Optim Eng (2009) 10: 313–330
DOI 10.1007/s11081-008-9060-4

Beamspace blind signal separation for speech enhancement

Siow Yong Low · Ka-Fai Cedric Yiu · Sven Nordholm

Received: 4 October 2008 / Accepted: 20 October 2008 / Published online: 6 November 2008

Abstract Signal processing methods for speech enhancement are of vital interest for communications equipments. In particular, multichannel algorithms, which perform spatial filtering to separate signals that have overlapping frequency content but different spatial origins, are important for a wide range of applications. Two of the most popular multichannel methods are blind signal separation (BSS) and beamforming. Briefly, (BSS) separates mixed sources by optimizing the statistical independence among the outputs whilst beamforming optimizes the look direction of the desired source(s). However, both methods have separation limitations, in that BSS succumbs to reverberant environments and beamforming is very sensitive to array model mismatch. In this paper, we propose a novel hybrid scheme, called beamspace BSS, which is intended to compensate the aforementioned separation weaknesses by jointly optimizing the spatial selectivity and statistical independence of the sources. We show that beamspace BSS outperforms the separation performance of the conventional sensor space BSS significantly, particularly in reverberant room environments.

Keywords Blind signal separation · Beamspace · Speech enhancement · Microphone arrays

K.F.C. Yiu is supported by RGC Grant PolyU. 7191/06E and the research committee of the Hong Kong Polytechnic University.

S.Y. Low · S. Nordholm
Western Australian Telecommunications Research Institute, Crawley, WA 6009, Australia

K.F.C. Yiu (✉)
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China
e-mail: macyiu@polyu.edu.hk

1 Introduction

The cocktail party effect which was coined by Cherry in the early fifties illustrates human's ability to focus on a specific talker in a multi-talker situations (Cherry 1953). Part of the explanation for this focussing capability lies in the spatial sampling performed by the two human ears. This spatial diversity makes use of the fact that the origins of the desired and interfering signals originate from different locations in space. Similarly, speech separation or extraction may be realized electronically by sampling in space, i.e., having more than one microphone. Two of the most popular multi-channel techniques are blind signal separation (BSS) and beamforming (Brandstein and Ward 2001; Haykin 2000; Benesty et al. 2005). Their popularity is evident from the numerous applications where these technologies fit in, namely, speech enhancement for hands-free communication systems, speech recognizers, hearing aids, to name a few.

The BSS as the name implies "blindly" recovers a set of unobserved signals from several observed mixtures (Cardoso 1998; Haykin 2000). The term "blind" refers to the fact that no information about the array geometry and source localization is needed in its formulation. As blind as BSS may seem, it does however make the assumption that the signals (to be separated) are statistically independent. This is a plausible assumption as speech signals originate from different sources. With this in mind, BSS separates the mixed sources by optimizing the statistical independence among the outputs. However, as pointed out by Araki et al. (2003), there is a fundamental limitation in the BSS separation quality especially in a reverberant environment. This is because BSS primarily separates signals due to the direct path and not the reflected paths.

Beamforming, on the other hand, requires array geometry and source localization information to form a beam to extract the source of interest. Contrary to optimizing the statistical independence in BSS, beamforming optimizes the spatial selectivity such that it has a maximal response on the desired (predefined) direction. A well known method is the generalized sidelobe canceller (GSC) where the signal enhancement is performed in two stages via two paths (Griffiths and Jim 1982). Firstly, the upper path points a beam towards the signal of interest. Conversely, the lower path blocks the desired signal and consists of mainly the undesired signals. Finally, the output of the upper path is then adaptively decorrelated from the outputs in the lower path. Nevertheless, the GSC succumbs to model mismatch which causes leakage or cross-talk in the overall output (Brandstein and Ward 2001). Alternatively, calibrated methods can be used to design the beamformer to compensate for any model mismatch (Nordholm et al. 1999). By doing so, all the information about the desired signal location and microphone positions are readily incorporated in the design. However, the calibration method only performs optimally when the desired signal matches the calibrated information. In most practical applications, the actual source location can never be determined exactly.

This contribution exploits the synergy between beamforming and BSS by introducing a new class of hybrid algorithm for signal separation called beamspace BSS, where the combined method complements the weaknesses of both beamforming and BSS alone. This is because the poor separation performance of the BSS in reverberant environments is compensated by the beamspace, which deflates the reflected paths. Moreover, the poor separation performance of the beamformer due to signal mismatch or miscalibration is compensated by the separation capability of the BSS. The proposed beamspace BSS is justified by a derivation of a comparison of performance measures, which shows that the error measure of beamspace BSS is lower to that of the element-space BSS. Further, through statistical analysis, the convergence behaviors and stability bounds for the learning functions of the beamspace BSS and the conventional element-space BSS are derived, analyzed and compared. Note that, element-space BSS refers to the conventional BSS approach whereby the observed signals from the microphones are processed with no a priori information about the sources' locations and geometry. It is shown that beamspace BSS has faster convergence and larger stability range than the element-space BSS. Experimental results with different reverberation time (RT60) demonstrate that the proposed beamspace BSS agrees with the theoretical findings and outperforms the element-space BSS.

The paper is organized as follows. Firstly, Sect. 2 gives the problem formulation. Following that, Sect. 3 explains the idea behind beamspace processing. Next, the idea of beamspace BSS is presented in Sect. 4. This section describes both BSS and beamspace BSS. Also, it provides a detailed proof, which shows the error measures for beamspace BSS is lower or equal to the element-space BSS. Section 5 analyzes and compares the convergence behaviors and stability bounds for the beamspace BSS and element-space BSS learning functions. Then, experimental settings and results are presented in Sect. 6. Finally, Sect. 7 concludes the paper.

2 Problem formulation

Consider an L-element linear array that receives a mixture of M sources originating from different spatial locations. The mixing model for frequency ω and time instant k can be written as

$$\mathbf{x}(\omega, k) = \mathbf{A}(\omega)\mathbf{s}(\omega, k) \tag{1}$$

where $\mathbf{A}(\omega)$ is a mixing matrix defined as

$$\mathbf{A}(\omega) = \begin{bmatrix} a_{11}(\omega) & \cdots & a_{1M}(\omega) \\ \vdots & \ddots & \vdots \\ a_{L1}(\omega) & \cdots & a_{LM}(\omega) \end{bmatrix} \tag{2}$$

in which $a_{lm}(\omega)$ is the response scalar from source m to element l at frequency ω. The source vector $\mathbf{s}(\omega, k)$ is

$$\mathbf{s}(\omega, k) = \left[s_1(\omega, k), \ldots, s_M(\omega, k)\right]^T \tag{3}$$

where $(\cdot)^T$ denotes transposition. The objective here is to unmix the mixtures so as to best recover $\mathbf{s}(\omega, k)$ from $\mathbf{x}(\omega, k)$.

