Blind Source Separation and Direction Estimation for Stereophonic Mixtures of Multiple Speech Signals Based on Time-Frequency Sparseness

March 2012

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Engineering

Keio University

Graduate School of Science and Technology
School of Integrated Design Engineering

Ning Ding

Abstract

With the development of modern society, digital signal processing has become increasingly important because it is changing our daily lives in many ways. In the field of digital signal processing, human-machine interfaces are regarded as a key topic. Among the various applications of human-machine interfaces, speech information is very useful. We often encounter questions such as "Where is the speech source?" and "Can we obtain the desired speech from many simultaneous speeches?". The former is known as the direction-of-arrival (DOA) problem, while the latter is known as the blind source separation (BSS) problem, for which a sensor array technique is essential.

In this dissertation, several approaches to blind source separation and DOA estimation using a microphone array are described. In this study, without knowledge of source localization, active time or mixing process (blind), a pair of microphones is used to estimate the source directions and separate multiple speech signals even when the number of sources is two or more.

1. Speaker localization and source separation by Principal Component Analysis (PCA) and harmonic structure

In conventional methods data are treated as a whole in the time-frequency domain, and the difference between time frames is not distinguished. However, these methods suffer from low separation performance when the sources are closely located.

Since the ratio of the principal eigenvalues obtained by principal component analysis (PCA) indicates the degree of data spread around the first principal axis, in the author's approach, using the mathematical tool of PCA to analyze the phase difference versus frequency distribution data in a single time frame, The observated time frames are classified according to the activity pattern of multiple source frames to non-source active (NSA), single-source active (SSA) and double-source active (DSA) frames. SSA frames are used for DOA estimation. A new separation algorithm is explored for use in DSA frames. Depending on the frequency band, two methods are combined to obtain the separated signals in DSA frames: a DOA-based method and a harmonic-structure-based method.

2. Reliable cell selection

The common-sense approach that the use of reliable data guarantees reliable results is adopted. The problem is how to check the reliability of data from the observation. In the author's approach, the consistency with neighborhood data is utilized and cells are selected using a newly defined reliability index. The reliability index of a T-F cell's phase difference exploits the consistency of the time difference of arrival (TDOA) in the local window of the underlying cell. The consistency of the TDOA in a window is evaluated using the variance of the TDOAs for all T-F cells in the window.

3. Use of kernel density estimator for DOA estimation

A model of the propagation of the statistical error between the estimated phase difference and the consequent DOA is introduced. The model leads to a probability density function (PDF) of the DOA, then the DOA estimation problem is reduced to finding the most probable points for the DOA. Finally, the kernel density estimator is applied to selected cells to calculate the PDF and estimate the source direction.

Some experiments were performed to evaluate the proposed methods. The results show that the proposed source separation method is superior to the conventional method, and the proposed DOA estimation method outperforms other methods in terms of both accuracy in the case of real observed data and robustness in the case of simulation with additional diffused noise.