HAO SHI

Researcher, at Speech and Audio Processing Lab, Kyoto University
✉ shi@sap.ist.i.kyoto-u.ac.jp / hshi.cca@gmail.com
Personal Website: `https://hshi-speech.github.io`

RESEARCH INTERESTS

Automatic Speech Recognition:
- Noise-robust
- Adaptation
- Multi-speaker
- Knowledge distillation

Speech Enhancement:
- Front-end for robust ASR
- Ensemble of complementary systems
- Probabilistic model
- Multi-model

Speech Separation:
- Target speaker extraction
- Blind source separation

EDUCATION

Ph.D. in Informatics, Kyoto University, Kyoto, Japan — 2021 – 2024
- Department of Intelligence Science and Technology, Graduate School of Informatics
- Supervisor: Prof. Tatsuya Kawahara

Master in Computer Science and Technology, Tianjin University, Tianjin, China — 2018 – 2021
- College of Intelligence and Computing
- Supervisor: Prof. Longbiao Wang

Working Experiences

Researcher, at Kyoto University — 10. 2024 – Present.

Research Fellow, at Kyoto University — 04. 2024 – 09. 2024

Research Intern, at NTT (CS Lab @ Keihanna) — 08. 2023 – 09. 2023

Research Intern, at Sony (R&D @ Osaki) — 01. 2023 – 03. 2023

Research Assistant, at Tianjin University — 08. 2021 – 01. 2022

HONORS

Fellowship, awarded by Japan Science and Technology Agency (JST) — 04. 2022 – 03. 2024

LANGUAGE SKILL

• Chinese (native) • English (fluent)

REVIEWER

• IEEE/ACM Trans. ASLP • Speech Communication • IEEE-ICASSP • INTERSPEECH • APSIPA ASC • SLT • WASPAA

PUBLICATIONS

Journal Papers (Reviewed):

- **Hao Shi**, Masato Mimura, Tatsuya Kawahara, "Time-domain Speech Enhancement Using Spectrogram Encoding for Robust Speech Recognition", IEEE/ACM Trans. Audio, Speech and Language Process, Vol.32, pp.3049–3060, 2024.

Conference Papers (Reviewed):

- **Hao Shi**, Yuan Gao, Zhaoheng Ni, Tatsuya Kawahara, "Serialized Speech Information Guidence with Overlapped Encoding Separation for Multi-Speaker Automatic Speech Recognition," in Proc. IEEE-SLT (Accepted).
- **Hao Shi**, Tatsuya Kawahara, "Dual-path Adaptation of Pretrained Feature Extraction Module for Robust Automatic Speech Recognition", in Proc. INTERSPEECH, 2024, pp.2850-2854.
- Yuan Gao, **Hao Shi**, Chenhui Chu, Tatsuya Kawahara, "Speech Emotion Recognition with Multi-level Acoustic and Semantic Information Extraction and Interaction", in Proc. INTERSPEECH, 2024, pp.1060-1064.
- Yuchun Shu, Bo Hu, Yifeng He, **Hao Shi**, Longbiao Wang, Jianwu Dang, "Error Correction by Paying Attention to Both Acoustic and Confidence References for Automatic Speech Recognition", in Proc. INTERSPEECH, 2024, pp.3500-3504.
- **Hao Shi**, Naoyuki Kamo, Marc Delcroix, Tomohiro Nakatani, and Shoko Araki, "Ensemble Inference for Diffusion Model-based Speech Enhancement", in Proc. IEEE-ICASSPW, 2024, pp.735-739.
- **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. IEEE-ICASSP, 2024, pp.12951–12955.
- Yuan Gao, **Hao Shi**, Chenhui Chu, and Tatsuya Kawahara, "Enhancing Two-stage Finetuning for Speech Emotion Recognition Using Adapters", in Proc. IEEE-ICASSP, 2024, pp.11316–11320.
- 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. WASPAA, 2023, pp.1–5.
- **Hao Shi**, Masato Mimura, Longbiao Wang, Jianwu Dang, and Tatsuya Kawahara, "Time-domain Speech Enhancement Assisted by Multi-resolution Frequency Encoder And Decoder," in Proc. IEEE-ICASSP, 2023, pp.1–5.
- Yanbing Yang, **Hao Shi**, 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. ISCSLP, 2022, pp.6–10.
- **Hao Shi**, Yuchun Shu, Longbiao Wang, Jianwu Dang, and Tatsuya Kawahara, "Fusing Multiple Bandwidth Spectrograms for Improving Speech Enhancement," in Proc. APSIPA ASC, 2022, pp.1935–1940.
- **Hao Shi**, Longbiao Wang, Sheng Li, Jianwu Dang, and Tatsuya Kawahara, "Subband-Based Spectrogram Fusion for Speech Enhancement by Combining Mapping and Masking Approaches," in Proc. APSIPA ASC, 2022, pp.286–292.
- **Hao Shi**, 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. INTERSPEECH, 2022, pp.221–225.
- Tongtong Song, Qiang Xu, Meng Ge, Longbiao Wang, **Hao Shi**, Yongjie Lv, Yuqin Lin, and Jianwu Dang, "Language-specific Characteristic Assistance for Code-switching Speech Recognition," in Proc. INTERSPEECH, 2022, pp.3924–3928.
- Qiang Xu, Tongtong Song, Longbiao Wang, **Hao Shi**, 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. INTERSPEECH, 2022, pp.1716–1720.
- **Hao Shi**, Longbiao Wang, Sheng Li, Cunhang Fan, Jianwu Dang, and Tatsuya Kawahara, "Spectrograms Fusion-based End-to-end Robust Automatic Speech Recognition," in Proc. APSIPA ASC, 2021, pp.438–442.
- Luya Qiang, **Hao Shi**, 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. ICONIP, 2021, pp.55–63.
- Haoran Yin, **Hao Shi**, Longbiao Wang, Luya Qiang, Sheng Li, Meng Ge, Gaoyan Zhang, and Jianwu Dang, "Simultaneous Progressive Filtering-based Monaural Speech Enhancement," in Proc. ICONIP, 2021, pp.213–221.
- **Hao Shi**, Longbiao Wang, Meng Ge, Sheng Li, and Jianwu Dang, "Spectrograms Fusion with Minimum Difference Masks Estimation for Monaural Speech Dereverberation," in Proc. IEEE-ICASSP, 2020, pp.7544-7548.
- **Hao Shi**, 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. INTERSPEECH, 2020, pp.2412-2416.
- Meng Ge, Longbiao Wang, Nan Li, **Hao Shi**, Jianwu Dang, and Xiangang Li, "Environment-dependent attention-driven recurrent convolutional neural network for robust speech enhancement," in Proc. INTERSPEECH, 2019, pp.3153-3157.

Reports:

- **Hao Shi**, and Tatsuya Kawahara, "Investigation of Adapter for Automatic Speech Recognition in Noisy Environment", in SIG Technical Reports, 2023, pp.1–6.