

# Time-Domain Blind Separation of Audio Sources on the Basis of a Complete ICA Decomposition of an Observation Space

Zbyněk Koldovský<sup>1,2</sup> and Petr Tichavský<sup>2</sup>

<sup>1</sup>Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic. E-mail: zbynek.koldovsky@tul.cz, fax:+420-485-353112, tel:+420-485-353534

<sup>2</sup>Institute of Information Theory and Automation, Pod vodárenskou věží 4, P.O. Box 18, 182 08 Praha 8, Czech Republic. E-mail: tichavsk@utia.cas.cz, fax:+420-2-868-90300, tel. +420-2-66052292

**Abstract**—Time-domain algorithms for blind separation of audio sources can be classified as being based either on a partial or complete decomposition of an observation space. The decomposition, especially the complete one, is mostly done under a constraint to reduce the computational burden. However, this constraint potentially restricts the performance. The authors propose a novel time-domain algorithm that is based on a complete unconstrained decomposition of the observation space. The observation space may be defined in a general way, which allows application of long separating filters, although its dimension is low. The decomposition is done by an appropriate independent component analysis (ICA) algorithm giving independent components that are grouped into clusters corresponding to the original sources. Components of the clusters are combined by a reconstruction procedure after estimating microphone responses of the original sources. The authors demonstrate by experiments that the method works effectively with short data, compared to other methods.

## I. INTRODUCTION

Blind separation of simultaneously active audio sources is a popular task of audio signal processing motivated by many emerging applications, such as hands-free and distant-talking speech communication, human/machine interactions, and so on. The goal is to retrieve  $d$  audio sources from their convolutive mixtures recorded by  $m$  microphones, which is described by

$$x_i(n) = \sum_{j=1}^d \sum_{\tau=0}^{M_{ij}-1} h_{ij}(\tau) s_j(n-\tau), \quad i = 1, \dots, m, \quad (1)$$

where  $x_1(n), \dots, x_m(n)$  are the observed signals on microphones and  $s_1(n), \dots, s_d(n)$  are the original (audio) signals unknown in the “blind” scenario.

In fact, the mixing system is a multi-input multi-output (MIMO) linear filter with source-microphone impulse responses  $h_{ij}$ , each of length  $M_{ij}$ . The responses characterize

<sup>0</sup>Parts of this work were presented at HSCMA 2008 [21]. The work was supported by Grant Agency of the Czech Republic through the projects 102/07/P384 and 102/09/1278 and by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572.

propagation of sound in the recording room and are also unknown. It is assumed that the system is time-invariant, which usually means that positions of the sources and the microphones do not change within recording of  $N$  samples.

The separation through a linear processing consists in seeking a MIMO filter that inverts the mixing process (1). Any estimate of the  $j$ th original signal  $s_j(n)$ ,  $j = 1, \dots, d$ , thus has the form

$$\hat{s}_j(n) = \sum_{i=1}^m \sum_{\tau=0}^{L-1} w_{ji}(\tau) x_i(n-\tau), \quad (2)$$

where  $L$  is the length of the separating filter. The blind separation that is based on the assumption of statistical independence of the original signals is addressed here. The separating filters will therefore be estimated via Independent Component Analysis (ICA) [1], [2].

Indeterminacies that are inherent to the ICA cause that each original signal is estimated up to an unknown filtering [3], [4]. Without any prior knowledge, that is not available in the blind scenario, an arbitrarily filtered source signal can also be considered a source signal. It is therefore meaningful to aim at estimating responses of sources at microphones, which only have properly defined colorations. Following from (1), the microphone response of the  $k$ th source at the  $i$ th microphone is

$$s_k^i(n) = \sum_{\tau=0}^{M_{ik}-1} h_{ik}(\tau) s_k(n-\tau). \quad (3)$$

Consequently, each source is estimated  $m$  times (all its responses are estimated). Once the responses  $s_k^i(n)$ ,  $i = 1, \dots, m$ , are estimated, it might be desirable to combine them in one-channel estimate of the  $k$ th signal denoted by  $\hat{s}_k(\cdot)$ .

Basically, the blind audio source separation can be performed either in the frequency-domain or in the time-domain (TD). In the frequency-domain approach [5], [6], [7], the signals are transformed by the Discrete Fourier Transform (DFT), and the convolution operation in (1) changes to the

ordinary multiplication<sup>1</sup>. This translates the convolutive model into a set of complex-valued instantaneous mixtures, one for each frequency, that can be separated by complex-domain ICA methods. The frequency-domain approach allows effective computation of long separating filters, which is favorable in audio applications. By contrast, the computation of long filters requires long recordings to generate sufficient amount of data for each frequency [8].

Time-domain approaches transform the convolutive model into an instantaneous one by constructing data vectors or matrices of a special structure, by which the convolution is translated into the vector/matrix product. The data structures, constructed from the only available signals from microphones, define the *observation space*. Most often a matrix is defined so that its rows contain the time-lagged copies of signals from microphones, and the observation space is spanned by these rows. In general, TD methods aim at finding subspaces of the observation space that correspond to separated signals [9].

Decomposition of the observation space can be either complete or partial [10]. In the former case, the original signals are represented by  $d$  independent subspaces spanning the whole observation space. In the latter case, the signals are estimated as one-dimensional subspaces (components) of the observation space. A reconstruction procedure must follow the decomposition to retrieve the microphone responses of separated signals.

Performance of methods doing the partial decomposition depends very much on initialization of a convergence scheme [11], [12], [13]. It might also happen that the method finds two components of the same source and skips another source. In this respect, the complete decomposition is more reliable, however, at higher computational demand.

To alleviate these problems, the decomposition may be done with some constraint. The complete decomposition is usually constrained by an assumption that the inverse of the decomposing transform (matrix) has a special structure, for example block-Toeplitz or block-Sylvester; see articles of Kellermann et al. and Belouchrani et al., e.g. [9], [14], [15]. Févotte et al. proposed a two-stage separation procedure in [10] doing the complete decomposition by an algorithm for the independent subspace analysis (ISA) through joint block diagonalization (JBD) [9] utilizing the orthogonal constraint [17]. The algorithm of Douglas et al. [18] is an example of a constrained partial decomposition. It uses a para-unitary filter constraint and is compared in experiments in this article.

A potential drawback of the constrained decomposition is that it assumes all independent subspaces to have the same dimension. The constraint might also cause some restrictions due to the finite length of data or the limited length of separating filters. In this respect, the complete *unconstrained* decomposition provides an effective way to utilize the available data as effectively as possible, but it was considered to be computationally too extensive [10]. For instance, the JBD algorithm applied in [10] appeared to fail with  $L > 6$ , which, in other words, means that this algorithm cannot work on

observation spaces of higher dimension. It is known that the stability and speed issues in high-dimensional spaces are the shortcomings of many ICA/ISA algorithms.

In this article, a novel method based on the complete unconstrained decomposition of the observation space is proposed. It utilizes modern ICA methods that allow fast, accurate and reliable separation of high-dimensional spaces. Especially, very fast ICA algorithms that are based on approximate joint diagonalization (AJD) by Tichavský and Yeredor [20] are used. Next, the method involves an effective reconstruction step, which yields effective results even when separating filters are much shorter than the mixing filter. Moreover, a general construction of the observation space is proposed, which allows the method to apply long (even infinite) separating filters while preserving its computational complexity (dimension of the observation space). In real-world experiments, the proposed method yields very good results in comparison with its competitors. It has several attractive features such as the ability to estimate the number of sources  $d$ , and it provides room for further development of its variants in future, such as a sub-band version or an on-line version.

The article is organized as follows. The following Section II provides a comprehensive description of a basic version of the proposed method, first introduced in [21], where classical time-lag construction of the observation space is used. The method is a five-step procedure, where each step can be solved in many alternative ways. A few basic variants are proposed. This also includes a novel *oracle* algorithm [8] that utilizes known responses of the sources and provides a reference solution that depends on the quality of the ICA decomposition only. In Section III, an extension of the method that comes from a generalized definition of the observation space is proposed. A special case of the definition leads to the application of infinite impulse response (IIR) Laguerre separating filters. In Section IV, results of various real-world experiments that demonstrate excellent performance of the proposed method in comparison with other existing methods are presented.

