Geometric Source Separation: Merging Convolutive Source Separation With Geometric Beamforming

Lucas C. Parra and Christopher V. Alvino

***Abstract*—Convolutive blind source separation and adaptive beamforming have a similar goal—extracting a source of interest (or multiple sources) while reducing undesired interferences. A benefit of source separation is that it overcomes the conventional cross-talk or leakage problem of adaptive beamforming. Beamforming on the other hand exploits geometric information which is often readily available but not utilized in blind algorithms. In this work we propose to join these benefits by combining cross-power minimization of second-order source separation with geometric linear constraints used in adaptive beamforming. We find that the geometric constraints resolve some of the ambiguities inherent in the independence criterion such as frequency permutations and degrees of freedom provided by additional sensors. We demonstrate the new method in performance comparisons for actual room recordings of two and three simultaneous acoustic sources.**

***Index Terms*—Cross-power spectra, frequency permutation, linear constraints, reverberation, source separation.**

I. Introduction

THE PURPOSE of this work is to merge for the first time adaptive beamforming [1] with convolutive blind source separation of broad-band signals (e.g., [2]–[6]). Both methods are concerned with the problem of optimizing a multichannel filter structure to improve the signals detected with a sensor array such a microphone array or an antenna array. Their goal is to reduce interference of simultaneous sources by optimizing the spatial selectivity of the filter array. While the optimization criteria in source separation reduces interference, adaptive beamforming minimizes signal power without reducing the beam's response for a desired source position. We propose to use the optimization criteria of source separation while constraining the responses of multiple beams based on readily available geometric information. More specifically, we introduce the concept of *geometric source separation* (GSS), which relies on cross-power spectral minimization (sufficient for nonstationary signals [6], [7]), while assuming that sources are localized in space. The method will be demonstrated on the separation of acoustic sources from multiple microphones in a reverberant environment. Multimicrophone speech enhancement is useful in a variety of applications such as automatic speech recognition, acoustic surveillance, or hands-free telephony.

The main limitation of beamforming is cross-talk. In beamforming the filter coefficients are optimized to produce a spatial pattern with a dominant response for the location of interest. Adaptive beamforming shapes the filter coefficients such that the response is minimized for the positions of interfering signals. In multipath or reverberant environments, however, the interfering signals may reach the sensor array from many directions, and so the optimization often alters the response for the region of interest also [8]. Source separation, in principle, circumvents this cross-talk problem as it generates multiple response pattern jointly optimized to give independent outputs.

The main limitation of source separation is the existence of ambiguities in the independence criterion. Robustness has been limited in the past due to frequency permutation ambiguities which can only be resolved by considering different frequency bands simultaneously [4], [9]. The required polyspectra are hard to estimate and the resulting algorithms are computationally expensive [9], [10]. In addition, increasing the number of sensors, which in principle improves performance, introduces free parameters that are not fully determined by separation criteria. Geometric constraints have the potential to resolve some of the ambiguities of convolutive blind source separation.

In Section II the task of source separation will be presented, independence criteria will be reviewed, and the concept of GSS will be introduced. Section III outlines how geometric information can be expressed through linear constraints, and interprets the results of source separation as geometric beam patterns. We will review the ambiguities of the independence criteria in Section IV and discuss the algorithm design options for GSS in Section V. Finally, in Section VI we will present specific implementation of GSS and experimental results on actual room recordings.

II. Convolutive Source Separation

A. Separation Problem

Consider M uncorrelated sources, $\mathbf{s}(t) \in \mathbb{R}^M$, originating from different spatial locations and $N > M$ sensors detecting signals $\mathbf{x}(t) \in \mathbb{R}^N$. In a multipath environment, each source i couples with sensor j through a linear transfer function $A_{ji}(\tau)$, such that $x_j(t) = \sum_{i=1}^{M} \sum_{\tau=0}^{P-1} A_{ji}(\tau) s_i(t-\tau)$. Using matrix notation and denoting the convolutions by $*$ we can write this briefly as $\mathbf{x}(t) = \mathbf{A}(t) * \mathbf{s}(t)$, or after applying the discrete time Fourier transform

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

Manuscript received January 19, 2001; revised May 31, 2002. This work was supported in part by Thomson Multimedia under Program 2000-007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dirk van Compernolle.

L. C. Parra is with Sarnoff Corporation, Princeton, NJ 08543 USA (e-mail: lparra@sarnoff.com).

C. V. Alvino is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: alvino@ece.gatech.edu).

Digital Object Identifier 10.1109/TSA.2002.803443.

The aim of convolutive source separation is to find filters $W_{ij}(\tau)$ that invert the effect of the convolutive mixing $\mathbf{A}(\tau)$ by producing recovered signals $\mathbf{y}(t)$ such that

$$\mathbf{y}(\omega) = \mathbf{W}(\omega)\mathbf{x}(\omega) \tag{2}$$

corresponds to the original sources $\mathbf{s}(t)$ up to a permutation and scaling. That is

$$\mathbf{y}(\omega) = \mathbf{W}(\omega)\mathbf{A}(\omega)\mathbf{s}(\omega) = \mathbf{PS}(\omega)\mathbf{s}(\omega) \tag{3}$$

where $\mathbf{P}$ represents an arbitrary permutation matrix and $\mathbf{S}(\omega)$ an arbitrary diagonal scaling matrix per frequency. Our notation implies that we are only considering finite impulse response (FIR) filters for separation. Conditions on invertibility of $\mathbf{A}(\omega)$ in the context of convolutive blind source separation are discussed in [11].

B. Blind Separation Criteria

Different criteria for convolutive separation have been proposed [2]–[7], [12]. All criteria can be derived from the assumption of statistical independence of the unknown signals. In fact, typically only pairwise independence of the recovered signals is used. Pairwise independence implies that all cross-moments factor, giving us a number of necessary conditions for the recovered signals

$$\forall t, n, m, \tau, i \neq j: \quad E\left[y_i^n(t) y_j^m(t+\tau)\right] = E\left[y_i^n(t)\right] E\left[y_j^m(t+\tau)\right]. \tag{4}$$

Convolutive separation requires that these conditions be satisfied for multiple delays τ, corresponding to the delays of the filter taps of $\mathbf{W}(\tau)$. For stationary signals, multiple n, m, i.e., higher-order criteria, are required [2]–[4]. For nonstationary signals, multiple t can be used and multiple decorrelation, i.e., $n = m = 1$, is sufficient [6], [7], [13].

C. From Blind to Geometric Source Separation

Multiple decorrelation can be implemented as cross-power minimization to reduce off-diagonal elements of

$$\mathbf{R}_{\mathbf{yy}}(t, \tau) = E\left[\mathbf{y}(t)\mathbf{y}^H(t+\tau)\right] \tag{5}$$

for multiple times t. In [6] this is implemented in the frequency domain by diagonalizing cross-power spectra $\mathbf{R}_{\mathbf{yy}}(t, \omega)$.

Note that from the perspective of beamforming, the filter coefficients in each row of $\mathbf{W}(\omega) = [\mathbf{w}_1(\omega), \ldots, \mathbf{w}_M(\omega)]^H$ represent a different geometric beam pattern (see Section III-B). The advantage of multiple outputs, $\mathbf{y}(t) \in \mathbb{R}^M$, is that one can overcome cross-talk problems of simultaneous sources by jointly optimizing multiple beams to capture each source, while reducing correlation between their outputs.

In contrast, adaptive beamforming relies mostly on a power criterion for a single output $y(t) \in \mathbb{R}$, i.e., a diagonal term of $\mathbf{R}_{\mathbf{yy}}(t, \omega)$. (Cross-power has been used primarily in the context of eigen-analysis which differs from source separation as will be discussed in Section IV-D.) Sometimes power is minimized such as in noise or sidelobe canceling, which aims to adaptively minimize the response to interfering signals [1]. Sometimes power is maximized such as in matched-filter approaches that seek to maximize the response of interest [14]. With simultaneous sources, however, these one-channel power criteria have a serious crosstalk or leakage problem especially in reverberant environments [8].

The advantage of beamforming over source separation lies in its use of geometric information. Information such as sensor positioning or source location is often readily available and can be used to design optimal array responses. In adaptive beamforming it is used to maintain desired geometric response patterns by constraining the filter coefficients during adaptation [15], [1].

