Shuo Han

Pittsburgh, PA
shuohan@andrew.cmu.edu
(518) 526 – 0507
Personal Webpage

EDUCATION

M.S. in Electrical & Computer Engineering, Carnegie Mellon University **Sep 2021 – May 2023**

B.S. in Computer & System Engineering, Rensselaer Polytechnic Institute **Sep 2018 – May 2021**

Concentrations: Speech Enhancement, Speech Recognition, Large Language Model, Signal Processing

RESEARCH EXPERIENCE

Research Assistant | Language Technologies Institute, Carnegie Mellon University **Sep 2024 – Nov 2024**

- Define comparative reasoning and ADIFF - Audio-Language Model (ALM) for explaining audio differences.
- Curated multi-tiered datasets with natural language explanations derived from annotated audio captions.
- Deployed an efficient data collection pipeline, cutting collection time by 90% and ensuring long-term scalability.
- ADIFF outperforms state-of-the-art ALMs by 57% in objective metrics and 3% in human evaluations.

Research Assistant | Language Technologies Institute, Carnegie Mellon University **May 2024 – Sep 2024**

- AudioEntailment: Define deductive reasoning for audio and build up benchmark dataset and framework.
- Create deductive reasoning datasets grounded in audio, with LLM-generated hypotheses validated by humans.
- Experimented with various prompting strategies to improve LLMs' audio understanding through textual captions.
- Benchmarked state-of-the-art ALMs, revealing logical reasoning gaps in zero-shot and linear-probe settings.

Research Assistant | MLSP & WiTech Lab, Carnegie Mellon University **Nov 2023 – Feb 2024**

- DeWinder: A novel wind noise reduction system leveraging ultrasound sensing for enhanced audio clarity.
- Implement a modular framework compatible with cutting-edge models (e.g., DEMUCS, DCCRN).
- Collect ~10 hours of ultrasound recordings under varied wind conditions.
- Integrate Doppler-based ultrasound characterization for superior low-SNR noise suppression.
- Demonstrate significant PESQ and STOI improvements across diverse SNR conditions.

Student Researcher | Carnegie Mellon University & Meta Reality Lab **Sep 2022 – Oct 2022**

- PAAPLoss: Incorporates phoneme-specific weights to tailor acoustic optimization to phonetic context.
- Fine-tuning Demucs and FullSubNet with PAAPLoss boosts PESQ by 7.5% and reduces WER from 15.0% to 13.2%.
- Phoneme-dependent analysis shows more significant loudness gains for plosives and F1 improvement for vowels.

Student Researcher | Carnegie Mellon University & Meta Reality Lab **June 2022 – Sep 2022**

- TAPLoss: A differentiable estimator for temporal acoustic parameters to optimize enhanced speech.
- Integrating TAPLoss with Demucs and FullSubNet improves acoustic parameters by 15% and PESQ by 7.4%.
- Integrating into AudioGenie enables an easy switch to any speech enhancement model.

PUBLICATIONS

ADIFF: Explaining audio difference using natural language — 2nd author, ICLR, 2025 (Under Review)

Audio Entailment: Assessing Deductive Reasoning for Audio Understanding — 2nd author, AAAI, 2025

Dewinder: A Temporal Acoustic Parameter Loss for Speech Enhancement. — co-author, Interspeech, 2024

Psychoacoustic Challenges Of Speech Enhancement On VoIP Platforms — SynData4GenAI Workshop, InterSpeech

TAPLoss: A Temporal Acoustic Parameter Loss for Speech Enhancement. — co-author, ICASSP, 2023

PAAPLoss: A Phonetic-Aligned Acoustic Parameter Loss for Speech Enhancement. — co-author, ICASSP, 2023

SKILLS

Programming Languages: Python, MATLAB, Java, C; Open-Source Framework: ESPNet, SUPERB, Fairseq, UniLM