

Education

Johns Hopkins University

Electrical and Computer Engineering
Advisor: Prof. Mounya Elhilali

Doctor of Philosophy
Aug 2022 – Present

Tsinghua University

M.E. in Civil Engineering
Member of Big Data Program

Master
Aug 2020 – Jun 2022

Tsinghua University

B.E. in Civil Engineering
B.S. in Business Analytics

Undergraduate
Aug 2016 – Jun 2020

Selected Projects

EzAudio: Enhancing Text-to-Audio Generation with Efficient Diffusion Transformer [1]

May 2024 – Aug 2024

Intern at Tencent AI Lab

Selected to Hugging Face's Daily Paper.  [Homepage](#)


Mentors: Dr. Yong Xu, Dr. Hao Zhang, Dr. Dong Yu

- Proposed a text-to-audio generation model featuring an efficient diffusion transformer and a multi-stage training strategy, surpassing open-source text-to-audio generators in both subjective and objective evaluations.
- Developed audio editing tools and integrated ControlNet for real-world sound and music production applications.

DreamVoice: Text-Guided Voice Conversion [4]

Feb 2024 – Apr 2024

Johns Hopkins University | Collaborated with Myshell

 [Homepage](#)

Advisor: Prof. Mounya Elhilali

- Created a dataset for fine-grained speaker timbre characterization.
- Proposed two models: (1) An end-to-end diffusion model for text-guided voice conversion. (2) A plugin model that adapts pre-trained voice conversion systems to support text-guided voice conversion.

Diff-Pitcher: Diffusion-based Singing Voice Pitch Correction [9]

Mar 2023 – May 2023

Johns Hopkins University

Highly Praised by Hollywood music producer David Kershenbaum.  [Homepage](#)

Advisor: Prof. Mounya Elhilali

- Proposed a diffusion-based automatic pitch correction system that generates natural-sounding singing voices.

Selected Pre-prints and Publications

- Jiarui Hai**, Yong Xu, Hao Zhang, Chenxing Li, Helin Wang, Mounya Elhilali, Dong Yu. EzAudio: Enhancing Text-to-Audio Generation with Efficient Diffusion Transformer. Submitted to ICASSP 2025.
- Helin Wang*, **Jiarui Hai***, Yen-Ju Lu, Karan Thakkar, Mounya Elhilali, Najim Dehak. SoloAudio: Target Sound Extraction with Language-oriented Audio Diffusion Transformer. Submitted to ICASSP 2025.
- Helin Wang, Meng Yu, **Jiarui Hai**, Chen Chen, Yuchen Hu, Rilin Chen, Najim Dehak, Dong Yu. SSR Speech: Towards Stable, Safe, and Robust Zero-shot Text-based Speech Editing and Synthesis. Submitted to ICASSP 2025.
- Jiarui Hai***, Karan Thakkar*, Helin Wang, Zengyi Qin, Mounya Elhilali. DreamVoice: Text-Guided Voice Conversion. Interspeech 2024.
- Helin Wang, Jesus Villalba, Laureano Moro-Velazquez, **Jiarui Hai**, Thomas Thebaud, Najim Dehak. Noise-robust Speech Separation with Fast Generative Correction. Interspeech 2024.
- Jiarui Hai***, Helin Wang*, Dongchao Yang, Karan Thakkar, Najim Dehak, Mounya Elhilali. DPM-TSE: A Diffusion Probabilistic Model for Target Sound Extraction. ICASSP 2024.
- Karan Thakkar, **Jiarui Hai**, Mounya Elhilali. Investigating Self-Supervised Deep Representations for EEG-based Auditory Attention Decoding. ICASSP 2024.
- Jiarui Hai***, Yu-jeh Liu*, Mounya Elhilali. Boosting Modality Representation with Pre-trained Models and Multi-task Training for Multimodal Sentiment Analysis. ASRU 2023.
- Jiarui Hai**, Mounya Elhilali. Diff-Pitcher: Diffusion-based Singing Voice Pitch Correction. WASPAA 2023.
- Rui Lu, Baigong Zheng, **Jiarui Hai**, Fei Tao, Zhiyao Duan, Ji Liu. Progressive Teacher-Student Training Framework for Music Tagging. ICASSP 2022.

Music Experiences

- Co-founder, InnerSup Records, a Chinese record label and record producer collective 2021
- Featured producer in a rap song on HIPHOP BANK, a Chinese TV show August 2021
- Delivered a lecture on music production at Modern Sky, China’s leading music company June 2021
- Won Top 1% New Producer at the BeatsHome Hip-hop Production Contest April 2021
- My music video ranked in the Top 10 on Bilibili’s trending board October 2020
- President, Rap/HipHop Association at Tsinghua University 2020–2021
- Founder and Leader, HipHop Production Group at Tsinghua University 2018–2022