Haoran Wang

wanghaor@sjtu.edu.cn / https://whr-a.github.io/

EDUCATION

Carnegie Mellon University — Pittsburgh, PA
Visiting scholar at WAVLab, Language Technologies Institute — 08/2025 – 12/2025
Advisor: Prof. Shinji Watanabe

Shanghai Jiao Tong University — Shanghai, China
B.Eng. in Computer Science & Technology — 09/2022 – 06/2026 (Expected)
Research Intern at X-LANCE Lab, School of Computer Science
Advisor: Prof. Kai Yu

Member of (ACM Honors Class), an elite CS program for top 5% students

- **Grade** (All): 87.1/100, 19 A/A+ courses
- Selected core courses: Computer Vision: 100/100, Intelligent Speech Technology: 100/100, Natural Language Processing: 97/100, Large Language Model: 99/100, Computational Complexity Theory: 97/100

RESEARCH INTERESTS

My research interests lie broadly in machine learning for speech and general audio, with a focus on generative modeling and representation learning for general-purpose audio systems, including but not limited to neural audio codecs and speech language models. My long-term goal is to pioneer a universal audio foundation model capable of holistically understanding and generating the full spectrum of sound, including speech, music, and complex acoustic scenes, while making these models high-fidelity, controllable, and efficient.

SELECTED PUBLICATIONS

BSCodec: A Band-Split Neural Codec for High-Quality Universal Audio Reconstruction
Haoran Wang, Jiatong Shi, Jinchuan Tian, Bohan Li, Kai Yu, Shinji Watanabe
Under review. [arxiv] [code] [demo]

Towards General Discrete Speech Codec for Complex Acoustic Environments: A Study of Reconstruction and Downstream Task Consistency
Haoran Wang, Guanyu Chen, Bohan Li, Hankun Wang, Yiwei Guo, Zhihan Li, Xie Chen, Kai Yu
IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2025. [arxiv]

PURE Codec: Progressive Unfolding of Residual Entropy for Speech Codec Learning
Jiatong Shi, Haoran Wang, William Chen, Chenda Li, Wangyou Zhang, Jinchuan Tian, Shinji Watanabe
IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2025.

Why Do Speech Language Models Fail to Generate Semantically Coherent Outputs? A Modality Evolving Perspective
Hankun Wang, Haoran Wang, Yiwei Guo, Zhihan Li, Chenpeng Du, Xie Chen, Kai Yu
Under review. [arxiv]

RESEARCH EXPERIENCE

Carnegie Mellon University, WAVLab LTI/CMU — Pittsburgh, PA
Advisor: Prof. Shinji Watanabe

Efficient Content-Adaptive Unified Audio Coding for multi-domain audio — 06/2025 – 12/2025

- Developed a highly effective solution to this problem by leveraging a band-splitting strategy to address the multi-domain challenge. (*first author, under review*)

- We apply band-splitting and use an independent VQ codebook for each frequency band. This approach is designed to capture far more meaningful information than the often uninterpretable residuals in RVQ, enabling better performance at lower bitrates.
- We apply band-splitting to the encoder and decoder to substantially reduce the model capacity required to adapt to diverse multi-domain data.

Entropy-Guided Robust Speech Codec Learning 04/2025 – 06/2025

- Developed PURE Codec, a novel framework that stabilizes residual vector quantization (RVQ) training by leveraging enhancement-guided supervision to anchor codebooks to low-entropy speech representations. (*second author, accepted by ASRU 2025*)
- We design a "Progressive Unfolding of Residual Entropy" mechanism that explicitly separates low-entropy denoised content from high-entropy residuals across quantization stages.

Shanghai Jiao Tong University, X-LANCE Lab Shanghai, China
Advisor: Prof. Kai Yu

General Discrete Speech Codec for Complex Acoustic Environments 11/2024 – 02/2025

- We propose the Environment-Resilient Speech Codec Benchmark (ERSB) to systematically evaluate the environmental resilience of neural speech codecs, addressing the lack of exploration in complex noise scenarios. (*first author, accepted by ASRU 2025*)
- We design a comprehensive evaluation pipeline that assesses two key capabilities: robust reconstruction quality for preserving non-speech acoustic details, and downstream task consistency to ensure stability in SE and ASR backends.

Semantic Coherence in End-to-End Speech Language Models 09/2024 – 12/2024

- We propose a "Modality Evolving" perspective to systematically analyze the underlying causes of semantic incoherence in Speech Language Models (SLMs), bridging the gap between text-based and speech-based generation. (*second author, under review*)
- We reveal that paralinguistic information exerts the most significant negative impact, particularly by hindering basic lexical modeling, while sequence length primarily affects syntactic and semantic consistency.

TEACHING EXPERIENCE

Computer Systems (Operating Systems) 09/2024 – 01/2025
Computer Systems (Architecture) 02/2024 – 06/2024

Role as teaching assistant: giving lectures and recitation classes, writing documents and sample solutions, grading homework, creating exam questions, designing final project lists, etc.

HONORS

2025 Zhiyuan Honorary Scholarship (**Top 2%** in SJTU)
2024 Zhiyuan Honorary Scholarship (**Top 2%** in SJTU)
2023 Zhiyuan Honorary Scholarship (**Top 2%** in SJTU)
2022 Zhiyuan Honorary Scholarship (**Top 2%** in SJTU)

TECHNICAL SKILLS

Language: TOEFL: Total 102 (Reading 27, Listening 25, Speaking 22, Writing 28)
Programming and Software: Python, C++, Bash, Git, Java, and Verilog
DL Frameworks/Toolkits: PyTorch, **ESPnet (Contributor & Codec module maintainer)**