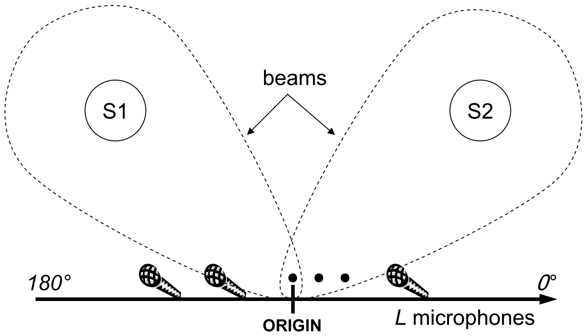


Fig. 1 An illustration of the beamspace, where two directional beams are formed towards the sectors of interest

3 Beamspace processing

3.1 Overview

Beamspace transformation is a dimension reduction method that focuses on a limited angular sector of the possible source(s) directions (Linebarger et al. 1995; Buckley and Xu 1990). This means that the sensor space data is mapped into a lower dimensional space by using a linear transformation before any desired signal processing algorithm. For instance, if the signals are known to be within a sector in space, the data can be reduced in such a way that information from that sector is retained whilst eliminating information from the other sectors. In other words, beamspace transformation can be viewed as a set of overlapping directional beams, which spans the entire direction of interest where each beam passes the sources in their respective directions.

Assuming that the location of the sources are known to be in some sectors or regions in space, beams can then be formed to point at those sectors (note that the sources' locations only need to be known approximately). This means that BSS will only separate the outputs of the multiple beams, which is collectively known as beamspace. As such, beamspace transformation does not deflate the signal subspace, but the noise/interference subspace, i.e., it retains the information from the sectors of interest and eliminates information from the other sectors. In other words, the sensor space is deflated via a beamspace transformation by forming beams towards the sources' directions or sectors (Lee and Wengrovitz 1990; Rodriguez et al. 2006).

To explain further, consider Fig. 1, which illustrates two directional beams that are formed towards the sectors of interest. Clearly, each of the beamspace will "pass" the source(s) that is in its path and suppress sources that are outside its path. As such, beamspace provides a "cleaner output" for the BSS to process. Beamspace processing can be viewed as a pre-processor in array processing, which reduces the dimensionality of the problem and consequently improves computational savings, source resolution and robustness against sensitivity (Hassanien et al. 2004; Tian and Van Trees 2001; Amin and Bhalla 1998). Typically, beamspace transformation involves a set of multiple beams that spans the entire region of interest. Each of the beams focuses on a limited angular sector, i.e., a directional beam, which has a unity gain from a desired angular sector and a lower gain elsewhere.

3.2 Beamspace data

Let B be the number of directional beams. The outputs of the directional beams at frequency ω can be expressed as

$$\mathbf{y}(\omega, k) = \mathbf{V}^H(\omega)\mathbf{x}(\omega, k) \tag{4}$$

where $\mathbf{V}(\omega)$ is the beamspace transformation matrix defined as

$$\mathbf{V}(\omega) = \left[\mathbf{v}_1(\omega) \ \cdots \ \mathbf{v}_B(\omega)\right]. \tag{5}$$

In principle, $\mathbf{V}(\omega)$ consists of a set of basis functions, which represent the space of interest. Here, it is assumed that the regions of interest are spanned by the B angular sectors. It can be observed that each column of the beamspace transformation matrix, $\mathbf{V}(\omega)$ represents a spatial filter pointing to the direction of the b-th sector. Thus, one may geometrically view the beamspace data as spanned by the columns of $\mathbf{V}(\omega)$. Note that if $B < L$, the size of the problem and consequently the required computation is reduced. Since the main focus is not in the design of the transformation matrix, the readers may refer to Brandstein and Ward (2001), Grbić and Nordholm (2002), Nordholm et al. (1999) for more details on designing the beams.

4 Beamspace blind signal separation

4.1 Blind signal separation

In this paper, the second-order BSS reported in Parra and Alvino (2000) is adopted. Briefly, the second-order BSS exploits non-stationarity of the sources to provide more information to separate the sources. For instance, if the sources are non-stationary, then their respective covariances at different time intervals are linearly independent. As such, additional information can be obtained to perform the separation (Parra and Alvino 2000; Weinstein et al. 1993). In this case, the covariance matrix $\mathbf{R}_x(\omega, n)$ of the received data can be estimated for the N number of intervals as

$$\mathbf{R}_x(\omega, n) = \frac{1}{K}\sum_{k=0}^{K-1} \mathbf{x}(\omega, nK + k)\mathbf{x}^H(\omega, nK + k) \tag{6}$$

where $n = 0, \ldots, N - 1$ and K is the length of the interval for estimating the cross covariance matrix. To achieve separation, the N number of covariance matrices in (6) are to be jointly diagonalized. Following the approach in Parra and Alvino (2000), the separation matrix, $\mathbf{W}_{ES}(\omega)$ can be obtained by using a least squares estimate as

$$\hat{\mathbf{W}}_{ES}(\omega) = \arg\min_{\mathbf{W}_{ES}(\omega)} \sum_{n=1}^{N} \|\mathbf{E}(\omega, n)\|_F^2 \tag{7}$$

where

$$\mathbf{E}(\omega, n) = \mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega) - \mathbf{\Lambda}_s(\omega, n) \tag{8}$$

and $\|\cdot\|_F^2$ is the squared Frobenius norm operator. The matrix, $\mathbf{\Lambda}_s(\omega, n)$ is the desired sources covariance matrix. Since the sources are independent, $\mathbf{\Lambda}_s(\omega, n)$ is a diagonal matrix with each diagonal element representing the power of each source.

The estimation of the frequency domain unmixing weights $\mathbf{W}_{ES}(\omega)$, leads to arbitrary permutation of each frequency bin. A simple method to solve this problem is to impose a constraint on the time-domain filter size of the unmixing weights, F such that $\mathbf{W}_{ES}(\tau) = 0, \tau > F$, where $F \ll M$, where M is the number of frequency bins. As shown in Parra and Alvino (2000), the constraint couples the otherwise independent frequencies, which provides a continuity of the spectra, hence reducing the permutation problem. Since permutation problem is not the main concentration in this paper, readers may refer to Sawada et al. (2004), Ikram and Morgan (2002) for further information.

