Statement of Purpose

I want to pursue a PhD and explore effective and novel approaches for speech processing. My passion in speech research stems from several research projects as an undergraduate. I have performed research at Johns Hopkins University (JHU), Human Language Technology Center of Excellence (HLTCoE), and University of Edinburgh (UoE). Through these experiences, I found myself fascinated with the most fundamental challenge inherent in spoken language processing: how do we design features for a given speech signal? I aspire to answer this question and in the past, I have investigated approaches from signal processing techniques, deep learning models and cognitive science ideas. Eventually, my goal is to lead a research laboratory and contribute to improving our everyday speech technology.

My interest in speech research first developed while working on **machine learning models for Voice Activity Detection (VAD)** with Professor Najim Dehak and Dr. Jesús Villalba at JHU. VAD is employed in all speech systems, yet the standard energy-thresholding method was effective only under high signal to noise ratio. To create a robust VAD model, I trained a Long Short-Term Memory Network (LSTM) under various noise conditions. I found that contextual information is crucial for speech identities detection, and presented the work at the 2017 JHU Undergraduate Research Symposium. To further improve my VAD model, I participated in the the 2017 Summer Camp for Applied Language Exploration (SCALE) at HLTCoE. I observed that outputs of my LSTM varied greatly between neighboring frames, contradicting the smoothness property of speech. To resolve the issue, I studied and implemented several machine learning models, and conversations with researchers at HLTCoE gave me new insights on my work. By the end of SCALE, I extended my LSTM with a two-state Hidden Markov Model with more diverse noise conditions from Google's AudioSet, and the model was evaluated on the National Institute of Standards and Technology OpenSAT.

The research at SCALE motivated me to develop a thorough understanding of the basics of speech processing, and I started to work with a graduate student on **applied signal and speech processing for Parkinson's Disease (PD) detection.** We designed automatic speech biomarkers for PD detection with Perceptual Linear Predictive and i-vectors features. To undertake the project, I took four graduate-level signal processing courses in my Junior year. As the research progressed, I improved my technical skill set from large-scale data processing to GPU and parallel computing. We were able to create the first PD detection system that utilizes an acoustic model, capturing the differences in articulatory movements between normal speakers and PD patients[1]. The design of psychophysics-inspired acoustic features and low-dimensional speaker embedding ignited my passion for finding optimal speech representations.

In my Junior year, I also collaborated with another graduate student on **speech bandwidth extension for automatic speaker verification (ASV).** Traditionally to avoid sampling mismatch in speech recordings, wideband data is first downsampled to match the sampling rate of narrowband data. However, downsampling incurs information loss and degrades downstream ASV performance. We proposed two upsampling techniques and demonstrated that our methods not only address sampling mismatches between recordings, but also improve accuracy by 11 % for ASV. The work was published at Interspeech 2018[2]. The research has a significant impact for speech processing research because the most common data such as microphone and

[1]Project presentation can be found at `https://goo.gl/MMUsyy`

[2]`https://www.isca-speech.org/archive/Interspeech_2018/pdfs/2394.pdf`

telephone speech are recorded at different rates. This project gave me the fundamentals to take challenges in spoken language research, as I became particularly experienced with three toolkits: Kaldi, Keras and PyTorch, allowing me to process data, build up speech pipelines from ground up, and train deep neural networks effectively.

To broaden my research horizon, I was awarded the Vredenburg Scholarship and interned at UoE working with Professor Simon King on **an attention model for replay attacks detection.** I collaborated with four researchers on developing countermeasures for speech adversarial attacks. These countermeasures are crucial for the development and deployment of ASV systems for the advent of voice assistants and smart home devices. I proposed Attentive Filtering Network, an attention model that automatically acquires and enhances speech feature representations that are helpful for the detection of replay attacks, and our system lowers the detection error rate by 20 % over the previous state-of-the-art system. I was able to explore a research question in depth with independence, and I gave talks on this work at UoE, JHU, and Rice University. The work was submitted to ICASSP 2019[3].

After working on several projects, I was excited to work on a different perspective of spoken language research for my senior thesis – **predictive coding for speech analysis.** My interest in predictive coding stems from taking classes on human auditory system with Professor Hynek Hermansky and vision as Bayesian inference with Professor Alan Yuille. Predictive coding states that human brains are constantly generating and updating hypothesis via a feedback loop for efficient coding, and I have been working on using the concept in designing unsupervised learning models to separate invariant information from temporally varying signals. I believe that research on cognitive science models for speech processing is a promising area to explore, as we could gain some insights from formal theories of cognition to inspect representations for speech.

Research has been the focus of my undergraduate studies. I have been fortunate to take part in projects that generated state-of-the-art results. More importantly, I found research an exceptionally enjoyable journey: identifying open problems, contemplating ideas from different angles, and collaborating with others. My motivation for applying to graduate school stems from the desire to build upon the work already done in applied signal processing, deep learning and cognitive science, and explore novel approaches for identifying the optimal speech representation. Looking forward, I intend to leverage my previous experiences in speech analysis and processing, and push the frontiers of spoken language research.

Advancing our knowledge in spoken language requires expertise from interdisciplinary fields such as statistics, computer science and electrical engineering. MIT Computer Science Artificial Intelligence Lab has not only demonstrated successes in speech but also in machine learning and cognitive science. MIT's committed faculty and research scientists as well as its collaboration with other research institutes and industries are unparalleled resources for me to continue develop as a researcher. In spoken language research, I am drawn to the work of Professor Jim Glass's group. My background in speech processing and speaker recognition would fit particularly well in their work on Speaker Verification and Diarization. In addition, their recent progress on speech representation with unsupervised learning methods is fascinating to me. I also look forward to exploring other research projects for my PhD study at MIT CSAIL. MIT CSAIL aligns well with my research goal and experiences, and I am confident that MIT stands out as the ideal place for my graduate studies.

[3] `https://arxiv.org/abs/1810.13048`