# 'Shadow BSS' for Blind Source Separation in Rapidly Time-Varying Acoustic Scenes

S. Wehr[1], A. Lombard[1], H. Buchner[2], and W. Kellermann[1]

[1] University Erlangen-Nuremberg
Multimedia Communications and Signal Processing
Cauerstraße 7, 91058 Erlangen, Germany
{Wehr,Lombard,WK}@LNT.de
[2] Deutsche Telekom Laboratories
Technical University Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
hb@buchner-net.com

**Abstract.** This paper addresses the tracking capability of blind source separation algorithms for rapidly time-varying sensor or source positions. Based on a known algorithm for blind source separation, which also allows for simultaneous localization of multiple active sources in reverberant environments, the source separation performance will be investigated for abrupt microphone array rotations representing the *worst case*. After illustrating the deficiencies in source-tracking with the given efficient implementation of the BSS algorithm, a method to ensure robust source separation even with abrupt microphone array rotations is proposed. Experimental results illustrate the efficiency of the proposed concept.

## 1 Introduction

This paper is motivated by the so-called *cocktail-party problem* which arises when convolutive mixtures of multiple simultaneously active speakers are recorded by multiple microphones. In many applications (e.g. hands-free human-machine interfaces, [1]), we need to focus on one single source and try to suppress interfering sources. We address this problem here by *blind source separation* (BSS) algorithms which can deal well with unknown microphone and source positions [2]. Furthermore, BSS provides us with several separated source signals which may be individually selected for further processing.

We briefly review the generic ICA-based BSS framework for convolutive mixtures called TRINICON [3,4], which is also capable of simultaneously localizing multiple active sources [5,6]. The motivation for considering it here is that most of the known state-of-the-art BSS algorithms may be seen as certain *approximations* of this concept. As a fairly recent and advanced approximate practical algorithm, we investigate here [7]. This algorithm serves thus as a good representative for many of the major ICA algorithms. It is based on a special choice of Sylvester constraint, the correlation method, the natural gradient, and on an

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 560–568, 2007.

approximated normalization. This allows for an efficient implementation, but may be responsible for the observed deficiencies in certain situations. Although the investigated algorithm is designed for $Q$ active sources, we consider only two active sources in this paper. In Section 2, we demonstrate by simulations that both the separation performance of the considered BSS algorithm as well as the performance of the BSS-based source localization may significantly degrade for rapidly time-varying sensor positions. Analysis of this scenario leads us to proposing the so-called *shadow-BSS* system, which runs in parallel to the main BSS algorithm. Simulation results confirm the efficiency of the proposed extension.

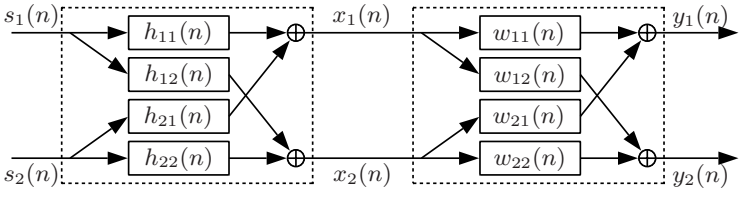


**Fig. 1.** 2-Channel Mixing and Demixing System

Figure 1 illustrates the BSS scheme for two sources with the source signals $s_i(n)$, the sensor signals $x_i(n)$, and the BSS output signals $y_i(n)$, respectively $(i = 1, 2)$. The unknown mixing system is modeled by $M$-tap room impulse responses $h_{ij}(n)$ and the demixing system determined by BSS is modeled by $L$-tap FIR filters $w_{ij}(n)$ $(j = 1, 2)$. The microphone signals and BSS output signals can then be written as:

$$x_i(n) = \sum_{j=1}^{2} \sum_{\kappa=0}^{M-1} h_{ji}(\kappa) s_j(n-\kappa) \tag{1}$$

$$y_i(n) = \sum_{j=1}^{2} \sum_{\kappa=0}^{L-1} w_{ji}(\kappa) x_j(n-\kappa). \tag{2}$$

The source separation problem is then solved by appropriately determined demixing filters. Further details on the adaptation of the demixing filters are given in, e.g., [7].

