HAO SHI

Research Scientist, at SB Intuitions, Tokyo, Japan
✉ hao.shi@sbintuitions.co.jp / hshi@ieee.org
Personal Website: https://sites.google.com/view/hshi-speech

EDUCATION

Ph.D. in Informatics Apr. 2021 - Sep. 2024
Department of Intelligence Science and Technology, Graduate School of Informatics
Kyoto University, Kyoto, Japan

Master in Computer Science and Technology Sep. 2018 - Jan. 2021
College of Intelligence and Computing
Tianjin University, Tianjin, China

B.Sc. in Computer Science and Technology Sep. 2014 - Jun. 2018
The School of Information Science and Technology
Southwest Jiaotong University, Sichuan, China

RESEARCH INTERESTS

Speech to Speech:
End-to-end, Pipeline, Adaptation

Speech Enhancement:
Front-end for ASR, Systems ensemble, Probabilistic model

Automatic Speech Recognition:
Noise-robust, Adaptation, Multi-speaker, Multi-lingual

WORKING EXPERIENCES

Research Scientist, at SB Intuitions, SoftBank, Tokyo, Japan Apr. 2025 - Present
Main Topic: Speech-to-Speech

Researcher, at Kyoto University, Kyoto, Japan Oct. 2024 - Mar. 2025
Main Topic: Multi-talker Robotics

Research Fellow, at Kyoto University, Kyoto, Japan Apr. 2024 - Sep. 2024
Main Topic: Noise-robust ASR

Research Intern, at NTT (CS Lab @ Keihanna), Kyoto, Japan Aug. 2023 - Sep. 2023
Main Topic: Systems fusion and diffusion model for speech enhancement

Research Intern, at Sony (R&D @ Osaki), Tokyo, Japan Jan. 2023 - Mar. 2023
Main Topic: Diffusion model for speech enhancement

HONORS

Fellowship, awarded by Japan Science and Technology Agency (JST) Apr. 2022 - Mar. 2024

SKILLS

Language skill Chinese (native), English (fluent)
Programming skills Python, C++, shell, matlab

ACADEMIC ACTIVITIES

Reviewer IEEE Trans. ASLP, Speech Communication
IEEE-ICASSP, INTERSPEECH, APSIPA ASC, SLT, WASPAA, IJCNN
Organizer Automatic Song Aesthetics Evaluation Challenge @ ICASSP 2026

PUBLICATIONS

Journal Papers (Reviewed)

30. Yuan Gao, Hao Shi, Yahui Fu, Chenhui Chu, and Tatsuya Kawahara, "Bridging Speech Emotion Recognition and Personality: Dataset and Temporal Interaction Condition Network," IEEE Transactions on Affective Computing, 2025 (accepted).
29. Hao Shi, Xugang Lu, Kazuki Shimada, and Tatsuya Kawahara, "Combining Deterministic and Diffusion Model to Meet the Partially Stochastic Function Property of Speech Enhancement," IEEE Trans. Audio, Speech and Language Process, Vol.33, pp.4253-4266, 2025.
28. Yuan Gao, Hao Shi, Chenhui Chu, and Tatsuya Kawahara, "Multi-Attribute Learning for Multi-Level Emotion Recognition from Speech," APSIPA Transactions on Signal and Information Processing, Vol.14, No. e20, pp.1–29, 2025.
27. Hao Shi, Masato Mimura, and Tatsuya Kawahara, "Waveform-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)