We propose to combine blind source separation and geometric beamforming by minimizing cross-power spectra for multiple t, while enforcing geometric constraints used in conventional adaptive beamforming. Geometric information incorporated as constraints or regularizations can, in addition to maintaining desired beam shapes, also reduce the ambiguities of blind separation which will be described in Section IV.

III. Geometric Information

A. Linear Constraints

In conventional geometric and adaptive beamforming, information such as sensor position and source location is often utilized (for a review see [1]). We want to emphasize that geometric assumptions can be incorporated and implemented as linear constraints to the filter coefficients. In a multiple sidelobe canceler (MSC), for example, the response of one of the sensors, say sensors j, is kept constant, which can be expressed as $\mathbf{w}^H(\omega)\mathbf{e}_i = const.$ The elements of the column vector $\mathbf{w}(\omega) \in \mathbb{C}^N$ are the filter to be applied to each sensor, and $\mathbf{e}_i$ is the ith column of the identity matrix. Rather than constraining a sensor, one can also constrain the response of a beamformer for a particular orientation. To see that let us introduce the concept of the array response. If the location and response characteristics of each sensor are known, one can compute the free-field response of a set of sensor and associated beamforming filters $\mathbf{w}(\omega)$. For position $\mathbf{q}$, the phase and magnitude response is given by

$$r(\omega, \mathbf{q}) = \mathbf{w}^H(\omega)\mathbf{d}(\omega, \mathbf{q}) \tag{6}$$

where $\mathbf{d}(\omega, \mathbf{q}) \in \mathbb{C}^N$ represents the phase and magnitude response of the N sensors for a source located at $\mathbf{q}$.

Constraining the response to a particular orientation is simply expressed by the linear constraint, $r(\omega, \mathbf{q}) = \mathbf{w}^H(\omega)\mathbf{d}(\omega, \mathbf{q}) = const.$ This concept is used in the linearly constrained minimum variance (LCMV) algorithm and is also the underlying idea for the generalized sidelobe canceler [15], [16]. To obtain a robust beam, it has also been suggested to require a smooth response around a desired orientation. This again can be implemented by constraining the derivative

$$\frac{\partial}{\partial \mathbf{q}} \mathbf{r}(\omega, \mathbf{q}) = \mathbf{w}^H(\omega) \frac{\partial}{\partial \mathbf{q}} \mathbf{d}(\omega, \mathbf{q}) = 0.$$

To summarize, all these conditions, or a combination of them, can be expressed as linear constraints on $\mathbf{w}(\omega)$.

B. Beam Patterns

For a linear array with omnidirectional sensors and a far-field source (much beyond the ratio of array aperture squared to the wavelength of interest) the sensor response depends approximately only on the angle, $\theta = \theta(\mathbf{q})$, between the source and the linear array

$$\mathbf{d}(\omega, \mathbf{q}) = \mathbf{d}(\omega, \theta) = e^{-\mathbf{i}(\mathbf{p}/c)\omega \sin(\theta)} \tag{7}$$

where $\mathbf{p} = [p_1, \ldots, p_N]$ are the sensor position of the linear array and c is the wave propagation speed.

The matrix $\mathbf{W}(\omega)$ used in source separation produces M output signals. The ith output signal is produced by the N FIR filters in $\mathbf{w}_i^H(\omega)$, the ith row of $\mathbf{W}(\omega)$. In the beamforming literature, each of these sets of N FIR filters is said to produce a beam or beam pattern. Each beam pattern can be viewed by displaying magnitude response as a function of two variables, frequency, ω, and incident angle on the array, θ. Therefore, by plotting the quantity

$$|r(\omega, \theta)| = |\mathbf{w}^H(\omega)\mathbf{d}(\omega, \theta)| \tag{8}$$

we obtain a convenient way to visualize the spatial and frequency response of a given beam, $\mathbf{w}^H$. We will use this to give a geometric interpretation to the resulting beam patterns in the rows of $\mathbf{W}$ and to compare the results of different separation and beamforming algorithms.

IV. Ambiguities of the Independence Criteria for Convolutive Source Separation

In this section we discuss some of the ambiguities of the independence criteria used in convolutive blind source separation. These ambiguities highlight the need to constrain the filter coefficients by introducing geometric information. For a quicker development the reader may directly skip to the description of the new algorithms and performance results Sections V and VI.

A. Scaling and Permutation

The independence criterion (4) has an ambiguity in terms of permutation and scaling; hence, source separation does not aim to identify $\mathbf{s}(t)$ directly and limits itself to recovered signals $\mathbf{y}(t)$ as in (3). The scaling ambiguity applies to each frequency bin, resulting in a arbitrary convolution of each source signal. This reflects the fact that any convolved versions of independent signals remain independent. It is not the goal of source separation nor beamforming to solve this blind deconvolution problem. Furthermore, when defining a frequency-domain independence criterion such as

$$E\left[y_i^n(\omega)y_j^m(\omega)\right] = E\left[y_i^n(\omega)\right]E\left[y_j^m(\omega)\right] \tag{9}$$

there is a permutation ambiguity per frequency for all orders n, m. The criteria (9) are equally well satisfied with arbitrary scaling and assignment of indices i, j to the recovered signals, i.e.,

$$\mathbf{W}(\omega)\mathbf{A}(\omega) = \mathbf{P}(\omega)\mathbf{S}(\omega) \tag{10}$$

where $\mathbf{P}(\omega)$ is now an arbitrary permutation matrix which could be different for *each frequency*. As a result, contributions of a given source may not be assigned consistently to a single recovered signal for different frequency bins. The problem is more severe with an increasing number of sensors as the number of possible permutations increases.

This problem has often been considered an artifact of the frequency-domain formulation of an separation criterion, since there the separation task is decoupled into an individual separation tasks for each frequency bin. We will show in the Appendix, however, that for $n = m = 1$ this ambiguity also applies to the time-domain independence criteria (4). Even for higher orders we will demonstrate through numerical simulations in the Appendix that time-domain criteria do not necessarily guarantee correct permutations.

B. Approaches to Overcome Permutation Ambiguity

Some source separation work in the past has simply ignored the problem. Others have proposed to exploit the continuity in the spectra of the recovered signals [17], or the co-modulation of different frequency bins [10]. A rigorous way of capturing these statistical properties of multiple frequency contributions are polyspectra [4], [9]. However, in practice, it is hard to obtain robust statistics at multiple frequencies, in particular for nonstationary signals such as speech. In addition, the algorithms that consider combinations of frequencies are by nature computationally very demanding [4], [10]. Smoothness constraints on the filter coefficients in the frequency domain have also been proposed [18], [6]. This is equivalent to constraining the length of the filter as compared to the size of the analysis window. However, this limitation on the filter size may not always be reasonable as rather long filters are required in strongly reverberant environments [19].

C. Ambiguity Due to Additional Sensors

In theory only N sensors are needed to separate $M = N$ sources, or to put it differently, N sensors are needed to place $N - 1$ zeros. In practice, however, one may want to use more sensors $(N > M)$ to improve the performance of a real system. If we ignore temporarily the permutation and scaling ambiguity, (10) reads $\mathbf{W}(\omega)\mathbf{A}(\omega) = \mathbf{I}$, where $\mathbf{I}$ represents the identity matrix. This suggests that for a given $\mathbf{A}(\omega)$, the independence criteria leave us with a $(N - M)$-dimensional linear space of equivalent solutions for $\mathbf{W}(\omega)$. In effect, this indicates that there are additional degrees of freedom to shape the beam patterns represented by filters $\mathbf{W}(\omega)$.

