SIMPLIFIED FORMULATION OF A DEPERMUTATION CRITERION IN 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

For the separation of convolutive mixtures, an often used approach is the transformation to the time-frequency domain, where the problem is reduced to multiple instantaneous mixtures. This allows for the employment of well-known ICA algorithms. The drawbacks of this method are the inherent permutation and scaling problems. These ambiguities have to be corrected before a transformation back to the time domain can be carried out. The scaling ambiguity is usually solved using the minimal distortion principle. For the permutation problem, several approaches have been proposed. In this paper we propose a modification of an existing algorithm with the aim of simplifying the depermutation criterion and the corresponding computational effort while maintaining the same performance.

1. INTRODUCTION

Different methods of independent component analysis (ICA) and blind source separation (BSS) have been proposed for the separation of linear instantaneous mixtures [1, 2, 3]. With real-world mixtures of audio signals, the situation becomes more complicated. As the signals arrive multiple times with different lags, the mixing process is convolutive. Usually, it can be modelled using FIR filters, but for realistic scenarios the length of the filters can reach up to several thousand. The task of BSS is then to estimate a system of unmixing filters, which ideally have at least the order of the mixing filters.

There exist methods for calculating such filters directly in the time domain [4, 5]. The drawback of these methods is the high computational cost and difficulties of convergence. A much more promising way is the transformation of the signals to the time-frequency domain where the convolution becomes a multiplication [6]. Using this approach, an instantaneous ICA method can be applied to each frequency bin independently. The problems arising from this approach are the arbitrary scaling and permutation in every bin. Without correction of the scaling ambiguity the restored signals are arbitrarily filtered, but with the minimal distortion principle, proposed in [7], an acceptable solution is found.

The correction of the different permutations is even more important, as otherwise the whole separation process will fail. There have been proposed different approaches for this problem. One class of algorithms utilizes the properties of the unmixing matrices. In [8] the authors propose to use these as beamformers. With the calculation of the direction of arrival, depermutation could be achieved for most of the bins. In [9] and [10], the authors propose an alternative formulation with the use of directivity patterns.

Another class of algorithms uses the time structure of the separated bins. The assumption of high correlation in neighboring bins has been used for the definition of a depermutation criterion in [11]. In [12] the authors propose to model every bin using a generalized Gaussian distribution (GGD) and to employ the small differences of the parameters between neighboring bins for a calculation of correct assignments. In this paper, we propose a modification of this algorithm with the aim of simplified formulation and reduced calculation, while maintaining the same overall separation performance.

2. MODEL AND METHODS

2.1 BSS for instantaneous mixtures

The instantaneous mixing process of N sources into N observations can be modeled by an $N \times N$ matrix $\mathbf{A}$. Neglecting the measurement noise, a given source vector $\mathbf{s}(n) = [s_1(n), \ldots, s_N(n)]^T$ is transformed to an observation $\mathbf{x}(n) = [x_1(n), \ldots, x_N(n)]^T$ by

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

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

$$\mathbf{y}(n) = [y_1(n), \ldots, y_N(n)]^T = \mathbf{B} \cdot \mathbf{x}(n) \tag{2}$$

The only sources of information for estimating $\mathbf{B}$ are the statistical properties of the observed signals $\mathbf{x}(n)$. When $\mathbf{BA} = \mathbf{D\Pi}$ with $\mathbf{\Pi}$ being a permutation matrix and $\mathbf{D}$ an arbitrary diagonal matrix, the separation is successful. The two matrices represent the ambiguities of BSS: (1) The order of the separated signals is arbitrary, and (2) they are only scaled versions of the sources.

In the present work, for learning unmixing matrices $\mathbf{B}$, we use the well-known gradient-based update rule [1] $\mathbf{B}_{k+1} = \mathbf{B}_k + \Delta\mathbf{B}_k$ with

$$\Delta\mathbf{B}_k = \mu_k(\mathbf{I} - E\left\{\boldsymbol{g}(\mathbf{y})\mathbf{y}^H\right\})\mathbf{B}_k \tag{3}$$

and $\boldsymbol{g}(\mathbf{y}) = [g_i(y_i), \ldots, g_n(y_n)]$ being a component-wise vector function of nonlinear score functions $g_i(s_i) = -p_i'(s_i)/p_i(s_i)$, where $p_i(s_i)$ are the assumed source probability densities.

It is necessary to know, or at least well approximate, the probability density functions of the sources in order to achieve good separation performance. In [13] the authors use the GGD with some fixed parameters, while in [14] the

parameters are estimated on the basis of the separated signals after each iteration of (3).

