Improving the Robustness of the Correlation Approach for Solving the Permutation Problem in the Convolutive Blind Source Separation

Radoslaw Mazur and Alfred Mertins
Institute for Signal Processing
University of Lübeck,
23538 Lübeck, Germany
{mazur,mertins}@isip.uni-luebeck.de

***Abstract*—In this paper, we propose a modification to the correlation approach in convolutive blind source separation to achieve an improved robustness. An often used approach for separation of convolutive mixtures is the transformation to the time-frequency domain. This allows for the use of an instantaneous ICA algorithm independently in each frequency bin, which greatly reduces complexity. The drawback of this approach are the so-called permutation and scaling problems. Here, we modify the well known correlation approach for making it more robust. We propose to incorporate a confidence function based on estimated SIR which allows for detection of frequency bins with high probability of wrong permutations. The results of the new algorithm will be shown on an real-world example.**

***Index Terms*—Blind source separation, convolutive mixture, frequency-domain ICA, permutation problem.**

I. Introduction

Blind Source Separation is a technique for restoring signals from observed mixtures. It is called blind as usually neither the mixing system nor the original signals are known. When dealing with instantaneous cases, a variety of existing algorithms [1], [2], [3] can be used.

This simple approach does not work when dealing with real-world acoustic scenarios. Because of the low speed of sound waves in air, the signals arrive at different times at the microphones. Furthermore, sound waves are reflected on different objects, so the signals arrive at multiple times. This convolutive mixing process can be described using FIR filters. For realistic cases these filters can reach lengths of several thousand coefficients. In such a scenario the task of BSS is then to estimate a set of unmixing filters with at least the same lengths.

These filters can be calculated directly in the time domain [4], [5]. The downside of this approach is the high computational cost and difficulties with convergence, as the algorithms often get trapped in local minima. Therefore an alternative approach can be used: After transformation to the time-frequency domain the convolution becomes a multiplication [6]. This greatly reduces the complexity of the problem as is allows for independent separation in each frequency bin using an instantaneous method. But this simplification has a major downside. The order of the separated signals may differ in every bin. Furthermore every bin has an arbitrary scaling.

Without correcting the scaling, a filtered version of the signals is recovered. The methods proposed in [7], [8] use a postfilter in order to restore the signals as the have been recorded by the microphones. This approach accepts the filtering done by the mixing system without adding new distortion. Alternative methods solve the scaling problem with the aim of filter shortening [9] or shaping [10].

The correction for the permutation problem is even more vital as otherwise the whole separation process will fail. The existing approaches can be divided into two groups. The algorithms for the first group use the properties of the unmixing matrices. The central idea is to see the vectors of the unmixing matrix as beamformers [11] and use them to calculate the direction of arrival. This allows then a depermutation for a plenty of bins, while the remaining bins have to be depermuted using some other method. In [12] and [13], the authors propose an alternative formulation with the use of directivity patterns.

The other group of algorithms use the time structure of the separated bins. Here, the most frequent idea is the assumption of high correlation between neighboring bins. This has been used for example in [7] and [14]. In [15] the authors used the amplitude modulation correlation for getting a separation criterion which avoids the permutation problem. Other approaches include a statistical modeling of the single bins using the generalized Gaussian distribution. Small differences of the parameters lead to a depermutation criterion in [16] and [17].

The depermutation based on the assumption of highly correlated bins is not always very robust. Any improvement of this method would be very beneficial. Here, we present such an improvement based on a confidence function, that is based on the blind estimation of the signal-to-interference ratio (SIR) as proposed in [18]. This approach is based on the observation, that the performance of the correlation approach is considerably degraded at frequency bins which are poorly separated. By leaving out these frequency bins during the primary calculation of the correlation coefficients, the overall depermutation performance can be greatly enhanced.

