

# HELIN WANG

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

## 📄 BASIC INFORMATION

---

Home Page: <https://wanghelin1997.github.io/helinwang/>

Google Scholar Page: [https://scholar.google.com/citations?user=I\\_V0zBMAAAAJ](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 – Present

*Ph.D student* 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

---

**Johns Hopkins University**, CLSP, Baltimore, USA August 2022 – Present

*Research Assistant* Supervisors: Najim Dehak, Laureano Moro-Velázquez and Jesús Villalba

**Tencent AI Lab**, Speech Group, Bellevue, USA May 2024 - August 2024

*Intern* Supervisors: Meng Yu and Dong Yu

**Amazon**, Amazon General Intelligence (AGI), Baltimore, USA December 2022 - December 2023

*Student Researcher* Supervisors: Venkatesh Ravichandran and Milind Rao

**Microsoft STCA**, NLP Group, Beijing, China February 2022 - May 2022

*Intern* Supervisors: Linjun Shou and Ming Gong

**Tencent AI Lab**, Speech Group, Shenzhen, China May 2020 - November 2021

*Intern* Supervisors: Bo Wu and Chao Weng

**Peking University**, ADSP LAB, Shenzhen, China August 2019 - July 2022

*Research Assistant* Supervisor: Yuexian Zou

**Ubtech Robotics Inc**, Speech Group, Shenzhen, China July 2019 - September 2019

*Intern* Supervisor: Dongyan Huang

**University of California Berkeley**, Berkeley, USA July 2018 - September 2018

*Summer Researcher* Supervisor: Masayoshi Tomizuka

1. **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.
2. **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.
3. 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.
4. Chen Chen, Yuchen Hu, Wen Wu, **Helin Wang**, Eng Siong Chng, Chao Zhang: Enhancing Zero-shot Text-to-Speech Synthesis with Human Feedback. Submitted to ICLR 2025.
5. **Helin Wang**, Jesus Villalba, Laureano Moro-Velazquez, Jiarui Hai, Thomas Thebaud, Najim Dehak: Noise-robust Speech Separation with Fast Generative Correction. INTRESPEECH 2024 (**Oral**). \* Nominated for the **Best Paper Award** and the **Best Student Paper Award** (5 out of 1,030 accepted papers).
6. Jiarui Hai\*, **Helin Wang\***, Dongchao Yang, Karan Thakkar, Najim Dehak, Mounya Elhilali: DPM-TSE: A Diffusion Probabilistic Model for Target Sound Extraction. ICASSP 2024: 1196-1200.
7. Jiarui Hai, Karan Thakkar, **Helin Wang**, Zengyi Qin, Mounya Elhilali: DreamVoice: Text-Guided Voice Conversion. INTERSPEECH 2024: 4373-4377.
8. Maliha Jahan, **Helin Wang**, Thomas Thebaud, Yinglun Sun, Giang Ha Le, Zsuzsanna Fagyal, Odette Scharenborg, Mark Hasegawa-Johnson, Laureano Moro Velazquez, Najim Dehak: Finding Spoken Identifications: Using GPT-4 Annotation for an Efficient and Fast Dataset Creation Pipeline. LREC-COLING 2024: 7296-7306.
9. **Helin Wang**, Venkatesh Ravichandran, Milind Rao, 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**).
10. **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: 1548-1552.
11. Dading Chong\*, **Helin Wang\***, Peilin Zhou, Qingcheng Zeng: Masked Spectrogram Prediction For Self-Supervised Audio Pre-Training. ICASSP 2023: 1-5.
12. Junling Liu, Peilin Zhou, Yining Hua, Dading Chong, Zhongyu Tian, Andrew Liu, **Helin Wang**, Chenyu You, Zhenhua Guo, Lei Zhu, Michael Lingzhi Li: Benchmarking large language models on cmexam-a comprehensive chinese medical exam dataset. NeurIPS 2023.
13. Dongchao Yang, Songxiang Liu, Jianwei Yu, **Helin Wang**, Chao Weng, Yuexian Zou: NoreSpeech: Knowledge Distillation based Conditional Diffusion Model for Noise-robust Expressive TTS. INTERSPEECH 2023: 4798-4802.
14. Dongchao Yang, Jianwei Yu, **Helin Wang**, Wen Wang, Chao Weng, Yuexian Zou, Dong Yu: Diffsound: Discrete diffusion model for text-to-sound generation. IEEE/ACM Transactions on Audio, Speech, and Language Processing 31, 1720-1733.
15. **Helin Wang**, Dongchao Yang, Yuexian Zou, Fan Cui, Yujun Wang: Detect What You Want: Target Sound Detection. DCASE 2022.
16. Dongchao Yang\*, **Helin Wang\***, Wenwu Wang, Yuexian Zou: A Mixed Supervised Learning Framework For Target Sound Detection. DCASE 2022.
17. Dongchao Yang\*, **Helin Wang\***, Zhongjie Ye, Yuexian Zou, Wenwu Wang: RaDur: A Reference-aware and Duration-robust Network for Target Sound Detection. INTERSPEECH 2022: 1511-1515.
18. **Helin Wang**, Dongchao Yang, Chao Weng, Jianwei Yu, Yuexian Zou: Improving Target Sound Extraction with Timestamp Information. INTERSPEECH 2022: 1526-1530.
19. Zhongjie Ye, Yuqing Wang, **Helin Wang**, Dongchao Yang, Yuexian Zou: Featurecut: An adaptive data augmentation for automated audio captioning. APSIPA ASC 2022: 313-318.
20. Peilin Zhou, Dading Chong, **Helin Wang**, Qingcheng Zeng: Calibrate and Refine! A Novel and Agile Framework for ASR-error Robust Intent Detection. INTERSPEECH 2022: 1096-1100.
21. Dongchao Yang, **Helin Wang**, Yuexian Zou, Zhongjie Ye, Wenwu Wang: A Mutual learning framework for Few-shot Sound Event Detection. ICASSP 2022: 811-815.
22. **Helin Wang**, Yuexian Zou, Wenwu Wang: A Global-Local Attention Framework for Weakly Labelled

