

YOU (NEIL) ZHANG

<https://zyyouzhang.com>

☎ 585-732-2916 ✉ you.zhang@rochester.edu [in](#) [LinkedIn](#) [GitHub](#) [Google Scholar](#)

EDUCATION

University of Rochester <i>Ph.D., Electrical and Computer Engineering</i>	Aug 2019 – Dec 2024 (Expected) <i>Rochester, NY</i>
University of Rochester <i>M.S., Electrical and Computer Engineering</i>	Aug 2019 – May 2021 <i>Rochester, NY</i>
University of California, Berkeley <i>Undergraduate Exchange Studies, Electrical Engineering and Computer Science</i>	Jan 2018 – Jan 2019 <i>Berkeley, CA</i>
University of Electronic Science and Technology of China <i>B.Eng., Automation</i>	Sep 2015 – Jun 2019 <i>Chengdu, Sichuan, China</i>

RESEARCH INTERESTS

Speech & Audio Processing, Spatial Audio, Audio-Visual Analysis, Virtual and Augmented Reality, Deep Learning

EXPERIENCE

University of Rochester – Audio Information Research Lab <i>Research Assistant, Advisor: Prof. Zhiyao Duan</i>	Aug 2019 – Present <i>Rochester, NY</i>
<ul style="list-style-type: none">• HRTF Modeling for Spatial Audio in Virtual and Augmented Reality<ul style="list-style-type: none">* Proposed a deep learning system to predict the personalized head-related transfer functions (HRTF) employing anthropometric measurements and scanned head geometry of subjects.* Proposed neural field representations for unifying measured HRTFs across existing databses. We also proposed a generative model for such representation and applied it on HRTF interpolation and generative tasks.• Enhance the Robustness of Speaker Verification<ul style="list-style-type: none">* Improved the generalization ability to unseen spoofing attacks with proposed one-class learning.* Hypothesized and verified that channel effect is a primary reason for cross-dataset performance degradation. We proposed training strategies to improve the channel robustness for anti-spoofing.* Jointly optimized speaker verification and anti-spoofing with a proposed probabilistic framework.* Extended the one-class idea with speaker attractor multi-center one-class learning to maintain speaker diversity in real speech.• Emotional Talking Face Generation<ul style="list-style-type: none">* Implemented and evaluated the baseline method and took charge of the subjective evaluation section, including the Amazon Mechanical Turk (AMT) setup, survey design, and data analysis, and proved the proposed method exceeds the baseline.	
Tencent America – Tencent AI Lab <i>Research Intern, Mentor: Dr. Shi-Xiong Zhang</i>	May 2022 – Aug 2022 <i>Bellevue, WA</i>
<ul style="list-style-type: none">• Multi-Channel Audio-visual Speaker Diarization<ul style="list-style-type: none">* Proposed a probabilistic framework to incorporate the spatial information from multi-channel audio, speaker characteristics, and visual information to perform speaker diarization.	
Bytedance / TikTok – Speech, Audio & Music Intelligence <i>Research Intern, Mentor: Dr. Ming Tu</i>	May 2021 – Aug 2021 <i>Mountain View, CA</i>
<ul style="list-style-type: none">• Audio-visual Active Speaker Detection<ul style="list-style-type: none">* Implemented state-of-the-art active speaker detection methods and adapted them to real-world data on short-video platforms with a semi-supervised learning method, noisy student training.	
Tencent – Tencent Media Lab <i>Research Intern, Mentor: Dr. Yannan Wang</i>	Jun 2019 – Aug 2019 <i>Shenzhen, Guangdong, China</i>
<ul style="list-style-type: none">• Perceptual Loss Design for Mask-based Speech Enhancement<ul style="list-style-type: none">* Improved the perceptual quality of the enhanced speech using multi-task learning with implementation of several perception-inspired losses using uncertainty.	

