Junkai Wu

junkaiwu@uw.edu wjk0925.github.io/ +1 206-482-8933
LinkedIn Google Scholar GitHub

Education

University of Washington Seattle, WA
Ph.D. Candidate in Electrical and Computer Engineering *Sep. 2023 – Present*

- Advisor: Mari Ostendorf

University of Illinois Urbana-Champaign Champaign, IL
B.S. in Computer Engineering, Minor in Mathematics *Aug. 2019 – May 2023*

- Advisors: Paris Smaragdis, Mark Hasegawa-Johnson

Research Interests

My research focuses on audio-centric multimodal AI, especially text and multimodal large language models that understand and generate speech, music, and sound, both end to end and as components of larger systems.

Industry Research Experience

Adobe Research - Music AI Group June 2026 – Sep. 2026
Research Scientist/Engineer Intern *San Francisco, CA*

- Supervisors: Zhepei Wang and Nicholas J. Bryan
- **Text-to-Music Generation with Structured Captions:** Improved music-text alignment through structured-caption training and LLM rewriting, and developed an interpretable evaluation metric that better matches human judgments. Manuscript under review at ICASSP 2027; preprint available on arXiv.

Academic Research Experience

University of Washington - TIAL Lab Oct. 2023 – Present
Research Assistant *Seattle, WA*

- Advisor: Mari Ostendorf
- **Accented Text-to-Speech:** Built a web-scale accented speech collection pipeline and trained flow-matching models for zero-shot accented speech synthesis.
- **LLM Navigation Agents:** Developed a tool-using dialogue agent and fine-tuned LLMs for user simulation and proactive navigation assistance.

University of Illinois Urbana-Champaign - Computational Audio Lab May 2021 – May 2023

Undergraduate Research Assistant *Champaign, IL*

- Advisor: Paris Smaragdis
- **Meta-Learning for Adaptive Filters:** Developed neural networks that learn adaptive filter update rules across frequencies, improving acoustic echo cancellation accuracy and computational efficiency.
- **Continual Self-Supervised Learning for Audio:** Implemented a framework that learns new sound classes from unlabeled audio while retaining representations of previously learned classes.

University of Illinois Urbana-Champaign - Statistical Speech Technology Group May 2022 – May 2023

Undergraduate Research Assistant *Champaign, IL*

- Advisor: Mark Hasegawa-Johnson
- **Speech Data Augmentation:** Generated synthetic speech with a speech-unit language model and unit-WaveNet for self-supervised training.
- **Speech Synthesis with Text to Unit Translation:** Built a text-to-speech system using discrete speech units and studied how unit representation and decoding affect synthesis quality, exploring potential benefits for noise robustness.

Teaching Experience

University of Washington *Seattle, WA*

- **EE P 598: Introduction to Digital Audio** — Teaching Assistant (Spring 2026)
- **EE 344: Data-Driven Modeling and Machine Learning** — Teaching Assistant (Fall 2026)

University of Illinois Urbana-Champaign *Champaign, IL*

- **CS 448: Audio Computing Lab** — Course Assistant (Spring 2023)

Honors & Awards

Best Paper Award - WASPAA 2025 Oct. 2025

- Awarded to "Bridging Ears and Eyes" at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA); selected as the best paper among 95 accepted papers.

Robert E. Rushmer Endowed Fellowship - UW ECE Mar. 2023

- Awarded by UW ECE to exceptional applicants to its doctoral program; covers tuition and a stipend for the 2023–2024 academic year.

Publications

[*]Equal contributions

[12] Text-to-Music Generation via Structured Captions and Interpretable Evaluation
Junkai Wu, Zhepei Wang, Ge Zhu, Sarah Hye-yeon Lee, Mari Ostendorf, Nicholas J. Bryan.
Submitted to ICASSP 2027; under review

[11] Rubric-Based Optimization for Text-to-Music Generation
Ping Wang[*], Guang Yang[*], Shao-Rong Su[*], **Junkai Wu**, Pang Wei Koh, Noah A. Smith.
Submitted to ICLR 2027; under review

[10] Cocktail-Talker: Multi-Speaker Dialog Modeling in Noisy Social Environments with Turn Action GRPO
Xilin Jiang, Riki Shimizu, Sukru Samet Dindar, **Junkai Wu**, Zhongweiyang Xu, Nima Mesgarani.
SLT 2026 (accepted)

[9] AVMeme Exam: A Multimodal Multilingual Multicultural Benchmark for LLMs' Contextual and Cultural Knowledge and Thinking
Xilin Jiang*, Qiaolin Wang*, **Junkai Wu***, Xiaomin He*, Zhongweiyang Xu*, et al.
COLM 2026 (accepted)

[8] SightSound-R1: Cross-Modal Reasoning Distillation from Vision to Audio Language Models
Qiaolin Wang, Xilin Jiang, Linyang He, **Junkai Wu**, Nima Mesgarani.
ICASSP 2026

[7] Bridging Ears and Eyes: Analyzing Audio and Visual Large Language Models to Humans in Visible Sound Recognition and Reducing Their Sensory Gap via Cross-Modal Distillation
Xilin Jiang*, **Junkai Wu***, Vishal Choudhari, Nima Mesgarani.
WASPAA 2025 (Best Paper Award)

[6] Just ASR + LLM? A Study on Speech Large Language Models' Ability to Identify and Understand Speaker in Spoken Dialogue
Junkai Wu*, Xulin Fan*, Bo-Ru Lu, Xilin Jiang, Nima Mesgarani, Mark A Hasegawa-Johnson, Mari Ostendorf.
SLT 2024

[5] Meta-AF Echo Cancellation for Improved Keyword Spotting
Jonah Casebeer, **Junkai Wu**, Paris Smaragdis.
ICASSP 2024

[4] Unsupervised Improvement of Audio-Text Cross-Modal Representations
Zhepei Wang, Cem Subakan, Krishna Subramani, **Junkai Wu**, Tiago Tavares, Fabio Ayres, Paris Smaragdis.
WASPAA 2023

[3] Listen, Decipher and Sign: Toward Unsupervised Speech-to-Sign Language Recognition
Liming Wang, Junrui Ni, Heting Gao, Jialu Li, Kai Chieh Chang, Xulin Fan, **Junkai Wu**, Mark Hasegawa-Johnson, Chang Yoo.
ACL Findings 2023

[2] Learning Representations for New Sound Classes With Continual Self-Supervised Learning
Zhepei Wang, Cem Subakan, Xilin Jiang, **Junkai Wu**, Efthymios Tzinis, Mirco Ravanelli, Paris Smaragdis.
IEEE Signal Processing Letters 2022

[1] Meta-Learning for Adaptive Filters with Higher-Order Frequency Dependencies
Junkai Wu, Jonah Casebeer, Nicholas J. Bryan, Paris Smaragdis.
IWAENC 2022

SKILLS

Languages: English, Mandarin

Programming: Python, PyTorch, JAX, NumPy, HuggingFace, SuperCollider