YUCHEN HU

(+65)-8039-2078 ⋄ yuchen005@e.ntu.edu.sg ⋄ Homepage ⋄ Google Scholar ⋄ GitHub

RESEARCH FOCUS

Automatic Speech Recognition (ASR), Large Language Model (LLM), Multimodal

EDUCATION

Nanyang Technological University 08/2021 - 08/2025
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 & INTERNSHIP

Nanyang Technological University 08/2021 - Present
Research Assistant, Supervisor: Eng Siong Chng *Singapore*

- **Generative Speech Recognition with LLM** (In NeurIPS, ICLR)

  - We propose an ASR **generative error correction (GER)** benchmark that leverages LLM to generate the ground-truth transcription from ASR N-best hypotheses, which significantly outperforms typical LM rescoring methods. To enable LLM finetuning, we also propose a **HyPoradise** dataset that contains over 334K pairs of N-best hypotheses and ground-truth transcription. Experiments on LLaMA shows that GER achieves remarkable improvements over Whisper baseline on various ASR domains, with up to 79.5% relative WER reduction. [2]

  - We extend GER to noisy ASR and propose a **language-space denoising** approach to improve its noise robustness. Experiments show that our method achieves a new breakthrough with up to 53.9% WER reduction. [1]

  - We propose a dynamic late fusion approach to incorporate acoustic information into LLM to mitigate the data uncertainty in GER, which significantly improves its performance by up to 23.0% relative WER reduction. [12]

- **Noise-Robust Speech Recognition** (In ACL, AAAI, IJCAI, TASLP, ICASSP, InterSpeech)

  - We propose several audio-visual speech recognition (AVSR) approaches to improve ASR noise-robustness with visual modality, including cross-modal interaction, multimodal discrete mapping, adversarial learning, and reinforcement learning, which achieve the state-of-the-art on the large-scale LRS3 and LRS2 datasets. [3] [4] [5] [14]

  - We propose several noise-robust ASR approaches to alleviate the speech distortion issue in popular joint SE-ASR system, including enhanced-noisy feature fusion, gradient remedy, and quantization methods, which achieve over 10% relative WER improvements on CHiME-4, LibriSpeech-FreeSound and RATS datasets. [6] [9] [10] [11]

- **Speech Enhancement and Separation** (In ICASSP)

  - We propose a noise-aware speech enhancement (SE) approach with classifier-guided diffusion model, which achieves promising improvements over various diffusion SE baselines on VoiceBank-DEMAND dataset. [7]

  - We propose a joint speech enhancement and separation framework for noise-robust speech separation, which achieves the state-of-the-art on Libri2Mix and Libri3Mix (noisy version) datasets. [8]

iFLYTEK AI Research & USTC NEL-SLIP 05/2020 - 07/2021
Research Intern, Supervisor: Lirong Dai *Hefei, China*

- Develop a Cross-Attention Augmented Transducer (CAAT) system with USTC-NELSLIP team for simultaneous speech translation, which achieves the 1-st Place at IWSLT 2021 Evaluation Campaign. [16]
- Improve streaming ASR decoding efficiency of Google's Hybrid Autoregressive Transducer by pruning.

PROJECT EXPERIENCE

ANPASSEN: Unseen Noise and Multilingual Speech Recognition 09/2023 - Present

- Scale: 2 years, over S$ 600K.
- Role: Project engineer, responsible for noise-robust ASR. [1] [12] [2]

ISSAC: Language Identification, Speaker Diarization, and Speech Separation 08/2021 - Present

- Scale: 3 years, over S$ 960K.
- Role: Project engineer, responsible for speech separation. [8]

MAISON2: Speech Recognition in Adverse Conditions 08/2021 - 12/2021

- Scale: 2.5 years, over S$ 600K.
- Role: Project engineer, responsible for noise-robust ASR. [10] [11]

PUBLICATIONS & PREPRINTS

[1] **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**. [Paper] [Code] [Data]

[2] 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]

[3] **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**. [Paper] [Code]

[4] **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**. [Paper] [Code]

[5] **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]

[6] **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]

[7] **Y. Hu**, C. Chen, R. Li, Q. Zhu, E. S. Chng, *"Noise-aware Speech Enhancement using Diffusion Probabilistic Model"*, **Under Review**. [Paper] [Code]

[8] **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]

[9] **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]

[10] **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]

[11] **Y. Hu**, N. Hou, C. Chen, E. S. Chng, *"Interactive Feature Fusion for End-to-End Noise-Robust Speech Recognition"*, **ICASSP 2022**. [Paper] [Code]

[12] C. Chen, R. Li, **Y. Hu**, C. H. H. Yang, S. M. Siniscalchi, P. Y. Chen, E. S. Chng, *"It's Never Too Late: Fusing Acoustic Information into Large Language Models for Automatic Speech Recognition"*, **ICLR 2024**. [Paper]

[13] Q. Zhu, J. Zhang, Y. Gu, **Y. Hu**, L. Dai, *"Multichannel AV-wav2vec2: A Framework for Learning Multichannel Multi-modal Speech Representation"*, **AAAI 2024**. [Paper] [Code]

[14] C. Chen, **Y. Hu**, Q. Zhang, H. Zou, B. Zhu, E. S. Chng, *"Leveraging Modality-specific Representations for Audio-visual Speech Recognition via Reinforcement Learning"*, **AAAI 2023**. [Paper]

[15] H. Zou, M. Shen, C. Chen, **Y. Hu**, D. Rajan, E. S. Chng, *"UniS-MMC: Multimodal Classification via Unimodality-supervised Multimodal Contrastive Learning"*, **ACL 2023**. [Paper] [Code]

[16] 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), ARR (23), EMNLP (23), AAAI (24), ICASSP (22,24), InterSpeech (22,23)
Volunteer	EMNLP (23), ICASSP (22)

SKILLS

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

AWARDS & HONORS

• 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