

YUCHEN HU

(+65)-8039-2078 [◇ yuchen005@e.ntu.edu.sg](mailto:yuchen005@e.ntu.edu.sg) [◇ Homepage](#) [◇ Google Scholar](#) [◇ GitHub](#)

EDUCATION

Nanyang Technological University

08/2021 - 10/2025

Ph.D. in Computer Science. Supervisor: [Eng Siong Chng](#).

Singapore

University of Science and Technology of China

09/2016 - 06/2020

B.Eng. in Automation. GPA: 3.76/4.3 (Rank: Top 5%). [\[Transcript\]](#)

Hefei, China

RESEARCH & INTERNSHIPS

Nanyang Technological University

08/2021 - Present

Research Assistant, Supervisor: Eng Siong Chng

- **Large Language Models (LLMs)**

- *Generative Seq2seq Learning*: We propose a generative error correction (GER) framework that uses LLM to generate the ground-truth transcription from N-best hypotheses in seq2seq tasks (e.g., ASR, AST, MT). [\[5\]](#) [\[4\]](#) [\[3\]](#) [\[2\]](#)

- *Text-to-Speech Synthesis with Human Feedback*: We introduce RLHF techniques into TTS models to improve the subjective quality of synthesized speech.

- **Speech Processing**

- *Speech Recognition*: We propose STAR, a source-free unsupervised domain adaptation approach, which can adapt speech foundation models (e.g., Whisper) to specific target domains using **less than one-hour data**. [\[1\]](#)

- *Robustness in Speech Recognition*: We propose several noise-robust ASR approaches under adverse conditions, including front-end enhancement, feature quantization, GER denoising, etc. [\[14\]](#) [\[13\]](#) [\[12\]](#) [\[9\]](#) [\[4\]](#)

- *Speech Enhancement and Separation*: We propose a speech enhancement (SE) approach using classifier-guided diffusion model, and we explore improving noise-robustness of speech separation with SE front-end. [\[10\]](#) [\[11\]](#)

- **Multimodal**

- *Audio-visual Representation Learning*: We propose several audio-visual speech recognition approaches. [\[6\]](#) [\[7\]](#) [\[8\]](#)

iFLYTEK AI Research & USTC NEL-SLIP

05/2020 - 07/2021

Research Intern, Supervisor: Lirong Dai

- *Simultaneous Speech Translation*: Develop a cross-attention augmented transducer (CAAT) system with USTC-NELSLIP team and achieve the 1-st Place at IWSLT 2021 Evaluation Campaign. [\[15\]](#)

- *Streaming Speech Recognition*: Improve decoding efficiency of Hybrid Autoregressive Transducer by pruning.

PUBLICATIONS & PREPRINTS

- [1] **Y. Hu**, C. Chen, C. H. H. Yang, C. Qin, P. Y. Chen, E. S. Chng, C. Zhang, “*Self-Taught Recognizer: Toward Unsupervised Adaptation for Speech Foundation Models*”, **Under Review**.
- [2] **Y. Hu**, C. Chen, C. H. H. Yang, R. Li, D. Zhang, Z. Chen, E. S. Chng, “*GenTranslate: Large Language Models are Generative Multilingual Speech and Machine Translators*”, **ACL 2024**. [\[Paper\]](#) [\[Code\]](#) [\[Data\]](#)
- [3] **Y. Hu**, C. Chen, C. Qin, Q. Zhu, E. S. Chng, R. Li, “*Listen Again and Choose the Right Answer: A New Paradigm for Automatic Speech Recognition with Large Language Models*”, **ACL 2024**. [\[Paper\]](#)
- [4] **Y. Hu**, C. Chen, C. H. H. Yang, R. Li, C. Zhang, P. Y. Chen, E. S. Chng, “*Large Language Models are Efficient Learners of Noise-Robust Speech Recognition*”, **ICLR 2024 (Spotlight, Top 5%)**. [\[Paper\]](#) [\[Code\]](#) [\[Data\]](#)
- [5] C. Chen*, **Y. Hu***, C. H. H. Yang, S. M. Siniscalchi, P. Y. Chen, E. S. Chng, “*HyPoradise: An Open Baseline for Generative Speech Recognition with Large Language Models*”, **NeurIPS 2023**. [\[Paper\]](#) [\[Code\]](#) [\[Data\]](#)