II. Model and methods

A. BSS for instantaneous mixtures

In this section, we describe the instantaneous unmixing process that we use in frequency bins of the convolutive one. The instantaneous mixing process of N sources into N observations is modeled by an $N \times N$ matrix $\mathbf{A}$. With the source vector $\boldsymbol{s}(n) = [s_1(n), \ldots, s_N(n)]^T$ and negligible measurement noise, the observation signals $\boldsymbol{x}(n) = [x_1(n), \ldots, x_N(n)]^T$ are given by

$$\boldsymbol{x}(n) = \mathbf{A} \cdot \boldsymbol{s}(n). \tag{1}$$

The separation is again a multiplication with a matrix $\mathbf{B}$:

$$\boldsymbol{y}(n) = \mathbf{B} \cdot \boldsymbol{x}(n) \tag{2}$$

with $\boldsymbol{y}(n) = [y_1(n), \ldots, y_N(n)]^T$. The only source of information for the estimation of $\mathbf{B}$ is the observed process $\boldsymbol{x}(n)$. The separation is successful when $\mathbf{B}$ can be estimated so that $\mathbf{BA} = \mathbf{D\Pi}$ with $\mathbf{\Pi}$ being a permutation matrix and $\mathbf{D}$ being an arbitrary diagonal matrix. These two matrices stand for the two ambiguities of BSS. The signals may appear in any order and can be arbitrarily scaled.

For the separation we use the well known gradient-based update rule [1]

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \Delta\mathbf{B}_k \tag{3}$$

with

$$\Delta\mathbf{B}_k = \mu_k(\boldsymbol{I} - E\left\{\boldsymbol{g}(\boldsymbol{y})\boldsymbol{y}^T\right\})\mathbf{B}_k. \tag{4}$$

The term $\boldsymbol{g}(\boldsymbol{y}) = (g_1(y_1), \ldots g_n(y_n))$ is a component-wise vector function of nonlinear score functions

$$g_i(s_i) = -\frac{p_i'(s_i)}{p_i(s_i)} \tag{5}$$

where $p_i(s_i)$ are the assumed source probability densities. These should be known or at least well approximated in order to achieve good separation performance [19].

B. Convolutive mixtures

When dealing with real-world acoustic scenarios it is necessary to consider reverberation. The mixing system can be modeled by FIR filters of length L. Depending on the reverberation time and sampling rate, L can reach several thousand. The convolutive mixing model reads

$$\boldsymbol{x}(n) = \mathbf{H}(n) * \mathbf{s}(n) = \sum_{l=0}^{L-1} \mathbf{H}(l)\boldsymbol{s}(n-l) \tag{6}$$

where $\mathbf{H}(n)$ is a sequence of $N \times N$ matrices containing the impulse responses of the mixing channels. For the separation we use FIR filters of length M and obtain

$$\boldsymbol{y}(n) = \mathbf{W}(n) * \boldsymbol{x}(n) = \sum_{l=0}^{M-1} \mathbf{W}(l)\boldsymbol{x}(n-l) \tag{7}$$

with $\mathbf{W}(n)$ containing the unmixing coefficients. In order to achieve satisfying performance we choose $M \geq L - 1$ [20]. Fig. 1 shows the scenario for two sources and sensors.

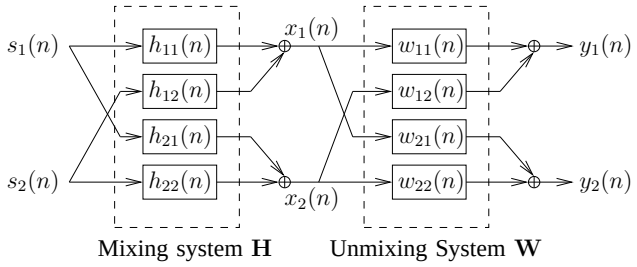

Fig. 1. BSS model with two sources and sensors.

It is possible to estimate $\mathbf{W}(n)$ in the time domain. However, because of the large number of unknowns, MN^2, the existing approaches [4], [5] often suffer problems with convergence.

Using the short-time Fourier transform (STFT), the signals can be transformed to the time-frequency domain, where the convolution approximately becomes a multiplication [6]:

$$\boldsymbol{Y}(\omega_k, \tau) = \boldsymbol{W}(\omega_k)\boldsymbol{X}(\omega_k, \tau), \qquad k = 0, 1, \ldots, K-1, \tag{8}$$

where K is the FFT length. The major benefit of this approach is the possibility to estimate the unmixing matrices for each frequency independently, however, at the price of possible permutation and scaling in each frequency bin:

$$\boldsymbol{Y}(\omega_k, \tau) = \boldsymbol{W}(\omega_k)\boldsymbol{X}(\omega_k, \tau) = \boldsymbol{D}(\omega_k)\mathbf{\Pi}(\omega_k)\boldsymbol{S}(\omega_k, \tau) \tag{9}$$

where $\mathbf{\Pi}(\omega)$ is a frequency-dependent permutation matrix and $\boldsymbol{D}(\omega)$ an arbitrary diagonal scaling matrix.

