You (Neil) Zhang

https://yzyouzhang.com

585-732-2916 you.zhang@rochester.edu LinkedIn GitHub Google Scholar

Research Interests: Machine Learning in Speech, Acoustics, and Audio Signal Processing

* **Security and Privacy in Speech & Audio**: Speech Anti-Spoofing, Singing Voice Deepfake Detection, Audio Watermarking

* **Spatial Audio for Virtual and Augmented Reality**: Head-Related Transfer Function (HRTF) Personalization

* **Multi-Modal Learning**: Audio-Visual Speech Analysis, Emotional Talking Face Generation, Emotional Text-to-Speech

Education

University of Rochester (UR) — **Aug 2019 – May 2025 (Expected)**
Ph.D., Electrical and Computer Engineering — *Rochester, NY*

University of Rochester (UR) — **Aug 2019 – May 2021**
M.S., Electrical and Computer Engineering — *Rochester, NY*

University of California, Berkeley (UCB) — **Jan 2018 – Jan 2019**
Undergraduate Exchange Studies, Electrical Engineering and Computer Science — *Berkeley, CA*

University of Electronic Science and Technology of China (UESTC) — **Sep 2015 – Jun 2019**
B.Eng., Automation — *Chengdu, Sichuan, China*

Honors & Awards

IEEE Signal Processing Society (SPS) Scholarship (*one of the 45 international recipients in 2024*) [link]	Fall 2024
National Institute of Justice (NIJ) Graduate Research Fellowship (*one of the 24 national awardees in 2023*) [link]	Fall 2023
Top 3% of all papers accepted at ICASSP 2023 (*one of the 75 awarded papers out of 2722 accepted papers*) [link]	Summer 2023
ICASSP Rising Stars in Signal Processing (*one of the 24 international awardees in 2023*) [link]	Summer 2023
Signal Processing at the ASA Student Paper Award - Second Place (*$200*)	Spring 2023
Travel Grant from AS&E Graduate Student Association @ UR (*$500 each*)	Fall 2021 & Summer 2022
Travel Grant from NSF-NRT AR/VR Training Program (*$1000*)	Spring 2022
Outstanding Graduate @ UESTC (*top 1% in the same year of graduation*)	Spring 2019
Renmin Scholarship (*top 3% in the same grade and major*)	Fall 2016 & Fall 2017 & Fall 2018

Academic & Industrial Research Experience

University of Rochester – *Audio Information Research Lab* — **Aug 2019 – Present**
PhD Candidate, Committee: Prof. Zhiyao Duan (Advisor), Prof. Mujdat Cetin, Prof. Jiebo Luo — *Rochester, NY*

- **Audio Deepfake Detection / Speaker Verification Anti-Spoofing**
  - Generalization Ability: Developed **one-class learning** for better detecting unseen spoofing attacks; Extended the one-class learning idea with speaker attractor multi-center one-class learning to maintain speaker diversity in real speech
  - Channel Robustness: Established that **channel effect** is a primary reason for cross-dataset performance degradation, and developed training strategies to improve the channel robustness for anti-spoofing
  - Joint Optimization: Developed a **probabilistic fusion framework** for spoofing aware speaker verification
  - Singing Voice Deepfake Detection (SVDD): Proposed novel SVDD task and identified challenges with the collected SingFake dataset; Organized SVDD 2024 Challenge at IEEE SLT2024 and MIREX@ISMIR2024
  - Algorithm Deployment: Impact real-world by working with IngenID to deploy the developed anti-spoofing algorithms
- **Personalized Head-Related Transfer Function (HRTF) Modeling for Spatial Audio**
  - Established learning **neural field representations** to unify measured HRTFs databases for upsampling and personalization
  - Developed a novel **position-dependent normalization strategy** that effectively mitigates the influence of cross-database differences to improve the learned representation further

  - Built a deep learning system to predict the **spherical harmonic coefficients** from anthropometric measurements and scanned head geometry of subjects for HRTF personalization