As shown in [6], TRINICON-based BSS inherently identifies the (unknown) mixing system up to a scaling for $Q = 2$:

$$w_{ji}(n) = -\alpha_{ji} \cdot h_{ji}(n) \qquad \text{and} \qquad w_{jj}(n) = \alpha_{ii} \cdot h_{ii}(n), \qquad i \neq j. \tag{3}$$

Based on this system identification, the TDOA (*Time Difference of Arrival*) can be derived from the demixing filters simultaneously for both active sources from the main peaks of $w_{ij}(n)$ [6].

## 2 BSS and Source Localization in Rapidly Time-Varying Scenarios

We now investigate the separation and localization performance of the chosen BSS algorithm for rapidly time-varying scenarios. The experimental setup is as follows: We assume abrupt rotations of a microphone array consisting of two sensors, which represents a *worst case* for time-variant scenarios. Dealing with rapidly rotating microphone arrays is important in many applications of BSS, e.g. when the microphone array is held and moved by a person. Figure 2 illustrates the DOAs (*Direction of Arrival*) for two typical scenarios: In the first scenario, the broadside of the microphone array points between the two sources and the array is rotated by $\pm 30°$. In the second scenario, one source is located broadside and the other source is on the side after each turn, where the rotations are $\pm 80°$. Note that the array orientation is abruptly changed and that the DOA is measured relative to the broadside direction of the array. We use clean speech signals as sources, which are convolved by measured impulse responses of a low-echoic chamber ($T_{60} \approx 50$ms) to represent the microphone array inputs. All simulations are performed with sampling rate $f_s = 16$kHz and demixing filter length $L = 1024$.

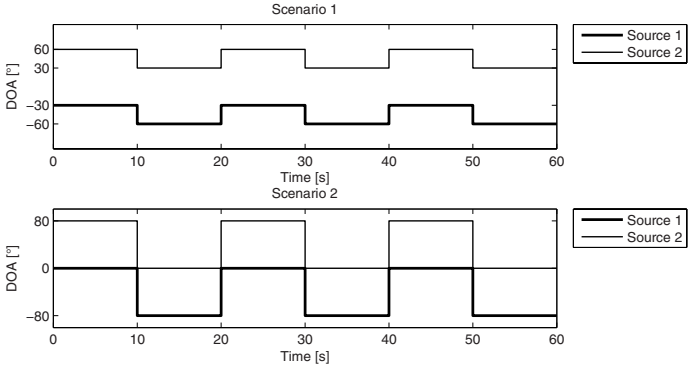


**Fig. 2.** DOAs for two scenarios with abrupt microphone array rotations

Figures 3 and 4 depict both the BSS gain and the results of the source localization obtained with the chosen BSS algorithm for both scenarios, where the BSS gain in dB represents the suppression of the interfering source in each BSS channel. The vertical dashed lines indicate the time instants, when the array orientation is changed. The channel-averaged BSS gain for each array orientation is displayed in the framed boxes. In scenario 1, the chosen BSS algorithm exhibits the expected good source separation and localization performance: After each array rotation, the demixing filters allow for good source separation and the estimated TDOAs follow the DOAs given by scenario 1. In scenario 2, we observe that the investigated BSS algorithm is only able to track the first array rotation (see the time period 10s-20s), but already with a degradation in separation and

