Helin Wang

hwang258@jhu.edu · (+1) 667-354-9059 · helin-wang-2a74671b3

Basic Information

Home Page: https://wanghelin1997.github.io/helinwang/
Google Scholar Page: https://scholar.google.com/citations?user=I_V0zBMAAAAJ
Github Page: https://github.com/WangHelin1997

Research Interest

My research interest majorly lies in AI for speech and audio signal processing, encompassing audio generation tasks such as source separation and text-to-speech synthesis, as well as audio understanding tasks like audio classification and captioning.

Educations

Johns Hopkins University, Baltimore, USA 2022 – Expected 2026
Ph.D candidate in Whiting School of Engineering (WSE)
Supervisor: Najim Dehak

Peking University, Beijing, China 2019 – 2022
Master student in School of Electronic and Computer Engineering (ECE)
Supervisor: Yuexian Zou

Tsinghua University, Beijing, China 2015 – 2019
B.S. in Department of Automation (DA)

Experiences

Meta, FAIR, New York, USA May 2025 – Present
Research on multi-modal audio separation
Research Scientist Intern Supervisor: Wei-Ning Hsu and Bowen Shi

Tencent AI Lab, Speech Group, Bellevue, USA May 2024 - August 2024
Research on multi-lingual speech editing and TTS
Intern Supervisors: Meng Yu and Dong Yu

Amazon, Amazon General Intelligence (AGI), Baltimore, USA December 2022 - December 2023
Research on spoken language understanding in atypical speech
Student Researcher Supervisors: Venkatesh Ravichandran and Milind Rao

Microsoft STCA, NLP Group, Beijing, China February 2022 - May 2022
Research on entity linking
Intern Supervisors: Linjun Shou and Ming Gong

Tencent AI Lab, Speech Group, Shenzhen, China May 2020 - November 2021
Research on speech enhancement
Intern Supervisors: Bo Wu and Chao Weng

Selected Publications

Speech and Audio Generation:

1. Multimodal unified audio separation: Bowen Shi*, Andros Tjandra*, John Hoffman*, **Helin Wang***, Yi-Chiao Wu*, Luya Gao*, Julius Richter, Matt Le, Apoorv Vyas, Sanyuan Chen, Christoph Feichtenhofer, Piotr Dollár, Wei-Ning Hsu, Ann Lee. *SAM Audio: Segment Anything in Audio*. Preprint, 2025.
2. TTS with rich speaker styles and sound events: **Helin Wang***, Jiarui Hai*, Dading Chong, Karan Thakkar, Tiantian Feng, Dongchao Yang, Junhyeok Lee, Laureano Moro Velazquez, Jesus Villalba, Zengyi Qin, Shrikanth Narayanan, Mounya Elhiali, Najim Dehak. *CapSpeech: Enabling Downstream Applications in Style-Captioned Text-to-Speech*. Preprint, 2025.
3. Fully-generative speech extraction: **Helin Wang**, Jiarui Hai, Dongchao Yang, Chen Chen, Kai Li, Junyi Peng, Thomas Thebaud, Laureano Moro-Velazquez, Jesus Villalba, Najim Dehak. *SoloSpeech: Enhancing Intelligibility and Quality in Target Speech Extraction through a Cascaded Generative Pipeline*. Preprint, 2025.
4. Multimodal generative audio extraction: **Helin Wang***, Jiarui Hai*, Yen-Ju Lu, Karan Thakkar, Mounya Elhilali, Najim Dehak. *SoloAudio: Target Sound Extraction with Language-oriented Audio Diffusion Transformer*. ICASSP 2025.
5. Robust and stable speech generation: **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*. ICASSP 2025.
6. Refinement for speech generation: **Helin Wang**, Jesus Villalba, Laureano Moro-Velazquez, Jiarui Hai, Thomas Thebaud, Najim Dehak. *Noise-robust Speech Separation with Fast Generative Correction*. INTERSPEECH 2024. (**Oral**). ⋆ **Best Paper Award** and **Best Student Paper Award** Nomination.
7. First generative audio extraction: Jiarui Hai*, **Helin Wang***, Dongchao Yang, Karan Thakkar, Najim Dehak, Mounya Elhilali. *DPM-TSE: A Diffusion Probabilistic Model for Target Sound Extraction*. ICASSP 2024.
8. Data Augmentation with TTS: **Helin Wang**, Venkatesh Ravichandran, Milind Rao, Becky Lammers, Becky Lammers, Myra Sydnor, Nicholas Maragakis, Ankur A. Butala, Jayne Zhang, Lora Clawson, Victoria Chovaz, Laureano Moro-Velazquez. *Improving fairness for spoken language understanding in atypical speech with Text-to-Speech*. NeurIPS 2023 Workshop on Synthetic Data Generation with Generative AI (**Oral**).
9. Duration-aware voice conversion: **Helin Wang**, Thomas Thebaud, Jesus Villalba, Myra Sydnor, Becky Lammers, Najim Dehak, Laureano Moro-Velazquez. *DuTa-VC: A Duration-aware Typical-to-atypical Voice Conversion Approach with Diffusion Probabilistic Model*. INTERSPEECH 2023.
10. Jointly audio detection and extraction: **Helin Wang***, Dongchao Yang*, Chao Weng, Jianwei Yu, Yuexian Zou. *Improving Target Sound Extraction with Timestamp Information*. INTERSPEECH 2022.
11. Attentive speech enhancement: **Helin Wang**, Bo Wu, Lianwu Chen, Meng Yu, Jianwei Yu, Yong Xu, Shi-Xiong Zhang, Chao Weng, Dan Su, Dong Yu. *TeCANet: Temporal-Contextual Attention Network for Environment-Aware Speech Dereverberation*. INTERSPEECH 2021.