## II. BASIC VERSION OF THE PROPOSED METHOD

In the following subsection, a brief description of main steps of the basic variant of the proposed method is given, and in the other subsections each step is further commented and illustrated by an example.

### A. Outline

Assume that  $N$  samples of simultaneously recorded signals from microphones  $x_1(n), \dots, x_m(n)$ ,  $n = 1, \dots, N$ , are available. The method proceeds in five consecutive steps.

- 1) Form a  $M \times (N_2 - N_1 + 1)$  data matrix  $\mathbf{X}$ , whose rows contain time-lagged copies of the signals from microphones. Each signal is delayed  $L$  times, thus,  $L$  rows correspond to each signal, and  $M = mL$ . The

<sup>1</sup>More precisely, the circular convolution changes to the ordinary multiplication.

matrix  $\mathbf{X}$  is given by

$$\mathbf{X} = \begin{bmatrix} x_1(N_1) & \dots & \dots & x_1(N_2) \\ x_1(N_1 - 1) & \dots & \dots & x_1(N_2 - 1) \\ \vdots & \vdots & \vdots & \vdots \\ x_1(N_1 - L + 1) & \dots & \dots & x_1(N_2 - L + 1) \\ x_2(N_1) & \dots & \dots & x_2(N_2) \\ x_2(N_1 - 1) & \dots & \dots & x_2(N_2 - 1) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_m(N_1 - L + 1) & \dots & \dots & x_m(N_2 - L + 1) \end{bmatrix}, \quad (4)$$

where  $N_1$  and  $N_2$ ,  $1 \leq N_1 < N_2 \leq N$ , determine a segment of recordings that is used for computations. The subspace of  $\mathbb{R}^{N_2 - N_1 + 1}$  spanned by rows of  $\mathbf{X}$  will be called the *observation space*.

- 2) Apply an ICA method to the mixture given by  $\mathbf{X}$  to obtain all independent components of  $\mathbf{X}$ . As there may be up to  $M$  (independent) components, the output is a  $M \times M$  de-mixing (decomposing) matrix  $\mathbf{W}$ , and the components are given by  $\mathbf{C} = \mathbf{W}\mathbf{X}$ . The rows of  $\mathbf{C}$  will be denoted  $\mathbf{c}_1^T, \dots, \mathbf{c}_M^T$  and the components (the signals) defined by them will be denoted by  $c_1(n), \dots, c_M(n)$ .
- 3) Group the components  $c_1(n), \dots, c_M(n)$  into  $d_{est}$  clusters, so that each cluster contains components that correspond to the same original source. The number  $d_{est}$  is either estimated or equal to an a priori known (if available) number of sources  $d$ . The grouping is done subject to a similarity measure between the components.
- 4) For each cluster and each component, a weight that characterizes a measure of confidence of the component to belong to the cluster is computed. Then for each cluster, a reconstructed version of the matrix  $\mathbf{X}$  is computed using weighted components subject to the cluster, and rows of the reconstructed matrix are used for estimation of microphone responses of a source corresponding to the cluster. Mathematically, the reconstructed matrix, for the  $k$ th cluster,  $k = 1, \dots, d_{est}$ , is

$$\begin{aligned} \widehat{\mathbf{S}}_k &= \mathbf{W}^{-1} \text{diag}[\lambda_1^k, \dots, \lambda_M^k] \mathbf{C} \\ &= \mathbf{W}^{-1} \text{diag}[\lambda_1^k, \dots, \lambda_M^k] \mathbf{W} \mathbf{X}, \end{aligned} \quad (5)$$

where  $\lambda_1^k, \dots, \lambda_M^k$  denote the weights, each one from  $[0, 1]$ , reflecting degrees of affiliation of components to the  $k$ th cluster. Their particular selection will be described later in this section. Finally, microphone responses (3) of an original source corresponding to the  $k$ th cluster are estimated as

$$\widehat{s}_k^i(n) = \frac{1}{L} \sum_{\ell=1}^L \psi_{k, (i-1)L + \ell}(n + \ell - 1), \quad i = 1, \dots, m, \quad (6)$$

where  $\psi_{k,p}(n)$  is the  $(p, n)$ th element of  $\widehat{\mathbf{S}}_k$ . Obviously  $\psi_{k,p}(n)$ ,  $p = (i-1)L + \ell$ , provides an estimate of  $s_k^i(n - \ell + 1)$ .

- 5) Apply a beamformer to the estimated responses of each source to get the one-channel estimate of the source.

In the following subsections, the steps of this method are discussed in more details. To make the presentation clearer,

an accompanying example is given with three original sources that were artificially mixed into three signals. The mixing system consists of filters of the length  $M_{ij} = 4$ ,  $i, j = 1, \dots, 3$ , whose coefficients were randomly generated according to Gaussian law with zero mean and unit variance. The original and the mixed signals are, respectively, shown in Figs. 1 and 2.

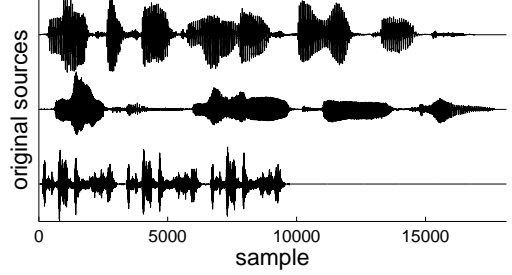


Fig. 1. Original sources considered in the demonstration example. The signals are, respectively, a man's speech, a woman's speech, and a typewriter sound, recorded at the sampling frequency 8kHz.

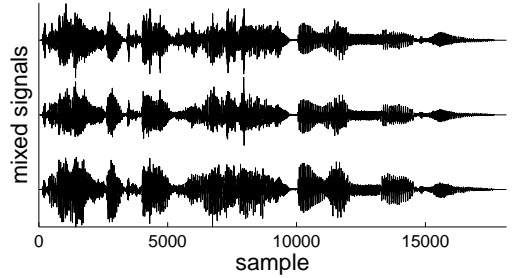


Fig. 2. Three artificial convolutive mixtures of the sources from Fig. 1 simulating signals obtained by three microphones.

## B. Step 1: Construction of $\mathbf{X}$

As mentioned in the introduction, constructing  $\mathbf{X}$  according to (4) allows to convey the separating convolution operation via multiplying  $\mathbf{X}$  by a de-mixing matrix.  $\mathbf{X}$  is usually interpreted as an instantaneous mixture,  $\mathbf{X} = \mathbf{A}\mathbf{S}$ , where  $\mathbf{S}$  is a matrix constructed of delayed original signals analogously to  $\mathbf{X}$ , and  $\mathbf{A}$  is a mixing matrix that has the block-Sylvester structure. However, such mixture is equivalent with (1) in full if only  $\mathbf{A}$  has more columns than rows,  $m > d$  and  $L$  is sufficiently large; see [9], [10].

In this article, none of the above conditions is assumed. The mixing or, equivalently, the de-mixing matrix is considered to be square without any special structure. The structure of  $\mathbf{S}$  is not specified either. It is only assumed that its rows consist of the filtered versions of original signals and form independent subspaces. Consequently,  $\mathbf{S}$  can be estimated, up to indeterminacies, as independent subspaces or components of  $\mathbf{X}$  via ISA or ICA. This approach proves to be more flexible, among others, because  $\mathbf{X}$  may be defined in different ways than (4) as proposed in Section III.

In the accompanying example, consider  $L = 4$ ,  $N_1 = L$  and  $N_2 = 8000 + L - 1$ . This means that the length of

separating filters is 4, and the matrix  $\mathbf{X}$  is  $12 \times 8000$ . Note that all computations made with  $\mathbf{X}$  use data contained in first 8000 samples (the first second) of recordings only. Once the separating MIMO filter is found, it can be applied to the entire data set.

