Xuanjun Chen

D12942018@ntu.edu.tw · xjchen.tech · Google Scholar · linkedin.com/in/jun-ntu

Summary

Xuanjun is a Ph.D. candidate at National Taiwan University, advised by Profs. Hung-yi Lee and Jyh-Shing Roger Jang. He focuses on **Speech & Audio Processing** and **Large Language Models**, spanning foundation models, pre/post-training, self-improving agents, and deepfakes. He has published **15+ first/co-first papers** in Speech and NLP (e.g., COLM, TASLP, TIST, ICASSP, INTERSPEECH) and is a core contributor to DeSTA2.5-Audio (TASLP) and the SUPERB benchmarks (ICLR, ACL). He holds a Google PhD Fellowship (Machine Perception), NVIDIA Academic Grant, and IEEE SPS Scholarship, plus an IEEE ASRU Best Student Paper Nomination and three Best/Outstanding Paper Awards.

Education

National Taiwan University, Ph.D. in Communication Engineering, GPA: 4.3/4.3 — Sep 2023 – Present
National Taiwan University, M.S. in Computer Science & Information Engineering (CSIE), GPA: 4.19/4.3 — Jan 2023
National Taiwan University of Science and Technology, B.S. in CSIE, 3-time Dean's List, GPA: 4.11/4.3 — Jun 2020

Research Experience

NVIDIA Academic Grant Program, Research Lead (Grant Recipient), Taipei Taiwan — Oct 2026 – Mar 2027

- Research on speech/audio understanding and generation, in collaboration with NVIDIA Taiwan.

National Taiwan University, Research Assistant (Ph.D. & M.S.), Taipei Taiwan — Oct 2020 – Present

- Published over 20 peer-reviewed publications with **15 as first/co-first author** and 5 under review; selected publications are listed below.
- **Speech Generation.** Explored LLM agents for automating RL post-training in zero-shot TTS, halving generation failure cases [1].
- **Speech and Audio Language Models.** Analyzed LALM post-training, showing gains stem largely from answer-format fitting, not audio understanding [2]; core contributor to DeSTA2.5-Audio, Dynamic-SUPERB, Codec-SUPERB, and a codec survey [3]–[7].
- **Retrieval and Agentic LLMs.** Introduced a dual-loop agentic framework for optimizing benchmark generation harnesses [8]; improved multi-hop RAG via associative graph navigation (+7–10% recall) [9] and grounded delta planning (22–68% lower cost) [10].
- **Deepfake Speech.** Built CodecFake+, the largest CodecFake proxy dataset [11] [12], and codec source tracing [13] [14]; achieved SOTA in SingFake detection [15] [16] with deep-dive analyses [17]–[19], plus adversarial attacks [20] [21] and audio-visual learning [21] [22].

Research Highlights

*Total citations **600+** as of Oct 2026.*

- Can LLM Agents Automate Reinforcement Learning for Text-to-Speech? *ICASSP '27 (Submitted)*
  TL;DR: *An LLM agent reconstructs and improves an RL post-training recipe for zero-shot TTS, cutting failure cases by half.*
- What Does LALM Post-training Learn? Rethinking Audio Understanding and Reasoning Capabilities *NAACL '27 (Submitted)*
  TL;DR: *Much of the post-training gain comes from fitting answer formats and textual cues, not from better audio understanding.*
- DeSTA2.5-Audio: Toward General-Purpose Large Audio Language Model with Self-Generated Cross-Modal Alignment *TASLP '26*
  TL;DR: *Aligns LLMs with audio via self-generated cross-modal data without task-specific instruction tuning; 28.9K+ HF downloads.*
- MetaBench-Harness: Unlocking End-to-End Optimization of Benchmark Harnesses *ICLR '27 (Submitted)*
  TL;DR: *A dual-loop framework where a meta-harness refines the harnesses that evolve benchmarks hard enough for frontier LLMs.*
- Only Ask What You Don't Know: Grounded Delta Planning for Efficient Multi-step RAG *COLM '26*
  TL;DR: *Plans only for missing information after a first retrieval; top multi-hop QA accuracy at 22–68% lower cost.*
- Towards Generalized Source Tracing for Codec-Based Deepfake Speech *ASRU '25* · Best Student Paper Nominee (Top 2.5%)
  TL;DR: *SASTNet fuses Whisper semantics with Wav2vec2/AudioMAE acoustics to trace CodecFake speech back to its source codec.*

Open Source Contributions

Lead	**CodecFake+**	Largest proxy dataset of codec-resynthesized speech against CodecFake; 5.9K+ downloads [HF]	*TASLP '26*
	DFADD	Mentored students on the first deepfake dataset from diffusion/flow-matching TTS; 9.8K+ downloads [HF]	*SLT '24*
	SingGraph	Achieved SOTA on SingFake detection, fusing MERT and wav2vec 2.0; 24 GitHub stars [GitHub]	*INTERSPEECH '24*
	MTDVocaLiST	Compact audio-visual lip-sync model distilled from VocaLiST; 29 GitHub stars [GitHub]	*ICASSP '24*
Collab.	**DeSTA2.5-Audio**	(Core Contributor) 144 GitHub stars, HF model (28.9K+) and data (11.5K+ downloads) [GitHub]	*TASLP '26*
	Dynamic-SUPERB Phase-2	(Task Editor) 203 GitHub stars, 325 HF datasets (303K+ downloads) [GitHub]	*ICLR '25*
	Codec-SUPERB	(Core Contributor) 311 GitHub stars and a public leaderboard [Website]	*ACL Findings '24 & SLT '24*