D. Subspace Analysis to Reduce Degrees of Freedom

The conventional approach for reducing the degrees of freedom in a linear system is subspace analysis, which will, given sufficient signal power, identify the M-dimensional origins of the N-dimensional sensor readings. Subspace analysis produces a transformation matrix $\mathbf{U}(\omega)$, such that

$$\mathbf{R}_{\mathbf{xx}}(\omega) = \mathbf{U}(\omega)\mathbf{\Lambda}(\omega)\mathbf{U}^H(\omega) \tag{11}$$

where $\mathbf{\Lambda}(\omega)$ is a diagonal matrix, and the transformations are constrained to be orthogonal, $\mathbf{U}^{-1}(\omega) = \mathbf{U}^H(\omega)$. The diagonal

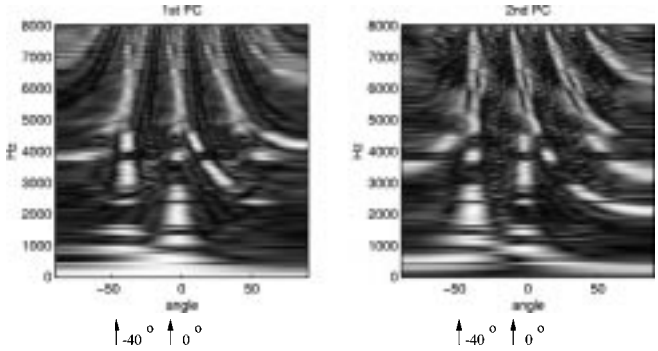

Fig. 1. Beam patterns generated by first two principal components, i.e., $\mathbf{w} = \mathbf{u}_1$, and $\mathbf{w} = \mathbf{u}_2$ in (8). Two sources were located at 0° and −40°. See Section VI-A for details on experimental conditions.

elements represent the eigenvalues corresponding to the column vectors in $\mathbf{U}(\omega)$. The matrix $\mathbf{\Lambda}(\omega)$ also corresponds, in approximation, to the power spectrum $\mathbf{R}_{\mathbf{zz}}(\omega)$ of the variables,[1]

$$\mathbf{z}(\omega) = \mathbf{U}^H(\omega)\mathbf{x}(\omega). \tag{12}$$

Conventionally, it is assumed that the eigenvectors corresponding to the maximal eigenvalues, the principal components, span the signal subspace while the eigenvectors with small eigenvalues correspond to the noise subspace [20]–[22]. That is, the first components of $\mathbf{z}(\omega)$ with largest powers contain the main signals, while the remaining contain mostly sensor noise and other low power signals. For example, the maximum power criterion used for matched-filtering converges to the first principal component $\mathbf{u}_1(\omega)$—the first column in $\mathbf{U}(\omega)$—with maximum eigenvalue. The matched-filter design uses that principal component to generate a beam that points into the direction of maximum power, $z_1(\omega) = \mathbf{u}_1^H(\omega)\mathbf{x}(\omega)$ [14].

Fig. 1 displays in gray scale, the values of (8) for the beam patterns produced by the first two principal component, where white and black correspond to the highest and lowest values of the response magnitude $|r(\omega, \theta)|$. The first principal component filter is shown on the left and the second is shown on the right. The sensors were in a linear array of four microphones, with a total array aperture of 70 cm. The sources were at incident angles of 0° and −40° to the array. Note that the response of the principal component filters is high for the orientations with high signal powers. Note also that in the presence of multiple sources the maximum power does not always correspond to the source of interest or even the same source for all frequency bands. Furthermore, due to the orthogonality assumption, second and higher components do not necessarily have a one-to-one correspondence with the sources.

For instantaneous mixtures, the separation algorithm is often applied only to the first M components. Though a similar approach can be taken in the convolutive case [23], our current experiments indicate that the signal of interest is often found also in higher components ($M+1$, $M+2$, ...). We have so far not been able to demonstrate that subspace analysis can give an additional gain over constraining the filter coefficients based on geometric assumptions.

[1]The factorization $\mathbf{R}_{\mathbf{xx}}(\omega) \approx \mathbf{U}(\omega)\mathbf{R}_{\mathbf{zz}}(\omega)\mathbf{U}^H(\omega)$, with $\mathbf{R}_{\mathbf{zz}}(\tau) = E[\mathbf{z}(t)\mathbf{z}^H(t+\tau)]$ is approximately correct when the filter length $\mathbf{U}(\tau)$ is small compared to the analysis window.

V. Geometric Source Separation—Algorithmic Design Options

There are a number of options for the implementation of the geometric source separation concept. Different choices can lead to different algorithms. Before we report some results on different implementations of the basic idea we want to discuss some of these options.

A. Frequency versus Time Domain

Diagonalization of (5) can be expressed in the time domain as well as in the frequency domain. For efficiency, we suggest implementing the algorithm in the frequency domain by using cross-power spectra $\mathbf{R}_{\mathbf{yy}}(t, \omega)$ as discussed in the following section. Additionally, in the frequency domain the filter parameters are well decoupled; as a result, power normalization per frequency can speed up convergence considerably [24].

B. Cross-Power Estimation

The cross-power spectra must be estimated from the data. In an on-line algorithm one may compute a running estimate of $\mathbf{R}_{\mathbf{yy}}(t, \omega)$ directly from the outputs $\mathbf{y}(t)$ as in [25]. If the filter coefficients change rapidly, however, such an estimate will lag behind. In the past we have used the approximate factorization

$$\mathbf{R}_{\mathbf{yy}}(t, \omega) \approx \mathbf{W}(\omega)\mathbf{R}_{\mathbf{xx}}(t, \omega)\mathbf{W}^H(\omega) \tag{13}$$

which facilitates an efficient on-line algorithm by computing the output cross-correlations using the most recent $\mathbf{W}(\omega)$ and a current estimate of the input cross-power spectra $\mathbf{R}_{\mathbf{xx}}(t, \omega)$ [6], [24]. Factorization (13) is based on the approximation of the linear convolution by the circular convolution. Therefore, it is only accurate when $Q \ll T$, where Q is the filter length and T is the analysis window length. Note also that independent signals will have vanishing empirical cross-power only for sufficiently long estimation times [26], which may go much beyond the stationarity time of the signals. On the other hand, also note that averaging over extended time periods may in effect lead to stationary estimates, which violates the nonstationarity requirement of second-order source separation [6]. This tradeoff has to be determined based on the statistics of the signals.

C. Optimization Criteria

A criterion and algorithm for simultaneous diagonalization of $\mathbf{R}_{\mathbf{yy}}(t, \omega)$ over multiple t must be chosen. Previously, we had proposed to minimize the sum of squares of the off-diagonal elements of $\mathbf{R}_{\mathbf{yy}}(t, \omega)$ [6]

$$J(\mathbf{W}) = \sum_{t,\omega} \alpha(\omega) \left\| \mathbf{R}_{\mathbf{yy}}(t, \omega) - \operatorname{diag}\left[\mathbf{R}_{\mathbf{yy}}(t, \omega)\right] \right\|^2 \tag{14}$$

with the factorization (13), and where $\|.\|$ refers to the Frobenius norm, defined as $\|\mathbf{M}\|^2 = \operatorname{Tr}(\mathbf{M}\mathbf{M}^H)$. For faster convergence using gradient descent, we normalized by the total input power per frequency with $\alpha(\omega) = \sum_t \|\mathbf{R}_{xx}(t, \omega)\|^{-2}$. In [25] a diagonalization criterion is suggested based on the coherence function, which can be written concisely[2] as $\|\operatorname{diag}(\mathbf{R})^{-1/2}\mathbf{R}\ \operatorname{diag}(\mathbf{R})^{-1/2}\|$. The diagonal elements of

[2]In the remainder of this paragraph we abbreviate $\mathbf{R}_{\mathbf{yy}}(t, \omega) = \mathbf{R}$ and omit the sums over t and ω for brevity.

