David Levi, Christopher Kim, Justin Jose

ECE-395A: Senior Electrical Engineering Projects

**Final Report: Audio Source Separation**

**Abstract**

This project focuses on the ability to separate and enhance one audio source amongst many different audio sources. The goal is to emulate the cocktail party effect, the phenomenon of a human’s ability to focus on one voice while filtering out other noises. This project attempt to isolate and amplify the desired audio source in real time by utilizing deep neural networks. The audio source separator aims to help people with hearing loss have clear conversations with others in a crowded area.

**Introduction**

Audio source separation is a challenging and exciting technological field that has captured the attention of many signal processing experts and machine learning researchers. In particular, many have tried to emulate the human brain’s ability to focus on a singular audio source while filtering out other noises using technology. This phenomenon is better known as the cocktail party effect; the name is associated to the idea that a person is able to have a conversation with someone else in a noisy room, such as a cocktail party, due to the person’s ability to discern the other’s voice to listen to and filter out the rest of the voices in that noisy room. Since technologically reproducing the cocktail party effect has been such a popular one in the past years [1], this challenge has been named as the cocktail party problem.

With the rise of deep learning, many known problems related to image, audio and text that engineers once believed to be impossible to attempt have made great progress by utilizing neural networks. The cocktail party problem is one of many that has seen a resurgence in popularity due to the promising results that deep learning has been able to produce. While there have been recent successes in performance with the use of neural networks, such as the deep attractor network [2] and the recent neural egg network [3], there have not been efforts to emulating the cocktail party effect in real time. Undoubtedly, achieving promising results is a significant feat that should not be undermined. However, we believe that truly reproducing the cocktail party effect also entails a real-time component, where the technology would be able to separate speech from a source in an almost instant manner that reflects the way in which humans are able to do so.

Our primary motivation for this project is not pushing the boundaries of deep learning implementations of the cocktail party problem. Instead, our main focus is to create a technology that would be able to emulate the cocktail party effect in real time and provide assistance to people who suffer from hearing loss. This technology will be based on a convolutional neural network model built by a Cooper Union alumnus. While we do make improvements on the model, we