

Xuanjun Chen

✉ d12942018@ntu.edu.tw | 🏠 xjchen.tech | 🐙 xjchenGit | 📺 jun-ntu | 🎓 Google Scholar

Research Interests

- Speech processing, Machine learning, Deepfake detection and localization

Education

National Taiwan University (NTU)

PH.D. STUDENT IN GRADUATE INSTITUTE OF COMMUNICATION ENGINEERING, GPA 4.30/4.30

M.S. IN COMPUTER SCIENCE AND INFORMATION ENGINEERING, GPA 4.19/4.30

- **Advisor:** Prof. [Jyh-Shing Roger Jang](#) and Prof. [Hung-yi Lee](#)

Taipei Taiwan

Sep. 2023 - Present

Sep. 2020 - Jan. 2023

National Taiwan University of Science and Technology (NTUST)

B.S. IN COMPUTER SCIENCE AND INFORMATION ENGINEERING, GPA 4.11/4.30

Taipei Taiwan

Sep. 2018 - Jun. 2020

Shenzhen Univ. of Information Technology (f.k.a. Shenzhen Institute of Information Technology, SZIT)

Shenzhen China

JUNIOR COLLEGE IN ELECTRONIC INFORMATION ENGINEERING TECHNOLOGY

Sep. 2015 - Jul. 2018

Selected Publications († indicates equal contribution.)

- [1] Chien-yu Huang, Wei-Chih Chen, Shu-wen Yang, Andy T. Liu, Chen-An Li, Yu-Xiang Lin, Wei-Cheng Tseng, Anuj Diwan, Yi-Jen Shih, Jiatong Shi, William Chen, **Xuanjun Chen**, et al., “[Dynamic-SUPERB Phase-2: A Collaboratively Expanding Benchmark for Measuring the Capabilities of Spoken Language Models with 180 Tasks](#)”. in *International Conference on Learning Representations (ICLR)*, Singapore, April 2025.
- [2] Chih-Kai Yang, Yu-Kuan Fu, Chen-An Li, Yi-Cheng Lin, Yu-Xiang Lin, Wei-Chih Chen, Ho Lam Chung, Chun-Yi Kuan, Wei-Ping Huang, Ke-Han Lu, Tzu-Quan Lin, Hsiu-Hsuan Wang, En-Pei Hu, Chan-Jan Hsu, Liang-Hsuan Tseng, I-Hsiang Chiu, Ulin Sanga, **Xuanjun Chen**, Po-chun Hsu, Shu-wen Yang, Hung-yi Lee, “[Building a Taiwanese Mandarin Spoken Language Model: A First Attempt](#)”. *Tech Report*, Nov. 2024.
- [3] **Xuanjun Chen**, Haibin Wu, Chung-Che Wang, Hung-yi Lee, Jyh-Shing Roger Jang “[Multimodal Transformer Distillation for Audio-Visual Synchronization](#)”. in the *2024 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2024)*, Seoul, Korea, April 2024.
- [4] **Xuanjun Chen**, Haibin Wu, Jyh-Shing Roger Jang, Hung-yi Lee “[Singing Voice Graph Modeling for SingFake Detection](#)”. in *Interspeech 2024*, Kos Island, Greece, Sept 2024.
- [5] **Xuanjun Chen**[†], Jiawei Du[†], Haibin Wu, Jyh-Shing Roger Jang, Hung-yi Lee “[Neural Codec-based Adversarial Sample Detection for Speaker Verification](#)”. in *Interspeech 2024*, Kos Island, Greece, Sept 2024.
- [6] Haibin Wu, **Xuanjun Chen**, Yi-Cheng Lin, Kai-wei Chang, Ho-Lam Chung, Alexander H. Liu, Hung-yi Lee “[Towards audio language modeling-an overview](#)”. *Overview Report*, Feb. 2024.
- [7] Haibin Wu, Ho-Lam Chung, Yi-Cheng Lin, Yuan-Kuei Wu, **Xuanjun Chen**, Yu-Chi Pai, Hsiu-Hsuan Wang, Kai-Wei Chang, Alexander H. Liu, Hung-yi Lee “[Codec-SUPERB: An In-Depth Analysis of Sound Codec Models](#)”. in *Findings of the Association for Computational Linguistics ACL 2024*, Bangkok, Thailand, Aug 2024.
- [8] Haibin Wu, **Xuanjun Chen**, Yi-Cheng Lin, Jiawei Du, Kai-Wei Chang, Ke-Han Lu, Alexander Liu, Ho-Lam Chung, Yuan-Kuei Wu, Dongchao Yang, Songxiang Liu, Yi-Chiao Wu, Xu Tan, James Glass, Shinji Watanabe, and Hung-yi Lee “[Codec-SUPERB @ SLT 2024: A lightweight benchmark for neural codec models](#)”. in the *2024 IEEE Spoken Language Technology Workshop (IEEE SLT 2024)*, Macao, China, 2024.
- [9] 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](#)”. in the *2024 IEEE Spoken Language Technology Workshop (IEEE SLT 2024)*, Macao, China, 2024.
- [10] Hsuan-Yu Lin, **Xuanjun Chen**, Jyh-Shing Roger Jang “[Singer separation for karaoke content generation](#)”. in *The 27th Oriental COCOSDA Conference (O-COCOSDA 2024)*, Hsinchu, Taiwan, Oct 2024.
- [11] **Xuanjun Chen**[†], Haibin Wu[†], Helen Meng, Hung-yi Lee, Jyh-Shing Roger Jang “[Push-Pull: Characterizing the Adversarial Robustness for Audio-Visual Active Speaker Detection](#)”. in the *2022 IEEE Spoken Language Technology Workshop (IEEE SLT 2022)*, Doha, Qatar, Jan 2023.
- [12] **Xuanjun Chen**[†], Yen-Lun Liao[†], Chung-Che Wang, Jyh-Shing Roger Jang “[Adversarial Speaker Distillation for Counter-measure Model on Automatic Speaker Verification](#)”. in *Proc. 2nd Symposium on Security and Privacy in Speech Communication (SPSC 2022)*, Incheon, Korea, Sept 2022.