4.2 Beamspace blind signal separation—an improved estimate

Following (7), the least squares estimate of the beamspace BSS is

$$\hat{\mathbf{W}}_{BS}(\omega) = \arg\min_{\mathbf{W}_{BS}(\omega)} \sum_{n=1}^{N} \|\mathbf{E}_{BS}(\omega, n)\|_F^2 \tag{9}$$

where

$$\mathbf{E}_{BS}(\omega, n) = \mathbf{W}_{BS}(\omega)\mathbf{R}_y(\omega, n)\mathbf{W}_{BS}^H(\omega) - \mathbf{\Lambda}_s(\omega, n). \tag{10}$$

Note that, (10) represents the error function in the beamspace and $\mathbf{R}_y(\omega, n)$ is the beamspace covariance information. Equivalently, the covariance information of the beamspace data at time interval n can be expressed as

$$\begin{aligned}
\mathbf{R}_y(\omega, n) &= \frac{1}{K}\sum_{k=0}^{K-1} \mathbf{y}(\omega, nK + k)\mathbf{y}^H(\omega, nK + k) \\
&= \mathbf{V}^H(\omega)\frac{1}{K}\sum_{k=0}^{K-1} \mathbf{x}(\omega, nK + k)\mathbf{x}^H(\omega, nK + k)\mathbf{V}(\omega) \\
&= \mathbf{V}^H(\omega)\mathbf{R}_x(\omega, n)\mathbf{V}(\omega)
\end{aligned} \tag{11}$$

where $\mathbf{V}(\omega)$ is the beamspace transformation matrix defined in (5). If the column vectors of $\mathbf{V}(\omega)$ are orthonormal, i.e., $\mathbf{V}^H(\omega)\mathbf{V}(\omega) = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix, then (11) is a unitary transformation of $\mathbf{R}_x(\omega)$. Note that if $\mathbf{V}(\omega) = \mathbf{I}$, the beamspace covariance matrix reverts back to the sensor space covariance matrix. For the case of $B < L$, $\dim[\mathbf{R}_y(\omega, n)]$ is smaller than $\dim[\mathbf{R}_x(\omega, n)]$, where $\dim[\cdot]$ denotes the dimension of a matrix.

The aim here is to show that the error measure for the beamspaced joint diagonalization technique is upper bounded by the non-beamspaced joint diagonalization i.e., $C_{BS}(\omega) \le C_{ES}(\omega)$ where

$$C_{ES}(\omega) = \sum_{n=1}^{N} \|\mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega) - \mathbf{\Lambda}_s(\omega, n)\|_F^2 \tag{12}$$

and

$$C_{BS}(\omega) = \sum_{n=1}^{N} \|\mathbf{W}_{BS}(\omega)\mathbf{R}_y(\omega, n)\mathbf{W}_{BS}^H(\omega) - \mathbf{\Lambda}_s(\omega, n)\|_F^2. \tag{13}$$

By using the fact that $\|\mathbf{A}\|_F^2 = \mathrm{Tr}\{\mathbf{A}\mathbf{A}^H\}$, where $\mathrm{Tr}\{\cdot\}$ denotes the trace operator, the least square criterion in (12) can be rewritten as

$$\begin{aligned}
C_{ES}(\omega) &= \sum_{n=1}^{N} \|\mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega) - \mathbf{\Lambda}_s(\omega, n)\|_F^2 \\
&= \sum_{n=1}^{N} \mathrm{Tr}\big\{\big[\mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega) - \mathbf{\Lambda}_s(\omega, n)\big] \\
&\quad \times \big[\mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega) - \mathbf{\Lambda}_s(\omega, n)\big]^H\big\} \\
&= \sum_{n=1}^{N} \mathrm{Tr}\big\{\mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega)\mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega) \\
&\quad - \mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega)\mathbf{\Lambda}_s^H(\omega, n) \\
&\quad - \mathbf{\Lambda}_s(\omega, n)\mathbf{W}_{ES}(\omega)\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega) + \mathbf{\Lambda}_s(\omega, n)\mathbf{\Lambda}_s^H(\omega, n)\big\}.
\end{aligned} \tag{14}$$

Assume that the weights are unitary, i.e.,

$$\mathbf{W}_{ES}(\omega)\mathbf{W}_{ES}^H(\omega) = \mathbf{W}_{ES}^H(\omega)\mathbf{W}_{ES}(\omega) = \mathbf{I} \tag{15}$$

and by using the identity $\mathrm{Tr}\{\mathbf{AB}\} = \mathrm{Tr}\{\mathbf{BA}\}$, (14) can be simplified to

$$C_{ES}(\omega) = \sum_{n=1}^{N} \mathrm{Tr}\big[\mathbf{R}_x(\omega, n)\mathbf{R}_x(\omega, n)\big] \tag{16}$$

$$\begin{aligned}
&- 2\mathrm{Tr}\big[\mathbf{R}_x(\omega, n)\mathbf{W}_{ES}^H(\omega)\mathbf{\Lambda}_s^H(\omega, n)\mathbf{W}_{ES}(\omega)\big] \\
&+ \mathrm{Tr}\big[\mathbf{\Lambda}_s(\omega, n)\mathbf{\Lambda}_s^H(\omega, n)\big].
\end{aligned} \tag{17}$$

Let the eigendecomposition of $\mathbf{R}_x(\omega, n)$ be

$$\mathbf{R}_x(\omega, n) = \mathbf{U}_x(\omega, n)\mathbf{\Lambda}_x(\omega, n)\mathbf{U}_x^H(\omega, n) \tag{18}$$

where $\mathbf{U}_x(\omega, n)$ and $\mathbf{\Lambda}_x(\omega, n)$ correspond to the eigenvector and eigenvalue of $\mathbf{R}_x(\omega, n)$, respectively. The eigenvalues in $\mathbf{\Lambda}_x(\omega, n)$ are assumed to be ordered in decreasing order, i.e., $\lambda_{x,1}(\omega, n) \geq \cdots \geq \lambda_{x,L}(\omega, n)$. Substituting (18) into (17) yields

$$\begin{aligned}
C_{ES}(\omega) = \sum_{n=1}^{N} &\mathrm{Tr}\big[\mathbf{U}_x(\omega, n)\mathbf{\Lambda}_x(\omega, n)\mathbf{U}_x^H(\omega, n)\mathbf{U}_x(\omega, n)\mathbf{\Lambda}_x(\omega, n)\mathbf{U}_x^H(\omega, n)\big] \\
&- 2\mathrm{Tr}\big[\mathbf{U}_x(\omega, n)\mathbf{\Lambda}_x(\omega, n)\mathbf{U}_x^H(\omega, n)\mathbf{W}_{ES}^H(\omega)\mathbf{\Lambda}_s^H(\omega, n)\mathbf{W}_{ES}(\omega)\big]
\end{aligned}$$

$$
\begin{aligned}
&+\mathrm{Tr}\big[\mathbf{\Lambda}_s(\omega,n)\mathbf{\Lambda}_s^H(\omega,n)\big] \\
&=\sum_{n=1}^{N}\mathrm{Tr}\big[\mathbf{\Lambda}_x(\omega,n)\mathbf{\Lambda}_x(\omega,n)\big]-2\,\mathrm{Tr}\big[\mathbf{\Lambda}_x(\omega,n)\mathbf{\Lambda}_s^H(\omega,n)\big] \\
&\quad+\mathrm{Tr}\big[\mathbf{\Lambda}_s(\omega,n)\mathbf{\Lambda}_s^H(\omega,n)\big] \\
&=\sum_{n=1}^{N}\left[\sum_{\ell=1}^{L}\lambda_{x,\ell}^2(\omega,n)-2\sum_{\ell=1}^{L}\lambda_{x,\ell}(\omega,n)\lambda_{s,\ell}(\omega,n)+\sum_{\ell=1}^{L}\lambda_{s,\ell}^2(\omega,n)\right] \\
&=\sum_{n=1}^{N}\sum_{\ell=1}^{L}\big[\lambda_{x,\ell}(\omega,n)-\lambda_{s,\ell}(\omega,n)\big]^2 \qquad (19)
\end{aligned}
$$

where $\lambda_{x,\ell}(\omega,n)$ and $\lambda_{s,\ell}(\omega,n)$ are the l-th eigenvalue of $\mathbf{R}_x(\omega,n)$ and $\mathbf{\Lambda}_s(\omega,n)$, respectively.

