College Honors Scholarship (TOP 10%), Zhiyuan College   2020-2024

Tencent Scholarship (TOP 2%), Tencent Technology (Shenzhen) Co., Ltd.   2021

PUBLICATIONS

[1] DRCap: Decoding CLAP Latents with Retrieval-Augmented Generation for Zero-shot Audio Captioning.
Xiquan Li, Wenxi Chen, Ziyang Ma, Xuenan Xu, Yuzhe Liang, **Zhisheng Zheng**, Qiuqiang Kong, Xie Chen
International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2025.

[2] SLAM-AAC: Enhancing Audio Captioning with Paraphrasing Augmentation and CLAP-Refine through LLMs.
Wenxi Chen, Ziyang Ma, Xiquan Li, Xuenan Xu, Yuzhe Liang, **Zhisheng Zheng**, Kai Yu, Xie Chen
International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2025.

BEFORE PH.D.

[3] EmoBox: Multilingual Multi-corpus Speech Emotion Recognition Toolkit and Benchmark.
Ziyang Ma, Mingjie Chen, Hezhao Zhang, **Zhisheng Zheng**, Wenxi Chen, Xiquan Li, Jiaxin Ye, Xie Chen, Thomas Hain
INTERSPEECH 2024.

[4] Emotion2vec: Self-Supervised Pre-Training for Speech Emotion Representation.
Ziyang Ma, **Zhisheng Zheng**, Jiaxin Ye, Jinchao Li, Zhifu Gao, Shiliang Zhang, Xie Chen
Findings of ACL, 2024.

[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.

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