Contribution of Statistical Tests to Sparseness-Based Blind Source Separation

Si Mohamed Aziz Sbaï, Abdeldjalil Aïssa-El-Bey and Dominique Pastor

Institut Télécom; Télécom Bretagne; UMR CNRS 3192 Lab-STICC, Technopôle Brest Iroise CS 83818 29238 Brest, France

Université européenne de Bretagne

Abstract

We address the problem of blind source separation in the underdetermined mixture case. Two statistical tests are proposed to reduce the number of empirical parameters involved in standard sparseness-based underdetermined blind source separation (UBSS) methods. The first test performs multisource selection of the suitable time-frequency points for source recovery and is full automatic. The second one is dedicated to autosource selection for mixing matrix estimation and requires fixing two parameters only, regardless of the instrumented SNRs. We experimentally show that the use of these tests incurs no performance loss and even improves the performance of standard weak-sparseness UBSS approaches.

Index Terms

Underdetermined blind source separation, sparse signals, time-frequency domain, noise variance estimation, weak sparseness, random distortion testing.

I. Introduction

Source separation is aimed at reconstructing multiple sources from multiple observations (mixtures) captured by an array of sensors. In what follows, we assume these sensors to be linear, which is acceptable in many applications. The problem is said to be *blind* when the observations are linearly mixed by the transfer medium and no prior knowledge on the transfer medium or the sources is available. Blind source separation (BSS) is an important research topic in a variety of fields, including radar processing [1], medical imaging [2], communication [3], [4], speech and audio processing [5]. BSS problems can be classified according to the nature of the mixing process (instantaneous, convolutive) and the ratio between the number of sources and the number of sensors of the problem (underdetermined, overdetermined).

If the sources are assumed to be statistically independent, solutions to the BSS problem are calculated so as to optimize separation criteria based on higher order statistics [6], [7]. Otherwise, when the sources have temporal coherency [8], are nonstationary [9], or possibly cyclostationary [10], the separation criteria to optimize are based on second-order statistics.

Although BSS algorithms exist in great profusion, the underdetermined case (UBSS for underdetermined blind source separation), where the number of sensors is smaller than the number of sources, is less addressed than the overdetermined case, where the number of sensors is greater than or equal to the number of sources. Therefore, the UBSS problem is still challenging.

In the UBSS case, one way to deal with the lack of information is to use an Expectation-Maximization-based method [11] to obtain a maximum likelihood estimation of the mixing matrix and sources. However, such an approach requires prior knowledge of the source distributions. In contrast, sparseness-based methods solve the UBSS problem [12]–[20] without prior knowledge on the source distribution, by exploiting the sparseness of the non-stationary sources in the time-frequency domain. Roughly speaking, sparseness-based approaches [21] involve transforming the mixtures into an appropriate representation domain. The transformed sources are then estimated thanks to their sparseness and, finally, the sources are reconstructed by inverse transform. A source is said to be sparse in a given signal representation domain if most of its coefficients, in this domain, are (almost) zero and only a few of them are big.

In the instantaneous mixture case, where each observation consists of a sum of sources with different signal intensity in presence of noise, the sparseness-based methods introduced in [12]–[17], among others, rely on parameters that are chosen empirically. The general question addressed in this paper is then to what extent this empirical parameter choice can be by-passed thanks to statistical methods, specifically designed to cope with sparse representations. This question is particularly relevant because a whole family of sparseness-based UBSS algorithms relies on assumptions very similar to those employed in theoretical frameworks dedicated to the detection and estimation of sparse signals. Our contribution to this question is then the following.

The UBSS algorithms proposed in [12]–[17] estimate the unknown mixing matrix by assuming the presence of only one single source at each time-frequency point. In practice, a selection of time-frequency points that probably pertain to one single source is expected to improve performance of the mixing matrix estimation. The mixing matrix estimate is then used to recover the source signals. Rejecting time-frequency points of noise alone and, thus, selecting and processing the time-frequency points where the possibly multiple sources are present only, should also improve the overall performance of the methods. Our contribution is then to perform the selection processes mentioned in the foregoing, by

considering them as statistical decision problems and reducing the number of empirical parameters for better robustness. Sparseness hypotheses are then particularly suitable for detecting the time-frequency points needed by the separation procedure, whereas such hypotheses are useless for selecting the time-frequency points used by the mixing matrix estimation.