- **Audio-Visual Rendering and Analysis**
  - Emotional Talking Face Generation: Implemented and evaluated a baseline method and took charge of the subjective evaluation section, including the Amazon Mechanical Turk (AMT) setup, survey design, and data analysis
  - Audio-Visual Speaker Diarization: Alleviated audio-visual **synchronization** and **off-screen speakers** problem for audio-visual speaker diarization
  - Audio-Visual Deepfake Detection: Developed a multi-stream fusion framework complemented with **one-class learning** to improve the generalization ability for audio-visual deepfake detection

Meta – *Reality Labs Research* **May 2024 – Sep 2024**
Research Intern, Mentor: Dr. Ishwarya Ananthabhotla *Redmond, WA*

- **Perceptual Head-Related Transfer Function (HRTF) Representation Learning**
  - Improved HRTF representation learning by incorporating **perceptually-informed loss functions** and analyzed perceptual features in machine learning models through alignment with **auditory model observers**.

Microsoft – *Applied Sciences* **May 2023 – Aug 2023**
Research Intern, Mentor: Dr. Kazuhito Koishida *Redmond, WA*

- **Audio-Visual Segmentation by Prompting Segment Anything Model**
  - Proposed a **prompting framework** to augment a vision foundation model (Segment Anything Model, SAM) with auditory understanding capabilities, enabling it to **simultaneously localize and segment** sounding sources in video frames.

Tencent America – *Tencent AI Lab* **May 2022 – Aug 2022**
Research Intern, Mentor: Dr. Shi-Xiong (Austin) Zhang *Bellevue, WA*

- **Multi-Channel Audio-Visual Speaker Diarization with Spatial Features**
  - 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* **May 2021 – Aug 2021**
Research Intern, Mentor: Dr. Ming Tu *Mountain View, CA*

- **Audio-Visual Active Speaker Detection with Noisy Student Training**
  - 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* **Jun 2019 – Aug 2019**
Research Intern, Mentor: Dr. Yannan Wang *Shenzhen, Guangdong, China*

- **Perceptual Loss Design for Mask-based Speech Enhancement**
  - Improved the perceptual quality of the enhanced speech using **multi-task learning** with the implementation of several perception-inspired losses using **uncertainty**.

Grant Writing Experiences

Audio Deepfake Detection for Forensics and Security (Awarded Fellow: You Zhang) **Jan 2024 – Apr 2025**
National Institute of Justice (NIJ) Graduate Research Fellowship *$64,003*

Safeguarding Generative AI by Audio-Visual Deepfake Detection (PI: Zhiyao Duan) **Jun 2024 – Dec 2024**
National AI Research Resource (NAIRR) Pilot *10,000 GPU hours*

Training Audio-Visual Foundation Models to Capture Fine-Grained Dependencies (PI: Zhiyao Duan) **Nov 2023 – Jun 2024**
Microsoft Accelerate Foundation Models Research (AFMR) initiative *20,000 Azure credits*

Developing and Deploying Spoofing Aware Speaker Verification Systems (PI: Zhiyao Duan) **Jan 2023 – Dec 2023**
New York State Center of Excellence in Data Science *$59,989*

Personalized Immersive Spatial Audio with Neural Field (PIs: Zhiyao Duan and Mark Bocko) **Nov 2022 – Oct 2023**
University of Rochester Goergen Institute for Data Science seed funding program *$20,000*

Media Coverage

Why AI-generated audio is so hard to detect [link] — NBC News
News10NBC Investigates: Here's what happened when we did a deep fake on Berkeley Brean's voice [link] — WHEC-TV Rochester
Audio deepfake detective developing new sleuthing techniques [link] — UR News Center

Invited Talks

[5] Generalizing Audio Deepfake Detection
Carnegie Mellon University (CMU) Speech Lunch, USA – Online — Nov 2024

[4] Audio Deepfake Detection [video]
Generative AI Spring School & Global AI Bootcamp, Ukraine – Online — Mar 2024

[3] Improving Generalization Ability for Audio Deepfake Detection
Learning And Mining from DatA (LAMDA) Lab, Nanjing University, China — Dec 2023

[2] Generalizing Voice Presentation Attack Detection to Unseen Synthetic Attacks
ISCA Special Interest Group (SIG) - Security and Privacy in Speech Communication (SPSC) webinar – Online — Feb 2023

