A Method for Filter Equalization 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. The separation of convolutive mixed signals can be carried out in the time-frequency domain, where the task is reduced to multiple instantaneous problems. This direct approach leads to the permutation and scaling problems, but it is possible to introduce an objective function in the time-frequency domain and minimize it with respect to the time domain coefficients. While this approach allows for the elimination of the permutation problem, the unmixing filters can be quite distorted due the unsolved scaling problem. In this paper we propose a method for equalization of these filters by using the scaling ambiguity. The resulting filters have a characteristic of a Dirac pulse and introduce less distortion to the separated signals. The results are shown on a real-world example.

1 Introduction

The blind source separation method (BSS) is used to recover signals from observed mixtures. It is called blind as neither the original signals nor the mixing system is known. For the instantaneous case, different methods have been proposed [2,6,4]. In the case of real world acoustic mixtures of speech the situation is more complicated, as the signals arrive multiple times with different lags. This behavior can be modeled using FIR filters, but for realistic scenarios the length can reach several thousand taps. In this case, the separation can be performed only using filters with similar lengths.

An often used approach is the transformation to the time-frequency domain where the convolution becomes a multiplication [18]. This allows the use of instantaneous methods in each frequency bin independently. The major drawback is the arbitrary permutation in each frequency bin which has to be corrected, or the whole process fails. Although different methods have been proposed [15,3,10,16,19,14], the correction can not be calculated reliable in all cases.

The unmixing filters can be calculated directly in the time domain [5,1] but these algorithms suffer from high computational costs. In [13] an alternative method has been proposed, where the objective function is formulated in the frequency domain and minimized with respect to the time coefficients. This method combines the effectiveness of the frequency domain approaches with the absence of the permutation problem of the time domain methods. However, the scaling in the different frequencies is not addressed and therefore quite arbitrarily. This leads to coloration and added reverberation in the separated signals.

V. Vigneron et al. (Eds.): LVA/ICA 2010, LNCS 6365, pp. 328–336, 2010.

The postfilter-method in [7] tries to recover the signals as they have been recorded at the microphones and thus accepts all filtering done by the mixing system without adding new distortions. In [9], with the minimal distortion principle, a similar technique has been proposed.

New approaches as proposed in [11] and [12] solve the scaling problem with the aim of filter shortening or shaping. The methods of [11] and [12] use the instantaneous separation in the time-frequency domain as in [18] and allow for a simple calculation of the scaling coefficients. As these approaches are able to enhance the separation performance and reduce the reverberation at the same time, we propose to use these methods in combination with the algorithm from [13]. As this algorithm calculates only time domain unmixing coefficients the calculation of the scaling coefficients has to be modified. In this paper we will show how to calculate the scaling coefficients in this setup and apply an equalization method. The results will be shown in a real world example.

2 Problem Statement

The mixing system of real-world acoustic scenarios is convolutive and can be described using FIR filters of length L where L can reach several thousand. With N sources and N mixtures, the source vector $\boldsymbol{s}(n) = [s_1(n), \ldots, s_N(n)]^T$, and negligible measurement noise, the observation signals are given by

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

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 $M \geq L-1$ and obtain

$$\boldsymbol{y}(n) = \mathbf{W}(n) * \boldsymbol{x}(n) = \sum_{l=0}^{M-1} \mathbf{W}(l)\boldsymbol{x}(n-l) \tag{2}$$

with $\boldsymbol{y}(n) = [y_1(n), \ldots, y_N(n)]^T$ being the vector of separated outputs and $\mathbf{W}(n)$ containing the time domain unmixing coefficients. Fig. 1 shows the scenario for two sources and sensors. The overall system is given by

$$\boldsymbol{y}(n) = \mathbf{W}(n) * \mathbf{H}(n) * \boldsymbol{s}(n) = \mathbf{G}(n) * \boldsymbol{s}(n), \tag{3}$$

which reduces to a multiplication in the time-frequency domain:

$$\boldsymbol{Y}(\omega, \tau) \approx \boldsymbol{W}(\omega)\boldsymbol{H}(\omega)\boldsymbol{S}(\omega, \tau) = \boldsymbol{G}(\omega)\boldsymbol{S}(\omega, \tau). \tag{4}$$

The only sources of information for estimating $\mathbf{W}(n)$ are the statistical properties of the observed signals $\boldsymbol{x}(n)$. Using the time-frequency approaches the overall system can only be estimated up to an arbitrary order and scaling:

$$\boldsymbol{G}(\omega) = \boldsymbol{P}(\omega)\boldsymbol{D}(\omega) \tag{5}$$