Awards

| | | |
|-----------|--|--------------------------|
| 2024 | Google Conference Scholarship , Google LLC | <i>Travel Grant</i> |
| 2024 | National Science and Technology Council (NSTC) Conference Subsidies , NSTC Taiwan | <i>Travel Grant</i> |
| 2024 | CTCI Foundation Bursary Award for Overseas Students , Taiwan CTCI Foundation | <i>Bursary Award</i> |
| 2020-2025 | Distinguished Academic Record Award (5 years) , Taipei Kwong Tong Community Associations | <i>Scholarship</i> |
| 2021 | Ranked 3rd of the logical access track of the ASVspoof 2021 challenge , Interspeech 2021 | <i>3rd/42 Worldwide</i> |
| 2018-2020 | Certificate of Achievement (Top 5%, 3 semesters) , NTUST | <i>Scholarship</i> |
| 2017 | National Bronze & Guangdong Provincial Gold Awards , 3rd China College Internet Entrep. Comp. | <i>3rd Nationwide</i> |
| 2016-2017 | National Encouragement Scholarship (Only 3%) & 3rd Prize of Academic Award (Top 20%) | <i>SZIIT Scholarship</i> |

Research Experiences ([~] indicates the research proficiency)

Speech Processing and Machine Learning Lab, NTU, Advisor: Prof. Hung-yi Lee

Sept. 2023 - Present

RESEARCH ASSISTANT

- **[Lead research]** Researched on **Singing Voice DeepFake Detection**
 - Proposed a SingGraph model for singing voice deepfake detection, which improves EER relatively for seen singers by 13.2%, for unseen singers by 24.3%, and unseen singers using different codecs by 37.1%. [\[Interspeech 2024\]](#)
- **[Collaboration]** Write a neural codecs overview [\[arXiv\]](#), build benchmark [\[Findings of ACL 2024\]](#), organize a challenge [\[SLT 2024\]](#).
- **[Mentorship]** Led 10 students on **TAIHUCAIS: TAIwan Humanities Conversational AI Knowledge Discovery System** project.

