Exploiting the Self-Steering Capability of Blind Source Separation to Localize Two or More Sound Sources in Adverse Environments

Anthony Lombard [1], *Tobias Rosenkranz* [1*], *Herbert Buchner* [2], *Walter Kellermann* [1]

[1] Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg,
Cauerstr. 7, 91058-Erlangen, Germany
E-Mail: {lombard,rosenkr,wk}@LNT.de

[2] Deutsche Telekom Laboratories, Technical University Berlin,
Ernst-Reuter-Platz 7, 10587-Berlin, Germany
E-Mail: hb@buchner-net.com

Abstract

Blind Source Separation (BSS) algorithms have often been interpreted as a set of blind adaptive beamformers. Although this interpretation does not entirely hold under realistic conditions, it gives some useful insights on the self-steering capacity of BSS techniques. Actually, while accurate source location information is usually necessary to steer a beamformer, BSS offers the possibility to recover the original source signals from a reverberant sound mixture without this prior knowledge. Intuitively, this self-steering capacity should therefore imply that the BSS demixing filters contain some useful information on the location of each source. In this paper, we discuss the relations between BSS and beamforming techniques, based on the general form of the ideal separation solution. Two possible ways to extract the location information from the BSS demixing system are then presented. Experiments using a broadband BSS algorithm based on the TRINICON framework demonstrate the efficiency of the proposed methods in a highly reverberant and noisy environment, for up to six simultaneously active speech sources.

1 Introduction

Acoustic source localization aims at estimating the position of one or several sound sources by exploiting the spatial diversity offered by an array of microphones. Accurate localization of one or several sound sources can serve in many applications as a preliminary step to other processes like, e.g., steering a beamformer or pointing a camera in the direction of a sound source.

To address the problem of localizing several simultaneously active sound sources in noisy and reverberant environments, a generic broadband approach, which jointly processes all frequency bins, has been proposed for Blind Source Separation (BSS) in [2, 3]. It performs blind adaptive Multiple-Input-Multiple-Output (MIMO) system identification of the acoustical system to extract the Time Differences Of Arrival (TDOAs) of multiple sources, hence explicitly accounting for the multipath propagation of sound occurring in real environments. This method has also been extended to the multi-dimensional case in [7, 8], where real-time 2D localization of two sources was demonstrated. Other BSS-based approaches for localizing two sources exist [5, 6]. Originally developed to solve a permutation problem specific to narrowband BSS, which considers each frequency bin separately, they exploit the directivity patterns of the demixing BSS filters to recover the Direction of Arrival (DOA) of each sound source, hence working under the far-field assumption. Whether based on a broadband or on a narrowband formulation, all these methods aim at extracting some location information from the BSS demixing filters. They therefore rely on the self-steering capability of BSS, i.e., its capacity to recover each sound source from a convolutive mixture, without any prior knowledge on the source position.

In this paper, we restrict ourselves to the single-dimension case but propose to treat the general case of $P \geq 2$ sources. The rest of the paper is organized as follows. We provide an overview on BSS in Sect. 2. We formulate the general form of the ideal signal separation solution and use it to show the relations existing between BSS and beamforming techniques. In Sect. 3, we review the TDOA estimation approach [3], and we propose a new method exploiting the BSS directivity patterns similar to [6], and extend these methods to the general case of $P \geq 2$ sources and to large microphone apertures. Experimental results are finally provided in Sect. 4, where up to six simultaneously active speech sources are localized in a reverberant and noisy environment.

* Now with Siemens Audiologische Technik, Erlangen

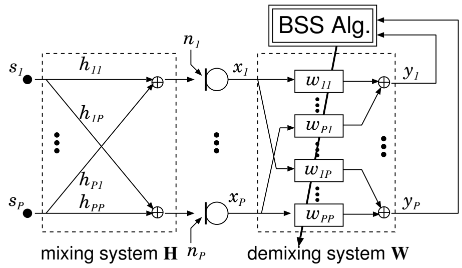


Figure 1: BSS signal model.

