論文内容の要旨

論文題目　Blind Separation of Audio Sources from Single and Stereo Mixtures with a Special Consideration on Underdetermined Condition
(劣条件下における単一および２個の混合信号からのブラインド信号分離)

氏名　モラ エムディ カデムル イスラム

Abstract

The separation of mixed audio signals is the problem of automated separation of audio sources present around a set of differently placed microphones, capturing the acoustical scene. The whole problem resembles the task a human can solve in a cocktail party situation, where using two sensors (ears), the brain can focus on a specific source of interest, suppressing all other sources present (cocktail party problem). In this thesis, we examine the audio source separation problem using a range of approaches to segregate the component sources from monophonic and stereo recordings. In particular, we consider underdetermined condition (i.e. the number of sensors is less than the number of sources) which is a challenging topic in the field of blind source separation.

Independent component analysis (ICA) is a recently developed statistical method to tackle the blind source separation (BSS) problem. ICA does not work directly in the underdetermined case. We have successfully implement ICA to separate speech and interfering signals from single mixture. The mixed signal is projected to time-frequency (TF) space. A finite set of independent basis vectors are derived from the TF space by applying principal component analysis (PCA) and ICA sequentially. The vectors are clustered to represent the independent subspaces corresponding to the component sources in the mixture. The time domain source signals are reconstructed from the subspaces. The separation efficiency is greatly affected by the method employed in TF representation of the mixed signal. The short-time Fourier transform (STFT) and Hilbert spectrum (HS), a fine-resolution TF representation are used here. The experimental results show that HS performs better than STFT in single mixture separation.

Only the spectral information of the individual source is used in single mixture separation. It

becomes difficult to separate the sources with overlapping spectra. The use small size microphone array can recover this problem increasing the separation performance. We have proposed stereo mixtures (two microphones) to perform localization based separation of the audio sources. The separation performance based on spatial localization of the sources is independent of signal contents. When the audio sources are spatially distributed, the stereo recording introduces time difference (TD) and intensity difference (ID) between two microphones' signals. These two cues (TD and ID) are used for spatial localization of the sources considering that the sources are stationary in spatial locations. Then two methods are used to segregate the sources from stereo mixtures: (i) spatial beamforming and (ii) binary masking. The linear constraint minimum variance beamformer (LCMVB) is implemented on multi-band representation of the mixture signals. We have proposed some modification of the beamformer to tackle the problem of underdetermined condition. The separation by binary masking is implemented in TF domain and HS is employed as TF representation.

In real-world application it is usual case that the sources are moving while emitting the audio signals. The stationary consideration makes the proposed separation algorithm narrow in the practical usage. The final step of this thesis is the separation of multiple moving sources from stereo mixtures. It is considered that the sources are stationary for small time frame. The separation is performed for that time slot by using binary masking method. There occurs a permutation problem among the sources separated from the successive time frames. A data-adaptive audio source discrimination method is proposed to resolve the problem. The sources separated from the consecutive frames are concatenated properly to obtain the overall separation. The proposed algorithm is applied to separate the sources from stereo mixtures recorded in the anechoic room of NTT communication research laboratory. The simulation results show a noticeable performance in the field of audio source separation. The thesis concludes by highlighting some of the as yet unsolved problems to tackle the actual audio source separation problem in full.