

# YOU (NEIL) ZHANG

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

## EDUCATION

### University of Rochester

*Ph.D., Electrical and Computer Engineering*

**Aug 2019 – Dec 2024 (Expected)**

*Rochester, NY*

### University of Rochester

*M.S., Electrical and Computer Engineering*

**Aug 2019 – May 2021**

*Rochester, NY*

### University of California, Berkeley

*Undergraduate Exchange Studies, Electrical Engineering and Computer Science*

**Jan 2018 – Jan 2019**

*Berkeley, CA*

### University of Electronic Science and Technology of China

*B.Eng., Automation*

**Sep 2015 – Jun 2019**

*Chengdu, Sichuan, China*

## RESEARCH INTERESTS

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

## EXPERIENCE

### University of Rochester – Audio Information Research Lab

*Research Assistant, Advisor: Prof. Zhiyao Duan*

**Aug 2019 – Present**

*Rochester, NY*

- **HRTF Modeling for Spatial Audio in Virtual and Augmented Reality**

- \* 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 databases. We also proposed a **generative model** for such representation and applied it on HRTF interpolation and generative tasks.

- **Enhance the Robustness of Speaker Verification**

- \* 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**

- \* 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

*Research Intern, Mentor: Dr. Shi-Xiong Zhang*

**May 2022 – Aug 2022**

*Bellevue, WA*

- **Multi-Channel Audio-visual Speaker Diarization**

- \* 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

*Research Intern, Mentor: Dr. Ming Tu*

**May 2021 – Aug 2021**

*Mountain View, CA*

- **Audio-visual Active Speaker Detection**

- \* 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

*Research Intern, Mentor: Dr. Yannan Wang*

**Jun 2019 – Aug 2019**

*Shenzhen, Guangdong, China*

- **Perceptual Loss Design for Mask-based Speech Enhancement**

- \* 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](#)]
- [11] Siwen Ding, **You Zhang**, and Zhiyao Duan. “SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing”, 2022. (submitted)
- [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](#)]
- [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<sup>A</sup>T<sub>E</sub>X, 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