More specifically, Section II recalls the source recovery and mixing matrix estimation steps in classical UBSS methods based on sparseness assumptions. By so proceeding, we highlight the empirical parameters required by these steps. Then, Section III is the main core of the paper because it introduces the statistical tests for the selection of the time-frequency points needed by source recovery and mixing matrix estimation. For source recovery, the selection of the time-frequency points relies on a weak notion of sparseness, exploited through an estimate-and-plug-in detector: We begin by estimating the noise standard deviation via the d-Dimensional Amplitude Trimmed Estimator (DATE), recently introduced in [22], especially designed for coping with noisy representations of weakly-sparse signals; then, the noise standard deviation estimate is used instead of the unknown true value in the expression of a statistical test, specifically designed for noisy representations of weakly-sparse signals as well. For the mixing matrix estimation, the physics of the signal suggest introducing a novel strategy. Indeed, the problem is to select time-frequency points whose energy is big enough in noise to consider that they pertain to one single source. We thus introduce a tolerance above which the energy of these relevant points must be regardless of noise. A statistical test involving this tolerance and based on Signal Norm Testing (SNT) recently introduced in [23] is then used to select these points in presence of noise.

Summarizing, we thus extend significantly [24], by introducing three new features of importance. First, we replace the Modified Complex Essential Supremum Estimate (MC-ESE) of the noise standard deviation by the DATE, which is as accurate, relies on an even stronger theoretical background and has a computational cost significantly lower. Second, the selection of the time-frequency points of interest for source recovery is performed by using a thresholding test, as in [24], but the value of the detection threshold is determined automatically on the basis of the results provided in [25] for the detection of signals satisfying the weak-sparseness model in noise. Third, the mixing matrix estimation is carried out by taking the physical nature of the signals into account.

In Section IV, we apply the statistical tests of Section III to several standard UBSS methods [15], [16], [18], [26], [27] in the instantaneous mixture case. We thus show that our statistical algorithms reduce the number of empirical parameters and improve the overall performance of the UBSS methods under consideration. For instance, by using these statistical algorithms, the subspace-based method presented in [15] can be significantly automatized so as to involve two parameters only. These two parameters are

adjusted once for all possible SNRs, in contrast to standard UBSS methods.

In Section V, these results are discussed. In particular, the convolutive mixture case is addressed for its importance in practice. Some perspectives of this work are then presented in the concluding Section VI.

II. Main steps of standard UBSS methods

A. Principles

We consider the instantaneous mixing system:

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t), \tag{1}$$

where t ranges in some finite set of sampling times such that, for every t in this set of sampling times, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_N(t)]^T$ is the vector of the N sources, $\boldsymbol{x}(t) = [x_1(t), x_2(t), \cdots, x_M(t)]^T$ is the M-dimensional mixture vector, $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_N]$ is the complex $M \times N$ mixing matrix and $\boldsymbol{n}(t) = [n_1(t), n_2(t), \cdots, n_M(t)]^T$ is additive noise. It is assumed that $(n_k(t))_{1\leq k\leq M}$ are random Gaussian processes, mutually decorrelated and independent of the sources. In the sequel, we address the underdetermined case where $N > M$. Without loss of generality, we assume that the column vectors of $\boldsymbol{A}$ have all unit norm, i.e., $\|\boldsymbol{a}_i\| = 1$ for all $i \in \{1, 2, \ldots, N\}$.

Time-frequency signal processing provides effective tools for analyzing nonstationary signals, whose frequency contents vary in time. It involves representing signals in a two-dimensional space, that is, the joint time-frequency domain, hence providing a distribution of the signal energy versus time and frequency simultaneously. The sparseness of the time-frequency coefficients of the source signals is one of the main keys to solve the UBSS problem.

One well-known time-frequency representation and most used in practice is the short-time discrete Fourier transform (STFT). The mixing process can be modeled in the time-frequency domain via the STFT as:

$$\mathcal{S}_{\boldsymbol{x}}(t, f) = \boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t, f) + \mathcal{S}_{\boldsymbol{n}}(t, f) \quad , \tag{2}$$

where $\mathcal{S}_{\boldsymbol{x}}(t, f)$, $\mathcal{S}_{\boldsymbol{s}}(t, f)$ and $\mathcal{S}_{\boldsymbol{n}}(t, f)$) are the vectors of the STFT coefficients at time-frequency bin (t, f) of the mixtures, the sources and noise, respectively.

