Abstract

The cocktail party problem is the task of extracting a speaker's voice from a noisy environment containing the voices of other speakers. The human brain does this quite well in most cases, but hearing impaired individuals often struggle to hear others due to background noise in a crowded bar or restaurant. Previous research efforts in the fields of signal processing and machine learning have presented frameworks to solve this task in supervised and unsupervised environments. One such implementation is Non-Negative Matrix Factorization (NMF).

NMF is an algorithm for creating an approximation of a non-negative matrix into the product of two smaller, non-negative matrices. NMF is a linear dimensionality reduction technique that is often applied in recommendation systems, text mining, spectral data analysis, and audio analysis. This thesis expands upon previous work demonstrating the benefits of using a neural network implementation of the NMF algorithm for the tangential problem of source separation. We present an end-to-end framework for approaching a semi-supervised case of the cocktail party problem. In the semi-supervised case, we attempt to extract a target speaker's speech signal from a noisy mixture signal only with the knowledge of the presence of noise, not the type of noise in the mixture. The framework presented in this thesis transforms a raw audio signal into an non-negative magnitude spectrogram