Speech and Audio Understanding:

12. Multimodal unified audio separation judgement: **Helin Wang**, Bowen Shi, Andros Tjandra, John Hoffman, Yi-Chiao Wu, Apoorv Vyas, Najim Dehak, Ann Lee and Wei-Ning Hsu. *SAM Audio Judge: A Unified Multimodal Framework for Perceptual Evaluation of Audio Separation*. Preprint, 2025.
13. Masked audio pretraining: Dading Chong*, **Helin Wang***, Peilin Zhou, Qingcheng Zeng. *Masked Spectrogram Prediction For Self-Supervised Audio Pre-Training*. ICASSP 2023.
14. First target audio detection: **Helin Wang**, Dongchao Yang, Yuexian Zou, Fan Cui, Yujun Wang. . *Detect What You Want: Target Sound Detection*. DCASE 2022.
15. Jointly weak and strong supervision: Dongchao Yang*, **Helin Wang***, Wenwu Wang, Yuexian Zou. *A Mixed Supervised Learning Framework For Target Sound Detection*. DCASE 2022.
16. Attentive reference for audio detection: Dongchao Yang*, **Helin Wang***, Zhongjie Ye, Yuexian Zou, Wenwu Wang. *RaDur: A Reference-aware and Duration-robust Network for Target Sound Detection*. INTERSPEECH 2022.
17. Global and local attention in audio: **Helin Wang**, Yuexian Zou, Wenwu Wang. *A Global-Local Attention Framework for Weakly Labelled Audio Tagging*. ICASSP 2021.
18. Hidden space data augmentation: **Helin Wang**, Yuexian Zou, Wenwu Wang. *SpecAugment++: A Hidden Space Data Augmentation Method for Acoustic Scene Classification*. INTERSPEECH 2021.
19. Label dependencies in audio: **Helin Wang**, Yuexian Zou, Dading Chong, Wenwu Wang. *Modeling Label*

Dependencies for Audio Tagging With Graph Convolutional Network. IEEE Signal Processing Letters.
20. Spectrogram processing: **Helin Wang**, Yuexian Zou, Dading Chong. *Acoustic Scene Classification with Spectrogram Processing Strategies*. DCASE 2020.
21. Temporal-spectral attention in audio: **Helin Wang**, Yuexian Zou, Dading Chong, Wenwu Wang. *Environmental Sound Classification with Parallel Temporal-Spectral Attention*. INTERSPEECH 2020.

♡ Services (Reviewer)

- IEEE/ACM Transactions on Audio, Speech, and Language Processing
- IEEE Signal Processing Letters
- Neurocomputing
- INTERSPEECH 2023, 2024, 2025
- ICASSP 2023, 2024, 2025
- NeurIPS 2024, 2025
- ICLR 2025

♡ Teaching

2024/01 - 2024/05, *Teaching Assistant*, Johns Hopkins University, Baltimore, USA:
EN.520.123: Computational Modeling for Electrical and Computer Engineering in Spring 2024 in the Department of Electrical and Computer Engineering
2025/01 - 2025/05, *Teaching Assistant*, Johns Hopkins University, Baltimore, USA:
EN.520.629: Machine Learning for Medical Applications in Spring 2025 in the Department of Electrical and Computer Engineering

♡ Honors and Awards

IEEE SPS Young Author Best Paper Award	2024
Best Paper Award Nomination at INTERSPEECH	2024
1st Team Ranking of DCASE Challenge Task 5 (Judges' award)	2021
Outstanding Graduate Student Award of Peking University	2022
Outstanding Graduate Thesis Award of Peking University	2022
Award for Scientific Research of Peking University	2020-2021
School Prize of Peking University	2019-2020
Merit Student of Peking University	2019-2020