Yusheng Dai

Hefei, Anhui, China
dalisondys@gmail.com Google Scholar Home Page

EDUCATION

University of Science and Technology of China (985 project) **Sep. 2022 – May 2025**
Master of Information Science and Technology ***GPA: 3.68*** *Hefei, China*

Sichuan University (985 project) **Sep. 2018 – June 2022**
Bachelor of Cyber Science and Engineering ***GPA: 3.78 Ranking: 4/172*** *Chengdu, China*

EXPERIENCE

Audio-Visual Speech Recognition (AVSR) Jan. 2022 – Present

Graduate Research Supervised by Prof. Jun Du and Prof. Chin-hui Lee

- My primary research focuses on leveraging visual cues (e.g., lip movement and facial expressions) to enhance speech applications in adverse acoustic environments. In early work, we propose a novel pre-training method that correlates lip shapes with syllable-level subword units to alleviate the discrepancies between audio and video inputs based on end-to-end AVSR training framework [2].
- In recent work, we explore dropout-induced modality bias on robustness to missing video frames for AVSR. We propose a novel Multimodal Distribution Approximation with Knowledge Distillation (MDA-KD) and Modality-Specific Adapter (MS-Adapter) to maintain performance and robustness simultaneously [1].

Video to Audio & Controlable Audio Generation Dec. 2023 – Present

Graduate Research Supervised by Prof. Jun Du

- We focus on researching controllable audio generation based on given silent video (like Sora), text, and customized format prompts. Semantic alignment and temporal alignment cross modality are our primary concerns.

Financial Data Movement Prediction March 2022 – Jan. 2024

Self-driven Graduate Research

- Researching financial data movement prediction has long been my interest. We formalize the problem within a standard meta-learning framework to tackle challenges including limited data availability and domain shift.
- We've developed MASSER, leveraging self-supervised learning and meta-learning for both offline and online day-trading scenarios. It achieves top performance on benchmark datasets and successful in real-world testing [4][5].

Bird Sound Recognition in Complex Acoustic Environments March 2020 – March 2021

Undergraduate Research Supervied by Prof. Jin Yang

- This project focuses on applying a blind source separation method to identify all foreground bird species within overlapping vocalization recordings, such as those found in a bird dawn chorus.
- Algorithms include Independent Vector Analysis (IVA), spectrogram recognition, multi-channel signal simulation. We finally develop a visual and simulation platform for ecological surveillance and further research [3].

Holding MISP Challenge 2021 - 2023 Dec. 2021 – Dec. 2023

Graduate Research Supervised by Prof. Jun Du

- The MISP challenges 2021, 2022, and 2023 have been successfully held as the grand challenges of ICASSP in recent three years. The three challenges focus on speech recognition, enhancement, and diarization tasks respectively. As a team member, I am responsible for building baseline systems [6][8].

Selected PUBLICATIONS

[1] **Yusheng Dai**, Hang Chen, Jun Du, Chin-Hui Lee, et.al. A Study of Dropout-Induced Modality Bias on Robustness to Missing Video Frames for Audio-Visual Speech Recognition. *IEEE / CVF Computer Vision and Pattern Recognition Conference (**CVPR**)*, 2024.

[2] **Yusheng Dai**, Hang Chen, Jun Du, Chin-hui Lee, et.al. Improving Audio-Visual Speech Recognition by Lip-Subword Correlation Based Visual Pre-training and Cross-Modal Fusion Encoder. *IEEE International Conference on Multimedia and Expo (**ICME**)*, 2023 as **oral**.

[3] **Yusheng Dai**, Yang Jin, Yiwei Dong et.al. Blind source separation-based IVA-Xception model for bird sound recognition in complex acoustic environments. ***Electronics Letters***.

[4] Dongli Zhan*, **Yusheng Dai***, et.al. Meta-Adaptive Stock Movement Prediction with Two-Stage Representation Learning. *NeurIPS Workshop on Distribution Shifts (**NeurIPS**)*, 2022. (* means equal contribution)

[5] Dongli Zhan*, **Yusheng Dai***, et.al. Meta-Adaptive Stock Movement Prediction with Two-Stage Representation Learning. *SIAM International Conference on Data Mining (**SDM**)*, 2024. (* means equal contribution)

[6] Hang chen, Jun Du, **Yusheng Dai**, et.al. Audio-Visual Speech Recognition in MISP2021 Challenge: Dataset Release and Deep Analysis. *In Proceedings of the Annual Conference of the International Speech Communication Association (**Interspeech**)*, 2022.

[7] Haotian Wang, Jun Du, **Yusheng Dai**, et.al. Improving Multi-Modal Emotion Recognition Using Entropy-Based Fusion and Pruning-Based Network Architecture Optimization. *IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, 2024.

[8] Shilong Wu, Chenxi Wang, Hang Chen, **Yusheng Dai**, et.al. The Multimodal Information Based Speech Processing (MISP) 2023 Challenge: Audio-Visual Target Speaker Extraction. *IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, 2024.

AWARDS

- CHIME 7 World Champion Group Member (Defeated NTT, NVIDIA, Cambridge, etc.) 2023
- Outstanding Graduate of Sichuan University 2022
- Distinguished Undergraduate Thesis of Sichuan University 2022
- National First Prize in the China College Students' Computer Design Competition 2021
- National-level Award in the College Students' Innovative Entrepreneurial Training Plan Program 2021
- Sichuan University Speech and Intersection Association Vice President 2021-2022
- Individual First-Class Scholarship for Sichuan University 2019-2021
- Student of Wu Yuzhang Honors College, Sichuan University 2018-2022
- Best Debater, Sichuan University College Union Debate Competition 2019

SKILLS

Languages	English: Advanced (IELTS 7.0), Mandarin: Native.
Coding	Python, Java, C, SQL, HTML, CSS, JavaScript, MATLAB.
Sports	Skiing, Badminton, Basketball, Frisbee.
Misc.	Hosting, Debating, Folk Guitar, Portrait Photography, Video Editing.