### C. Step 2: ICA Decomposition

At the heart of the proposed separation procedure is a suitable ICA algorithm to be applied to  $\mathbf{X}$ . Because no constraint is applied to the de-mixing matrix, many of the known ICA and ISA algorithms can be considered including those based on Non-Gaussianity, nonstationarity or spectral diversity (distinct coloration) of signals; a survey of ICA algorithms is provided, for example, by ICALAB [22].

The problem of the selection of ICA/ISA algorithm for this purpose exceeds the scope of this paper. The study in [23] showed that ISA algorithms do not have any obvious advantage over ICA algorithms that are followed by clustering. Potentially, ICA methods are computationally inefficient since they not only separate independent subspaces, but also signals within the subspaces. However, the ICA methods considered here are computationally still much faster than up-to-date ISA methods.

Owing to the need to separate mixtures whose dimension is frequently 40 or more, two algorithms are considered: the Non-Gaussianity based EFICA algorithm from [24] and the nonstationarity based algorithm from [20] called BGSEP.

EFICA is an improved version of the well-known FastICA algorithm [25]. BGSEP consists in a special approximate joint diagonalization of a set of covariance matrices of signals in data matrix divided in blocks. Both methods achieve asymptotical optimality within respective models of signals and perform very well in [23].

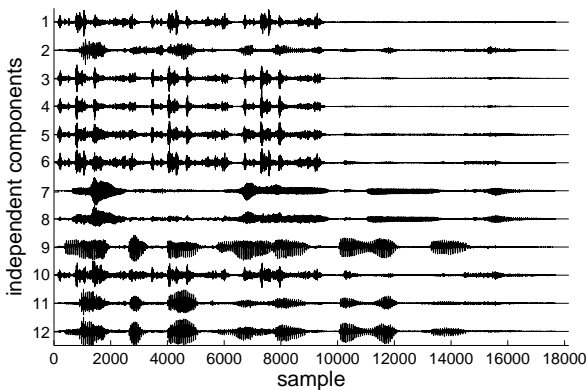


Fig. 3. Independent components obtained by the BGSEP algorithm in the demonstration example. As can be seen by comparing signals from Fig. 1, some components clearly correspond to separated signals.

### D. Step 3: Clustering of Components

An independent component obtained by the ICA algorithm equals, in an ideal case, to a filtered copy of an original source. As the number of components is higher than the number of sources, i.e.  $M > d$ , there should be  $d$  clusters of components

where each cluster contains components of one source. The utilization of the ICA algorithm in the second step should be therefore followed by the clustering of components.

As already discussed above, the alternative way is to apply an ISA method instead of ICA, which does not need the clustering step [19], [26], [27]. However, the “ICA+clustering” approach used here has the following advantages.

- ICA methods work reliably without knowing or estimating the number of components of clusters.
- The approach is flexible because various criteria of similarity of components and clustering methods can be used.

1) *Similarity of components*: If the  $i$ th and the  $j$ th component belong to the same source and contain no interference, it holds that there exists a filter  $f$  such that

$$c_i(n) = \sum_{\tau=-\infty}^{+\infty} f(\tau)c_j(n-\tau) = \{f \star c_j\}(n). \quad (7)$$

In practice, (7) holds approximately only, and  $f$  can be searched by minimizing the mean square distance between the two sides of (7). Therefore, the value of

$$\min_f \hat{E}[c_i(n) - \{f \star c_j\}(n)]^2, \quad (8)$$

where  $\hat{E}$  denotes the sample mean operator, reveals whether the two components belong to the same source. In practice, the minimization in (8) proceeds over filters of length  $2L$ .

Therefore, the similarity of the  $i$ th and the  $j$ th component,  $i \neq j$ , is defined as the  $ij$ th element of matrix  $\mathbf{D}$ , where

$$D_{ij} = \hat{E}[\mathbf{P}_i c_j]^2 + \hat{E}[\mathbf{P}_j c_i]^2, \quad (9)$$

where  $\mathbf{P}_i$  denotes a projector on a subspace spanned by delayed copies of the  $i$ th component, that is, by signals  $c_i(n-L+1), \dots, c_i(n+L-1)$ . Diagonal elements of  $\mathbf{D}$  have no significance here and are set to zero. The computation of (9) can be done efficiently using the FFT and Levinson-Durbin algorithm; see [21]. An example of the similarity matrix  $\mathbf{D}$  is shown in Fig. 4.

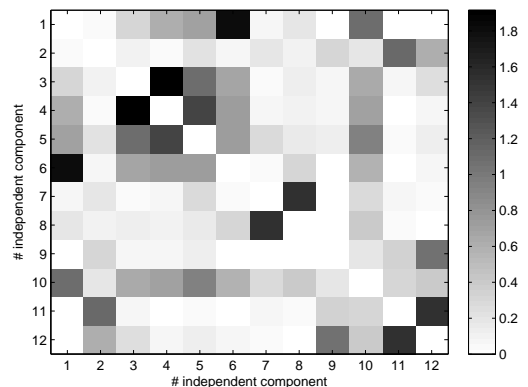


Fig. 4. Similarity matrix  $\mathbf{D}$  between components from Fig. 3 computed according to the definition (9).

2) *Clustering*: The task now is to cluster the  $M$  components subject to the similarity matrix  $\mathbf{D}$ . This general task can be solved by various methods. In this study, attention is restricted to the agglomerative hierarchical clustering algorithm that appears to perform well in this application.

The algorithm consists of  $M$  levels, each giving a partitioning of components. In the beginning (the first level), each component forms a cluster, called singleton, thus, there are  $M$  clusters. At each subsequent level, the method merges two clusters whose similarity is maximal. The number of clusters is thus always equal to the level. In the last level, all components form one cluster.

Finally, the most satisfactory level (partitioning) should be chosen. If the number of sources  $d$  is known in advance, the level giving the desired number of clusters is selected. Otherwise, it is possible to select the level according to a criterion such as

$$\max_p \left\{ \frac{1}{M-p+1} \sum_{k=1}^{M-p+1} \frac{M - |K_k^p| \sum_{i \in K_k^p, j \in K_k^p} \mathbf{D}_{ij}}{|K_k^p| \sum_{i \in K_k^p, j \notin K_k^p} \mathbf{D}_{ij}} \right\}, \quad (10)$$

where  $p$  is the level index within  $M - m + 1, \dots, M - 1$  (i.e., the maximum number of estimated sources corresponds to the number of microphones  $m$ ),  $K_k^p$  is a set of indices of components in the  $k$ th cluster of the  $p$ th partitioning level, and  $|K_k^p|$  is the number of those indices. The argument of sum in (10) evaluates the ratio between the average intra-similarity of components of the  $k$ th cluster to the average inter-similarity of components from the other clusters. The criterion thus reflects the quality of the  $p$ th partitioning as it averages the argument over all of its clusters.

Maximization of (10) can be interpreted as a method of estimating the number of sources. However, since the results are not always satisfactory in practice, there is room for further improvement. In this paper we assume, for simplicity, that the number of active sources is known a priori.

What is left is to define the similarity between clusters, called the *linkage strategy*. A modified average linking strategy is to be used, which is defined as follows. Let  $Q$  and  $R$  contain indices of components of two different clusters. The similarity of the clusters is given by

$$d(Q, R) = \frac{1}{\min(|Q|, |R|)} \frac{1}{|Q|} \frac{1}{|R|} \sum_{q \in Q} \sum_{r \in R} \mathbf{D}_{qr}, \quad (11)$$

where  $|Q|$  is the number of indices in  $Q$ . The modification of the average linkage strategy consists in the division by  $\min(|Q|, |R|)$ . It penalizes mutual similarity of “large” clusters and highlights the similarity of “small” clusters with “large” ones, which is preferable to this application. Pseudocode 1 summarizes the clustering algorithm.

