

# Xuanjun Chen

PH.D. STUDENT · NATIONAL TAIWAN UNIVERSITY

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

## 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 and Preprints ( † indicates equal contribution)

### PUBLICATIONS

#### [8] 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]

#### [7] Singing Voice Graph Modeling for SingFake Detection

Interspeech 2024

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

[ISCA] [arXiv] [Code]

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

Interspeech 2024

XUANJUN CHEN<sup>†</sup>, JIAWEI DU<sup>†</sup>, HAIBIN WU, JYH-SHING ROGER JANG AND HUNG-YI LEE

[ISCA] [arXiv] [Code]

#### [5] 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]

#### [4] Codec-SUPERB @ SLT 2024: A lightweight benchmark for neural codec models

IEEE SLT 2024

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

[arXiv] [Code] [Huggingface]

WATANABE, HUNG-YI LEE

#### [3] DFADD: The Diffusion and Flow-Matching based Audio Deepfake Dataset

IEEE SLT 2024

JIAWEI DU, I-MING LIN, I-HSIANG CHIU, XUANJUN CHEN, HAIBIN WU, WENZE REN, YU TSAO, HUNG-YI LEE,

JYH-SHING ROGER JANG

[arXiv]

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

IEEE SLT 2022

XUANJUN CHEN<sup>†</sup>, HAIBIN WU<sup>†</sup>, 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

XUANJUN CHEN<sup>†</sup>, YEN-LUN LIAO<sup>†</sup>, CHUNG-CHE WANG, AND JYH-SHING ROGER JANG

[ISCA] [arXiv]

### PREPRINTS

#### [2] Towards audio language modeling-an overview

Preprints 2024

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

[arXiv] [Awesome List]

#### [1] Singer separation for karaoke content generation

Preprints 2021

HSUAN-YU CHEN, XUANJUN CHEN, JYH-SHING ROGER JANG

[arXiv] [Project]

## Honors and Awards

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

Travel Grant

2024 **National Science and Technology Council (NSTC) Conference Subsidies**, NSTC Taiwan

Travel Grant

2020-2025 **Distinguished Academic Record Award (5 years)**, Taipei Kwong Tong Community Associations

Scholarships

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

Top 5%

## Honors and Awards

2021	<b>Ranked 3rd of the logical access track of the ASvspoof 2021 challenge</b> , Interspeech 2021	3rd/42 Worldwide
	<b>National Bronze Award and Guangdong Provincial Gold Award</b> , the 3rd China College Students' "Internet Plus" Innovation and Entrepreneurship Competition	3rd Nationwide
2017	<b>3rd Prize of Academic Award</b> , Department of EIET, SZIIT	Top 20%
2016	<b>National Encouragement Scholarship</b> , 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 **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], and organize a challenge [SLT 2024].
- **[Mentorship]** Led 10 students on **TAIHUCAIS: TAIwan Humanities Conversational AI Knowledge Discovery System** project.

### 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.
  - Proposed neural-codec based adversarial sample detection method [Interspeech 2024].
  - Built the diffusion and flow-matching based audio deepfake dataset [SLT 2024].

## Communication and Leadership

### PROFESSIONAL

- 2023-Now **Reviewer/Program Committee**, EMNLP (2025), ACL (2024), ICASSP (2023-2024), 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 Spr. **Teaching Assistant**, EE5200: **Introduction to Generative AI**, 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 languages & Frameworks:** Python, C/C++, LaTeX, PyTorch, PyTorch Lightning, Numpy, Pandas, Matplotlib
- **Languages:** Mandarin, Cantonese, English