Underdetermined Blind Separation of Nondisjoint Sources in the Time-Frequency Domain

Abdeldjalil Aissa El Bey, Nguyen Linh-Trung, Karim Abed-Meraim, Adel Belouchrani, Yves Grenier

Underdetermined Blind Separation of Nondisjoint Sources in the Time-Frequency Domain

Abdeldjalil Aïssa-El-Bey, Nguyen Linh-Trung, Karim Abed-Meraim, Adel Belouchrani and Yves Grenier

***Abstract*— This paper considers the blind separation of nonstationary sources in the underdetermined case, when there are more sources than sensors. A general framework for this problem is to work on sources that are sparse in some signal representation domain. Recently, two methods have been proposed with respect to the time-frequency (TF) domain. The first uses quadratic time-frequency distributions (TFDs) and a clustering approach, and the second uses a linear TFD. Both of these methods assume that the sources are disjoint in the TF domain; i.e. there is at most one source present at a point in the TF domain. In this paper, we relax this assumption by allowing the sources to be TF-nondisjoint to a certain extent. In particular, the number of sources present at a point is strictly less than the number of sensors. The separation can still be achieved thanks to subspace projection that allows us to identify the sources present and to estimate their corresponding TFD values. In particular, we propose two subspace-based algorithms for TF-nondisjoint sources, one uses quadratic TFDs and the other a linear TFD. Another contribution of this paper is a new estimation procedure for the mixing matrix. Finally, then numerical performance of the proposed methods are provided highlighting their performance gain compared to existing ones.**

***Index Terms*— blind source separation, underdetermined/overcomplete representation, spatial time-frequency representation, vector clustering, subspace projection, speech signals, sparse signal decomposition/representation.**

I. Introduction

SOURCE SEPARATION aims at recovering multiple sources from multiple observations (mixtures) received by a set of linear sensors. The problem is said to be 'blind' when the observations have been linearly mixed by the transfer medium, while having no *a priori* knowledge of the transfer medium or the sources. Blind source separation (BSS) has applications in several areas, such as communication, speech/audio processing, and biomedical engineering [1]. A fundamental and necessary assumption of BSS is that the sources are statistically independent and thus are often sought solutions using higher-order statistical information [2]. If some information about the sources is available at hand, such as temporal coherency [3], source nonstationarity [4], or source cyclostationarity [5] then one can remain in the second-order statistical scenario.

Manuscript received November 7, 2005; revised February 28, 2006.

A. Aïssa-El-Bey, K. Abed-Meraim and Y. Grenier are with the TSI department, ENST-Paris, 46 rue Barrault 75634, Paris Cedex 13, France. Email: {elbey, trung, abed, grenier}@tsi.enst.fr

N. Linh-Trung is at the College of Technology, Vietnam National Univ., 144 Xuanthuy, Caugiay, Hanoi, Vietnam. Email: linhtrung@ieee.org

A. Belouchrani is at the ENP, 10 avenue Hassan Badi 16200, El Harache, Algiers, Algeria. Email: adel.belouchrani@enp.edu.dz

The BSS is said to be *underdetermined* if there are more sources than sensors. In that case, the mixing matrix is not invertible and, consequently, a solution for source estimation must also be found even if the mixing matrix has been estimated. A general framework for underdetermined blind source separation (UBSS) is to exploit the sparseness, if it exists, of the sources in a given signal representation domain [6]. The mixtures are then transformed to this domain; one may then, estimate the transformed sources using their sparseness, and finally recover their time waveforms by source synthesis. For more information on BSS and UBSS methods, see for example a recent survey [7].

Recently, several UBSS methods for *nonstationary sources* have been proposed, given that these sources are sparse in the time-frequency (TF) domain [8]–[10]. The first method uses quadratic 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 want to relax 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 [11] (which corresponds to part of this work) and in [12] for the parametric mixing matrix case. 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 sources present, and hence, to estimate the corresponding TFD values of these sources.

The main contribution of this paper is proposing two subspace-based algorithms for UBSS in the TF domain; one uses quadratic TFDs while the other uses linear TFD. In line with the cluster-based quadratic algorithm proposed in [8], we also propose here a cluster-based algorithm but using a linear TFD, which is not a block-based technique like the quadratic one. Therefore, its low cost computation is useful for processing speech and audio sources. Another contribution of the paper is a method of estimation for the mixing matrix.

The paper is organized as follows. Section II-A formulates the UBSS problem, introduces the underlying TF tools, and states some TF conditions necessary for the separation of nonstationary sources in the TF domain. Section III deals with the TF-disjoint sources. It reviews the cluster-based

quadratic TF-UBSS algorithm [8], and from that, proposes a cluster-based linear TF-UBSS algorithm. Section IV proposes two subspace-based TF-UBSS algorithms for TF-nondisjoint sources, using quadratic and linear TFDs. In this section, we propose also a method for the blind estimation of mixing matrix. There is some discussion of the proposed methods in Section V. The performance of the above methods are numerically evaluated in Section VI.

II. Problem Formulation

A. Data model

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

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \tag{1}$$

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, and $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$ is the mixing matrix of size $M \times N$ that satisfies:

Assumption 1: *The column vectors of* $\mathbf{A}$ *are pair-wise linearly independent.*

That is, for any index pair $i, j \in \mathcal{N}$, where $\mathcal{N} = \{1, \ldots, N\}$, and $i \neq j$, we have $\mathbf{a}_i$ and $\mathbf{a}_j$ linearly independent. This assumption is necessary because if otherwise, we have $\mathbf{a}_1 = \alpha\mathbf{a}_2$ for example, then the input/output relation (1) can be reduced to

$$\mathbf{x}(t) = [\mathbf{a}_1, \mathbf{a}_3, \ldots, \mathbf{a}_N]\,[s_1(t) + \alpha s_2(t), s_3(t), \ldots, s_N(t)]^T,$$

and hence the separation of $s_1(t)$ and $s_2(t)$ is inherently impossible.

It is known that BSS is only possible up to some scaling and permutation. We take advantage of these indeterminacies to further assume, without loss of generality, that the column vectors of $\mathbf{A}$ all have unit norm, i.e. $\|\mathbf{a}_i\| = 1$ for all $i \in \mathcal{N}$.

The sources are nonstationary, that is their frequency spectra vary in time. Often, nonstationarity imposes more difficulties on a problem, however, in this case it actually offers a solution: one can solve the BSS problem without using higher-order approaches by directly exploiting the additional information of this TF diversity across the spectra; this solution was proposed in [4]. We defer to a little later making TF assumptions on the sources, and for now we introduce the concept of TF signal processing.

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 involve 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 time-frequency distribution (TFD).

The general class of quadratic TFDs of an analytic signal $z(t)$ is defined as [13]:

$$\rho_{zz}(t,f) \triangleq \iiint_{-\infty}^{\infty} e^{j2\pi\nu(u-t)}\,\Gamma(\nu,\tau) \times z(u+\frac{\tau}{2})z^*(u-\frac{\tau}{2})\,e^{-j2\pi f\tau}\,d\nu\,du\,d\tau, \tag{2}$$

where $\Gamma(\nu,\tau)$ is a two-dimensional function in the so-called ambiguity domain and is called the Doppler-lag kernel, and the superscript $(^*)$ denotes the conjugate operator. We can design a TFD with certain desired properties by properly constraining Γ.

When $\Gamma(\nu,\tau) = 1$ we have the following famous Wigner-Ville distribution (WVD):

$$\rho_{zz}^{\text{wvd}}(t,f) \triangleq \int_{-\infty}^{\infty} z(t+\frac{\tau}{2})z^*(t-\frac{\tau}{2})\,e^{-j2\pi f\tau}\,d\tau. \tag{3}$$

The WVD is the most widely studied TFD. It achieves maximum energy concentration in the TF plane around the instantaneous frequency for linear frequency-modulated (LFM) signals. However, it is in general non-positive and it introduces the so-called "cross-terms" when multiple frequency laws (e.g. two LFM components) exist in the signals, due to the quadratic multiplication of shifted versions of the signals.

