SEPARATING CONVOLUTIVE MIXTURES WITH TRINICON

Walter Kellermann, Herbert Buchner, and Robert Aichner

Multimedia Communications and Signal Processing
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
{wk, buchner, aichner}@LNT.de

ABSTRACT

Blind source separation (BSS) algorithms are often categorized as either narrowband or broadband algorithms depending on whether their respective cost functions aim at individual DFT bins or the entire broadband signal. In this contribution, we present comparable general natural gradient-based formulations of both concepts based on the TRINICON framework. As a distinctive feature, narrowband algorithms imply an internal permutation and scaling problem. We show that the common DOA estimation-based methods for aligning the permutations effectively rely on geometric a-priori knowledge, and we explain why they need to be complemented by additional repair mechanisms for robust BSS. The latter can already be viewed as approximations of the generic TRINICON broadband algorithm. As a conclusion, we propose to always use a generic broadband algorithm as a starting point for the design of new BSS algorithms.

1. INTRODUCTION

Over the last decade, blind separation of convolutive signal mixtures has become a major research area in signal processing, notably for separating speech and audio signals in hands-free communication environments. The main idea of BSS is to retrieve the separated source signals from convolutive mixtures as recorded by several sensors (see Fig.1). The demixing system should be identified by exclusively exploiting the statistical independence of the sources, which leads to the notion of independent component analysis (ICA) [1].

For our acoustic communication context, we use the same number P of audio source signals s_p, sensor signals x_p, and outputs y_p as shown in Fig.1. Assuming time-invariance for the mixing system $\mathbf{H}$ and the demixing system $\mathbf{W}$, the microphone signals $x_p(n)$ and the outputs $y_q(n)$ can be written as

$$x_p(n) = \sum_{q=1}^{P} \sum_{\kappa=0}^{M-1} h_{qp,\kappa} s_q(n-\kappa), \qquad (1)$$

$$y_q(n) = \sum_{p=1}^{P} \sum_{\kappa=0}^{L-1} w_{pq,\kappa} x_p(n-\kappa), \qquad (2)$$

respectively, where the length M of the impulse response models h_{qp} is on the order of at least several hundred samples for realistic acoustic environments, even at relatively low sampling rates of $f_\mathrm{s} =$ 8kHz.

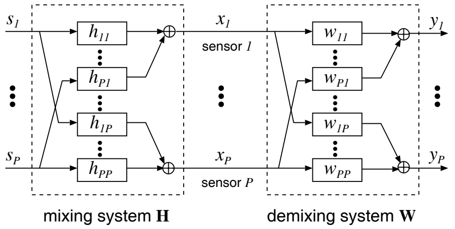

Fig. 1. Generic setup for BSS

In an attempt to compare and generalize the multitude of seemingly different BSS algorithms, TRINICON ('TRIple-N ICA for CONvolutive mixtures') [11, 12] has been formulated as a general framework for blind MIMO signal processing [13], and lead already to several new BSS algorithms [11, 16, 17], but also to new algorithms for speech dereverberation [12] and source localization [15]. In this contribution, we use the TRINICON framework to study some aspects of the so-called *internal permutation problem* as it arises in *narrowband BSS algorithms*. For that, we present in Section 2 a general formulation of natural gradient-based narrowband BSS algorithms as a specialization of TRINICON and confront it in Section 3 with the generic natural gradient-based BSS formulation of TRINICON, which implies a broadband signal model. In Section 4, we compare the relevance of the sensor arrangement to narrowband and broadband algorithms, and, in Section 5, point at new hybrid algorithms as a promising research direction.

2. NARROWBAND BSS

The general idea of narrowband BSS for convolutive mixtures is to transform the convolutive mixtures in the time domain $x_p(n)$ into instantaneous mixtures in the DFT domain, so that for the demixing system $\mathbf{W}$ only P^2 scalars have to be identified for each frequency bin, instead of P^2 impulse responses for the broadband signals. The according instantaneous BSS problem (e.g. [1]) can then be solved individually in each frequency bin, which intuitively should be conceptually simpler and computationally less complex compared to the broadband approach.