[1] One-class Learning Towards Synthetic Voice Spoofing Detection
National Institute of Informatics (NII) Yamagishi Lab, Japan – Online — Jan 2021

Tutorials

[3] Tutorial on Machine Learning for Acoustics (Co-organized with Ryan McCarthy, Samuel A. Verburg, Peter Gerstoft)
Acoustical Society of America (ASA) 187th Meeting, Online — Nov 2024

[2] Multimedia Deepfake Detection (Co-organized with Menglu Li, Luchuan Song, Xiao-Ping Zhang, Chenliang Xu, Zhiyao Duan)
IEEE International Conference on Multimedia and Expo (ICME), Niagra Falls, Canada — July 2024

[1] Machine Learning for Personalized Head-Related Transfer Functions (HRTFs) Modeling in Gaming [slides]
AES 6th International Conference on Audio for Games, Tokyo, Japan — Apr 2024

Teaching

Teaching Assistant

• ECE 411	Selected Topics in Augmented and Virtual Reality	Spring 2024
• ECE 277 / 477	Computer Audition	Fall 2020 & Fall 2023
• ECE 208 / 408	The Art of Machine Learning	Spring 2022 & Spring 2023
• ECE 440	Introduction to Random Processes	Fall 2022
• ECE 272 / 472	Audio Signal Processing	Spring 2020 & Spring 2021
• ECE 216	Microprocessor & Data Conversion	Fall 2019

Guest Lectures

• ECE 277 / 477	Room Acoustics and Spatial Audio	Fall 2024
• ECE 411	Audio Deepfake Detection and Watermarking	Spring 2024
• ECE 277 / 477	Python Programming for Audio	Fall 2023
• ECE 277 / 477	Speech Technology	Fall 2023
• ECE 277 / 477	Speech Anti-Spoofing	Fall 2023
• ECE 208 / 408	Support Vector Machines (SVM)	Spring 2023
• ECE 208 / 408	Neural Network Training	Spring 2023
• ECE 208 / 408	Generative Adversarial Networks (GAN)	Spring 2022
• ECE 277 / 477	Introduction to Speech Technology	Fall 2020

Students Mentored

• Ye In (Brynn) Lee	DS master @ UR	Audio Deepfake Detection
• Kyungbok Lee	CS undergraduate @ UR	Audio-Visual Deepfake Detection
• Yutong Wen	AME undergraduate @ UR	HRTF Personalization with Neural Fields
• Enting Zhou	CS undergraduate @ UR	Speech Emotion Representation Learning
• Yongyi Zang	AME undergraduate @ UR	Audio Deepfake Detection
• Siwen (Sivan) Ding	DS master @ Columbia University	Audio Deepfake Detection
• Abudukelimu Wuerkaixi	PhD student @ Tsinghua University	Audio-Visual Speaker Diarization
• Xinhui Chen	CS master @ UR	Audio Deepfake Detection

PUBLICATIONS (* Equal contribution (EC), ‡ Student mentored) Google Scholar: nYtHcRAAAAAJ

Under Review / Preprint

[U3] Kun Zhou, **You Zhang**, Shengkui Zhao, Hao Wang, Zexu Pan, Dianwen Ng, Chong Zhang, Chongjia Ni, Yukun Ma, Trung Hieu Nguyen, Jia Qi Yip, and Bin Ma. "Emotional Dimension Control in Language Model-Based Text-to-Speech: Spanning a Broad Spectrum of Human Emotions", *arXiv preprint 2409.16681*, 2024. [arXiv] [demo]

[U2] 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", *arXiv preprint 2207.14352*, 2024. [arXiv] [code]

[U1] **You Zhang***, Ge Zhu*, Julia M. Soto*, Samantha E. Lettenberger*, Maryam Zafar, Peggy Auinger, Abigail Arky, Emma Waddell, Kelsey Spear, Rajbir Toor, Grace Nkrumah, Emily A. Hartman, Jacob Epifano, Michael J. Hasselberg, Anton P. Porsteinsson, Rich Christie, Zhiyao Duan, Aaron J. Masino, and E. Ray Dorsey, "Words Spoken Daily among Individuals with Neurodegenerative Conditions: A Pilot Study", 2023.