Another well-known TFD and most used in practice is the short-time Fourier transform (STFT):

$$\mathcal{S}_z(t,f) \triangleq \int_{-\infty}^{\infty} z(\tau)h(\tau-t)\,e^{-j2\pi f\tau}d\tau, \tag{4}$$

where $h(t)$ is a window function. Note that the STFT is a *linear* TFD[1], and its quadratic version, called the spectrogram (SPEC), is defined as:

$$\rho_{zz}^{\text{spec}}(t,f) \triangleq |\mathcal{S}_z(t,f)|^2. \tag{5}$$

Clearly, from the definition, there is no cross-terms effect present in STFT, hence in the SPEC. However, these distributions have very low TF resolution in comparison with the WVD. The low cost of implementation for the STFT, hence for the SPEC, in comparison with that for the WVD and, together with the advantage of being free of cross-terms, justifies the fact that the STFT is most used in practice, especially for speech or audio signals. But when it comes to FM-like signals, the WVD is preferred.

To combine the high resolution of the WVD while using the free cross-term property of the SPEC, the masked Wigner-Ville distribution (MWVD) is derived so that:

$$\rho_{zz}^{\text{mwvd}}(t,f) \triangleq \rho_{zz}^{\text{wvd}}(t,f) \cdot \rho_{zz}^{\text{spec}}(t,f). \tag{6}$$

There are many other useful TFDs in the literature, notably those that give high TF resolution while effectively minimizing the cross-terms, for example the B distribution [14]. However, we only introduce here the TFDs above since they will be used in the later sections.

[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.

C. TF conditions on sources

Now, as we have introduced the concept of TF signal processing as a useful tool for analyzing nonstationary signals, some TF conditions need to be applied to the sources. Note that the TF method in [4] does not work for UBSS because the mixing matrix is not invertible. 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 the separation is to be carried out in that domain to exploit the sparseness.

1) TF-disjoint sources: Recently, there have been several UBSS methods, notably those in [8] and [9], 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 Ω_1 and Ω_2 are the TF supports of two sources $s_1(t)$ and $s_2(t)$ then $\Omega_1 \cap \Omega_2 = \emptyset$. This condition can be illustrated in Figure 1. However, this is a rather strict assumption. A more practical assumption is that the sources are almost-disjoint in the TF domain [8], allowing some small overlapping in the TF domain, for which the above two methods also worked.

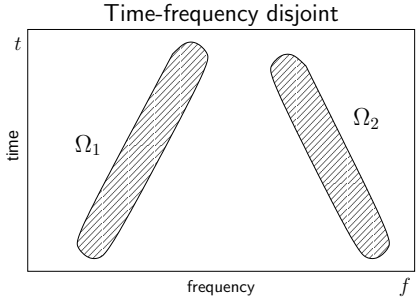


Fig. 1. Source TF-disjoint condition: $\Omega_1 \cap \Omega_2 = \emptyset$ (when $\Omega_1 \cap \Omega_2 \approx \emptyset$, sources are said to be TF-almost-disjoint).

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; as illustrated in Figure 2.

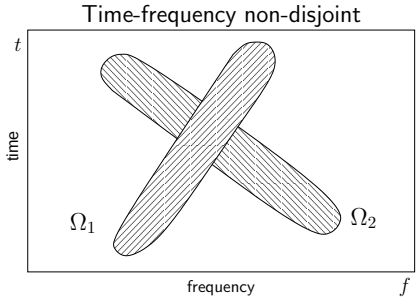


Fig. 2. TF nondisjoint condition: $\Omega_1 \cap \Omega_2 \neq \emptyset$

This is motivated by a drawback of the method in [8]. Although this method worked well under the TF-almost-disjoint condition, it did not explicitly treat the TF regions where the sources were allowed to have some small 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 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:

Assumption 2: *The number of sources that contribute their energy at any TF point is strictly less than the number of sensors.*

In other words, for the configuration of M sensors, there exist 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.

We also make another assumption on the TF conditioning of the sources.

Assumption 3: *For each source, there exists a region in the TF domain, where this source exists alone.*

Note that, this assumption is easily met and hence not restrictive for audio sources and FM-like signals. Also, it should be noted that this last assumption is, however, not a restriction on the use of subspace projection, because it will only be used later for the estimation of the mixing matrix. If otherwise, the mixing matrix can be obtained by another method, for example the one in [15], then Assumption 3 can be omitted.

III. Cluster-based TF-UBSS approach for disjoint sources

A. Quadratic TFD approach

In this section, we review a method proposed in [8] based on the idea of clustering; hence, it is now referred to as the *cluster-based quadratic TF-UBSS algorithm*. For a signal vector $\mathbf{z}(t) = [z_1(t), \ldots, z_N(t)]^T$, the Spatial Time Frequency Distribution (STFD) matrix is given by [4]:

$$\mathbf{D_{zz}}(t,f) \triangleq \begin{bmatrix} \rho_{z_1 z_1}(t,f) & \ldots & \rho_{z_1 z_N}(t,f) \\ \vdots & \ddots & \vdots \\ \rho_{z_N z_1}(t,f) & \ldots & \rho_{z_N z_N}(t,f) \end{bmatrix}, \quad (7)$$

where, for $i,j \in \mathcal{N}$, $\rho_{z_i z_j}(t,f)$ is the quadratic cross-TFD between $z_i(t)$ and $z_j(t)$ as obtained by (2), but with the first z being replaced by z_i and the second by z_j. By definition, the STFD takes into account the spatial diversity.

By applying the STFD defined in (7) on both sides of the BSS model in (1), we obtain the following TF-transformed structure:

$$\mathbf{D_{xx}}(t,f) = \mathbf{A}\mathbf{D_{ss}}(t,f)\mathbf{A}^H \quad (8)$$

where $\mathbf{D_{ss}}(t,f)$ and $\mathbf{D_{xx}}(t,f)$ are, respectively, the source STFD matrix and mixture STFD matrix.

TABLE I
CLUSTER-BASED QUADRATIC TF-UBSS ALGORITHM USING STFD

1) Mixture STFD computation by (10); noise thresholding by (11).
2) Noise thresholding and auto-source point selection by (11).
3) Vector clustering by (12) and k-means algorithm; source TFD estimation by (13).
4) Source TF synthesis by [16].

Let us call an *auto-source TF point* a point at which there is a true energy contribution/concentration of source or sources in the TF domain, and a *cross-source point* a point at which there is a 'false' energy contribution (due to the cross-term effect of quadratic TFDs). Note that, at other points with no energy contribution, the TFD value is ideally equal to zero. Under the assumption that all sources are disjoint in the TF domain, there is only one source present at any auto-source point. Therefore, the structure of $\mathbf{D}_{\mathbf{xx}}(t,f)$ is reduced to

$$\mathbf{D}_{\mathbf{xx}}(t_a,f_a) = \rho_{s_i s_i}(t_a,f_a)\,\mathbf{a}_i\mathbf{a}_i^H, \forall (t_a,f_a)\in\Omega_i, \quad (9)$$

where Ω_i denotes, hereafter, the TF support of source $s_i(t)$.

The observation (9) suggests that for all $(t_a,f_a)\in\Omega_i$, the corresponding set of STFD matrices $\{\mathbf{D}_{\mathbf{xx}}(t_a,f_a)\}$ will have the same principal eigenvector $\mathbf{a}_i$. It is this observation that leads to the general separation method using quadratic TFDs in [8]. Indeed, [8] proposed several algorithms and pointed out that the choice of the TFD should be made carefully in order to have a 'clean' (cross-term free) TFD representation of the mixture, and chose the MWVD as a good candidate. This algorithm is summarized in Table I, and further detailed below for later use.

1) STFD mixture computation and noise thresholding: The STFD of the mixtures using the MWVD is computed by the following:

$$\left[\mathbf{D}_{\mathbf{xx}}^{\text{wvd}}(t,f)\right]_{k,l} = \rho_{x_k x_l}^{\text{wvd}}(t,f) \quad (10a)$$

$$\left[\mathbf{D}_{\mathbf{xx}}^{\text{stft}}(t,f)\right]_{k,l} = \begin{cases}\mathcal{S}_{x_k}(t,f), & \text{for } k=l,\\ 0, & \text{otherwise,}\end{cases} \quad (10b)$$

$$\mathbf{D}_{\mathbf{xx}}^{\text{mwvd}}(t,f) = \mathbf{D}_{\mathbf{xx}}^{\text{wvd}}(t,f)\odot\left|\mathbf{D}_{\mathbf{xx}}^{\text{stft}}(t,f)\right|^2 \quad (10c)$$

In (10), $k,l\in\mathcal{N}$, and $\odot$ denotes the Hadamard product.

