A Clustering Approach for Solving the Spatial Aliasing Problem in Convolutive Blind Source Separation

Radoslaw Mazur, Huy Phan, and Alfred Mertins
Institute for Signal Processing
University of Lübeck
Ratzeburger Allee 160, 23562 Lübeck, Germany
{mazur,phan,mertins}@isip.uni-luebeck.de

***Abstract*—In this paper we propose to extend a recently introduced clustering approach for solving the permutation ambiguity in convolutive blind source separation to a case where spatial aliasing occurs. A well known approach for separation of sources is the transformation to the time-frequency domain, where the task can be reduced to multiple instantaneous problems. While these may be easily solved using independent component analysis, this approach has the drawback of the inherent permutation and scaling ambiguities, which have to be corrected before the transformation to the time domain or otherwise the whole process will fail. Here, we extend an existing clustering approach to cope with the case where spatial aliasing occurs. This is achieved by exploiting the direction information of whole clusters instead of single bins. The performance of the proposed method is evaluated on real-room recordings.**

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

I. Introduction

When dealing with linear and instantaneous mixtures of non-Gaussian signals, blind separation may be performed using the Independent Component Analysis (ICA). For this case, several algorithms have been proposed [1], [2], [3]. The approach is called blind, as typically neither the sources nor the mixing system are known.

Real-world mixtures of acoustic signals such as speech are not separable using this simple approach. With finite speed of sound and multiple reflections in closed rooms, the signals arrive at the microphones multiple times with different delays. This convolutive mixing process is usually modeled using FIR filters. For realistic scenarios, filters with several thousand coefficients are needed. For the separation, a set of unmixing filters with at least the same length is needed.

It is possible to calculate these filters directly in the time domain [4], [5], but these methods usually suffer from high computational load and often poor convergence. Therefore, an often used approach is the transformation to the time-frequency domain, where the convolution becomes a multiplication and instantaneous ICA algorithms can be applied in each frequency bin independently. However, with this approach, the discrete bins usually have different scalings, and they can be arbitrarily permuted. Without correction of the different scalings, a filtered version of the sources will be retrieved. A widely used solution is the minimal distortion principle [6] or inverse postfilters [7]. These methods do not add any additional distortions while accepting the filtering of the mixing system. Other approaches solve the scaling ambiguity with the aim of filter shortening [8] or shaping [9], [10].

Without the correction of the varying permutations in the single frequency bins, different source signals will appear at different frequencies in different permutations and the whole process will fail. There are two main approaches for the solution of this permutation problem. The first group of algorithms relies on the characteristics of the unmixing matrices. By interpretation as beamformer, the direction of arrival (DOA) can be used to design a depermutation criterion [11]. Alternative formulations evaluate directivity patterns [12] or time differences of arrivals (TDOA) [13], [14], [15]. Here, an assumption of specific directions of the sources is exploited, which is only true when the reverberation is low enough. Otherwise only part of the frequencies can be assigned and the remaining ones have to be aligned using other techniques [11]. Time and phase differences exhibit a circularity property, which result in spatial aliasing. Therefore, in order to deal with higher frequencies and larger microphone arrays, it is necessary to resolve this additional ambiguity. For example, in [16] the authors used a circular-linear model and clustering with a sequential variant of Random Sample Consensus for using up to six-fold larger array compared to the no aliasing case.

The second group of algorithms exploits the similarities of the time structure of the separated bins, for example by assuming a high correlation between neighboring bins [7]. This method has been extended in [17], [18] to use activity patterns. In [19] the authors proposed a dyadic sorting scheme by comparing an increasing number of frequency bins in each iteration. The dyadic sorting has also been used in [20] together with a sparsity criterion and in [21] in a combination of non-decimating filter bank and spectral summation. 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 [22] and [23].

In [24] a two stage method has been introduced which employs both of the above mentioned properties. At the first stage, clusters of robust depermuted bins are found. The robustness is achieved by a very conservative criterion of a cluster being non ambiguous by containing only bins which are all positively correlated to each other. In the second stage, by calculating an average TDOA of a cluster, a robust depermutation could be achieved.