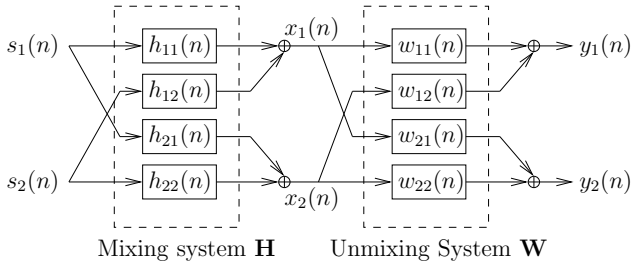

Fig. 1. BSS model with two sources and sensors

with $\boldsymbol{P}(\omega)$ being a permutation matrix and $\boldsymbol{D}(\omega)$ an arbitrary diagonal matrix. For a successful separation the permutation matrices $\boldsymbol{P}(\omega)$ have to be the same at all frequencies. All matrices $\boldsymbol{D}(\omega)$ which are not the identity introduce filtering to the separated signals.

3 Blind Separation Algorithm

The method from [13] does not suffer from the previously addressed permutation problem. This is achieved by using the integrated Kullback-Leibler divergence in the frequency domain as the objective function and minimizing it with respect to the time-domain matrices $\mathbf{W}(n)$. The update rule reads

$$\mathbf{W}^{l+1}(n) = \mathbf{W}^{l}(n) - \mu \frac{\partial f(\boldsymbol{\mathcal{W}})}{\partial \mathbf{W}^{l}(n)} \tag{6}$$

with l being the iteration index, $f(\cdot)$ the integrated Kullback-Leibler divergence and $\boldsymbol{\mathcal{W}} = [\mathbf{W}(0), \mathbf{W}(1), \ldots, \mathbf{W}(M-1)]$. The gradient is calculated in [13] as

$$\frac{\partial f(\boldsymbol{\mathcal{W}}^{l})}{\partial \mathbf{W}^{l}(n)} = \int_{-\pi}^{\pi} [\boldsymbol{I} - \boldsymbol{D}^{-1}(l,\omega)\boldsymbol{P}(l,\omega)]\boldsymbol{W}^{l}(e^{j\omega})e^{j\omega n}d\omega \tag{7}$$

where

$$\boldsymbol{D}(l,\omega) = \operatorname{diag}\left([\sigma_1^{r_1}(l,\omega), ..., \sigma_N^{r_N}(l,\omega)]^{\mathrm{T}}\right) \tag{8}$$

and

$$\boldsymbol{P}(l,\omega) = \boldsymbol{Y}^{r-1}(l,e^{j\omega})\boldsymbol{Y}^{\mathrm{H}}(l,e^{j\omega}), \tag{9}$$

$$\begin{aligned}\mathbf{Y}^{r-1}(l,e^{j\omega}) = \Big[&\left|Y_1(l,e^{j\omega})\right|^{r_1-1} e^{j\theta(Y_1(l,e^{j\omega}))}, \ldots \\ &\ldots, \left|Y_N(l,e^{j\omega})\right|^{r_N-1} e^{j\theta(Y_N(l,e^{j\omega}))}\Big]^{\mathrm{T}}\end{aligned} \tag{10}$$

with $Y_i(e^{j\omega})$ being the short-time Fourier transforms of $y_i(n)$, $i = 1, 2, \ldots, N$ and

$$\sigma_p^{r_p}(l,\omega) = \beta\sigma_p^{r_p}(l,\omega) + (1-\beta)\left|y_p(l,e^{j\omega})\right|^{r_p}. \tag{11}$$

The parameter β with $0 < \beta < 1$ is a moving-average parameter, and r_p is the order of an assumed generalized Gaussian source model.

This method is able to separate real-room recordings as it is capable of dealing with long filters, but it suffers from linear distortions which are introduced by the unmixing filters. A new method for resolving this problem will be presented in the next section.

4 Resolving the Scaling Ambiguity

A commonly used method for solving the scaling ambiguity is the minimal distortion principle (MDP) as proposed in [9]. The frequency unmixing matrices are calculated as

$$\boldsymbol{W}'(\omega) = \mathrm{dg}(\boldsymbol{W}^{-1}(\omega)) \cdot \boldsymbol{W}(\omega) \tag{12}$$

with $\mathrm{dg}(\cdot)$ returning the argument with all off-diagonal elements set to zero. In conjunction with the BSS algorithm presented in the last section there are two possibilities of employing it. The simple way is to carry out the iteration as in equation (6) and after convergence, transform the filters to the frequency domain by the Discrete Fourier Transform (DFT) where (12) can be carried out for all frequencies. The time-domain filters are then obtained by the inverse DFT. A better approach is to apply the MDP after every step of (6). As it will be shown in the simulations section, this method is able to greatly enhance the separation performance.