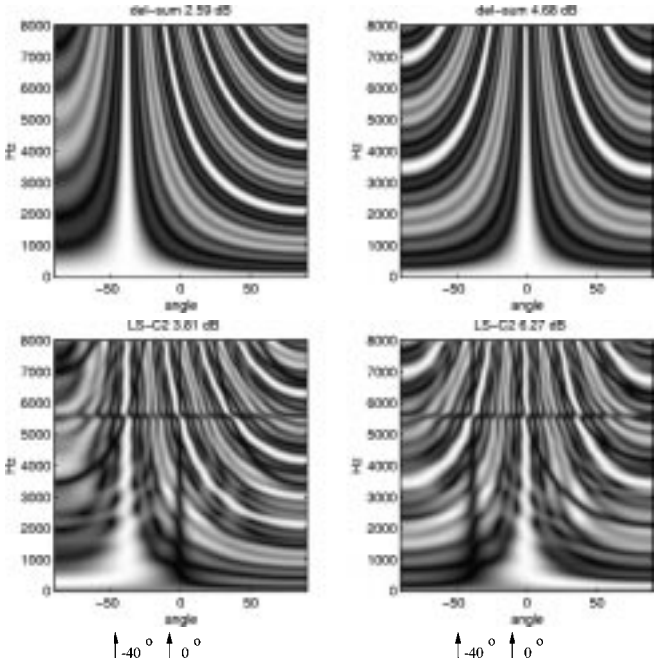

Fig. 2. Responses for data independent beamforming. Results for algorithms delay-sum and LS-C2 are shown. The dB values here and in the following beam plots represent the SIR obtained for separation of two sources. All beam plots shown in this paper assume the same four-microphone configuration corresponding to the recording array.

the coherence function are normalized to 1, which results in a fast converging gradient algorithm. In [27] a diagonalization criterion based on the Hadamard inequality, $\mathrm{Tr}[\log(\mathbf{R})] - \log[\det(\mathbf{R})]$ was proposed. A similar diagonalization criterion, based on the maximum likelihood (ML) cost assuming multivariate Gaussian distributions, was presented in [28], $\mathrm{Tr}[\mathrm{diag}(\mathbf{R})^{-1}\mathbf{R}] - \log\{\det[\mathrm{diag}(\mathbf{R})^{-1}\mathbf{R}]\}$. All of these criteria are to be minimized with respect to the filter coefficients $\mathbf{W}$. Their lower bound of zero is obtained if and only if $\mathbf{R}$ is diagonal.

D. Geometric Constraints

We will assume, without loss of generality, that the sources we are trying to recover are localized at angles $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_M]$ and at sufficient distance for a far-field approximation to apply. Following Section III, the response of the M filters in $\mathbf{W}$ for the M directions in $\boldsymbol{\theta}$ is given by $\mathbf{W}(\omega)\mathbf{D}(\omega, \boldsymbol{\theta})$, where $\mathbf{D}(\omega, \boldsymbol{\theta}) = [\mathbf{d}(\omega, \theta_1), \ldots, \mathbf{d}(\omega, \theta_M)]$. We will consider linear constraints such as

$$\text{C1:} \quad \mathrm{diag}\left(\mathbf{W}(\omega)\mathbf{D}(\omega, \boldsymbol{\theta})\right) = \mathbf{I} \tag{15}$$

or,

$$\text{C2:} \quad \mathbf{W}(\omega)\mathbf{D}(\omega, \boldsymbol{\theta}) = \mathbf{I}. \tag{16}$$

Constraint (15) restricts each filter $\mathbf{w}_i^H(\omega)$—the ith row vector in $\mathbf{W}(\omega)$—to have unit response in direction θ_i. Constraint (16) enforces in addition that each filter has zero response in the direction of interfering signals θ_j, $i \neq j$.

Note that the conventional delay-sum beamformer, given by $\mathbf{w}_i^H(\omega) = \mathbf{d}^H(\omega, \theta_i)$, satisfies constraint C1 strictly. Examples are shown on the top two plots of Fig. 2. The top left and top right plots are the magnitude response, $|r(\omega, \theta)|$, with steering angles $\theta_1 = 0°$ and $\theta_2 = -40°$. Condition C2 requires that $\mathbf{D}(\omega, \boldsymbol{\theta})$ is invertible for the given set of angles. However, this is not always possible. A data independent filter structure that guarantees condition C2 can be computed with a least squares approach and is given by the pseudo-inverse of $\mathbf{D}^H(\omega, \theta)$, or including a regularization term $\beta\mathbf{I}$ to compensate for noninvertibility; the solution is given by $\mathbf{W}(\omega) = \mathbf{D}^H(\omega, \boldsymbol{\theta})[\mathbf{D}(\omega, \boldsymbol{\theta})\mathbf{D}^H(\omega, \boldsymbol{\theta}) + \beta\mathbf{I}]^{-1}$. We denote this solution by LS-C2. The bottom two plots of Fig. 2 show the resulting beam patterns. At the frequencies where the grating lobes[3] of a beam pattern coincide with the interfering angles, $\mathbf{D}(\omega, \boldsymbol{\theta})$ is not invertible (e.g., Fig. 2, bottom at about 5.6 kHz). It is therefore not reasonable to try to enforce C2 as a hard constraint. Rather, as we confirmed in our experiments, it is beneficial to regularize the inverse problem by adding a penalty term of the form $J_{\mathrm{C2}}(\omega) = ||\mathbf{W}(\omega)\mathbf{D}(\omega, \boldsymbol{\theta}) - \mathbf{I}||^2$ to the optimization criterion (14). Note also that power or cross-power minimization will minimize the response at the interference angles. This will lead to an equivalent singularity at those frequencies. It is therefore beneficial to enforce constraint (15) also only as a regularization by using a penalty term of the form $J_{\mathrm{C1}}(\omega) = ||\mathrm{diag}[\mathbf{W}(\omega)\mathbf{D}(\omega, \boldsymbol{\theta})] - \mathbf{I}||^2$.

VI. Experimental Results

Now we present a number of different instantiations of geometric source separation on recorded audio data. We present details of the recordings and data in Section VI-A. In Sections VI-B and VI-C, we will describe the specific algorithms. In Section VI-D, the resulting beam patterns will be discussed and in Section VI-E, a more complete performance evaluation will be given in terms of signal-to-interference ratio (SIR) for two and three simultaneous acoustic sources and various microphone configurations. We defer the details of source localization to Section VI-F.

A. Data and Methods

All recordings used a linear array of cardioid condenser microphones with an aperture of 70 cm in moderately reverberant rooms ($T_{30} \approx 50$ ms). The number of microphones was varied from two to eight. The angles θ_i of the multiple sources were identified based on the response profiles using eight sensors as outlined in Section VI-F. All source locations were reliably identified from the data despite the fact that one source is located outside the recording room; its radiation characteristic is rather diffuse as the signal is radiating through the open door of the room. Fig. 3 shows an overhead view of the experiment setups.

The data was stored and processed at 16 kHz sampling rate and a filter length of $Q = 512$ was mostly sufficient in these experiments. We introduced a delay of $Q/2$ in all filters, which can be simply implemented by introducing a delay in the definition of the array vectors $\mathbf{d}$. Since the resulting filters are rather compact around that central delay, $T = Q$ was sufficient to guarantee an accurate approximation in (13).

[3]Periodic replication of the main lobe due to limited spatial sampling.

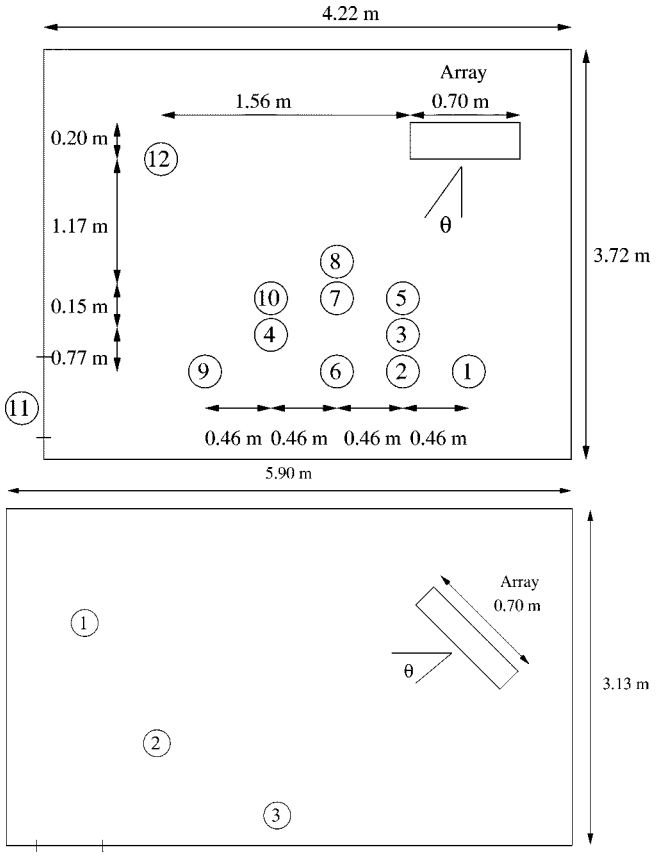

Fig. 3. Overhead views of experimental set-up for two-source recordings (top) and three-source recordings (bottom). Numbered circles indicate various source positions.

