Jiawei Du

+886976413341 • Email: isjiawei.du@gmail.com • Homepage: jiaweidu.top • Google Scholar: J.W. Du

EDUCATION

National Taiwan University — **Taipei, Taiwan**
M.S. in Computer Science and Information Engineering, GPA 4.03/4.3 — *Sep. 2022 - Jun. 2025*
- Ranked 1/189 in the department in the 2023/24 academic year
- Supervised by Prof. Jyh-Shing Roger Jang and work intensively with Prof. Hung-yi Lee

Ming Chuan University — **Taoyuan, Taiwan**
B.S. in Information and Telecommunications Engineering, GPA 3.98/4.0 — *Sep. 2018 - Jun. 2022*
- Cumulatively ranked 1/79 in the department
- Supervised by Prof. Shu-Yin Chiang in Robotics

Shanghai Jiao Tong University — **Shanghai, China**
Exchange Student in Computer Science and Technology — *Sep. 2020 - Jan. 2021*

RESEARCH INTERESTS

Jiawei's research interests center on neural audio codecs, AI security, and audio-centric multimodal large language models. His work will focus on developing advanced techniques for deepfake detection and localization, exploring token-based speech representation learning for speech generation and understanding, and designing robust and trustworthy speech models. These technologies are critical for distinguishing truth from falsehood and promoting the generation of authentic content in the era of AIGC.

PUBLICATIONS (* is equal contribution) *59 citations*

TPAMI×1, AAAI×1 (both under review); Interspeech×2, SLT×3, APSIPA×1, DCASE×1, Preprint×1

- Xuanjun Chen*, Shih-Peng Cheng*, **Jiawei Du**, Lin Zhang, Xiaoxiao Miao, Chung-Che Wang, Haibin Wu, Hung-yi Lee, Jyh-Shing Roger Jang, "Localizing Audio-visual Temporal Deepfakes via Bidirectional Hierarchical Boundary Matching," 2026 Association for the Advancement of Artificial Intelligence (Submitted to AAAI'26)
- **Jiawei Du***, Xuanjun Chen*, Haibin Wu, Lin Zhang, I-Ming Lin, I-Hsiang Chiu, Wenze Ren, Yuan Tseng, Yu Tsao, Jyh-Shing Roger Jang, Hung-yi Lee, "CodecFake+: A Large-Scale Codec-based Deepfake Speech Dataset," 2025 IEEE Transactions on Pattern Analysis and Machine Intelligence (Submitted to TPAMI'25)
- Xuanjun Chen*, I-Ming Lin*, Lin Zhang, **Jiawei Du**, Haibin Wu, Hung-yi Lee, Jyh-Shing Roger Jang, "Codec-Based Deepfake Source Tracing via Neural Audio Codec Taxonomy," 2025 ISCA Proc. Interspeech (Interspeech'25)
- **Jiawei Du***, Xuanjun Chen*, Haibin Wu, Jyh-Shing Roger Jang, Hung-yi Lee, "Neural Codec-based Adversarial Sample Detection for Speaker Verification," 2024 ISCA Proc. Interspeech (Interspeech'24)
- **Jiawei Du***, I-Ming Lin*, I-Hsiang Chiu*, Xuanjun Chen, Haibin Wu, Wenze Ren, Yu Tsao, Hung-yi Lee, Jyh-Shing Roger Jang, "DFADD: The Diffusion and Flow-Matching Based Audio Deepfake Dataset," 2024 IEEE Spoken Language Technology Workshop (SLT'24)
- Haibin Wu, **Jiawei Du***, Xuanjun Chen*, Yi-Cheng Lin*, Kaiwei Chang*, Ke-Han Lu*, Alexander H. Liu*, Ho-Lam Chung*, Yuan-Kuei Wu*, Dongchao Yang*, Songxiang Liu, Yi-Chiao Wu, Xu Tan, James Glass, Shinji Watanabe, Hung-yi Lee, "Codec-SUPERB@ SLT 2024: A lightweight benchmark for neural audio codec models," 2024 IEEE Spoken Language Technology Workshop (SLT'24)
- Haibin Wu*, Huang-Cheng Chou*, Kai-Wei Chang, Lucas Goncalves, **Jiawei Du**, Jyh-Shing Roger Jang, Chi-Chun Lee, Hung-Yi Lee, "Open-Emotion: A Reproducible EMO-SUPERB for Speech Emotion Recognition Systems," 2024 IEEE Spoken Language Technology Workshop (SLT'24)
- Haibin Wu*, Huang-Cheng Chou*, Kai-Wei Chang, Lucas Goncalves, **Jiawei Du**, Jyh-Shing Roger Jang, Chi-Chun Lee, Hung-Yi Lee, "Empower Typed Descriptions by Large Language Models for Speech Emotion," 2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA'24)
- Haibin Wu*, Huang-Cheng Chou*, Kai-Wei Chang, Lucas Goncalves, **Jiawei Du**, Jyh-Shing Roger Jang, Chi-Chun Lee, Hung-Yi Lee, "EMO-SUPERB: An In-depth Look at Speech Emotion Recognition," 2024 Preprint (arXiv'24)
- **Jiawei Du***, Chung-Che Wang*, Jyh-Shing Roger Jang "Dcase 2023 task 6b: Text-to-audio retrieval using pretrained models," 2023 Detection and Classification of Acoustic Scenes and Events (DCASE'23)