Multimedia Information Retrieval Lab, NTU, Advisor: Prof. Jyh-Shing Roger Jang

Oct. 2020 - Present

RESEARCH ASSISTANT

- **[Lead research]** Researched on leveraging **Cross-attention Distillation for Audio-Visual Learning**
 - Proposed multimodal Transformer distillation for the audio-visual synchronization model, reducing teacher parameters by 83.52%, achieving competitive performance and providing comprehensive analysis. [\[ICASSP 2024\]](#)
 - Received the [Google Conference Scholarship \(Asia-Pacific\)](#) of Google LLC
- **[Lead research]** Researched on the **Adversarial Robustness for Audio-Visual Learning** [\[IEEE SLT 2022\]](#)
 - Revealed the vulnerability of the audio-visual active speaker detection in many ways, such as single- and multi-modal attacks, three-attack algorithms, white- and black-box attackers, and training-aware and inference-aware scenarios.
 - Proposed audio-visual interaction loss enables inter-class dispersion and intra-class compactness, which outperforms the adversarial training by 33.14 mAP (%) under multi-modal attacks.
- **[Collaboration]** Researched on **Automatic Speaker Verification Spoofing Countermeasures (ASVspoof)**
 - Proposed ASD-ResNetSE model combines generalized end-to-end pre-training and adversarial fine-tuning, achieving competitive performance with only 22.5% teacher parameters and 19.4% teacher MACs [\[ISCA SPSC 2022\]](#).
 - Ranked **3rd out of 42 submissions** in the logical access track of the ASVspoof 2021 challenge of **Interspeech 2021**.
- **[Mentorship]** Led 5 students on **Advanced Technologies for Designing Trustable AI Services** industry cooperation project.
 - Proposed neural-codec based adversarial sample detection method [\[Interspeech 2024\]](#).
 - Built the diffusion and flow-matching based audio deepfake dataset [\[SLT 2024\]](#).

Leadership and Service

PROFESSIONAL SERVICES

| | |
|-----------|--|
| 2023-Now | Reviewer / Program Committee , EMNLP (2025), ACL (2024), ICASSP (2023-2025), LREC-COLING/COLING (2024-2025), MLSP (2024), IALP (2024), ECCV AVGenL (2024) |
| 2024 | Technical Committee , Codec-SUPERB Challenge at IEEE 2024 Spoken Language Technology Workshop (SLT 2024) |
| 2024-2025 | Teaching Assistant , EE5200: Introduction to Generative AI (2024 Spr.) and EE5184: Machine Learning (2025 Spr.) , NTU |

CROSS-FUNCTIONAL SERVICES

| | |
|-----------|--|
| 2023-Now | Administrative Assistant , NVIDIA-NTU Artificial Intelligence Joint Research Center, NTU |
| 2021-2022 | Vice Director , Mainland Alumni Association, NTU |
| 2019 | Head Volunteer , The Fourth Cross-Strait Youth Maker Competition, Tongji University, Shanghai & Suzhou, China |
| 2019 | Propaganda Minister and Teaching Material Leader , Asia-Pacific University Volunteers Association, NTUST |
| 2016-2018 | Team Leader , founded a photography studio and led the propaganda department of the speaking club at SZIIT |

Open Source and Skills

| | |
|-------------------------|--|
| Open Source | Official repository of MTDVocaLiST model for Audio-Visual Synchronization task (Starred 23) |
| Open Source | Official repository of SingGraph model for SingFake Detection task (Starred 21 , Fork 4) |
| PLs and Toolkits | Python, LaTeX, Shell, PyTorch, Huggingface, PyTorch Lightning, Numpy, Pandas, Matplotlib. |