While being computationally easy, the method from [24] fails in the case of spatial aliasing. In this work we extend this approach to cope with this ambiguity. The proposed circular model allows for a direct estimation of TDOA in this case, while still having negligible computational cost compared to the ICA stage. The performance of the proposed method will be shown on real world examples.

II. Model and methods

The instantaneous mixing and unmixing processes form the basis for the convolutive case. Both methods will be described in the following.

A. BSS for instantaneous mixtures

The instantaneous mixing process of N sources into N observations is modeled by an $N \times N$ matrix $\mathbf{A}$. With the source vector

$\boldsymbol{s}(n) = [s_1(n), \dots, s_N(n)]^T$ and negligible measurement noise, the observation signals $\boldsymbol{x}(n) = [x_1(n), \dots, x_N(n)]^T$ are given by

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

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

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

with $\boldsymbol{y}(n) = [y_1(n), \dots, y_N(n)]^T$. 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

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

The term $\boldsymbol{g}(\boldsymbol{y}) = (g_1(y_1), \dots g_n(y_n))$ is 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.

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:

$$\boldsymbol{x}(n) = \mathbf{H}(n) * \mathbf{s}(n) = \sum_{l=0}^{L-1} \mathbf{H}(l)\boldsymbol{s}(n-l) \tag{5}$$

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

$$\boldsymbol{y}(n) = \mathbf{W}(n) * \boldsymbol{x}(n) = \sum_{l=0}^{M-1} \mathbf{W}(l)\boldsymbol{x}(n-l) \tag{6}$$

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

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

$$\boldsymbol{Y}(\omega_k, \tau) = \boldsymbol{W}(\omega_k)\boldsymbol{X}(\omega_k, \tau), \qquad k = 0, 1, \dots, K-1 \tag{7}$$

with K being 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:

$$\boldsymbol{Y}(\omega_k, \tau) = \boldsymbol{W}(\omega_k)\boldsymbol{X}(\omega_k, \tau) = \boldsymbol{D}(\omega_k)\mathbf{\Pi}(\omega_k)\boldsymbol{S}(\omega_k, \tau) \tag{8}$$

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. Using the minimal distortion principle [6] to resolve this ambiguity, the unmixing matrix reads

$$\boldsymbol{W}'(\omega) = \mathrm{dg}(\boldsymbol{W}^{-1}(\omega)) \cdot \boldsymbol{W}(\omega) \tag{9}$$

with $\mathrm{dg}(\cdot)$ returning the argument with all off-diagonal elements set to zero.

Without correction of the permutation, different signals will be restored at different frequencies and the whole separation process will fail. In the next section, we will propose a new scheme for calculation of the depermutation.

III. Depermutation Algorithms

In this section, we describe the basic algorithms for depermutation. At first, the basics of the correlation approach with the use of activity patterns will be revised and the robust clustering method from [24] will be shortly summarized. Using this clustering results a new method for explicit estimation for TDOAs in the presence of spatial aliasing will be derived.

A. Correlation approaches

Many depermutation algorithms exploit the statistics of the separated signals. For example, in [7] the criterion is based on the assumption of high correlation of envelopes of neighboring bins. With $\boldsymbol{V}(\omega, \tau) = |\boldsymbol{Y}(\omega, \tau)|$, the correlation between two bins k and l is defined as

$$\rho_{qp}(\omega_k, \omega_l) = \frac{\sum_{\tau=0}^{\mathcal{T}-1} V_q(\omega_k, \tau)V_p(\omega_l, \tau)}{\sqrt{\sum_{\tau=0}^{\mathcal{T}-1} {V_q}^2(\omega_k, \tau)}\sqrt{\sum_{\tau=0}^{\mathcal{T}-1} {V_p}^2(\omega_l, \tau)}} \tag{10}$$

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

$$r_{kl} = \frac{\rho_{pp}(\omega_k, \omega_l) + \rho_{qq}(\omega_k, \omega_l)}{\rho_{pq}(\omega_k, \omega_l) + \rho_{qp}(\omega_k, \omega_l)}. \tag{11}$$

With $r_{kl} > 1$ the bins are assumed to be correctly aligned and otherwise a permutation has occurred. The simple method, where consecutive bins are examined, is not robust, as single wrong permutations lead to whole blocks of falsely permuted bins.

