HAO SHI

PhD Candidate, at Speech and Audio Processing Lab, Kyoto University
✉ shi@sap.ist.i.kyoto-u.ac.jp / hshi.cca@gmail.com

RESEARCH INTERESTS

Automatic Speech Recognition:

- Robust ASR (Under noise conditions, with front-end)
- Knowledge distillation for ASR (Model compression)
- Accents-based ASR (Accent adaptation)

Speech Enhancement:

- Spectrograms fusion (Complementary between different systems, frequency-domain models)
- Front-end for robust ASR (Improve both human hearing and recognition performance)

Speech Separation:

- Blind source separation
- Tatget speaker extraction

EDUCATION

Ph.D. in Informatics, Kyoto University, Kyoto, Japan — Apr. 2021 – Present.

- Department of Intelligence Science and Technology, Graduate School of Informatics
- Supervisor: Prof. Tatsuya Kawahara

Master in Computer Science and Technology, Tianjin University, Tianjin, China — Sep. 2018 – Jan. 2021

- College of Intelligence and Computing
- Supervisor: Prof. Longbiao Wang

Research Internships

Research Intern, at NTT (CS Lab @ Keihanna) — Aug. 2023 – Sep. 2023

- Mentor: Naoyuki Kamo, Shoko Araki, Tomohiro Nakatani, Marc Delcroix

Research Intern, at Sony (R&D @ Osaki) — Jan. 2023 – Mar. 2023

- Mentor: Kazuki Shimada, Shusuke Takahashi

Research Assistant, at Tianjin University (Cognitive Computing and Application Lab.) — Aug. 2021 – Jan. 2022

- Mentor: Longbiao Wang, Jianwu Dang

LANGUAGE SKILL

- Chinese (native)
- English (fluent)

PUBLICATIONS

Conference Papers (Reviewed Paper):

- **Hao Shi**, Naoyuki Kamo, Marc Delcroix, Tomohiro Nakatani, and Shoko Araki, "Ensemble Inference for Diffusion Model-based Speech Enhancement", in Proc. of HSCMA, 2024, pp. xx-xx.
- **Hao Shi**, Kazuki Shimada, Masato Hirano, Takashi Shibuya, Yuichiro Koyama, Zhi Zhong, Shusuke Takahashi, Tatsuya Kawahara, and Yuki Mitsufuji, "Diffusion-Based Speech Enhancement with Joint Generative and Predictive Decoders", in Proc. of ICASSP, 2024, pp. xx-xx.
- Yuan Gao, **Hao Shi**, Chenhui Chu, and Tatsuya Kawahara, "Enhancing Two-stage Finetuning for Speech Emotion Recognition Using Adapters", in Proc. of ICASSP, 2024, pp. xx-xx.
- Zhi Zhong, **Hao Shi**, Masato Hirano, Kazuki Shimada, Kazuya Tateishi, Takashi Shibuya, Shusuke Takahashi, and Yuki Mitsufuji, "Extending Audio Masked Autoencoders Toward Audio Restoration", in Proc. of WASPAA, 2023, pp. 1–5.

- **<u>Hao Shi</u>**, Masato Mimura, Longbiao Wang, Jianwu Dang, and Tatsuya Kawahara, "Time-domain Speech Enhancement Assisted by Multi-resolution Frequency Encoder And Decoder," in Proc. of ICASSP, 2023, pp. 1–5.
- Yanbing Yang, **<u>Hao Shi</u>**, Yuqin Lin, Meng Ge, Longbiao Wang, Qingzhi Hou and Jianwu Dang, "Adaptive Attention Network with Domain Adversarial Training for Multi-Accent Speech Recognition," in Proc. of ISCSLP, 2022, pp. 6–10.
- **<u>Hao Shi</u>**, Yuchun Shu, Longbiao Wang, Jianwu Dang, and Tatsuya Kawahara, "Fusing Multiple Bandwidth Spectrograms for Improving Speech Enhancement," in Proc. of APSIPA ASC, 2022, pp. 1935–1940.
- **<u>Hao Shi</u>**, Longbiao Wang, Sheng Li, Jianwu Dang, and Tatsuya Kawahara, "Subband-Based Spectrogram Fusion for Speech Enhancement by Combining Mapping and Masking Approaches," in Proc. of APSIPA ASC, 2022, pp. 286–292.
- **<u>Hao Shi</u>**, Longbiao Wang, Sheng Li, Jianwu Dang, and Tatsuya Kawahara, "Monaural speech enhancement based on spectrogram decomposition for convolutional neural network-sensitive feature extraction," in Proc. of Interspeech, 2022, pp. 221–225.
- Tongtong Song, Qiang Xu, Meng Ge, Longbiao Wang, **<u>Hao Shi</u>**, Yongjie Lv, Yuqin Lin, and Jianwu Dang, "Language-specific Characteristic Assistance for Code-switching Speech Recognition," in Proc. of Interspeech, 2022, pp. 3924–3928. (Corresponding Author)
- Qiang Xu, Tongtong Song, Longbiao Wang, **<u>Hao Shi</u>**, Yuqin Lin, Yongjie Lv, Meng Ge, Qiang Yu, and Jianwu Dang, "Self-Distillation Based on High-level Information Supervision for Compressing End-to-End ASR Model," in Proc. of Interspeech, 2022, pp. 1716–1720. (Corresponding Author)
- **<u>Hao Shi</u>**, Longbiao Wang, Sheng Li, Cunhang Fan, Jianwu Dang, and Tatsuya Kawahara, "Spectrograms Fusion-based End-to-end Robust Automatic Speech Recognition," in Proc. of APSIPA ASC, 2021, pp. 438–442.
- Luya Qiang, **<u>Hao Shi</u>**, Meng Ge, Haoran Yin, Nan Li, Longbiao Wang, Sheng Li, and Jianwu Dang, "Speech Dereverberation Based on Scale-aware Mean Square Error Loss," in Proc. of ICONIP, 2021, pp. 55–63. (Joint first author, equal contribution)
- Haoran Yin, **<u>Hao Shi</u>**, Longbiao Wang, Luya Qiang, Sheng Li, Meng Ge, Gaoyan Zhang, and Jianwu Dang, "Simultaneous Progressive Filtering-based Monaural Speech Enhancement," in Proc. of ICONIP, 2021, pp. 213–221. (Joint first author, equal contribution)
- **<u>Hao Shi</u>**, Longbiao Wang, Meng Ge, Sheng Li, and Jianwu Dang, "Spectrograms Fusion with Minimum Difference Masks Estimation for Monaural Speech Dereverberation," in Proc. of ICASSP, 2020, pp. 7544-7548.
- **<u>Hao Shi</u>**, Longbiao Wang, Sheng Li, Chenchen Ding, Meng Ge, Nan Li, Jianwu Dang, and Hiroshi Seki, "Singing Voice Extraction with Attention based Spectrograms Fusion," in Proc. of Interspeech, 2020, pp. 2412-2416.
- Meng Ge, Longbiao Wang, Nan Li, **<u>Hao Shi</u>**, Jianwu Dang, and Xiangang Li, "Environment-dependent attention-driven recurrent convolutional neural network for robust speech enhancement," in Proc. of Interspeech, 2019, pp. 3153-3157.