Title	動的モード分解を用いた音源分離の基礎的検討
Author(s)	宮崎, 陽祐
Citation	
Issue Date	2024-06
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19081
Rights	
Description	Supervisor: 鵜木 祐史, 先端科学技術研究科, 修士(情報科学)

Abstract

A single source usually produces a single sound. Based on this assumption, it is possible to separate sources by separating individual sounds from a signal waveform in which multiple sounds are mixed. This processing technology is called source separation. It is important to study methods to improve source separation performance and to enable separation under conditions that have been difficult to achieve in the past, because these methods can improve the performance of other acoustic signal processing applications, such as speech recognition. In this study, we focus on the separation of two sound sources. Therefore, we study techniques for separating individual acoustic signals from a mixture of two different acoustic signals.

The source separation strategy depends on the number of sources and the number of channels in the recorded sound. If the number of sources exceeds the number of channels, it is relatively easy to perform source separation. However, separation becomes more difficult if the number of sources is greater than the number of channels. Attempting source separation on a monaural sound is always difficult because the number of channels is always less than the number of sources. However, the monaural source separation method can be implemented with a single microphone, which provides a much greater degree of freedom in implementing the methods in real-world environments. Moreover, it is possible to extend the separation model that works on monaural sounds to multichannel sounds, or to integrate it with existing multichannel separation methods. Because of these potential applications, the study of source separation methods for monaural sounds is a very important topic because of its significant contribution to the overall sound source separation technology.

The separation of non-stationary signals is a long-standing problem in source separation of monaural signals. Research is still ongoing to solve this problem. Currently, the latest methods are based on Deep Neural Network (DNN), but these methods are limited by the large amount of training data required. On the other hand, the source separation method based on Principal Component Analysis (PCA) does not require any training. Therefore, PCA can be applied even in situations where sufficient training data is not available, making it an effective method in this regard. However, by its nature, PCA cannot capture local temporal changes in the amplitude and frequency of signals. Therefore, it is not suitable for analyzing non-stationary signals with such local changes. This is expected to hurt the source separation of non-stationary signals. Since the separation of non-stationary signals is an issue in the source separation of monaural signals, it is desirable to process sounds with a method suitable for the analysis of non-stationary signals.

This study focuses on a method called Dynamic Mode Decomposition (DMD), which, like PCA, decomposes signals into orthogonal components in a data-driven manner. In addition, DMD has a feature not found in PCA: the time evolution of the decomposed components can be observed as a parameter. This feature makes it possible to capture local changes in the amplitude and frequency of the sound. DMD has been used to analyze non-stationary signals in various research fields and has been applied in the field of acoustics. Therefore, it is expected that DMD can be effectively used in source separation tasks. The purpose of this study is to investigate the feasibility of source separation using DMD in the task of extracting only the target sound from a monaural sound. To this end, this study investigated how sounds are analyzed by DMD. Based on the results, this study designed a method of source separation using DMD and evaluated its separation performance.

The relationship between DMD and sounds was investigated, and the following three results were found: (1) The signal to be analyzed is represented by DMD as a linear sum of modes with unique frequencies and attenuation rates; (2) The distribution of time evolution features obtained by DMD analysis is significantly different between acoustic signals and noise; and (3) When a noise mixed signal is mode-decomposed by DMD, the distribution of time evolution features is divided into two groups, and it is possible to separate the sound by extracting only one of the groups. Based on the above findings, this study designed a source separation method using DMD by the following steps: (1) Mode decomposition and extraction of features related to the time evolution of the sounds/noise mixture by DMD; (2) Grouping the features estimated to correspond to the acoustic signal by focusing on the distribution of the features; (3) Synthesize the signal using only the modes associated with the features of grouping.

To investigate the feasibility of the DMD source separation method, DMD and PCA source separation methods were compared using computer simulations to compare the performance. As a result, the DMD source separation method showed higher performance than the PCA source separation method, indicating that DMD has good potential to be applied to source separation. However, it was also found that the low signal-to-noise ratio (SNR) of the noise mixture sounds distorts the features.

If this distortion of the features can be compensated, the performance of the DMD source separation method can be further improved. Therefore, it is necessary to investigate the relationship between noise level and the features distortion in the future. In addition, although this simulation investigated the separation of non-stationary artificial sound from stationary noise, it is necessary to further investigate whether this separation can also be performed for non-stationary noise. For this purpose, it is necessary to investigate the relationship between non-stationary noise and DMD in the future.