B. Activity patterns

In [17], [18] an alternative method to the correlation of the envelopes has been proposed. Here, the authors exploit the sparsity of speech signals and compute the dominance of the i-th single separated signal as

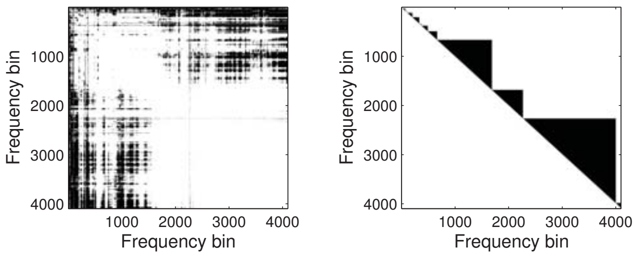
$$\mathrm{powRatio}_i(\omega_k, \tau) = \frac{\|\mathbf{w}_i(\omega_k)y_i(\omega_k, \tau)\|^2}{\sum_{k=1}^{N} \|\mathbf{w}_k(\omega_k)y_k(\omega_k, \tau)\|^2}. \tag{12}$$

The values of these activity patterns are normalized to $[0; 1]$. A value of approximately one indicates a dominance of the given signal, while low values denote the dominance of some other signals. The comparison of activity patterns instead of envelopes by (10) and (11) is usually more robust. However, this assumption is violated when one signal is dominant the whole time. This can be problematic for speech signals, which usually have no energy below the fundamental frequency.

C. Robust clustering

The assumption of correlated envelopes or activity patterns is usually only valid for neighboring bins. In Fig. 1 (a) an example using the dataset from [25] is shown. Here, a white point indicates a correct and a black point ($r_{kl} < 1$) a false decision for a perfectly depermuted case. The upper right corner with a high number of black points indicates a very low similarity of low and high frequencies for the signals.

In [24] a new clustering method has been introduced. Based on the alignment coefficients r_{kl}, clusters of bins are identified which are correctly depermuted. In order to achieve robustness a conservative criterion is proposed. Using a sequential greedy algorithm a bin is added to a cluster only if it has the same correlation to all previous bins. Otherwise, a new cluster is started. In Fig. 1 (b) the result of the clustering procedure is shown. Here, black areas indicate nonambiguous correlation coefficients and mark the boundaries of



(a) (b)

Fig. 1. Visualization of the robust clustering method. a) Alignment coefficients for all bins from (11) using activity patterns (12). White points ($r_{kl} > 1$) indicate a correct and black ($r_{kl} < 1$) a false decision for a perfectly depermuted case. b) The detected clusters with nonambiguous coefficients.

the clusters. In this case, 19 clusters have been identified, which is a substantial reduction from 4097 bins.

D. Average TDOA in the case of no spatial aliasing

For the case of no spatial aliasing, the authors of [13] calculate TDOA for the 2×2-case for the single bins as

$$\mathrm{TDOA}_i(\omega_k) = \frac{1}{2\pi f_k} \arg\left(\frac{[H(\omega_k)]_{1i}}{[H(\omega_k)]_{2i}}\right) \tag{13}$$

with $[H(\omega_k)]_{li}, l \in \{1,2\}$ being the coefficients of the mixing matrix corresponding to the i-th source, f_k the frequency and $\arg(\cdot)$ calculating the phase of a complex number in the range $(-\pi;\pi]$. With this formulation, no information about the microphone distance is needed.