Given $\boldsymbol{x}(t)$, our purpose is to recover $\boldsymbol{s}(t)$ or equivalently $\mathcal{S}_{\boldsymbol{s}}(t, f)$. As formalized in [28], the UBSS problem is generally decomposed in two separate subproblems. First, in the so called **mixing matrix estimation**, the normalized columns $(\boldsymbol{a}_i)_{1\leq i\leq N}$ are estimated so as to obtain an estimate of $\boldsymbol{A}$. Then,

on the basis of this estimate, the second step called **signal recovery**, provides a solution to equation (2). Figure 1 presents the flowchart of such a two-step approach.

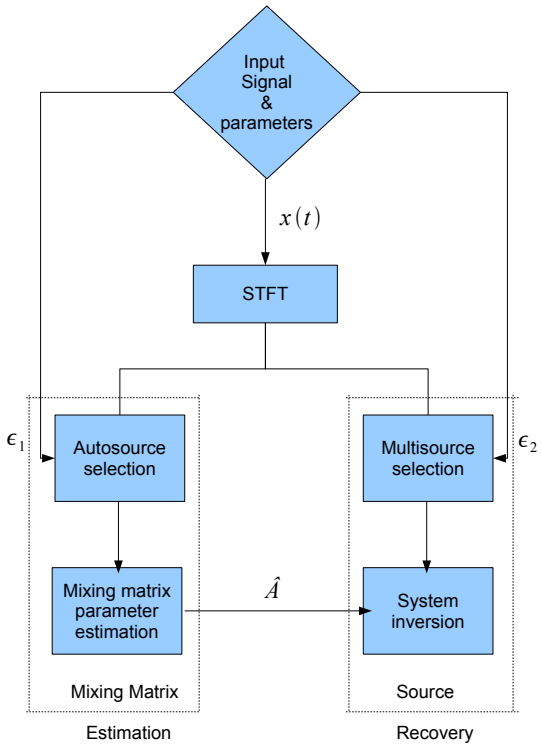


Fig. 1. Flowchart of standard two-step BSS algorithms.

We now detail the mixing matrix estimation and the source recovery based on sparseness assumptions.

B. Mixing matrix estimation

The UBSS methods based on sparse signal representations in the time-frequency domain share the following main assumption:

Assumption 1 *For each source, there exists a set of time-frequency points where this source exists alone.*

The elements of this set can be assumed to be isolated time-frequency points as in DUET (Degenerate Unmixing Estimation Technique) [26] and [15] or to form a time-frequency box as in TIFROM (TIme-Frequency Ratio Of Mixtures) [16] and TIFCORR (TIme-Frequency CORRelation) [27]. Assumption 1 is

often reasonable thanks to the sparseness of the time-frequency representation of the sources, especially when this number of sources is moderate.

As mentioned above, the first step in UBSS methods is to estimate the mixing matrix $\boldsymbol{A}$ to achieve source recovery. In most two-step source separation algorithms [12], [13], [15]–[18] an autosource selection is performed. By autosource selection, it is meant the detection of regions where only one source occurs. The methods for estimating $\boldsymbol{A}$ on the basis of assumption 1 can then be summarized as follows.

Jourjine et al. [26] present the DUET method, which is restricted to two mixtures ($M = 2$). They address the anechoic case, where source transmission attenuations and delays between sensors are taken into account. The columns of the mixing matrix are estimated by finding picks in a $2D$ histogram of amplitude-delay estimates.

In [16], the mixing matrix estimation of the TIFROM method is based on the complex ratios $\frac{\mathcal{S}_{x_j}(t,f)}{\mathcal{S}_{x_k}(t,f)}$, where, given $m \in \{1, 2, \ldots, M\}$, $\mathcal{S}_{x_m}(t, f)$ stands for the m^{th} coordinate of $\mathcal{S}_{\boldsymbol{x}}(t, f)$. These ratios are computed for each time-frequency point and for two arbitrarily chosen indices j and k in $\{1, 2, \ldots, M\}$. A first limitation of this method is to assume non-null matrix coefficients. A second limitation is the use of an empirical threshold to select the smallest empirical variances of these ratios.

