Junan Zhang

junanzhang@link.cuhk.edu.cn
Homepage
Netease Artist Page

Education

•Fudan University *September 2019 - June 2024*
Bachelor in Software Engineering. Advisor: Prof. Bihuan Chen Shanghai, China

•The Chinese University of Hong Kong, Shenzhen *September 2024 - Present*
PhD in Computer Science. Advisor: Prof. Zhizheng Wu Shenzhen, China
Research Topics: Audio Enhancement, Music Generation.

Selected Publications

Full List at *Google Scholar*, *Equal Contribution.

- **Junan Zhang**, Jing Yang, Zihao Fang, Yuancheng Wang, Zehua Zhang, Zhuo Wang, Fan Fan, Zhizheng Wu.
  AnyEnhance: A Unified Generative Model with Prompt-Guidance and Self-Critic for Voice Enhancement
  Accepted. *IEEE Transactions on Audio, Speech, and Language Processing (TASLP), 2025.*
- **Junan Zhang**, Xueyao Zhang, Jing Yang, Yuancheng Wang, Fan Fan, Zhizheng Wu.
  Multi-Metric Preference Alignment for Generative Speech Restoration
  Under review. AAAI Conference on Artificial Intelligence (AAAI), 2026.
- **Junan Zhang***, Yunjia Zhang*, Xueyao Zhang, Zhizheng Wu.
  AnyAccomp: Generalizable Accompaniment Generation via Quantized Melodic Bottleneck
  Under review. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2026.
- **Junan Zhang***, Yunjia Zhang*, Shaohan Jiang*, Jing Yang, Fan Fan, Zhizheng Wu.
  SingVERSE: A Diverse, Real-World Benchmark for Singing Voice Enhancement Evaluation
  Under review. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2026.
- Xueyao Zhang, **Junan Zhang**, Yuancheng Wang, Chaoren Wang, Yuanzhe Chen, Dongya Jia, Zhuo Chen, Zhizheng Wu.
  Vevo2: Bridging Controllable Speech and Singing Voice Generation via Unified Prosody Learning
  Under review, 2025.
- Yicheng Gu, **Junan Zhang**, Chaoren Wang, Jerry Li, Zhizheng Wu, Lauri Juvela.
  Aliasing Free Neural Audio Synthesis
  Under review. IEEE Transactions on Audio, Speech, and Language Processing (TASLP), 2025.
- Yuancheng Wang, Jiachen Zheng, **Junan Zhang**, Xueyao Zhang, Huan Liao, Zhizheng Wu.
  Metis: A Foundation Speech Generation Model with Masked Generative Pre-training
  Accepted. *Annual Conference on Neural Information Processing Systems (NeurIPS), 2025.*
- Yicheng Gu*, Chaoren Wang*, **Junan Zhang***, Xueyao Zhang, Zihao Fang, Haorui He, Zhizheng Wu.
  SingNet: Towards a Large-Scale, Diverse, and In-the-Wild Singing Voice Dataset
  Technical Report, 2024.
- Junyan Ye*, Baichuan Zhou*, Zilong Huang*, **Junan Zhang***, Tianyi Bai, Hengrui Kang, Jun He, Honglin Lin, Zihao Wang, Tong Wu, Zhizheng Wu, Yiping Chen, Dahua Lin, Conghui He, Weijia Li.
  LOKI: A Comprehensive Synthetic Data Detection Benchmark using Large Multimodal Models
  Accepted **(*Spotlight*).** *International Conference on Learning Representations (ICLR), 2025.*

Research Experience

•CCF Advanced Audio Technology Competition 2025 *June 2025 - August 2025*
Lead Organizer (Track 1: Speech Restoration) Shenzhen, China

- Spearheaded the complete technical design and execution of the Speech Restoration track for *CCF Advanced Audio Technology Competition* (AATC 2025) hosted by the China Computer Federation (CCF).
- Engineered and open-sourced the official baseline system based on the AnyEnhance model, providing a robust and competitive starting point for participating teams.
- Defined the competition's entire technical architecture, from problem formulation and data curation to the comprehensive evaluation protocol (incl. DNSMOS, PESQ, complexity constraints), attracting over 100 registered teams.

•CUHKSZ Speech Lab *July 2024 - Present*

PhD Research Assistant, supervised by Prof. Zhizheng Wu Shenzhen, China

- Led the development of AnyEnhance, a unified generative model for enhancing both speech and singing voices, capable of handling tasks such as denoising, dereverberation, declipping, super-resolution, and target speaker extraction simultaneously without fine-tuning.
- Introduced a prompt-guidance mechanism for in-context learning, both improving enhancement performance and enabling target speaker extraction without architectural changes. And a self-critic mechanism into enhancing masked generative process.
- Conducted extensive experiments demonstrating that AnyEnhance outperforms existing methods in both objective metrics and subjective listening tests across various enhancement tasks.
- Published a first-author paper to TASLP 2025.

•Shanghai AI Lab *January 2024 - June 2024*

Undergraduate Research Intern Shanghai, China

- Co-developed a comprehensive benchmark for detecting synthetic data via large multimodal models.
- Conducted extensive research on synthetic datasets and open/closed-source models across four domains: speech, singing, audio, and music. Collected and generated 2,280 real/synthetic parallel audios, curated corresponding QA datasets for benchmarking.
- Evaluated 7 audio multimodal large models, including Qwen-Audio, SALMONN, AnyGPT, and Gemini-1.5-Flash.
- Discovered that current audio multimodal large models struggle to distinguish synthetic audio from real audio (all ∼50% accuracy), with this benchmark forming the foundation of the LOKI project, accepted by ICLR 2025 (Spotlight).

•Software Engineering Lab, Fudan University *May 2022 - June 2023*

Undergraduate Research Assistant, supervised by Prof. Bihuan Chen Shanghai, China

- Led a project leveraging pre-trained language models (BERT, RoBERTa) and malicious behavior sequence modeling to detect malicious packages in PyPI/NPM, outperforming state-of-the-art methods by 10.8% in precision and 7.3% in recall.
- Discovered 1,482 novel malicious packages in the wild (683 PyPI, 799 NPM), leading to 707 official acknowledgements from the PyPI and NPM security teams.
- Published the complete methodology and findings as a first-author paper in TOSEM 2024.

Project Experience

•Amphion Development *November 2023 - Present*

Core Contributor

- Acted as a core contributor to *Amphion*, an audio generation toolkit with over 9,000 GitHub stars.
- Led the development and integration of multiple key modules, including Speech Enhancement (SE), Singing Voice Conversion (SVC), and Text-to-Music (TTM) generation.
- Contributed two original, novel models to the framework: AnyEnhance for unified voice enhancement and AnyAccomp for accompaniment generation.
- Implemented and integrated a VITS-based model for singing voice conversion.

Honors & Awards

•Merit Fellowship for PhD Students, The Chinese University of Hong Kong, Shenzhen *September 2024*

•Scholarship for Outstanding Students, Fudan University (Top 20%) *November 2023, November 2022*

Skills

Programming Languages: Python, Bash, LaTeX
Developer Tools: Linux, Git
Framework: PyTorch, Hugging Face, Transformers
Language: Chinese(Native), English(Fluent, TOEFL 102(S 23))
Music Production: Semi-Professional, over 2,000,000 streams on Netease Cloud Music