2) Noise thresholding and auto-source point selection: A 'noise thresholding' procedure is used to keep only those points having sufficient energy, i.e. auto-source points. One way to do this is: for each time-slice (t_p,f) of the TFD representation, apply the following criterion for all the frequency points f_q belonging to this time-slice:

$$\text{If}\quad \frac{\left\|\mathbf{D}_{\mathbf{xx}}^{\text{mwvd}}(t_p,f_q)\right\|}{\max_f\left\{\|\mathbf{D}_{\mathbf{xx}}^{\text{mwvd}}(t_p,f)\|\right\}} > \epsilon_1, \quad \text{keep}\,(t_p,f_q), \quad (11)$$

where ϵ_1 is a small threshold (typically, $\epsilon_1 = 0.05$). This 'hard thresholding' procedure has been preferred to the 'soft thresholding' using power-weighting of [9] as it contributes also to reducing the computation complexity. The set of all the auto-source points is denoted by Ω. Since sources are TF-disjoint, we have $\Omega = \bigcup_{i=1}^N \Omega_i$. This partition is found in the following way:

3) Vector clustering and source TFD estimation: For each point $(t_a,f_a)\in\Omega$, compute its corresponding spatial direction $\mathbf{a}(t_a,f_a)$

$$\mathbf{a}(t_a,f_a) = \frac{\text{diag}\left\{\underline{\mathbf{D}}_{\mathbf{xx}}^{\text{stft}}(t_a,f_a)\right\}}{\left\|\text{diag}\left\{\underline{\mathbf{D}}_{\mathbf{xx}}^{\text{stft}}(t_a,f_a)\right\}\right\|}, \quad (12)$$

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

Having the set of spatial direction $\{\mathbf{a}(t_a,f_a)|(t_a,f_a)\in\Omega\}$ one can cluster them into N classes using any unsupervised clustering algorithm (see [17] for different clustering methods). The clustering algorithm used in [8] is rather sensitive due to the threshold in use; a robust method should be investigated, and this deserves another contribution. If the number of sources has been well estimated, one can use the so-called k-means clustering algorithm [17] to achieve a good clustering performance. The output of the clustering algorithm is a set of N classes $\{C_i|i\in\mathcal{N}\}$. Also, the collection of all the points that correspond to all the vectors in the class C_i forms the TF support Ω_i of the source $s_i(t)$.

Then, estimate the TFD of the source $s_i(t)$ (up to a scalar constant) as:

$$\hat{\rho}_{s_i}^{\text{wvd}}(t,f) = \begin{cases}\text{trace}\left\{\underline{\mathbf{D}}_{\mathbf{xx}}^{\text{wvd}}(t,f)\right\}, & (t,f)\in\Omega_i,\\ 0, & \text{otherwise.}\end{cases} \quad (13)$$

4) Source TF synthesis: Having obtained the source TFD estimate $\hat{\rho}_{s_i}^{\text{wvd}}(t,f)$, the estimation of the source $s_i(t)$ can be done through a TF synthesis algorithm. The method in [16] is used for TF synthesis from a WVD estimate, based on the following inversion property of the WVD [13]:

$$x(t) = \frac{1}{x^*(0)}\int_{-\infty}^{\infty}\rho_x^{\text{wvd}}(\frac{t}{2},f)\,e^{j2\pi ft}\,df\,,$$

which implies that the signal can be reconstructed to within a complex exponential constant $e^{j\alpha} = x^*(0)/|x(0)|$ given $|x(0)|\neq 0$.

It can be observed that in this version of the quadratic TF-UBSS algorithm, the STFD matrices are not fully needed as only their diagonal entries are used in the algorithm. This should be taken into account to reduce the computational cost.

B. Linear TFD approach

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 briefly review the STFT method in [9], and propose simultaneously a *cluster-based linear TF-UBSS algorithm* using the STFT to avoid some of the drawbacks in [9].

First, under the transformation into the TF domain using the STFT, the model in (1) becomes:

$$\mathcal{S}_{\mathbf{x}}(t,f) = \mathbf{A}\mathcal{S}_{\mathbf{s}}(t,f), \quad (14)$$

where $\mathcal{S}_{\mathbf{x}}(t,f)$ is the mixture STFT vector and $\mathcal{S}_{\mathbf{s}}(t,f)$ is the source STFT vector. Under the assumption that all sources are disjoint in the TF domain, (14) is reduced to

$$\mathcal{S}_{\mathbf{x}}(t_a,f_a) = \mathbf{a}_i\mathcal{S}_{s_i}(t_a,f_a), \forall(t_a,f_a)\in\Omega_i, \forall i\in\mathcal{N}. \quad (15)$$

TABLE II
CLUSTER-BASED LINEAR TF-UBSS ALGORITHM USING STFT

1) Mixture STFT computation by (17); noise thresholding by (18)
2) Vector clustering by (19) and (20).
3) Source STFT estimation by (21).
4) Source TF synthesis by [18].

Now, in [9], the structure of the mixing matrix is particular in that it has only 2 rows (i.e. the method uses only 2 sensors) and the first row of the mixing matrix contains all 1. Then, (15) is expanded to

$$\begin{bmatrix}\mathcal{S}_{x_1}(t_a,f_a)\\ \mathcal{S}_{x_2}(t_a,f_a)\end{bmatrix}=\begin{bmatrix}1\\ a_{2,i}\end{bmatrix}\mathcal{S}_{s_i}(t_a,f_a),$$

which results in

$$a_{2,i}=\frac{\mathcal{S}_{x_2}(t_a,f_a)}{\mathcal{S}_{x_1}(t_a,f_a)}. \tag{16}$$

Therefore, all the points for which the ratios on the right-hand side of (16) have the same value form the TF support Ω_i of a single source, say $s_i(t)$. Then, the STFT estimate of $s_i(t)$ is computed by:

$$\hat{\mathcal{S}}_{s_i}(t,f)=\begin{cases}\mathcal{S}_{x_1}(t,f), & \forall (t,f)\in\Omega_i,\\ 0, & \text{otherwise.}\end{cases}$$

The source estimate $\hat{s}_i(t)$ is then obtained by converting $\hat{\mathcal{S}}_{s_i}(t,f)$ to the time domain using inverse STFT [18]. Note that, the extension of the UBSS method in [9] to more than two sensors is a difficult task. Second, the division on the right-hand side of (16) is prone to error if the denominator is close to zero.

To avoid the above mentioned problems, we propose here a modified version of the previous method referred to as the cluster-based linear TF-UBSS algorithm. In particular, from the observation (15), we can deduce the separation algorithm as shown next, and summarized in Table II.

1) Mixture STFT computation and noise thresholding: Compute the STFT of the mixtures, $\mathcal{S}_{\mathbf{x}}(t,f)$, by applying (4) for each of the mixture in $\mathbf{x}(t)$, as follows:

$$\mathcal{S}_{x_i}(t,f)=\int_{-\infty}^{\infty}x_i(\tau)h(\tau-t)e^{-j2\pi f\tau}d\tau,\quad i=1,\ldots,M, \tag{17a}$$

$$\mathcal{S}_{\mathbf{x}}(t,f)=[\mathcal{S}_{x_1}(t,f),\ldots,\mathcal{S}_{x_M}(t,f)]^T. \tag{17b}$$

Since the STFT is totally free of cross-terms, a point with a nonzero TFD value is ideally an auto-source point. Practically, we can select all auto-source points by only applying a noise thresholding procedure as that in the cluster-based quadratic TF-UBSS algorithm. In particular, for each time-slice (t_p,f) of the TFD representation, apply the following criterion for all the frequency points f_k belonging to this time-slice

$$\text{If}\quad\frac{\|\mathcal{S}_{\mathbf{x}}(t_p,f_k)\|}{\max_f\{\|\mathcal{S}_{\mathbf{x}}(t_p,f)\|\}}>\epsilon_2,\quad\text{then keep }(t_p,f_k), \tag{18}$$

where ϵ_2 is a small threshold (typically, $\epsilon_2=0.05$). Then, the set of all selected points, Ω, is expressed by $\Omega=\bigcup_{i=1}^{N}\Omega_i$, where Ω_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. Hence, by equation (18) (or equation (11)), we would keep only time-frequency points where the signal energy is significant, 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, negligeable at least for moderate and high SNRs.