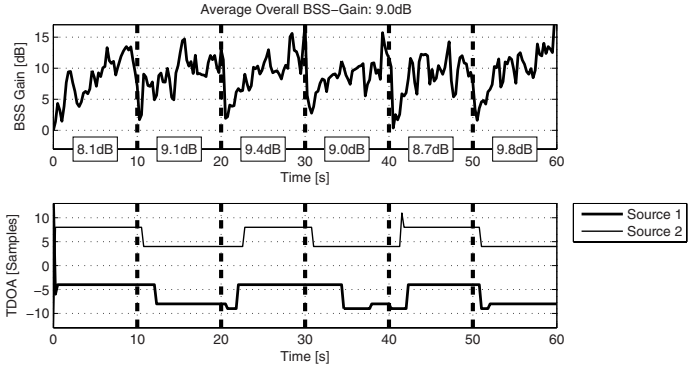


**Fig. 3.** BSS gain (top) and Estimated TDOAs (bottom) obtained with the chosen BSS algorithm, Scenario 1

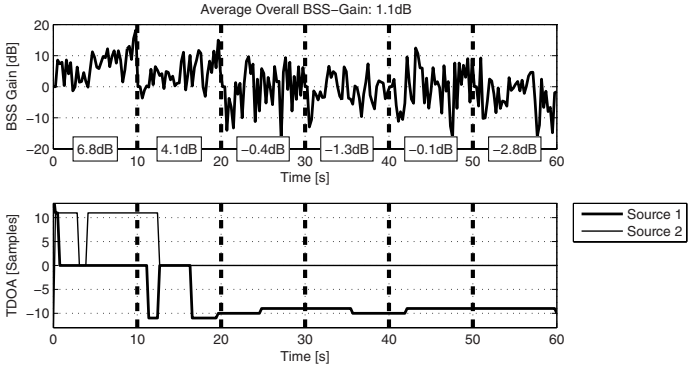


**Fig. 4.** BSS gain (top) and Estimated TDOAs (bottom) obtained with the chosen BSS algorithm, Scenario 2

localization performance. However, the chosen BSS algorithm fails to separate and locate the two sources after the subsequent array turns: The BSS gain is even negative and the estimated TDOAs seem to be "frozen". The results of the TDOA estimation suggest that the chosen BSS algorithm is not able to adapt the demixing filters after the second array rotation. Therefore, we investigate the demixing filters. The first 32 coefficients of the demixing filters $w_{12}(n)$ and $w_{21}(n)$ are depicted in Figure 5. The filter coefficients of $w_{11}(n)$ and $w_{22}(n)$ are not depicted, because they mainly consist of distinct positive peaks at 16 samples, which leads to a delayed but mainly unfiltered contribution of the microphone signals to the two BSS outputs. The vertical dashed lines indicate again array rotations. The horizontal dashed lines mark the filter coefficients which are important for suppressing sources from -80°, 0°, and 80°, respectively. By appropriately placing negative peaks in the demixing filters $w_{12}(n)$ and $w_{21}(n)$, spatial nulls are formed and source 1 and source 2 are suppressed in BSS output

2 and in BSS output 1, respectively. We observe that after the second array rotation, the spatial null in 0°, which is caused by the negative peak at filter coefficient 16 in $w_{21}(n)$, is fixed. This spatial null cancels the source located in broadside direction and thus the source on the side is enhanced. However, the BSS adaptation does not form a significant negative peak in $w_{12}(n)$ to cancel the source at ±80°. Instead, a minor positive peak at filter coefficient 16 is formed, which basically corresponds to a filter-and-sum beamformer at 0° enhancing the broadside source.

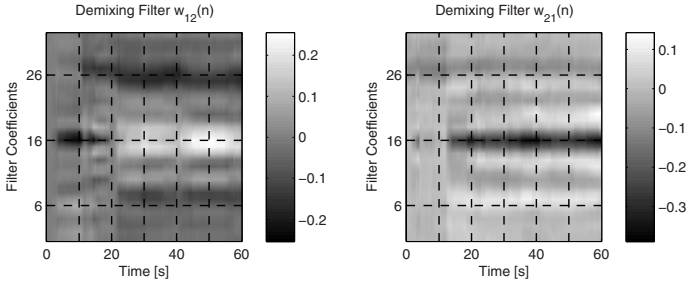


**Fig. 5.** Demixing filters obtained with the chosen BSS algorithm in scenario 2

In order to further analyze and understand the encountered problem, we take a close look at the update rule of the chosen BSS algorithm given in [4]. For each online block $m$, matrix $\mathbf{W}(m)$, which contains the demixing filters $w_{ij}(n)$ in a so-called *Sylvester structure*, is updated as follows:

$$\mathbf{W}(m) = \mathbf{W}(m-1) - \mu \triangle \mathbf{W}(m). \tag{4}$$

Incorporating the *natural gradient*, the update $\triangle\mathbf{W}$ becomes

$$\triangle \mathbf{W} = 2 \sum_{i=0}^{\infty} \beta(i,m) \mathbf{W} \left\{ \mathbf{R_{yy}} - \mathrm{bdiag}\left\{ \mathbf{R_{yy}} \right\} \right\} \cdot \left( \mathrm{bdiag}\left\{ \mathbf{R_{yy}} \right\} \right)^{-1}, \tag{5}$$

where the weighting function $\beta(i,m)$ allows offline and online implementations. The operator $\mathrm{bdiag}\{\mathbf{R_{yy}}\}$ returns the block-diagonal elements of $\mathbf{R_{yy}}$, which is the cross-correlation matrix of the BSS outputs. After abrupt array rotations, the update $\triangle\mathbf{W}$ is then based on badly adapted demixing filters and on BSS output signals, which will not exhibit any source separation. Due to this improper data, the chosen BSS algorithm is not able to adapt the demixing filters and thus fails to track the sources after abrupt array rotations in scenario 2. This deficiency might be caused by the approximations described in [7].

## 3 Shadow-BSS

In Section 2, we studied a scenario, where the chosen BSS algorithm was not able to track abrupt array rotations. However, we could always observe that

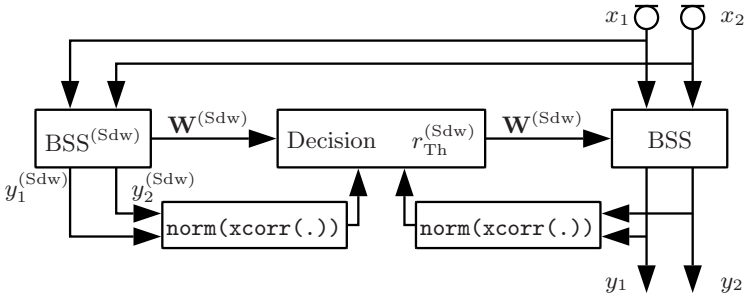


**Fig. 6.** Block diagram of the Shadow-BSS algorithm

the chosen BSS algorithm is capable to converge to well-separating demixing filters after blind initialization. Therefore, we propose the usage of a so-called *shadow-BSS* system, which is periodically blindly initialized and – in the case of outperforming the signal separation of the main BSS system – the demixing filters of the shadow-BSS are transferred for use in the main BSS system. The shadow-BSS system is motivated by the successful usage of shadow systems in, e.g., adaptive echo cancellation [8]. The effectiveness of the shadow-BSS scheme is demonstrated by simulations.

### 3.1 Algorithm

Figure 6 shows the block diagram of the proposed shadow-BSS system, where the dependency on time has been omitted for notational convenience. BSS and BSS$^{(\mathrm{Sdw})}$ denote the main BSS system and the shadow-BSS system. Note that the demixing filter length in the shadow-BSS system $L^{(\mathrm{Sdw})}$ can be chosen independently from $L$. Both systems use the two microphone signals $x_1$ and $x_2$ and perform source separation, which leads to the output signals $y_{1,2}$ and $y_{1,2}^{(\mathrm{Sdw})}$. The shadow-BSS system is periodically blindly reinitialized at multiples of period $T^{(\mathrm{Sdw})}$. We now investigate the method for replacing the demixing filters of BSS by the demixing filters of BSS$^{(\mathrm{Sdw})}$, if BSS$^{(\mathrm{Sdw})}$ performs better source separation.