In TIFCORR [27], the mixing matrix estimation is similar by selecting the empirical covariance coefficients above a certain threshold chosen manually.

The subspace-based UBSS (SUBSS) method [15] relies on another type of mixing matrix estimation. Let Ω_k stand for the set of all the time-frequency points (t, f) where the k^{th} source is present and Ω stand for the union of all these sets Ω_k for $k = 1, 2, \ldots, N$. According to assumption 1, the sets Ω_k are non-empty and so is Ω. For $(t, f) \in \Omega_k$, (2) reduces to

$$\mathcal{S}_{\boldsymbol{x}}(t, f) = \mathcal{S}_{s_k}(t, f)\boldsymbol{a}_k + \mathcal{S}_{\boldsymbol{n}}(t, f). \tag{3}$$

According to this result, the mixing matrix can be estimated as follows. First, all the spatial direction vectors $\boldsymbol{d}(t, f) = \frac{\mathcal{S}_{\boldsymbol{x}}(t,f)}{\|\mathcal{S}_{\boldsymbol{x}}(t,f)\|}$, with $(t, f) \in \Omega$, are clustered by using an unsupervised clustering algorithm and taking into account that the number of sources is supposed to be known. Since (3) shows that for all the time-frequency points (t, f) of Ω_k, the STFT vectors $\mathcal{S}_{\boldsymbol{x}}(t, f)$ have same spatial direction $\boldsymbol{a}_k$, the column vectors of the mixing matrix $\boldsymbol{A}$ are then estimated as the centroids of the N classes returned by the clustering algorithm. In [15], the authors propose the use of the k-means algorithm but other techniques could be employed. The set Ω required for the clustering procedure is determined by comparing the ratio $\|\mathcal{S}_{\boldsymbol{x}}(t, f)\|/\max_{\xi} \|\mathcal{S}_{\boldsymbol{x}}(t, \xi)\|$ to a threshold height, whose value is chosen empirically.

C. Source recovery

This section presents a number of techniques used in the source recovery stage of two-step UBSS algorithms. In the underdetermined case, the system (2) has less equations than unknowns, and thus it has (in general) infinitely many solutions. In order to recover the original sources, additional assumptions are needed.

The DUET method [26] assumes the sources to be (approximately) W-disjoint orthogonal in the time-frequency domain, that is, the supports of the STFTs of any two sources present in the observations are disjoints. The source recovery is performed by partitioning the time-frequency plane using the mixing parameter estimates. This procedure assigns a source to each time-frequency point, even if this point is due to noise alone, which is detrimental to the method overall performance.

Although TIFROM and TIFCORR do not require the sources to be W-disjoint orthogonal for source recovery, they however suffer from the same limitation as DUET in that they also assign time-frequency points of noise alone to sources.

Bofill and Zibulevsky [18] use the ℓ_1-norm minimization to recover the sources. In the noiseless case, this can be accomplished by solving the convex optimization

$$\min_{\mathcal{S}_{\boldsymbol{s}}(t,f)} \|\mathcal{S}_{\boldsymbol{s}}(t,f)\|_1 \quad \text{subject to} \quad \mathcal{S}_{\boldsymbol{x}}(t,f) = \boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f), \tag{4}$$

where $\|\cdot\|_1$ is the ℓ_1 norm. In presence of noise, the foregoing constraint must be modified so as to take the noise standard deviation into account. In practice, this noise standard deviation is unknown and must be estimated.

For the SUBSS approach in [15], the source recovery is based on the following assumptions:

Assumption 2 *The number of active sources at any (t, f) is strictly less than the number M of sensors.*

Assumption 3 *Any $M \times M$ sub-matrix of the mixing matrix has full rank, that is , for all $J \subset \{1, 2, \cdots, N\}$ with cardinality less than M, $(\boldsymbol{a}_j)_{j\in J}$ are linearly independent.*

The subspace approach then performs multisource selection, that is, the selection of time-frequency points pertaining to a mixture and then, identifies the sources present at a multisource time-frequency points. Thanks to assumption 2, the method then involves solving the resulting locally overdetermined linear problem. By construction, the methods requires rejecting time-frequency points of noise alone. In [15], the time-frequency points with energy below some empirically chosen threshold are rejected.

