Digital Signal Processing 31 (2014) 79–92

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Multi-band multi-centroid clustering based permutation alignment for frequency-domain blind speech separation

Lin Wang [a,b]

[a] Institute of Physics, University of Oldenburg, Germany
[b] School of Electronic and Information Engineering, Dalian University of Technology, China

ARTICLE INFO

Article history:
Available online 28 April 2014

Keywords:
Blind source separation
Convolutive mixture
Frequency domain
Permutation ambiguity

ABSTRACT

This paper investigates the permutation ambiguity problem in frequency-domain blind source separation and proposes a robust permutation alignment algorithm based on inter-frequency dependency, which is measured by the correlation coefficient between the time activity sequences of separated signals. To calculate a global reference for permutation alignment, a multi-band multi-centroid clustering algorithm is proposed where at first the permutation inside each subband is aligned with multi-centroid clustering and then the permutation among subbands is aligned sequentially. The multi-band scheme can reduce the dynamic range of the activity sequence and improve the efficiency of clustering, while the multi-centroid clustering scheme can improve the precision of the reference and reduce the risk of wrong permutation among subbands. The combination of two techniques enables to capture the variation of the time–frequency activity of a speech signal precisely, promising robust permutation alignment performance. Extensive experiments are carried out in different testing scenarios (up to reverberation time of 700 ms and 4×4 mixtures) to investigate the influence of two parameters, the number of subbands and the number of clustering-centroids, on the performance of the proposed algorithm. Comparison with existing permutation alignment algorithms proves that the proposed algorithm can improve the robustness in challenging scenarios and can reduce block permutation errors effectively.

1. Introduction

Blind source separation has attracted considerable attention in research communities in recent years. Its main objective is to separate multiple sources mixed through unknown channels using only the observations of their mixtures. It has found a lot of potential applications including noise robust speech recognition, crosstalk separation in telecommunications, biomedical signal analysis and so on. The term "blind" means that separation is done without using any information about the mixing channels and the sources. The simplest BSS model assumes the existence of independent signals and the observation of their mixtures being linear and instantaneous. This problem can be solved by independent components analysis (ICA) [1,2]. A more challenging case is when the sources are mixed through convolutive channels [3]. This situation is common for audio applications where signals are recorded in a reverberant environment. Long reverberation time and non-stationary mixing conditions make the estimation of the original source signals a challenging task [4,5]. In addition, the high computational complexity hampers the application of the algorithms in real-time devices.

Traditional approaches to solve the convolutive blind source separation problem can be classified into two categories: time-domain and frequency-domain. In time-domain BSS, the separation network is derived by optimizing a time-domain cost function [6–8]. These approaches may not be effective due to slow convergence and large computation load. In frequency-domain BSS, the observed time-domain signals are converted into the time–frequency domain by short-time Fourier transform (STFT), and then an instantaneous BSS algorithm is applied to each frequency bin, after which the separated signals of all frequency bins are combined and inverse-transformed back to the time domain [9–13]. Although satisfactory instantaneous separation may be achieved at all frequency bins, combining them to recover the original sources is a challenge because there are unknown permutations associated with individual frequency bins. This permutation ambiguity should be dealt with properly so that the separated frequency components from the same source are grouped together.

There are generally three strategies to tackle the permutation ambiguity problem:

E-mail addresses: lin.wang@uni-oldenburg.de, wanglin_2k@sina.com.

http://dx.doi.org/10.1016/j.dsp.2014.04.009

- The first strategy is to exploit the continuity of the separation matrices across frequencies. In [7,14,15], separation filters of shorter length relative to the fast Fourier transform (FFT) block size are used so that the separation filters are smooth across frequencies and hence the permutation ambiguities are avoided. In [16], the magnitude continuity of the separation filters is used to align the permutation. In [17] a recursively regularized ICA (RR-ICA) algorithm is proposed which uses a predicted separation matrix from previous frequency bins as the initialization for the current bin. With this recursive initialization scheme, the phase continuity of the separation filters is maintained and the permutation ambiguities are minimized.
- The second strategy is to use the time structure of separated frequency bins, such as the inter-frequency dependency of the amplitude of separated signals [18,19], assuming high correlation between neighboring bins. Aligning consecutive bins using inter-frequency dependency may be precise but not robust, since a single wrong permutation leads to whole blocks of falsely permuted bins. This is referred to as misalignment spread. To solve this problem, clustering-based and region-wise permutation alignment schemes are proposed [19–23]. As a particular formulation of joint blind source separation [24,25], independent vector analysis (IVA) algorithms are proposed which directly incorporate the inter-frequency dependency measure into instantaneous ICA so that permutation ambiguities can be minimized by a joint optimization across all the frequency bins [26–30].
- The last strategy is to use position information of the sources such as direction of arrival (DOA) or time difference of arrival (TDOA) [31–34]. It is believed that contributions from the same source are likely to come from the same direction. By estimating the arriving delay of the sources or analyzing the directivity pattern formed by a separation matrix, source directions can be estimated and permutations aligned. The major drawback of this approach is that the assumption of sources originating from specific directions is only valid in low reverberation. Although robust in low reverberation, the performance of this strategy degrades significantly in highly reverberant environments. In addition, it fails to align the permutation when the two sources are closely spaced. In [35–38], source direction information and inter-frequency dependency of the separated signals are combined to get a precise and robust permutation result.