With the GGD defined as

$$p_y(y) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|y|/\alpha)^\beta} \tag{4}$$

with $\alpha, \beta > 0$ and $\Gamma(\cdot)$ being the Gamma function given by

$$\Gamma(y) = \int_0^\infty x^{y-1} e^{-x} dx \tag{5}$$

the nonlinear score function reduces to

$$g_i(x_i) = \frac{x_i}{|x_i|^{2-\beta}}. \tag{6}$$

For the complex case the GGD is assumed to be spherical symmetric in the z-plane around the origin. This assumption yields the same nonlinear score function as in (6). The validity of this approach is shown in [15].

2.2 Convolutive mixtures

For real-world acoustic scenarios, the mixing model has to be modified due to the convolutive properties. It can be modeled by FIR filters of length L where L can go beyond 2000, depending on the sampling rate and reverberation time. The convolutive mixing model reads

$$\mathbf{x}(n) = \sum_{l=0}^{L-1} \mathbf{H}(l)\mathbf{s}(n-l) \tag{7}$$

where $\mathbf{H}(n)$ is a sequence of $N \times N$ matrices containing the impulse responses of the mixing channels. For the separation, one can use FIR filters of length $M \geq (N-1)(L-1)+1$ [16] and obtain

$$\mathbf{y}(n) = \sum_{l=0}^{M-1} \mathbf{W}(l)\mathbf{x}(n-l) \tag{8}$$

with $\mathbf{W}(n)$ containing the unmixing coefficients.

The direct estimation of $\mathbf{W}(n)$ in the time domain is very difficult due to the large number of coefficients. The existing approaches [4, 5] are not satisfying because of distortion introduced by the unmixing system. To circumvent this problem, an approach in the time-frequency domain is often used. Using the blockwise short-time Fourier transform (STFT), the convolution becomes a multiplication [6]:

$$\boldsymbol{Y}(\omega_k, \tau) = \boldsymbol{W}(\omega_k)\boldsymbol{X}(\omega_k, \tau). \tag{9}$$

With this formulation, the coefficients for each frequency bin can be estimated separately. This simplification comes at the price of each frequency bin being differently scaled and permuted:

$$\boldsymbol{Y}(\omega_k, \tau) = \boldsymbol{W}(\omega_k)\boldsymbol{X}(\omega_k, \tau) = \boldsymbol{D}(\omega_k)\boldsymbol{\Pi}(\omega_k)\boldsymbol{S}(\omega_k, \tau) \tag{10}$$

where $\boldsymbol{\Pi}(\omega)$ is a frequency-dependent permutation matrix and $\boldsymbol{D}(\omega)$ is an arbitrary diagonal scaling matrix. If the permutations are not corrected, the whole separation process will fail, as parts of all signals can appear in every output channel. Without correcting the scaling, a filtered version of the sources is recovered. In [17] the authors proposed to use inverse postfilters for restoring the signals as they have been recorded by the microphones. This approach accepts the filtering done by the mixing system without adding new distortion. In [7] a similar method was proposed which is called the minimal distortion principle. Newer approaches solve the scaling ambiguity with the aim of filter shortening [18] or shaping [19]. In the following, we use the unmixing matrix

$$\boldsymbol{W}'(\omega) = \text{diag}[\boldsymbol{W}^{-1}(\omega)] \cdot \boldsymbol{W}(\omega), \tag{11}$$

which corresponds to the method in [7], with diag[·] returning the argument with all off-diagonal elements set to zero.

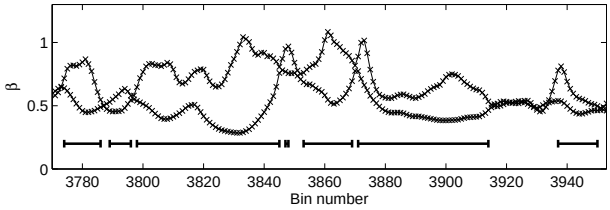

Figure 1: Beta values of two signals over the frequency index. The detected clusters are indicated with bars ⊢⊣.