III. STATISTICAL TESTS FOR SPARSENESS-BASED UBSS

This section is the main core of the paper since it is dedicated to a series of improvements brought to the classical UBSS methods presented in Section II. These improvements concern the selection of the time-frequency points of interest for source separation (multisource selection) and the selection of the time-frequency points suitable for mixing matrix estimation (autosource selection). The crux of the approach followed bellow is to consider the aforementioned selections of time-frequency points as statistical testing problems of accepting or rejecting the presence of sources in noise. These two hypothesis testing problems are different in that mixing matrix estimation requires selecting points where only one single source is present, whereas this constraint is useless for denoising and source recovery.

The issue in these binary hypothesis testing problems is twofold. On the one hand, the observation in each problem has unknown distribution because basically the possible source signal distributions are themselves unknown. On the other hand, the noise standard deviation is unknown as well. Because of this lack of prior knowledge, standard likelihood theory or extensions such as generalized likelihood ratios or invariance-based approaches do not apply.

For source recovery, our solution is an estimate-and-plug-in detector. Based on a weak-sparseness model for the signal sources in noise, it begins by estimating the noise standard deviation via the DATE introduced in [22] . Then, the noise standard deviation estimate is used instead of the unknown true value in the expression of a statistical test, also designed for noisy sparse signal representations.

For mixing matrix estimation, we exploit the physical nature of the signals to detect the time-frequency points where one single source is present. For signals with high overlapping rate, SNT is appropriate to select such time-frequency points. When the signals have low overlapping rate, we directly use the time-frequency points provided by the source recovery procedure.

Figure 2 presents the flowchart of the proposed approach based on the DATE and SNT.

A. *Weak-sparseness-based time-frequency detection for source recovery (multisource selection)*

Recovering sources involves detecting the time-frequency points that pertain to signals. Therefore, time-frequency points due to noise alone are useless to recover sources. Detecting the time-frequency points appropriate for source recovery thus amounts to deciding whether any given time-frequency point (t, f) pertains to some signal of interest or not. It is thus natural to state this problem as the binary hypothesis testing, where the null hypothesis $\mathcal{H}_0$ is that $\mathcal{S}_{\boldsymbol{x}}(t, f) \sim \mathcal{N}_c(0, \sigma^2)$ is complex Gaussian noise and the alternative hypothesis $\mathcal{H}_1$ is that $\mathcal{S}_{\boldsymbol{x}}(t, f) = \Theta(t, f) + \mathcal{S}_n(t, f)$ is a source mixture in independent

