YAOYUN ZHANG

X-LANCE Lab, SJTU ◇ yaoyunzhang897@gmail.com ◇ Homepage ◇ Google Scholar

Research Interests

My research interests mainly lie in **universal audio intelligence**, with an emphasis on multimodal audio generation and unified audio representation learning. I am particularly interested in developing unified audio codecs and generative models that support speech, music, and general sounds while **remaining efficient, robust, and generalizable across modalities and domains**.

Education

School of Computer Science, Shanghai Jiao Tong University *Shanghai, China*
· M.Eng. in Computer Technology, Advised by Prof. Mengyue Wu Sep 2023 – Mar 2026

Chengdu University of Technology *Chengdu, China*
◇ Dual Bachelor's Degrees Sep 2019 – Jun 2023
· B.Eng. in Computer Science and Technology, School of Computer Science
· B.A. in English, School of Foreign Languages and Cultures

Selected Publications

Smooth-Foley: Creating Continuous Sound for Video-to-Audio Generation Under Semantic Guidance
Yaoyun Zhang, Xuenan Xu, Mengyue Wu*.
ICASSP 2025 | Page | Paper

FakeSound2: A Benchmark for Explainable and Generalizable Deepfake Sound Detection
Zeyu Xie, **Yaoyun Zhang**, Xuenan Xu, Yongkang Yin, Chenxing Li, Mengyue Wu*, Yuexian Zou*.
INTERSPEECH 2026 | Page | Paper

UniFlow-Audio: A Unified Flow Matching Framework for Audio Generation from Omni-Modalities
Xuenan Xu, Jiahao Mei, Zihao Zheng, Ye Tao, Zeyu Xie, **Yaoyun Zhang**, Haohe Liu, Yuning Wu, Ming Yan, Wen Wu, Chao Zhang, Mengyue Wu*.
NeurIPS 2026 (under review) | Page | Paper

Domain Invariant Learning for Gaussian Processes and Bayesian Exploration
Zhaoxi Long, Siyuan Bian, **Yaoyun Zhang**, Yuliang Zhang, Qinying Gu, Xinbing Wang, Chenghu Zhou, Nanyang Ye*.
AAAI 2024 | Page | Paper

Research Experience

Meituan, Beijing Sankuai Online Technology Co., Ltd. *Beijing, China*
Research Intern — Audio Generation and Neural Codec Jun 2025 – Sep 2025

- **Developed a unified neural audio tokenizer/codec pipeline** based on BEST-RQ-style masked representation learning, followed by learnable vector quantization and flow-matching-based reconstruction for speech, music, and general audio.
- **Designed and evaluated a three-stage training curriculum over nearly one million hours of audio**; systematically investigated intermediate-layer VQ, multi-codebook/RVQ configurations, and codebook-utilization strategies, and proposed local feature fusion for low-bitrate representations.
- **Conducted large-scale ablations on multimodal data mixtures and codec configurations**, identifying a 1:10:10 speech/music/general-audio curriculum and achieving consistent improvements in PESQ/STOI and Mel/STFT reconstruction metrics.

Research Experience (continued)

X-LANCE Lab, Shanghai Jiao Tong University *Shanghai, China*
Graduate Researcher Sep 2023 – Mar 2026
Research Collaborator Apr 2026 – Present
Advisor: Prof. Mengyue Wu

- **Led Smooth-Foley (ICASSP 2025) end-to-end as first author**, independently handling problem formulation, data construction, model development, experiments, and paper writing, and developed frame-level visual fusion and semantic temporal conditioning for continuous video-to-audio generation.
- **Contributed to the early development of UniFlow-Audio**, by validating backbone architectures and flow-matching/diffusion implementations across heterogeneous tasks, including text-to-audio, audio super-resolution, and video-to-audio generation.
- **Led the video-to-audio branch of UniFlow-Audio**, including video data preprocessing, multimodal conditioning and training pipeline implementation, model training, and evaluation.
- **Contributed to FakeSound2 (INTERSPEECH 2026)**, including the large-scale data construction pipeline, manipulation/source taxonomy, and out-of-domain evaluation protocol for generalizable deepfake sound detection.

Honors & Awards

National Encouragement Scholarship (Top 3% nationwide) *2020*
Second Prize, National Computer Game Competition *2021*

Skills

Skills: Python, PyTorch, Linux, LaTeX
Languages: English (CATTI International Certificate C-Level), Chinese (Native)