PUBLICATIONS

- [12] **You Zhang**, Yuxiang Wang, and Zhiyao Duan. “HRTF Field: Unifying Measured HRTF Magnitude Representation with Neural Fields”, *arXiv preprint arXiv:2210.15196*, 2022. (submitted) [[link](#)][[code](#)]
- [11] Siwen Ding, **You Zhang**, and Zhiyao Duan. “SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing”, *arXiv preprint arXiv:2211.02718*, 2022. (submitted) [[link](#)][[code](#)]
- [10] Abudukelimu Wuerkaixi, Kunda Yan, **You Zhang**, Zhiyao Duan, and Changshui Zhang. “DyViSE: Dynamic Vision-Guided Speaker Embedding for Audio-Visual Speaker Diarization”, in *Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2022. (accepted) [[link](#)][[code](#)]
- [9] Yuxiang Wang, **You Zhang**, Zhiyao Duan, and Mark Bocko. “Predicting Global Head-Related Transfer Functions From Scanned Head Geometry Using Deep Learning and Compact Representations”, *IEEE/ACM Transactions on Audio Speech and Language Processing*, 2022. (submitted) [[link](#)][[code](#)]
- [8] Abudukelimu Wuerkaixi, **You Zhang**, Zhiyao Duan, and Changshui Zhang. “Rethinking Audio-visual Synchronization for Active Speaker Detection”, in *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2022. [[link](#)]
- [7] **You Zhang**, Ge Zhu, and Zhiyao Duan. “A Probabilistic Fusion Framework for Spoofing Aware Speaker Verification”, in *Proc. The Speaker and Language Recognition Workshop (Odyssey)*, pp. 77-84, 2022. [[link](#)][[code](#)]
- [6] **You Zhang**, Fei Jiang, Ge Zhu, Xinhui Chen, and Zhiyao Duan. “Generalizing Voice Presentation Attack Detection to Unseen Synthetic Attacks and Channel Variation”, *Handbook of Biometric Anti-spoofing*, Springer, 2022. (to be published) [[code](#)]
- [5] Sefik Emre Eskimez, **You Zhang**, and Zhiyao Duan. “Speech Driven Talking Face Generation from a Single Image and an Emotion Condition”, *IEEE Transactions on Multimedia*, vol. 24, pp. 3480-3490, 2021. [[link](#)][[project webpage](#)][[code](#)]
- [4] Xinhui Chen*, **You Zhang***, Ge Zhu*, and Zhiyao Duan. “UR Channel-Robust Synthetic Speech Detection System for ASVspoof 2021”, in *Proc. ASVspoof 2021 Workshop*, pp. 75-82, 2021. (* equal contribution) [[link](#)][[code](#)][[video](#)]
- [3] **You Zhang**, Ge Zhu, Fei Jiang, and Zhiyao Duan. “An Empirical Study on Channel Effects for Synthetic Voice Spoofing Countermeasure Systems”, in *Proc. Interspeech*, pp. 4309-4313, 2021. [[link](#)][[code](#)][[video](#)]
- [2] **You Zhang**, Fei Jiang, and Zhiyao Duan. “One-class Learning Towards Synthetic Voice Spoofing Detection”, *IEEE Signal Processing Letters*, vol. 28, pp. 937-941, 2021. [[link](#)][[code](#)][[video](#)][[project webpage](#)]
- [1] Yuxiang Wang, **You Zhang**, Zhiyao Duan, and Mark Bocko. “Global HRTF Personalization Using Anthropometric Measures”, in *Audio Engineering Society (AES) 150th Convention*, 2021. [[link](#)][[code](#)][[video](#)]

SKILLS

Programming: Python (PyTorch, Numpy, Pandas), Java, MATLAB, R, VHDL, C, L^AT_EX, Markdown

Platforms: Linux, Git, Jupyter Notebook, PyCharm, IntelliJ, Xilinx Vivado, Multisim

TEACHING

Teaching Assistant

• ECE 440	Introduction to Random Processes	Fall 2022
• ECE 208 / 408	The Art of Machine Learning	Spring 2022
• ECE 272 / 472	Audio Signal Processing	Spring 2020 & Spring 2021
• ECE 477	Computer Audition	Fall 2020
• ECE 216	Microprocessor & Data Conversion	Fall 2019

Students Mentored / Mentoring

• Yongyi Zang	AME undergraduate @ UR	Summer 2022 - Present
• Siwen Ding	DS master @ Columbia University	Summer 2022 - Fall 2022
• Abudukelimu Wuerkaixi	PhD student @ Tsinghua University	Fall 2021 - Summer 2022
• Xinhui Chen	CS master @ UR	Spring 2021 - Summer 2021

SERVICE & AWARD

Reviewer

- Audio Engineering Society (AES) 152nd, 153rd Convention
- IEEE Transactions on Computational Imaging (TCI)
- International Journal of Electrical and Computer Engineering Systems (IJECES)

Co-chaired

- Western New York Virtual and Augmented Reality Mini-Conference 2022 [[link](#)]

Awarded

• Travel Grant from AS&E Graduate Student Association	Fall 2021 & Summer 2022
• Travel Grant from NSF-NRT AR/VR Training Program	Spring 2022
• Outstanding Fresh Graduate @ UESTC	Spring 2019
• Renmin Scholarship	Fall 2016 & Fall 2017 & Fall 2018