For a formal description, we apply the DFT to a windowed version of a length-R segment of the time-domain signal $x_p(n)$, so that we obtain for frequency bin ν ($\nu = 0, \ldots, R-1$) of the signal segment m:

$$\underline{x}_p^{(\nu)}(m) = \sum_{r=0}^{R-1} x_p\left(r + m\frac{R}{\alpha}\right) v(r) e^{-j2\pi\nu r/R}, \qquad (3)$$

where $v(r)$ denotes the window function and R/α is a shifting interval of the window, with α as overlap factor of successive segments. Based on this, we formulate the demixing system (2) in the DFT domain as follows:

$$\underline{\mathbf{y}}^{(\nu)}(m) = \underline{\mathbf{x}}^{(\nu)}(m)\underline{\mathbf{W}}^{(\nu)}(m), \qquad (4)$$

1-4244-0469-X/06/$20.00 ©2006 IEEE

where

$$\underline{\mathbf{y}}^{(\nu)}(m) = \left[\underline{y}_1^{(\nu)}(m),\dots,\underline{y}_P^{(\nu)}(m)\right], \quad (5)$$

$$\underline{\mathbf{x}}^{(\nu)}(m) = \left[\underline{x}_1^{(\nu)}(m),\dots,\underline{x}_P^{(\nu)}(m)\right], \quad (6)$$

and where $\underline{\mathbf{W}}^{(\nu)}$ is a $P \times P$ matrix whose elements are the DFT domain counterparts $\underline{w}_{pq}^{(\nu)}$ of the coefficients $w_{pq,\kappa}$ in (2). As the main trait of narrowband ICA algorithms, the update of the coefficient matrix is performed separately for each bin according to

$$\underline{\mathbf{W}}^{(\nu)}(m) = \underline{\mathbf{W}}^{(\nu)}(m-1) - \mu\Delta\underline{\mathbf{W}}^{(\nu)}(m), \quad (7)$$

where the update term can be written in the general form

$$\Delta\underline{\mathbf{W}}^{(\nu)}(m) = \underline{\mathbf{W}}^{(\nu)}(m-1)\cdot\left(\sum_{j=0}^{m}\gamma(j,m)\left(\underline{\mathbf{y}}^{(\nu)}\right)^{H}(j)\underline{\mathbf{\Phi}}(\underline{\mathbf{y}}^{(\nu)}(j)) - \mathbf{I}\right) \quad (8)$$

with γ as a weighting function and the score function $\underline{\mathbf{\Phi}}$, which reads in its optimum general form [6]:

$$\underline{\mathbf{\Phi}}(\underline{\mathbf{y}}^{(\nu)}(m)) = -\left[\frac{\frac{\partial\hat{p}(\underline{y}_1^{(\nu)}(m))}{\partial\underline{y}_1^{(\nu)}(m)}}{\hat{p}(\underline{y}_1^{(\nu)}(m))},\dots,\frac{\frac{\partial\hat{p}(\underline{y}_P^{(\nu)}(m))}{\partial\underline{y}_P^{(\nu)}(m)}}{\hat{p}(\underline{y}_P^{(\nu)}(m))}\right]. \quad (9)$$

Here, $\hat{p}(\underline{y}_p^{(\nu)}(m))$ is the estimated or assumed probability density function of the ν-th bin of output channel p. Known narrowband algorithms approximate this score function usually by nonlinear functions such as $\tanh$ [6] leading to 'higher order statistics' (HOS) algorithms.

Independent processing of different frequency bins implies the problem of inconsistent scaling and the so-called internal permutation problem: Inconsistent scaling results from the fact that BSS algorithms produce outputs which are unique only up to a scaling factor. With independent BSS in each frequency bin, the scaling will in general not be the same for different bins of the same source. As a remedy, the minimum distortion principle (MDP) [18] is commonly applied. The term 'internal permutation' describes the effect that frequency components in different bins belonging to the same source s_q do not necessarily appear at the same output channel. To align the frequency bins correctly, three classes of repair mechanisms have been developed:

- The separated components are aligned according to the phases of the DFT bins, which corresponds to a classification according to the estimated direction of arrival (DOA) of the sources (e.g., [2, 3]).
- A second class of repair mechanisms exploits the correlation of the temporal evolution of spectral magnitudes for a given source. For a 'local' version the correlation between the temporal envelopes of neighboring frequency bins is used to align the components (e.g., in [3, 4]), whereas for the computationally expensive, optimum 'global' version [5] the correlations between all frequency bins are accounted for.
- For a third, computationally efficient but suboptimum method, a spectral smoothness constraint is imposed on the demixing filters by windowing (shortening) the corresponding impulse responses in the time domain. [6, 7].