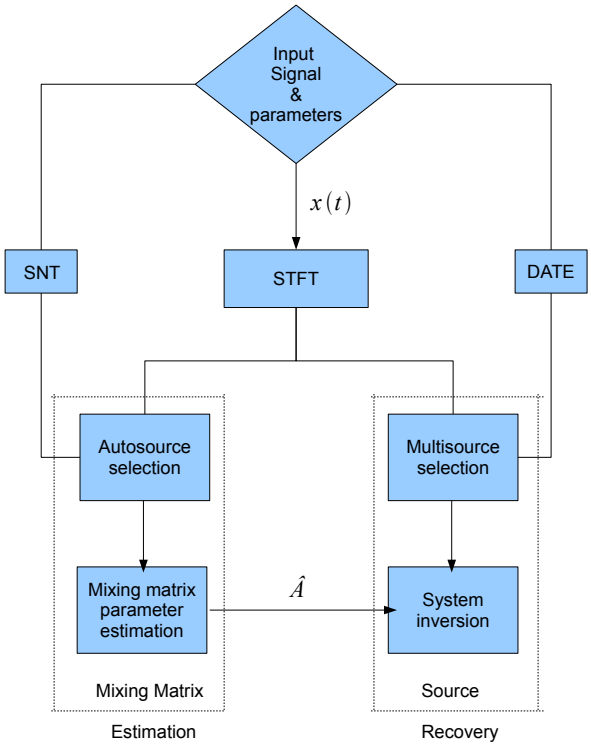


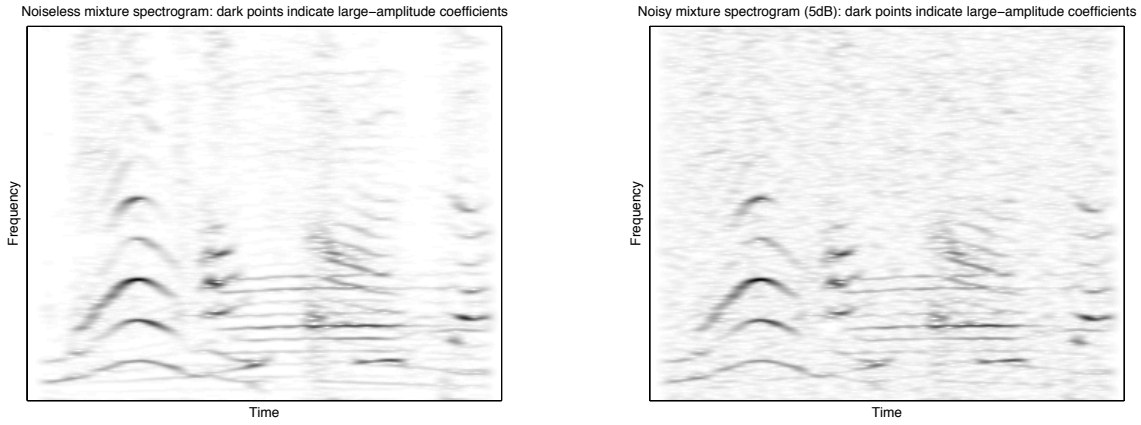
Fig. 2. Flowchart of the proposed two-step BSS algorithms.

and additive complex Gaussian noise, where $\mathcal{S}_n(t,f) \sim \mathcal{N}_c(0,\sigma^2)$ and $\Theta(t,f)$ stands for the mixture of signals possibly present at time-frequency point (t,f).

The issue is then the following. Although $\mathcal{S}_{\boldsymbol{x}}(t,f)$ can reasonably be modeled as a random complex variable, the distribution of $\mathcal{S}_{\boldsymbol{x}}(t,f)$ can hardly be known and standard likelihood theory thus becomes useless. This difficulty can however be overcome by resorting to a weak-sparseness model that can be introduced as follows.

Figure 3-(a) displays the spectrogram obtained by STFT of a mixture of audio signals. This spectrogram exhibits many time-frequency components with small or even null amplitudes. When this mixture is corrupted by additive and independent noise as in Figure 3-(b), small components are masked and only big ones are still visible. We must also note that the proportion of these big components remains seemingly less than or equal to one half. In other words, it is reasonable to assume that 1) the signal components are either present or absent in the time-frequency domain with a probability of presence less than or

equal to one half and 2) when present, the signal components are *relatively big* in that their amplitude is above some minimum value. These two assumptions specify the weak sparseness model by bounding our lack of prior knowledge on the signal distribution. The weak-sparseness model slightly differs from the "strong" sparsity model encountered in compressive sensing, where it is assumed that the non-null significant signal components are very few. In the weak sparseness model, we do not restrict our attention to very small proportions of big time-frequency components.



(a) (b)

Fig. 3. (a) Noiseless audio signal mixture in the time-frequency domain. Many time-frequency coefficients are close to 0. (b) Noisy audio signal mixture in the time-frequency domain. The time-frequency coefficients with small amplitudes are masked by noise. Only big time-frequency coefficients remain visible. They are not really affected by noise as long as the signal to noise ratio is large enough. The proportion of these significant coefficients is less than one half.

To take the weak-sparseness model into account in our binary hypothesis problem statement, we assume that 1) the probability of occurrence of hypothesis $\mathcal{H}_1$ is less than or equal to one half and 2) there exists some positive real value α such that $|\Theta(t,f)| > \alpha$. The value α can be regarded as the minimum signal amplitude. We thus write that

$$\begin{cases} \mathcal{H}_0: \quad \mathcal{S}_{\boldsymbol{x}}(t,f) \sim \mathcal{N}_c(0,\sigma^2) \\ \mathcal{H}_1: \quad \mathcal{S}_{\boldsymbol{x}}(t,f) = \Theta(t,f) + \mathcal{S}_n(t,f), \end{cases} \tag{5}$$