2 Separation of convolutive mixtures

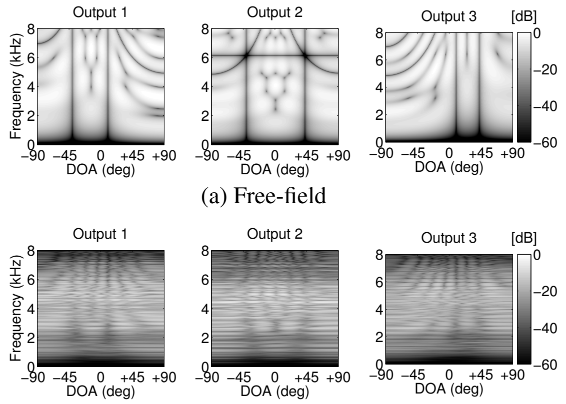
2.1 Problem formulation

Fig. 1 shows the general BSS setup. Because of the reverberation in the acoustical environment, P source signals s_q $(q = 1 \dots P)$ are filtered by a MIMO mixing system $\mathbf{H}$ modeled by M-tap Finite Impulse Response (FIR) filters h_{qp} between the q-th source and the p-th sensor [1]. P signal mixtures x_p are picked up by a microphone array, together with some background or sensor noise n_p. We assume here that the number of active sources is less or equal to the number of microphone signals (the P sources in Fig. 1 might or might not all be simultaneously active). Furthermore, the sources are assumed to be mutually independent, which in general holds for speech and audio signals.

To separate the source signals s_q without access to the acoustical mixing system $\mathbf{H}$, BSS algorithms force the output signals y_q to become statistically independent by suitably adapting the weights of the BSS demixing system $\mathbf{W}$, which captures the L-tap FIR separating filters w_{pq} between the p-th sensor and the q-th output. In this paper, we consider a broadband second-order-statistics realization of the generic TRINICON-based update rule [1]. This is a block-online algorithm based on a broadband BSS coefficient update, which processes signals on a block-by-block basis. Note that this broadband approach does not suffer from the internal permutation ambiguity encountered by narrowband BSS techniques, where an output permutation alignment problem has to be solved in each frequency bin (see, e.g., [5, 6, 9]).

The general form of the ideal separating filter matrix based on the mixing system was shown in [2]. Expressed in the frequency domain to substitute convolutions by scalar multiplications, this relationship reads:

$$\mathbf{W}_{\text{ideal}}(f) \quad = \quad \text{Adj}\{\mathbf{H}(f)\} \cdot \mathbf{\Lambda} \cdot \mathbf{P}, \qquad (1)$$



(a) Free-field

(b) Lecture room with $T_{60} \approx 1$sec

Figure 2: Directivity patterns of the ideal separation solution.

where the Adj$\{\cdot\}$ operator computes the adjoint of a square matrix. The *frequency-independent* matrix $\mathbf{P}$ (a matrix with a one in each row and each column, and zeros elsewhere) and the diagonal matrix $\Lambda = \mathrm{Diag}\{\alpha_1, \ldots, \alpha_P\}$ describe an arbitrary permutation and scaling of the BSS outputs, respectively. An algorithm converging to a solution satisfying (1) achieves *perfect* separation of the competing sources since it forces the overall system

$$\begin{aligned} \mathbf{C}_{\text{ideal}}(f) &= \mathbf{H}(f) \cdot \mathbf{W}_{\text{ideal}}(f) \\ &= \mathbf{H}(f) \cdot \mathrm{Adj}\{\mathbf{H}(f)\} \cdot \Lambda \cdot \mathbf{P} \\ &= \det\{\mathbf{H}(f)\} \cdot \Lambda \cdot \mathbf{P} \end{aligned} \tag{2}$$

to become diagonal, up to a permutation among the BSS outputs. det$\{\cdot\}$ refers to the determinant (i.e., a scalar) of a square matrix. The uncritical (since frequency-independent) permutation and scaling ambiguities will be ignored in the rest of this paper, i.e., $\Lambda = \mathbf{P} = \mathbf{I}$, the identity matrix.

2.2 Relations between BSS and beamforming

Beamformers are processors exploiting the spatial diversity offered by a microphone array to enhance signals coming from a given direction of interest while rejecting interfering signals coming from other directions. Hence, BSS has often been heuristically interpreted as a set of blind adaptive beamformers [4]. This interpretation can be partly justified by observing the directivity pattern, i.e., the response to monochromatic plane waves coming from all possible directions θ (measured with respect to the normal of the array axis) of each BSS output:

$$B_q(\mathbf{W}, \theta, f) = \left| \sum_{p=1}^{P} w_{pq}(f) e^{-j2\pi f d_p \sin(\theta)/c} \right|^2, \tag{3}$$

where c is the sound velocity and d_p is the distance from the p-th sensor to the first sensor of the assumed linear array.