Studies based on the TRINICON framework in [14] reveal that the latter two mechanisms result actually as special cases of the generic broadband algorithm if some constraints in its DFT-domain formulation are removed, so that the resulting algorithms represent hybrids of strictly narrowband and strictly broadband algorithms.

3. BROADBAND BSS

As opposed to the narrowband algorithms, broadband BSS algorithms are derived from a time-domain representation and, thus, inherently avoid the internal permutation problem. Early publications (e.g., [8, 9]) aimed at multichannel blind deconvolution and lead to distortion of the desired signals ('whitening effect'). Within the TRINICON framework, new broadband BSS algorithms were developed that avoid the whitening effect and, with proper design, are able to blindly identify the theoretically optimum signal separation filters [10]. Note that the latter property can also be exploited for simultaneous localization of multiple sources in strongly reverberant environments [15].

For comparison to the narrowband algorithms above, we summarize the corresponding broadband representation in the time domain (see e.g.[11, 13]). The convolution in the demixing system (2) is captured by the following matrix form:

$$\mathbf{y}(m,j) = \mathbf{x}(m,j)\mathbf{W}(m), \quad (10)$$

where m denotes the block index, and $j = 0,\cdots,N-1$ is a time-shift index within a data segment of length $N+D-1$, and where

$$\begin{aligned}
\mathbf{x}(m,j) &= [\mathbf{x}_1(m,j),\dots,\mathbf{x}_P(m,j)], & (11)\\
\mathbf{y}(m,j) &= [\mathbf{y}_1(m,j),\dots,\mathbf{y}_P(m,j)], & (12)\\
\mathbf{W}(m) &= \begin{bmatrix}\mathbf{W}_{11}(m) & \cdots & \mathbf{W}_{1P}(m)\\ \vdots & \ddots & \vdots\\ \mathbf{W}_{P1}(m) & \cdots & \mathbf{W}_{PP}(m)\end{bmatrix}, & (13)\\
\mathbf{x}_p(m,j) &= [x_p(mL+j),\dots,x_p(mL-2L+1+j)] & (14)\\
\mathbf{y}_q(m,j) &= [y_q(mL+j),\dots,y_q(mL-D+1+j)] & (15)\\
&= \sum_{p=1}^{P}\mathbf{x}_p(m,j)\mathbf{W}_{pq}(m). & (16)
\end{aligned}$$

The $2L \times D$ matrix $\mathbf{W}_{pq}(m)$ denotes a Sylvester matrix that contains the L coefficients $w_{pq,l}$ $(l = 0,\dots,L-1)$ of the demixing FIR filter, while D $(1 \leq D \leq L)$ indicates the number of time lags used for exploiting the nonwhiteness of the sources:

$$\mathbf{W}_{pq}(m) = \begin{bmatrix}
w_{pq,0} & 0 & \cdots & 0\\
w_{pq,1} & w_{pq,0} & \ddots & \vdots\\
\vdots & w_{pq,1} & \ddots & 0\\
w_{pq,L-1} & \vdots & \ddots & w_{pq,0}\\
0 & w_{pq,L-1} & \ddots & w_{pq,1}\\
\vdots & & \ddots & \vdots\\
0 & \cdots & 0 & w_{pq,L-1}\\
0 & \cdots & 0 & 0
\end{bmatrix}. \quad (17)$$

