# Blind Separation of Underdetermined Convolutive Mixtures Using Their Time–Frequency Representation

Abdeldjalil Aïssa-El-Bey, Karim Abed-Meraim, *Senior Member, IEEE*, and Yves Grenier, *Member, IEEE*

***Abstract*—This paper considers the blind separation of nonstationary sources in the underdetermined convolutive mixture case. We introduce, two methods based on the sparsity assumption of the sources in the time–frequency (TF) domain. The first one assumes that the sources are disjoint in the TF domain, i.e., there is at most one source signal present at a given point in the TF domain. In the second method, we relax this assumption by allowing the sources to be TF-nondisjoint to a certain extent. In particular, the number of sources present (active) at a TF point should be strictly less than the number of sensors. In that case, the separation can be achieved thanks to subspace projection which allows us to identify the active sources and to estimate their corresponding time–frequency distribution (TFD) values. Another contribution of this paper is a new estimation procedure for the mixing channel in the underdetermined case. Finally, numerical performance evaluations and comparisons of the proposed methods are provided highlighting their effectiveness.**

***Index Terms*—Blind source separation (BSS), convolutive mixture, sparse signal decomposition/representation, speech signals, subspace projection, time–frequency distribution (TFD), underdetermined/overcomplete representation, vector clustering.**

## I. Introduction

THE OBJECTIVE of blind source separation (BSS) is to extract the original source signals from their mixtures and possibly to estimate the unknown mixing channel using only the information of the observed signal with no, or very limited, knowledge about the source signals and the mixing channel. The BSS problem arises in many fields of study including speech processing, data communication, biomedical signal processing, etc. [1]. Most approaches to blind source separation assume the sources are statistically independent and thus are often seek solutions of separation criteria using higher order statistical information [2] or using only second-order statistical information in the case where the sources have temporal coherency [3], are nonstationary [4], or eventually are cyclostationary [5].

Although the plethora of existing BSS algorithms, the underdetermined case (UBSS for underdetermined blind source separation) where the number of sources is greater than the number of sensors remains relatively poorly treated especially in the convolutive case, and its resolution is one of the challenging problems of blind source separation. In the instantaneous mixture case, some methods exploiting the sparseness of the sources in certain transform domain have been proposed for UBSS [6]–[10]. Other methods consider similarly underdetermined mixtures of delayed sources [11], [12]. All these methods proceed "roughly" as follows: The mixtures are first transformed to an appropriate representation domain; the transformed sources are then estimated using their sparseness, and finally one recovers their time waveforms by source synthesis (for more information, see the recent survey work [13]).

The UBSS methods for *nonstationary sources* have been proposed, given that these sources are sparse in the time–frequency (TF) domain [7], [11]. The first method uses time–frequency distributions (TFDs), whereas the second one uses a linear TFD. The main assumption used in these methods is that the sources are TF-disjoint. In other words, there is at most one source present at any point in the TF domain. This assumption is rather restrictive, though the methods have also showed that they worked well under a quasi-sparseness condition, i.e., sources are TF-almost-disjoint.

In this paper, we focus on the UBSS in convolutive mixtures case and target the relaxation of the TF-disjoint condition by allowing the sources to be *nondisjoint* in the TF domain; that is, multiple sources are possibly present at any point in the TF domain. This case has been considered in [8] for the separation of instantaneous mixtures, in [12] for the deconvolution of single-path channels with nonzero delays, in [14] where *a priori* information about the location of the considered sources as well as an approximation of the filter impulse response are considered, and in [15] where binary TF-masking (or directivity pattern based masking [16], [17]) and the independent component analysis (ICA) technique are jointly used. In particular, we limit ourselves to the scenario where the number of sources present at any point is smaller than the number of sensors. Under this assumption, the separation of TF-nondisjoint sources is achieved thanks to *subspace projection*. Subspace projection allows us to identify at any point the active sources, and then to estimate their corresponding TFD values.

The main contribution of this paper consists in two new algorithms for UBSS in the TF domain; the first one uses vector clustering while the other uses subspace projection. Another side contribution of the paper is an estimation method for the mixing channel matrix.

The paper is organized as follows. Section II-A formulates the UBSS problem, introduces the underlying TF tools, and

Manuscript received July 1, 2006; revised January 31, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shoji Makino.

The authors are with the TSI Department, ENST-Paris, 75634 Paris Cedex 13, France (e-mail: elbey@tsi.enst.fr; abed@tsi.enst.fr; grenier@tsi.enst.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2007.898455

states some TF conditions necessary for the separation of nonstationary sources in the TF domain. In Section III-A, we propose a new method for the blind estimation of mixing channel. Section III-B deals with the TF-disjoint sources. It proposes a cluster-based TF-CUBSS (time–frequency convolutive underdetermined blind source separation) algorithm. Section III-C proposes the subspace-based TF-CUBSS algorithm for TF-nondisjoint sources. Some comments and remarks on the proposed methods are provided in Section IV. Finally, the performance of the above methods are numerically evaluated in Section V while Section VI is devoted for the concluding remarks.

## II. Problem Formulation

### A. Data Model

Let $s_1(t), \ldots, s_N(t)$ be the desired sources to be recovered from the convolutive mixtures $x_1(t), \ldots, x_M(t)$ given by

$$\mathbf{x}(t) = \sum_{k=0}^{K} \mathbf{H}(k)\mathbf{s}(t-k) + \boldsymbol{\eta}(t) \tag{1}$$

where $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^T$ is the source vector with the superscript $^T$ denoting the transpose operation, $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T$ is the mixture vector, $\boldsymbol{\eta}(t)$ is the observation noise, and $\mathbf{H}(k) \stackrel{\text{def}}{=} [\mathbf{h}_1(k), \ldots, \mathbf{h}_N(k)]$ are $M \times N$ matrices for $k \in [0, K]$ representing the impulse response coefficients of the channel that satisfies the following.

*Assumption:* The channel is such that each column vector of

$$\mathbf{H}(z) \stackrel{\text{def}}{=} \sum_{k=0}^{K} \mathbf{H}(k) z^{-k} \stackrel{\text{def}}{=} [\mathbf{h}_1(z), \ldots, \mathbf{h}_N(z)]$$

is irreducible, i.e., the entries of $\mathbf{h}_i(z)$ denoted $h_{ij}(z), j = 1, \ldots, M$, have no common zeros $\forall i$. Moreover, any $M$ column vectors of $\mathbf{H}(z)$ form a polynomial matrix $\tilde{\mathbf{H}}(z)$ that it full rank over the unit-circle, i.e., $\text{rank}(\tilde{\mathbf{H}}(f)) = M \, \forall f$.

The sources are nonstationary, that is their frequency spectra vary in time. Often, nonstationarity gives rise to more difficulties in a problem; however, in this case it actually offers certain diversity that allows us to achieve the BSS without using higher order approaches by directly exploiting the additional information of this TF diversity across the spectra [4]. In that case, we often use the powerful tool of time–frequency signal analysis which basic concept is introduced next.

### B. Time–Frequency Distributions

TF signal processing provides effective tools for analyzing nonstationary signals, whose frequency content varies in time. This concept is a natural extension of both the time domain and the frequency domain processing that involves representing signals in a two-dimensional space, the joint TF domain, hence providing a distribution of signal energy versus time and frequency simultaneously. For this reason, a TF representation is commonly referred to as a TFD.

Well-known TFD[1] and most used in practice is the short-time Fourier transform (STFT)

$$\mathcal{S}_x(t, f) \triangleq \int_{-\infty}^{\infty} x(\tau) w(\tau - t)\, e^{-j2\pi f \tau} d\tau \tag{2}$$

where $w(t)$ is a windowing function, and $x(t)$ a given nonstationary signal. Note that the STFT is a *linear* TFD and thus has the advantage of simplicity compared to other nonlinear (quadratic) TFDs, e.g., Wigner–Ville and Cohen's class distributions [18].

### C. TF Conditions on the Sources

In order to deal with UBSS, one often seeks for a sparse representation of the sources [6]. In other words, if the sources can be sparsely represented in some domain, then their separation can be carried out in that domain by exploiting their sparseness.

*1) TF-Disjoint Sources:* Recently, there have been several UBSS methods, notably those in [7] and [11], in which the TF domain has been chosen to be the underlaying sparse domain. These two papers have based their solutions on the assumption that the sources are disjoint in the TF domain. Mathematically, if $\Omega_1$ and $\Omega_2$ are the TF supports of two sources $s_1(t)$ and $s_2(t)$ then the sources are said TF-disjoint if $\Omega_1 \cap \Omega_2 = \emptyset$. However, this is a rather strict assumption. A more practical assumption is that the sources are almost disjoint in the TF domain [7], allowing some small overlapping in the TF domain, for which the above two methods (in [7] and [11]) also worked.