Under the free-field assumption and neglecting attenuation in the propagation medium, the (single-path) filters of the mixing system $\mathbf{H}$ can be written as:

$$h_{qp}^{\text{FF}}(f) = e^{-j2\pi f \tau_{qp}(\theta_q)}, \tag{4}$$

where $\tau_{qp}(\theta_q) = d_p \sin(\theta_q)/c$ is the TDOA of the q-th source (with DOA θ_q) at the p-th sensor. This allows to compute the BSS directivity patterns of the ideal separating filters (1) in the free-field case. They are shown in Fig. 2a, for $P = 3$, $[d_1\ d_2\ d_3] = [0\text{m}\ 0.042\text{m}\ 0.21\text{m}]$, $[\theta_1\ \theta_2\ \theta_3] = [-40^\circ\ 10^\circ\ 40^\circ]$ and sampling frequency f_s= 16kHz. Values smaller than -60dB have been clipped. Moreover, Fig. 2b shows the BSS directivity patterns of the ideal separation solution computed for a reverberant lecture room with identical geometric parameter values. The mixing system was here generated from measured Room Impulse Responses (RIRs) for sources located two meters away from the center of the sensor array.

We see from Fig. 2a that the ideal separation solution under the free-field assumption consists of a set of P null-beamformers, each placing $P-1$ perfect spatial nulls (i.e., with infinite attenuation) in the direction of $P-1$ competing sources. The ideal separation solution can be seen, in this case, as a generalization of the 2-channel delay-and-subtract beamformer introduced in [5]. However, in realistic conditions, considering BSS as a set of blind beamformers is a bit misleading and somehow reducing since, contrary to a beamformer, the ideal BSS solution (1) still guarantees *perfect* interference rejection. This can be seen in Fig. 2b, where the spatial nulls are hardly visible, although *perfect* source separation is provided. In reverberant environments, multiple paths exist between each source and each sensor. While methods based on a single-path propagation model (like classical beamformers) see multiple sources for several propagation paths of the same source, a BSS algorithm sees only the actual P coherent point sources [4]. Extraction of one source in each BSS output is then performed by compensating simultaneously for every acoustical propagation path (so not only the direct path) coming from the remaining $P-1$ interfering sources. This corresponds to the joint diagonalization (2) of $\mathbf{C}(f)$ in all frequencies and for an arbitrary $\mathbf{H}(f)$.

3 Localization based on BSS

While beamformers need some prior knowledge on the source positions to be correctly steered, a BSS algorithm can recover the original source signals without this prior knowledge. Intuitively, this self-steering capacity should therefore imply that the BSS demixing filters contain some useful information on the location of each source. In the following, we present two methods to extract this information.

3.1 TDOA extraction using blind system identification

One way to retrieve the localization information from BSS demixing filters has been presented in [3]. It relies on the ability of the broadband BSS algorithm [1] to converge to a solution of the form described by (1). Expanding (1) for the case of $P = 2$ sources and $\Lambda = \mathbf{P} = \mathbf{I}$:

$$\mathbf{W}_{\text{ideal}}(f) = \begin{bmatrix} h_{22}(f) & -h_{12}(f) \\ -h_{21}(f) & h_{11}(f) \end{bmatrix}, \tag{5}$$

we see actually that the ideal separation solution allows to directly identify the filters of the acoustical mixing system. For localization purposes, the TDOA of each source can be extracted from (5), simply by identifying the position of the direct path component (i.e., the dominant one) in each filter estimate [3].

In the general case $P > 2$, expanding (1) does not directly provide a one-to-one relation between each acoustical filter and the filters of the ideal separation solution. An additional step is therefore required after BSS in order to obtain the estimate of the mixing system from the demixing system [2].

3.2 DOA extraction using averaged BSS directivity patterns

As a simplified approximate solution for $P > 2$, we follow here another approach inspired from existing works on narrowband BSS. It was shown in Sect. 2.2 that BSS acts as a set of null-beamformers under the free-field assumption. Although this interpretation does not entirely hold anymore under reverberant conditions, it has often been used to perform localization in each frequency bin and to address the internal permutation problem specific to narrowband BSS approaches, based on the BSS directivity patterns [5, 6].