In this paper we aim at solving the permutation ambiguity problem based on inter-frequency dependency of separated signals, which can be measured by the correlation of the time activity sequences at individual frequency bins. The frequency-dependent time activity sequence, as can be calculated from the power ratio of the separated signals, has proven to show strong dependency between two frequencies if they come from the same source [19]. The key of permutation alignment is to find a frequency-independent global reference for each source respectively. By aligning to the reference the permutation can be corrected across the whole frequency band. Several ways have already been proposed to estimate such a reference:

- The simplest way is to use the time-activity sequence at an adjacent bin as a reference to align the permutation at the current bin [22]. Obviously, this bin-wise processing is sensitive to the permutation error at an individual bin, leading to misalignment spread.
- In [22], a region-growing algorithm is proposed, which divides the full frequency band into multiple regions based on the bin-wise permutation alignment result and then merges these regions together in a region-growing way, and shown to prevent misalignment spread effectively. The region-growing algorithm calculates the reference adaptively by doing permutation alignment and reference update simultaneously during its region-growing procedure. One disadvantage is that the permutation errors at individual bins may accumulate during the region-growing procedure, leading to possibly wrong updates of the reference.
- Clustering-based algorithms, which perform one-centroid [19] or multi-centroid clustering across the whole frequency band [20], are proposed to estimates centroid sequences for each source and uses them as a global reference for permutation alignment. In [21] this work is extended to an underdetermined separation problem. Compared with the region-growing algorithm, the clustering-based algorithm performs the two tasks of reference calculation and permutation alignment separately. It is less affected by the permutation alignment results at individual bins and thus tends to be more robust especially when a multi-centroid scheme is employed. However, although the time activity of a speech signal shows strong inter-frequency similarity, it still varies slowly with frequency. In some case, it is difficult to find a global reference that is consistent to all the bins throughout the frequency band. As a result, permutation errors still occur at some bins or even throughout a block of bins.

Given the deficiencies of the existing algorithms above, we propose a multi-band multi-centroid clustering algorithm to better estimate the reference. The proposed algorithm in essence is a combination of multi-band processing and multi-centroid clustering, i.e., the whole frequency band is split into multiple bands, and in each subband the permutation reference is estimated in a multi-centroid way. After the permutation correction inside each subband, the permutation between these subbands is aligned sequentially to recover the full frequency band. The multi-band scheme can reduce the dynamic range of the activity sequences and improve the efficiency of clustering; while the multi-centroid clustering scheme can improve the precision of the reference and reduce the risk of wrong permutation among subbands. The combination of two techniques can capture the variation of the time–frequency activity of a speech signal more precisely and hence promises better permutation alignment results. Based on the principle of the proposed algorithm, the number of subbands and the number of centroids play important roles on the permutation alignment performance. The impact of the two parameters will be investigated in this paper.

One issue which has not been studied well in previous papers is the robustness of a permutation alignment algorithm, i.e., it may perform well for one mixing scenario but fails in another case. For this reason, the performance of the permutation alignment algorithm is investigated extensively in various mixing scenarios with different reverberation time (up to 700 ms), number of sources (up to 4), and testing files. To make a comprehensive evaluation, four objective measures (the mean performance and robustness performance in terms of permutation error and signal-interference-ratio) are newly defined and computed with multiple testing files. Experimental results clearly show the superior performance of the proposed method over the single-band multi-centroid algorithm. Especially, it can improve the robustness and reduce block permutation errors effectively. Finally, comparison with other referenced permutation alignment algorithms in both simulated and real environments also proves the advantage of the proposed algorithm.

The rest of the paper is organized as follows. The principle of frequency-domain blind source separation is reviewed in Section 2. The proposed multi-band multi-centroid permutation alignment scheme is described in detail in Section 3. Experimental results are presented in Section 4. Computational cost analysis of the