*2) TF-Nondisjoint Sources:* In this paper, we want to relax the TF-disjoint condition by allowing the sources to be nondisjoint in the TF domain.

This is motivated by a drawback of the methods in [7] and [11]. Although these methods worked under the TF-almost-disjoint condition, they did not explicitly treat the TF regions (points) where the sources were overlapping. A point at the overlapping of two sources was assigned "by chance" to belong to only one of the sources. As a result, the source that picks up this point will have some information of the other source while the latter loses some information of its own. The loss of information can be recovered to some extent by the interpolation at the intersection point using TF synthesis. However, for the other source, there is an interference at this point; hence, the separation performance may degrade if no treatment is provided. If the number of overlapping points increases (i.e., the TF-almost-disjoint condition is violated), the performance of the separation is expected to degrade unless the overlapping points are properly treated.

This paper will give such a treatment using subspace projection. Therefore, we will allow the sources to be *nondisjoint* in the TF domain; that is, multiple sources are allowed to be present at any point in the TF domain. However, instead of being inevitably nondisjoint, we limit ourselves by making the following constraint.

[1]In fact, the STFT does not represent an energy distribution of the signal in the TF plane. However, for simplicity, we still refer to it as a TFD.

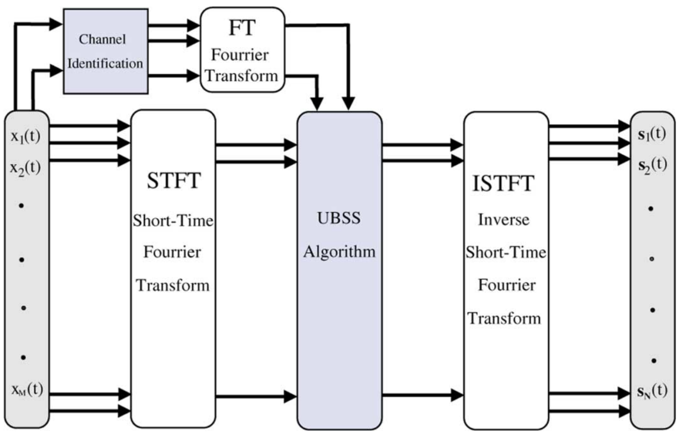


Fig. 1. Diagram of proposed TF-CUBSS algorithm combining channel identification and UBSS technique in TF domain.

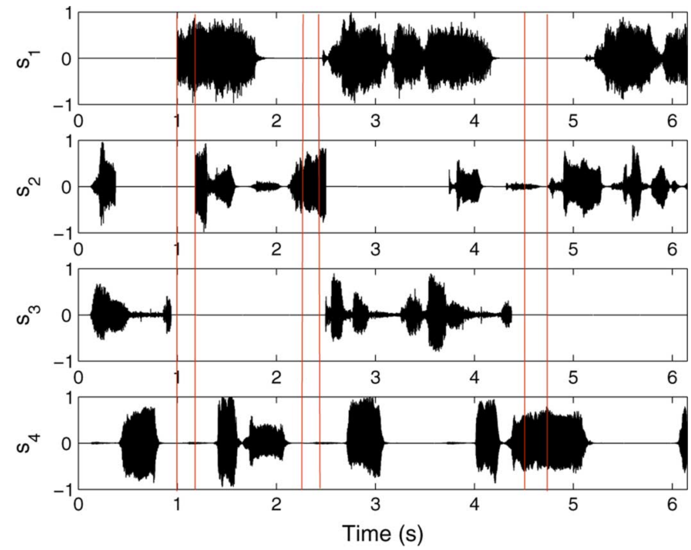


Fig. 2. Time representation of four audio sources: This representation illustrates the audio signal sparsity (i.e., there exists time intervals where only one source is present).

*Assumption 2:* The number of active sources (i.e., sources that overlap) at any TF point is strictly less than the number of sensors.

In other words, for the configuration of $M$ sensors, there exists at most $(M-1)$ sources at any point in the TF domain. For the special case when $M=2$, Assumption 2 reduces to the disjoint condition.

Note that in [15]–[17], the case of $M$ overlapping sources has been treated thanks to additional strong assumptions that we do not consider in our work. More specifically, the channels are assumed to be of single path with a given direction of arrival, and the sources are such that one of them is present alone at certain time instant and can be removed by binary masking.

## III. TF-CUBSS Algorithm

In order to solve the UBSS problem in the convolutive case, we propose to identify first the impulse response of the channels (see the algorithm's diagram in Fig. 1). This problem in overdetermined case is very difficult and becomes almost impossible in the underdetermined case without side information on the considered sources. In this paper and similarly to [19], we exploit the sparseness property of the audio sources by assuming that from time to time only one source is present. In other words, we consider the following assumption.

*Assumption 3:* There exists, periodically, time intervals where only one source is present in the mixture. This occurs for all source signals of the considered mixtures (see Fig. 2).

To detect these time intervals, we propose to use information-criteria-based testing for the estimation of the number of sources present in the signal (see Section III-A for more details).

### A. Channel Estimation

Based on Assumption 3, we propose here to apply single-input multiple-output (SIMO)-based techniques to blindly estimate the channel impulse response. Regarding the problem at hand, we have to solve three different problems: first, we have to select time intervals where only one source signal is effectively present; then, for each selected time interval one should apply an appropriate blind SIMO identification technique to estimate the channel parameters; finally, the way we proceed, the same channel may be estimated several times and hence one has to group together (cluster) the channel estimates into $N$ classes corresponding to the $N$ source channels.

*1) Source Number Estimation:* Let define the spatio-temporal vector

$$\mathbf{x}_d(t) = [\mathbf{x}^T(t), \ldots, \mathbf{x}^T(t-d+1)]^T = \sum_{k=1}^{N} \mathbf{H}_k \mathbf{s}_k(t) + \boldsymbol{\eta}_d(t) \tag{3}$$

where $\mathbf{H}_k$ are block-Sylvester matrices of size $dM \times (d+K)$

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{h}_k(0) & \cdots & \mathbf{h}_k(K) & & \mathbf{0} \\ & \ddots & & \ddots & \\ \mathbf{0} & & \mathbf{h}_k(0) & \cdots & \mathbf{h}_k(K) \end{bmatrix}$$

$\mathbf{s}_k(t) \stackrel{\text{def}}{=} [s_k(t), \ldots, s_k(t-K-d+1)]^T$, and $d$ is a chosen processing window size. Under the no-common zeros assumption (Assumption 1) and for large window sizes (see [20] for more details), matrices $\mathbf{H}_k$ are full column rank.

Hence, in the noiseless case, the rank of the data covariance matrix $\mathbf{R} \stackrel{\text{def}}{=} E[\mathbf{x}_d(t)\mathbf{x}_d^H(t)]$ is equal to $\min(p(d+K), dM)$, where $p$ is the number of sources present in the considered time interval over which the covariance matrix is estimated. In particular, for $p=1$, one has the minimum rank value equal to $(d+K)$.

Therefore, our approach consists in estimating the rank of the sample averaged covariance matrix $\mathbf{R}$ over several time slots (intervals) and select those corresponding to the smallest rank value $r = d+K$.

In the case where $p$ sources are active (present) in the considered time slot, the rank would be $r = p(d+K)$ and hence $p$ can be estimated by the closest integer value to $(r/d+K)$.

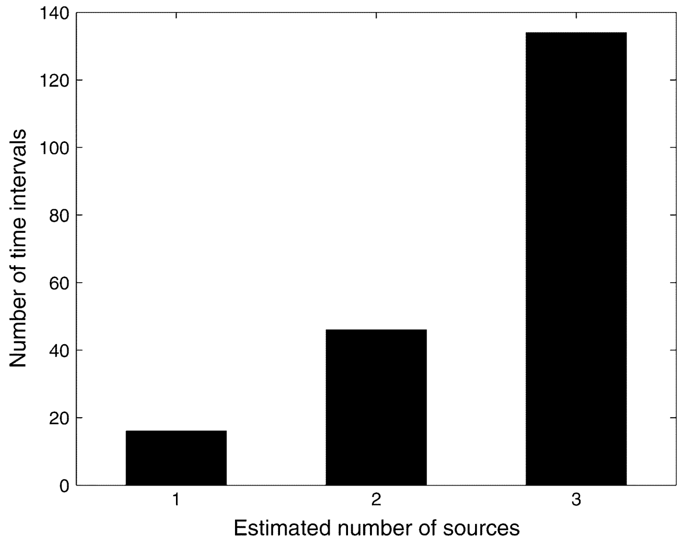


Fig. 3. Histogram representing the number of time intervals for each estimated number of sources for four audio sources and three sensors in convolutive mixture case.

