

Personal Particulars

Name: Pan Zexu **Phone:** +65 98945592
Gender: Male **Email:** pan_zexu@u.nus.edu / xu325504@hotmail.com
Nationality: Chinese **Address:** 198 Pasir Panjang Road, #01-08, Singapore 117526



Research Interests

I am interested in artificial intelligence, deep learning, and their applications on but not limited to speech processing, such as multi-modal speech enhancement, speaker extraction, speaker diarization, robust automatic speech recognition, multi-modal representation learning, auditory attention detection, etc.

Education

Ph.D. Degree: National University of Singapore (NUS), Institute of Data Science Aug 2019 - Jul 2023
Bachelor's Degree: Nanyang Technological University (NTU), (GPA 4.87/5.0) Aug 2015 - Jul 2019
University of Maryland, USA. (Exchange program) Jul 2016 - Dec 2016

Work Experiences

Visiting research scientist @ Mitsubishi Electric Research Laboratories, USA. (MERL) May 2023-now
Speech and audio processing, with a focus on audio-visual and neuro-steered speaker extraction

Visiting Researcher @ Cognitive system lab, University of Bremen, Germany Mar 2023-Apr 2023
Research on Neuro-steered speaker extraction with EEG signals and more effective online deployment

Research internship @ Mitsubishi Electric Research Laboratories, USA. (MERL) Aug 2022-Dec 2022
Derive and implement new models and optimization methods for Audio-visual speaker diarization for in-the-wild video recordings.

Intern @ Delta-NTU Cyber Physical Security Laboratory Jan 2018-Jun 2018
Designing control and automation algorithm for robot arms

Student researcher @ NTU Internet of things lab May 2017- Jun 2018
Data acquisition and analysis with Millimeter-wave radar for obstacle/target detection and tracking. Designing autonomous obstacle bypass, target tracking, and pursuit algorithm for unmanned aerial vehicle.

Journal Publications

- [1] **Z. Pan**, R. Tao, C. Xu, and H. Li, "Selective Listening by Synchronizing Speech with Lips," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 30, pp. 1650-1664, 2022.
- [2] **Z. Pan**, M. Ge, and H. Li, "USEV: Universal Speaker Extraction with Visual Cue," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 30, pp. 3032-3045, 2022
- [3] **Z. Pan**, X. Qian, and H. Li, "Target Speaker Extraction with Co-speech Gestures Cue," IEEE Signal Processing Letter, vol. 29, pp. 1467-1471, 2022.
- [4] T. Wang, **Z. Pan**, M. Ge, Z. Yang, and H. Li, "Time-domain Speech Separation Networks with Graph Encoding Auxiliary," IEEE Signal Processing Letter, vol. 30, pp. 110-114, 2023.

Conference & Workshop Publications

- [1] **Z. Pan**, G. Wichern, F. Germain, A. Subramanian, and J. Le Roux, "Towards End-to-end Speaker Diarization in the Wild," submitted to ASRU 2023.
- [2] **Z. Pan**, W. Wang, M. Borsdorf, and H. Li, "ImagineNET: Target Speaker Extraction with Intermittent Visual Cue through Embedding Inpainting," in Proc. ICASSP, 2023.
- [3] **Z. Pan**, M. Ge, and H. Li, "A Hybrid Continuity Loss to Reduce Over-Suppression for Time-domain Target Speaker Extraction," in Proc. INTERSPEECH, 2022, pp. 1786-1790.
- [4] **Z. Pan**, R. Tao, C. Xu, and H. Li, "Muse: Multi-modal target speaker extraction with visual cues," in Proc. ICASSP, 2021, pp. 6678-6682.

- [5] **Z. Pan**, Z. Luo, J. Yang, and H. Li, "Multi-Modal Attention for Speech Emotion Recognition," in Proc. INTERSPEECH, 2020, pp. 364–368.
- [6] Y. Jiang, R. Tao, **Z. Pan**, and H. Li, "Target Active Speaker Detection with Audio-visual Cues," in Proc. INTERSPEECH, 2023.
- [7] K. Zhang, M. Borsdorf, **Z. Pan**, H. Li, Y. Wei, and Y. Wang, "Speaker Extraction with Detection of Presence and Absence of Target Speakers," in Proc. INTERSPEECH, 2023.
- [8] J. Li, M. Ge, **Z. Pan**, R. Cao, L. Wang, J. Dang, and S. Zhang, "Rethinking the visual cues in audio-visual speaker extraction," in Proc. INTERSPEECH, 2023.
- [9] J. Li, M. Ge, **Z. Pan**, L. Wang, and J. Dang, "VCSE: Time-domain visual-contextual speaker extraction network," in Proc. INTERSPEECH, 2022, pp. 906-910.
- [10] R. Tao, **Z. Pan**, RK Das, Z. Shou, H. Li "Is Someone Speaking? Exploring Long-term Temporal Features for Audio-visual Active Speaker Detection." ACM MM, 2021, pp. 3927-3935
- [11] X. Qian, M. Madhavi, **Z. Pan**, J. Wang and H. Li, "Multi-Target DoA Estimation with an Audio-Visual Fusion Mechanism," in Proc. ICASSP 2021, pp. 4280-4284.

Awards

- | | |
|---|-------------------|
| 1. 2nd Runner up for ActivityNet Challenge at CVPR workshop | 2021 |
| 2. NUS Integrative Science and Engineering Scholarship | Aug 2019-Jul 2023 |
| 3. NTU Science and Engineering Undergraduate Scholarship | Aug 2015-Jul 2019 |
| 4. Ministry of Education Senior Middle 1 Scholarship | Jan 2011-Dec 2014 |

Teaching Experiences

- | | |
|---|-------------------|
| <u>NUS Graduate Teaching Assistant</u> | Jan 2020-Jul 2022 |
| EE2211 Introduction to Machine Learning: Tutor for tutorials. | |
| EE2026 Digital Design: Tutor for Verilog programming. | |
| <u>Primary school teacher</u> | Nov 2014-Jul 2015 |
| Voluntary teaching in an impoverished countryside school in Anhui Province for 1 semester | |
| Voluntary teaching in an impoverished countryside school in Henan Province for 1 semester | |