In [13] the TDOAs have been used directly for clustering and with some additional confidence functions a depermutation for almost all bins has been achieved. In [24] the authors calculated an average cluster TDOA:

$$\mathrm{acTDOA}_i(C_m) = \mathrm{mean}(\mathrm{TDOA}_i(\omega_k)), \quad k \in C_m \tag{14}$$

with C_m being a set of indices of the bins of the m-th cluster obtained by the above robust clustering. The averaged cluster TDOAs could be easily arranged for a global alignment by a simple distance measurement

$$\sum_{i=1}^{2}(\mathrm{acTDOA}_j(C_m) - \mathrm{acTDOA}_i(C_M))^2, \quad j \in \{1,2\} \tag{15}$$

with C_M being the largest cluster.

This method is robust due to the fact that averaging the TDOAs removes the need to deal with outliers which are quite often a problem in the lower frequencies.

E. Average TDOA in the case of spatial aliasing

In Fig. 2 the results of the TDOA calculation in the case of spatial aliasing are shown. Here, we can visualize three major areas. The lower frequencies up to around bin 600 show a high variance of the estimated TDOAs. This is the area, where the averaging used to calculate cluster TDOA is very effective at dealing with outliers and giving a good estimate. The second area up to about frequency bin 1600 is quite unproblematic. Even a simple clustering procedure can yield a correct depermutation. The third part above 1600 shows an example of spatial aliasing. Here, the estimated TODAs do not have a consistent value and even overlap with the other channel.

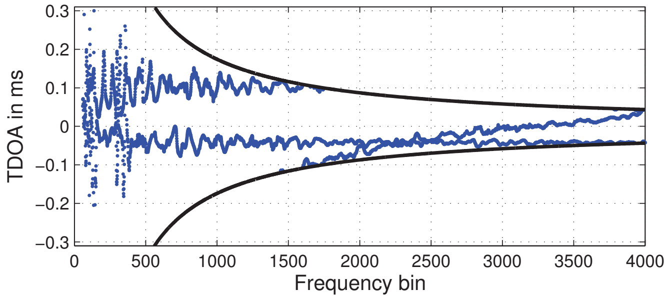


Fig. 2. TDOAs of single frequency bins for a 2×2 case calculated using (13). The additional lines indicate the range of aliasing free TDOA values. The discontinuity at frequency bin 1600 indicates spatial aliasing.

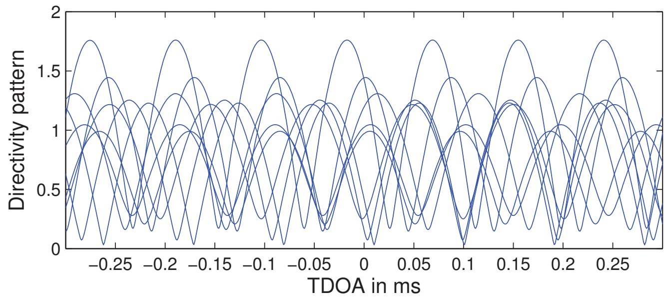


Fig. 3. Directivity plot for some frequency bins between 3000 and 4000. The single patterns are periodic with different periods. The overall picture is very chaotic. The positions of common minima are hard to estimate.

In the following, we propose a method to deal with this case. First we recall, that the estimated TDOA using (13) is the minimum of the central period of a directivity pattern [12]

$$F_i(\omega_k, d) = \left|[H(\omega_k)]_{1i} - [H(\omega_k)]_{2i}\, e^{-j\omega d/c}\right| \tag{16}$$

with c the speed of sound and d the different values for TDOAs.

In Fig. 3 several directivity patterns for frequencies between 3000 and 4000 Hz are shown. The directivity patterns are periodic and the period is varying with frequency. When looking at multiple directivity patterns at once, the situation is quite chaotic. The typical way to estimate the TDOAs is to cluster the minima either in one period [13] or try to find a global minimum for all frequencies at once [16]. This is a computationally demanding approach and needs to deal with noise and outliers.

Here, we propose another approach which is inspired by the directivity patterns and calculate the TDOA by using the information from the robust clustering from Section III-C.