The estimation of the rank value is done here by Akaike's criterion [20] according to

$$r = \arg\min_k \left[ -2\log\left( \frac{\prod_{i=k+1}^{Md} \lambda_i^{1/(Md-k)}}{\frac{1}{Md-k}\sum_{i=k+1}^{Md}\lambda_i} \right)^{(Md-k)T_s} + 2k(2Md-k) \right] \quad (4)$$

where $\lambda_1 \geq \cdots \geq \lambda_{Md}$ represent the eigenvalues of $\mathbf{R}$, and $T_s$ is the time slot size. This criterion represents the maximum-likelihood estimate of the system parameters (given here by the signal eigenvectors and eigenvalues of the covariance matrix) penalized by the number of free adjusted parameters under the asymptotic Gaussian distribution of the latter (see [20] for more details).

Note that it is not necessary at this stage, to know exactly the channel degree $K$ as long as $d > K$ (i.e., an over-estimation of the channel degree is sufficient) in which case the presence of one signal source is characterized by

$$d < r < 2d.$$

Fig. 3 illustrates the effectiveness of the proposed method where a recording of 6 s of $M = 3$ convolutive mixtures of $N = 4$ sources is considered. The sampling frequency is 8 kHz, and the time slot size is $T_s = 200$ samples. The sources consist of three speech signals corresponding to two men and one woman plus a guitar signal. The convolutive channel is of order $K = 6$ and its coefficients are generated randomly using Gaussian law. One can observe that the case $p = 1$ (one signal source) occurs approximatively 10% of the time in the considered context.

*2) Blind Channel Identification:* To perform the blind channel identification, we have used in this paper the cross-relation (CR) technique described in [21] and [22]. Consider a time interval where we have only the source $s_i$ present. In this case, we can consider a SIMO system of $M$ outputs given by

$$\mathbf{x}(t) = \sum_{k=0}^{K} \mathbf{h}_i(k)s_i(t-k) + \boldsymbol{\eta}(t) \quad (5)$$

where $\mathbf{h}_i(k) = [h_{i1}(k)\ldots h_{iM}(k)]^T, k = 0,\ldots,K$. From (5), the noise-free outputs $x_j(k), 1 \leq j \leq M$ are given by

$$x_j(k) = h_{ij}(k) * s_i(k), \quad 1 \leq j \leq M \quad (6)$$

where "$*$" denotes the convolution. Using commutativity of convolution, it follows

$$h_{il}(k) * x_j(k) = h_{ij}(k) * x_l(k), \quad 1 \leq j \neq l \leq M. \quad (7)$$

This is a linear equation satisfied by every pair of channels. It was shown that reciprocally, the previous $M(M-1)/2$ cross-relations characterize uniquely the channel parameters. We have the following theorem [21].

*Theorem 1:* Under the no-common zeros assumption (Assumption 1), the set of cross-relations (in the noise-free case)

$$x_l(k) * h'_j(k) - x_j(k) * h'_l(k) = 0, \quad 1 \leq l < j \leq M \quad (8)$$

where $\mathbf{h}'(z) = [h'_1(z)\ldots h'_M(z)]^T$ is a $M \times 1$ polynomial vector of degree $K$, is satisfied if and only if $\mathbf{h}'(z) = \alpha\mathbf{h}_i(z)$ for a given scalar constant $\alpha$.

By collecting all possible pairs of $M$ channels, one can easily establish a set of linear equations. In matrix form, this set of equations can be expressed as

$$\boldsymbol{\mathcal{X}}_M\mathbf{h}_i = \mathbf{0} \quad (9)$$

where $\mathbf{h}_i \stackrel{\text{def}}{=} [h_{i1}(0)\ldots h_{i1}(K),\ldots,h_{iM}(0)\ldots h_{iM}(K)]^T$ and $\boldsymbol{\mathcal{X}}_M$ is defined by

$$\boldsymbol{\mathcal{X}}_2 = [\boldsymbol{\mathcal{X}}_{(2)}, \quad -\boldsymbol{\mathcal{X}}_{(1)}]$$

$$\boldsymbol{\mathcal{X}}_n = \left[\begin{array}{ccc|c} & \boldsymbol{\mathcal{X}}_{n-1} & & \mathbf{0} \\ \hline \boldsymbol{\mathcal{X}}_{(n)} & & \mathbf{0} & -\boldsymbol{\mathcal{X}}_{(1)} \\ & \ddots & & \vdots \\ \mathbf{0} & & \boldsymbol{\mathcal{X}}_{(n)} & -\boldsymbol{\mathcal{X}}_{(n-1)} \end{array}\right] \quad (10)$$

with $n = 3,\ldots,M$ and

$$\boldsymbol{\mathcal{X}}_{(n)} = \begin{bmatrix} x_n(K) & \ldots & x_n(0) \\ \vdots & & \vdots \\ x_n(T_s-1) & \ldots & x_n(T_s-K-1) \end{bmatrix}. \quad (11)$$

In the presence of noise, (9) can be naturally solved in the least-squares (LS) sense according to

$$\hat{\mathbf{h}}_{\text{CR}} = \arg\min_{\|\mathbf{h}_i\|=1} \mathbf{h}_i^H\boldsymbol{\mathcal{X}}_M^H\boldsymbol{\mathcal{X}}_M\mathbf{h}_i \quad (12)$$

which solution is given by the least unit-norm eigenvector of matrix $\boldsymbol{\mathcal{X}}_M^H\boldsymbol{\mathcal{X}}_M$. It is shown in [21] that the noise term in the quadratic form (12) has a mean value proportional to the identity

TABLE I
CLUSTER-BASED TF-CUBSS ALGORITHM USING STFT

| |
|---|
| 1) Channel estimation; AIC criterion [20] to detect the number of source, then application of the blind identification algorithm in [21], [22] followed by vector clustering. |
| 2) Mixture STFT computation by (15) and noise thresholding by (16) |
| 3) Vector clustering by (17) and (18). |
| 4) Source STFT estimation by (19). |
| 5) Source TF synthesis by [27]. |

matrix. Consequently, the channel estimates remains unbiased under white additional noise assumption.

*Remark:* We have presented here a basic version of the CR method. In [23], an improved version of the method (introduced in the adaptive scheme) is proposed exploiting the quasi-sparse nature of acoustic impulse responses. Other channel estimation techniques in the overcomplete case, e.g., [24], can be used as well at this stage.

*3) Clustering of Channel Vector Estimates:* The first step of our channel estimation method consists in detecting the time slots where only one single source signal is "effectively" present. However, the same source signal $s_i$ may be present in several time intervals (see Figs. 2 and 3) leading to several estimates of the same channel vector $\mathbf{h}_i$.

We end up, finally, with several estimates of each source channel that we need to group together into $N$ classes. This is done by clustering the estimated vectors using $k$-means algorithm [25]. The $i$th channel estimate is evaluated as the centroid of the $i$th class.

### B. UBSS Algorithm With TF-Disjoint Assumption

As we have seen before, the STFT is often used for speech/audio signals because of its low computational cost. Therefore, in this section we propose a new *cluster-based TF-CUBSS algorithm* using the STFT for convolutive mixture case. Note that the STFT is a particular form of wavelet transforms which have been used in [26] for the UBSS of image signals.

After transformation into the TF domain using the STFT, the model in (1) becomes (in the noiseless case)

$$\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) = \mathbf{H}(f)\boldsymbol{\mathcal{S}}_{\mathbf{s}}(t,f) \tag{13}$$

where $\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f)$ is the mixture STFT vector, $\boldsymbol{\mathcal{S}}_{\mathbf{s}}(t,f)$ is the source STFT vector, and $\mathbf{H}(f) = [\mathbf{h}_1(f)\ldots\mathbf{h}_N(f)]$ is the channel Fourier transform matrix. Under the assumption that all sources are disjoint in the TF domain, (13) reduces to

$$\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) = \mathbf{h}_i(f)\boldsymbol{\mathcal{S}}_{s_i}(t,f), \quad \forall (t,f) \in \Omega_i, \forall i \in \boldsymbol{\mathcal{N}} \tag{14}$$

where $\boldsymbol{\mathcal{N}} = \{1,\ldots,N\}$ and $\Omega_i$ is the TF support of the $i$th source.