3. DEPERMUTATION ALGORITHM

There exist different methods [17, 11] which utilize 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

$$\rho_{qp}(\omega_k, \omega_l) = \frac{\sum_{\tau=0}^{\mathcal{T}-1} V_q(\omega_k, \tau) V_p(\omega_l, \tau)}{\sqrt{\sum_{\tau=0}^{\mathcal{T}-1} {V_q}^2(\omega_k, \tau)} \sqrt{\sum_{\tau=0}^{\mathcal{T}-1} {V_p}^2(\omega_l, \tau)}} \tag{12}$$

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

$$r_{kl} = \frac{\rho_{pp}(\omega_k, \omega_l) + \rho_{qq}(\omega_k, \omega_l)}{\rho_{pq}(\omega_k, \omega_l) + \rho_{qp}(\omega_k, \omega_l)}. \tag{13}$$

It is assumed that with $r_{kl} > 1$ the bins are correctly aligned and otherwise a permutation has occurred. The problem arising here is that the assumption of highly correlated bins cannot always be made, especially when the bins are poorly separated. In [11] the authors proposed to use a dyadic sorting scheme. They start with pairwise comparisons and then arrange these pairs to new larger groups. By repeating this procedure recursively, all bins can be grouped, and single false permutations do not unbalance the overall structure. Unfortunately, this is not true if too many errors occur in the early stages.

In [12], the authors propose an alternative strategy. The main difference is the calculation of the starting clusters, which are then depermuted using a correlation approach similar to the one mentioned above. Although not all bins are sorted in the first step, the calculated clusters have only correct assignments. Using such clusters, the subsequent dyadic sorting is much more effective.

The idea is to model every frequency bin using the GGD and to use the small differences of the parameters α and β in neighboring bins for defining a depermutation criterion. In Figure 1, such a situation is shown. Here we see β values

for the two signals and marked areas where the clustering procedure was successful.

The clustering procedure as described in [12] is implemented independently for both parameters α and β. After building the clusters, the overlapping parts are used to create larger ones. This way, more bins are assigned in fewer clusters, which is a better starting point for the depermutation using the correlation coefficients.

The rules for clustering in [12] have the following style:

$$\begin{aligned} \beta_{\mathrm{H}}(\omega_l) &> k_1 \cdot \beta_{\mathrm{L}}(\omega_l) \\ \text{and} \quad \beta_{\mathrm{H}}(\omega_{l+1}) &> k_1 \cdot \beta_{\mathrm{L}}(\omega_{l+1}) \end{aligned} \tag{14}$$

with

$$\beta_{\mathrm{H}}(\omega) = \max[\beta(\omega,p),\beta(\omega,q)] \tag{15}$$
$$\beta_{\mathrm{L}}(\omega) = \min[\beta(\omega,p),\beta(\omega,q)] \tag{16}$$

and some constant k_1. If (14) is fulfilled, the next bin can be added to the cluster. For clustering using the parameter α, a similar procedure is utilized, but due to different properties, the logarithm of the values is used. Overall there are eight equations with nine constants which makes the procedure quite complicated.

4. NOVEL ALTERNATIVE FORMULATION

Here we propose to use an alternative formulation that makes the calculation drastically easier. Instead of comparing $\beta_{\mathrm{H}}(\omega_l)$ and $\beta_{\mathrm{L}}(\omega_l)$, we use the difference:

$$\beta_{pq}(\omega) = \beta(\omega,p) - \beta(\omega,q) \tag{17}$$

For the clustering using the α parameter, we define correspondingly

$$\alpha_{pq}(\omega) = \log[\alpha(\omega,p)] - \log[\alpha(\omega,q)] \tag{18}$$

The values of $\beta_{pq}(\omega)$ and $\alpha_{pq}(\omega)$ are quite similar in adjacent bins. Therefore it is possible to make a prediction based on preceding bins. With $\beta_{\mathrm{pred}_{pq}}(\omega)$ being the predicted value, the comparison of

$$\beta_{\mathrm{d}_1}(\omega_{l+1}) = \beta_{\mathrm{pred}_{pq}}(\omega_{l+1}) - \beta_{pq}(\omega_{l+1}) \tag{19}$$
$$\beta_{\mathrm{d}_2}(\omega_{l+1}) = \beta_{\mathrm{pred}_{pq}}(\omega_{l+1}) - \beta_{qp}(\omega_{l+1}) \tag{20}$$

yields a depermutation criterion. When β_{d_1} and β_{d_2} differ substantially, then the correct permutation can be determined. The ratio

$$r_l = \frac{\beta_{\mathrm{d}_1}(\omega_{l+1})}{\beta_{\mathrm{d}_2}(\omega_{l+1})} \tag{21}$$

shows the correct permutation. With $r_l < 1$ the bins are correctly aligned and $r_l > 1$ indicates a permutation. For more reliable clustering, an error margin is advisable, where the comparisons $r_l < 1/\rho$ and $r_l > \rho$ with $\rho > 1$ are used.