Similarly, the least square criterion for the beamspace BSS can be derived as

$$
C_{BS}(\omega)=\sum_{n=1}^{N}\sum_{\ell=1}^{B}\big[\lambda_{y,\ell}(\omega,n)-\lambda_{s,\ell}(\omega,n)\big]^2 \qquad (20)
$$

where $\lambda_{y,\ell}(\omega,n)$ is the ℓ-th eigenvalue of $\mathbf{R}_y(\omega,n)$. Since $B<L$, then the beamspace covariance matrix, $\mathbf{R}_y(\omega,n)$ can be regarded as a principal submatrix of the element-space covariance matrix, $\mathbf{R}_x(\omega,n)$, i.e., a unitary transformation of $\mathbf{R}_y(\omega,n)$. Thus, it follows from the Poincaré Separation Theorem (Horn and Johnson 1985) that

$$
\lambda_{x,\ell}(\omega,n)\geq\lambda_{y,\ell}(\omega,n) \qquad (21)
$$

where $l=1,\ldots,N_S$. Here, N_S corresponds to the number of sources in the observation and $N_S\leq B$. Evidently from (21), for every n-th term, there is a corresponding term in $\sum_{\ell=1}^{L}[\lambda_{x,\ell}(\omega,n)-\lambda_{s,\ell}(\omega,n)]^2$, which is greater than or equal to $\sum_{\ell=1}^{B}[\lambda_{y,\ell}(\omega,n)-\lambda_{s,\ell}(\omega,n)]^2$. Thus,

$$
C_{ES}(\omega)\geq C_{BS}(\omega). \qquad (22)
$$

Equation (22) shows that the error measure of the beamspace BSS is upper bounded by the element-space BSS.

5 Analysis of element-space BSS and beamspace BSS

This section aims to analyze and compare the convergence behavior of the element-space BSS and beamspace BSS. Also, the stability regions for the range of step-sizes to provide convergence for both algorithms are derived.

5.1 Convergence analysis

By using the gradient descent approach, the gradient update of the least square cost in (12) can be expressed as

$$\mathbf{W}_{ES}^{(m+1)}(\omega) = \mathbf{W}_{ES}^{(m)}(\omega) - 2\mu \sum_{n=0}^{N-1} \left[\mathbf{W}_{ES}^{(m)}(\omega)\mathbf{R}_x(\omega,n)\mathbf{W}_{ES}^{(m)H}(\omega) - \mathbf{\Lambda}_s(\omega,n)\right]\mathbf{W}_{ES}^{(m)}(\omega)\mathbf{R}_x(\omega,n) \tag{23}$$

where μ is the step-size and m is the iteration index. In the following derivation, it is assumed that the field is homogeneous and that all sources have the same power, then $\mathbf{\Lambda}_s(\omega,n) = S(\omega,n)\mathbf{I}$, where $S(\omega,n)$ is the sources' power constant. Note that, the data can also be pre-whitened to achieve this assumption (Cichocki and Amari 2002). Also, it is assumed that the sources' power constant satisfies, $S(\omega,n) \le \lambda_{y,\ell}(\omega,n)$ and consequently from (21), $S(\omega,n) \le \lambda_{x,\ell}(\omega,n)$. Equation (23) can be re-written as

$$\begin{aligned}
\mathbf{W}_{ES}^{(m+1)}(\omega) &= \mathbf{W}_{ES}^{(m)}(\omega) - 2\mu \sum_{n=0}^{N-1} \left[\mathbf{W}_{ES}^{(m)}(\omega)\mathbf{R}_x(\omega,n)\mathbf{W}_{ES}^{(m)H}(\omega) - S(\omega,n)\mathbf{I}\right]\mathbf{W}_{ES}^{(m)}(\omega)\mathbf{R}_x(\omega,n) \\
&= \mathbf{W}_{ES}^{(m)}(\omega) - 2\mu \sum_{n=0}^{N-1} \left[\mathbf{W}_{ES}^{(m)}(\omega)\mathbf{R}_x(\omega,n)\mathbf{R}_x(\omega,n) - S(\omega,n)\mathbf{W}_{ES}^{(m)}(\omega)\mathbf{R}_x(\omega,n)\right] \\
&= \mathbf{W}_{ES}^{(m)}(\omega) \left\{\mathbf{I} - 2\mu \sum_{n=0}^{N-1} \left[\mathbf{R}_x(\omega,n)\mathbf{R}_x(\omega,n) - S(\omega,n)\mathbf{R}_x(\omega,n)\right]\right\}.
\end{aligned} \tag{24}$$

Similarly, the update equation for the beamspace BSS weight, $\mathbf{W}_{BS}(\omega)$, is given as

$$\mathbf{W}_{BS}^{(m+1)}(\omega) = \mathbf{W}_{BS}^{(m)}(\omega) \left\{\mathbf{I} - 2\mu \sum_{n=0}^{N-1} \left[\mathbf{R}_y(\omega,n)\mathbf{R}_y(\omega,n) - S(\omega,n)\mathbf{R}_y(\omega,n)\right]\right\}. \tag{25}$$

Following the analysis in Douglas and Cichocki (1997), the following mean behavior of the averaged systems in (24) and (25)

$$E\left[\mathbf{W}_{ES}^{(m+1)}(\omega)\right] = E\left[\mathbf{W}_{ES}^{(m)}(\omega)\right] \left\{\mathbf{I} - 2\mu \sum_{n=0}^{N-1} \left[\mathbf{R}_x(\omega,n)\mathbf{R}_x(\omega,n) - S(\omega,n)\mathbf{R}_x(\omega,n)\right]\right\}, \tag{26}$$

$$E\left[\mathbf{W}_{BS}^{(m+1)}(\omega)\right] = E\left[\mathbf{W}_{BS}^{(m)}(\omega)\right] \left\{\mathbf{I} - 2\mu \sum_{n=0}^{N-1} \left[\mathbf{R}_y(\omega,n)\mathbf{R}_y(\omega,n) - S(\omega,n)\mathbf{R}_y(\omega,n)\right]\right\} \tag{27}$$

respectively, are studied. The operator $E[\cdot]$ denotes the statistical expectation. Let $\mathbf{Q}^H(\omega)\mathbf{R}_x(\omega,n)\mathbf{Q}(\omega) = \mathbf{\Lambda}_x(\omega,n)$, $\forall n$. Then by pre- and post-multiplying (24) with $\mathbf{Q}^H(\omega)$ and $\mathbf{Q}(\omega)$, respectively, yields