Consequently, two TF points $(t_1,f_1)$ and $(t_2,f_2)$ belonging to the same region $\Omega_i$ (i.e., corresponding to the source signal $s_i$) are "associated" with the same channel $\mathbf{h}_i$. It is this observation that is used to derive the separation algorithm summarized in Table I and detailed next.

First, we compute the STFT of the mixtures $\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f)$ by applying (2) for each of the mixture in $\mathbf{x}(t)$, as follows:

$$\boldsymbol{\mathcal{S}}_{x_i}(t,f) = \sum_{m=-(L-1)/2}^{m=(L-1)/2-1} w(t-m)x_i(m)e^{-j2\pi fm}, \quad i = 1,\ldots,M \tag{15a}$$

$$\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) = [\boldsymbol{\mathcal{S}}_{x_1}(t,f),\ldots,\boldsymbol{\mathcal{S}}_{x_M}(t,f)]^T. \tag{15b}$$

where $w(t)$ is a chosen window (in our simulations we chose Hamming window) of length $L$.

Then, we apply a noise thresholding procedure which mitigates the noise effect and reduces the computational cost as only the selected TF points are further treated by our algorithm. In particular, for each frequency $f_0$, we apply the following criterion for all the time points $t_k$ belonging to the frequency slice $(t,f_0)$

$$\text{If} \quad \frac{\|\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t_k,f_0)\|}{\max_t\{\|\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f_0)\|\}} > \epsilon_1, \quad \text{then keep } (t_k,f_0) \tag{16}$$

where $\epsilon_1$ is a small threshold (typically, $\epsilon_1 = 0.01$). Then, the set of all selected points $\Omega$ is expressed by $\Omega = \bigcup_{i=1}^{N}\Omega_i$, where $\Omega_i$ is the TF support of the source $s_i(t)$. Note that, the effects of spreading the noise energy while localizing the source energy in the time–frequency domain amounts to increasing the robustness of the proposed method with respect to noise (see [18, Part IV]). Hence, by (16), we would keep only time–frequency points where the signal energy is non-negligible, the other time–frequency points are rejected, i.e., not further processed, since considered to represent noise contribution only. Also, due to the noise energy spreading, the contribution of the noise in the source time–frequency points is relatively negligible, at least for moderate and high signal-to-noise ratios (SNRs).

On the other hand, note that the noise thresholding as well as TF masking induce nonlinear distortion in the reconstructed signal. Now, how this distortion affects the source estimates is an open problem that still raises many questioning and research works including those which try to mitigate this distortion in the TF domain, e.g., [28].

After noise thresholding, the clustering procedure can be done as follows: For each TF point, we obtain the spatial direction vectors by

$$\mathbf{v}(t,f) = \frac{\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f)}{\|\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f)}\|, \quad (t,f) \in \Omega \tag{17}$$

and force them, without loss of generality, to have the first entry real and positive.

Next, we cluster these vectors into $N$ classes $\{C_i \,|\, i \in \boldsymbol{\mathcal{N}}\}$ by minimizing the criterion

$$\mathbf{v}(t,f) \in C_i \Longleftrightarrow i = \arg\min_k \left\| \mathbf{v}(t,f) - \frac{\hat{\mathbf{h}}_k(f)e^{-j\theta_k}}{\|\hat{\mathbf{h}}_k(f)\|} \right\|^2 \tag{18}$$

where $\hat{\mathbf{h}}_k(f)$ is the Fourier transform of the $k$th channel vector estimate [given by (12) and the proposed clustering procedure]

TABLE II
SUBSPACE-BASED TF-CUBSS ALGORITHM USING STFT

| |
|---|
| 1) Channel estimation; AIC criterion [20] to detect the number of source, then application of the blind identification algorithm in [21], [22] followed by vector clustering. |
| 2) STFT computation and noise thresholding. |
| 3) For all selected TF points, detect the active sources by (22) and (25). |
| 4) Source STFT estimation by (26). |
| 5) Source TF synthesis by [27]. |

and $\theta_k$ is the phase argument of $\hat{h}_{k1}(f)$ (this is to force the first entry to be real positive). The collection of all points, whose vectors belong to the class $C_i$, now forms the TF support $\Omega_i$ of the source $s_i(t)$.

Therefore, we can estimate the STFT of each source $s_i(t)$ by

$$\hat{\boldsymbol{\mathcal{S}}}_{s_i}(t,f) = \begin{cases} \frac{\hat{\mathbf{h}}_i^H(f)}{\|\hat{\mathbf{h}}_i(f)\|^2}\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f), & \forall (t,f) \in \Omega_i \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

since, from (14), we have

$$\begin{aligned} \frac{\hat{\mathbf{h}}_i^H(f)}{\|\hat{\mathbf{h}}_i(f)\|^2}\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) &= \frac{\hat{\mathbf{h}}_i^H(f)\mathbf{h}_i(f)}{\|\hat{\mathbf{h}}_i(f)\|^2}\boldsymbol{\mathcal{S}}_{s_i}(t,f) \\ &\approx \boldsymbol{\mathcal{S}}_{s_i}(t,f), \quad \forall (t,f) \in \Omega_i. \end{aligned}$$

### *C. UBSS Algorithm With TF-Nondisjoint Assumption*

We have seen the cluster-based TF-CUBSS methods, using the STFT, as summarized in Table I. This method relies on the assumption that the sources were TF-disjoint, which led to the TF-transformed structure in (14). The latter is no longer valid, when the sources are nondisjoint in the TF domain.

Under the TF-nondisjoint condition, stated in Assumption 2, we propose in this section an alternative method using subspace projection.

Recall that the first two steps of the cluster-based TF-CUBSS algorithm do not rely on the assumption of TF-disjoint sources (see Table I). Therefore, we can reuse these steps to obtain the channel estimation and all the TF points of $\Omega$. Under the TF-nondisjoint condition, consider a TF point $(t,f) \in \Omega$ at which there are $\mathcal{J}$ sources $s_{\alpha_1}(t),\ldots,s_{\alpha_{\mathcal{J}}}(t)$ present, with $\mathcal{J} < M$ where $\alpha_1,\ldots,\alpha_{\mathcal{J}} \in \boldsymbol{\mathcal{N}}$ denote the indices of the sources present at $(t,f)$. Our goal is to identify the sources that are present at $(t,f)$, i.e., $\alpha_1,\ldots,\alpha_{\mathcal{J}}$, and to estimate the STFT of each of these contributing sources.

We define the following:

$$\tilde{\mathbf{s}} = \left[s_{\alpha_1}(t),\ldots,s_{\alpha_{\mathcal{J}}}(t)\right]^T \tag{20a}$$

$$\tilde{\mathbf{H}}_\alpha(f) = \left[\mathbf{h}_{\alpha_1}(f),\ldots,\mathbf{h}_{\alpha_{\mathcal{J}}}(f)\right]. \tag{20b}$$

Then, (13) is reduced to the following:

$$\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) = \tilde{\mathbf{H}}_\alpha(f)\boldsymbol{\mathcal{S}}_{\tilde{\mathbf{s}}}(t,f). \tag{21}$$

Let $\tilde{\mathbf{H}}_\beta(f) = [\mathbf{h}_{\beta_1}(f),\ldots,\mathbf{h}_{\beta_{\mathcal{J}}}(f)]$ and $\mathbf{Q}_\beta(f)$ be the orthogonal projection matrix onto the noise subspace of $\tilde{\mathbf{H}}_\beta(f)$ expressed by

$$\mathbf{Q}_\beta(f) = \mathbf{I} - \tilde{\mathbf{H}}_\beta(f)\left(\tilde{\mathbf{H}}_\beta^H(f)\tilde{\mathbf{H}}_\beta(f)\right)^{-1}\tilde{\mathbf{H}}_\beta^H(f). \tag{22}$$

