Neurocomputing 185 (2016) 73–81

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Dynamic blind source separation based on source-direction prediction ☆

Yangjie Wei [a,*], Yi Wang [b]

[a] College of Information Science and Engineering, Northeastern University, Wenhua Str. 3, Shenyang, China
[b] Department of Information Technology, Uppsala University, S:t Olofsgatan 10 B, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 1 September 2014
Received in revised form
14 July 2015
Accepted 15 December 2015
Communicated by: Y. Liu
Available online 23 December 2015

Keywords:
Deterministic techniques
Blind source separation (BSS)
Kalman filter
Direction prediction
Three-sigma rule

ABSTRACT

Deterministic techniques are based on the source-directions and multipath characteristics of the reverberant environment for different source signals. However, searching for the desired directions of the time-block sequence of an acoustic signal is time consuming, and existing deterministic methods rarely consider the motion properties of the acoustic source. In this paper, a dynamic source-direction prediction method for real-time blind convolutive mixtures based on a Kalman filter is proposed. First, the convolutive mixture signals captured by the coincident array geometry are formulated, and the relationship between source-direction and source separation is developed. Second, motion prediction based on a Kalman filter is theoretically analyzed, and the motion of a source is modeled as a noise-driven position integrator with enough samples. Then, a dynamic source-direction prediction method for real-time blind source separation based on a Kalman filter is proposed to predict the directions of a time sequential signal. Combined with the local direction searching method, our proposed method has a self-correction ability according to the three-sigma rule. Finally, extensive experiments are performed with three-source convolutive mixtures of speeches in English and Chinese, whose direction varies in linear and nonlinear motions. The signal-to-distortion and signal-to-interference of the separated signals are calculated, and the experimental results demonstrate the feasibility and validity of the proposed method.

1. Introduction

In blind source separation (BSS), the interference-free versions of simultaneously active sound sources are obtained without any information about their number and positions or the acoustic environment [1–4]. As sound mixtures are almost convolutive in enclosures, BSS is a useful preprocessing stage for many application fields such as hearing aids, teleconferencing, multichannel audios, and acoustical surveillance. Currently, there are three types of techniques including stochastic, adaptive, and deterministic techniques that are widely used in the separation of convolutive sound mixtures.

Stochastic techniques such as independent component analysis (ICA) [5–7], are based on the assumption that the source signals are probabilistically independent and distributed with respect to a non-Gaussian distribution. Given these assumptions, contrast features representative of both non-Gaussianity and independence are maximized, leading to the recovery of separated signals. In fact, ICA is one of the most widely used BSS techniques for revealing hidden factors that underlie sets of random variables, measurements, or signals. Besides separation of speech sources, recently, ICA is also used in other applications such as mechanical systems [8,9], Biomedical Engineering [10–12], and image processing [13]. However, because several iterations are required for the computation of the demixing filters, stochastic methods are usually computationally expensive. Furthermore, ICA-based techniques in the frequency domain cannot avoid the scaling and permutation problems resulting from the independent application of separation algorithms in each frequency bin [14,15].

Adaptive techniques such as adaptive beamforming (ABF) [16,17] optimize a multichannel filter structure according to the properties of the signals and source geometry; therefore, they utilize spatial selectivity to suppress interference and improve the acquisition of the target source. These adaptive algorithms are similar to stochastic methods in the sense that they both depend on the properties of the signals to reach a solution, but the ABF algorithms require the geometry of the microphone array. The most obvious disadvantage of the ABF algorithms is that its solution adaption may stop at a suboptimal position. In addition, the null beamforming applied to the interference signal is not very

☆This work is supported by the National Natural Science Foundation of China (No. 61305025) and the Fundamental Research Funds for the Central Universities (N130504011).
* Corresponding author.
E-mail address: weiyangjie@ise.neu.edu.cn (Y. Wei).

http://dx.doi.org/10.1016/j.neucom.2015.12.040
0925-2312/

effective under reverberant conditions owing to the reflections; therefore, an upper bound for the performance of BSS techniques exists [18].