Both methods yield filters that have quite arbitrary form. Besides the main peak the filters have lots of large coefficients, which leads to coloration and reverberation in the separated signals. For reducing this coloration we propose to adapt the method from [12]. For this, it needs to be changed from a frequency by frequency method to an algorithm in which all frequencies are processed jointly. The time domain unmixing filters $\boldsymbol{w}_{ij}$ can be calculated in the dependency of the scaling coefficients $\boldsymbol{c}_j = [c_j(\omega_0), \ldots, c_j(\omega_{K-1})]^T$ as

$$\begin{aligned} \boldsymbol{w}_{ij} &= \bar{\mathcal{F}} \cdot \boldsymbol{E}_{ij} \cdot \boldsymbol{B} \cdot \boldsymbol{c}_j \\ &= \boldsymbol{V}_{ij} \cdot \boldsymbol{c}_j \end{aligned} \tag{13}$$

with

$$\boldsymbol{E}_{ij} = \mathrm{diag}([\mathcal{R}(\mathbf{W}_{ij})\mathcal{I}(\mathbf{W}_{ij})] \tag{14}$$

being a diagonal matrix of the frequency domain unmixing coefficients with separated real and imaginary parts. As the resulting filters $\boldsymbol{w}_{ij}$ and the scaling coefficients $\boldsymbol{c}_j$ are real, it is possible to take advantage of the symmetry properties of the DFT. With M being the length of $\boldsymbol{w}_{ij}$ there are only $K = M/2+1$ scaling coefficients. $\bar{\mathcal{F}}$ is obtained from the $(M \times M)$-IDFT matrix $\mathcal{F}$ by concatenation of the real and imaginary parts such that the multiplication with $\boldsymbol{E}_{ij}$ is real again. Finally, $\boldsymbol{B}$ consists of concatenated identity matrices and is responsible

Table 1. Comparison of the signal-to-interference ratios in dB and the distortions measured by the SFM

	Left	Right	Overall	SFM
MDP (1)	2.85	7.37	5.77	0.32
MDP (2)	8.26	8.88	8.75	0.33
New Alg.	9.91	12.10	11.46	0.80

for aligning the scaling coefficients to both the real and imaginary parts of the transformation matrices.

The scaling factors $c_j(\omega)$ have to be calculated for all filters belonging to the same output j simultaneously. This can be achieved by stacking $\boldsymbol{V}_{ij}$ into $\bar{\boldsymbol{V}}_j$ and minimizing

$$||\bar{\mathbf{w}}_j \bar{\boldsymbol{V}}_j - \boldsymbol{c}_j||_{\ell_2} \tag{15}$$

where $\bar{\mathbf{w}}_j$ is the vertical concatenation of some desired filters. For the proposed equalization, these desired filters $\mathbf{w}_{ij}$ are defined to consist of zeros and have a single one at the position where the corresponding $\boldsymbol{w}_{ij}$ have the main peak when calculated using the MDP. The solution is given by $\boldsymbol{c}_j = \bar{\boldsymbol{V}}_j^+ \bar{\mathbf{w}}_j$, with $\bar{\boldsymbol{V}}_j^+$ being the pseudoinverse of $\bar{\boldsymbol{V}}_j$.

5 Simulations

Simulations have been done on real-world recordings of eight seconds of speech sampled at 16 kHz. The length of the unmixing filters was $M = 1024$. As single contributions of signals at the microphones are available, the separation performance can be calculated as in [17]. The coloration done by the unmixing system is measured in the terms of spectral flatness measure (SFM) [8]. With SFM being one a filter is an all-pass and does not color the signals. A value near zero indicates very strong distortions.

The separation was successful and no permutation occurred. The results after applying the normalization after convergence are shown in the first line of Table 1. The separation performance is quite poor and with an average SFM = 0.32 the signals are colorated. Applying the normalization in every step enhances the separation performance, but the coloration is still the same. In the last line of Table 1 the results for the equalized filters are shown. The new algorithm is able to enhance the separation performance even more and the coloration is reduced. With an average SFM = 0.80 the filters have a lot more all-pass characteristic which also can be seen in the Figs. 2 and 3 where the filter set before and after equalization is compared. The main peak has been enhanced, while the other coefficients are scaled down. The energy of these coefficients has been reduced by approximately 10 dB.