2) Vector clustering and source TFD estimation: The clustering procedure can be done in a similar manner as in the quadratic algorithm. First, we obtain the spatial direction vectors by:

$$\mathbf{v}(t_a,f_a)=\frac{\mathcal{S}_{\mathbf{x}}(t_a,f_a)}{\|\mathcal{S}_{\mathbf{x}}(t_a,f_a)\|},\quad(t_a,f_a)\in\Omega, \tag{19}$$

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

Next, we cluster these vectors into N classes $\{C_i\,|\,i\in\mathcal{N}\}$, using the k-means clustering algorithm. The collection of all points, whose vectors belong to the class C_i, now forms the TF support Ω_i of the source $s_i(t)$. Then, the column vector $\mathbf{a}_i$ of $\mathbf{A}$ is estimated as the centroid of this set of vectors:

$$\hat{\mathbf{a}}_i=\frac{1}{|C_i|}\sum_{(t,f)\in\Omega_i}\mathbf{v}(t,f), \tag{20}$$

where $|C_i|$ is the number of vectors in this class.

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

$$\hat{\mathcal{S}}_{s_i}(t,f)=\begin{cases}\hat{\mathbf{a}}_i^H\mathcal{S}_{\mathbf{x}}(t,f), & \forall\,(t,f)\in\Omega_i,\\ 0, & \text{otherwise.}\end{cases} \tag{21}$$

since, from (15), we have

$$\hat{\mathbf{a}}_i^H\mathcal{S}_{\mathbf{x}}(t,f)=\hat{\mathbf{a}}_i^H\mathbf{a}_i\mathcal{S}_{s_i}(t,f)\approx\mathcal{S}_{s_i}(t,f),\quad\forall\,(t,f)\in\Omega_i.$$

Note that the STFT is a particular form of wavelet transforms which have been used in [19] for the UBSS of image signals.

IV. SUBSPACE-BASED TF-UBSS APPROACH FOR NONDISJOINT SOURCES

We have seen the cluster-based TF-UBSS methods, using either quadratic TFDs such as the MWVD or linear TFDs such as the STFT, as summarized in Table I or Table II, respectively. These methods relied on the assumption that the sources were TF-disjoint, which has led to the enabling TF-transformed structures in (9) or (15). When the sources are nondisjoint in the TF domain, then these equations are no longer true.

Under the TF-nondisjoint condition, stated in Assumption 2, we propose in this section two alternative methods, one for quadratic TFDs and the other for linear TFDs, for the UBSS problem using subspace projection.

A. Subspace-based quadratic TF-UBSS algorithm

Recall that the first two steps of the cluster-based quadratic TF-UBSS algorithm do not rely on the assumption of TF-disjoint sources (see Table I). Therefore, we can reuse these steps to obtain the set of auto-source points Ω. Now, under the TF-nondisjoint condition, consider an auto-source point $(t_b, f_b) \in \Omega$ such that there are K sources, $K < M$, present at this point. Our goal is to identify the sources present at (t_b, f_b) and to estimate the energy each of these sources contributes.

Denote $\alpha_1, \ldots, \alpha_K \in \mathcal{N}$ the indices of the sources present at (t_b, f_b), and define the following:

$$\tilde{\mathbf{s}} = [s_{\alpha_1}(t), \ldots, s_{\alpha_K}(t)]^T, \tag{22a}$$

$$\tilde{\mathbf{A}} = [\mathbf{a}_{\alpha_1}, \ldots, \mathbf{a}_{\alpha_K}]. \tag{22b}$$

Then, under Assumption 2, (8) is reduced to

$$\mathbf{D}^{\text{wvd}}_{\mathbf{xx}}(t_b, f_b) = \tilde{\mathbf{A}} \mathbf{D}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(t_b, f_b) \tilde{\mathbf{A}}^H, \tag{23}$$

Consequently, given that $\mathbf{D}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}$ is of full rank, we have

$$\text{Range}\left\{\mathbf{D}_{\mathbf{xx}}(t_b, f_b)\right\} = \text{Range}\{\tilde{\mathbf{A}}\}. \tag{24}$$

Let $\mathbf{P}$ be the orthogonal projection matrix onto the noise subspace of $\mathbf{D}^{\text{wvd}}_{\mathbf{xx}}(t_b, f_b)$. Then, from (24), we obtain:

$$\mathbf{P} = \mathbf{I} - \mathbf{V}\mathbf{V}^H, \tag{25}$$

and

$$\begin{cases} \mathbf{P}\mathbf{a}_i = \mathbf{0}, & \forall\, i \in \{\alpha_1, \ldots, \alpha_K\}, \\ \mathbf{P}\mathbf{a}_i \neq \mathbf{0}, & \forall\, i \in \mathcal{N} \backslash \{\alpha_1, \ldots, \alpha_K\} \end{cases} \tag{26}$$

In (25), $\mathbf{V}$ is the matrix formed by the K principal singular eigenvectors of $\mathbf{D}_{\mathbf{xx}}(t_b, f_b)$.

Assuming that $\mathbf{A}$ has been estimated by some method, the observation in (26) enables us to identify the indices $\alpha_1, \ldots, \alpha_K$; and hence, the sources present at (t_b, f_b). In practice, to take into account the estimation noise, one can detect these indices by detecting the K smallest values from the set $\{\|\mathbf{P}\mathbf{a}_i\| \mid i \in \mathcal{N}\}$, as mathematically expressed by:

$$\{\alpha_1, \ldots, \alpha_K\} = \arg\min{}^{K} \{\|\mathbf{P}\mathbf{a}_i\| \mid i \in \mathcal{N}\}, \tag{27}$$

where $\min^K$ denotes the minimization to obtain the K smallest values. The TFD values of the K sources at (t_b, f_b) are estimated as the diagonal elements of the following matrix:

$$\hat{\mathbf{D}}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(t_b, f_b) \approx \tilde{\mathbf{A}}^{\#} \mathbf{D}_{\mathbf{xx}}(t_b, f_b) \left(\tilde{\mathbf{A}}^{\#}\right)^H, \tag{28}$$

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

Here, we propose also an estimation method for $\mathbf{A}$ by using Assumption 3. This assumption states that, for each source $s_i(t)$, there exists a TF region $\mathcal{R}_i$ where $s_i(t)$ exists alone. In other words, $\mathcal{R}_i$ contains all the single-source auto-source points of $s_i(t)$. Therefore, we can reuse the observation (9) in the TF-disjoint case, but for some TF regions, as below:

$$\mathbf{D}_{\mathbf{xx}}(t, f) = \rho_{s_i s_i}(t, f) \mathbf{a}_i \mathbf{a}_i^H, \quad \forall (t, f) \in \mathcal{R}_i, \forall i \in \mathcal{N}.$$

The union of these regions, $\mathcal{R} = \bigcup_{i=1}^{N} R_i$, is detected by the following:

$$\text{If } \left| \frac{\lambda_{max}\{\mathbf{D}^{\text{wvd}}_{\mathbf{xx}}(t, f)\}}{\text{trace}\{\mathbf{D}^{\text{wvd}}_{\mathbf{xx}}(t, f)\}} - 1 \right| < \epsilon_3, \text{ then } (t, f) \in \mathcal{R}, \tag{29}$$

TABLE III
SUBSPACE-BASED QUADRATIC TF-UBSS ALGORITHM USING MWVD

1) Mixture STFD computation by (10).
2) Noise thresholding and auto-source point selection by (11).
3) Single-source auto-source point selection by (29); mixing matrix estimation by (30) and (31)
4) For all auto-source points, perform subspace-based TFD estimation of sources by (25), (27) and (28)
5) Source TF synthesis by [16].