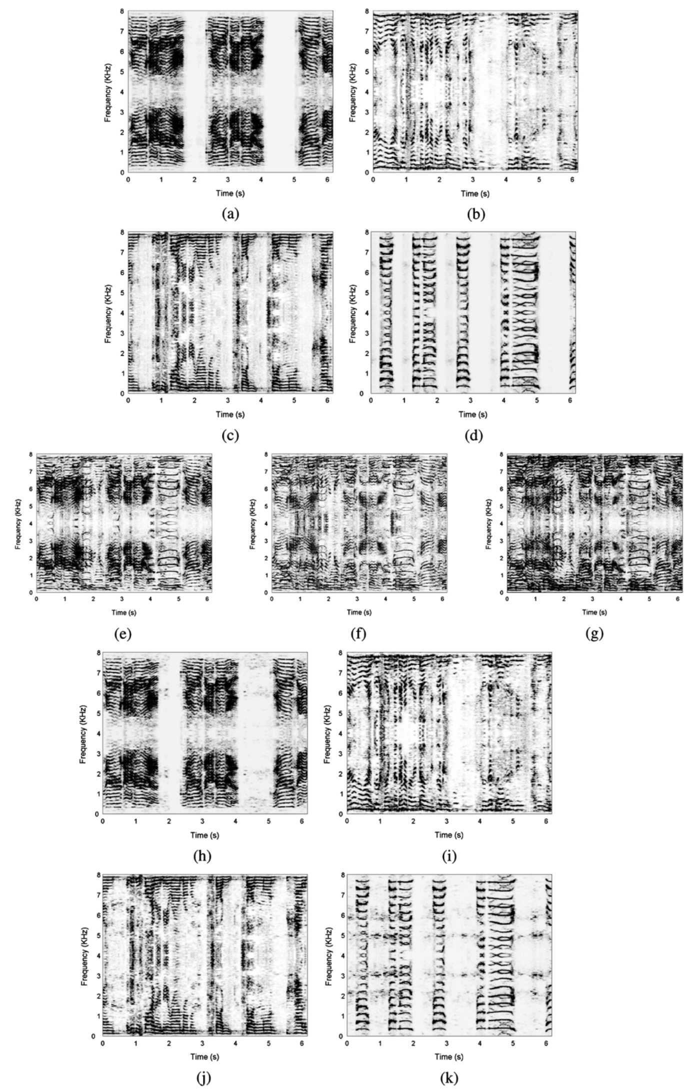


Fig. 4. Simulated example (viewed in TF domain) for the subspace-based TF-CUBSS algorithm in the case of four speech sources and three sensors. The top four plots represent the original source signals, the middle three plots represent the three mixtures, and the bottom four plots represent the source estimates. (a) $S_{s_1}(t,f)$. (b) $S_{s_2}(t,f)$. (c) $S_{s_3}(t,f)$. (d) $S_{s_4}(t,f)$. (e) $S_{x_1}(t,f)$. (f) $S_{x_2}(t,f)$. (g) $S_{x_3}(t,f)$. (h) $S_{\hat{s}_1}(t,f)$. (i) $S_{\hat{s}_2}(t,f)$. (j) $S_{\hat{s}_3}(t,f)$. (k) $S_{\hat{s}_4}(t,f)$.

We have the following observation:

$$\begin{cases} \mathbf{Q}_\beta(f)\mathbf{h}_i(f) = 0, & i \in \{\beta_1,\ldots,\beta_{\mathcal{J}}\} \\ \mathbf{Q}_\beta(f)\mathbf{h}_i(f) \neq 0, & i \in \boldsymbol{\mathcal{N}}\backslash\{\beta_1,\ldots,\beta_{\mathcal{J}}\}. \end{cases} \tag{23}$$

Consequently, as $\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) \in \mathrm{Range}\{\tilde{\mathbf{H}}_\alpha(f)\}$, we have

$$\begin{cases} \mathbf{Q}_\beta(f)\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) = 0, & \text{if } \{\beta_1,\ldots,\beta_{\mathcal{J}}\} = \{\alpha_1,\ldots,\alpha_{\mathcal{J}}\} \\ \mathbf{Q}_\beta(f)\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) \neq 0, & \text{otherwise.} \end{cases} \tag{24}$$

If $\mathbf{H}(f)$ has already been estimated by the method presented in Section III-A, then this observation gives us the criterion to detect the indices $\alpha_1,\ldots,\alpha_{\mathcal{J}}$; and hence, the contributing

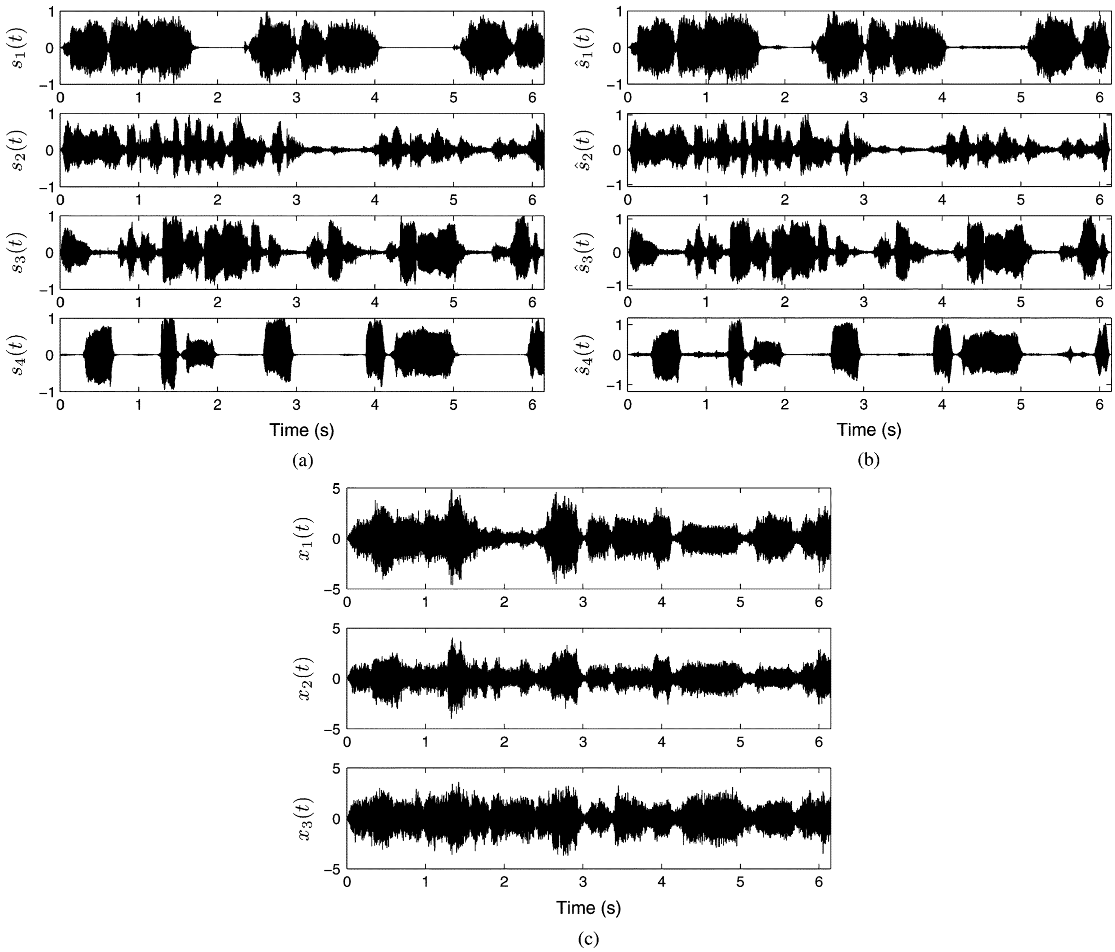


Fig. 5. Simulated example (viewed in time domain) for the subspace-based TF-CUBSS algorithm in the case of four speech sources and three sensors. (a) Original signals. (b) Estimated signals. (c) Mixture signals.

sources at the considered TF point. In practice, to take into account noise, one detects the column vectors of $\tilde{\mathbf{H}}_\alpha(f)$ by minimizing

$$\{\alpha_1, \dots, \alpha_{\mathcal{J}}\} = \arg \min_{\beta_1,\dots,\beta_{\mathcal{J}}} \{\|\mathbf{Q}_\beta(f)\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f)\|\}. \tag{25}$$

Next, TFD values of the $\mathcal{J}$ sources at TF point $(t,f)$ are estimated by

$$\hat{\boldsymbol{\mathcal{S}}}_{\tilde{\mathbf{s}}}(t,f) \approx \tilde{\mathbf{H}}_\alpha^{\#}(f)\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f) \tag{26}$$

where the superscript $(^{\#})$ represents the Moore–Penrose's pseudo-inversion operator.

In the simulation, the optimization problem of (25) is solved using exhaustive search. This is computationally tractable for small array sizes but would be prohibitive if $M$ is very large.

Table II provides a summary of the subspace projection based TF-CUBSS algorithm using STFT.

## IV. Discussion

We discuss here certain points relative to the proposed TF-CUBSS algorithms and their applications.

*1) Number of Sources:* The number of sources $N$ is assumed known in the clustering method that we have used. However, there exist clustering methods [25] which perform the class estimation as well as the estimation of the number $N$. In our simulation, we have observed that most of the time the number of classes is overestimated, leading to poor source separation quality. Hence, robust estimation of the number of sources in the UBSS case remains a difficult open problem that deserves particular attention in future works.