- [6] **Y. Hu**, R. Li, C. Chen, C. Qin, Q. Zhu, E. S. Chng, “*Hearing Lips in Noise: Universal Viseme-Phoneme Mapping and Transfer for Robust Audio-Visual Speech Recognition*”, **ACL 2023 (Oral)**. [\[Paper\]](#) [\[Code\]](#)
- [7] **Y. Hu**, C. Chen, R. Li, H. Zou, E. S. Chng, “*MIR-GAN: Refining Frame-Level Modality-Invariant Representations with Adversarial Network for Audio-Visual Speech Recognition*”, **ACL 2023 (Oral)**. [\[Paper\]](#) [\[Code\]](#)
- [8] **Y. Hu**, R. Li, C. Chen, H. Zou, Q. Zhu, E. S. Chng, “*Cross-Modal Global Interaction and Local Alignment for Audio-Visual Speech Recognition*”, **IJCAI 2023**. [\[Paper\]](#) [\[Code\]](#)
- [9] **Y. Hu**, C. Chen, Q. Zhu, E. S. Chng, “*Wav2code: Restore Clean Speech Representations via Codebook Lookup for Noise-Robust ASR*”, **IEEE/ACM TASLP, 2023**. [\[Paper\]](#)
- [10] **Y. Hu**, C. Chen, R. Li, Q. Zhu, E. S. Chng, “*Noise-aware Speech Enhancement using Diffusion Probabilistic Model*”, **Under Review**. [\[Paper\]](#) [\[Code\]](#)
- [11] **Y. Hu**, C. Chen, H. Zou, X. Zhong, E. S. Chng, “*Unifying Speech Enhancement and Separation with Gradient Modulation for End-to-End Noise-Robust Speech Separation*”, **ICASSP 2023**. [\[Paper\]](#) [\[Code\]](#)
- [12] **Y. Hu**, C. Chen, R. Li, Q. Zhu, E. S. Chng, “*Gradient Remedy for Multi-Task Learning in End-to-End Noise-Robust Speech Recognition*”, **ICASSP 2023**. [\[Paper\]](#) [\[Code\]](#)
- [13] **Y. Hu**, N. Hou, C. Chen, E. S. Chng, “*Dual-Path Style Learning for End-to-End Noise-Robust Speech Recognition*”, **InterSpeech 2023**. [\[Paper\]](#) [\[Code\]](#)
- [14] **Y. Hu**, N. Hou, C. Chen, E. S. Chng, “*Interactive Feature Fusion for End-to-End Noise-Robust Speech Recognition*”, **ICASSP 2022**. [\[Paper\]](#) [\[Code\]](#)
- [15] D. Liu, M. Du, X. Li, **Y. Hu**, L. Dai, “*The USTC-NELSLIP Systems for Simultaneous Speech Translation Task at IWSLT 2021*”, **IWSLT 2021**. [\[Paper\]](#)

SERVICES

Reviewer	ACL (23-24), ARR (23-24), EMNLP (23), AAAI (24), ICASSP (22,24), InterSpeech (22-24)
Volunteer	EMNLP (23), ICASSP (22)

SKILLS

Programming Languages	Python, C, Matlab
Deep Learning	PyTorch, HuggingFace, Fairseq, ESPnet, SpeechBrain, lit-gpt
English Levels	TOEFL (104, R30/L28/S22/W24), GRE (329+4.0), CET-6 (619), CET-4 (620)

HONORS & AWARDS

• ACL 2023 Area Chair Award	07/2023
• Winner of IWSLT 2021 Evaluation Campaign	08/2021
• USTC Excellent Graduate (Top 10%)	06/2020
• Scholarship of SIMIT, Chinese Academy of Sciences (Top 5%)	10/2018
• USTC Outstanding Student Scholarship (Top 5%)	10/2017 & 10/2019