The clustering algorithm was applied to the components from Fig. 3. Three clusters shown by Figs. 5(a)-5(c) were found; reordered similarity matrix  $\mathbf{D}$  according to the clustering is shown by Fig. 6. This example demonstrates clearly that each source may consist of different number of components. In other words, independent subspaces corresponding to the original sources may have different dimensions.

---

#### Pseudocode 1 Hierarchical clustering of components

---

```

 $K_i^1 = \{i\}, i = 1, \dots, M$ 
 $\mathcal{K}^1 = \{K_1^1, \dots, K_M^1\}$ 
for  $p = 1$  to  $M - 1$  do
     $k, \ell = \arg \min_{k, \ell=1, \dots, M-p+1} d(K_k^p, K_\ell^p)$ 
     $\mathcal{K}^{p+1} = \{K_i^p, i \neq k, \ell\} \cup \{K_k^p \cup K_\ell^p\}$ 
end for
if  $d$  is known then
     $p = M - d + 1$ 
else
    Select  $p$  from  $M - m + 1$  to  $M - 1$  according to (10)
end if
return  $\mathcal{K}^p = \{K_1^p, \dots, K_{M-p+1}^p\}$ 

```

---

It can also be seen that some components often exhibit certain closeness to more than one cluster. This is because of the residual interference between components caused by various practical limitations such as the finite length of separating filters. The method takes this important phenomenon into account in the reconstruction step discussed in the following subsection.

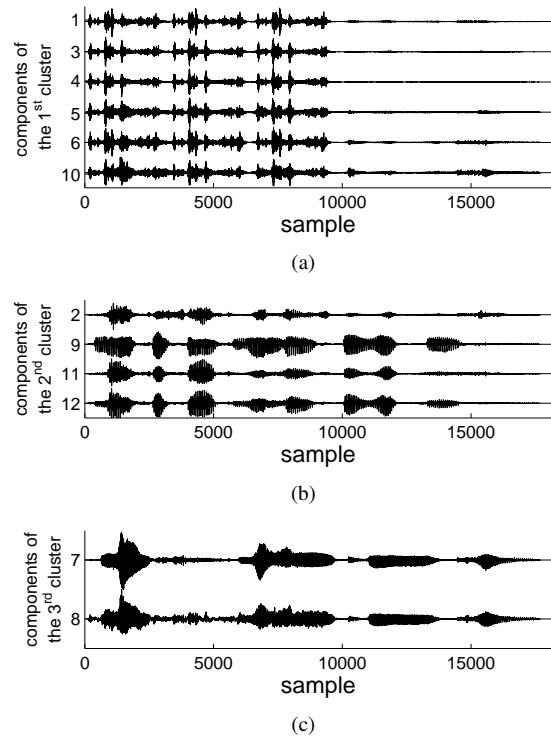


Fig. 5. Components assigned to the three founded clusters.

#### E. Step 4: Reconstruction

The goal of the reconstruction step is to obtain the responses of sources on microphones defined by (3). The response is a signal observed by the microphone if the source is sounding solo. Since all sources sound simultaneously, it holds that

$$x_i(n) = s_1^i(n) + \dots + s_d^i(n), \quad i = 1, \dots, m. \quad (12)$$

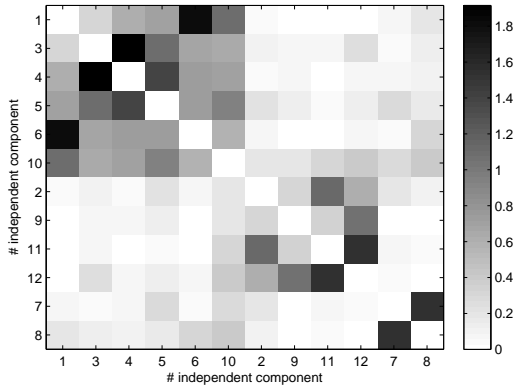


Fig. 6. Reordered similarity matrix  $\mathbf{D}$  from Fig. 4 after the hierarchical agglomerative clustering. Three detected clusters can be markedly seen from the picture.

Hence,  $\mathbf{X}$  can be written as a sum of matrices  $\mathbf{S}_1, \dots, \mathbf{S}_d$ , where  $\mathbf{S}_k$  is constructed in the same way as  $\mathbf{X}$  but using responses of the  $k$ th source only,

$$\mathbf{X} = \mathbf{S}_1 + \dots + \mathbf{S}_d. \quad (13)$$

First, the binary weighting that reflects the results of the clustering of components is introduced by setting the weights introduced in (5) to

$$\lambda_\ell^k = \begin{cases} 1 & \ell \in K_k \\ 0 & \text{otherwise} \end{cases}, \quad (14)$$

where  $K_k$  contains indices of components assigned to the  $k$ th cluster.

If there is no interference between components, and the clustering proceeds without errors, then  $\widehat{\mathbf{S}}_k$  obtained by (5) satisfies  $\widehat{\mathbf{S}}_k = \mathbf{S}_k$ ,  $k = 1, \dots, d_{est}$ , and rows of  $\widehat{\mathbf{S}}_k$  contain delayed microphone responses of the  $k$ th source  $s_k^1(n), \dots, s_k^m(n)$ . Equation (6) means that the  $p$ th response is estimated as an average of  $(p-1)L + 1, \dots, pL$  rows of  $\widehat{\mathbf{S}}_k$  with restored delays.

It is worth noting here that the length of filters found by ICA producing independent components is  $L$ . The reconstruction formula (6) can be interpreted as a filtering by another FIR filter of the length  $L$ . Therefore the final separating filter has the length up to  $2L - 1$ .

#### F. Computation of Weights

A natural extension of the “hard” weighting given by (14) is to consider  $\lambda_\ell^k$  as positive numbers from  $[0, 1]$  selected according to an appropriate rule. The rule introduced in [21] is used, which is given by

$$\lambda_\ell^k \leftarrow \left( \frac{\sum_{j \in K_k, j \neq \ell} \mathbf{D}_{\ell j}}{\sum_{j \notin K_k, j \neq \ell} \mathbf{D}_{\ell j}} \right)^\alpha \quad (15)$$

$$\lambda_\ell^k \leftarrow \lambda_\ell^k / \left( \max_{\ell=1, \dots, M} \lambda_\ell^k \right),$$

where  $\alpha$  is an adjustable positive parameter. The denominator in (15) reflects the similarity of the  $\ell$ th component to components from different clusters than the  $k$ th one. If the component

clearly belongs to the  $k$ th cluster, the denominator is close to zero, and the value of (15) becomes large.

If  $\alpha \rightarrow +\infty$ , the reconstruction proceeds practically from a single component with the maximum value of the fraction in (15). On the other hand, with  $\alpha$  close to zero the weighting becomes uniform, which means no separation.

An example in Section IV. indicates that a good choice of  $\alpha$  is  $\alpha = 1$ . Figure 7 shows resulting weights obtained in the demonstration example for this choice.

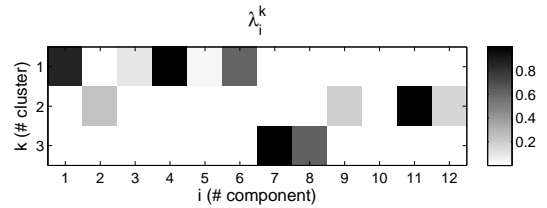


Fig. 7. Weights of components from Fig. 3 computed according to the rule (15) with  $\alpha = 1$ .

#### G. Oracle Weighting

It is interesting to know what would be the best possible weights for separation in theory, given the ICA decomposition of the observation space. In other words, what are the best possible weights independent of the similarity given by  $\mathbf{D}$  and the clustering algorithm. Such weights can be derived using known responses of sources. The authors call it an “oracle weighting”, and the corresponding algorithm an “oracle algorithm”, following the work of Vincent et al. [8].

