Abstract

Speech enhancement seeks to improve the quality of speech degraded by noise. Its importance can be found in applications such as mobile phone communication, speech recognition, and hearing aids. An example of speech enhancement relates to the famous cocktail party problem. This problem deals with extracting a target speaker's voice from a mixture of background conversations. In such a situation, the human brain tends to do a good job focusing in on the target speech while blocking out the noisy environment surrounding it. The goal of solving the cocktail party problem is to find a computer algorithm that functionally mimics how the brain extracts the target speaker's voice. In this master's thesis, a novel approach to solving the cocktail party problem is presented that relies on a fully convolutional neural network (FCN) architecture. The FCN takes noisy, raw audio data as input and performs nonlinear, filtering operations to produce clean, raw audio data of the target speech at the output. Results from experimentation indicate the ability to generalize to new speakers and robustness to new noise environments of varying signal-to-noise ratios.