For the prediction of $\beta_{\mathrm{pred}_{pq}}(\omega)$, a linear predictor can be used. Some tests with different lengths showed that even very short linear predictors deliver good results. Using longer linear predictors can be better for particular signals, but then the generalization suffers. Therefore we use a linear predictor with two coefficients which just do a linear extrapolation. The prediction error under the two possible permutations then reads

$$\beta_{\mathrm{d}_1}(\omega_{l+1}) = -\beta_{pq}(\omega_{l-1}) + 2\beta_{pq}(\omega_l) - \beta_{pq}(\omega_{l+1}), \tag{22}$$
$$\beta_{\mathrm{d}_2}(\omega_{l+1}) = -\beta_{pq}(\omega_{l-1}) + 2\beta_{pq}(\omega_l) - \beta_{qp}(\omega_{l+1}). \tag{23}$$

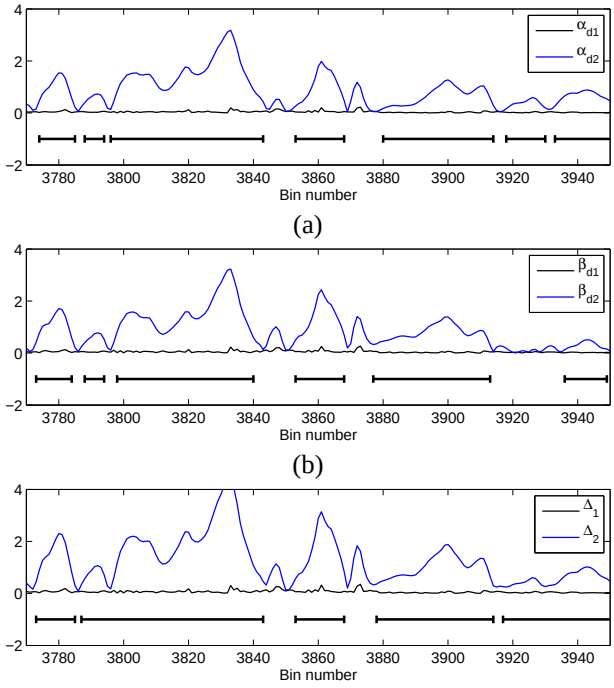

Figure 2: Detected Clusters. (a) New algorithm using α. (b) New algorithm using β. (c) Joint algorithm.

For clustering using α, we propose correspondingly:

$$\alpha_{\mathrm{d}_1}(\omega_{l+1}) = -\alpha_{pq}(\omega_{l-1}) + 2\alpha_{pq}(\omega_l) - \alpha_{pq}(\omega_{l+1}) \tag{24}$$
$$\alpha_{\mathrm{d}_2}(\omega_{l+1}) = -\alpha_{pq}(\omega_{l-1}) + 2\alpha_{pq}(\omega_l) - \alpha_{qp}(\omega_{l+1}) \tag{25}$$

In Figures 2(a) and 2(b), the results for a real-world case are shown. As one can see, the values α_{d_1} and β_{d_1} are quite small for almost all frequency bins, while the values β_{d_2} and α_{d_2} are large for most of the frequencies. Since small α_{d_1} and β_{d_1} and large β_{d_2} and α_{d_2} indicate equal permutations of adjacent bins, this means that the method correctly identifies most permutations as correctly aligned. Thus, using either α or β, most of the bins can be clustered, as marked with the underlying bars. The next step according to [12] would be to join both cluster types by estimating overlapping parts.

In this paper, we use another approach. With

$$\mathbf{z}_{pq}(\omega) = \left[\frac{\alpha_{pq}(\omega)}{\sigma_\alpha}, \frac{\beta_{pq}(\omega)}{\sigma_\beta}\right]^T \tag{26}$$

and σ_α and σ_β being the standard deviations for α_{pq} and β_{pq}, respectively, a joint criterion can be derived. For this, we define

$$\Delta_1(\omega_{l+1}) = \|-\mathbf{z}_{pq}(\omega_{l-1}) + 2\mathbf{z}_{pq}(\omega_l) - \mathbf{z}_{pq}(\omega_{l+1})\|_{\ell_2}, \tag{27}$$
$$\Delta_2(\omega_{l+1}) = \|-\mathbf{z}_{pq}(\omega_{l-1}) + 2\mathbf{z}_{pq}(\omega_l) - \mathbf{z}_{qp}(\omega_{l+1})\|_{\ell_2}. \tag{28}$$

The final decision is made in analogy to (21) on the basis of

$$r_l = \frac{\Delta_1(\omega_{l+1})}{\Delta_2(\omega_{l+1})} \tag{29}$$

with $r_l < 1/\rho$ showing correct alignment and $r_l > \rho$ indicating a permutation.