The generic broadband BSS algorithm, which simultaneously exploits nonwhiteness, nonstationarity, and nongaussianity, is derived by computing the natural gradient of a general cost function minimizing mutual information of the outputs [13]. The resulting natural gradient-based TRINICON update rule reads:

$$\mathbf{W}(m) = \mathbf{W}(m-1) - \mu \Delta \mathbf{W}(m), \tag{18}$$

$$\Delta \mathbf{W}(m) = \sum_{i=0}^{m} \beta(i,m) \sum_{j=0}^{N-1} \mathbf{W}(i-1) \cdot \left(\mathbf{y}^H(i,j) \mathbf{\Phi}(\mathbf{y}(i,j)) - \mathbf{I} \right) \tag{19}$$

with $\beta(i,m)$ as a general weighting function allowing for both offline and online implementations, and with the score function

$$\mathbf{\Phi}(\mathbf{y}(i,j)) = - \left[\frac{\frac{\partial \hat{p}_D(\mathbf{y}_1(i,j))}{\partial \mathbf{y}_1(i,j)}}{\hat{p}_D(\mathbf{y}_1(i,j))}, \ldots, \frac{\frac{\partial \hat{p}_D(\mathbf{y}_P(i,j))}{\partial \mathbf{y}_P(i,j)}}{\hat{p}_D(\mathbf{y}_P(i,j))} \right], \tag{20}$$

which is based on the estimated or assumed *multivariate* probability density functions (pdfs) $\hat{p}_D(\cdot)$ and $\hat{p}_{PD}(\cdot)$ of dimensions D and PD, respectively. It should be emphasized here that, for improved computational efficiency, this generic broadband algorithm and various approximations will in practice partly be implemented in an equivalent DFT-domain representation [14, 16].

4. DOA ESTIMATION AND INTERNAL PERMUTATION

As a special aspect when comparing narrowband and broadband algorithms, we investigate the usability of DOA information for solving the internal permutation problem and, thereby, discuss a crucial feature of narrowband algorithms. For $P = 2$, we compare the outputs of offline versions [14] of a narrowband algorithm and a broadband algorithm, respectively, after convergence (200 iterations) and for identical experimental conditions: The mixtures were recorded with sampling frequency $f_s = 8$kHz in a reverberant roomwith a reverberation time of $T_{60} = 250$ms and the sources were emulated by loudspeakers placed at a distance of 2m and at $\pm 45°$ relative to the microphone array axis.

As a representative narrowband algorithm we use an HOS-based algorithm with DFT length $R = 2048$, $\alpha = 8$, after [6] (using $\tanh$ as score function) and address inconsistent scaling by the minimum distortion principle according to [18]. The permutations are combatted by an entirely DOA-based approach after [2].

As a representative broadband algorithm we use a computationally efficient SOS-based algorithm from [16] with $N = 2048$, $L = 1024$, $D = L$, approximating the required inverse autocorrelation matrix by a diagonal matrix ('NLMS-like' normalization).

In Fig. 2, we show the minima in the directivity patterns for each frequency bin and both outputs after convergence of the narrowband algorithm for two different microphone spacings, $d = 4$, and $d = 20$cm. As the BSS algorithms should place a spatial minimum in the direction of the respective other source, all minima should appear at $\pm 45°$ for an ideal free-field propagation, and the internal permutation repair mechanism then only had to align the same minimum locations to the same output.

Fig. 2 confirms that realistic rooms do not conform with the free-field propagation model and minima in the directivity pattern will be shifted away from the correct DOA due to multipath propagation. Note that this effect becomes more pronounced with increasing distance between sources and sensors, and of course with increasing noise level that is neglected here. Furthermore, it becomes obvious from Fig. 2 that the severity of the permutation problem with narrowband algorithms strongly depends on the array geometry: As long as the microphone spacing is less than half a wavelength, spatial aliasing is precluded for sources impinging from any direction, which means that only a single minimum will occur in the angular range from $-90°$ to $90°$. For sources impinging from $\pm 45°$ spatial aliasing is precluded for frequencies $f = c/\lambda < c/(\sqrt{2}d)$ (with $c \approx 340$m/sec as the sound velocity, and λ as wavelength) which corresponds to approximately 6kHz for $d = 4$cm and 1.2kHz for $d = 20$cm. Above this frequency spatial aliasing occurs and a purely DOA-based alignment method will necessarily be prone to errors as can be seen from the lower plot in Fig. 2. As a consequence, DOA-based alignment methods ideally have to concentrate on the frequency ranges without spatial aliasing to determine the DOA of the source and will determine a reference for aligning the bins at higher frequencies from that. This can be interpreted as a local decision in a reliable frequency range which is then used as a global decision for all frequency bins. Due to the reduced number of observations for this decision, it must be expected to be less reliable relative to a global decision based on all frequency bins or the broadband signal. Moreover, in realistic scenarios, environmental noise power is often concentrated in the low frequency range, so that it becomes difficult to define reliable frequency ranges for the DOA estimation-based repair mechanisms. Obviously, the alignment will also be more difficult if the angular distance between the

