Available online at www.sciencedirect.com
ScienceDirect
Signal Processing 87 (2007) 1833–1847
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors

Shoko Araki[a,b,*], Hiroshi Sawada[a], Ryo Mukai[a], Shoji Makino[a,b]

[a] *NTT Communication Science Laboratories, NTT Corporation, 2–4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan*
[b] *Graduate School of Information Science and Technology, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo-shi, Hokkaido 060-0814, Japan*

Received 31 July 2006; received in revised form 7 February 2007; accepted 9 February 2007
Available online 1 March 2007

Abstract

This paper presents a new method for blind sparse source separation. Some sparse source separation methods, which rely on source sparseness and an anechoic mixing model, have already been proposed. These methods utilize level ratios and phase differences between sensor observations as their features, and they separate signals by classifying them. However, some of the features cannot form clusters with a well-known clustering algorithm, e.g., the *k*-means. Moreover, most previous methods utilize a linear sensor array (or only two sensors), and therefore they cannot separate symmetrically positioned sources. To overcome such problems, we propose a new feature that can be clustered by the *k*-means algorithm and that can be easily applied to more than three sensors arranged non-linearly. We have obtained promising results for two- and three-dimensionally distributed speech separation with non-linear/non-uniform sensor arrays in a real room even in underdetermined situations. We also investigate the way in which the performance of such methods is affected by room reverberation, which may cause the sparseness and anechoic assumptions to collapse.

Keywords: Blind source separation; Sparseness; Clustering; Normalization; Binary mask; Speech separation; Reverberation

1. Introduction

Blind source separation (BSS) [1] is an approach for estimating source signals that uses only the mixed signal information observed at each sensor. The BSS technique for speech dealt with in this paper has many applications including hands-free teleconference systems and automatic conference minute generators.

Two approaches have been widely studied and employed to solve the BSS problem; one is based on independent component analysis (ICA) (e.g., [2]) and the other relies on the sparseness of source signals (e.g., [3]). Recently, many ICA methods have been proposed even for the convolutive BSS problem [2,4–10]. ICA works well even in a reverberant condition when the number of sources N is less than or equal to the number of sensors M. On the other hand, the sparseness-based approaches are attractive because they can handle the underdetermined problem, i.e., $N>M$.

The sparseness-based approaches can be divided into two main categories. One method is based on

*Corresponding author. NTT Communication Science Laboratories, NTT Corporation, 2–4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan. Tel.: +81 774 93 5319; fax: +81 774 93 5158.
E-mail address: shoko@cslab.kecl.ntt.co.jp (S. Araki).

0165-1684/$ - see front matter
doi:10.1016/j.sigpro.2007.02.003

MAP estimation, where the sources are estimated after mixing matrix estimation [11–17], and the other extracts each signals with time-frequency binary masks [3,18–20]. The former method includes mixing matrix estimation and l_1-norm minimization in the frequency domain (i.e., for complex numbers), both of which still present difficulties [16]. The latter, the binary mask approach, has the advantage of being implemented in real time [21]. In this paper we focus on the binary mask approach.

In the binary mask approach, we assume that signals are sufficiently sparse, and therefore, we can assume that at most one source is dominant at each time–frequency slot. If this assumption holds, a histogram of the level and frequency normalized phase differences between two sensor observations has N clusters [3,18,20]. Because an individual cluster in the histogram corresponds to an individual source, we can separate each signal by selecting the observation signal at time–frequency points in each cluster with a binary mask. The best-known approach may be the Degenerate Unmixing Estimation Technique (DUET) [3,18,21].

Previously, such clustering was performed manually [3,18], by using kernel density estimation [20], or with an ML-based gradient method [21]. On the other hand, if clustering could be performed with a well-known algorithm such as the k-means clustering or hierarchical clustering [22], the clustering will be automated and simplified. To employ a widely utilized clustering algorithm such as the k-means, we should be careful about the variances of multiple variables, in this case the level ratios and phase differences. However, frequency normalization of the phase difference, which is important in terms of avoiding the permutation problem among frequencies [16,17], sometimes makes the phase difference much smaller than the level ratio as shown in Section 3.2. Such different variances between the features make clustering with the k-means difficult. This is the prime motivation for this work.

Our second motive is to employ more than three sensors arranged two- or three-dimensionally, which could have a non-linear/non-uniform alignment. Only a few authors have generalized [16,17,23] a method for more than two sensors. Authors of [23] used up to eight sensors, however, their sensors were still linearly arranged. The paper [24] has already tried a multichannel DUET (DESPRIT) by combining the sparse assumption and the Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT); however, their method still limits the array shape: a linear array or two sets of congruent arrays. A two-sensor system and a linear sensor array limits the separation ability on a two-dimensional half-plane, e.g., the previous methods cannot separate sources placed in a mirror image arrangement. To allow the free location of sources, we need more than three sensors arranged two- or three-dimensionally.

Based on these two motivations, we propose a new binary mask approach MENUET (Multiple sENsor dUET), which employs the well-known k-means clustering algorithm. As a feature, our method utilizes the level ratios and phase differences between multiple observations. To realize level ratio and phase difference variances of a comparable level, we propose a way of weighting the phase term for successful clustering. Moreover, our proposed method does not require sensor location information. This allows us to employ freely arranged multiple sensors easily. Therefore, the proposed method can separate signals that are distributed two- or three-dimensionally. Our previous paper, [16], utilized a two-dimensional sensor array to test the MAP approach proposed in [16]. However, that work did not employ the frequency normalization, and therefore, suffered from the abovementioned permutation problem. On the other hand, in this paper, we employ appropriate frequency normalization for the k-means algorithm. Moreover, we also apply our proposed method to a three-dimensional sensor array, and describe the result.

An additional contribution of this paper is that it undertakes an investigation of the separation performance in real world acoustic environments. Both our proposed method and previous methods employ assumptions of source sparseness and anechoic mixing (i.e., a simple attenuation and delay model for a room impulse response). Such assumptions can easily be affected by reverberation. We show how the performance is affected when the problem does not satisfy the assumed conditions.

The organization of this paper is as follows. Section 2 presents the basic framework of the binary mask-based BSS method. In Section 3, we describe some features for clustering, and test how each feature will be clustered by the k-means clustering algorithm. In Section 4, we propose a novel method MENUET, which includes the estimation of geometric features from multiple sensor observations. Our proposed feature is suitable for k-means clustering. Section 5 reports some experimental results obtained with non-linearly arranged sensors