Based on the two pairs of output signals, the blocks `norm(xcorr(.))` compute the norms of the cross-correlations as follows:

$$\mathrm{norm}\left\{R_{yy}(m)\right\} = \sqrt{\sum_{\tau=-D}^{D} \left|R_{yy}(m,\tau)\right|^2} \tag{6}$$

$$\mathrm{norm}\left\{R_{yy}^{(\mathrm{Sdw})}(m)\right\} = \sqrt{\sum_{\tau=-D}^{D} \left|R_{yy}^{(\mathrm{Sdw})}(m,\tau)\right|^2}. \tag{7}$$

The parameter $\tau$ denotes the time lags of the cross-correlation. The cross-correlation norms may now be considered as a quantity measuring the separation performance of BSS and BSS$^{(\mathrm{Sdw})}$. In the case of good source separation, the

two output signals of a separation system are sufficiently uncorrelated and the according norm in Equations (6), (7) is small. Hence the ratio of both cross-correlation norms $r(m)$,

$$r(m) = \frac{\operatorname{norm}\left\{R_{yy}^{(\mathrm{Sdw})}(m)\right\}}{\operatorname{norm}\left\{R_{yy}(m)\right\}} \tag{8}$$

is used as a decision variable, which indicates the separation system (BSS or $\mathrm{BSS}^{(\mathrm{Sdw})}$) with the better separation performance. To avoid unnecessary transfers of the demixing filters from $\mathrm{BSS}^{(\mathrm{Sdw})}$ to BSS, we can average $r(m)$ with an exponentially decaying forgetting factor $\lambda^{(\mathrm{Sdw})}$:

$$r(m) = \lambda^{(\mathrm{Sdw})} r(m-1) + \left(1-\lambda^{(\mathrm{Sdw})}\right) \frac{\operatorname{norm}\left\{R_{yy}^{(\mathrm{Sdw})}(m)\right\}}{\operatorname{norm}\left\{R_{yy}(m)\right\}}. \tag{9}$$

Comparing $r(m)$ to the threshold $r_{\mathrm{Th}}^{(\mathrm{Sdw})}$ allows for a decision:

$$r(m) < r_{\mathrm{Th}}^{(\mathrm{Sdw})} \Rightarrow \text{Transfer demixing filters from shadow-BSS to BSS}$$
$$r(m) \geq r_{\mathrm{Th}}^{(\mathrm{Sdw})} \Rightarrow \text{Keep demixing filters of BSS}$$

The sensitivity of the overall system may be adjusted by the threshold $r_{\mathrm{Th}}^{(\mathrm{Sdw})}$. Note that the latter $L - L^{(\mathrm{Sdw})}$ filter coefficients of BSS are set to zero, if the demixing filters are transferred from shadow-BSS to BSS.

### 3.2 Simulations

We now present both the separation performance and the results of the TDOA estimation obtained with the proposed algorithm based on a shadow-BSS system for scenario 2 described in Section 2. The demixing filter lengths are $L = 1024$