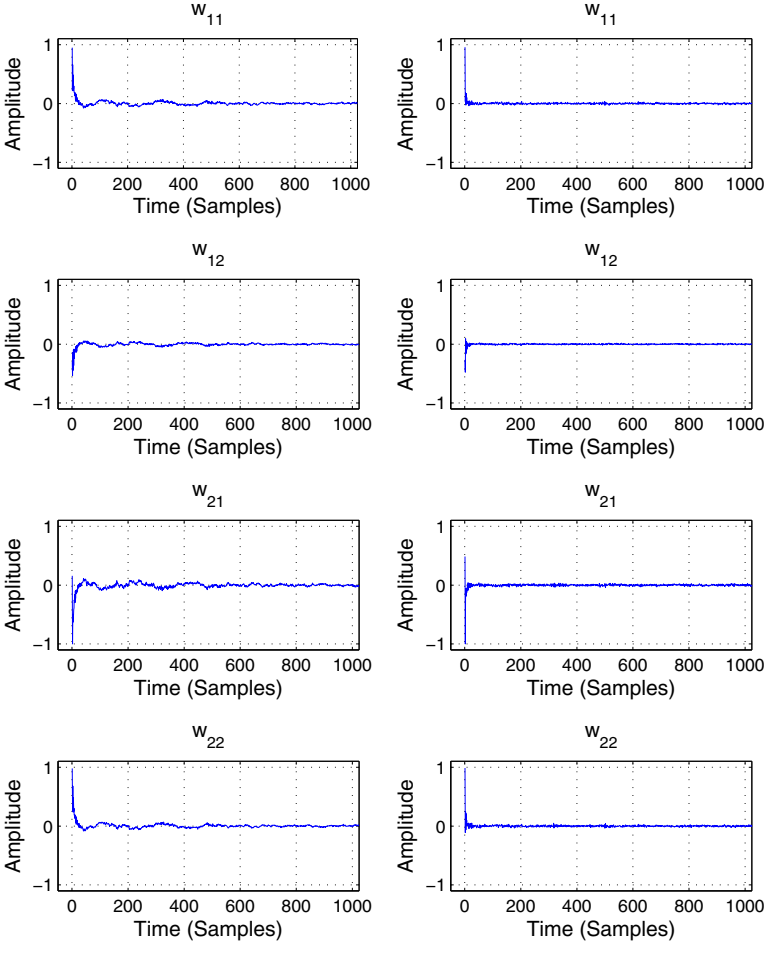

Fig. 2. Comparison of filter sets using the minimal distortion principle (left) and the new method (right)