Without correction of scaling, a filtered version of the sources is recovered. In [7] the authors recovered the sources as they had been recorded by the microphones by using inverse postfilters. This approach does not add any new distortion while accepting the filtering done by the mixing system. A similar technique has been proposed in [8] under the paradigm of the minimal distortion principle, which uses the unmixing matrix

$$\boldsymbol{W}'(\omega) = \mathrm{dg}(\boldsymbol{W}^{-1}(\omega)) \cdot \boldsymbol{W}(\omega) \tag{10}$$

with $\mathrm{dg}(\cdot)$ returning the argument with all off-diagonal elements set to zero. Alternative techniques based on filter-shortening and shaping methods have been proposed in [9][10].

The correction for permutation is essential, as otherwise different signals will be restored at different frequencies and the whole process will fail. In the next section we will review the correlation approach for solving the permutation problem and propose a confidence function in order to make the whole process more robust.

III. Depermutation Algorithm

There are a lot of different depermutation algorithms. Here we use the so-called correlation approach [7], [14]. The key assumption is the high correlation of neighboring bins. With $\boldsymbol{V}(\omega, \tau) = |\boldsymbol{Y}(\omega, \tau)|$, the correlation between two bins k and l is defined as

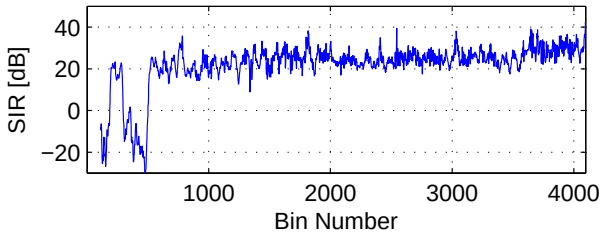

Fig. 2. The bin-wise separation performance using the simple correlation approach. There are three positions where the depermutation fails, which leads to block permutations.

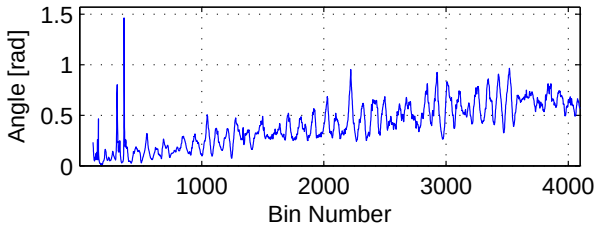

Fig. 3. The calculated angles between the unmixing vectors. The block permutation from Fig. 2 occur where the angles are very small.

$$\rho_{qp}(\omega_k,\omega_l)=\frac{\sum_{\tau=0}^{\mathcal{T}-1}V_q(\omega_k,\tau)V_p(\omega_l,\tau)}{\sqrt{\sum_{\tau=0}^{\mathcal{T}-1}{V_q}^2(\omega_k,\tau)}\sqrt{\sum_{\tau=0}^{\mathcal{T}-1}{V_p}^2(\omega_l,\tau)}} \tag{11}$$

where p,q are the indices of the separated signals, $V_q(\omega_k,\tau)$ is the qth element of $\boldsymbol{V}(\omega_k,\tau)$, and $\mathcal{T}$ is the number of frames. The decision on aligning the bins is made on the basis of the ratio

$$r_{kl}=\frac{\rho_{pp}(\omega_k,\omega_l)+\rho_{qq}(\omega_k,\omega_l)}{\rho_{pq}(\omega_k,\omega_l)+\rho_{qp}(\omega_k,\omega_l)}. \tag{12}$$

By aligning consecutive bins, a correct depermutation for all frequencies should be achieved. In Fig. 2, a result for a real world example is given. Here, the bin-wise separation performance (SIR) [21] for all frequencies is given. A change in the sign of the SIR indicates a permutation. It this example, most bins have been correctly depermuted, but on three positions the procedure failed. As only neighboring bins are compared, the wrong permutations lead to block permutations. In this case, the overall separation performance has been immensely reduced.