26. Atsushi Kojima, Yusuke Fujita, Hao Shi, Tomoya Mizumoto, Mengjie Zhao, Yui Sudo, "Conversation Context-aware Direct Preference Optimization for Style-Controlled Speech Synthesis," in Proc. APSIPA ASC, 2025, 573-578.
25. Hao Shi, Yusuke Fujita, Tomoya Mizumoto, Lianbo Liu, Atsushi Kojima, and Yui Sudo, "Serialized Output Prompting for Large Language Model-based Multi-Talker Speech Recognition," in Proc. IEEE-ASRU, 2025 (Accepted).
24. Tomoya Mizumoto, Yusuke Fujita, Hao Shi, Lianbo Liu, Atsushi Kojima, and Yui Sudo, "Evaluating Japanese Dialect Robustness across Speech and Text-based Large Language Models," in Proc. IEEE-ASRU, 2025 (Accepted).
23. Jiahui Zhao, Hao Shi, Tianrui Wang, Hexin Liu, Zhaoheng Ni, Lingxuan Ye, and Longbiao Wang, "Adapting Pretrained Speech Recognition Models for Code-Switching through Encoding Refining and Language-Aware Attention-based Decoding," in Proc. IEEE-ICASSP, 2025.
22. Zhongjian Cui, Chenrui Cui, Tianrui Wang, Mengnan He, Hao Shi, Meng Ge, Caixia Gong, Longbiao Wang, and Jianwu Dang, "Reducing the Gap between Pretrained Speech Enhancement and Recognition Models Using a Real Speech-Trained Bridging Module," in Proc. IEEE-ICASSP, 2025.
21. Hao Shi, Yuan Gao, Zhaoheng Ni, and Tatsuya Kawahara, "Serialized Speech Information Guidence with Overlapped Encoding Separation for Multi-Speaker Automatic Speech Recognition," in Proc. IEEE-SLT, 2024, pp.193–199.
20. Hao Shi, and Tatsuya Kawahara, "Dual-path Adaptation of Pretrained Feature Extraction Module for Robust Automatic Speech Recognition," in Proc. INTERSPEECH, 2024, pp.2850–2854.
19. Yuan Gao, Hao Shi, Chenhui Chu, and Tatsuya Kawahara, "Speech Emotion Recognition with Multi-level Acoustic and Semantic Information Extraction and Interaction," in Proc. INTERSPEECH, 2024, pp.1060–1064.
18. Yuchun Shu, Bo Hu, Yifeng He, Hao Shi, Longbiao Wang, and Jianwu Dang, "Error Correction by Paying Attention to Both Acoustic and Confidence References for Automatic Speech Recognition," in Proc. INTERSPEECH, 2024, pp.3500–3504.
17. 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.
16. 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.
15. 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.
14. 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.
13. 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.
12. 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.

11. **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.
10. **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.
9. **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.
8. 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.
7. 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.
6. **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.
5. 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.
4. 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.
3. **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.
2. **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.
1. 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.

Under Review

- **Hao Shi**, Yusuke Fujita, Mengjie Zhao, Tomoya Mizumoto, and Yui Sudo, "CTC-based Multi-talker Speech Recognition with Talker Counting-based Branch Selection," (Submitted to IEEE-ICASSP 2026).
- Mengjie Zhao, **Hao Shi**, Yusuke Fujita, and Yui Sudo, "S2S-ja: Datasets, Benchmarks, and Baselines for Japanese Speech-to-Speech Systems," (Submitted to IEEE-ICASSP 2026).
- Reo Yoneyama, Yusuke Fujita, Haesung Jeon, Lianbo Liu, Atsushi Kojima, **Hao Shi**, Mengjie Zhao, and Yui Sudo, "Balancing Acoustic Modeling and Knowledge Retention in LLM-Based Prompt-TTS via RVQ Granularity Control," (Submitted to IEEE-ICASSP 2026).

Reports and Pre-print

- **Hao Shi**, and Tatsuya Kawahara, "Investigation of Adapter for Automatic Speech Recognition in Noisy Environment," in SIG Technical Reports, 2023, pp.1–6.
- Tongtong Song, Qiang Xu, Haoyu Lu, Longbiao Wang, **Hao Shi**, Yuqin Lin, Yanbing Yang, Jianwu Dang, "Monolingual Recognizers Fusion for Code-switching Speech Recognition," arXiv preprint arXiv:2211.01046, 2022.