where ϵ_3 is a small threshold value (typically, $\epsilon_3 \leq 0.1$) and $\lambda_{max}\{\mathbf{D}^{\text{wvd}}_{\mathbf{xx}}(t, f)\}$ denotes the maximum eigenvalue of $\mathbf{D}^{\text{wvd}}_{\mathbf{xx}}(t, f)$. Then, we can apply the same vector clustering procedure as in Section III-A.3 to estimate $\mathbf{A}$. In particular, we first obtain all the spatial direction vectors:

$$\mathbf{a}(t, f) = \frac{\text{diag}\left\{\underline{\mathbf{D}}^{\text{stft}}_{\mathbf{xx}}(t, f)\right\}}{\left\|\text{diag}\left\{\underline{\mathbf{D}}^{\text{stft}}_{\mathbf{xx}}(t, f)\right\}\right\|}, \quad \forall (t, f) \in \mathcal{R}. \tag{30}$$

Next, we cluster these vectors into N classes $\{D_i | i \in \mathcal{N}\}$ using the k-means clustering algorithm. The collection of all points, whose vectors belong to the class D_i, now forms the TF region $\mathcal{R}_i$ of the source $s_i(t)$. Finally, the column vectors $\mathbf{A}$ are estimated as the centroid vectors of these classes as:

$$\hat{\mathbf{a}}_i = \frac{1}{|D_i|} \sum_{(t,f) \in \mathcal{R}_i} \mathbf{a}(t, f), \quad \forall i \in \mathcal{N} \tag{31}$$

where D_i is the number of points in $\mathcal{R}_i$.

Table III gives a summary of the subspace-based quadratic TF-UBSS algorithm.

B. Subspace-based linear TF-UBSS algorithm

Similarly, we propose here a subspace-based linear TF-UBSS algorithm for TF-nondisjoint sources using STFT. We also use the first step of the cluster-based linear TF-UBSS algorithm (see Table II) to obtain all the auto-source points Ω. Under the TF-nondisjoint condition, consider an auto-source point $(t_b, f_b) \in \Omega$ at which there are K sources $s_{\alpha_1}(t), \ldots, s_{\alpha_K}(t)$ present, with $K < M$. Then, (8) is reduced to the following

$$\mathcal{S}_{\mathbf{x}}(t_b, f_b) = \tilde{\mathbf{A}} \mathcal{S}_{\tilde{\mathbf{s}}}(t_b, f_b), \quad \forall (t_b, f_b) \in \Omega \tag{32}$$

where $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{s}}$ are as previously defined in (22).

Let $\mathbf{Q}$ represent the orthogonal projection matrix onto the noise subspace of $\tilde{\mathbf{A}}$. Then, $\mathbf{Q}$ can be computed by:

$$\mathbf{Q} = \mathbf{I} - \tilde{\mathbf{A}} \left(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}}\right)^{-1} \tilde{\mathbf{A}}^H. \tag{33}$$

We have the following observation:

$$\begin{cases} \mathbf{Q}\mathbf{a}_i = 0, & i \in \{\alpha_1, \ldots, \alpha_K\} \\ \mathbf{Q}\mathbf{a}_i \neq 0, & i \in \mathcal{N} \backslash \{\alpha_1, \ldots, \alpha_K\} \end{cases}. \tag{34}$$

If $\mathbf{A}$ has already been estimated by some method, then this observation gives us the criterion to detect the indices

TABLE IV
SUBSPACE-BASED LINEAR TF-UBSS ALGORITHM USING STFT

1) STFT computation.
2) Noise thresholding and auto-source point selection
3) Mixing matrix estimation by (20) and (37), and k-means algorithm.
4) For all auto-source points, perform subspace-based TFD estimation of sources by (33), (35) and (36).
5) Source TF synthesis by [18].

$\alpha_1, \ldots, \alpha_K$; and hence, the contributing sources at the auto-source point (t_b, f_b). In practice, to take into account noise, one detects the column vectors of $\tilde{\mathbf{A}}$ minimizing:

$$\{\alpha_1, \ldots, \alpha_K\} = \arg \min_{\beta_1,\ldots,\beta_K} \left\{ \|\mathbf{Q}\mathcal{S}_{\mathbf{x}}(t,f)\| \mid \tilde{\mathbf{A}}_\beta \right\} \quad (35)$$

where $\tilde{\mathbf{A}}_\beta = [\mathbf{a}_{\beta_1}, \ldots, \mathbf{a}_{\beta_K}]$.
Next, TFD values of the K sources at TF point (t, f) are estimated by:

$$\hat{\mathcal{S}}_{\tilde{\mathbf{s}}}(t,f) \approx \tilde{\mathbf{A}}^{\#} \mathcal{S}_{\mathbf{x}}(t,f). \quad (36)$$

Here we propose a method for estimating the mixing matrix $\mathbf{A}$. This is performed by clustering all the spatial direction vectors in (19) as for the preview TF-UBSS algorithm. Then within each class C_i we eliminate the far-located vectors from the centroid (in the simulation we estimate vectors $\mathbf{v}(t,f)$ such that:

$$\|\mathbf{v}(t,f) - \hat{\mathbf{a}}_i\| > 0.8 \max_{\mathbf{v}(t,f)\in\Omega_i} \|\mathbf{v}(t,f) - \hat{\mathbf{a}}_i\|, \quad (37)$$

leading to a size-reduced class $\tilde{C}_i$. Essentially this is to keep the vectors corresponding to the TF region $\mathcal{R}_i$, which are ideally equal to the spatial direction $\mathbf{a}_i$ of the considered source signal. Finally, the i^{th} column vector of $\mathbf{A}$ is estimated as the centroid of $\tilde{C}_i$.

Table IV provides a summary of the subspace projection based TF-UBSS algorithm using STFT.

V. DISCUSSION

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

1) *Number of sources:* The number of sources N is assumed known in the clustering method (k-means) that we have used. However, there exist clustering methods [17] 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, we have to evaluate the number K of overlapping sources at a given TF point. This can be done by finding out the number of non-zero eigenvalues of $\mathbf{D}_{\mathbf{xx}}^{\text{wvd}}(t,f)$ using criteria such as Minimum Description Length (MDL) or Akaike Information Criterion (AIC) [20]. It is also possible to consider a fixed (maximum) value of K that is used for all auto-source TF points. Indeed, if the number of overlapping sources is less than K, we would estimate close-to-zero source STFT values. For example, if we assume $K = 2$ 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 criterion. In our simulation, we did choose this solution with $K = 2$ or $K = 3$.

3) *Quadratic versus linear TFDs:* We have proposed two algorithms using quadratic and linear TFDs. The one using the quadratic TFD should be preferred when dealing with FM-like signals and for small or moderate sample sizes. For audio source separation often the case the sample size is large, and hence, to reduce the computational cost one should prefer the linear TFD based UBSS algorithm. Overall, the quadratic version performs slightly better than the linear one but costs much more in computations.