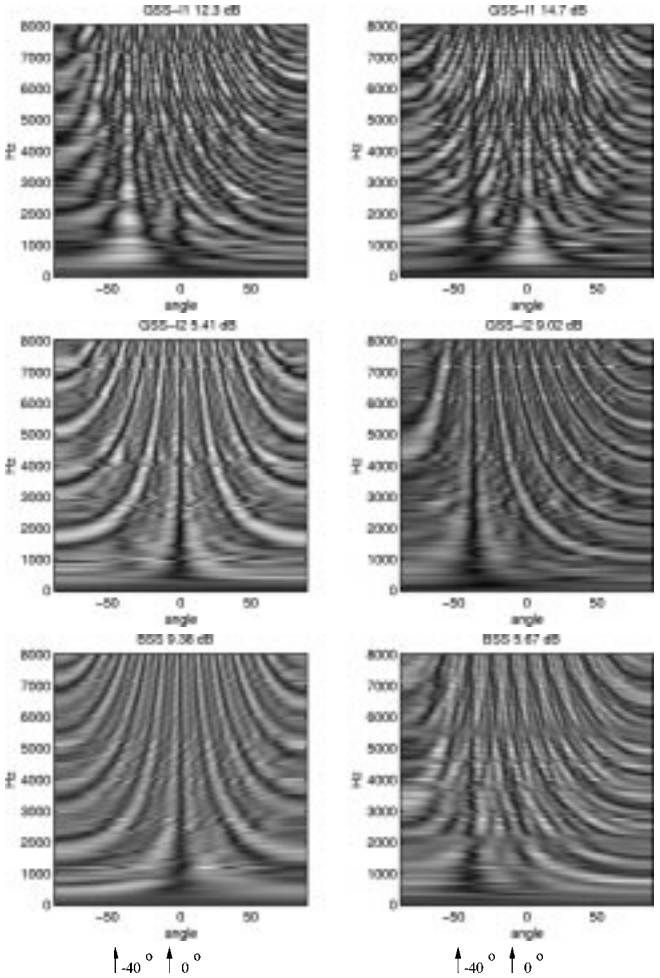

Fig. 4. Resulting responses of geometrically initialized source separation. Results of initializations with delay-sum beamformer (GSS-I1), minima at interferences (GSS-I2), and unit filter (BSS) are shown.

B. Geometrically Initialized Source Separation

Many researchers have obtained reasonable separation results despite ignoring the permutation problem. We argue that this is due to the specific initialization within an optimization procedure. When constraints (9) are expressed as an optimization criterion such as (14), the permutation ambiguity is represented by a set of optimal filters of which only one corresponds to the correct permutation. Since solutions are typically close for neighboring frequencies, close-by initial conditions typically converge to solutions with consistent permutations. Therefore, consistent permutations are expected within large frequency bands, and proper initialization of the filter structure may adequately address the permutation problem.

We optimized criterion (14) with different initializations. In all cases we constrained $\mathbf{w}_i^H(\omega)\mathbf{e}_i = 1$, for all i, to normalize the scale during optimization. We obtained the best performance when initializing with the filter structure corresponding to a delay-sum beamformer pointing to the individual sources. Using the orientations θ_i, we initialized the filter coefficients of the ith beam with

$$\text{I1:} \quad \mathbf{w}_i^H(\omega) = \mathbf{d}^H(\omega, \theta_i). \tag{17}$$

Example beam patterns corresponding to this initialization are shown in Fig. 2 (top). Minimization of (14) with initialization I1 will be referred to as GSS-I1. Alternatively, in GSS-I2 we initialized with beams that place zeros at all orientations of interfering sources. The initialization filters with minimum norm that satisfy those constraints can be computed explicitly with a least-squares (LS) approach resulting in

$$\text{I2:} \quad \mathbf{w}_i^H(\omega) = [\mathbf{e}_i, \mathbf{d}_1, \ldots, \mathbf{d}_{i-1}, \mathbf{d}_{i+1}, \ldots, \mathbf{d}_M]^{\#}\,\mathbf{e}_1 \tag{18}$$

where $\#$ indicates the Hermitian transpose of the pseudo-inverse, and $\mathbf{d}_i$ is the ith column of $\mathbf{D}(\omega, \boldsymbol{\theta})$. In Fig. 4, the resulting response patterns are shown for one example of two simultaneous sources and angles $0°$ and $-40°$. For comparison, we report also the results obtained with a unit filter initialization, $\mathbf{W}(\omega) = \mathbf{I}$, which is the initialization conventionally used in blind source separation, and corresponds to our previous blind source separation (BSS) algorithm [6]. In the performance comparison of Section VI-E we will see that the algorithm GSS-I1 currently gives the best SIR, while the conventional unit filter initialization is not robust across different sources and sensor configurations.

C. Geometrically Constrained Source Separation

In an online implementation of a separation algorithm, the concept of introducing geometric information through an initialization is not feasible as the source positions in the environment may be changing dynamically. It is more realistic to regularize the filters with the geometric information through a

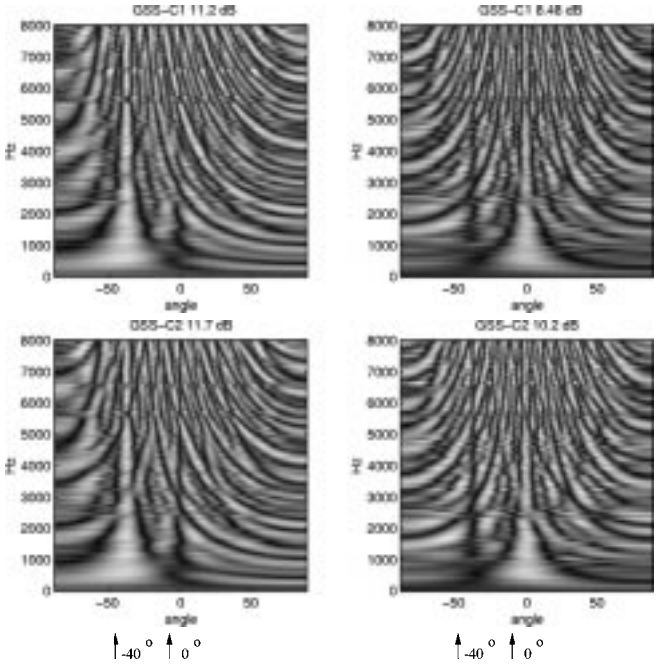

Fig. 5. Responses for geometrically constrained source separation. Algorithms GSS-C1 and GSS-C2 minimize (19) with constraints C1 and C2 respectively.

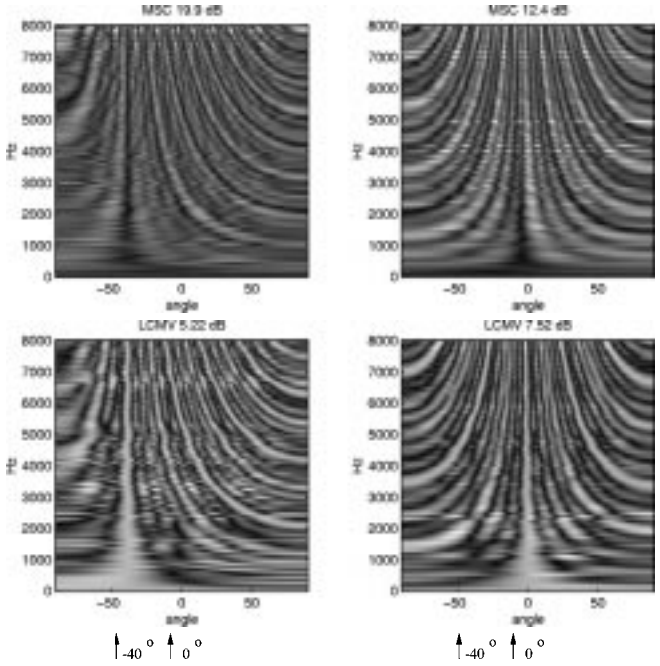

Fig. 6. Responses for multiple sidelobe canceling (MSC) and linearly constrained minimum variance (LCMV).

