

YUSHENG DAI

Chengdu, Sichuan, China

✉ dalison@mail.ustc.edu.cn  Google Scholar  Homepage

EDUCATION

University of Science and Technology of China (985 project)

Sep. 2022 – July 2025

Master of Information Science and Technology || GPA: 3.68/4.3 Score: 86.77/100

Hefei, China

National Scholarship (2023 - 2024) & IELTS, Overall 7.0 (L 7.0, R 8.0, W 6.0, S 6.0)

Sichuan University (985 project)

Sep. 2018 – June 2022

Bachelor of Cyber Science and Engineering || GPA: 3.78/4.0 Score: 89.35/100

Chengdu, China

Certificate of Honor (Top-2%) & Distinguished Undergraduate Thesis (Top-4%)

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] Dongli Zhan*, **Yusheng Dai***, et.al. Meta-Adaptive Stock Movement Prediction with Two-Stage Representation Learning. *NeurIPS Workshop on Distribution Shifts (NeurIPS)*, 2022, and at SIAM International Conference on Data Mining (**SDM**), 2024. (* means equal contribution)
- [4] **Yusheng Dai**, Yang Jin, et.al. Blind source separation-based IVA-Xception model for bird sound recognition in complex acoustic environments. *Electronics Letters (SCI)*, 2021.
- [5] 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.
- [6] Haotian Wang, Jun Du, **Yusheng Dai**, et.al. Improving Multi-Modal Emotion Recognition Using Entropy-Based Fusion and Pruning-Based Network Architecture Optimization. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.
- [7] Hang Chen, Shilong Wu, **Yusheng Dai**, et.al. Summary on the MSIP 2022 Challenge. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.
- [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. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.

RESEARCH & INTERN EXPERIENCE

Audio-Visual Speech Recognition

Sep. 2022 – Feb. 2024

The CAS National Pioneer Project (Team Leader)

Advisor: Prof. Jun Du & Prof. Chin-hui Lee (Georgia Tech)

- In this program, we use talking face videos to enhance speech assistants, particularly in smart home and in-car scenarios with multi-speakers. In the early work, we successfully alleviate disparities in convergence and representation across modalities in multimodal training by introducing: (1) A two-stage modality decoupled training framework, and (2) A visual pre-training that correlates lip shapes with syllable units [2].
- In recent work, we try to enhance video frame missing robustness by exploring an excessive modality bias caused by data dropout technique. To maintain performance and robustness simultaneously, we propose the novel Multimodal Distribution Approximation framework and Modality Specific Adapter for partially and totally missing situations, achieving SOTA performance in both the MISP2021 and MISP2022 Challenges [1].

Video to Audio & Text to Audio Generation

Feb. 2024 – Present

The Pioneer Research Project of IFLYTEK (Team Leader)

Advisor: Prof. Jun Du & Prof. Chin-hui Lee

- I am leading a research group in IFLYTEK focusing on video and audio generation in movies and we-media products. Currently, we have achieved a cascade and controllable sound effects generation system for silent videos, generated either by AI or humans. The following two first-author papers are in progress.
- Firstly, we aim to improve the cross-modality perceptual abilities of MLLMs that use a visual prior to generate descriptions of other modalities like audio and touch. In addition, we enhance generative models' prompt robustness and zero-shot generation capabilities by Retrieval Augmentation Generation (RAG).
- Secondly, we implement the atomic controllability in temporal and semantic aspects with flexibility in time length for T2A task. We open-source a new universal and unbiased benchmark with temporally-strong labels.

Financial Data Movement Prediction 🔄

March 2022 – Jan. 2024

Self-interest Driven Research (Main Contribution)

Advisor: Prof. James Anderson (Columbia)

- 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 is successful in real-world testing [4].

CHALLENGE EXPERIENCE

Holding MISP 2021 - 2023 Challenges 🔄🔄🔄

Dec. 2021 – Dec. 2023

ICASSP Signal Processing Grand Challenges (Main contribution)

Advisor: Prof. Jun Du & Prof. Chin-hui Lee

- We have held the MISP 2021, 2022, and 2023 challenges as the grand challenges of ICASSP in recent three years. The three challenges focus on speech recognition, enhancement, and diarization tasks respectively. As a core member, I was responsible for dataset collecting, building baseline systems and leaderboard website [5][7][8].

CHIME7 Challenge World Championship

Feb. 2022 – July. 2022

Interspeech Signal Processing Grand Challenges (Main contribution)

Advisor: Dr. Lei Sun (iFlyTek)

- For the CHIME7 competition, the challenge focuses on multi-channel multi-speaker in home scenarios. My main work involves optimizing the GSS algorithm for source separation in the frontend and enhancing the ASR self-supervised/supervised models in the backend. Our system achieves world championship in all tracks.

Bird Sound Recognition in Complex Acoustic Environments 🔄

March 2020 – March 2021

National First Prize Program in Chinese Computer Design Competition (Team Leader)

Prof. Jin Yang (SCU)

- This project focuses on optimizing a blind source separation method (IVA) to identify all foreground bird species within overlapping vocalization recordings, and develop a visual multichannel signal simulation platform [4].

AWARDS

- | | |
|--|-----------|
| • First-Class Scholarship for University of Science and Technology of China | 2023-2024 |
| • Distinguished Undergraduate Thesis of Sichuan University (Top-4%) | 2022 |
| • National First Prize in the China College Students' Computer Design Competition | 2021 |
| • First-Class Scholarship for Sichuan University | 2019-2021 |
| • Certificate of Honor from the Yuzhang Wu Honors College at Sichuan University (Top-2%) | 2018-2022 |