

Yiwen Shao

Tel: 410-960-1058 | Email: yshao18@jhu.edu

EDUCATION

Johns Hopkins University, Baltimore, USA

Aug.2017 - **Present**

Whiting School of Engineering

- MSE in Computer Science
- Ph.D in Computer Science

Southeast University, Nanjing, China

Sep.2013 - Jun.2017

School of Electronic Science and Engineering

- B.Eng in Internet of Things Engineering
- GPA 3.81/4 RANK: 1/30

PUBLICATIONS

- [1] **Y Shao**, Q Lin, "Use of Pitch Continuity for Robust Speech Activity Detection", **ICASSP 2018**.
- [2] Q Lin, **Y Shao**, "A Novel Normalization Method for Autocorrelation Function for Pitch Detection and for Speech Activity Detection", **Interspeech 2018**.
- [3] A Arora, CC Chang, B Rekadbar, B BabaAli, D Povey, D Etter, D Raj, H Hadian, J Trmal, P Garcia, S Watanabe, V Manohar, **Y Shao**, S Khudanpur, "Using ASR methods for OCR", **ICDAR 2019**.
- [4] Y Wang, T Chen, H Xu, S Ding, H Lv, **Y Shao**, N Peng, L Xie, S Watanabe, S Khudanpur, "Espresso: A fast end-to-end neural speech recognition toolkit", **ASRU 2019**.
- [5] Z Huang, S Watanabe, Y Fujita, P García, **Y Shao**, D Povey, S Khudanpur, "Speaker diarization with region proposal network", **ICASSP 2020**.
- [6] **Y Shao**, Y Wang, D Povey, S Khudanpur, "PyChain: A Fully Parallelized PyTorch Implementation of LF-MMI for End-to-End ASR", **Interspeech 2020**.
- [7] P Żelasko, S Joshi, **Y Shao**, J Villalba, J Trmal, N Dehak, S Khudanpur, "Adversarial Attacks and Defenses for Speech Recognition Systems", **arXiv 2021** preprint arXiv:2103.17122.
- [8] **Y Shao***, SX Zhang*, D Yu, "Multi-Channel Multi-Speaker ASR Using 3D Spatial Feature". **ICASSP 2022**.
- [9] **Y Shao**, J Villalba, S Joshi, S Kataria, N Dehak, S Khudanpur, "Chunking Defense for Adversarial Attacks on ASR", **Interspeech 2022**.
- [10] S Joshi, S Kataria, **Y Shao**, P Zelasko, J Villalba, S Khudanpur, N Dehak, "Defense against adversarial attacks on hybrid speech recognition using joint adversarial fine-tuning with denoiser", **Interspeech 2022**.
- [11] **Y Shao**, "Challenges and Insights: Exploring 3D Spatial Features and Complex Networks on the MISP Dataset", **arXiv 2023** preprint arXiv:2310.03901.
- [12] **Y Shao**, SX Zhang, D Yu, "RIR-SF: Room Impulse Response based spatial feature for multi-channel multi-talker ASR", **arXiv 2023** preprint arXiv:2311.00146.
- [13] Z Huang, **Y Shao**, SX Zhang, D Yu, "UNIX-encoder: a universal x-channel speech encoder for ad-hoc microphone array speech processing", **ICASSP 2024**.

RESEARCH

CLSP, JHU, Baltimore, MD

Aug.2017 - **Present**

Research Assistant, advised by Daniel Povey, and Sanjeev Khudanpur

MCMS ASR

- Multi-channel Multi-speaker ASR based on spatial information

GARD

- Adversarial defense/attack on ASR systems
- IBM [adversarial-robustness-toolbox](#) speech recognition modules contribution

PyChain <https://github.com/YiwenShaoStephen/pychain>

- Implemented Lattice-free MMI loss for ASR in PyTorch

Natural Gradient <https://github.com/YiwenShaoStephen/NGD-SGD>

- Implemented a modified version of natural gradient in PyTorch.

Espresso <https://github.com/freewym/espresso>

- A toolkit for end-to-end ASR based on fairseq.

Auxiliary Augmentation <https://github.com/YiwenShaoStephen/Auxiliary-Augmentation>

- A self-supervised data-augmentation method that adds an auxiliary label of the permutation (e.g. rotation) done during training.

Waldo <https://github.com/waldo-seg/waldo> (extended: <https://github.com/YiwenShaoStephen/mergeNet>)

A toolkit for text localization and instance segmentation in PyTorch using a new non-detection method.

Tencent AI Lab, Seattle, USA

Research Intern, advised by Shi-Xiong Zhang and Dong Yu

- Developed 3D spatial features and an All-in-one model for multi-channel multi-speaker ASR. May.2021- Aug.2021
 - Tested the All-in-one model on real and simulated data with strong reverberation. June.2022 - Aug.2022
 - Extended 3D spatial feature to RIR-based spatial features for multi-channel multi-speaker ASR. May.2023- Aug.2023
 - Contributed to the development of a universal x-channel encoder for multi-channel speech.
-

Baihu Co., Ltd, Wuxi, China

Jun.2016 - Aug.2017

Speech Processing & Software Engineering Intern, advised by Qiguang Lin

- Implemented a Speech Activity Detection (SAD) system in C++/Java based on the Combo-SAD for feature extraction and the Ring Processing and SVMs for classification.
 - Proposed a novel algorithm that utilizes pitch continuity to improve SAD performance
 - Proposed exponential ACF that gets better results on both SAD and Pitch Detection.
-

TEACHING

- **CA**, Natural Language Processing
 - **CA**, Information Extraction
 - **TA**, Information Extraction
-

SKILLS

- **Languages:** Python, Shell, C/C++, Java, Perl
- **Deep Learning:** Pytorch, Tensorflow, Kaldi, K2