The oracle weighting can be derived as the one that minimizes  $\|\mathbf{S}_k - \widehat{\mathbf{S}}_k\|_F^2$ ,  $k = 1, \dots, d_{est}$ , given the true responses of sources on the microphones forming the matrix  $\mathbf{S}_k$ . Here  $\|\cdot\|_F$  denotes the Frobenius norm, and  $d_{est} = d$ . Using (5), the oracle weights are defined by

$$\mathbf{l}_k = \arg \min_{\mathbf{l}} \|\mathbf{S}_k - \mathbf{W}^{-1} \text{diag}(\mathbf{l}) \mathbf{W} \mathbf{X}\|_F^2, \quad (16)$$

where  $\mathbf{l}_k = [\lambda_1^k, \dots, \lambda_M^k]^T$ . After some computations it can be shown that

$$\mathbf{l}_k = \left( (\mathbf{W} \mathbf{X} \mathbf{X}^T \mathbf{W}^T) \odot (\mathbf{W} \mathbf{W}^T)^{-1} \right)^{-1} \cdot \text{diag} [\mathbf{W}^{-T} \mathbf{S}_k \mathbf{X}^T \mathbf{W}^T], \quad (17)$$

and  $\odot$  denotes the Hadamard (element-wise) product. The rest of the oracle algorithm (reconstruction and beamforming) proceeds normally.

#### H. Step 5: Beamforming

A beamformer can be applied to the multi-channel estimate of each source (microphone responses) to yield a mono-channel estimate of the source. This problem is not addressed here, because it exceeds the scope of this article. The beamforming requires an additional definition of a principle that is not given in the blind scenario considered here. The reader is referred to [16].

Results obtained in the demonstration example after delay-and-sum beamforming of estimated microphone responses are shown in Fig. 8. The order of estimated signals with respect to the original ones is arbitrary.

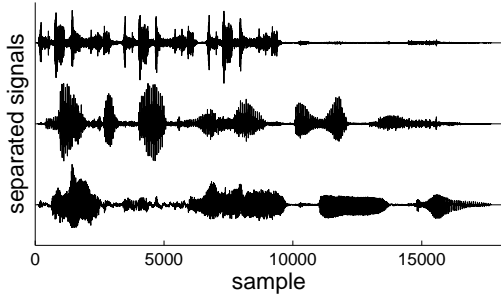


Fig. 8. Estimates of original sources obtained by the delay-and-sum beamformer applied to estimated microphone responses.

### III. GENERALIZED OBSERVATION SPACE

In this section, the previous definition of the observation space is altered (generalized) in order to enable the application of long (IIR) separating filters while keeping the same dimension of the observation space. This makes the method computationally affordable even when long separating filters are considered.

It is worth pointing out that maintaining a reasonable dimension of the observation space is also desirable from the probabilistic theory point of view. The ICA is a stochastic method whose accuracy measured in terms of the interference-to-signal ratio is, in theory, proportional to the number of components  $M$ ; see for example [24], [28], [29].

The original construction of  $\mathbf{X}$  given by (4) means that FIR filters are applied to the mixed signals, because the outputs of FIR filters can be seen as linear combinations of time-shifted versions of the input signals (rows of (4)). The proposed generalization consists in constructing  $\mathbf{X}$  so that separating filters have a well-known generalized feed-forward (FF) structure [30], [31], which also embodies FIR filters as a special case.

The output of a generalized FF filter applied to an input signal  $x(n)$  can be written as

$$\begin{aligned} y(n) &= \sum_{\ell=1}^L w_{\ell} \sum_{\tau=-\infty}^{+\infty} f_{\ell}(\tau)x(n-\tau) \\ &= \sum_{k=1}^L w_{\ell} \{f_{\ell} \star x\}(n), \end{aligned} \quad (18)$$

where  $\star$  denotes the convolution,  $w_{\ell}$  are the filter weights, and the filters  $f_{\ell}$  are called *eigenmodes* of the filter. The definition of MIMO filters with the generalized FF structure is analogous.

#### A. Generalized Observation Matrix $\mathbf{X}$

For a given set of invertible eigenmodes  $f_{\ell}$ , the  $i$ th block of the observation matrix  $\mathbf{X}$  can be defined as

$$\mathbf{X}_i = \begin{bmatrix} \{f_1 \star x_i\}(N_1) & \dots & \dots & \{f_1 \star x_i\}(N_2) \\ \{f_2 \star x_i\}(N_1) & \dots & \dots & \{f_2 \star x_i\}(N_2) \\ \vdots & \vdots & \vdots & \vdots \\ \{f_{\ell} \star x_i\}(N_1) & \dots & \dots & \{f_{\ell} \star x_i\}(N_2) \end{bmatrix}, \quad (19)$$

The whole  $\mathbf{X}$  is then given by

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{bmatrix}. \quad (20)$$

If  $f_{\ell}$  is the all-pass filter that realizes backward time-shift by  $\ell$  samples, the construction in (20) coincides with the one in (4).

*Example of perfect separation:* Consider the general  $2 \times 2$  scenario

$$x_1(n) = \{h_{11} \star s_1\}(n) + \{h_{12} \star s_2\}(n) \quad (21)$$

$$x_2(n) = \{h_{21} \star s_1\}(n) + \{h_{22} \star s_2\}(n). \quad (22)$$

Almost perfect separation can be achieved when taking  $L = 2$  and applying special eigenmodes for each matrix  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , namely,  $f_{11} = g \star h_{22}$ ,  $f_{12} = -g \star h_{21}$ ,  $f_{21} = -g \star h_{12}$ , and  $f_{22} = g \star h_{11}$ , where  $g = (h_{11} \star h_{22} - h_{21} \star h_{12})^{-1}$  assuming that the inversion exists. A trivial verification shows that combinations of signals  $\{f_{11} \star x_1\}(n) + \{f_{21} \star x_2\}(n)$  and  $\{f_{12} \star x_1\}(n) + \{f_{22} \star x_2\}(n)$  are, respectively, equal to the original independent sources  $s_1$  and  $s_2$ . In other words,  $s_1$  and  $s_2$  can be found as independent components in the observation space. ■

The example demonstrates the great potential of the general construction of  $\mathbf{X}$  in theory. For instance, it is indicative of the possibility to tailor the eigenmodes to room acoustics.

After the ICA decomposition of  $\mathbf{X}$  the method proceeds normally up to the fourth reconstruction step. Let  $\psi_{k,p}(n)$  be the  $(p, n)$ th element of  $\hat{\mathbf{S}}_k$ . Then,  $\psi_{k,p}(n)$ ,  $p = (i-1)L + \ell$ , provides an estimate of  $\{f_{\ell} \star s_k^i\}(n)$ . Let  $f_{\ell}^{-1}$  be the inverse of  $f_{\ell}$  so that  $f_{\ell} \star f_{\ell}^{-1} = \delta$ . The authors estimate the response of the  $k$ th separated source at the  $i$ th microphone as

$$\hat{s}_k^i(n) = \frac{1}{L} \sum_{\ell=1}^L \{f_{\ell}^{-1} \star \psi_{k,(i-1)L+\ell}\}(n). \quad (23)$$

Obviously, (23) is a generalization of (6).

#### B. Laguerre Filters

A good example of FF filters for this study are Laguerre filters parametrized by  $\mu$  from  $(0, 2)$ , which were considered in [31]. They are defined recursively, through their transfer functions

$$F_1(z) = 1, \quad (24)$$

$$F_2(z) = \frac{\mu z^{-1}}{1 - (1 - \mu)z^{-1}}, \quad (25)$$

$$F_n(z) = F_{n-1}(z)G(z), \quad n = 3, \dots, L, \quad (26)$$

where

$$G(z) = \frac{(\mu - 1) + z^{-1}}{1 - (1 - \mu)z^{-1}}. \quad (27)$$

Note that  $f_2$  is either a low-pass filter (for  $0 < \mu < 1$ ) or a high-pass filter (for  $1 < \mu < 2$ ), and  $g$  is an all-pass filter.