The new algorithm now depends on just one parameter ρ. In Figure 2(c) the resulting clustering for this method is shown. Especially, some bins could be clustered which had been left out by the single approaches.

Table 1: Comparison of cluster sizes using the algorithms from [12]. Only clusters with size larger than eight are counted.

	Number	Clustered bins	Avg. Cluster Sizes
α-Cluster	92	3085	33.53
β-Cluster	90	3236	35.96
Res. Cluster	62	3402	54.87

Table 2: Comparison of cluster sizes using the new algorithm. Only clusters with size larger than eight are counted.

	Number	Clustered bins	Avg. Cluster Sizes
α-Cluster	106	3371	31.80
β-Cluster	98	3159	32.23
Joint Cluster	71	3706	52.20

5. SIMULATIONS

The simulations have been performed using data available at [20]. This data set consists of eight seconds long speech recordings sampled at 8 kHz with individual contributions from the sources to the micropohones. The chosen parameters were a Hanning window of length 2048, a window shift of 256, and an FFT-length of 8192. After 400 iterations of (3), the depermutation has been performed using either the old or the new algorithm for the first clustering stage. The following stage of cluster correlation was carried out using the method from [18].

The results of the algorithm from [12] are shown in Table 1. The results for the new algorithm are given in Table 2. They show that even more bins could be clustered, but that the average cluster size is slightly smaller. This is due to the effect of additional small clusters which actually improve the performance. The next step, in which the cluster permutations are aligned using the correlation approach described in [18], could depermute all bins. The overall separation performance for channels 1 and 2 from [20] was 18.07 dB.

6. SUMMARY

In this paper we have proposed a modification of a depermutation algorithm that is used in convolutive blind source separation. The depermutation criterion is greatly simplified, while the overall performance is maintained. Results have been shown using real-world data.

REFERENCES

[1] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 757–763. The MIT Press, 1996.

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

[3] J.-F. Cardoso and A. Soulomiac. Blind beamforming for non-Gaussian signals. *Proc. Inst. Elec. Eng., pt. F.*, 140(6):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*, 13(1):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)*, pages 987–992, Nara, Japan, April 2003.

[6] P. Smaragdis. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22(1-3):21–34, 1998.

[7] K. Matsuoka. Minimal distortion principle for blind source separation. In *Proceedings of the 41st SICE Annual Conference*, volume 4, pages 2138–2143, 5-7 Aug. 2002.

[8] 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*, 12(5):530–538, Sept. 2004.

[9] 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*, volume 3195, pages 532–539. Springer, 2004.

[10] M. Z. Ikram and D. R. Morgan. Permutation inconsistency in blind speech separation: investigation and solutions. *IEEE Transactions on Speech and Audio Processing.*, 13(1):1–13, Jan. 2005.

[11] K. Rahbar and J. P. Reilly. A frequency domain method for blind source separation of convolutive audio mixtures. *IEEE Trans. Speech and Audio Processing*, 13(5):832–844, Sept. 2005.

[12] 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*, 17(1):117–126, Jan. 2009.

[13] S. Choi, A. Cichocki, and S. Amari. Flexible independent component analysis. In T. Constantinides, S. Y. Kung, M. Niranjan, and E. Wilson, editors, *Neural Networks for Signal Processing VIII*, pages 83–92, 1998.

[14] K. Kokkinakis and A. K. Nandi. *Multichannel Speech Separation Using Adaptive Parameterization of Source PDFs*, volume 3195 of *Lecture Notes in Computer Science*. Springer, 2004.

[15] I. Lee, T. Kim, and T.-W. Lee. Independent vector analysis for convolutive blind speech separation. In *Blind Speech Separation*, pages 169–192. Springer Netherlands, 2007.

[16] 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*, volume 5, pages 2745–2748, 7-11 May 2001.

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

[18] 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.

[19] R. Mazur and A. Mertins. A method for filter shaping in convolutive blind source separation. In *Independent Component Analysis and Signal Separation (ICA2009)*, LNCS. Springer, 2009.

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