Audio Tagging. ICASSP 2021: 351-355.

23. **Helin Wang**, Yuexian Zou, Wenwu Wang: SpecAugment++: A Hidden Space Data Augmentation Method for Acoustic Scene Classification. INTERSPEECH 2021: 551-555.
24. **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: 1109-1113.
25. Zhongjie Ye, **Helin Wang**, Dongchao Yang, Yuexian Zou: Improving the Performance of Automated Audio Captioning via Integrating the Acoustic and Textual Information. DCASE2021 Challenge.
26. Dongchao Yang, **Helin Wang**, Zhongjie Ye, Yuexian Zou: Few-shot Bioacoustic Event Detection: A Good Transductive Inference is All You Need. DCASE2021 Challenge.
27. Haoran Zhang, Yuexian Zou, **Helin Wang**: Contrastive self-supervised learning for text-independent speaker verification. ICASSP 2021: 6713-6717.
28. Dongchao Yang, **Helin Wang**, Yuexian Zou: Unsupervised Multi-Target Domain Adaptation for Acoustic Scene Classification. INTERSPEECH 2021: 1159-1163.
29. Zhiqi Huang, Fenglin Liu, Xian Wu, Shen Ge, Helin Wang, Wei Fan, Yuexian Zou: Audio-Oriented Multimodal Machine Comprehension via Dynamic Inter-and Intra-modality Attention. AAAI 2021: 13098-13106.
30. **Helin Wang**, Yuexian Zou, Dading Chong, Wenwu Wang: Modeling Label Dependencies for Audio Tagging With Graph Convolutional Network. IEEE Signal Process. Lett. 27: 1560-1564 (2020).
31. **Helin Wang**, Yuexian Zou, Dading Chong: Acoustic Scene Classification with Spectrogram Processing Strategies. DCASE 2020: 210-214.
32. **Helin Wang**, Yuexian Zou, Dading Chong, Wenwu Wang: Environmental Sound Classification with Parallel Temporal-Spectral Attention. INTERSPEECH 2020: 821-825.
33. **Helin Wang**, Dading Chong, Dongyan Huang, Yuexian Zou: What Affects the Performance of Convolutional Neural Networks for Audio Event Classification. ACII Workshops 2019: 140-146.

## ♡ SERVICES (REVIEWER)

---

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

## ♡ TEACHING

---

2024/02 - 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

## ♡ HONORS AND AWARDS

---

Nominated for the <i>Best Paper Award</i> (5 out of 1,030 accepted papers) in INTERSPEECH	2024
<i>1<sup>st</sup> Team Ranking</i> of DCASE Challenge Task 5 (Judges' award)	2021
<i>4<sup>th</sup> Team Ranking</i> of DCASE Challenge Task 6 (Judges' award)	2021
<i>3<sup>rd</sup> Team Ranking</i> of DCASE Challenge Task 6	2020
<i>Outstanding Graduate Student Award</i> of Peking University	2022
<i>Outstanding Graduate Thesis Award</i> of Peking University	2022
<i>Award for Scientific Research</i> of Peking University	2020-2021
<i>School Prize</i> of Peking University	2019-2020
<i>Merit Student</i> of Peking University	2019-2020