Book Chapters

[B1] **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 (3rd ed.)*, Springer, 2023. [DOI] [code]

Journals

[J2] 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. [DOI] [code] [project webpage]

[J1] **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. [DOI] [code] [video] [project webpage]

Peer-Reviewed Conferences and Workshops

[P15] **You Zhang**, Yongyi Zang, Jiatong Shi, Ryuichi Yamamoto, Tomoki Toda, and Zhiyao Duan. "SVDD 2024: The Inaugural Singing Voice Deepfake Detection Challenge", *Proc. IEEE Spoken Language Technology Workshop (SLT)*, 2024. [link] [code]

[P14] Kyungbok Lee‡, **You Zhang**, and Zhiyao Duan, "A Multi-Stream Fusion Approach with One-Class Learning for Audio-Visual Deepfake Detection", *Proc. IEEE 26th International Workshop on Multimedia Signal Processing (MMSP)*, 2024. [DOI] [code] (Kyungbok received MMSP 2024 travel grant and UR Undergraduate Research Presentation Award for this paper.)

[P13] Yongyi Zang‡, Jiatong Shi, **You Zhang**, Ryuichi Yamamoto, Jionghao Han, Yuxun Tang, Shengyuan Xu, Wenxiao Zhao, Jing Guo, Tomoki Toda, and Zhiyao Duan. "CtrSVDD: A Benchmark Dataset and Baseline Analysis for Controlled Singing Voice Deepfake Detection", *Proc. Interspeech*, 2024. [DOI] [code] [dataset]

[P12] Yongyi Zang*‡, **You Zhang* (EC)**, Mojtaba Heydari, and Zhiyao Duan. "SingFake: Singing Voice Deepfake Detection", in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024. [DOI] [code] [project webpage]

[P11] Enting Zhou‡, **You Zhang**, and Zhiyao Duan. "Learning Arousal-Valence Representation from Categorical Emotion Labels of Speech", in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024. [DOI] [code]

[P10] Yutong Wen‡, **You Zhang**, and Zhiyao Duan. "Mitigating Cross-Database Differences for Learning Unified HRTF Representation", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2023. [DOI] [code] [video] (Yutong received WASPAA 2023 travel grant and UR Undergraduate Research Presentation Award for this paper.)

[P9] Yongyi Zang‡, **You Zhang**, and Zhiyao Duan. "Phase Perturbation Improves Channel Robustness for Speech Spoofing Countermeasures", in *Proc. Interspeech*, pp. 3162-3166, 2023. [DOI] [code]

[P8] Siwen Ding‡, **You Zhang**, and Zhiyao Duan. "SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing", in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023. [DOI][code][video]

[P7] **You Zhang**, Yuxiang Wang, and Zhiyao Duan. "HRTF Field: Unifying Measured HRTF Magnitude Representation with Neural Fields", in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023. [DOI] [code] [video] (**Recognized as one of the top 3% of all papers accepted at ICASSP 2023**)

[P6] 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 24th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1-6, 2022. [DOI] [code]

[P5] Abudukelimu Wuerkaixi‡, **You Zhang**, Zhiyao Duan, and Changshui Zhang. "Rethinking Audio-visual Synchronization for Active Speaker Detection", in *Proc. IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP)*, 2022. [DOI][code]

[P4] **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. [DOI] [code]

[P3] Xinhui Chen*‡, **You Zhang* (EC)**, Ge Zhu*, and Zhiyao Duan. "UR Channel-Robust Synthetic Speech Detection System for ASVspoof 2021", in *Proc. ASVspoof 2021 Workshop*, pp. 75-82, 2021. [DOI] [code] [video]

[P2] **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. [DOI] [code] [dataset] [video]

[P1] Yuxiang Wang, **You Zhang**, Zhiyao Duan, and Mark Bocko. "Global HRTF Personalization Using Anthropometric Measures", in *Proc. Audio Engineering Society (AES) 150th Convention*, 2021. [DOI] [code] [video]

Technical Reports

[T2] **You Zhang**, Yongyi Zang, Jiatong Shi, Ryuichi Yamamoto, Jionghao Han, Yuxun Tang, Tomoki Toda, and Zhiyao Duan. "SVDD Challenge 2024: A Singing Voice Deepfake Detection Challenge Evaluation Plan", 2024. [link] [challenge webpage]