First, we define a cluster directivity pattern $G_i(m,d)$ for the m-th cluster as

$$G_i(C_m, d) = \sum_{k \in C_m} \mathrm{pdist}\left(\frac{[H(\omega_k)]_{1i}}{[H(\omega_k)]_{2i}}, e^{-j\omega d/c}\right)^2 \tag{17}$$

with

$$\mathrm{pdist}(a,b) = \mathrm{mod}(\angle(a) - \angle(b) + \pi, 2\pi) - \pi \tag{18}$$

being the distance of the phases of two complex numbers with regard to the 2π periodicity, $\angle(\cdot)$ the phase in the range $(-\pi;\pi]$, and $\mathrm{mod}(a,b)$ the remainder of the division a/b.

In (17) the ideal phases of a time delay d are compared to the actual phases calculated using the TDOAs, while correctly considering the

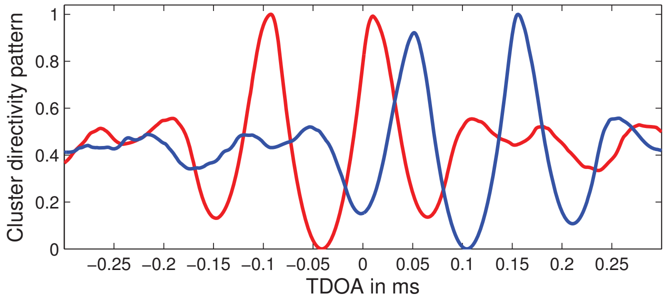


Fig. 4. The cluster directivity pattern for the largest cluster from Fig. 1 normalized to the range $[0, 1]$. The global minima of both functions are correctly indicating the time delay of both channels.

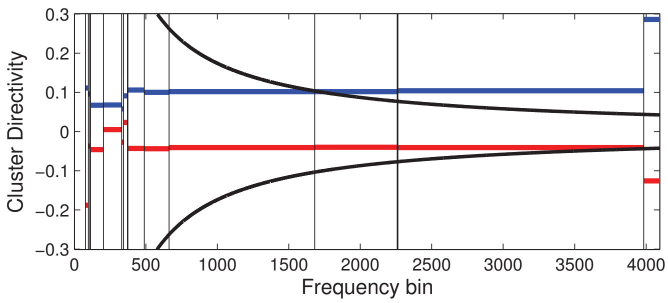


Fig. 5. The final clustering result. Due to the averaging effect of clustering, lower frequencies are easier to depermute. The spatial aliasing has been correctly resolved for the higher frequencies.

2π periodicity. Therefore, it is feasible to distinguish between the different minima in case of spatial aliasing, which is not possible using single frequency bins as in (13) or (16).

In Fig. 4 the normalized values of $G_i(C_m, d)$ for the biggest cluster (bins 2263 to 3983) from the previous example are shown. The global minima of both functions for both channels clearly indicate to correct TDOAs and are able to resolve the problem of spatial aliasing. The other local minima show the periodic repetitions, but compared to Fig. 3, where they have the same value as the global ones, they do not pose a problem.

The cluster TDOAs can be estimated by finding the position of the minimum by

$$\text{acTDOA}_i(C_m) = \arg\min_d G_i(C_m, d) \tag{19}$$

Due to the periodicities in $\text{pdist}(a, b)$ this minimum cannot be found directly. Still, being a one-dimensional problem the minimum may be found, for example, by dense sampling. Due to the small number of clusters, the computational cost is negligible compared to the ICA stage.

Finally, in the last stage of the algorithm the clusters are arranged as the biggest one using (15) as in [24]. In Fig. 5 the result is shown. Here, the spatial aliasing has been resolved correctly. Additionally, the lower frequencies are much easier to depermute due to the averaging effect of clustering.

IV. Simulations

The experiments using the proposed algorithm have been performed using real-world data available at [25]. The setup was chosen to be similar to that in [24] and [11]. With a sampling rate of 8 kHz, the FFT length was chosen to be 8192, and a 2048 point Hann analysis window has been used. For the ICA stage 400 iterations of (4) in each frequency bin have been performed.