*2) Number of Overlapping Sources:* In the subspace-based approach, it is also possible to consider a fixed (maximum) value of $\mathcal{J}$ that is used for all TF points. Indeed, if the number of overlapping sources is less than $\mathcal{J}$, we would estimate close-to-zero source STFT values. For example, if we assume $\mathcal{J} = 2$ sources are present at a given TF point while only one source is effectively contributing, then we estimate one close-to-zero source STFT value. This approach increases slightly the estimation error of the source signals (especially at low SNRs) but has the advantage of simplicity compared to using information theoretic-based criteria for estimating the value of $\mathcal{J}$.

*3) Separation Quality Versus Number of Sources:* Although we are in the underdetermined case, the number of sources $N$ should not exceed too much the number of sensors. Indeed, when $N$ increases, the level of source interference increases, and hence, the source quasi-disjointness assumption is ill-satisfied. Moreover, for a large number of sources, the likelihood of having two closely spaced sources, i.e., such that the spatial directions $\mathbf{h}_i$ and $\mathbf{h}_j$ are "close" to linear dependency, increases. In that case, vector clustering performance degrades

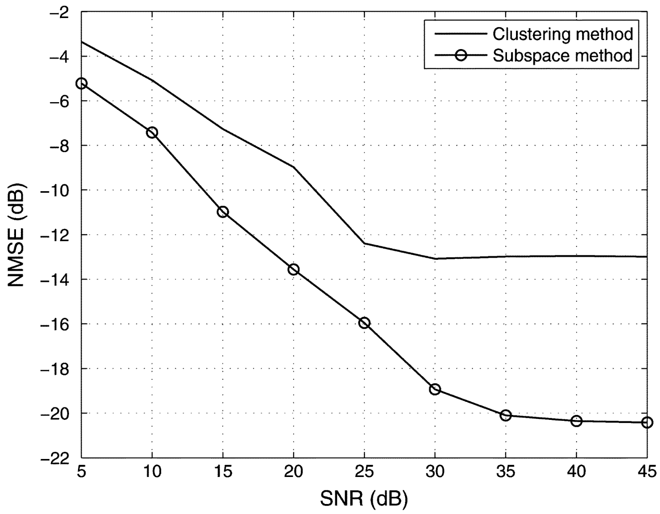


Fig. 6. Comparison between subspace-based and cluster-based TF-CUBSS algorithms: normalized MSE (NMSE) versus SNR for four speech sources and three sensors.

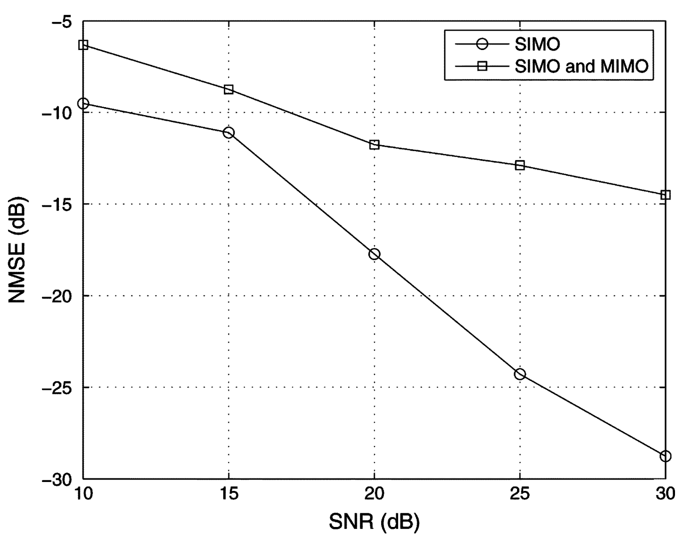


Fig. 7. NMSE versus SNR for four audio sources and three sensors in convolutive mixture case : Comparison of the performance of identification algorithm using only SIMO system and the algorithm using SIMO and MIMO system.

significantly. In brief, sparseness and spatial separation are the two limiting factors against increasing the number of sources.

*4) Overdetermined Case:* Our algorithm can be further simplified in the overdetermined case where $M \geq N$. In that context, the algorithm can be reduced to the channel estimation step, the STFT computation and noise thresholding then source STFT estimation using the channel matrix pseudo-inversion at each frequency

$$\hat{\boldsymbol{\mathcal{S}}}_{\mathbf{s}}(t,f) = \hat{\mathbf{H}}^{\#}(f)\boldsymbol{\mathcal{S}}_{\mathbf{x}}(t,f).$$

## V. Simulation Results

In the simulations, we have considered an array of $M = 3$ sensors, that receives signals from $N = 4$ independent audio sources (three speech signals corresponding to two men and

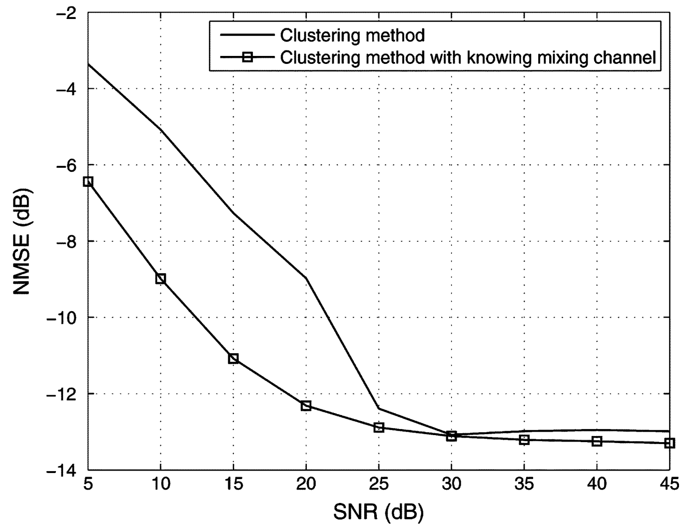


Fig. 8. Comparison, for the cluster-based TF-CUBSS algorithm, when the mixing channel $\mathbf{H}$ is known or unknown: NMSE of the source estimates.

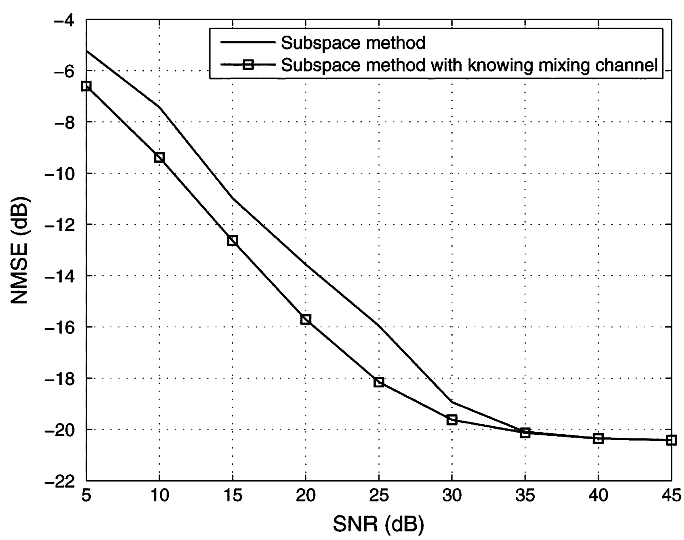


Fig. 9. Comparison, for the subspace-based TF-CUBSS algorithm, when the mixing channel $\mathbf{H}$ is known or unknown: NMSE of the source estimates.

one woman plus a guitar signal). The filter coefficients are chosen randomly and the channel order is $K = 6$. The sample size is $T = 8192$ samples (corresponding approximately to 1-s recording of speech signals sampled at 8 kHz). The separation quality is measured by the normalized mean squares estimation errors (NMSE) of the sources evaluated over $N_r = 200$ Monte Carlo runs and defined as

$$\mathrm{NMSE}_i \stackrel{\mathrm{def}}{=} \frac{1}{N_r}\sum_{r=1}^{N_r} \min_{\alpha}\left(\frac{\|\alpha\hat{\mathbf{s}}_{i,r} - \mathbf{s}_i\|^2}{\|\mathbf{s}_i\|^2}\right) \tag{27}$$

$$\mathrm{NMSE}_i = \frac{1}{N_r}\sum_{r=1}^{N_r} 1 - \left(\frac{\hat{\mathbf{s}}_{i,r}\mathbf{s}_i^H}{\|\hat{\mathbf{s}}_{i,r}\|\|\mathbf{s}_i\|}\right)^2 \tag{28}$$

$$\mathrm{NMSE} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{NMSE}_i \tag{29}$$

