Yutong Wen

(+1) 585-754-3106 | yutong12@illinois.edu | Github: AgentCooper2002 | Github: YutongWen

EDUCATION

University of Illinois Urbana-Champaign Aug 2024 - present

Phd in Computer Science

- Advisor: Paris Smaragdis and Minje Kim
- Research interests: audio source separation, controllable audio generation and editing, spatial audio

University of Rochester Aug 2020 - May 2024

BS in Audio and Music Engineering

- GPA: 3.82
- Advisor: Zhiyao Duan

PUBLICATIONS

- **Yutong Wen**, You Zhang, and Zhiyao Duan. "Mitigating Cross-Database Differences for Learning Unified HRTF Representation." *2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2023. [DOI][code]
- Ge Zhu, **Yutong Wen (co-first author)**, Marc-André Carbonneau, and Zhiyao Duan. "EDMSound: Spectrogram Based Diffusion Models for Efficient and High-Quality Audio Synthesis." *The Machine Learning for Audio workshop at Neural Information Processing Systems Conference*, 2023. [arxiv][code] [project webpage]

ACADEMIC EXPERIENCES

Audio Lab Aug 2024 - Present

Research Assistant, Advisor: Paris Smaragdis and Minje Kim

Audio Information Research Lab Dec 2022 - May 2024

Research Assistant, Advisor: Zhiyao Duan

A Complex Spectrogram Domain Diffusion Framework

- Proposed a versatile audio diffusion model utilizing conditional inputs to address diverse audio tasks, including audio restoration, TTS, and label conditioned generation other tasks;
- Modularized components of diffusion schedule and samplers;
- Implemented auxiliary conditioner including controlNet-like condition network.

Audio Sound Effects Replication Detection

- Designed data augmentation methods that covers a range of common audio transformations;
- Implemented an audio encoder for feature extraction, and applied embedding normalization and whitening for similarity computation;
- Train the model in a self-supervised manner employing modified InfoNCE loss and Entropy Loss.

Estimating the Direction of Arrival of an Acoustic Wave

Supervised by Michael Heilemann

- Co-designed and Recorded the dataset for speech command that was used to train the neural network that learns to estimate the direction of arrival of a speech command.

SELECTED PROJECTS

UniNav:A Unified Framework for Fine and Coarse-grained Instruction in Vision Language Navigation

- Participated in the development of UniNav, a unified AI framework designed to interpret and execute both fine-grained and coarse-grained natural language navigation instructions within diverse environments.

Source Separation Project
- Developed a source separation algorithm based on Non-Negative Matrix Factorization, and crafted a model employing a CNN-LSTM architecture. This model efficiently computed spectrogram masks running on the DSD100 dataset.

FM Synthesis Parameter Optimization Program
- Designed and coded in C++ with JUCE a program that is able to learn a given instrument's waveform both amplitude-wise and frequency-wise using Genetic Algorithm based on FM synthesis and other supporting algorithms. The program is able to learn at a reliability up to 93% measured by MSE.

Virtual Loudspeaker Music Player
- Designed and coded in C++ a music player that is able to create the sense of virtual loudspeaker in headphones listening utilizing methods of HRTF and room IR or delay line.

Solid State Amplifier Simulating Tube Sounds
- Designed an amplifier output stage based on the complementary output stage to simulate the distortion pattern of a tube sound using a voltage follower and special design of asymmetric amplification factors in the positive and negative branches.

Dynamic Storage Allocator Implementation
- Developed a dynamic storage allocator in C, encompassing the creation of personalized versions of malloc, free, and realloc routines. Emphasized creative exploration within the design space to implement an allocator that prioritizes correctness, efficiency, and speed.

PRESENTATIONS AND AWARDS

2024 University of Rochester Donald M. Barnard Fund

Mitigating Cross-Database Differences for Learning Unified HRTF Representation
2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (Oral Presentation)
WASPAA 2023 travel grant

EDMSound: Spectrogram Based Diffusion Models for Efficient and High-Quality Audio Synthesis
Speech and Audio in the Northeast (SANE) 2023 (Poster Presentation)

2023 University of Rochester Undergraduate Research Presentation Award

INTERNSHIP

Polycom, Plantronics, Inc, Chengdu, China May 2021 - July 2021
Engineer Assistant
- Assisted the design for solutions to A/V conference rooms;
- Helped with the setup of the A/V conference system in Industrial Bank Chengdu Branch.

SKILLS

- Python, C/C++, JUCE, Java, Matlab, Max/Msp, Faust
- optimization algorithms, including DP, graph, FFT, and so on
- Circuit analysis, analog and digital circuitry and acoustics
- professional audio products, recording techniques, and mixing techniques