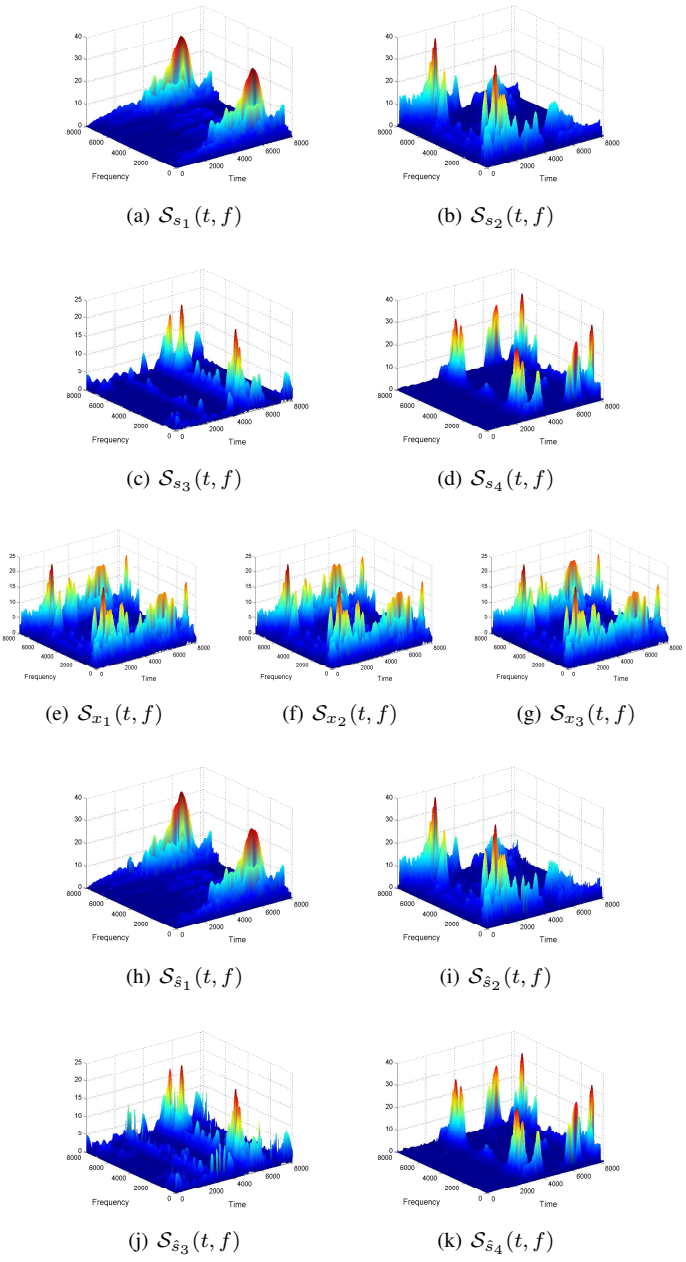
4) *Separation quality versus number of sources:* Although we are in the underdetermined case, the number of sources N should not exceed too much the number of sensors. Indeed, when N increases, the level of source interference increases, and hence, the source disjointness assumption is ill-satisfied. Moreover, for a large number of sources, the likelihood of having two sources closely spaced, i.e. such that the spatial directions $\mathbf{a}_i$ and $\mathbf{a}_j$ are 'close' to linear dependency, increases. In that case, vector clustering performance degrades significantly. In brief, sparseness and spatial separation are the two limiting factors against increasing the number of sources. Figure 8 illustrates the performance degradation of source separation versus the number of sources.

VI. SIMULATION RESULTS

A. Simulation results of subspace-based TF-UBSS algorithm using STFT

In the simulations, we use a uniform linear array of $M = 3$ sensors. It receives signals from $N = 4$ independent speech sources in the far field from directions $\theta_1 = 15$, $\theta_2 = 30$, $\theta_3 = 45$ and $\theta_4 = 75$ degrees respectively. The sample size is $T = 8192$ samples. In Figure 3, the top four plots represent the TF representation of the original sources signal, the middle three plots represent the TF representation of the M mixture signals and the bottom four plots represent the TF representation of the estimate of sources by the subspace-based algorithm using STFT (Table IV). Figure 4 represents the same disposition of signals but in the time domain. In Figure 5, we compare the separation performance obtained by the subspace-based algorithm with $K = 2$ and the cluster-based algorithm (Table II). It is observed that subspace-based algorithm provides much better separation results than those obtained by the cluster-based algorithm.

In the subspace-based method, one first needs to estimate the mixing matrix $\mathbf{A}$. This is done by the cluster-based method presented previously. The plot in Figure 6 represents the normalized estimation error of $\mathbf{A}$ versus the SNR in dB. Clearly, the proposed estimation method of the mixing matrix provides satisfactory performance, while the plot in Figure 7 presents



(a) $\mathcal{S}_{s_1}(t,f)$ (b) $\mathcal{S}_{s_2}(t,f)$

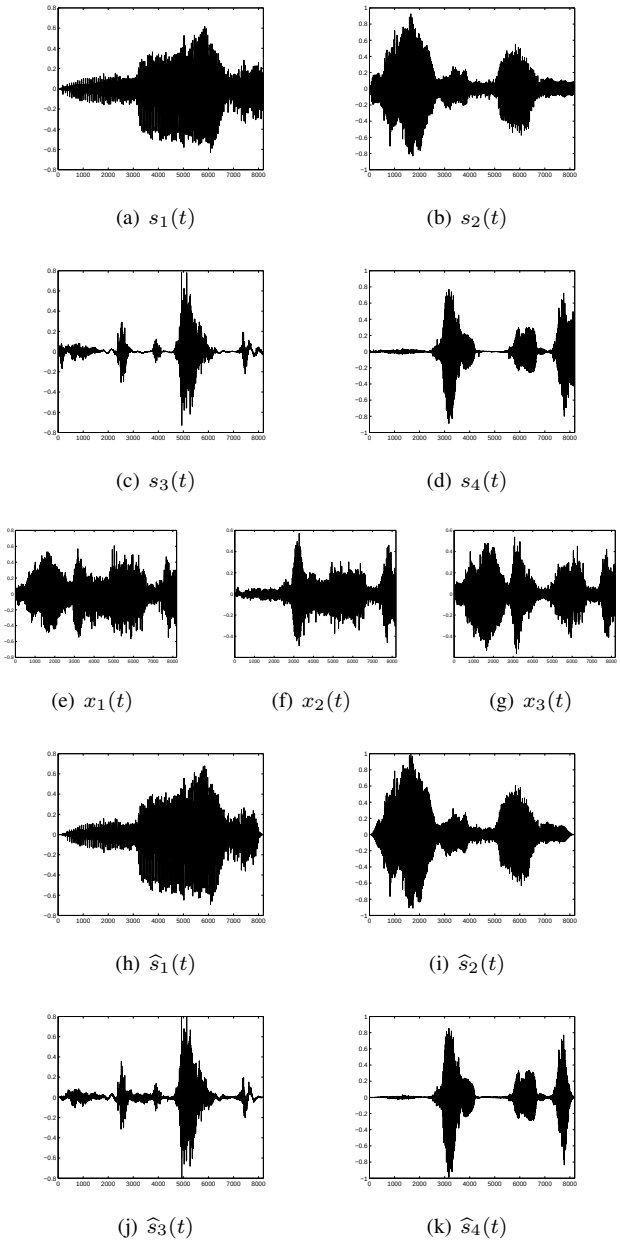
(c) $\mathcal{S}_{s_3}(t,f)$ (d) $\mathcal{S}_{s_4}(t,f)$

(e) $\mathcal{S}_{x_1}(t,f)$ (f) $\mathcal{S}_{x_2}(t,f)$ (g) $\mathcal{S}_{x_3}(t,f)$

(h) $\mathcal{S}_{\hat{s}_1}(t,f)$ (i) $\mathcal{S}_{\hat{s}_2}(t,f)$

(j) $\mathcal{S}_{\hat{s}_3}(t,f)$ (k) $\mathcal{S}_{\hat{s}_4}(t,f)$

Fig. 3. Simulated example (viewed in TF domain) for the subspace-based TF-UBSS algorithm with STFT in the case of 4 speech sources and 3 sensors. The top four plots represent the original source signals, the middle three plots represent the 3 mixtures, and the bottom four plots represent the source estimates.



(a) $s_1(t)$ (b) $s_2(t)$

(c) $s_3(t)$ (d) $s_4(t)$

(e) $x_1(t)$ (f) $x_2(t)$ (g) $x_3(t)$

(h) $\widehat{s}_1(t)$ (i) $\widehat{s}_2(t)$

(j) $\widehat{s}_3(t)$ (k) $\widehat{s}_4(t)$

Fig. 4. Simulated example (viewed in time domain) for the subspace-based TF-UBSS algorithm with STFT in the case of 4 speech sources and 3 sensors. The top four plots represent the original source signals, the middle three plots represent the 3 mixtures, and the bottom four plots represent the source estimates.

the separation performance when using the exact matrix $\mathbf{A}$ compared to that obtained with the proposed estimate $\hat{\mathbf{A}}$.

Figure 8 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 V.

In Figure 9, we compare the performance obtained with the subspace-based method for $K = 2$ and $K = 3$. In that experiment, we have used $M = 4$ sensors and $N = 5$ source signals. One

can observe that, for high SNRs, the case of $K = 3$ leads to a better separation performance than for the case of $K = 2$. However, for low SNRs, a large value of K increases the estimation noise (as mentioned in Section V) and hence degrades the separation quality.