[T1] **You Zhang**, Ge Zhu, and Zhiyao Duan. "UR Spoofing Aware Speaker Verification System for the SASV Challenge", 2022. [link]

Conference Abstracts

[C3] **You Zhang**, Yuxiang Wang, Mark Bocko, and Zhiyao Duan. "Grid-Agnostic Personalized Head-Related Transfer Function Modeling with Neural Fields", in *Acoustical Society of America (ASA) 184th Meeting*, 2023. [DOI] (**Recognized by Signal Processing at the ASA Student Paper Award - Second Place**)

[C2] Samantha E. Lettenberger, Maryam Zafar, Julia M. Soto, **You Zhang**, Ge Zhu, Aaron J. Masino, Grace Nkrumah, Emma Waddell, Kelsey Spear, Abigail Arky, Rajbir Toor, Emily Hartman, Jacob Epifano, Rich Christie, Zhiyao Duan, and Ray Dorsey. "Words Spoken Daily: A Novel Measure of Cognition", in *International Congress of Parkinson's Disease and Movement Disorders (MDS)*, 2023. [DOI]

[C1] Yuxiang Wang, **You Zhang**, Zhiyao Duan, and Mark Bocko. "Employing Deep Learning Method to Predict Global Head-Related Transfer Functions from Scanned Head Geometry", in *Acoustical Society of America (ASA) 181st Meeting*, 2021. [DOI]

Professional Services

Leadership

- IEEE ICASSP 2024 Student Volunteer — Spring 2024
- Executive Committee Member for AR/VR PhD training program — Fall 2023 – Spring 2024
- Western New York AR/VR Mini-Conference Co-chair [link] — Spring 2022 & Spring 2023
- Diversity, Equity, and Inclusion (DEI) Committee Member for ECE Department — Fall 2022 – Spring 2023
- IEEEXtreme 16.0 Ambassador [link] — Fall 2022

Reviewer

- **Journals**:
  - IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP) — 2023-2024
  - IEEE Journal of Selected Topics in Signal Processing (JSTSP) — 2024
  - IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) — 2024
  - IEEE Signal Processing Letters — 2024
  - IEEE Open Journal of Signal Processing (OJSP) — 2022-2024
  - ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) — 2024
  - EURASIP Journal on Audio, Speech, and Music Processing — 2024
  - EURASIP Journal on Advances in Signal Processing — 2023
  - Computer Speech & Language — 2024
  - Transactions of the International Society for Music Information Retrieval (TISMIR) — 2022-2023
  - ACM Computing Surveys — 2024
  - Neural Networks — 2023-2024
  - IEEE Access — 2023-2024
  - IEEE Transactions on Computational Imaging (TCI) — 2021
- **Peer-Reviewed Conferences and Workshops**:
  - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) — 2024-2025
  - ISCA Interspeech — 2023-2024

- ∗ IEEE Spoken Language Technology Workshop 2024
- ∗ Audio Engineering Society (AES) 152nd - 157th Convention 2022-2024
- ∗ ISCA Automatic Speaker Verification and Spoofing Countermeasures (ASVspoof) Workshop 2024
- ∗ ISCA 4th Symposium on Security and Privacy in Speech Communication 2024
- ∗ CVPR Multimodal Learning and Applications Workshop (MULA) 2023-2024
- ∗ IJCAI Workshop on Deepfake Audio Detection and Analysis (DADA) 2023

Membership

- Acoustical Society of America (ASA) Student Member 2023-2025
- IEEE Graduate Student Member 2021-2025
- Audio Engineering Society (AES) Student Member 2019-2025

Skills

Programming: Python (PyTorch, Numpy, Pandas), MATLAB, Java, R, VHDL, C, LaTeX, Markdown

Platforms: Linux, Git, Jupyter Notebook, Slurm, Visual Studio Code, PyCharm, IntelliJ, Xilinx Vivado, Multisim

Languages: English (Fluent), Mandarin Chinese (Native)

Miscellaneous

Activities: Half Marathon Finisher [Certificate]

Hobbies: Stand-up Paddleboarding, Traveling, Swimming, Badminton