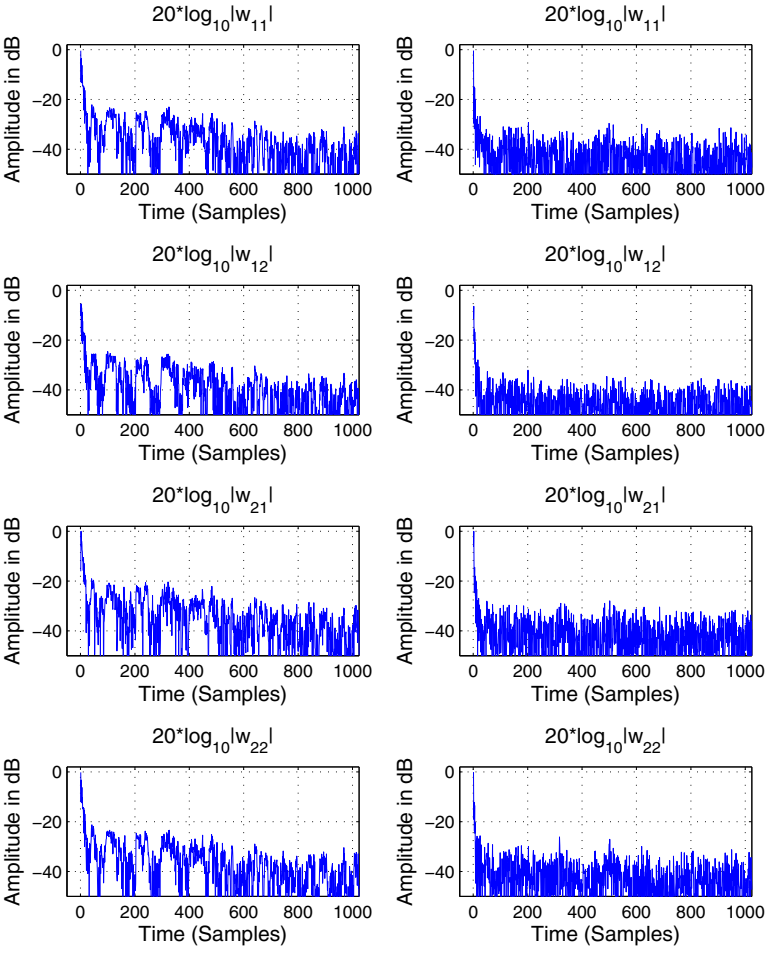

Fig. 3. Magnitudes of filters designed via the minimal distortion principle (left) and the new method (right)

6 Summary

In this paper, we have proposed to use the scaling ambiguity of convolutive blind source separation for equalization of the unmixing filters. We calculated a set of scaling factors that lead to unmixing filters with a more all-pass characteristic. This leads to less coloration of the separated signals and enhanced separation performance. The algorithm has been tested on a real-world example.

References

1. Aichner, R., Buchner, H., Araki, S., Makino, S.: On-line time-domain blind source separation of nonstationary convolved signals. In: Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA 2003), Nara, Japan, pp. 987–992 (April 2003)
2. Amari, S.I., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. In: Advances in Neural Information Processing Systems, vol. 8. MIT Press, Cambridge (1996)
3. Anemüller, J., Kollmeier, B.: Amplitude modulation decorrelation for convolutive blind source separation. In: Proceddings of the second international workshop on independent component analysis and blind signal separation, pp. 215–220 (2000)
4. Cardoso, J.F., Soulomiac, A.: Blind beamforming for non-Gaussian signals. Proc. Inst. Elec. Eng., pt. F. 140(6), 362–370 (1993)
5. Douglas, S.C., Sawada, H., Makino, S.: Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters. IEEE Trans. Speech and Audio Processing 13(1), 92–104 (2005)
6. Hyvärinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neural Computation 9, 1483–1492 (1997)
7. Ikeda, S., Murata, N.: A method of blind separation based on temporal structure of signals. In: Proc. Int. Conf. on Neural Information Processing, pp. 737–742 (1998)
8. Johnston, J.D.: Transform coding of audio signals using perceptual noise criteria. IEEE Journal on Selected Areas in Communication 6(2), 232–314 (1988)
9. Matsuoka, K.: Minimal distortion principle for blind source separation. In: Proceedings of the 41st SICE Annual Conference, August 5-7, vol. 4, pp. 2138–2143 (2002)
10. Mazur, R., Mertins, A.: 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 (2009)
11. Mazur, R., Mertins, A.: A method for filter shaping in convolutive blind source separation. In: Adali, T., Jutten, C., Romano, J.M.T., Barros, A.K. (eds.) ICA 2009. LNCS, vol. 5441, pp. 282–289. Springer, Heidelberg (2009)
12. Mazur, R., Mertins, A.: Using the scaling ambiguity for filter shortening in convolutive blind source separation. In: Proc. IEEE Int. Conf. Acoust, Taipei, Taiwan, pp. 1709–1712 (April 2009)
13. Mei, T., Xi, J., Yin, F., Mertins, A., Chicharo, J.F.: Blind source separation based on time-domain optimizations of a frequency-domain independence criterion. IEEE Trans. Audio, Speech, and Language Processing 14(6), 2075–2085 (2006)
14. Mukai, R., Sawada, H., Araki, S., Makino, S.: Blind source separation of 3-d located many speech signals. In: 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 9–12 (October 2005)

15. Rahbar, K., Reilly, J.P.: A frequency domain method for blind source separation of convolutive audio mixtures. IEEE Trans. Speech and Audio Processing 13(5), 832–844 (2005)
16. Sawada, H., Mukai, R., Araki, S., Makino, S.: 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 (2004)
17. Schobben, D., Torkkola, K., Smaragdis, P.: Evaluation of blind signal separation methods. In: Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation, Aussois, France (January 1999)
18. Smaragdis, P.: Blind separation of convolved mixtures in the frequency domain. Neurocomputing 22(1-3), 21–34 (1998)
19. Wang, W., Chambers, J.A., Sanei, S.: A novel hybrid approach to the permutation problem of frequency domain blind source separation. In: Puntonet, C.G., Prieto, A.G. (eds.) ICA 2004. LNCS, vol. 3195, pp. 532–539. Springer, Heidelberg (2004)