$$\begin{aligned}
&\mathbf{Q}^H(\omega)E\big[\mathbf{W}_{ES}^{(m+1)}(\omega)\big]\mathbf{Q}(\omega)\\
&\quad= \mathbf{Q}^H(\omega)E\big[\mathbf{W}_{ES}^{(m)}(\omega)\big]\\
&\qquad\times\left\{\mathbf{I} - 2\mu\sum_{n=0}^{N-1}\big[\mathbf{R}_x(\omega,n)\mathbf{R}_x(\omega,n) - S(\omega,n)\mathbf{R}_x(\omega,n)\big]\right\}\mathbf{Q}(\omega)\\
&\quad= \mathbf{Q}^H(\omega)E\big[\mathbf{W}_{ES}^{(m)}(\omega)\big]\mathbf{Q}(\omega)\\
&\qquad- 2\mu\sum_{n=0}^{N-1}\big[\mathbf{Q}^H(\omega)E\big[\mathbf{W}_{ES}^{(m)}(\omega)\big]\mathbf{R}_x(\omega,n)\mathbf{R}_x(\omega,n)\mathbf{Q}(\omega)\\
&\qquad- S(\omega,n)\mathbf{Q}^H(\omega)E\big[\mathbf{W}_{ES}^{(m)}(\omega)\big]\mathbf{R}_x(\omega,n)\mathbf{Q}(\omega)\big].
\end{aligned}\tag{28}$$

By letting $\mathbf{\Lambda}_{ES}^{(m)}(\omega) = \mathbf{Q}^H(\omega)[\mathbf{W}_{ES}^{(m)}(\omega)]\mathbf{Q}(\omega)$ and plugging (18) into (28), (28) reduces to

$$\begin{aligned}
&\mathbf{\Lambda}_{ES}^{(m+1)}(\omega)\\
&\quad= \mathbf{\Lambda}_{ES}^{(m)}(\omega) - 2\mu\sum_{n=0}^{N-1}\big[\mathbf{\Lambda}_{ES}^{(m)}(\omega)\mathbf{\Lambda}_x(\omega,n)\mathbf{\Lambda}_x(\omega,n) - S(\omega,n)\mathbf{\Lambda}_{ES}^{(m)}(\omega)\mathbf{\Lambda}_x(\omega,n)\big]\\
&\quad= \mathbf{\Lambda}_{ES}^{(m)}(\omega)\left\{\mathbf{I} - 2\mu\sum_{n=0}^{N-1}\big[\mathbf{\Lambda}_x(\omega,n)\mathbf{\Lambda}_x(\omega,n) - S(\omega,n)\mathbf{\Lambda}_x(\omega,n)\big]\right\}.
\end{aligned}\tag{29}$$

If the initial condition of the BSS weight is set such that $\mathbf{W}_{ES}^{(0)}(\omega) = c\mathbf{I}$, where c is an arbitrary constant, then the eigenvalues of $E[\mathbf{W}_{ES}^{(m)}(\omega)]$ approximately evolve according to

$$\lambda_{ES,\ell}^{(m+1)}(\omega) = \lambda_{ES,\ell}^{(m)}(\omega)\left\{1 - 2\mu\sum_{n=0}^{N-1}\big[\lambda_{x,\ell}^2(\omega,n) - S(\omega,n)\lambda_{x,\ell}(\omega,n)\big]\right\}\tag{30}$$

where $\lambda_{ES,\ell}^{(m)}(\omega)$ is the ℓ-th eigenvalue of $\mathbf{W}_{ES}^{(m)}(\omega)$. Note that for a uniform convergence, it is required that $\lambda_{ES,\ell}^{(m+1)}(\omega) < \lambda_{ES,\ell}^{(m)}(\omega)$. Likewise, the evolution of the eigenvalues of the beamspace BSS weights is

$$\lambda_{BS,\ell}^{(m+1)}(\omega) = \lambda_{BS,\ell}^{(m)}(\omega)\left\{1 - 2\mu\sum_{n=0}^{N-1}\big[\lambda_{y,\ell}^2(\omega,n) - S(\omega,n)\lambda_{y,\ell}(\omega,n)\big]\right\}\tag{31}$$

where $\lambda_{BS,\ell}^{(m)}(\omega)$ is the ℓ-th eigenvalue of $\mathbf{W}_{BS}^{(m)}(\omega)$. From the evolution of the eigenvalues, the potential convergence speeds between the element-space BSS and the beamspace BSS can be compared. Recall from (21) that $\lambda_{x,\ell}(\omega,n) \geq \lambda_{y,\ell}(\omega,n)$, thus

$$\begin{aligned}\sum_{n=0}^{N-1}\big[\lambda_{x,\ell}^2(\omega,n) - S(\omega,n)\lambda_{x,\ell}(\omega,n)\big] \\ \geq \sum_{n=0}^{N-1}\big[\lambda_{y,\ell}^2(\omega,n) - S(\omega,n)\lambda_{y,\ell}(\omega,n)\big].\end{aligned} \tag{32}$$

From the above, it can be deduced that

$$\begin{aligned}\left|1 - 2\mu\sum_{n=0}^{N-1}\big[\lambda_{x,\ell}^2(\omega,n) - S(\omega,n)\lambda_{x,\ell}(\omega,n)\big]\right| \\ \geq \left|1 - 2\mu\sum_{n=0}^{N-1}\big[\lambda_{y,\ell}^2(\omega,n) - S(\omega,n)\lambda_{y,\ell}(\omega,n)\big]\right|\end{aligned} \tag{33}$$

for the case of $\mu\sum_{n=0}^{N-1}[\lambda_{x,\ell}^2(\omega,n) - S(\omega,n)\lambda_{x,\ell}(\omega,n)] > 0.5$. Thus, from (33), the convergence of $\lambda_{BS}^{(m)}(\omega)$ is faster or equal to that of $\lambda_{ES}^{(m)}(\omega)$, i.e., beamspace BSS converges faster or equal to its counterpart element-space BSS. Note that since the step-size μ is a free parameter, it can be chosen such that the above condition is satisfied. The next section details the stability regions for μ to obtain uniform convergence.