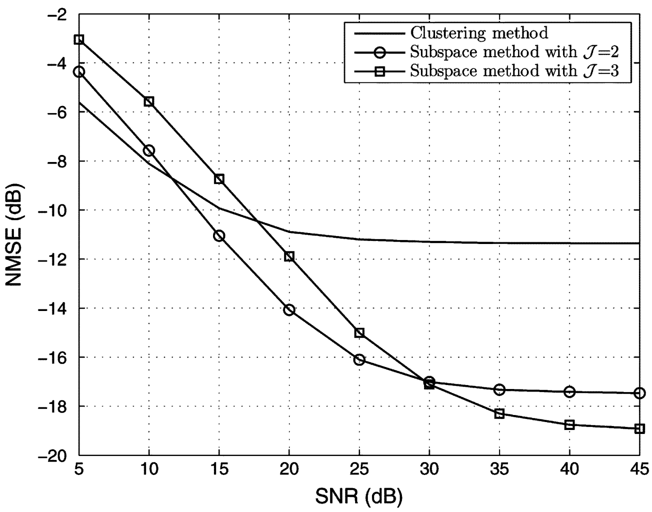


Fig. 10. Comparison between subspace-based and cluster-based TF-CUBSS algorithms: NMSE of the source estimates for different ranks of the projection subspace, for the case of five sources and four sensors.

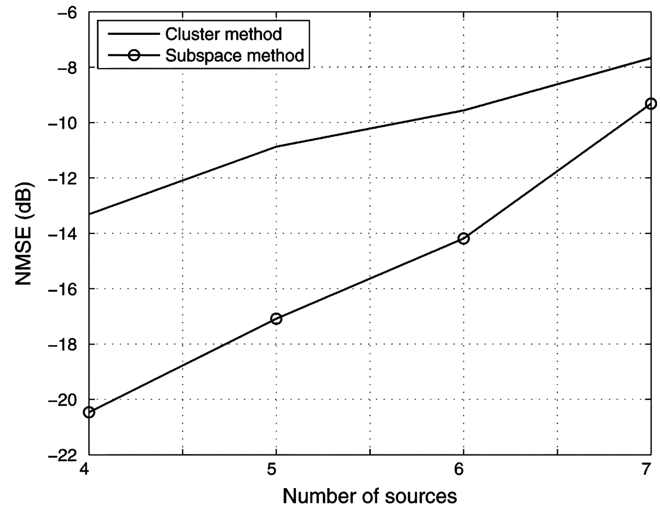


Fig. 11. Comparison between subspace-based and cluster-based TF-CUBSS algorithms: NMSE versus number of sources.

where $\mathbf{s}_i \stackrel{\text{def}}{=} [s_i(0), \ldots, s_i(T-1)]$, $\hat{\mathbf{s}}_{i,r}$ (defined similarly) represents the $r$th estimate of source $\mathbf{s}_i$ and $\alpha$ is a scalar factor that compensate for the scale indeterminacy of the BSS problem.

In Fig. 4, the top four plots represent the TF representation of the original source signals, the middle three plots represent the TF representation of the $M$ mixture signals and the bottom four plots represent the TF representation of the source estimates by the subspace-based algorithm (Table II) using STFT of length 1024. Fig. 5 represents the same disposition of signals but in the time domain.

In Fig. 6, we compare the separation performance obtained by the subspace-based algorithm with $\mathcal{J} = 2$ and the cluster-based algorithm (Table I). It is observed that subspace-based algorithm provides much better separation results than those obtained by the cluster-based algorithm. This is mainly due to the high occurrence of overlapping sources in the TF domain for this type of signals so that the "TF-disjointness" assumption used by the TF-CUBSS algorithm is poorly satisfied. This can be observed also from Fig. 4, where we can see that the TFD supports of the four audio sources are clearly overlapping.

In Fig. 7, we present the performance of channel identification obtained by using SIMO identification algorithm [in this case, we choose only the time intervals where only one source is present using Akaike Information Criterion (AIC) criterion] with SIMO and multiple-input multiple output (MIMO) identification algorithms[2] (in this case, we choose the time intervals where we are in the overdetermined case, i.e., where $p = 1$ or $p = 2$). It is observed that SIMO-based identification provides better results than those obtained by SIMO and MIMO identification algorithms. Indeed, the advantage of considering overdetermined MIMO system identification resides in the fact that the occurrence of MIMO (i.e., number of time intervals where this situation occurs as shown in Fig. 3) is much higher than that of SIMO case. However, as we observe it, this does not compensate for the higher estimation error of MIMO systems compared to SIMO systems.

[2]For the identification of MIMO system, we have used the subspace method [29] for the equalization step followed by second-order blind identification (SOBI) algorithm [3] for the separation step.

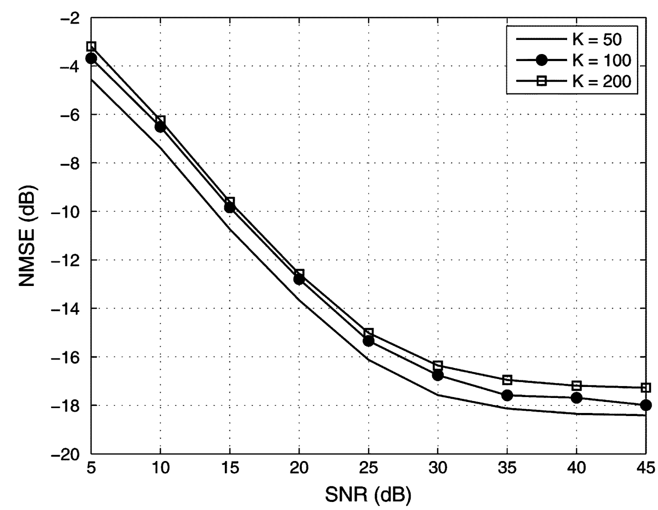


Fig. 12. NMSE versus SNR for four audio sources and three sensors: Comparison, for the subspace-based TF-CUBSS algorithm, for different filter size $K$.

The plot in Fig. 8 (respectively, in Fig. 9) presents the separation performance when using the exact matrix $\mathbf{H}$ compared to that obtained with the proposed estimate $\hat{\mathbf{H}}$ using the cluster-based method (respectively, the subspace-based method). The observed performance loss is due to the channel estimation error which is relatively high for low SNRs and becomes negligible for high SNRs.

In Fig. 10, we compare the performance obtained with the subspace-based method for $\mathcal{J} = 2$ and $\mathcal{J} = 3$. In that experiment, we have used $M = 4$ sensors and $N = 5$ source signals. One can observe that, for high SNRs, the case of $\mathcal{J} = 3$ leads to a better separation performance than for the case of $\mathcal{J} = 2$. However, for low SNRs, a large value of $\mathcal{J}$ increases the estimation noise (as mentioned in Section IV) and hence degrades the separation quality.

Fig. 11 illustrates the rapid degradation of the separation quality when we increase the number of sources from $N = 4$ to $N = 7$. This confirms the remarks made in Section IV.

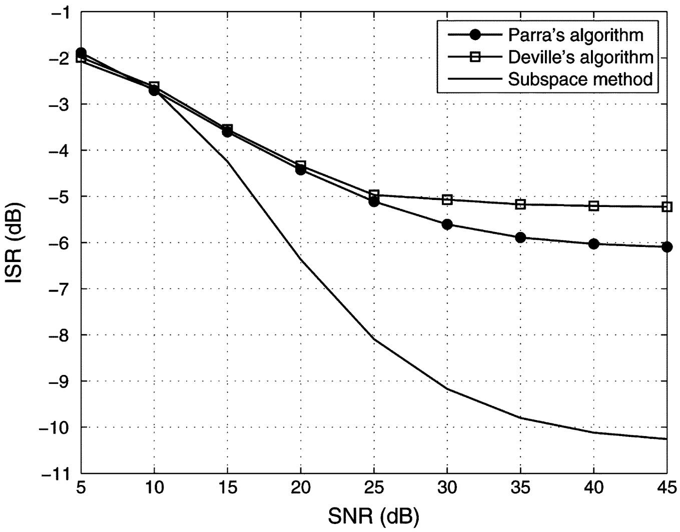


Fig. 13. ISR versus SNR for two audio sources and two sensors: Comparison between the subspace-based TF-CUBSS algorithm, Parra's algorithm, and Deville's algorithm.