RESEARCH IMPACT

- **DFADD** – Dataset from SLT'24 with 2,200+ downloads on Hugging Face
- **PFlow-TTS** – Contributor to 220+ stars flow-matching TTS repo on GitHub
- **Spot-adv-by-vocoder** – 80+ stars GitHub repo citing my Interspeech'24 paper for its method extension
- **EMO-SUPERB** – Benchmark from arXiv'24 with 40+ stars on GitHub

WORK EXPERIENCE

Research Internship **Beijing, China**
Speech Lab, Samsung Research *Feb. 2025 – May. 2025*

- **Neural Audio Codec – Streaming Efficiency and High Fidelity**
  - Designed a multi-band processing and group vector quantization framework with integrated intra and inter frame LSTM, and developed proficiency with mainstream vector quantization methods
  - Improved POLQA from 4.24 to 4.38 at 16 kbps with only 500k model parameters and 1 GFLOPs complexity

RESEARCH EXPERIENCE

Research Assistant **Taipei, Taiwan**
MIR LAB, National Taiwan University *Sep. 2022 – Jun. 2025*

- **AI Security – Audio-Centric Multimodal Deepfake and Adversarial Attack Detection**
  - Proposed a multi-task framework to trace codec-resynthesized audio sources via taxonomy (Interspeech'25)
  - Presented codec-based adversarial attack defender for Automatic Speaker Verification System (Interspeech'24)
  - Collected a dataset for detecting TTS-generated deepfake audio from diffusion- and flow-matching-based models, achieving up to a 47% reduction in EER on out-of-distribution test sets (SLT'24)
  - Boosted mAP@50 from 13.1% to 75.7% on private audio-visual deepfake data via LoRA-tuned SoTA model
- **Neural Audio Codec – Evaluation and Benchmark**
  - Proposed the largest dataset of codec-synthesized speech, and established a taxonomy based on decoder domains, vector quantization types, and semantic information extraction methods (Submitted to TPAMI'25)
  - Developed a lightweight benchmark for evaluating the application of neural audio codecs on speech tasks with 10 metrics and license-compliant datasets (SLT'24)
- **Speech Emotion Recognition**
  - Proposed a benchmark for speech emotion recognition (SLT'24)
  - Investigated the role of textual descriptions in influencing recognition accuracy (APSIPA'24)

COMPETITIONS

IEEE SLT Challenge on Singing Voice Deepfake Detection Challenge **Macau, China**
Ranked 13/49 (total 74 participants, 130 submissions) *Jun. 2024*

- Designed Conformer-based classifier using SSL features with codec-based data augmentation
- Achieved 5.33% EER, outperforming the baseline by 56% (results)

IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events **Tampere, Finland**
Ranked 3/11 *May. 2023*

- Explored audio-text retrieval using VALOR with cross-attention, achieved 27.3% mAP@10 (results)

AWARDS

- Student Travel Grant (top 10 of 42 applicants, $500 USD), IEEE SLT, 2024
- Exchange Program Nomination to NUS (declined by candidate), National Taiwan University, 2023
- 6 Academic Scholarships (rank 1/79, $4,000 USD), Ming Chuan University, 2018–2022

SERVICES

- Reviewer, ACM International Conference on Multimedia (ACMMM), 2025
- Reviewer, IEEE Transactions on Audio, Speech, and Language Processing (TASLP), 2025
- Reviewer, IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2025
- Reviewer, World Sci. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 2024–2025
- Reviewer, IEEE Journal of Selected Topics in Signal Processing (JSTSP), 2024
- Reviewer, IEEE Spoken Language Technology Workshop (SLT), 2024
- Technical Committee, Codec-SUPERB Special Session @ SLT, 2024

SKILLS

Programming Languages: Python, C/C++
ML Libraries & Tools: PyTorch, Numpy, SpeechBrain, OpenCV, Git, WSL, Linux, LaTeX
Language: Mandarin (native), English (IELTS 7.0; no band below 6.5)
Hobbies: music (electronic keyboard, piano, guitar, singing), basketball, traveling