The dataset contains four recordings of four different speech signals in a low reverberant room. As the signals do not have meaningful energy below 110 Hz, only bins above this frequency are taken into consideration. In Table I the results are shown.

In dataset 1 microphones one and two have been used. With a distance of 4cm there was no spatial aliasing. The depermutaion has been successful and the results are very close to the algorithm from [24]. In dataset 2 microphones one and three have been used. Here the distance is 8cm and there is spatial aliasing above 1700Hz. This is the dataset used to visualize the algorithm in Fig. 1 to 5. Again the overall procedure has been successful. Dataset 3 contains recordings from microphone one and four with a distance of 12cm. Again the performance of is very good. The last dataset is an average

TABLE I
Comparison of the results for different depermutation algorithms in terms of separation performance (SIR) in dB.

Algorithm	Set 1	Set 2	Set 3	Set 4
Proposed	17.3	18.0	17.2	20.1
Sparsity [20]	15.4	15.9	15.0	13.8
Dyadic sorting [19]	2.7	3.5	2.8	4.0
Non blind	17.6	18.8	17.9	21.8

calculated for all combinations of the four available signals and the three available distances of microphones. In comparison to the other used algorithms from [19] and [20], the proposed one is performing significantly better.

V. Conclusions

In this paper we proposed a new approach for solving the permutation ambiguity in convolutive blind source separation for the case where spatial aliasing occurs. The new method is using a previously introduced robust clustering method on single frequency bins. Using this information a new solution for calculating the time differences of arrival for a cluster is proposed. The new method is able to resolve the ambiguities of spatial aliasing. The performance of the proposed method is evaluated on real-room recordings.

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] 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.

[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] 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 (ICASSP)*, Taipei, Taiwan, April 2009, pp. 1709–1712.

[9] ——, "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.

[10] ——, "A method for filter equalization in convolutive blind source separation," in *Proc. 9th Int. Conf. on Latent Variable Analysis and Signal Separation*, St. Malo, France, Sept. 2010.

[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] 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.

[13] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Grouping separated frequency components with estimating propagation model parameters in frequency-domain blind source separation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. vol.15, no.5, pp. 1592–1604, July 2007.

[14] F. Nesta and M. Omologo, "Approximated kernel density estimation for multiple tdoa detection," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 149 –152.

[15] ——, "Generalized state coherence transform for multidimensional tdoa estimation of multiple sources," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 246 –260, Jan 2012.

[16] J. Traa and P. Smaragdis, "Blind multi-channel source separation by circular-linear statistical modeling of phase differences," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, May 2013, pp. 4320–4324.

[17] H. Sawada, S. Araki, and S. Makino, "Measuring dependence of binwise separated signals for permutation alignment in frequency-domain bss," in *IEEE International Symposium on Circuits and Systems (ISCAS 2007)*, May 2007, pp. 3247 –3250.

[18] ——, "Mlsp 2007 data analysis competition: Frequency-domain blind source separation for convolutive mixtures of speech/audio signals," in *IEEE Workshop on Machine Learning for Signal Processing*, aug. 2007, pp. 45 –50.

[19] 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.

[20] R. Mazur and A. Mertins, "A sparsity based criterion for solving the permutation ambiguity in convolutive blind source separation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, Prague, Czech Republic, May 2011, pp. 1996–1999.

[21] R. Mazur, J. O. Jungmann, and A. Mertinss, "Optimized dyadic sorting for solving the permutation ambiguity in acoustic blind source separation," in *Proc. European Signal Processing Conference*, Marrakech, Maroc, Sept 2013.

[22] 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.

[23] ——, "Simplified formulation of a depermutation criterion in convolutive blind source separation," in *Proc. European Signal Processing Conference*, Glasgow, Scotland, Aug 2009, pp. 1467–1470.

[24] R. Mazur, J. O. Jungmann, and A. Mertins, "A new clustering approach for solving the permutation problem in convolutive blind source separation," in *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, USA, Oct. 2013.

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