penalty term. We have further addressed the problem of noninvertibility discussed in Section V-D by introducing a frequency dependent weighting of the penalty term. The idea is to eliminate the constraints from the optimization for those frequency bands in which $\mathbf{D}(\omega, \boldsymbol{\theta})$ is not invertible. A rather straightforward metric for invertibility is the condition number. We therefore weight the penalty term with the inverse of the condition number of $\mathbf{D}$, i.e., $\lambda(\omega) = \{\mathrm{cond}[\mathbf{D}(\omega, \boldsymbol{\theta})]\}^{-1}$, which is zero when $\mathbf{D}(\omega, \boldsymbol{\theta})$ is not invertible and remains bounded otherwise, i.e., $0 \leq \lambda(\omega) \leq 1$. The total cost function including frequency dependent weighting of the geometric penalty term is given by

$$J(\mathbf{W}) + \lambda_0 \sum_{\omega} \lambda(\omega) J_{\mathrm{C1/2}}(\mathbf{W}(\omega)) \tag{19}$$

where $J(\mathbf{W})$ is given by (14), and λ_0 is scaling constant, chosen to be $1/T$, such that the two terms are of the same order of magnitude. Optimization of (19) is referred to as algorithm GSS-C1 or GSS-C2 depending on which penalty term is used. The example of Fig. 5 shows that in algorithm GSS-C1, the penalty term J_{C1} maintains the response of filters i in orientation θ_i. In algorithm GSS-C2 the penalty term J_{C2} minimizes in addition the response for the orientations of the interfering sources.

All GSS algorithms reported here minimize the cost function using conventional gradient descent for complex variables [29]. The gradient of the cost function and two regularizers are given with $\mathbf{E}(t, \omega) = \mathbf{R}_{\mathbf{yy}}(t, \omega) - \mathrm{diag}[\mathbf{R}_{\mathbf{yy}}(t, \omega)]$ by

$$\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}^*(\omega)} = 4 \sum_t \alpha(\omega) \mathbf{E}(t, \omega) \mathbf{W}(\omega) \mathbf{R}_{\mathbf{xx}}(t, \omega)$$

$$\frac{\partial J_{\mathrm{C1}}(\mathbf{W})}{\partial \mathbf{W}^*(\omega)} = 2\,\mathrm{diag}\,[\mathbf{W}(\omega)\mathbf{D}(\omega, \boldsymbol{\theta}) - \mathbf{I}]\,\mathbf{D}^H(\omega, \boldsymbol{\theta})$$

$$\frac{\partial J_{\mathrm{C2}}(\mathbf{W})}{\partial \mathbf{W}^*(\omega)} = 2\,[\mathbf{W}(\omega)\mathbf{D}(\omega, \boldsymbol{\theta}) - \mathbf{I}]\,\mathbf{D}^H(\omega, \boldsymbol{\theta}).$$

In all experiments the cross-power spectra $\mathbf{R}_{\mathbf{xx}}(t, \omega)$ are estimated at five time instances with a time window of about 3 s each, such that a total of about 15 s of data is analyzed. We iterate the gradient descent algorithm 400 to 1600 times depending on the number of microphones used.

D. Discussion of Resulting Beam Patterns

First we would like to qualitatively discuss the resulting response patterns of Figs. 4 and 5, and compare them with the results of some conventional beamforming algorithms as shown in Fig. 6.

Linearly constrained minimum variance (LCMV) minimizes power at all times while strictly constraining the response for a known source location [15]. In this specific LCMV algorithm we have in addition implemented weight decay to avoid spurious maxima in orientations from which no power is detected [30]. The weights are decayed toward the delay-sum beamformer with a regularization term of the form, $\beta \|\mathbf{W}(\omega) - \mathbf{D}^H(\omega, \boldsymbol{\theta})\|^2$. Multiple sidelobe canceling (MSC) uses knowledge of silent periods of each source during which it minimizes power. For the MSC algorithm, the speech segments were labeled by hand. It represents a best-case result which can be obtained only using this additional information and is given here only as a reference.

Figs. 4–6 show in all cases that the main lobe and zeros are at the desired angles consistently across frequency. Without geometric initializations or constraints that is not necessarily the case (see Fig. 1). Algorithms BSS, GSS-I2, and MSC mainly place a zero at the angles of interfering sources. Their response in other directions is not further specified. It is therefore possible that side zeros cancel the sources at bands that are important for the application. The results for GSS-I1, GSS-C1, GSS-C2 exhibit in addition a main lobe in the direction of the corresponding source. For conflicting frequency bands, where

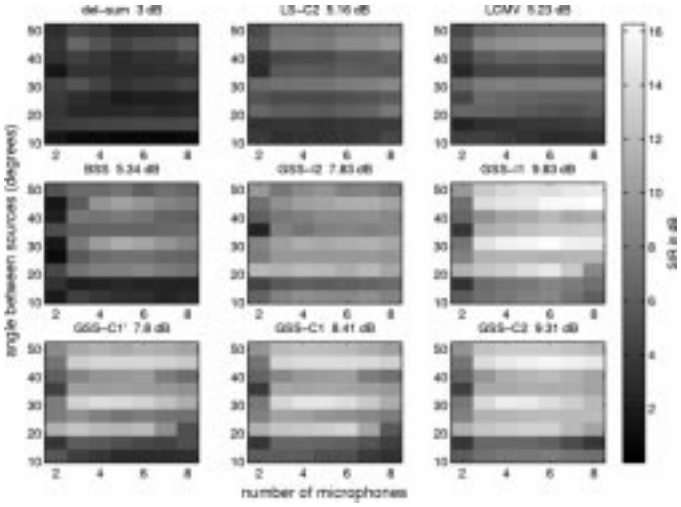

Fig. 7. Performance comparison of the proposed algorithms and geometric beamforming algorithms.

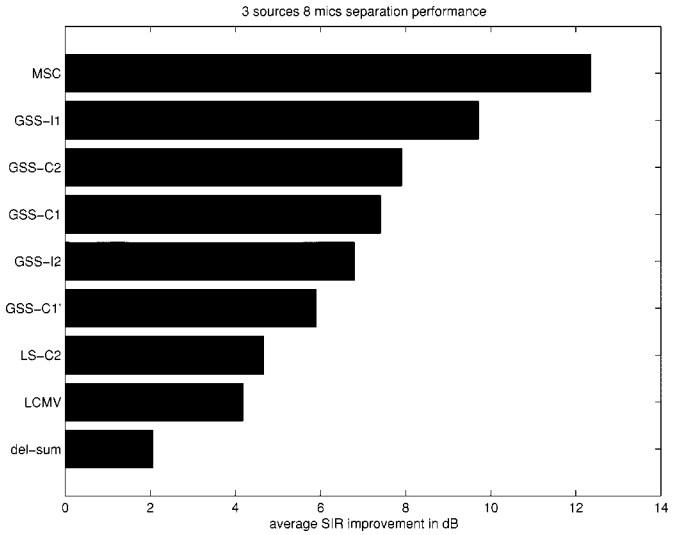

Fig. 8. Performance for the separation of three sources using eight sensors. SIR improvement are averaged over three configurations with angles $-78°$, $-41°$, $0°$; $-60°$, $0°$, $60°$; and $-43°$, $0°$, $36°$. The initial SIR is in average about -3 dB.

a grating lobe coincides with the location of an interfering source, multiple cross-power minimization cancels the main lobe for GSS-I1, while conserving it somewhat for GSS-C1 and GSS-C2, due to the geometric penalty. In LCMV, the power minimization and the strict constraints lead to some discontinuities in the response. In particular, the constraint and power minimization are contradictory criteria when signal power originates from the same direction as the unit-gain constraint. Qualitatively, the results for the data independent LS-C2 algorithm seem to capture both main lobe and zeros in the correct locations. However, its performance in terms of SIR is inferior to the data-adaptive algorithms.

E. Performance Comparison

A systematic performance evaluation of the algorithms for the case of two sources is presented in Fig. 7. We varied source locations and the number of sensors. Grayscale represents SIR performance in dB for the different algorithms. SIR performance, averaged over all positions and number of sensors is given next to the algorithm name. An array with aperture of 70 cm and a variable number of sensors was used (two to eight microphones varied on the horizontal axis). Two sources are located at variable angles. The results are sorted by their relative angle along the vertical axis [12°, 18°, 19°, 25°, 33°, 37°, 38°, 41°, 50° which correspond, in Fig. 3 (top), to locations pairs (1,2), (3,4), (1,5), (1,6), (1,7), (1,8), (1,9), (1,10), and (11,12), respectively.]