Selected Honors, Service & Skills

Full list on xjchen.tech/misc

Honors	**Google PhD Fellowship** Recognized as one of 13 Google PhD Fellows in Machine Perception worldwide	2026
	NVIDIA Academic Grant 35K H100 GPU-hours (8xH100) via competitive selection by NVIDIA	2026
	IEEE SPS Scholarship Selected 44 out of nearly 300 qualified applicants worldwide	2026
	Best Student Paper Nominee IEEE ASRU 2025, selected 11 out of 447 submitted papers (Top 2.5%)	2025
	Best/Outstanding Paper Awards ISCSLP 2026 (Top 5%), TAAI 2026 (Top 1.5%), ROCLING 2025 (Top 2.8%)	2025–26
Service	**Co-Organizer,** Responsible Speech and Audio Generative AI, Special Session at IEEE ASRU 2025	2025
	Technical Committee, Codec Speech Proc. Universal PERformance Benchmark Challenge, IEEE SLT 2024	2024
	Reviewer: AAAI, ICLR, NeurIPS, ICML; ACL, EMNLP, NAACL, COLM; ICASSP, INTERSPEECH, ASRU, SLT, etc.	2023–Now
Skills	**Coding & Toolkit:** Python, LaTeX; PyTorch, PyTorch Lightning, Hugging Face Transformers	

Selected Publications

*Selected from 25+ papers, with the full list on Google Scholar; * equal contribution*

Speech & Audio Language Models and Generation

[1] **X. Chen**, et al., "Can LLM Agents Automate Reinforcement Learning for Text-to-Speech?" *ICASSP 2027 (Submitted)*

[2] H.H. Chen*, **X. Chen***, et al., "What Does LALM Post-training Learn? Rethinking Audio Understanding and Reasoning Capabilities," *NAACL 2027 (Submitted)*

[3] K.H. Lu, Z. Chen, ... **X. Chen**, ... H. Lee, "DeSTA2.5-Audio: Toward General-Purpose Large Audio Language Model with Self-Generated Cross-Modal Alignment," *TASLP, 2026*

[4] C.Y. Huang, ... **X. Chen**, ... S. Watanabe, H. Lee, "Dynamic-SUPERB Phase-2: A Collaboratively Expanding Benchmark for Measuring the Capabilities of Spoken Language Models with 180 Tasks," *ICLR 2025*

[5] H. Wu*, ... **X. Chen**, ... H. Lee, "Codec-SUPERB: An In-Depth Analysis of Sound Codec Models," *ACL Findings 2024*

[6] H. Wu, **X. Chen**, et al., "Codec-SUPERB @ SLT 2024: A Lightweight Benchmark for Neural Codec Models," *SLT 2024*

[7] H. Wu, **X. Chen**, et al., "Towards Audio Language Modeling – An Overview," *Technical Report, Feb. 2024*

Retrieval and Agentic Large Language Models

[8] **X. Chen**, et al., "MetaBench-Harness: Unlocking End-to-End Optimization of Benchmark Harnesses," *ICLR '27 (Submitted)*

[9] C.Y. Li*, **X. Chen***, et al., "CodaRAG: Connecting the Dots with Associativity Inspired by Complementary Learning," *ACM TIST, 2026*

[10] W.C. Chou*, **X. Chen***, et al., "Only Ask What You Don't Know: Grounded Delta Planning for Efficient Multi-step RAG," *COLM 2026*

Audio Deepfakes: Neural Codec, Singing Voice, Adversarial Attack, and Audio-Visual Scenarios

[11] **X. Chen***, et al., "CodecFake+: Codec-Based Resynthesized Data as a Proxy for Detecting CodecFake Speech," *TASLP, 2026*

[12] **X. Chen**, et al., "Mitigating Proxy-to-Wild Domain Gap in Deepfake Speech," *ISCSLP 2026* · Outstanding Paper

[13] **X. Chen***, I.M. Lin*, et al., "Towards Generalized Source Tracing for Codec-Based Deepfake Speech," *ASRU 2025* · Best Student Paper Nominee

[14] **X. Chen**, et al., "Codec-Based Deepfake Source Tracing via Neural Audio Codec Taxonomy," *INTERSPEECH 2025*

[15] **X. Chen***, C.Y. Hu*, et al., "Joint Fullband-Subband Modeling for High-Resolution SingFake Detection," *INTERSPEECH 2026* · Long Oral

[16] **X. Chen**, et al., "Singing Voice Graph Modeling for SingFake Detection," *INTERSPEECH 2024* · Oral

[17] **X. Chen**, et al., "Why Is Spectrogram ResNet Still Effective in Singing Voice Deepfake Detection?" *ICASSP '27 (Submitted)*

[18] Y.A. Hsieh*, **X. Chen***, et al., "On Bona Fide Audio Discrepancies in Singing Voice Deepfake Detection," *ICASSP '27 (Submitted)*

[19] **X. Chen**, et al., "How Does Instrumental Music Help SingFake Detection?" *ICASSP 2026*

[20] **X. Chen**, et al., "Neural Codec-based Adversarial Sample Detection for Speaker Verification," *INTERSPEECH 2024*

[21] **X. Chen**, et al., "Push-Pull: Characterizing the Adversarial Robustness for Audio-Visual Active Speaker Detection," *SLT 2022*

[22] **X. Chen**, et al., "Multimodal Transformer Distillation for Audio-Visual Synchronization," *ICASSP 2024*