The construction discussed here is a generalization of (4), because for  $\mu = 1$ ,  $F_2(z) = G(z) = z^{-1}$ , that is  $f_2(n) =$

$g(n) = \delta(n-1)$ . This is the only case where separating filters are FIR of the length  $L$ . For  $\mu \neq 1$ , the filters are IIR.

The so-called *memory depth* denoted by  $L_*$  is defined as the minimum length needed to capture 90% of the total energy contained in the impulse response. For the Laguerre filters it approximately holds that [31]

$$L_* = (1 + 0.4|\mu - 1| \log_{10} L)L/\mu. \quad (28)$$

From here on, we will name the proposed method T-ABCD (Time-domain Audio sources Blind separation based on the Complete Decomposition of the observation space) keeping in mind that its given variant must be specified at the place where the acronym is used.

#### IV. EXPERIMENTS

##### A. Experiments with short recordings of two sources

The experiments described in this section were designed to compare T-ABCD in setups for which the method was specialized. Those are mainly situations where only short recordings are available, and short separating filters ( $L = 3, \dots, 40$ ) are used (more precisely, the dimension of the observation space is in the range of tens).

Data used for testing of T-ABCD consist of nine recordings of two simultaneously talking persons articulating short commands. The length of each recording is 2 s, which gives 16000 samples at 8kHz sampling. Different genders are considered so that there are three recordings of male/male, three of female/male, and three of female/female speakers.

The recordings were obtained by two closely spaced microphones when playing the speakers' commands over two loudspeakers. Microphone responses of each source were obtained by recording the source when the other sources were silent<sup>2</sup>.

Three different positions of loudspeakers were considered that differ in distance and angle between sources; see Fig. 9 and Table I. Each scenario was situated in an ordinary living room with the reverberation time of about 300 ms.

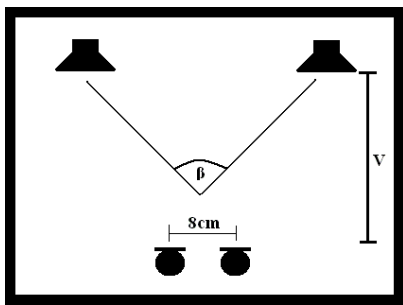


Fig. 9. Illustration of positions of sources (loudspeakers) and microphones.

Two variants of T-ABCD were tested using the BGSEP and the EFICA algorithm in the second step, respectively, marked as  $T-ABCD_b$  and  $T-ABCD_e$ . For theoretical reasons performances of the oracle algorithms (Section II.G) based

<sup>2</sup>Equipment used for recordings consists of external sound device EDIROL FA-101 and condenser omnidirectional microphones Audio-technica AT831b.

TABLE I  
TECHNICAL DETAILS OF SCENARIOS

scenario	$\beta$	$V$ [cm]	Average SNR [dB]
1	$70^\circ$	50	40.6
2	$60^\circ$	100	38.6
3	$75^\circ$	200	35.9

on these ICA algorithms were also studied. They are denoted as *Oracle-b* and *Oracle-e*, respectively. Finally, an ultimate performance bound determined by the MMSE estimator [31] is shown, which is computed for the separating filter length  $2L - 1$ .

Two other algorithms were used for comparisons. The first one is the STFICA algorithm from [18] using two stages of preprocessing (prewhitening) of the length 300. The observation space separated by STFICA is set to have the same dimension as the proposed method, that is,  $2L$  for two microphones. The second algorithm is that of Parra from [6] with two lengths of FFT, respectively, 512 (*Parra-1*) and 128 (*Parra-2*); the other parameters had the default values.

Results of these experiments are evaluated by two standard measures [32]: Signal-to-Interference ratio (SIR) and Signal-to-Distortion ratio (SDR). The SIR determines the ratio of energies of the desired signal and the interference in the separated signal. SIR is highly influenced by a filtering of the measured signal, which might be misleading, especially, in audio separation. It is also influenced by the input SIR, which is the SIR measured before the separation. In our experiments, the input SIR was always about 0 dB, which means that both sources were approximately equally loud. The SDR provides a supplementary criterion of SIR that reflects the difference between the desired and the estimated signal in the mean-square sense. SDR is, by contrast, highly sensitive to the filtering, which may yield a rigid evaluation of methods applying long separating filter. It is therefore advisable to consider both criteria.

Hereinafter, all results are evaluated in terms of averaged SIR improvement and SDR over all separated sources and over all their estimated microphone responses. The results are also averaged over the nine recorded combinations of signals to reduce the effect of statistical properties of the recorded signals.

1) *Performance versus  $L$* : Here, the separation is done for different lengths of separating filter  $L$ , and the other parameters are fixed. Namely, only 8000 samples of data (the first second) are used for the computation. T-ABCD utilizes the basic construction of the observation space corresponding to  $\mu = 1$ .

Figure 10 shows results of separation obtained by processing signals from scenario 1. SDR of T-ABCD improves with growing  $L$ , similar to the SDR of the MMSE estimator. SIR does not improve with growing  $L$ , but is good for all  $L$ . This is explained by the fact that for small  $L$ , few components are used to reconstruct sources, and SIR remains good, but SDR is poorer, because the reconstructed sources have different coloration than the original responses. In this respect, the behavior of oracle algorithms is different as they primarily optimize SDR. The gap between the SIR/SDR of the oracle

algorithms and T-ABCD indicates that there might still be a room for improvements of the performance through clustering and weighting.

Separated signals obtained by the other algorithms, STFICA, Parra-1 and Parra-2 are perceptually not bad, but not as good as those of the proposed method. It does not improve with growing  $L$  neither in terms of SIR nor in terms of SDR. STFICA failed to converge for  $L \geq 20$ .

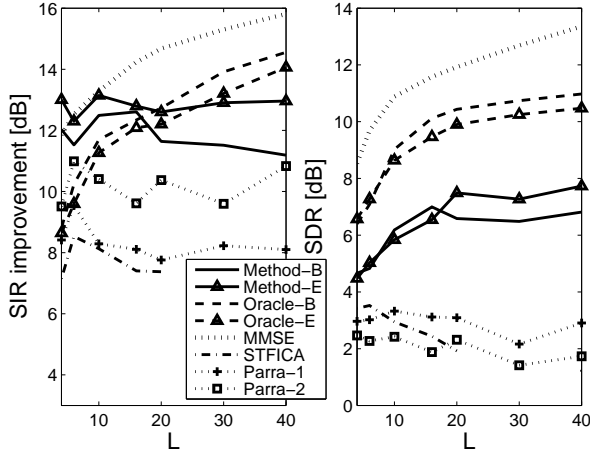


Fig. 10. SIR and SDR as functions of the length of separating filter  $L$ . The results were obtained by using data from scenario 1.

2) *Performance versus Length of Data*: A similar experiment to the previous one was repeated for a fixed  $L = 20$  and varying the length of data used for computations of separating filter in scenario 2. This scenario is more difficult for the separation because of the higher distance of sources and lower ratio between the energy of the direct-path source signals and the energy of their reflections. The results are shown in Fig. 11.

It is noted that for a fixed filter length, there is a certain length of the data beyond which performance of the algorithms does not improve at all. In this experiment, the length was about 0.8-1s. T-ABCD performs better than the other algorithms and demonstrates its superior capability to work with short data.

3) *Performance versus  $\alpha$* : The parameter  $\alpha$  was introduced in (15), and provides some trade-off between SIR and SDR. This is demonstrated by separating signals from scenario 2 by T-ABCD with  $L = 20$ , 8000 samples of data for computations, and various  $\alpha$ . Results are shown in Fig. 12.

It is noted that SIR is an increasing function of  $\alpha$ , whereas SDR achieves its maximum at a certain value in the vicinity of  $\alpha = 1$ . This points to the need of using SIR and SDR simultaneously to evaluate the separation fairly.