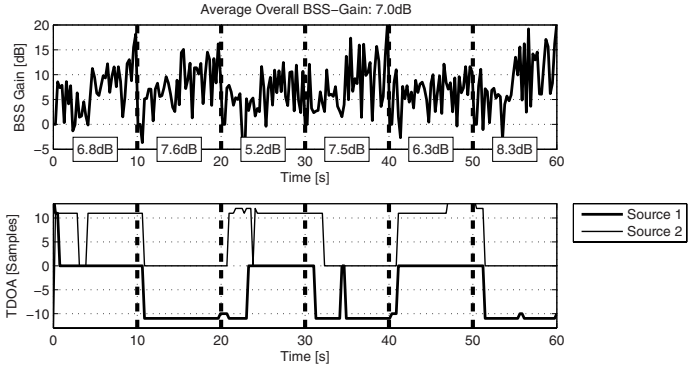


**Fig. 7.** BSS gain (top) and Estimated TDOAs (bottom) obtained with the shadow-BSS system

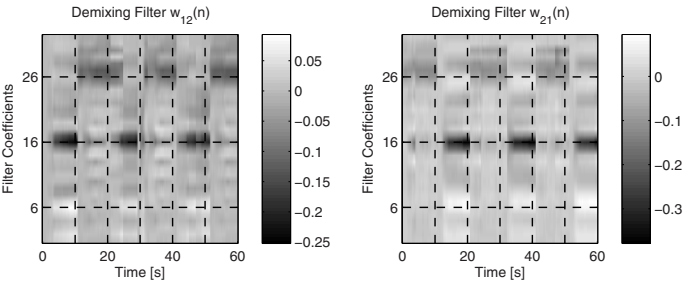


**Fig. 8.** Demixing filters of BSS influenced by shadow-BSS

and $L^{(\mathrm{Sdw})} = 30$. Moreover, the configuration of the shadow-BSS system is $T^{(\mathrm{Sdw})} = 2\mathrm{s}$, $\lambda^{(\mathrm{Sdw})} = 0.6$, and $r_{\mathrm{Th}}^{(\mathrm{Sdw})} = 0.8$.

Both the separation performance of BSS and the estimated TDOAs depicted in Figure 7 illustrate that the proposed shadow-BSS system is capable to track even the worst case of abrupt microphone array rotations. After a few seconds, the estimated TDOAs represent the true DOAs illustrated by Figure 2. Again, we investigated the demixing filters of BSS influenced by shadow-BSS, where we especially focus on the two cross-filters $w_{12}(n)$ and $w_{21}(n)$. As we see from Figure 8, the spatial nulls formed by the demixing filters now follow the alternating DOAs of the sources caused by the abrupt array rotations. Finally, it should be mentioned that no audible artefacts could be observed when the demixing filters are transferred from shadow-BSS to BSS.

## 4 Conclusions

In this paper, we first investigated the source separation and localization performance of the chosen BSS algorithm in the case of abrupt array rotations. We found that for certain relevant cases the chosen BSS algorithm was not able to maintain the usually high separation performance and thus fails to localize the sources correctly. We then proposed an extension to BSS, which incorporates a periodically blindly initialized shadow-BSS system. If the separation performance of the shadow-BSS system outperforms the main BSS system, the demixing filters were transferred from the shadow-BSS system to the main BSS system. An appropriate method for comparing the separation performance of both systems was introduced. Finally, we would like to mention that the proposed shadow-BSS system may also be applied in the case of rapidly moving sources.

## References

1. Brandstein, M.S., Ward, D.B. (eds.): Microphone Arrays: Signal Processing Techniques and Application. Springer, Berlin (2001)
2. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons, New York (2001)

3. Buchner, H., Aichner, R., Kellermann, W.: Audio Signal Processing for Next-Generation Multimedia Communication Systems. Kluwer Academic Publishers, Boston (2004)
4. Buchner, H., Aichner, R., Kellermann, W.: A generalization of blind source separation algorithms for convolutive mixtures based on second order statistics. IEEE Transactions on Speech and Audio Processing 13(1), 120–134 (2005)
5. Buchner, H., Aichner, R., Kellermann, W.: Relation between blind system identification and convolutive blind source separation. In: Conf. Rec. Joint Workshop for Hands-Free Speech Communication and Microphone Arrays (HSCMA) (March 2005)
6. Buchner, H., Aichner, R., Stenglein, J., Teutsch, H., Kellermann, W.: Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering. In: IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP) (March 2005)
7. Aichner, R., Buchner, H., Kellermann, W.: Exploiting narrowband efficiency for broadband convolutive blind source separation. EURASIP Journal on Applied Signal Processing 2007, 1–9 (2006)
8. Ochiai, K., Araseki, T., Ogihara, T.: Echo canceler with two echo path models. IEEE Transactions On. Communications COM-25(6), 589–595 (1977)