with $\mathcal{S}_n(t,f) \sim \mathcal{N}_c(0,\sigma^2)$, $|\Theta(t,f)| > \alpha$ and $\mathbb{P}(\mathcal{H}_1) \leqslant 1/2$. Furthermore, we do not assume that the probability distribution of $\Theta(t,f)$ is known. In what follows, we prefer summarizing this testing problem by introducing a Bernoulli distributed random variable $\varepsilon(t,f)$, valued in $\{0,1\}$, independent of $\Theta(t,f)$ and $\mathcal{S}_n(t,f)$, but defined on the same probability space, so as to write that $\mathcal{S}_{\boldsymbol{x}}(t,f) = \varepsilon(t,f)\Theta(t,f) + \mathcal{S}_n(t,f)$.

We thus have $\mathbb{P}(\mathcal{H}_1) = \mathbb{P}[\varepsilon(t,f) = 1]$. Given any test $\mathcal{T}$, that is, any measurable map of $\mathbb{C}^M$ into $\{0,1\}$, we then say that $\mathcal{T}$ accepts (resp. rejects) the null hypothesis $\mathcal{H}_0$ if $\mathcal{T}(\mathcal{S}_{\boldsymbol{x}}(t,f)) = 0$ (resp. $\mathcal{T}(\mathcal{S}_{\boldsymbol{x}}(t,f)) = 1$). In other words, $\mathcal{T}$ is said to return the expected value of the true hypothesis. The error probability of $\mathcal{T}$ is then defined as the probability $\mathcal{P}_e\{\mathcal{T}\} = \mathbb{P}[\mathcal{T}(\mathcal{S}_{\boldsymbol{x}}(t,f)) \neq \varepsilon(t,f)]$.

According to [25, Theorem VII.1], the decision should then be performed by using the thresholding test with threshold height $\lambda_D(\alpha,\sigma) = (\sigma/\sqrt{2})\xi(\alpha\sqrt{2}/\sigma)$ where, for any positive ρ, $\xi(\rho) = \mathrm{I}_0^{-1}(e^{\rho^2/2})/\rho$ and I_0 is the zeroth order modified Bessel function of the first kind. By thresholding test with threshold height $h \in [0,\infty)$, we mean the test $\mathcal{T}_h$ such that

$$\mathcal{T}_h(u) = \begin{cases} 1 & \text{if } |u| \geqslant h \\ \\ 0 & \text{if } |u| < h. \end{cases} \tag{6}$$

The reasons for which this test is recommended are the following ones. Let $\mathcal{L}_{\text{MPE}}$ be the Minimum-Probability-of-Error (MPE) test, that is, the likelihood ratio test that guarantees the least possible probability of error among all possible tests and that could be computed if the probability distribution of $\Theta(t,f)$ and the prior probability of presence $\mathbb{P}(\mathcal{H}_1)$ were known. Two facts follow from [25, Theorem VII.1]. First, the error probability of $\mathcal{T}_{\lambda_D(\alpha,\sigma)}$ is above the error probability of the Minimum-Probability-of-Error (MPE) test and less than or equal to the error probability of an explicit function $V(\alpha\sqrt{2}/\sigma)$, whose expression is useless in the sequel. Second, $V(\alpha\sqrt{2}/\sigma)$ is a sharp upper-bound since it is attained by the error probabilities of tests $\mathcal{L}_{\text{MPE}}$ and $\mathcal{T}_{\lambda_D(\alpha,\sigma)}$ in the least favorable case where $\mathbb{P}[\varepsilon = 1] = 1/2$ and $\Theta(t,f) = \alpha e^{\mathrm{i}\Phi(t,f)}$ with $\Phi(t,f)$ uniformly distributed in $[0,2\pi)$ and i is the imaginary unit ($\mathrm{i}^2 = -1$). To carry out this test, we must choose an appropriate value for α and perform an estimate of σ.