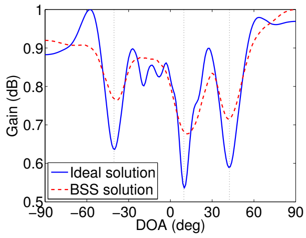


Figure 3: Averaged BSS directivity patterns in a lecture room.

Working under the far-field assumption, the directivity patterns of the adaptive demixing systems delivered by the broadband BSS algorithm [1] can be exploited in a similar way. However, instead of considering each frequency bin and each output separately, we average here the BSS directivity patterns over the frequencies, and over the $P-1$ "best" BSS outputs, i.e., discarding for each frequency point and each (discrete) angle, the output with maximum array response. The averaged directivity pattern $\bar{B}(\mathbf{W},\theta)$ is therefore computed as follows:

$$q^*(\theta,f) = \arg\max_q B_q(\mathbf{W},\theta,f), \quad (6)$$

$$\bar{B}(\mathbf{W},\theta) = \frac{1}{C}\int_f \sum_{\substack{q=1 \\ q\neq q^*(\theta,f)}}^{P} B_q(\mathbf{W},\theta,f)\, df, \quad (7)$$

with C being an arbitrary constant. In practice, the integral is replaced by a summation over a finite number of frequency points. This averaging procedure allows to largely attenuate the impact of spatial aliasing occuring at high frequencies since only the "true" spatial nulls (as opposed to the unwanted grating lobes) add up coherently when summing over the BSS outputs and frequencies. This allows to gather very useful localization information also from higher frequency regions, even with large microphone spacings, contrary to the approach presented in [5, 6] where these regions are simply discarded. Source localization can then be achieved by identifying the angles corresponding to the P deepest local minima in $\bar{B}(\mathbf{W},\theta)$.

Note that the approach originally presented in [6] is limited to the case of two sources only [9], whereas the proposed method is easily applicable to three or more sources. This is illustrated by Fig. 3 for the localization of $P=3$ sources in the lecture room already introduced in Sect. 2.2. The figure shows the averaged directivity patterns of the ideal solution (solid line), and of the BSS demixing filters after convergence of the broadband BSS algorithm (dashed line). The former was computed by applying the averaging procedure (7) to the directivity patterns of Fig. 2b. The geometrical setup considered for Fig. 2 and Fig. 3 were identical. Vertical dotted lines indicate the true source DOAs. We see that although the spatial nulls are hardly noticeable in Fig. 2b, $P=3$ local minima pointing towards each source become clearly visible when applying the above averaging procedure (see the solid line). The local minima are also easily detectable in the averaged directivity pattern identified by the adaptive demixing filters (the dashed line).

This shows that the broadband BSS algorithm converges indeed to the desired solution and can serve for localization purposes, without prior knowledge on the mixing process. Reverberation might disturb the DOA estimation in some cases but in general, as long as the direct propagation path is sufficiently strong compared to the reflection paths and the far-field assumption is sufficient for the intended application, the DOA estimation based on directivity patterns is applicable, as shown by the experimental results presented in the next section.

4 Experimental evaluations

4.1 Experimental setup

The localization performance of the presented BSS-based schemes have been assessed under reverberant and noisy conditions. To this aim, a linearly and equally-spaced array of six omnidirectional microphones has been placed in the lecture room already mentioned in Sect. 2 and Sect. 3, with reverberation time $T_{60} \approx 1$sec. RIRs have been measured at the sampling frequency $f_s = 16$kHz, for source positions located on a circle at a distance of two meters from the center of the array, as depicted in Fig. 4. Microphone signals were then generated by convolving speech signals with the measured RIRs and possibly by adding some spatio-temporally uncorrelated white noise at 5dB and 15dB SNR.

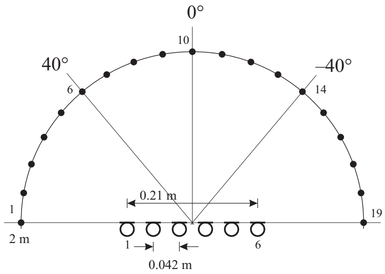


Figure 4: Experimental setup.

