Yutong Wen

(+1) 585-754-3106 | yutong12@illinois.edu | Github: AgentCooper2002 | Github: YutongWen |
Personal Webpage | Google Scholar | LinkedIn

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

University of Rochester Aug 2020 - May 2024

BS in Audio and Music Engineering

- GPA: 3.82
- Advisor: Zhiyao Duan

PUBLICATIONS

[6] **Yutong Wen**, Ke Chen, Prem Seetharaman, Oriol Nieto, Jiaqi Su, Rithesh Kumar, Minje Kim, Paris Smaragdis, Zeyu Jin, Justin Salamon. "PromptSep:Generative Audio Separation via Multimodal Prompting." *Under Review*, 2025. [project webpage]

[5] Yurii Halychanskyi, Cameron Churchwell, **Yutong Wen**, Volodymyr Kindratenko. "FAC-FACodec: Controllable Zero-Shot Foreign Accent Conversion with Factorized Speech Codec." *Under Review*, 2025. [project webpage]

[4] Zhu, Ge, **Yutong Wen**, and Zhiyao Duan. "A Review on Score-based Generative Models for Audio Applications." *Under Review,* 2025. [arxiv][code]

[3] **Yutong Wen**, Minje Kim, and Paris Smaragdis. "User-guided Generative Source Separation." *Proc. of the 26th International Society for Music Information Retrieval Conference*, 2025. [arxiv] [code][project webpage]

[2] **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]

[1] 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]

RESEARCH WORK EXPERIENCE

Adobe Research May 2025 - Nov 2025

Speech AI Lab & SODA. Research Scientist/Engineer Intern. Supervisor: Ke Chen

- Proposed a separation model with guidance via text operators and vocal imitation (publ. [6]);
- Invented removal text operator enabling sound removal and extraction in one system;
- Proposed new metrics for language-guided audio source separation for better evaluation;
- Enabled audio-visual separation by object clicking via integration with SAM2 and VLM.

ACADEMIC EXPERIENCES

Audio Lab Aug 2024 - Present

Research Assistant, Advisor: Paris Smaragdis and Minje Kim

Diffusion-Based User-Guided Music Source Separation

- Proposed one of the first Diffusion framework for music source separation (publ. [3]);
- Enabled stem-agnostic guidance by user-inputs including humming and spectral masking;

- Attained high Signal-to-Distortion Ratio that is comparable to predictive approaches.

Diffusion-based Music Editing Model
- Proposed a framework for music generation, separation, timbre transfer, and pitch correction;
- Implemented multi-channel complex-spectrogram Diffusion model;
- Proposed auxiliary separation regularization loss for Diffusion Score-prediction objective.

Full-Song-Length Music Source Separation with Signal-Domain Diffusion
- Invented infinite-length diffusion generation for full-song-length music separation;
- Solved signal-domain diffusion short window issue for music tasks.

Audio Information Research Lab Dec 2022 - May 2024
Research Assistant, Advisor: Zhiyao Duan

A Complex Spectrogram Domain Diffusion Framework
- Proposed an audio Diffusion model with conditional inputs to address diverse audio tasks including audio restoration, TTS, and others (publ. [1][4]);
- Modularized components of Diffusion schedule and Samplers;
- Implemented auxiliary conditioner including controlNet-like condition network.

Cross HRTF Database Normalization
- Proposed an algorithm to normalize the difference among HRTF databases (publ. [2]);
- Improved HRTF representation learning accuracy in mixed-databased training by 15%;

ACADEMIC SERVICE

Conference Reviewer: ICASSP 2026

PRESENTATIONS AND AWARDS

User-guided Generative Source Separation
2025 International Society for Music Information Retrieval Conference (Oral and poster presentation)

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

Spectrogram Based Diffusion Models for Efficient and High-Quality Audio Synthesis
Machine Learning for Audio Workshop at NeurIPS 2023 (Poster Presentation)
Speech and Audio in the Northeast (SANE) 2023 (Poster Presentation)

2023 University of Rochester Undergraduate Research Presentation Award

SKILLS

- Machine Learning Programing: Python, PyTorch, PyTorch Lightening, Huggingface
- Audio Software Programing: C/C++, JUCE, Java, Matlab, Max/Msp, Faust
- Hardware Developing Skills: KiCad, LTspice, Circuit analysis, analog circuitry and acoustics
- Audio Production Skills: DSP audio products, multi-channel recording, and mixing techniques