4) *Performance of T-ABCD using Laguerre filters versus  $\mu$* : The signals recorded in scenario 3 were separated with  $L = 20$  using the first second of data. The parameter  $\mu$  was gradually decreased from 1.9 to 0.1, which corresponds to changing the separating filter memory  $L_*$  defined by (28) from 15 to 293 samples; see Fig. 13.

The results indicate a minor (about 0.7 dB) improvement of performance of T-ABCD at the optimum value  $\mu = 0.2$

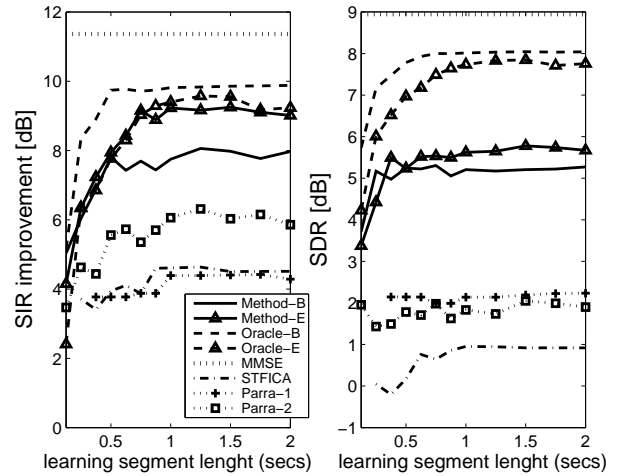


Fig. 11. SIR and SDR as functions of the length of data used for the computation of separating filter. Results corresponds to data from scenario 2.

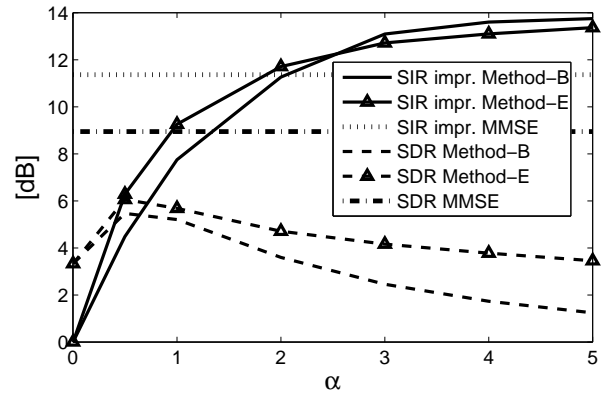


Fig. 12. Performance of T-ABCD as a function of  $\alpha$ . In this example, STFICA, Parra-1 and Parra-2 performed, respectively, with 4.6dB, 4.4dB and, 6dB of SIR improvement and 1dB, 1.8dB, and 2.2dB of SDR.

compared to  $\mu = 1$ . A higher potential improvement in performance is indicated by increased SIR improvement of the MMSE bound and of the oracle algorithms (about 2dB). Again, it indicates a room for improvement through a different clustering and weighting.

## B. Experiments with Hiroshi Sawada's Recordings

In this subsection, the above methods were tested by separating data available on the Internet<sup>3</sup>. These data were recorded in a room with the reverberation time 130ms. A linear microphone array with the distance of 4cm between microphones was used to record 2-4 simultaneous speeches coming from different directions from the distance of 1.2 m at the sampling rate 8kHz. The length of the recordings is 7 s.

The data were processed by T-ABCD with Laguerre filters with  $\mu = 0.2$  and  $L = 30$ . Therefore, the dimension of the observation space was equal to  $30m$ , where  $m$  is the number

<sup>3</sup><http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/bss2to4/index.html>

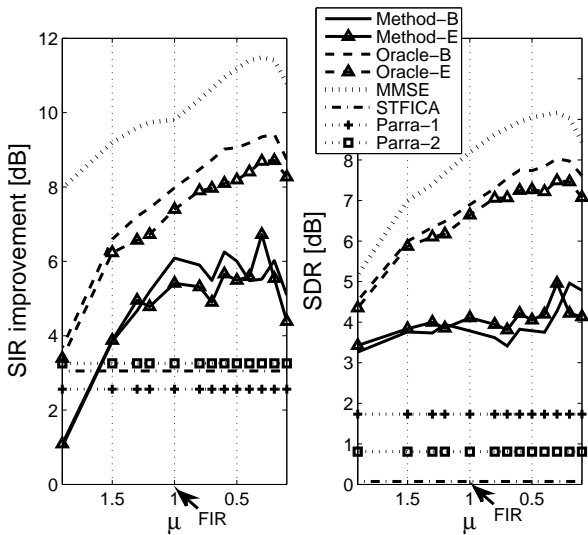


Fig. 13. SIR and SDR as functions of the parameter  $\mu$  of Laguerre filters. Results corresponds to data from scenario 3. Note that for  $\mu = 1$  the Laguerre filters are FIR, and the generalized T-ABCD coincides with the original one described in Section II.

of microphones (2-4). The algorithm used 8000 samples (1s) to perform the separation, beginning at 4.6s of the recordings. For the Parra's algorithm, the whole (7 s long) recordings were used, the length of the FFT was 1024 and the time-domain filter had 400 taps (the same setting as in [18]). The STFICA algorithm had the preprocessing length of 50 taps, and the separating system had  $L = 15$  taps. (The algorithm did not converge with a larger  $L$  and longer preprocessing.)

Results of the comparison are summarized in Fig. 14. It contains performance of the Sawada's algorithm [7], which works in the frequency domain, here, using the FFT length 2048 with the overlap of 512 samples.

It can be seen that T-ABCD outperforms STFICA and the Parra's algorithm, but is worse than that of Sawada's algorithm, whose results were taken from the web-site. The latter algorithms take the advantage of utilizing the whole data for the separation. In the case of four sources, performance of the proposed algorithm and the Sawada's algorithm are almost equal in the terms of SDR.

The SIR and the SDR of the Sawada's algorithm can be higher than these quantities for the MMSE, because the MMSE is computed for the filter length equal to  $30m$ , while the Sawada's algorithm applies filters of the length 2048 taps.

## V. CONCLUSIONS

The novel time-domain algorithm has been proposed for blind separation of audio sources that is based on the complete unconstrained ICA decomposition of the observation space. The algorithm, named T-ABCD, is suitable for situations where only short data records are available. In this respect, it outperforms other known time-domain BSS algorithms.

T-ABCD consists of five steps, each one providing a room for other variants and improvements. In particular, the selection of eigenmodes may lead to a more effective definition of

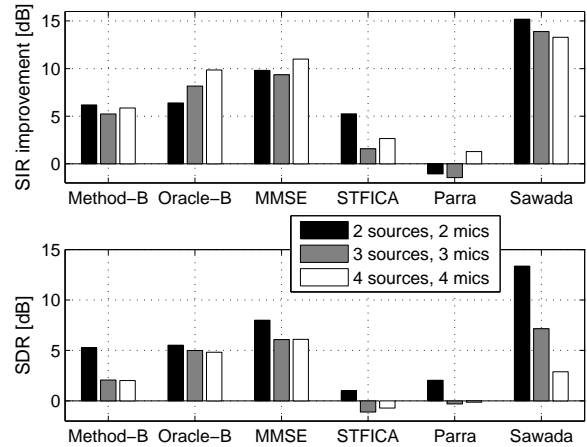


Fig. 14. Performance of separation methods applied to Sawada's data with 2-4 sources and microphones.

the observation space. The comparison with the oracle algorithm showed that the measure of the similarity of components, their clustering, and weighting might be still significantly improved.

Finally, as T-ABCD works with short data segments, it has great potential to be modified for online or batch processing needed in situations with moving sources.

