Junkai Wu

206-482-8933 | junkaiwu@uw.edu | wjk0925.github.io/ | github.com/wjk0925

Education

University of Washington Seattle, WA
Ph.D. in Electrical & Computer Engineering *Sep. 2023 – present*
- Advisor: Mari Ostendorf
- Robert E. Rushmer Endowed Fellowship

University of Illinois at 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 interests are in audio, speech & language processing. I am currently working on text-to-speech, audio/speech large language models, and large language model agents. I have also focused on machine learning for audio signal processing.

Research Experience

Transformation, Interpretation and Analysis of Language (TIAL) Group Oct 2023 – Present
Advisor: Mari Ostendorf *Seattle, WA*
- **Project 1: Text-To-Speech for Accented English**
- Designed and implemented an automated pipeline to collect large-scale, high-quality spontaneous accented English speech from the web.
- Trained and evaluated zero-shot text-to-speech models using flow-matching techniques, focusing on accented speech synthesis.
- **Project 2: LLM Agent for Navigation**
- Developed an LLM-based navigation agent leveraging MCP server integration and tool calling for multi-turn dialogue support.
- Fine-tuned open-source LLMs to simulate user behavior more effectively in interactive navigation tasks.
- Enhancing baseline agent LLMs to be more effective and proactive via supervised fine-tuning and reinforcement learning.

Computational Audio Lab May 2021 – May 2023
Advisor: Paris Smaragdis *Champaign, IL*
- **Project 1: Meta-Learning for Adaptive Filters**
- Proposed Higher-Order Meta-AF with improved performance and efficiency in double-talk acoustic echo cancellation.
- Built a pipeline for simulating challenging room acoustic scenes with pyroomacoustics and helping develop meta-optimizers trained with supervised loss.
- **Project 2: Continual Self-Supervised Learning for Sound**
- Implemented CSSL algorithm with MOCO and distillation loss for the project Continual Self-Supervised Learning (CSSL) of New Sound Classes.

Statistical Speech Technology Group May 2022 – May 2023
Advisor: Mark Hasegawa-Johnson *Champaign, IL*
- **Project 1: Speech Dataset Augmentation with Speech Generation Models**
- Worked on speech dataset augmentation for self-supervised training by generating synthetic speech with speech-unit language model and unit-WaveNet.
- Built evaluation pipeline for unit language models with Fairseq framework.

Publications (*equal contributions)

[**8**] SightSound-R1: Cross-Modal Reasoning Distillation from Vision to Audio Language Models. Qiaolin Wang, Xilin Jiang, Linyang He, **Junkai Wu**, Nima Mesgarani. **Under Review**
[**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**.
[**1**] Meta-Learning for Adaptive Filters with Higher-Order Frequency Dependencies. **Junkai Wu**, Jonah Casebeer, Nicholas J. Bryan, Paris Smaragdis. **IWAENC 2022**

Projects

Melody Transcription with Self-Supervised Music Features Feb 2024 – Mar 2024

- Reproduced the melody transcription model from "Melody transcription via generative pre-training".
- Evaluated the model on out-of-domain Irish traditional music.
- Experimented with separate onset prediction loss and pitch prediction loss, replacing Jukebox features with MERT features.

Speech Synthesis with Text to Unit Translation Sep 2022 – Dec 2022

- Developed a speech synthesis system that consists of a HubERT + KMeans speech to discrete units (s2u) model, a transformer text to discrete units (t2u) model, a HiFiGAN discrete units to speech (u2s) model.
- Studied how KMeans size and t2u beam search size influence speech synthesis quality. Explored the potential advantage brought by the robustness to noise property of s2u.

Wav2vec 2.0 Pretraining from Scratch on Non-Western Languages Feb 2022 – May 2022

- Pretrained Wav2vec 2.0 model on UN Proceedings Mandarin corpus and finetuned it for automatic speech recognition (ASR) on GlobalPhone Mandarin corpus with Fairseq toolkit.
- Fine-tunied pretrained English w2v2 and XLSR w2v2 models on the same Mandarin corpus for comparison.

Unsupervised Incremental Learning for Acoustic Scene Classification Oct 2021 – Dec 2021

- Implemented an acoustic scene classification model with WAV2CLIP on UrbanSound8k.
- Developed a confusion based novelty detection mechanism and a dataloader for generating unlabeled training data exposures, trained the acoustic scene classifier incrementally without supervision.

Learning to Learn Implementation in JAX May 2021 – June 2021

- Implemented the optimization algorithm from the paper Learning to Learn by Gradient Descent by Gradient Descent with JAX framework.
- Tested the algorithm's performance on quadratic problems and classification with multilayer perceptron for MNIST.

Skills & Courses

Languages: Python, Java, C, C++
Frameworks & Tools: PyTorch, JAX, deepmind-Haiku, SciPy, Git, Fairseq, AutoGen
Courses: Machine Learning, Deep Learning, Digital Signal Processing, Audio Computing, Speech Processing, Natural Language Processing, AI for Music, Linear Algebra, Optimization, Complex Variables, Random Process