Yuxiang Mei

E-mail: m153517@icloud.com |Portfolio: yuceong-may.github.io

EDUCATION

Sep 2024 – Jun 2027 **Shanghai Normal University** **M.Eng. in Communication Engineering（Supervisor: Yanhua Long）**

Sep 2020 – Jun 2024 **Xi'an Peihua University** **B.S. in Software Engineering**

RESEARCH INTERESTS

Automatic Speech Recognition (ASR); Speech Large Language Model (Speech-LLM).

INTERNSHIP EXPERIENCE

May 2026 – Present **AI Speech Algorithm R&D Intern | Baidu, ACG, Shanghai, China**

Conducted research on Text-Guided Target-Speaker ASR (TS-ASR) for far-field smart-home voice interaction. Developed cross-modal attention between text and speech encoders to guide the Paraformer model in target-speaker recognition and interference suppression. Extended the approach into **a unified TS-ASR framework supporting text-only, enrollment-only, and dual-cue inference.** Further explored embedding-free enrollment and low-frame-rate modeling for efficient speaker conditioning.

This work has been submitted to **ICLR 2027.**

Oct 2025 – Mar 2026 **Speech Foundation Model Research Intern | Unisound, AI Labs, Xiamen, China**

Proposed **Zipper-LoRA,** a parameter-efficient fine-tuning method that dynamically decouples shared and language-specific parameters to reduce cross-language interference in multilingual Speech-LLMs. Evaluated the method on a Whisper + LLM architecture across 12 languages, achieving significant improvements in low-resource ASR while maintaining competitive high-resource performance.

This work was published in **IEEE Transactions on Audio, Speech and Language Processing (TASLP 2026).**

GitHub： https://github.com/YuCeong-May/Zipper-LoRA

PROJECT EXPERIENCE

Jul 2025 – Sep 2025 **Bridging the Gap: Speech-LLM vs. End-to-End ASR**

Compared Speech-LLM and E2E ASR architectures for multilingual conversational ASR. Introduced cross-attention fusion into a Whisper + mHuBERT dual-encoder system and demonstrated the effectiveness of speech encoder pre-tuning before LLM integration. Investigated encoder fusion strategies and their impact on multilingual recognition performance, providing insights into Speech-LLM architecture design and optimization.

This work was published in **ICASSP 2026.**

GitHub： https://github.com/YuCeong-May/MLC-SLM

Mar 2025 – Jun 2025 **MLC-SLM Challenge, INTERSPEECH 2025**

Designed a Whisper + mHuBERT-147 dual-encoder architecture integrating supervised and self-supervised speech representations with Qwen2.5-7B. Applied LoRA to both speech encoders and the LLM, and developed a three-stage training strategy with language-specific prompting. Without additional data, achieved 33.1% / 41.7% relative improvements on development/test sets, ranking 7th among 25 teams.

This work was published in Workshop on Multilingual Conversational Speech Language Model **(MLC-SLM).**

Sep 2024 – Feb 2025 **End-to-End ASR and Speech-LLM Reproduction**

Reproduced Conformer-Transducer and HuBERT-CTC models on LibriSpeech 100h using ESPnet, covering data preprocessing, training, decoding, and evaluation. Gained practical experience with CTC and Transducer modeling and decoding mechanisms. Reproduced the first-stage training of SALMONN, exploring feature alignment and cross-modal interactions between speech encoders and large language models.

PUBLICATIONS & PREPRINTS

[1] **Yuxiang Mei,** Yuchen Yan, Dongxing Xu, Jiaen Liang, Yanhua Long, “Unified Target-Speaker ASR with Text and Enrollment Speech Cues,” **Submitted** to the ICLR 2027. ***Arxiv***

[2] **Yuxiang Mei**, Delai Qiu, Shengping Liu, Jiaen Liang, and Yanhua Long. "Zipper-LoRA: Dynamic Parameter Decoupling for Speech-LLM based Multilingual Speech Recognition." IEEE Transactions on Audio, Speech, and Language Processing, vol. 34, pp. 3956–3969, 2026. **(TASLP2026)** ***IEEE Xplore***

[3] **Yuxiang Mei,** Dongxing Xu, Jiaen Liang, and Yanhua Long. "Bridging the Gap: A Comparative Exploration of Speech-LLM and End-to-End Architecture for Multilingual Conversational ASR." Proc. IEEE ICASSP, 2026, pp. 18797-18801. **(ICASSP2026)** ***IEEE Xplore***

[4] **Yuxiang Mei**, Yuang Zheng, Dongxing Xu, and Yanhua Long, “SHNU Multilingual Conversational Speech Recognition System for INTERSPEECH 2025 MLC-SLM Challenge,” **in Workshop on Multilingual Conversational Speech Language Model (MLC-SLM), 2025, pp. 38–42.** ***ISCA Archive***

HONORS & AWARDS

Sep 2026	**National Scholarship for Graduate Students, Ministry of Education of China**
Oct 2026	**First-Class Graduate Academic Scholarship, Shanghai Normal University**

TECHNICAL SKILLS

ASR: Experienced with ESPnet, WeNet, and FunASR for multilingual ASR development and optimization.

Speech-LLM: Familiar with Transformers, DeepSpeed, and PEFT for model training and fine-tuning.

Programming & Engineering: Proficient in Python, PyTorch, and Linux, with experience in model deployment and AI-assisted development.