

Xuanjun Chen

PH.D. STUDENT · NATIONAL TAIWAN UNIVERSITY

✉ CHEN, XUANJUN | 🏠 xjchen.tech | 💻 xjchenGit | 📺 jun-ntu | ✉ d12942018@ntu.edu.tw

Research Interest

- My research interests include deep learning and speech processing, especially in audio coding [6, 7], model compression [1, 3], multimodal learning [2, 3], and robustness [1, 2, 4, 5]. I am honored to receive a **Google Conference Scholarship** in 2024.

Education

National Taiwan University (NTU)

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

Taipei Taiwan

Sep. 2023 - Present

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

Sep. 2020 - Jan. 2023

- **Advisor:** Professor **Hung-yi Lee** and **Jyh-Shing Roger Jang**

National Taiwan University of Science and Technology (Taiwan Tech)

Taipei Taiwan

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

Sep. 2018 - Jun. 2020

Shenzhen Institute of Information Technology (SZIT)

Shenzhen, China

J.C. IN ELECTRONIC INFORMATION ENGINEERING TECHNOLOGY

Sep. 2015 - Jul. 2018

Publications († indicates equal contribution)

[7] Towards audio language modeling-an overview

Submitted to SPL 2024

HAIBIN WU, XUANJUN CHEN, YI-CHENG LIN, KAI-WEI CHANG, HO-LAM CHUNG, ALEXANDER H LIU, HUNG-YI LEE

[arXiv]

[6] Codec-SUPERB: An In-Depth Analysis of Sound Codec Models

Findings of ACL 2024

HAIBIN WU, HO-LAM CHUNG, YI-CHENG LIN, YUAN-KUEI WU, XUANJUN CHEN, YU-CHI PAI, HSIU-HSUAN WANG,

[arXiv] [Code] [Leaderboard]

KAI-WEI CHANG, ALEXANDER H LIU, HUNG-YI LEE

[Huggingface]

[5] Singing Voice Graph Modeling for SingFake Detection

Interspeech 2024

XUANJUN CHEN, HAIBIN WU, JYH-SHING ROGER JANG AND HUNG-YI LEE

[ISCA] [arXiv]

[4] Neural Codec-based Adversarial Sample Detection for Speaker Verification

Interspeech 2024

XUANJUN CHEN, JIAWEI DU[†], HAIBIN WU[†], JYH-SHING ROGER JANG AND HUNG-YI LEE

[ISCA] [arXiv]

[3] Multimodal Transformer Distillation for Audio-Visual Synchronisation

IEEE ICASSP 2024

XUANJUN CHEN, HAIBIN WU, CHUNG-CHE WANG, HUNG-YI LEE, AND JYH-SHING ROGER JANG

[IEEE] [arXiv] [Code] [Poster]

[2] Push-Pull: Characterizing the Adversarial Robustness for Audio-Visual Active Speaker Detection

IEEE SLT 2022

XUANJUN CHEN[†], HAIBIN WU[†], HELEN MENG, HUNG-YI LEE, AND JYH-SHING ROGER JANG

[IEEE] [arXiv] [Demo] [Poster]

[1] Adversarial Speaker Distillation for Countermeasure Model on Automatic Speaker Verification

ISCA SPSC 2022

YEN-LUN LIAO[†], XUANJUN CHEN[†], CHUNG-CHE WANG, AND JYH-SHING ROGER JANG

[ISCA] [arXiv]

Honors and Awards

2024 **Google Conference Scholarship (Asia-Pacific)**, Google LLC

Travel Grant

2020-2023 **Distinguished Academic Record Award (4 years)**, Taipei Kwong Tong Community Associations

Scholarships

2018-2020 **Certificate of Achievement (3 semesters)**, Department of CSIE, Taiwan Tech

Top 5%

2021 **Ranked 3rd of the logical access track of the ASVspoof 2021 challenge**, Interspeech 2021

3rd/42 Worldwide

2017 **National Bronze Award and Guangdong Provincial Gold Award**, the 3rd China College Students'

3rd Nationwide

"Internet Plus" Innovation and Entrepreneurship Competition

2017 **3rd Prize of Academic Award**, Department of EIET, SZIT

Top 20%

2016 **National Encouragement Scholarship**, Chinese Ministry of Education

Only 3%

Research Experience ([~] indicates the research proficiency)

Speech Processing and Machine Learning Lab, NTU

Taipei Taiwan

RESEARCH ASSISTANT, ADVISOR: PROF. HUNG-YI LEE

Sep. 2023 - Present

- **[Lead research]** Researched on **Adversarial Robustness of Automatic Speaker Verification**
 - Designed neural codec-based adversarial sample detection method based on Descript-audio-codec, surpassing seven prior SOTA detection methods. [\[Interspeech 2024\]](#)
- **[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]** Neural codecs overview survey [\[arXiv\]](#) and benchmarking [\[arXiv\]](#).
- **[Mentorship]** Led 10 students on **TAIHUCAIS: TAIwan HUmanities Conversational AI Knowledge Discovery System** project.
 - Built a transitional Chinese Retrieval Augmented Generation evaluation dataset
 - In progressing

Multimedia Information Retrieval Lab, NTU

Taipei Taiwan

RESEARCH ASSISTANT, ADVISOR: PROF. JYH-SHING ROGER JANG

Oct. 2020 - Present

- **[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 model 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.
 - Built a Chinese audio-visual spoof-aware speaker verification dataset
 - In progressing

Communication and Leadership

PROFESSIONAL

2023 - Now **Reviewer**, ACL (2024), ICASSP (2023, 2024), LREC-COLING (2024), MLSP (2024)

2024 **Technical Committee**, Codec-SUPERB Challenge at IEEE 2024 Spoken Language Technology Workshop (SLT 2024)

2024 Spr. **Teaching Assistant**, EE5200: [Introduction to Generative AI](#) (Prof. Hung-yi Lee), NTU

CROSS-FUNCTIONAL

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, Taiwan Tech

2017-2018 **Team Leader**, Started a photography studio with 5 members while studying at SZIIT

2016-2017 **Propaganda Minister**, Led 20 members in Speaking Club's Propaganda Department of SZIIT

Skills

Programming	Python, C, C++, LaTeX
Machine Learning Related	PyTorch, PyTorch Lightning, Numpy, Pandas, Matplotlib
Languages	Mandarin, Cantonese, English