5.2 Stability regions

To obtain a uniform convergence, the evolution of the eigenvalues of the element-space BSS weights in each iteration must be such that $|\lambda_{ES,\ell}^{(m+1)}| < |\lambda_{ES,\ell}^{(m)}|$. This implies from (30) that the following must be satisfied

$$\left|1 - 2\mu_{ES}(\omega)\sum_{n=0}^{N-1}\big[\lambda_{x,\ell}^2(\omega,n) - S(\omega,n)\lambda_{x,\ell}(\omega,n)\big]\right| < 1 \tag{34}$$

where $\mu_{ES}(\omega)$ is the step-size for the element-space BSS and $|\cdot|$ represents the absolute value operator. Convergence of $\lambda_{ES,\ell}^{(m)}$ is thus guaranteed if the step-size is chosen such that

$$0 < \mu_{ES}(\omega) < \frac{1}{\sum_{n=0}^{N-1}[\lambda_{x,\ell}^2(\omega,n) - S(\omega,n)\lambda_{x,\ell}(\omega,n)]}. \tag{35}$$

Similarly, for the beamspace BSS weights to converge, the following must be satisfied

$$\left|1 - 2\mu_{BS}(\omega)\sum_{n=0}^{N-1}\big[\lambda_{y,\ell}^2(\omega,n) - S(\omega,n)\lambda_{y,\ell}(\omega,n)\big]\right| < 1 \tag{36}$$

where $\mu_{BS}(\omega)$ is the step-size for the beamspace BSS. Thus from (36), the stability condition on its step-size is then

$$0 < \mu_{BS}(\omega) < \frac{1}{\sum_{n=0}^{N-1}[\lambda_{y,\ell}^2(\omega, n) - S(\omega, n)\lambda_{y,\ell}(\omega, n)]}. \tag{37}$$

Again, from Poincaré Theorem (see (21)), it can be concluded that the stability range for learning function of the beamspace BSS is larger or equal than the element-space BSS. This is because $\mu_{BS}(\omega)$ is equal or larger than $\mu_{ES}(\omega)$, since the right hand term in (35) is smaller or equal to the right hand term in (37). Hence, the stability region for the step-size of the element-space BSS is upper bounded by the beamspace BSS as

$$0 < \mu_{ES}(\omega) \le \mu_{BS}(\omega). \tag{38}$$

It is interesting to note (38) implies that the learning function for the beamspace BSS can achieve a faster convergence than its element-space counterpart as larger step-sizes can be chosen. The derivation above also shows that the range of optimal step-sizes is approximately inverse to the eigenvalues of the covariance matrices. As beamspace deflates the signal subspace, the stability range for beamspace BSS to converge is increased.

6 Evaluations

6.1 Experimental settings

A separation experiment with two sources was conducted in a room of dimensions 3 m × 4 m × 3 m. The room was simulated by using the image model method with the reverberation time (RT) varying from 100 ms to 500 ms (Peterson 1986). The aim of this experiment was to compare the separation performance of the proposed beamspace BSS and the element-space BSS as reverberation increases. The experimental layout of the room detailing the positions of the sources and microphone array is illustrated in Fig. 2. In this experiment, the location of the sources were assumed to be known and two directional beams based on the calibration method (Nordholm et al. 1999) were designed to point at them (see also Fig. 1).

6.2 Performance measure

The normalized interference suppression, $\mathcal{S}$ in dB is

$$\mathcal{S} = 10\log_{10}\left[\frac{\sum_{m=0}^{\mathcal{M}-1}\hat{P}_{\mathrm{in},n}(\omega_m)}{\sum_{m=0}^{\mathcal{M}-1}\hat{P}_{\mathrm{out},n}(\omega_m)}\right] - 10\log_{10}(\mathcal{C}) \tag{39}$$

where $\hat{P}_{\mathrm{in},n}(\omega_m)$ and $\hat{P}_{\mathrm{out},n}(\omega_n)$ are the spectral power estimates of the reference sensor observation and the output, respectively when the interference is active alone.

Fig. 2 The experimental layout of the room with a linear array and two sources. The position is stated in $x \times y \times z$ coordinates

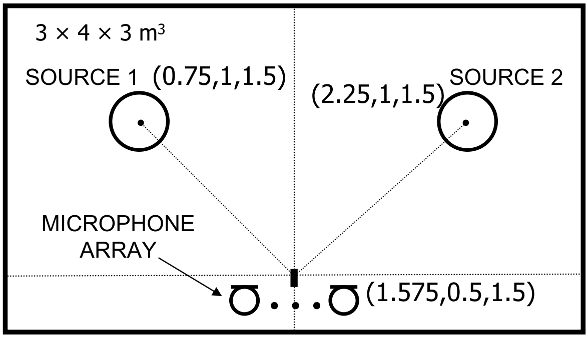
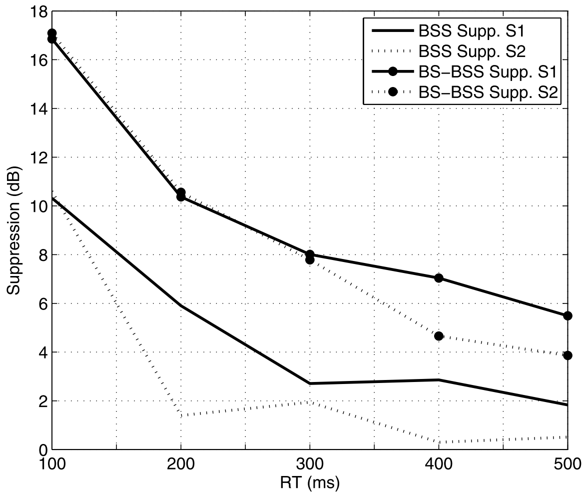


Fig. 3 The suppression measure $\mathcal{S}$ (dB) for source 1 and source 2 as a function of RT with two microphones by using element-space BSS and beamspace BSS



The normalizing constant $\mathcal{C}$ is given as

$$\mathcal{C} = \frac{\sum_{m=0}^{\mathcal{M}-1} \hat{P}_{\mathsf{in},s}(\omega_m)}{\sum_{m=0}^{\mathcal{M}-1} \hat{P}_{\mathsf{out},s}(\omega_m)} \tag{40}$$

where $\hat{P}_{\mathsf{in},s}(\omega_m)$ is the spectral power estimate of one sensor's input observation (which is the reference microphone observation) and $\hat{P}_{\mathsf{out},s}(\omega_m)$ is the spectral power estimate of the structure's output signal.

6.3 Discussion of results

Figures 3, 4 and 5 compare the suppression performance, $\mathcal{S}$, of the element-space BSS and the proposed beamspace BSS for different reverberation time and number of microphones. The results evidently show that the beamspace BSS results in a significant improvement over the element-space BSS in the separation quality for both sources. An average improvement of around 6 dB is achieved for all cases. The separation improvement suggests that the beamspace BSS benefits from a good ini-

Fig. 4 The suppression measure $\mathcal{S}$ (dB) for source 1 and source 2 as a function of RT with three microphones by using element-space BSS and beamspace BSS