Fig. 12 illustrates the algorithm's performance when we consider long impulse response channels. More specifically, the plots represent the separation performance for channels of length 50, 100, and 200, respectively. The channel taps are generated randomly using Gaussian law. We observe a slight performance degradation when the channel order increases but the separation quality remains quite good.

In Fig. 13, we compare the separation performance of our algorithm, Deville's algorithm in [30], and Parra's algorithm in [31] in the overdetermined case of two sensors and two speech signals of one man and one woman (selected among the four previous sources). The algorithms in [30] and [31] separate the sources only up to an unknown filter and, hence, we use in this experiment the *interference-to-signal ratio* (ISR) criterion defined in [31] instead of the NMSE. We observe a significant performance gain in favor of the proposed method especially at high SNR values. Moreover, our method has the following advantages: 1) it can treat the underdetermined case; 2) it estimates the sources up to a constant not to an unknown filter like in [30] and [31]; and 3) the proposed frame selection procedure does not involve any thresholding (the choice of an appropriate threshold value is a difficult problem, as it is strongly dependent on the context) or ad hoc selection of frequency range like in [30].

## VI. Conclusion

This paper introduces new methods for the UBSS of TF-disjoint and TF-nondisjoint nonstationary sources in the convolutive mixture case using their time–frequency representations. The first proposed method has the advantage of simplicity, while the second uses a weaker assumption on the source "sparseness," i.e., the sources are not necessarily TF-disjoint, and proposes an explicit treatment of the overlapping points using subspace projection, leading to significant performance improvements. Simulation results illustrate the effectiveness of our algorithms in different scenarios.

## References

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2003.
[2] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.
[3] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
[4] A. Belouchrani and M. G. Amin, "Blind source separation based on time–frequency signal representations," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 2888–2897, Nov. 1998.
[5] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source separation using second order cyclostationary statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 694–701, Apr. 2001.
[6] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," in *Signal Process.*, Nov. 2001, vol. 81, no. 11, pp. 2353–2362.
[7] N. Linh-Trung, A. Belouchrani, K. Abed-Meraim, and B. Boashash, "Separating more sources than sensors using time—Frequency distributions," *EURASIP J. Appl. Signal Process. (also in Proc. ISSPA 2001)*, vol. 2005, no. 17, pp. 2828–2847, 2005.
[8] N. Linh-Trung, A. Aïssa-El-Bey, K. Abed-Meraim, and A. Belouchrani, "Underdetermined blind source separation of non-disjoint nonstationary sources in time–frequency domain," in *Proc. 8th Int. Symp. Signal Process. Applicat. (ISSPA)*, Sydney, Australia, Aug. 2005, vol. 1, pp. 46–49.
[9] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, "Blind separation of audio sources using modal decomposition," in *Proc. 8th Int. Symp. Signal Process. Applicat. (ISSPA)*, Sydney, Australia, Aug. 2005, pp. 451–454.
[10] F. Abrard and Y. Deville, "A time–frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," in *Signal Process.*, Jul. 2005, vol. 85, no. 7, pp. 1389–1403.
[11] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time–frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.
[12] J. Rosca, C. Borss, and R. Balan, "Generalized sparse signal mixing model and application to noisy blind source separation," in *Proc. ICASSP*, 2004, vol. 3, pp. 877–880.
[13] P. O'Grady, B. Pearlmutter, and S. Rickard, "Survey of sparse and non-sparse methods in source separation," *Int. J. Imaging Syst. Tech.*, vol. 15, no. 1, pp. 18–33, 2005.
[14] S. Araki, H. Sawada, R. Mukai, and S. Makino, "A novel blind source separation method with observation vector clustering," in *Proc. IWAENC*, 2005, pp. 117–120.
[15] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, "Blind separation of more speech than sensors with less distortion by combining sparseness and ICA," in *Proc. IWAENC*, Kyoto, Japan, Sep. 2003, pp. 271–274.
[16] S. Araki, S. Makino, H. Sawada, and R. Mukai, "Underdetermined blind separation of convolutive mixtures of speech with directivity pattern based mask and ICA," in *Proc. Int. Conf. Ind. Compon. Anal. Blind Source Separation ICA*, Sep. 2004, pp. 898–905.
[17] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, "Underdetermined blind separation of convolutive mixtures of speech by combining time–frequency masks and ICA," in *Proc. EUSIPCO*, 2004.
[18] B. Boashash, Ed., *Time Frequency Signal Analysis and Processing: Method and Applications*. Oxford, U.K.: Elsevier, 2003.
[19] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 13, pp. 882–895, Sep. 2005.
[20] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustic, Speech, Signal Process.*, vol. ASSP-33, no. 1, pp. 387–392, Feb. 1985.
[21] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.
[22] A. Aïssa-El-Bey, M. Grebici, K. Abed-Meraim, and A. Belouchrani, "Blind system identification using cross-relation methods: Further results and developments," in *Proc. 8th Int. Symp. Signal Process. Applicat. (ISSPA)*, Paris, France, Jul. 2003, vol. 1, pp. 649–652.
[23] R. Ahmad, A. W. H. Khong, and P. Naylor, "Proportionate frequency domain adaptive algorithms for blind channel identification," in *Proc. ICASSP*, Toulouse, France, May 2006, vol. 5, pp. 29–32.

[24] S. Winter, H. Sawada, S. Araki, and S. Makino, "Overcomplete BSS for convolutive mixtures based on hierarchical clustering," in *Proc. Int. Conf. Ind. Compon. Anal. Blind Source Separation ICA*, Sep. 2004, pp. 652–660.

[25] I. E. Frank and R. Todeschini, *The Data Analysis Handbook*. New York: Elsevier, 1994.

[26] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev, "Blind Source Separation by Sparse Decomposition," in *Independent Component Analysis: Principles and Practice*, S. J. Roberts and R. M. Everson, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[27] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 2, pp. 236–243, Apr. 1984.

[28] J. Rosca, T. Gerkmann, and D. C. Balcan, "Statistical inference of missing speech data in the ICA domain," in *Proc. ICASSP*, May 2006, vol. 5, pp. 617–620.

[29] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 499–511, Mar. 1997.

[30] B. Albouy and Y. Deville, "Alternative structures and power spectrum criteria for blind segmentation and separation of convolutive speech mixtures," in *Proc. ICA*, 2003, pp. 361–366.

[31] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.


**Abdeldjalil Aïssa-El-Bey** was born in Algiers, Algeria, in 1981. He received the State Engineering degree from École Nationale Polytechnique (ENP), Algiers, in 2003 and the M.S. degree in signal processing from Supélec and Paris XI University, Orsay, France, in 2004. Currently, he is working towards the Ph.D. degree in the Signal and Image Processing Department, École Nationale Supérieure des Télécommunications (ENST) Paris, France.

His research interests are blind source separation, blind system identification and equalization, statistical signal processing, wireless communications, and adaptive filtering.

**Karim Abed-Meraim** (SM'04) was born in 1967. He received the State Engineering degree from École Polytechnique, Paris, France, in 1990, the State Engineering degree from École Nationale Supérieure des Télécommunications (ENST), Paris, in 1992, the M.S. degree from Paris XI University, Orsay, France, in 1992, and the Ph.D. degree from École Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1995 (in the field of signal processing and communications).

From 1995 to 1998, he was a Member of Research Staff in the Electrical Engineering Department, University of Melbourne, Melbourne, Australia, where he worked on several research projects related to "blind system identification for wireless communications," "blind source separation," and "array processing for communications," respectively. He is currently an Associate Professor (since 1998) in the Signal and Image Processing Department, ENST. His research interests are in signal processing for communications and include system identification, multiuser detection, space-time coding, adaptive filtering and tracking, array processing, and performance analysis.

Dr. Abed-Meraim is a past Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

**Yves Grenier** (M'81) was born in Ham, Somme, France, in 1950. He received the Ingénieur degree from École Centrale de Paris, Paris, France, in 1972, the Docteur-Ingénieur degree from École Nationale Supérieure des Télécommunications (ENST), Paris, in 1977, and the Doctorat d'État es Sciences Physiques, from the University of Paris-Sud, Paris, in 1984.

He has been with ENST since 1977, as an Assistant Professor, and since 1984 as a Professor. He has been Head of the TSI Department since January 2005. Until 1979, his interests had been in speech recognition, speaker identification, and speaker adaptation of recognition systems. He has then been working on signal modeling, spectral analysis of noisy signals, with applications in speech recognition and synthesis, estimation of nonstationary models, and time–frequency representations. He is currently interested in audio signal processing (acoustic echo cancellation, noise reduction, signal separation, microphone arrays, and loudspeaker arrays).

Prof. Grenier is a member of AES.