Carnegie Mellon University
Electrical & Computer Engineering

Ph.D. Thesis Defense

Incorporating Modulation Information into Deep Neural Networks for Robust Speech Processing

April 26, 2023 | 9:30 a.m. ET | PH B09

SPEAKER

Tyler Vuong

COMMITTEE

Richard Stern (*Advisor*)
CMU-ECE

Rita Singh
CMU-LTI

Bhiksha Ramakrishnan
CMU-ECE/LTI

Aswin Sankaranarayanan
CMU-ECE

ABSTRACT

Interacting with speech-related technologies has become an integral part of our everyday lives. Whether suppressing background noise during a remote video meeting at home or using one's voice to interact with a smart device in the presence of other sounds, it is essential for speech processing systems to be robust to the presence of background interference. Although the current data-driven speech-processing systems have shown tremendous improvements in recent years, their performance in the presence of background noise and other types of degradation is still substantially worse than the corresponding performance with clean input speech. To bridge this gap, this thesis describes several novel ways of applying spectral-temporal receptive fields (STRFs) to integrate modulation information into deep neural networks (DNN) and improve the robustness of speech processing systems. STRFs respond to a range of patterns of temporal modulation and spectral modulation and are motivated by structures that are believed to describe processing in the brainstem and the auditory cortex.

In this thesis, we first developed a learnable front end that leverages deep learning to discover insights about the important temporal modulations in speech. We found that the low temporal modulations were important for the speech activity detection system, which is consistent with the results of prior research. We designed a loss function that has the objective of preserving the important modulations in speech. This loss function is used to guide the training of deep-learning-based speech enhancement systems. We obtained the important modulation parameters through the use of a neural network layer parameterized as a learnable STRF. Finally, we show that the incorporation of the learnable STRFs into DNN-based speech processing systems improves robustness performance for multiple downstream tasks including speech recognition, voice type discrimination, and speech activity detection.

PUBLIC DEFENSE

Event Contact: Tyler Vuong (tvuong@andrew.cmu.edu)

Zoom Link: https://cmu.zoom.us/j/93498149034?pwd=bFJiM0xYOU00ajMwRS9ZTU5QOVMrZz09

Visit https://engineering.cmu.edu/education/graduate-programs/phd-defenses/index.html
for a calendar of public Ph.D. defenses.