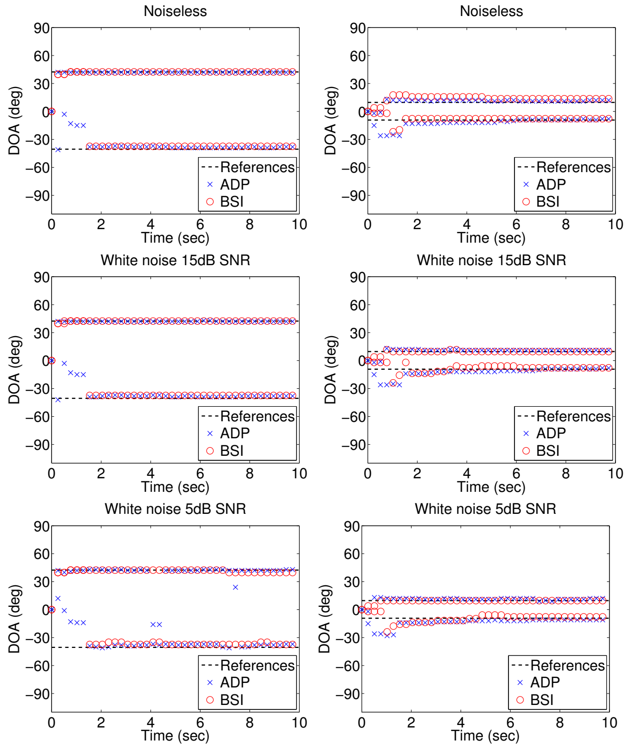


Figure 5: Localization of two sources.

In the following, the methods based on Blind System Identification (Sect. 3.1) and Averaged Directivity Patterns (Sect. 3.2) are labelled BSI and ADP, respectively. A BSS update followed by a DOA estimation was realized every 4096 samples at the sampling frequency $f_s = 16$kHz. The directivity patterns were computed with an angular resolution of 1° and a frequency resolution of 125Hz. To reduce the computational complexity of the BSS part, short BSS filters of length $L = 128$ samples have been used. This forces the BSS algorithm to concentrate on the early parts of the desired solution, which contain most of the information on the direct acoustical propagation paths.

4.2 DOA estimation results

Figure 5 shows the localization results obtained with both methods, using microphones 1 and 6 (i.e., for a microphone spacing $d = 0.21$m, see Fig. 4), for $P = 2$ sources placed far apart (left column) or very close to each other (right column). Note that contrary to the BSI method, the ADP method relies on the far-field

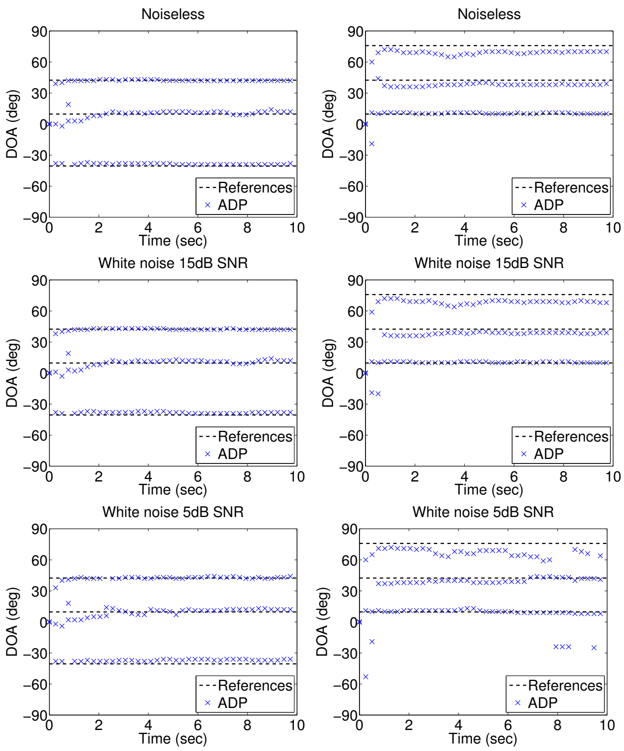


Figure 6: Localization of three sources.