B. Simulation results of subspace-based TF-UBSS algorithm using STFD

In this simulation, we use a uniform linear array of $M = 3$ sensors with half wavelength spacing. It receives signals from $N = 4$ independent LFM sources, each has 256 samples, in the presence of additive Gaussian noise where the SNR=20 dB.

We compare the cluster-based (Table I) and the proposed subspace-based (Table III) TF-UBSS algorithms. Fig-

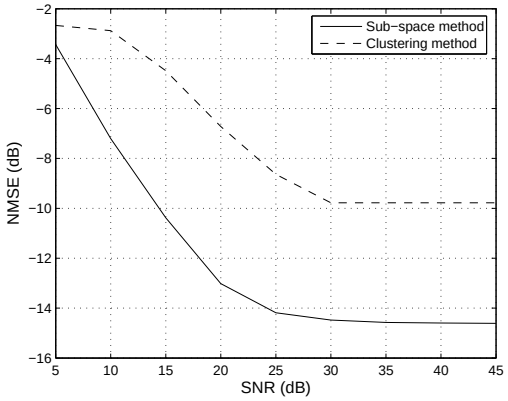


Fig. 5. Comparison between subspace-based and cluster-based TF-UBSS algorithms using STFT: normalized MSE (NMSE) versus SNR for 4 speech sources and 3 sensors.

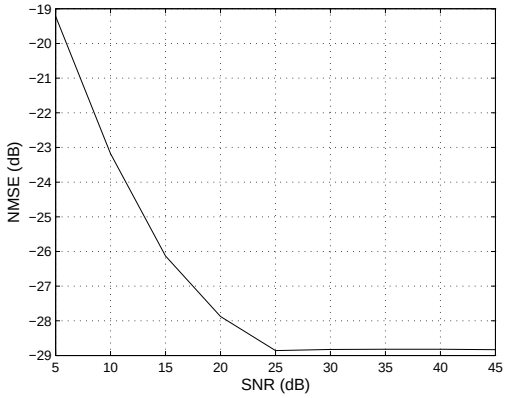


Fig. 6. Mixing matrix estimation: normalized MSE versus SNR for 4 speech sources and 3 sensors.

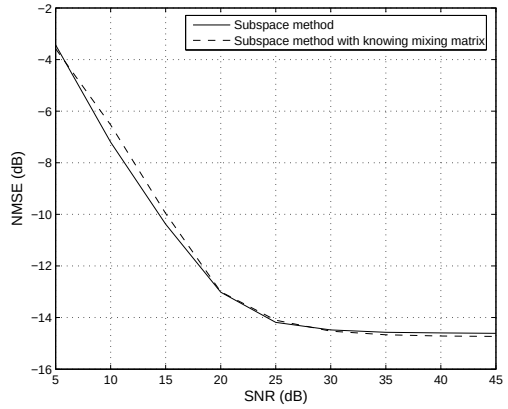


Fig. 7. Comparison, for the subspace-based TF-UBSS algorithm using STFT, when the mixing matrix $\mathbf{A}$ is known or unknown: NMSE of the source estimates.

ures 10-(a,d,g,j) represent the TFDs (using WVD) of the four sources. Figures 10-(b,e,h,k) show the estimated source TFDs using the cluster-based algorithm, whereas Figures 10-(c,f,i,l) are those obtained by the subspace-based algorithm.

From Figures 10-(b,e) we can see that the overlapping points

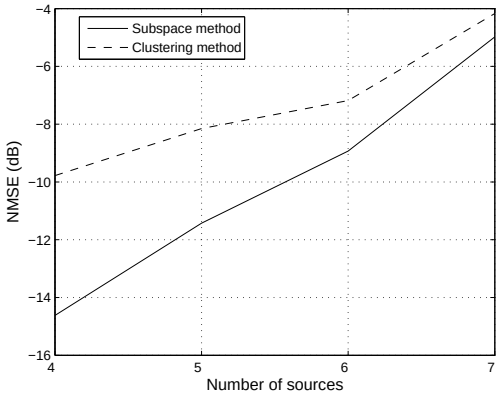


Fig. 8. Comparison between subspace-based and cluster-based TF-UBSS algorithms using STFT: NMSE versus number of sources.

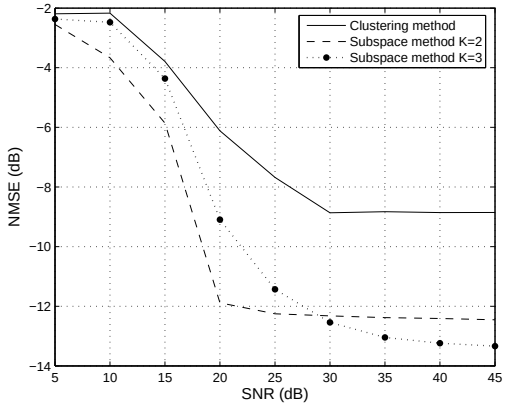


Fig. 9. Comparison between subspace-based and cluster-based TF-UBSS algorithms using STFT: NMSE of the source estimates for different sizes of the projector, for the case of 5 sources and 4 sensors.

between source $s_1(t)$ and source $s_2(t)$ were picked up by source $s_2(t)$ with the cluster-based algorithm. On the other hand, using the subspace-based algorithm, the intersection points have been redistributed to the two sources (Figure 10-(c,f)).

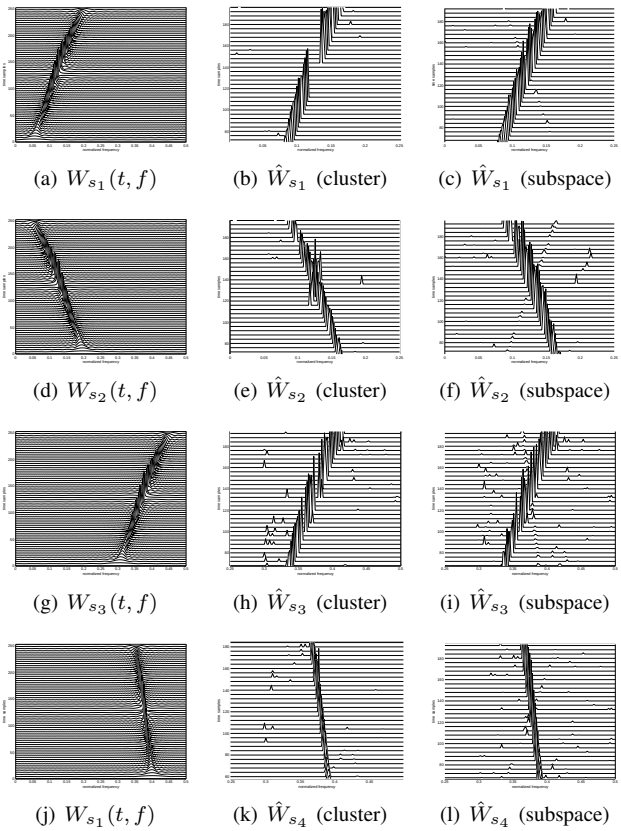
In general, the overlapping points in the nondisjoint case have been explicitly treated. This provides a visual performance comparison.

In Figure 11, we compare the statistical separation performance between the subspace-based algorithm and the cluster-based algorithm using STFD, evaluated over 1000 Monte-Carlo runs.

One can also notice that the gain here is smaller than the one obtained previously for audio sources. This is due to the fact that the overlapping region of the considered signals is smaller. This result confirms the previous visual observation with respect to the performance gain in favor of our subspace-based method.

VII. Conclusions

This paper introduces new methods for the UBSS of TF-nondisjoint nonstationary sources using time-frequency repre-



(a) $W_{s_1}(t,f)$ (b) $\hat{W}_{s_1}$ (cluster) (c) $\hat{W}_{s_1}$ (subspace)

(d) $W_{s_2}(t,f)$ (e) $\hat{W}_{s_2}$ (cluster) (f) $\hat{W}_{s_2}$ (subspace)

(g) $W_{s_3}(t,f)$ (h) $\hat{W}_{s_3}$ (cluster) (i) $\hat{W}_{s_3}$ (subspace)

(j) $W_{s_1}(t,f)$ (k) $\hat{W}_{s_4}$ (cluster) (l) $\hat{W}_{s_4}$ (subspace)

Fig. 10. Simulated example (viewed in TF domain) for the subspace-based TF-UBSS algorithm with STFT in the case of 4 LFM sources and 3 sensors. From left to right, the figures respectively represent the original source TF signatures, the estimated source TF signatures using the cluster-based algorithm, and the estimated source TF signatures using the subspace-based algorithm.

sentations. The main advantages over the proposed separation algorithms are, first, a weaker assumption on the source 'sparseness', i.e. the sources are not necessarily TF-disjoint, and second, 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 compared to those existing in the literature.