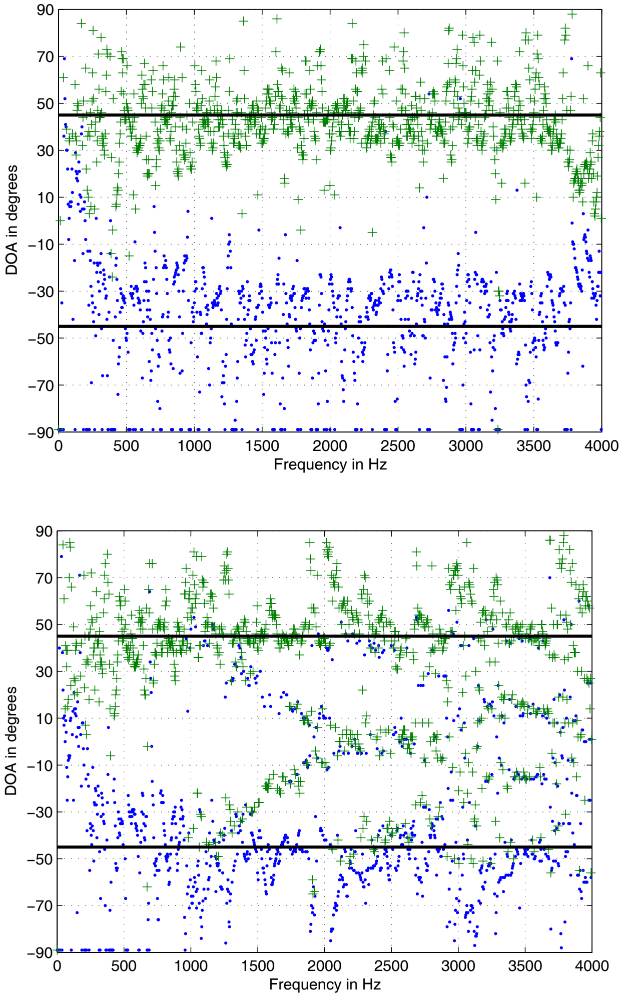

Fig. 2. Minima in the directivity pattern for both outputs of the narrowband approach after DOA-based permutation alignment (microphone spacing $d = 4$cm (upper plot), $d = 20$cm (lower plot))

Spacing d in cm	NB ch1	NB ch2	BB ch1	BB ch2
4	13.7	10.0	15.5	16.7
20	7.5	8.9	12.7	14.2

Table 1. SIR improvement in dB for $P = 2$ channels (NB: narrowband algorithm, BB: broadband algorithm)

sources decreases. Thus, for its successful employment, the DOA-based alignment requires that the microphone spacing is explicitly known and that the free-field propagation model is sufficiently reliable. Clearly, such algorithms are not completely blind in the sense that no geometric knowledge on positions of sensors and sources is required.

As opposed to such semi-blind narrowband algorithms, the broadband algorithms inherently avoid internal permutation and scaling inconsistency and require no geometric knowledge nor do they impose any restrictions on the array geometry. In Table 1 we show that this also implies improved performance by comparing the SIR improvement of both sources as achieved by the demixing systems for two different microphone spacings, for the narrowband and the broadband algorithm, respectively.

Clearly, these results support the strategy to supplement the DOA-based alignment by additional repair methods (as e.g. proposed in [3]) in order to increase robustness against internal permutations with narrowband algorithms.

