YUSHENG DAI

Hefei, Anhui, China

✉ dalisondys@gmail.com   🌐 google scholar   ⧉ github.com/dalision

EDUCATION

University of Science and Technology of China (985 project)   Sep. 2022 – May 2025
Master of Information Science and Technology **GPA: 3.64**   *Hefei, China*

Sichuan University (985 project)   Sep. 2018 – Jun. 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; Supervisor: Prof. Jun Du

- 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 pretraining method that correlates lip shapes with syllable-level subword units to alleviate the discrepancies between audio and video inputs in the end-to-end AVSR training framework [1].
- The recent work explores an interesting phenomenon caused by the dropout technique in audio-visual applications when dealing with missing video frame input from the perspective of modality bias [6].

Holding MISP Challenge 2021 - 2023 ⧉   Jan. 2022 – Present
Graduate Research; Supervisor: Prof. Jun Du

- The MISP challenges 2021, 2022, and 2023 have been successfully held as the grand challenges in recent ICASSP years. As a core team member, I actively involve in dataset recording and developing baseline systems [4][5].

Financial Data Movement Prediction ⧉   March 2022 – Present
Graduate Research; Self-directed Research

- Researching Financial Data Movement Prediction, particularly in stock trading, has long been my interest. We formalize the stock movement prediction problem within a standard meta-learning framework to address challenges like limited data availability and potential domain shift.
- We have developed an algorithm, MASSER, leveraging self-supervised learning and meta-learning for both offline and online day-trading scenarios. This algorithm not only achieves SOTA performance on benchmark datasets but has also proven successful in real-world testing, assisting me in covering my graduate tuition [3].

Bird Sound Recognition in Complex Acoustic Environments ⧉   March 2020 – March 2021
Undergraduate Research; Supervisor: 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, spectrogram recognition, and multichannel signal simulation.
- We finally develop a platform to assist biological researchers in ecological surveillance and build a simulation experiment environment to generate synthesized multi-channel signals for further research. [2]

IoT Firmware Dynamic Vulnerability Discovery System ⧉   Feb. 2021 – June 2021
Undergraduate Research; Supervisor: Prof. Cheng Huang

- In this project, we design an IoT firmware vulnerability discovery system, employing static code auditing and dynamic fuzz testing to establish a universal large-scale firmware simulation and management framework.
- This system integrates technologies such as device virtualization, network virtualization, Docker container technology, dynamic network crawling, packet parsing, taint analysis, and fuzzy testing.

PUBLICATIONS

[1] **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**.

[2] **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***.

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

[4] 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.

[5] 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.

[6] One study related to the Robustness of Audio-Visual Applications under Modality Bias View has been submitted to ***CVPR*** 2024 as the first author.

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.