In contrast, deterministic techniques such as geometrical methods are solely based on the deterministic aspects of a problem, such as the source-directions and the multipath characteristics of the reverberant environment [19–21], rather than any assumptions about the source signals themselves, to separate signals. In general, deterministic methods do not exploit the source statistics, but they can solve a source separation problem based on the deterministic characteristics of the source mixture in the absence of noise, and then make sure that the solution algorithm still behaves robustly when noise is added. For example, a deterministic technique based on intensity vector direction exploitation (IVDE) for acoustic source separation has been previously proposed to provide a nearly closed-form solution for the separation of convolutive mixtures captured by a compact, coincident microphone array [22]. Because no assumption is needed for deterministic techniques, they are more likely to be used in practice, especially when no information about the sources and signals is known beforehand. However, deterministic techniques are difficult to be used in real-time applications because of several factors. For example, 1) the preprocessing stage before separation is time consuming. The solving formulation of deterministic techniques requires the prior knowledge of the source-directions and produces as many channels as the number of sources, therefore, a preprocessing stage such as the multiple signal classification (MUSIC) algorithm is needed to estimate the number and directions of sources before the separation algorithm begins. Although the dimension of the matrix formed by the signals for each direction was reduced by singular value decomposition (SVD) in [23], and the local maxima of the signal energies were selected as the target source channels, selection of the maximum root-mean-square (RMS) energy is mostly based on a closed-form energy comparison, and preprocessing stages such as SVD and highest-singular-value searching are necessary to estimate the desired direction before source separation. Therefore, the computational burden of this method based on dimension reduction is still quite large. 2) The dynamic properties of acoustic sources have not been considered in traditional deterministic techniques, especially when the acoustic source is moving while speaking. Generally, a source mixture is divided into many time blocks before separation, and the direction of each time block is calculated individually. Besides the huge computational burden, the result of direction estimation is easily influenced by instantaneous noise owing to the lack of relationship between continuous time blocks. While the direction of each time block is logically related to that of its neighboring time blocks, and if the motion velocity of a source is known, it is possible to predict and track its source-directions of the next time blocks. Therefore, direction tracking is important self-correction ability for BSS based on deterministic techniques.

On the other hand, moving speech tracing by a microphone array on a mobile robot has become an area of research in recent years. In [24], the researchers studied the localization of simultaneous moving sound sources using a frequency-domain steered beamformer approach and showed that it located moving sound sources for a mobile robot. Because a direction search on a spherical grid is needed, it is difficult to process real-time signals. Nakadai et al. [25] implemented a real-time tracking system for multiple speakers by integrating audio and visual signals for robots, but the auditory localization by a pair of microphones was poor. Asoh et al. developed a system that tracks human speech signals using a particle filter [26]. However, it did not run in real time on conventional hardware. In 2005, Muraseet et al. proposed multiple moving speech tracking by a microphone array on a mobile robot based on Kalman filters with different history lengths, and their method can overcome the disambiguation problem of crossing speakers or the occurrence of approaching-then-leaving speakers [27]. However, they used eight-channel microphones installed at different positions on a mobile robot to obtain the minimum difference between estimated and observed positions; therefore, the power spectrum of the separated sound should be calculated during estimation, and the application of this algorithm is limited. In order to achieve accurate real-time BSS, a dynamic moving-direction tracing method for a microphone array is necessary before source separation.

In this paper, a dynamic source-direction prediction and tracking method based on a Kalman filter and intensity vector statistics is proposed for real-time BBS. Our present approach is novel in several ways and provides a mathematical relation between source-direction prediction and source separation. Firstly, the basic principles of acoustic source separation based on intensity vector statistics are introduced, and the relationship between source-direction and source separation is developed. Secondly, Kalman filter based motion prediction is theoretically analyzed, and the motion of a source is modeled as a noise-driven position integrator with enough samples in the time domain. Thirdly, a dynamic source-direction prediction method for real-time blind source separation based on a Kalman filter is proposed to predict the directions of a time sequential signal, taking into account self-correction ability and the three-sigma rule. Because the proposed method uses a Kalman filter to estimate the velocity of a time block and predict the source-direction of the next time block, the repeated circular searching process for blind source localization is omitted. Finally, extensive experiments are performed with three-source convolutive mixtures of speeches in English and Chinese, whose direction varies in linear and nonlinear motions, and the results show that our proposed method is a promising source separation method for real-time applications.

This paper is organized as follows. In Section 2, the convolutive mixture signals captured by a coincident array are formulated and the basic principle of the acoustic source separation based on intensity vector statistics is presented. Then, the basic theory of Kalman filter is introduced, and a real-time BSS based on dynamic source-direction prediction is proposed in Section 3. Section 4 describes the experimental test conditions and provides the obtained results. Section 5 concludes the paper.

2. Acoustic source separation based on intensity vector statistics

The mixture sounds are recorded by M omnidirectional microphones positioned arbitrary on a plane as shown in Fig. 1, where p_i denotes the pressure signal recorded by the ith omnidirectional microphone; d_m denotes the distance between the mth microphone and the center of microphones array; φ_m denotes the angle of the mth microphone respect to the center of the array, θ_n is the direction of the nth source respect to the center of the microphone array; $d_m\cos(\varphi_m-\theta_n)$ denotes the wave path difference between the ith microphone and the center of the microphones array.

In the time domain, the pressure signal recorded by the mth microphone for N sources can be written as,

$$p_m(t)=h_{mn}(t)s_n(t) \tag{1}$$

where $h_{mn}(t)$ is the time representation of the transfer function from the nth source signal to the mth microphone pressure signal; $s_n(t)$ is the time representation of the nth original source.