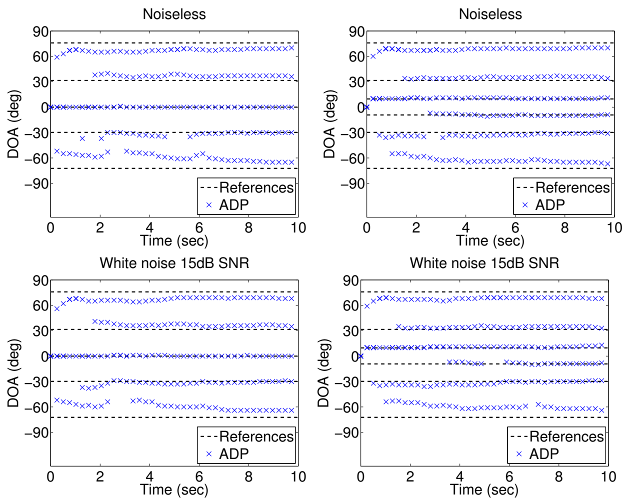


Figure 7: Localization of five sources (left column) and six sources (right column).

assumption to compute the BSS directivity patterns (Sect. 3.2). We therefore present the localization results in terms of DOA estimation performance, thereby mapping each TDOA estimate τ delivered by the BSI method into a DOA $\theta = \arcsin(c\tau/f_s d)$. Moreover, Fig. 6 and Fig. 7 show the results obtained from the ADP method for $P = 3$, 5 and 6 sources when using sensors [1 2 6], [1 2 3 4 6] and [1 2 3 4 5 6], respectively.

In this highly reverberant environment, the BSI method for two sources, and the ADP method for up to six sources, performed well, even for closely spaced sources and at relatively high background noise levels, although a few outliers appeared in the difficult 5dB-SNR case. From Fig. 7, we also see that the ADP method sometimes detected less sources than the expected P, but this happened only for short periods of time and can be attributed to speech pauses. The localization was generally very accurate in all considered conditions, except at large off-broadside angles of incidence where the spatial resolution of the linear array is the lowest.

5 Conclusions

Building upon the general form of the ideal separation solution and the relations existing between BSS and beamforming, we studied two methods to retrieve the localization information contained in the demixing system of a TRINICON-based broadband BSS algorithm. The first, already existing, approach [3] relies on the ability of the BSS algorithm to blindly identify the MIMO acoustical system. The second approach relies on the far-field assumption and uses the directivity patterns of the BSS demixing filters, similar to [6]. But we apply here an averaging procedure which allows us to treat the general case of two or more sources and to gather very useful localization information from every frequency region, including those corrupted by spatial aliasing for large microphone spacings.

Both techniques showed their effectiveness in a highly reverberant and noisy environment for two sources. The proposed method based on averaged directivity patterns also showed its applicability in the presence of up to six simultaneously active sound sources. Future works will focuss on extending the method to nonlinear microphone array geometries for multidimensional localization.

Literature

[1] R. Aichner, H. Buchner, F. Yan, and W. Kellermann. A real-time blind source separation scheme and its application to reverberant and noisy environments. *Signal Processing*, 86(6):1260–1277, June 2006.

[2] H. Buchner, R. Aichner, and W. Kellermann. TRINICON-based blind system identification with application to multiple-source localization and separation. In S. Makino T.-W. Lee and S. Sawada, editors, *Blind Speech Separation*. Springer-Verlag, Berlin, 2007.

[3] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellermann. Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering. In *IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Philadelphia, PA, USA, Mar. 2005.

[4] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE- Proc.-F*, 140(6):362–370, Dec. 1993.

[5] M. Ikram and D. Morgan. A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation. In *IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, volume 1, pages 881–884, 2002.

[6] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura. Evaluation of blind signal separation method using directivity pattern under reverberant conditions. In *IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000.

[7] A. Lombard, H. Buchner, and W. Kellermann. Multidimensional localization of multiple sound sources using blind adaptive MIMO system identification. In *Proc. IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Heidelberg, Germany, Sep. 2006.

[8] A. Lombard, H. Buchner, and W. Kellermann. A real-time demonstrator for the 2D localization of two sound sources using blind adaptive MIMO system identification. In *Proc. Joint Workshop on Hands-Free Communication and Microphone Arrays (HSCMA)*, Trento, Italy, May 2008.

[9] H. Sawada, R. Mukai, S. Araki, and S. Makino. Direction of arrival estimation for multiple source signals using independent component analysis. In *Proc. ISSPA*, Paris, France, 2003.