This method can be improved by comparing more than one bin. Following this approach in [14] a dyadic sorting scheme has been proposed. Here, in a first step, pairs of neighboring bins are compared. In the next step, these pairs get aligned. By repeating this procedure, the assigned groups get larger in every step and eventually all bins get sorted. The assumption here is, that single false alignments do not outbalance the overall structure. Unfortunately, this is not true if too many errors occur in the early stages.

IV. New algorithm

The approach proposed in this paper is based on the observation, that false alignments usually happen at positions where the separation performance is poor. So, by leaving out these positions during depermutation, a more robust algorithm can be derived. The separation performance in every bin is not known but a blind estimation can be made, as proposed in [18]. The blind estimation of the binwise SIR is based on the observation that the SIR is poor when the vectors $\boldsymbol{w}_i(\omega)$ of the unmixing matrix $\boldsymbol{W}(\omega)$ are similar. This similarity can be measured by the angle between the complex unmixing vectors:

$$\alpha(\omega)=\arccos\left(\frac{|\boldsymbol{w}_1^H(\omega)\boldsymbol{w}_2(\omega)|}{\|\boldsymbol{w}_1(\omega)\|\,\|\boldsymbol{w}_2(\omega)\|}\right) \tag{13}$$

Using the calculated angle $\alpha(\omega)$, an SIR estimate is then computed as in [18]. For the purpose of the algorithm proposed in this paper, this is not needed. The more important fact is the observation that the block permutations happen in close adjacency to the minima of $\alpha(\omega)$, as shown in Fig. 3. This allows for a formulation of a simple binary confidence function $f_c(\omega)$ which is zero for frequency bins, that are in close adjacency to the local minima of $\alpha(\omega)$ and one otherwise.

Using the confidence function leads to the following depermutation algorithm:

1) Calculate $f_c(\omega)$ for all frequencies using $\alpha(\omega)$ and a threshold for the estimation of the uncertain frequencies. The threshold is a trade off, as large values are more robust, but the ranges may become to small for the block depermutation in step 3.
2) Calculate a depermutation using (12) for all frequencies which are confident as indicated by $f_c(\omega)=1$, see Fig. 4 (a).
3) Calculate a block depermutation as in [14] for the depermuted blocks, see Fig. 4 (b).
4) Align the remaining bins by comparing them to the depermuted blocks, see Fig. 4 (c).

TABLE I
Comparison of the results for different depermutation algorithms.

Algorithm	SIR	Algorithm	SIR
Proposed	17.0	DOA-Approach [11]	17.3
Correlation [7]	3.1	$\alpha\beta$-Algorithm [16]	18.4
Dyadic sorting [8]	2.7	Non blind	18.4

V. Simulations

The simulations were performed using data available at [22]. This data set consists of speech recordings that are eight seconds long and sampled at 8 kHz with individual contributions from the sources to the microphones. The chosen parameters were a Hann window of length 2048, a window shift of 256, and an FFT-length of 8192. After 400 iterations

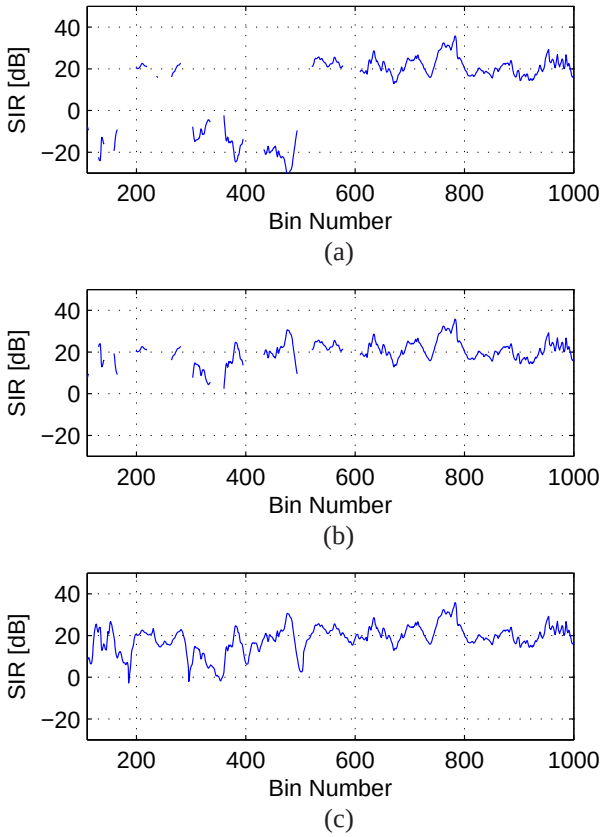

Fig. 4. The new depermutation method shown on the first 1000 bins. (a) Depermutation result at positions where the blind estimated SIR is high enough. (b) Depermutation result after calculation of block correlations. (c) Final result after aligning the remaining bins.

of (4), the depermutation has been performed. The results are shown in Table I.

