LEYING ZHANG

✉zhangleying@sjtu.edu.cn ✆(+86) 18621098717
3-520 SEIEE Building, Shanghai Jiao Tong University
800 Dongchuan Road, Minhang District, Shanghai, China 200240

RESEARCH INTERESTS

Text-to-speech, Multi-modality, Audio generation, Speaker Verification

EDUCATION

Shanghai Jiao Tong University 09/01/2023 - Present
PhD, Computer Science and Engineering Supervisor: Prof. Yanmin Qian

Shanghai Jiao Tong University 09/01/2021 - 06/30/2023
Master, Electronic Information Supervisor: Prof. Yanmin Qian

Télécom Paris (Institut polytechnique de Paris) 09/03/2021 - 02/11/2022
Exchange Student, Data science and Image processing

Shanghai Jiao Tong University 09/01/2017 - 06/30/2021
Bachelor of Information Engineering and French (double degree)

SELECTED PUBLICATIONS

[C1] **Leying Zhang**, Yao Qian, Xiaofei Wang, Manthan Thakker, Dongmei Wang, Jianwei Yu, Haibin Wu, Yuxuan Hu, Jinyu Li, Yanmin Qian, Sheng Zhao. “CoVoMix2: Advancing Zero-Shot Dialogue Generation with Fully Non-Autoregressive Flow Matching”. **NeurIPS**, *Dec. 2025*

[C2] **Leying Zhang**, Wangyou Zhang, Zhengyang Chen and Yanmin Qian. “Advanced Zero-Shot Text-to-Speech for Background Removal and Preservation with Controllable Masked Speech Prediction”. **ICASSP**, *April. 2025*

[C3] **Leying Zhang**, Yao Qian, Long Zhou, Shujie Liu, Dongmei Wang, Xiaofei Wang, Midia Yousefi, Yanmin Qian, Jinyu Li, Lei He, Sheng Zhao, Michael Zeng. “CoVoMix: Advancing Zero-Shot Speech Generation for Human-like Multi-talker Conversations”. **NeurIPS**,*Dec. 2024*

[C4] **Leying Zhang**, Yao Qian, Linfeng Yu, Heming Wang, Hemin Yang, Shujie Liu, Long Zhou, Yanmin Qian. “ DDTSE: Discriminative Diffusion Model for Target Speech Extraction”. **SLT**, *Dec. 2024*

[C5] **Leying Zhang**, Zhengyang Chen and Yanmin Qian. “Adaptive Large Margin Fine-tuning for Speaker Verification”. **ICASSP**, *June. 2023*

[C6] **Leying Zhang***, Zhengyang Chen* and Yanmin Qian. “Enroll-Aware Attentive Statistics Pooling for Target Speaker Verification ”. **InterSpeech**, *Sep. 2022*

[C7] **Leying Zhang**, Zhengyang Chen and Yanmin Qian. “Knowledge Distillation from Multi-Modality to Single-Modality for Person Verification”. **InterSpeech**, *Sep. 2021*

[C8] Yihan Liu, Zhengyang Chen, **Leying Zhang**, Yanmin Qian. “E2E-BPVC: End-to-End Background-Preserving Voice Conversion via In-Context Learning“. **InterSpeech**, *Aug. 2025*

[C9] Tingxiao Zhou; **Leying Zhang**; Yanmin Qian. “Knowledge Distillation from Discriminative Model to Generative Model with Parallel Architecture for Speech Enhancement“. **ISCSLP**, *Aug. 2024*

[C9] Haitian Lu, Gaofeng Cheng, Liuping Luo, **Leying Zhang**, Yanmin Qian, Pengyuan Zhang. “Slide: Integrating speech language model with llm for spontaneous spoken dialogue generation“. **ICASSP**, *April. 2025*

[C10] Linfeng Yu, Wangyou Zhang, Chenpeng Du, **Leying Zhang**, Zheng Liang, Yanmin Qian. “Generation-Based Target Speech Extraction with Speech Discretization and Vocoder”. **ICASSP**, *April. 2024*

[C11] Yichong Leng, Zhifang Guo, Kai Shen, Xu Tan, Zeqian Ju, Yanqing Liu, Yufei Liu, Dongchao Yang, **Leying Zhang**, Kaitao Song, Lei He, Xiang-Yang Li, Sheng Zhao, Tao Qin, Jiang Bian. “Prompttts 2: Describing and generating voices with text prompt”. **ICLR**, *May. 2024*

INDUSTRY EXPERIENCE

Research Intern Meta Superintelligence Labs
Supervised by Bowen Shi *New York, USA. 10/13/2025 - present*

Audio Evaluation: Design and implemented an efficient audio evaluation model.

Research Intern Microsoft Core AI
Supervised by Yao Qian *Remote. 10/29/2024 - 08/15/2025*

Text-to-Dialogue Generation: Design and implemented a purely non-autoregressive dialogue generation framework that support zero-shot multi-speaker, multi-turn and fine-grained temporal control. This system has been incorporated into the Azure TTS product.

Research Intern Microsoft Azure Research
Supervised by Yao Qian *Remote. 03/31/2023 - 03/29/2024*

Target speech extraction: Investigated diffusion-based model for target speech extraction. Proposed an efficient approach by combining diffusion and discriminative methods for handling multi- and single-speaker scenarios in both noisy and clean conditions.
Text-to-Dialogue Generation: Investigated Conversational Voice Mixture Generation, a novel model for zero-shot, human-like, multi-speaker, multi-round dialogue speech generation

Research Intern Microsoft Research Asia
Supervised by Xu Tan *Beijing, China. 11/01/2022 - 03/30/2023*

Audio generation: Implemented vector-quantized diffusion model with classifier-free guidance. Achieved 10% improvement over baseline. Investigated latent diffusion model’s effects by fine-tuning Stable diffusion.
Text-to-speech: Utilized vector-quantized diffusion model for text-to-speech on large-scale dataset with different neural audio codecs. Generated high-quality speech and get improvements on zero-shot text-to-speech.

TEACHING EXPERIENCE

Teaching Assistant - Intelligent Speech Techonolgy	Spring, 2025
Teaching Assistant - Machine Learning	Fall, 2022

HONORS AND AWARDS

National Scholarship	2022
ICASSP 2025 Travel Grant	2025
NeurIPS 2024 Scholar Award	2024
First place in CN-Celeb Speaker Recognition Challenge 2022	2022
ISCA and Interspeech Travel Grant	2021
Outstanding Graduates of Shanghai	2021
Outstanding student leader of SJTU	2021
Guanghua Scholarship	2020
SJTU Class B Scholarship	2019