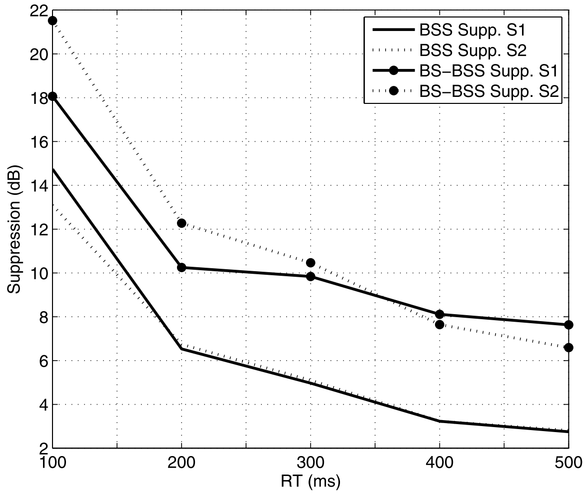
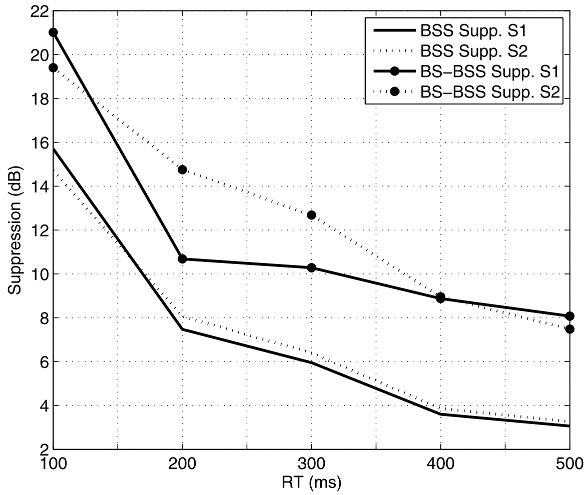


Fig. 5 The suppression measure $\mathcal{S}$ (dB) for source 1 and source 2 as a function of RT with four microphones by using element-space BSS and beamspace BSS



tialization condition, i.e., a deflated signal space for BSS to converge. As reported in Parra and Alvino (2002), Aichner et al. (2002), proper initial conditions typically converge to solutions with consistent permutations for neighbouring frequencies, which results in consistent permutation within large frequency bands. Similarly, beamspace provides well localized observations that correspond to consistent close-by initializations thus reducing permutation ambiguity.

Also, it can be noted from the results that the separation performance of BSS drops as the reverberation time increases. This degradation is attributed to the fact that there are many more sources due to more reflections (Peterson 1986; Chin et al. 2003). Consequently, the problem becomes under-determined and difficult to solve. However, results show that the beamspace BSS obtains a separation improvement even if the reverberation increases compared to the element-space BSS.

Table 1 The error measure values for the element-space BSS and the proposed beamspace BSS for different number of elements and reverberation time

Two elements	Error measure values				
	100 ms	200 ms	300 ms	400 ms	500 ms
Element space BSS	0.1325	0.5332	0.5277	0.5357	0.4027
Beamspace BSS	0.09560	0.1897	0.1165	0.06200	0.06470
Three elements	100 ms	200 ms	300 ms	400 ms	500 ms
Element space BSS	0.3223	0.7193	0.8408	0.7460	0.6624
Beamspace BSS	0.08450	0.2291	0.1349	0.07160	0.06290
Four elements	100 ms	200 ms	300 ms	400 ms	500 ms
Element space	0.3886	0.6413	0.7210	0.6742	0.6435
Beamspace BSS	0.04290	0.1185	0.1045	0.07480	0.06400

This is because beamspace deflates the subspace due to the reverberation components and reduces the dimension of the problem. Interestingly from the results, it can be observed that the beamspace BSS achieves greater separation quality for the case of three and four microphones as the reverberation time grows. This can be explained by the following. Firstly, the increased number of microphones introduces additional free parameters, which may contribute to a greater permutation indeterminacy (Fancourt and Parra 2001). Secondly, the additional free parameters coupled with longer reverberation time will result in a very poor separation performance by the element-space BSS. Here, beamspace BSS becomes advantageous as the free parameters due to more microphones along with the increased complexity of the problem due to reverberation are deflated prior to processing. Subsequently, greater separation problem can be achieved. Note that separation performance of the beamspace BSS is similar for the cases considered.

6.4 Error measure values

Table 1 tabulates the error measure values of the least squares criterion with RT = 300 ms for the case of two microphones, three microphones and four microphones, respectively. The experimental results are in agreement with (22), where the error measure of the beamspace BSS is upper bounded by the element-space BSS. Coupled with the suppression results in Figs. 3, 4 and 5, a lower error suggests a greater suppression capability. The results show that since beamspace BSS has a lower error value compared to the element-space BSS, it will achieve a greater suppression capability. Figures 6(a), (b) and (c) plot the convergence of the error measures for element-space BSS and beamspace BSS with different number of microphones at RT = 300 ms. The plots show that the proposed beamspace BSS achieves a faster convergence to an even lower error value compared to the element-space BSS.

6.5 Real room results

For completeness, an experiment was also conducted in a real room of dimensions $3.5 \times 3.1 \times 2.3$ m^3. The inter-element distance for the linear array was 0.04 m and

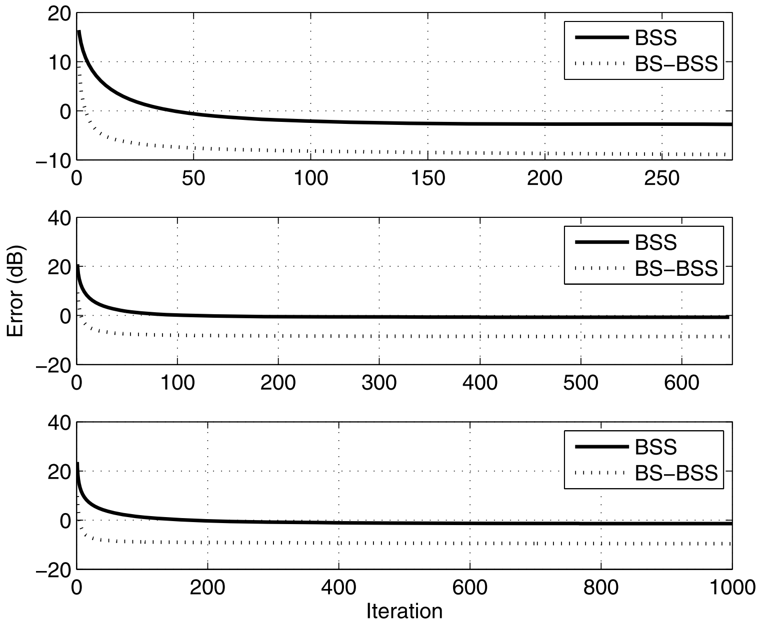
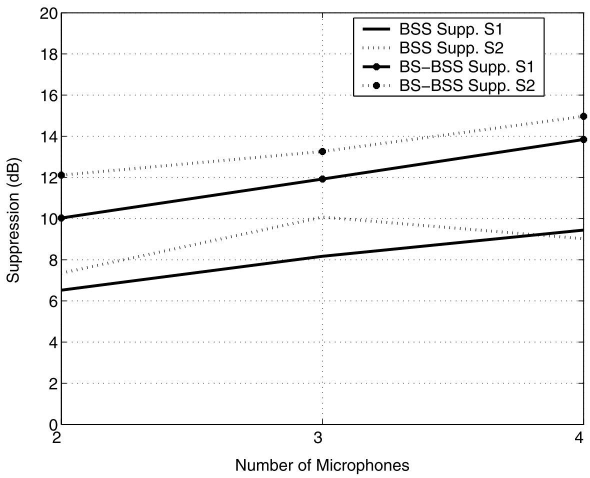