The results of the plain correlation-based approach could not depermute enough bins, so the overall SIR is very low. The dyadic sorting approach has the same problems, because the early wrong permutations outbalance the whole schema. The new method is able to overcome the problems, and the overall performance is comparable to other state-of-the-art algorithms.

VI. Summary

In this paper we have proposed an extension to the correlation-based depermutation algorithm based on a simple confidence function in order to make it more robust. The new approach is performing better than the simple approach and the dyadic sorting scheme and is comparable to other state-of-the-art algorithms. Results have been shown using real-world data.

References

[1] S.-I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, vol. 8, MIT Press, Cambridge, MA, 1996.

[2] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483–1492, 1997.

[3] J.-F. Cardoso and A. Soulomiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elec. Eng., pt. F.*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[4] S. C. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 1, pp. 92–104, Jan 2005.

[5] R. Aichner, H. Buchner, S. Araki, and S. Makino, "On-line time-domain blind source separation of nonstationary convolved signals," in *Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, Apr. 2003, pp. 987–992.

[6] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain." *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, 1998.

[7] S. Ikeda and N. Murata, "A method of blind separation based on temporal structure of signals." in *Proc. Int. Conf. on Neural Information Processing*, 1998, pp. 737–742.

[8] K. Matsuoka, "Minimal distortion principle for blind source separation," in *Proceedings of the 41st SICE Annual Conference*, vol. 4, 5-7 Aug. 2002, pp. 2138–2143.

[9] R. Mazur and A. Mertins, "Using the scaling ambiguity for filter shortening in convolutive blind source separation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, Taipei, Taiwan, April 2009, pp. 1709–1712.

[10] R. Mazur and A. Mertins, "A method for filter shaping in convolutive blind source separation," in *Independent Component Analysis and Signal Separation (ICA2009)*, ser. LNCS, vol. 5441. Springer, 2009, pp. 282–289.

[11] 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 and Audio Processing*, vol. 12, no. 5, pp. 530–538, Sept. 2004.

[12] W. Wang, J. A. Chambers, and S. Sanei, "A novel hybrid approach to the permutation problem of frequency domain blind source separation," in *Lecture Notes in Computer Science*, vol. 3195. Springer, 2004, pp. 532–539.

[13] M. Z. Ikram and D. R. Morgan, "Permutation inconsistency in blind speech separation: investigation and solutions," *IEEE Transactions on Speech and Audio Processing.*, vol. 13, no. 1, pp. 1–13, Jan. 2005.

[14] K. Rahbar and J. P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, pp. 832–844, Sept. 2005.

[15] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proceddings of the second international workshop on independent component analysis and blind signal separation*, 2000, pp. 215–220.

[16] R. Mazur and A. Mertins, "An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 117–126, Jan. 2009.

[17] R. Mazur and A. Mertins, "Simplified formulation of a depermutation criterion in convolutive blind source separation," in *Proc. European Signal Processing Conference*, Glasgow, Scotland, Aug 2009, pp. 1467–1470.

[18] R. Mazur and A. Mertins, "On separation performance enhancement in convolutive blind source separation," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Oct. 2008, pp. 1718–1721.

[19] S. Choi, A. Cichocki, and S. Amari, "Flexible independent component analysis," in *Neural Networks for Signal Processing VIII*, T. Constantinides, S. Y. Kung, M. Niranjan, and E. Wilson, Eds., 1998, pp. 83–92. [Online]. Available: citeseer.ist.psu.edu/choi00flexible.html

[20] K. Rahbar and J. P. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, 7-11 May 2001, pp. 2745–2748.

[21] D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," in *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation*, Aussois, France, Jan. 1999.

[22] http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/bss2to4/index.html.