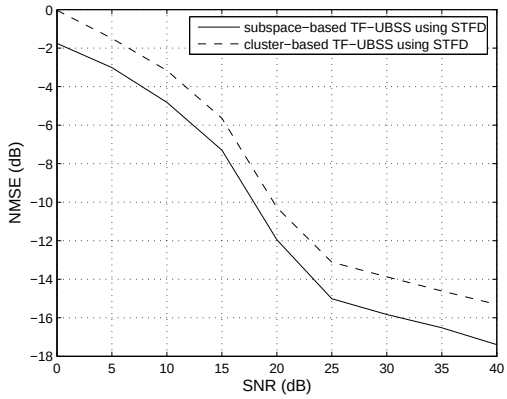


Fig. 11. Comparison between subspace-based and cluster-based TF-UBSS algorithms using STFD: normalized MSE (NMSE) versus SNR for 4 LFM sources and 3 sensors.

REFERENCES

[1] A. K. Nandi, Ed., *Blind estimation using higher-order statistics*. Boston: Kluwer Academic Publishers, 1999.

[2] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proc. of the 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 Transaction on Signal Processing*, 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 Transaction on Signal Processing*, 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 Transaction on Signal Processing*, vol. 49, no. 4, pp. 694–701, Apr. 2001.

[6] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, Nov. 2001.

[7] P. O'Grady, B. Pearlmutter, and S. Rickard, "Survey of sparse and non-sparse methods in source separation," *International Journal Imaging Sys. Tech.*, vol. 15, no. 1, pp. 18–33, 2005.

[8] N. Linh-Trung, A. Belouchrani, K. Abed-Meraim, and B. Boashash, "Separating more sources than sensors using time–frequency distributions," *EURASIP Journal Applied Signal Processing*, vol. 2005, no. 17, pp. 2828–2847, 2005.

[9] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transaction on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July 2004.

[10] B. Barkat and K. Abed-Meraim, "Algorithms for blind components separation and extraction from the time-frequency distribution of their mixture," *EURASIP Journal Applied Signal Processing*, vol. 2004, no. 13, pp. 2025–2033, 2004.

[11] 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. ISSPA*, vol. 1, Sydney, Australie, Aug. 2005, pp. 46–49.

[12] S. Rickard, T. Melia, and C. Fearon, "Desprit - histogram based blind source separation of more sources than sensors using subspace methods," in *Proc. Applications of Signal Processing to Audio and Acoustics IEEE Workshop*, Oct. 2005, pp. 5–8.

[13] B. Boashash, Ed., *Time Frequency Signal Analysis and Processing: Method and Applications*. Oxford: Elsevier, 2003.

[14] B. Barkat and B. Boashash, "A high-resolution quadratic time-frequency distribution for multicomponent signal analysis," *IEEE Transaction on Signal Processing*, vol. 49, no. 10, pp. 2232–2239, Oct. 2001.

[15] L. D. Lathauwer, B. Moor, and J. Vandewalle, "ICA techniques for more sources than sensors," in *HOS Proc. of IEEE Signal Processing Workshop*, June 1999, pp. 121–124.

[16] G. F. Boudreaux-Bartels and T. W. Parks, "Time-varying filtering and signal estimation using Wigner distributions," *IEEE Transaction on Acoustic, Speech, Signal Processing*, vol. ASSP-34, no. 3, pp. 442–451, Mar. 1986.

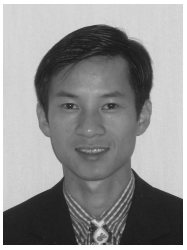
[17] I. E. Frank and R. Todeschini, *The data analysis handbook*. Elsevier, Sci. Pub. Co., 1994.

[18] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Transaction on Acoustic, Speech, Signal Processing*, vol. ASSP-32, no. 2, pp. 236–243, Apr. 1984.

[19] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev, *Independent Component Analysis: Principles and Practice*. Cambridge: S. J. Roberts and R. M. Everson eds., 2001, ch. Blind Source Separation by Sparse Decomposition.

[20] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transaction on Acoustic, Speech, Signal Processing*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.

Abdeldjalil Aïssa-El-Bey was born in Algiers, Algeria, in 1981. He received the State Engineering degree from École Nationale Polytechnique (ENP), Algiers, Algeria, in 2003, 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 at the Signal and Image Processing Department of 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.



Nguyen Linh-Trung was born in 1973 in Vietnam. He completed his undergraduate education (Bachelor of Electrical Engineering) and postgraduate education (PhD of Electrical Engineering) in 1997 and 2002 respectively, both from the Queensland University of Technology (Brisbane, Australia). He has visited the École Nationale Supérieure des Télécommunications (Paris, France) several times (2001, 2002, 2003) during and after his PhD, where he worked on the problem of time-frequency based underdetermined blind source separation. He was with the Information Group of the Aston University (Birmingham, UK) from October 2002 to January 2003, as a post-doctoral research associate, working on optimal biorthogonal representation of signals. From September 2003 to September 2005, he joined the Centre National d'Études Spatiales (Toulouse, France) as a post-doctoral research fellow, investigating mechanisms for priority access in emergency communications over public satellite networks. Since January 2006, he is a faculty member at the College of Technology of the Vietnam National University, Hanoi.

Karim Abed-Meraim 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, France, 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 research staff at the Electrical Engineering Department of the University of Melbourne 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) at the Signal and Image Processing Department of 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 an IEEE Senior Member and a past Associate Editor for the IEEE Transactions on Signal Processing.

Adel Belouchrani received the State Engineering degree in 1991 from École Nationale Polytechnique (ENP), Algiers, Algeria, the M.S. degree in signal processing from the Institut National Polytechnique de Grenoble (INPG), France, in 1992, and the Ph.D. degree in signal and image processing from Télécom Paris (ENST), France, in 1995. He was a Visiting Scholar at the Electrical Engineering and Computer Sciences Department, University of California, Berkeley, from 1995 to 1996. He was with the Department of Electrical and Computer Engineering, Villanova University, Villanova, Pa, as a Research Associate from 1996 to 1997. He also served as a Consultant to Comcast, Inc., Philadelphia, Pa, during the same period. In February 1997, he was a Visiting Scientist at the Laboratory for Artificial Brain System, Riken, Japan. From August 1997 to October 1997, he was with Alcatel ETCA, Belgium. From May to June 2001, he was a Visiting Scholar at the Department of Cognitive Science of the University of California at San Diego, USA. From August to September 2001, he was a Visiting Researcher at Brain Signal Processing Laboratory, Riken, Japan. From July to September 2003, he had a sabbatical stay at the Signal and Image Processing Department of École Nationale Supérieure des Télécommunications (ENST), Paris (France). From December 2003 to January 2004, he was an Invited Professor at Telekom Malaysia RnD, Kuala Lumpur, Malaysia. He is currently and since 1998 with the Electrical Engineering Department of ENP as an Associate Professor. His research interests are in statistical signal processing and (blind) array signal processing with applications in biomedical and communications, time-frequency analysis, time-frequency array signal processing, and wireless and spread spectrum communications.

Yves Grenier was born in Ham, Somme, France, in 1950. He received the Ingénieur degree from École Centrale de Paris, in 1972, the Docteur-Ingénieur degree from École Nationale Supérieure des Télécommunications, Paris, in 1977, and the Doctorat d'État es Sciences Physiques, from University of Paris-Sud in 1984. He has been with École Nationale Supérieure des Télécommunications, Paris, since 1977, as Assistant Professor, and since 1984 as Professor. He has been Head of the TSI Department since january 2005. Until 1979, his interests have 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, time frequency representations. He is presently interested in audio signal processing (acoustic echo cancellation, noise reduction, signal separation, microphone arrays, loudspeaker arrays). He is a member of IEEE and AES.