Fig. 6 A comparison of the cost function error curves between the element-space BSS (*solid line*) and beamspace BSS (*dotted line*) for RT=300 ms (**a**) two microphones, (**b**) three microphones and (**c**) four microphones

Fig. 7 The suppression measure $\mathcal{S}$ (dB) for source 1 and source 2 in a real room with different number of microphones by using element-space BSS and beamspace BSS



the RT of the room was measured to be around 200 ms. Figure 7 shows the suppression performance for both the element-space BSS and beamspace BSS for different number of microphones. On average, around 5 dB of suppression improvement was achieved by using beamspace BSS. Also, as explained earlier, beamspace BSS gains steadily from the increment of the number of microphones.

6.6 Remarks

From the results, it is clear that the motivation for the proposed beam- space BSS is twofold. Firstly, beamspace BSS improves the separation performance of BSS even in reverberant environments. As explained in Araki et al. (2003), BSS has its fundamental limitation when it comes to reverberation. This is because in a reverberant environment, there will be contribution due to the reflected paths. Since beamspace transformation is a dimension reduction method (Rodriguez et al. 2006), it readily reduces the dimension of the problem by deflating the subspace due to reverberation components and noise. Thus, the dimension of the separation problem decreases and better performance can be achieved.

Secondly, since each dimension in the beamspace provides a well localized information, beamspace transformation will give BSS a clear direction to work on, i.e, a pre-steered BSS. This means that each of the beamspace provides a good start for the BSS as any source in the beam will be enhanced (well demarcated energy to drive convergence), hence giving a good head start to converge as opposed to having a start from a non-preferred direction. Note that whilst the separation of only two sources is shown, the technique can be extended to arbitrary number of sources.

7 Conclusion

This paper introduced the beamspace BSS and compared the performances of the conventional element-space BSS and the beamspace BSS. Contrary to element-space BSS, beamspace BSS first reduces the dimension of the problem and then separates the lower dimension observation. Analysis indicates that the proposed beamspace BSS has faster convergence and larger stability range for its step-sizes than the element-space BSS. Experimental results in a room with different reverberation time are in good agreement with the theory and reveals improved separation performance compared to the element-space BSS. As a suggestion, if hardware permits, directional microphones can be used instead. As each directional microphone provides a steered response or a directional or localized observation, the directional microphones can be arranged to form the beamspace. Thus, no additional signal processing algorithm is needed for steering and BSS can be readily applied on the observed signals.

References

Aichner R, Araki S, Makino S, Nishikawa T, Saruwatari H (2002) Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming. In: PIEEE workshop on neural networks for signal processing, pp 445–454, September

Amin MG, Bhalla N (1998) Minimum bias spatial filters for beamspace direction-of-arrival estimation. J Franklin Inst 335(1):35–52

Araki S, Mukai R, Makino S, Nishikawa T, Saruwatari H (2003) The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. IEEE Trans Speech Audio Process 11(2):109–116

Benesty J, Makino S, Chen J (2005) Speech enhancement. Springer, Berlin

Brandstein M, Ward D (2001) Microphone arrays: Signal processing techniques and applications. Springer, Berlin

Buckley K, Xu XL (1990) Spatial spectrum estimation in a location sector. IEEE Trans Acoust Speech Signal Process 38(11):1842–1852

Cardoso JF (1998) Blind signal separation: Statistical principles. Proc IEEE 86(10):2009–2025

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25(5):975–979

Chin DN (2003) Blind source separation of convolutive mixtures of speech. In: Kobsa A, Wahlster W (eds) Adaptive signal processing: Applications to the real world. Signals and communications technology. Springer, Berlin

Cichocki A, Amari S (2002) Adaptive blind signal and image processing. Wiley, West Sussex

Douglas SC, Cichocki A (1997) Neural networks for blind decorrelation of signals. IEEE Trans Signal Process 45(11):2829–2842

Fancourt C, Parra LC (2001) The generalized sidelobe decorrelator. In: IEEE workshop on the apps of signal processing to audio and acoustics, pp 167–170, October

Grbić N, Nordholm S (2002) Soft constrained subband beamforming for handsfree speech enhancement. In: IEEE int conf on acoustics, speech and signal processing, pp 885–888, May

Griffiths LJ, Jim CW (1982) An alternative approach to linearly constrained adaptive beamforming. IEEE Trans Antennas Propag 30(1):27–34

Hassanien A, Elkader SA, Gershman AB, Wong KM (2004) Beamspace preprocessing with an improved robustness against out-of-sector sources using second-order cone programming. In: IEEE workshop on sensor array and multichannel signal processing, pp 347–351, July

Haykin S (2000) Unsupervised adaptive filtering: Blind source separation. Wiley, New York

Horn R, Johnson C (1985) Matrix analysis. Cambridge University Press, Cambridge

Ikram MZ, Morgan DR (2002) A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation. In: IEEE int conf on acoustics, speech and signal processing, pp 881–884, May

Lee H, Wengrovitz M (1990) Resolution threshold of beamspace MUSIC for two closely spaced emitters. IEEE Trans Acoust Speech Signal Process 38(9):1545–1559

Linebarger DA, DeGroat RD, Dowling EM, Stoica P, Fudge GL (1995) Incorporating a priori information into MUSIC algorithm and analysis. Signal Process 46(1):85–104

Nordholm S, Claesson I, Dahl M (1999) Adaptive microphone array employing calibration signals: Analytical evaluation. IEEE Trans Speech Audio Process 7(3):241–252

Parra LC, Alvino CV (2002) Geometric source separation: Merging convolutive source separation with geometric beamforming. IEEE Trans Speech Audio Process 6(10):352–362

Parra LC, Alvino CV (2000) Convolutive blind separation of non-stationary sources. IEEE Trans Speech Audio Process 8(3):320–327

Peterson PM (1986) Simulating the response of multiple microphones to a single acoustic source in a reverberant room. J Acoust Soc Am 80(5):1527–1529

Rodriguez A, Baryshnikov BV, Van Veen BD, Wakai RT (2006) MEG and EEG source localization in beamspace. IEEE Trans Biomed Eng 53(3):430–441

Sawada H, Mukai R, Araki S, Makino S (2004) A robust and precise method for solving the permutation problem of frequency domain blind source separation. IEEE Trans Speech Audio Process 12(5):530–538

Tian Z, Van Trees HL (2001) Beamspace MODE. In: Asilomar conf on signals, systems and computers, pp 926–930, November

Weinstein E, Feder M, Oppenheim AV (1993) Multi-channel signal separation by decorrelation. IEEE Trans Speech Audio Process 1(4):405–413