The top row of Fig. 7 shows the results for some known beamforming algorithms (delay-sum, LS-C2, LCMV). The center row represents the results for unconstrained multiple cross-power minimization with different initializations including the geometric initialization (BSS, GSS-I2, GSS-I1). The bottom row shows the results for the geometrically constrained separation algorithms (GSS-C1′, GSS-C1, GSS-C2). Algorithm GSS-C1′ is the same as GSS-C1 only with constant penalty term λ. Within each row, the algorithms are sorted by average performance. Comparison of the results for GSS-C1′ and GSS-C1 show the advantage of the frequency dependent weighting of the penalty term. Due to the limited angular resolution, all algorithms perform poorly when the sources are too close.

Fig. 8 reports the performance of separating three sources, two speech sources and babble noise, using eight microphones. As indicated in Section IV-A, the permutation problem becomes worse as the number of sources increases. To our knowledge this is the first report of successful source separation for more than two sources from simultaneous real room recordings. The performance ranking mostly mirrors the results obtained for the separation of two sources.

F. Source Localization

We used a source localization method based on subspace analysis similar to what is used in the MUSIC algorithm [31]. Orthogonality of the eigenvectors guarantees that minor components, $\mathbf{u}_{M+1}(\omega), \ldots, \mathbf{u}_N(\omega)$, do not capture high power signals covered by the principal components. Therefore we use minima in the response of minor components to identify source locations θ. To obtain a more robust angle estimate we combine the response for all frequencies

$$r(\theta) = \sum_{\omega} \sum_{i=M+1}^{N} \left| \mathbf{u}_i^H(\omega) \mathbf{d}(\omega, \theta) \right|. \tag{20}$$

Source locations can be identified in the minima of this response profile as shown in Fig. 9. We defined source locations as the M angles with the smallest values of $r(\theta)$ within a window of 7°.

This profile reflects the spatial distribution of the main power observed by the array, and will depend on the radiation characteristic and environment of the sources. The width of these minima gives us an indication of the adequacy of the localization assumptions used in this work. It is to be compared to the achievable spatial resolution for a given aperture and frequencies of interest. A simple aggregated metric for the spatial resolution of an array is the response profile for a delay-sum beamformer, i.e., $r(\theta') = \sum_{\omega} \sum_{i=1}^{M} |\mathbf{d}^H(\omega, \theta_i) \mathbf{d}(\omega, \theta')|$. In Fig. 9, one can verify that the width of the extrema are in the same

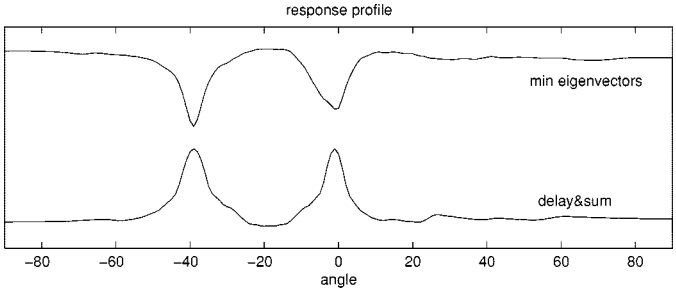

Fig. 9. Top: Response profile (20) of the six smallest eigenvectors for the same signals as shown Fig. 1. The source location can readily be identified. Bottom: Response profile for two delay-sum beamformer pointing to the corresponding angles. The vertical axis has an arbitrary shift and scale.

range, indicating that the sources are localized judged by the resolution of this array.

VII. Conclusion

We have proposed a method that, for the first time, combines convolutive blind source separation with adaptive beamforming. The key assumption of this work is that the sources are localized, at least up to the spatial resolution of a given array. This allows us to formulate geometric constraints on the filter coefficients, much as is done in conventional beamforming. A number of different geometric constraints can be introduced, leading to a class of algorithms we called geometric source separation (GSS). The performance of these GSS algorithms are superior to that of conventional beamforming algorithms. The geometric information was introduced into the algorithm as initialization of the filter parameters and as regularizations using penalty terms. Neither approach is specifically limited to the second-order criterion we used; both can be equally well combined with other higher-order criteria. In this work we have restricted ourselves to linear arrays and source locations specified by azimuth. The discussion extends in a straightforward manner to the more general case of arbitrary sensor arrangements and full 3-dimensional source localization. Though our algorithms rely on specific source locations, they are still blind algorithms since the source locations have been identified from the data.

Appendix

A. Permutation Ambiguity in Second-Order Criteria

Permutation ambiguity affects not only frequency-domain independence criteria but also the corresponding time-domain criteria. Here we show analytically that for $n = m = 1$, the time-domain criterion in (4) is invariant to frequency permutation for wide-sense stationary (w.s.s.) signals. Permutations convert cross-correlation terms into auto-correlation terms. We then use the well-known fact that the frequency components of the autocorrelation function are independent. A similar argument can be made for nonstationary signals when summing (4) over t.

To understand the effect of frequency permutation we express the signals in terms of their T-point frequency representation, $y_i(t) = \sum_{\omega=0}^{T-1} e^{(2\pi/T)\omega t} y_i(\omega)$, whereby

$$E[y_i^*(t)y_j(t+\tau)] = E\left[\sum_{\omega=0}^{T-1}\sum_{\omega'=0}^{T-1} e^{(j2\pi/T)(-\omega t+\omega' t+\omega'\tau)} y_i^*(\omega)y_j(\omega')\right] \quad (21)$$

$$\begin{aligned} &= \sum_{\omega\neq\omega_0}\sum_{\omega'\neq\omega_0} e^{(j2\pi/T)(-\omega t+\omega' t+\omega'\tau)} E\left[y_i^*(\omega)y_j(\omega')\right] \\ &\quad + \sum_{\omega'\neq\omega_0} e^{(j2\pi/T)(-\omega_0 t+\omega' t+\omega'\tau)} E\left[y_i^*(\omega_0)y_j(\omega')\right] \\ &\quad + \sum_{\omega\neq\omega_0} e^{(j2\pi/T)(-\omega t+\omega_0 t+\omega_0\tau)} E\left[y_i^*(\omega)y_j(\omega_0)\right] \\ &\quad + e^{(j2\pi/T)(-\omega_0 t+\omega_0 t+\omega_0\tau)} E\left[y_i^*(\omega_0)y_j(\omega_0)\right] = 0 \end{aligned} \quad (22)$$

provided that both $y_i(t)$ and $y_j(t)$ are zero mean and mutually independent. The purpose of this analysis is to identify the effect of a singly permuted frequency, ω_0, on the second-order statistic. Under permutation of frequency bin ω_0, (22) becomes

$$\begin{aligned} &\sum_{\omega\neq\omega_0}\sum_{\omega'\neq\omega_0} e^{(j2\pi/T)(-\omega t+\omega' t+\omega'\tau)} E\left[y_i^*(\omega)y_j(\omega')\right] \\ &\quad + \sum_{\omega'\neq\omega_0} e^{(j2\pi/T)(-\omega_0 t+\omega' t+\omega'\tau)} E\left[y_j^*(\omega_0)y_j(\omega')\right] \\ &\quad + \sum_{\omega\neq\omega_0} e^{(j2\pi/T)(-\omega t+\omega_0 t+\omega_0\tau)} E\left[y_i^*(\omega)y_i(\omega_0)\right] \\ &\quad + e^{(j2\pi/T)(-\omega_0 t+\omega_0 t+\omega_0\tau)} E\left[y_j^*(\omega_0)y_i(\omega_0)\right]. \end{aligned} \quad (23)$$

The first and fourth terms of (23) are still zero due to the assumption of independence. The second and third terms of (23) are equal to zero for w.s.s. signals, since $E[y^*(\omega)y(\omega')] \propto \mathbf{R}_{yy}(\omega)\delta(\omega-\omega')$ [32]. This shows that criterion (4) is insensitive to frequency permutation, making it an inadequate penalty term to assign the signal consistently across frequencies to the appropriate recovered signal.