5. HYBRIDS OF NARROWBAND AND BROADBAND ALGORITHMS

Given the challenges for DOA estimation-based alignment methods for combatting internal permutations and the need for additional repair mechanisms, which can already be seen as simplifications of the generic broadband algorithms (see Section 2), we advocate to use the TRINICON framework for designing hybrids of narrowband and broadband algorithms that combine the advantages of narrowband (computational efficiency resulting from operating in independent frequency bins) and broadband algorithms (preclusion of internal permutation and scaling inconsistency by preserving the broadband nature of the signal) by selectively approximating the generic wideband algorithms. Thereby, they should remain truly blind and be suited for a wide variety of microphone configurations, as they are desirable, e.g., when BSS is used for localization [15] where large spacings are needed for sufficient spatial resolution. As a recent successful example, the broadband/narrowband hybrid presented in [17] exhibits fast convergence, allows for large sensor spacings, and is already implemented as a real-time version on a laptop for $f_\mathrm{S} = 16$kHz.

6. CONCLUSIONS

Confronting narrowband and broadband BSS algorithms in general but comparable formulations based on the TRINICON framework supports the investigation of general structural properties and reveals fundamental limitations. In this contribution, we focused on the use of DOA information for permutation alignment in narrowband algorithms. As a result, the performance of DOA-based permutation alignment methods was found to be strongly dependent on the geometry of the experimental setup. Successful DOA estimation-based algorithms require small microphone spacings and rely on the exploitation of such geometric information, so that they are not fully blind any more. Moreover, for sufficient robustness they need to use additional repair mechanisms for the internal permutation problem, which can be seen as approximations of a generic broadband algorithm. As a consequence, we propose to use the TRINICON framework for truly blind signal separation to design hybrid algorithms which combine the advantages of narrowband and broadband concepts.

7. REFERENCES

[1] A. Hyvaerinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[2] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. ICASSP*, Istanbul, Turkey, Jun. 2000.

[3] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech Audio Proc.*, vol. 12, no. 8, Sep. 2004.

[4] S. Ikeda and N. Murata, "An approach of blind source separation of speech signals," in *Proc. Int. Conf. on Artificial Neural Networks (ICANN)*, Skövde, Sweden, Sep. 1998, pp. 761–767.

[5] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. ICA*, Helsinki, Finland, Jun. 2000, pp. 215–220.

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

[7] L. Parra and C. Spence, "Convolutive blind source separation of nonstationary sources," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[8] S.-I. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, Apr. 1997, pp. 101–104.

[9] X. Sun and S. Douglas, "A natural gradient convolutive blind source separation algorithm for speech mixtures," in *Proc. ICA*, San Diego, CA, USA, Dec. 2001.

[10] H. Buchner, R. Aichner, and W. Kellermann, "Relation between blind system identification and convolutive blind source separation," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Piscataway, NJ, USA, Mar. 2005.

[11] H. Buchner, R. Aichner, and W. Kellermann, "Blind Source Separation for Convolutive Mixtures Exploiting Nongaussianity, Nonwhiteness, and Nonstationarity," *Proc. Int. Workshop on Acoustic Echo and Noise Control*, 2003.

[12] H. Buchner, R. Aichner, and W. Kellermann, "TRINICON: A versatile framework for multichannel blind signal processing," in *Proc. ICASSP*, Montreal, Canada, May 2004.

[13] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures: A unified treatment," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, J. Benesty and Y. Huang, Eds., pp. 255–293. Kluwer, 2004.

[14] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second order statistics," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 1, pp. 120–134, Jan. 2005.

[15] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellermann, "Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering," in *Proc. ICASSP*, Philadelphia, PA, Mar. 2005.

[16] R. Aichner, H. Buchner, F. Yan, and W. Kellermann, "Real-time convolutive blind source separation based on a broadband approch," *Proc. ICA*, Granada, Spain, Sept. 2004.

[17] R. Aichner, H. Buchner, and W. Kellermann, "A novel normalization and regularization scheme for broadband convolutive blind source separation," *Proc. ICA*, Charleston, NC, USA, Mar. 2006.

[18] K. Matsuoka and S. Nakashima, "Minimal distortion principle for Blind Source Separation," *Proc. ICA*, San Diego, CA, USA, Dec. 2001.