## REFERENCES

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley-Interscience, New York, 2001.
- [2] A. Cichocki and S.-I. Amari, *Adaptive Signal and Image Processing: Learning Algorithms and Applications*, Wiley, New York, 2002.
- [3] S. C. Douglas, A. Cichocki, and S. Amari, "Self-Whitening Algorithms for Adaptive Equalization and Deconvolution," *IEEE Trans. Signal Processing*, vol. 47, no. 4, pp. 1161-1165, April 1999.
- [4] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," *Proceedings of 3rd International Conference on Independent Component Analysis and Blind Source Separation (ICA '01)*, pp. 722-727, San Diego, Calif, USA, Dec. 2001.
- [5] N. Mitianoudis and M. E. Davies, "Audio Source Separation of Convulsive Mixtures," *IEEE Trans. on Speech Audio Process.*, vol. 11, no. 5, pp. 489-497, Sep. 2003.
- [6] Parra, L., and Spence, C.: "Convulsive Blind Separation of Non-Stationary Sources", *IEEE Trans. on Speech and Audio Processing*, Vol. 8, No. 3, pp. 320-327, May 2000.
- [7] 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 Audio Processing*, vol. 12, no. 5, pp. 530-538, 2004.
- [8] E. Vincent, R. Gribonval, and M.D. Plumbley, "Oracle estimators for the benchmarking of source separation algorithms," *Signal Processing*, 87(8):1933-1950, 2007.
- [9] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meraim, "Jacobi-like algorithm for blind signal separation of convulsive mixtures," *IEE Elec. Letters*, vol. 37, no. 16, pp. 1049-1050, Aug. 2001.
- [10] C. Févotte, A. Debiolles, and C. Doncarli. "Blind separation of FIR convulsive mixtures: application to speech signals," *Proc. 1st ISCA Workshop on Non-Linear Speech Processing*, 2003.

- [11] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel Blind Deconvolution and Equalization Using the Natural Gradient," *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, pp. 101-104, April 1997.
- [12] S. C. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 1, pp. 92-104, Jan. 2005.
- [13] M. Gupta and S. C. Douglas, "Beamforming Initialization and Data Prewhitening in Natural Gradient Convolutional Blind Source Separation of Speech Mixtures," *Proc. of ICA 2007*, LNCS 4666, pp. 462-470, Sept. 2007.
- [14] H. Buchner, R. Aichner, and W. Kellermann, "A Generalization of Blind Source Separation Algorithms for Convolutional Mixtures Based on Second-Order Statistics," *IEEE Trans. on Speech and Audio Processing*, Vol. 13, No. 1, pp. 120-134, Jan. 2005.
- [15] H. Buchner, R. Aichner, and W. Kellermann, "TRINICON: A Versatile Framework for Multichannel Blind Signal Processing," *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, vol. 3, pp. 889-892, May 2004.
- [16] J. Benesty, S. Makino and J. Chen, *Speech Enhancement*, Springer, March 2005.
- [17] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," *Proc. EUSIPCO*, pp. 776-779, Edinburgh, Sept. 1994.
- [18] S. Douglas, M. Gupta, H. Sawada, and S. Makino, "Spatio-temporal FastICA algorithms for the blind separation of convolutional mixtures," *IEEE Trans. Audio, Speech and Language Processing*, vol. 15, no. 5, pp. 1511-1520, July 2007.
- [19] J.-F. Cardoso, "Multidimensional independent component analysis," *Proc. of ICASSP 1998*, Vol. 4, pp. 1941-1944, 12-15 May 1998.
- [20] P. Tichavský and A. Yeredor, "Fast Approximate Joint Diagonalization Incorporating Weight Matrices," *IEEE Transactions of Signal Processing*, vol. 57, no.3, pp. 878-891, March 2009.
- [21] Z. Koldovský, P. Tichavský, "Time-domain blind audio source separation using advanced component clustering and reconstruction," *Proc. of HSCMA 2008*, Trento, Italy, pp. 216-219, May 2008.
- [22] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, Anh Huy Phan, and R. Zdunek, ICALAB MATLAB Toolbox Ver. 3 for Signal Processing, [online], <http://www.bsp.brain.riken.jp/ICALAB>.
- [23] Z. Koldovský and P. Tichavský, "A Comparison of Independent Component and Independent Subspace Analysis Algorithms," *Proc. of EUSIPCO 2009*, pp. 1447-1451, Glasgow, August 24-28, 2009.
- [24] Z. Koldovský, P. Tichavský and E. Oja, "Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound," *IEEE Trans. on Neural Networks*, Vol. 17, No. 5, Sept 2006.
- [25] A. Hyvärinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," *IEEE Transactions on Neural Networks*, 10(3):626-634, 1999.
- [26] K. Abed-Meraim and A. Belouchrani, "Algorithms for Joint Block Diagonalization," *Proc. of EUSIPCO 2004*, pp. 2092-212, Vienna, Austria, 2004.
- [27] C. Févotte and F. J. Theis, "Orthonormal approximate joint block-diagonalization," *Technical Report GET/TIcom Paris 2007D007*, 2007.
- [28] Shalvi, O.; Weinstein, E., "Maximum likelihood and lower bounds in system identification with non-Gaussian inputs," *IEEE Tr. Information Theory*, Vol. 40, No. 2, March 1994, pp. 328 - 339.
- [29] D-T. Pham and J-F. Cardoso. "Blind separation of instantaneous mixtures of non stationary sources," *IEEE Trans. Signal Processing*, pp. 1837-1848, vol. 49, no. 9, 2001.
- [30] J. C. Principe, B. de Vries, Guedes de Oliveira, "Generalized feedforward structures: a new class of adaptive filters," *Proc. of ICASSP 1992*, Vol. 4, pp. 245-248, March 1992.
- [31] K. E. Hild II, D. Erdogmuz, J. C. Principe, "Experimental Upper Bound for the Performance of Convolutional Source Separation Methods", *IEEE Trans. on Signal Processing*, Vol. 54, No. 2, pp. 627-635, Feb. 2006.
- [32] D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," *Int. Workshop on Independent Component Analysis and Signal Separation (ICA '99)*, Aussois, France, pp. 261-266, Jan. 1999.
- [33] Z. Koldovský, Matlab<sup>TM</sup> p-code of the proposed algorithm, [online], <http://itakura.ite.tul.cz/zbynek/downloads.htm>
- [34] Z. Koldovský and P. Tichavský, "Způsob slepé separace akustických signálů z jejich konvolutorní směsi," Czech Patent Application PV 2008-752, Nov. 2008.



**Zbyněk Koldovský** (S'03-M'04) was born in Jablonec nad Nisou, Czech Republic, in 1979. He received the M.S. degree and Ph.D. degree in mathematical modeling from Faculty of Nuclear Sciences and Physical Engineering at the Czech Technical University in Prague in 2002 and 2006, respectively. He is currently with the Institute of Information Technology and Electronics, Technical University of Liberec. He has also been with the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic since 2002.

His main research interests are in audio signal processing, blind source separation and independent component analysis.



**Petr Tichavský** (M'98, SM'04) received the M.S. degree in mathematics in 1987 from the Czech Technical University, Prague, Czechoslovakia and the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences in 1992. Since that time he is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic in Prague. In 1994 he received the *Fulbright grant* for a 10 month fellowship at Yale University, Department of Electrical Engineering, in New Haven, U.S.A. In 2002 he received the *Otto Wichterle Award* from Academy of Sciences of the Czech Republic.

He is author and co-author of research papers in the area of sinusoidal frequency/frequency-rate estimation, adaptive filtering and tracking of time varying signal parameters, algorithm-independent bounds on achievable performance, sensor array processing, independent component analysis and blind signal separation.

Petr Tichavský served as associate editor of the *IEEE Signal Processing Letters* from 2002 to 2004, and as associate editor of the *IEEE Transactions on Signal Processing* from 2005 to 2009. Since 2009 he is a member of the IEEE Signal Processing Society's Signal Processing Theory and Methods (SPTM) Technical Committee. Petr Tichavský also serves as a general chair of the 36th IEEE Int. Conference on Acoustics, Speech and Signal Processing ICASSP 2011 in Prague, Czech Republic.