B. Effect of Frequency Permutation on Time-Domain Independence Criteria

Here we give numerical evidence that some higher-order time-domain criteria are also adversely affected by frequency permutation. To measure the accuracy of assumption (4) for signals $y_i(t)$ we define

$$\gamma_{y_k,y_l}^{n,m}(\tau) = E\left[y_k^n(t)y_l^m(t+\tau)\right] - E\left[y_k^n(t)\right]E\left[y_l^m(t+\tau)\right]. \quad (24)$$

For an incorrectly permuted signal $x(t)$ and a correctly separated signal $s(t)$ we report the relative amplitudes for specific values of n and m

$$10 * \log_{10}\left(\frac{\sum_\tau \gamma_{x_i,x_j}^2(\tau) + \gamma_{x_j,x_i}^2(\tau)}{\sum_\tau \gamma_{s_i,s_j}^2(\tau) + \gamma_{s_j,s_i}^2(\tau)}\right) \quad (25)$$

as a function of the percentage of incorrectly permuted frequency bins in Fig. 10. For independent signals and perfect estimation, the denominator in (25) vanishes. Nonvanishing values

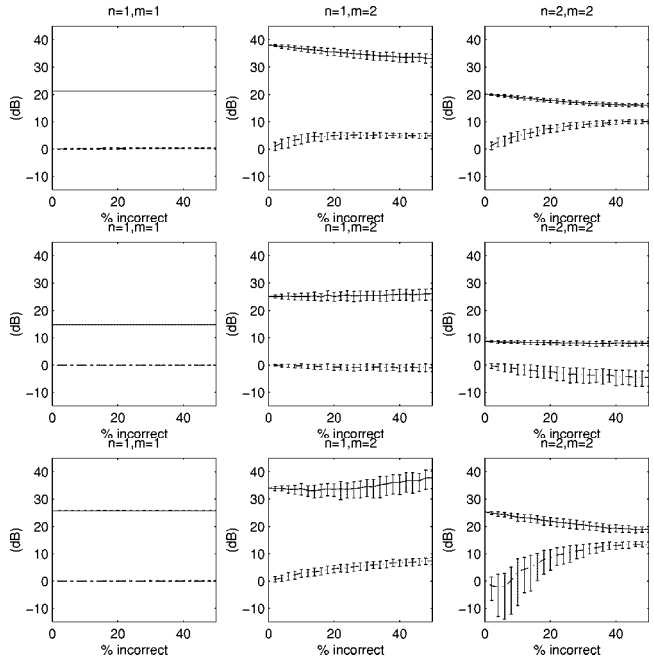

Fig. 10. Moment and cross-moment terms versus percentage of incorrectly permuted frequency bins. The solid lines illustrate the moments of the incorrectly permuted signals. The dashed lines illustrate the cross-moments.

correspond to the estimation error, and can be used as a reference value for the sample averages of these quantities.

In Fig. 10, the solid lines illustrate the moments of the incorrectly permuted signals. The dashed lines illustrate the cross-moments. Both are measured in decibels over the correctly permuted cross-moments as is shown in (25). Each row represents different signal pairs. The top row corresponds to two speech signals, the center row corresponds to two different music signals, and bottom row corresponds to music and speech. All signals were 7 s long and sampled at 16 kHz. They were normalized to unit variance and zero mean. We summed over 2048 taps.

As show in the previous section, the time-domain criterion (4) is invariant with respect to permutations of individual frequency bins for the case $n = m = 1$. This is confirmed numerically and shown in the left column of Fig. 10. Therefore, second-order criteria can not be used to find filters with correct permutations. Fig. 10 also demonstrates numerically that for $n = 1$, $m = 2$, and $n = m = 2$, the time-domain criterion (4) is not robust to permutations. For some signals (top row), the higher-order cross-moments are in fact minimized by the correct permutations. However, for some other signals (middle and bottom row) the higher-order cross-moments are minimal for incorrect permutations. This simulation indicates that third- and fourth-order time-domain criteria are not robust to the permutation problem, as the criteria are limited by estimation accuracy, in particular for nonstationary signals.

ACKNOWLEDGMENT

The authors would like to thank C. Fancourt and C. Spence for numerous discussions during the preparation of this work.

REFERENCES

[1] B. Van Veen and K. Buckley, "Beamforming techniques for spatial filtering," in *Digital Signal Processing Handbook*. Boca Raton, FL: CRC, 1997.

[2] H.-L. N. Thi and C. Jutten, "Blind source separation for convolutive mixtures," *Signal Process.*, vol. 45, no. 2, pp. 209–229, 1995.

[3] D. Yellin and E. Weinstein, "Multichannel signal separation: Methods and analysis," *IEEE Trans. Signal Processing*, vol. 44, pp. 106–118, Jan. 1996.

[4] S. Shamsunder and G. Giannakis, "Multichannel blind signal separation and reconstruction," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 515–528, Nov. 1997.

[5] H. Sahlin and H. Broman, "Separation of real-world signals," *Signal Process.*, vol. 64, pp. 103–104, 1998.

[6] L. Parra and C. Spence, "Convolutive blind source separation of nonstationary sources," *IEEE Trans. Speech Audio Processing*, pp. 320–327, May 2000.

[7] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multichannel signal separation by decorrelation," *IEEE Trans. Speech Audio Processing*, vol. 1, pp. 405–413, Apr. 1993.

[8] K. Duvall, "Signal cancellation in adaptive arrays: Phenomena and a remedy," Ph.D. dissertation, Stanford Univ., Stanford, CA, Sept. 1983.

[9] K. I. Diamantaras, A. P. Petropulu, and B. Chen, "Blind two-input–two-output FIR channel identification based on frequency domain second-order statistics," *IEEE Trans. Signal Processing*, vol. 48, pp. 534–542, Feb. 2000.

[10] J. Anenmueller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. ICA 2000*, 2000, pp. 215–220.

[11] D. Schobben, *Real-Time Adaptive Concepts in Acoustics: Blind Signal Separation and Multichannel Echo Cancellation*. Norwell, MA: Kluwer, 2001.

[12] F. Ehlers and H. G. Schuster, "Blind separation for convolutive mixtures and an application in automatic speech recognition in a noisy environment," *IEEE Trans. Signal Processing*, vol. 45, no. 10, pp. 2608–2612, Oct. 1997.

[13] M. Kawamoto, "A method of blind separation for convolved nonstationary signals," *Neurocomput.*, vol. 22, no. 1–3, pp. 157–171, 1998.

[14] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 425–437, Sept. 1997.

[15] O. L. Frost, "An algorithm for linear constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.

[16] L. Griffiths and C. Jim, "An alternative approach to linearly constrainted adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 27–34, Jan. 1982.

[17] V. Capdevielle, C. Serviere, and J. L. Lacoume, "Blind separation of wide-band sources in the frequency domain," in *Proc. ICASSP*, 1995, pp. 2080–2083.

[18] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomput.*, vol. 22, pp. 21–34, 1998.

[19] M. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *Proc. ICASSP 2000*, vol. II, 2000, pp. 1041–1044.

[20] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[21] Y. Kung, R. E. Hudson, C. W. Reed, D. Chen, and F. Lorenzelli, "Blind beamforming on a randomly distributed sensor array system," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1555–1567, Oct. 1998.

[22] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 497–507, Sept. 2000.

[23] F. Asano, M. Yoichi, H. Asoh, and T. Matsui, "Effect of PCA filter in blind source separation," in *Proc. ICA 2000*, 2000, pp. 57–62.

[24] L. Parra and C. Spence, "On-line convolutive blind source separation of nonstationary sources," *J. VLSI Signal Process.*, vol. 26, pp. 39–46, Aug. 2000.

[25] C. Fancourt and L. Parra, "Coherence function as a criterion for blind source separation," in *IEEE Int. Workshop on Neural Networks for Signal Processing*, 2001, pp. 303–312.

[26] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, pp. 2737–2740, 2001.

[27] J. Principe, "Simultaneous diagonalization in the frequency domain (SDIF) for source separation," in *Proc. ICA'99*, 1999, pp. 245–250.
[28] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of non stationary sources," *IEEE Trans. Signal Processing*, vol. 49, pp. 1837–1848, Sept. 2001.
[29] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. Inst. Elect. Eng.*, vol. 130, no. 1, pp. 11–16, Feb. 1983.
[30] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 10, pp. 1365–1375, 1987.
[31] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.
[32] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.

Lucas C. Parra, photograph and biography not available at the time of publication.

Christopher V. Alvino, photograph and biography not available at the time of publication.