Xiquan Li

andreasli0912@gmail.com Linkedin Github Google Scholar

Education

- **Shanghai Jiao Tong University** *Sep. 2020 - June 2024*
  BS., Information Engineering; BA., French Shanghai, China
  - GPA: 92.05, 1/85
- **Shanghai Jiao Tong University** *Sep. 2024 - Mar. 2027*
  MS., Electronic Engineering Shanghai, China
  - GPA: 3.77/4.0
- **Telecom Paris - Polytechnic Institute of Paris** *Sep. 2023 - June 2026*
  ME., Electronic Engineering (Joint Master Program between SJTU and Telecom Paris) Paris, France
  - GPA: 3.77/4.0

Publications

[1] **DRCap: Decoding CLAP Latents with Retrieval-Augmented Generation for Zero-shot Audio Captioning**
Xiquan Li, *Wenxi Chen, Ziyang Ma, Xuenan Xu, Yuzhe Liang, Zhisheng Zheng, Qiuqiang Kong and Xie Chen*
in Proc. ICASSP 2025. [Paper] [Code]

[2] **SLAM-AAC: Enhancing Audio Captioning with Paraphrasing Augmentation and CLAP-Refine through LLMs**
Wenxi Chen, Ziyang Ma, ***Xiquan Li****, Xuenan Xu, Yuzhe Liang, Zhisheng Zheng, Kai Yu, Xie Chen*
in Proc. ICASSP 2025. [Paper] [Code]

[3] **EmoBox: Multilingual Multi-corpus Speech Emotion Recognition Toolkit and Benchmark**
Ziyang Ma, Mingjie Chen, Hezhao Zhang, Zhisheng Zheng, Wenxi Chen, ***Xiquan Li****, et al.*
in Proc. InterSpeech 2024. [Paper] [Code]

[4] **SLAM-Omni: Timbre-Controllable Voice Interaction System with Single-Stage Training**
Wenxi Chen, Ziyang Ma, Ruiqi Yan, Yuzhe Liang, ***Xiquan Li****, et al.*
in arXiv [Paper] [Code]

Experience

- **X-LANCE Lab - Shanghai Jiao Tong University** Shanghai, China
  Research intern supervised by Prof. Xie Chen *Feb. 2023 - Now*
  – Leveraged the pre-trained audio encoder EAT and the LLM Vicuna-7B for automated audio captioning (AAC). Fine-tuned the model using the LoRA method and introduced back-translation to address data scarcity. Developed a CLAP-Refine strategy to further improve caption quality, achieved 3rd place in DCASE 2024 Task 6: Automated Audio Captioning. A related paper [**2**] has been accepted by ICASSP 2025
  – Keywords: Automated Audio Captioning, Large Language Model
- **DSP Lab - The Chinese University of Hong Kong** Hong Kong SAR, China
  Research Assistant, working with Prof. Qiuqiang Kong *June 2024 - Sep. 2024*
  – Leveraged the CLAP (Contrastive Language-Audio Pre-training) model in conjunction with the Large Language Model (LLM) to achieve zero-shot audio captioning. Developed a hybrid approach integrating projection decoding from the encoder side with retrieval-augmented generation (RAG) from the decoder side to bridge the modality gap and improve captioning accuracy.
  – Our system achieved state-of-the-art performance in both in-domain and cross-domain scenarios. A paper outlining these results [**1**] has been accepted by ICASSP 2025.
  – Key words: Large Language Model, Retrieval-augmented Generation, Zero-shot audio Captioning
- **ADASP Group - Telecom Paris** Paris, France
  In research track (a program selecting 10 students per year), supervised by Prof. Slim Essid *Sep. 2024 - June. 2025*
  – Conducted a comprehensive literature review on large audio language models (LALM), uncovering their basic structure, training paradigm and evaluation benchmarks. Defining typical hallucinations in state-of-the-art LALMs.
  – Key words: Large Audio Language Model, LLM Hallucination

Selected Awards

- **DCASE 2024 Task 6: Automated Audio Captioning** Ranked 3rd DCASE Org. *(2024)*
- **ICASSP 2024 ICMC-ASR Grand Challange** Ranked 7/36 IEEE Org. *(2024)*
- **SPEIT Excellent Scholarship (First Class)** Top 3% SPEIT, SJTU *(2022, 2023)*
- **Mathematical Contest in Modeling** Finalist Winner - Top 2% World Wide COMAP Org. *(2022)*
- **ARDIAN Scholarship** Top 3% ARDIAN Cop. *(2021)*

Technical skills

Languages: English (CET6: 624), French (DELF B2), Chinese (Native)

Programming: Python, MATLAB, LaTeX, SQL, PyTorch, C/C++