The value of α is fixed by following the same reasoning as in [29] and considering that the minimum amplitude of the signal to detect is the noise maximum value. More specifically, given m random variables $X_1, X_2, \ldots, X_m$ that are independent and identically distributed with $X_k \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)$ for $1 \leqslant k \leqslant m$, it is known [30, Eqs. (9.2.1), (9.2.2), Section 9.2, p. 187] [31, p. 454] [32, Section 2.4.4, p. 91] that

$$\lim_{m\to+\infty} \mathbb{P}\left[\lambda_u - \frac{\sigma \ln\ln m}{\ln m} \leqslant \max\{|X_k|, 1 \leqslant k \leqslant m\} \leqslant \lambda_u\right] = 1,$$

where $\lambda_u = \sigma\sqrt{2\ln m}$ is often called the universal threshold [33]. The maximum amplitude of $(X_k)_{1\leqslant k\leqslant m}$ has thus a strong probability of being close to λ_u when m is large and the universal threshold can be regarded as the noise maximum amplitude of m noise samples. In our case, we have M sensors so that each observation $\mathcal{S}_{\boldsymbol{x}}(t,f)$ is an M-dimensional complex vector. Let L stand for the number of time-frequency points (t,f) obtained for each sensor. We thus have $M \times L$ time-frequency points (t,f) and,

therefore, $2ML$ random variables — the real and imaginary parts of $\mathcal{S}_n(t,f)$ — that are $\mathcal{N}(0,\sigma^2/2)$. The maximum amplitude of these $2ML$ Gaussian independent and identically distributed random variables with standard deviation $\sigma/\sqrt{2}$ will then be considered as the minimum signal amplitude so that we set $\alpha = \sigma\sqrt{\log(2ML)}$. The threshold height used to detect the relevant time-frequency points is then $\lambda_D(\sigma) = \sigma\xi(\sqrt{\log(2ML)})$, which is henceforth called the detection threshold.

As far as the estimation of the noise standard deviation is concerned, usual solutions based on standard robust estimators such as the MAD (Median Absolute Deviation) [34] , the trimmed or the winsorized estimators [35] do not apply. Indeed, by considering the spectrogram of Figure 3-(b), it can easily be guessed that such standard estimators would fail because the proportion of significant noisy time-frequency points pertaining to the signals is large. Therefore, the noisy time-frequency points are not very few and cannot play the role of outliers with respect to the main core data distribution. In a recent paper [22], a new noise standard deviation estimator called the DATE has been proposed. This estimator relies on the weak-sparseness model presented before. An exhaustive presentation of the theoretical background on which this estimator is based is beyond the scope of the present paper and the reader is asked to refer to [22] for an heuristic presentation and a complete mathematical description of the DATE. In the context addressed in the present paper, this algorithm applies as follows.

With the notation used so far, each $\mathcal{S}_{\boldsymbol{x}}(t,f)$ is an M-dimensional complex vector. Let $\mathcal{S}_{x_j}(t,f)$, $j = 1,2,\ldots,M$, be the components of $\mathcal{S}_{\boldsymbol{x}}(t,f)$. For any given $j = 1,2,\ldots,M$, we assume that the L time-frequency components $\mathcal{S}_{x_j}(t,f)$ for the j^{th} sensor are independent and that each time-frequency component obeys the binary hypothesis model of (5) with $\alpha = \sigma\sqrt{\log(2ML)}$. According to [22] and setting $\kappa = \sqrt{2}\Gamma\left(3/2\right)$ where Γ is the standard Gamma function, there exists a specific convergence criterion, for which we have:

$$\frac{\displaystyle\sum_{(t,f)} |\mathcal{S}_{x_j}(t,f)|\,\mathbb{1}(|\mathcal{S}_{x_j}(t,f)| \leqslant \lambda_D(\sigma))}{\displaystyle\sum_{(t,f)} \mathbb{1}(|\mathcal{S}_{x_j}(t,f)| \leqslant \lambda_D(\sigma))} \approx \kappa\sigma \tag{7}$$

when the number L of time-frequency bins (t,f) is large enough. In the previous equation, $\mathbb{1}(|\mathcal{S}_{x_j}(t,f)| \leqslant \lambda_D(\sigma))$ stands for the indicator function of event $|\mathcal{S}_{x_j}(t,f)| \leqslant \lambda_D(\sigma)$, The specific convergence criterion involved in (7) is specified in [22] and is not given here because of its intricateness. It also turns out that the noise standard deviation σ is the unique solution of (9) with respect to the convergence criterion involved. Therefore, the DATE basically performs an estimate of the noise standard deviation by solving (7) with regard to this convergence criterion. The several steps involved in the computation are then the following ones.

The DATE:

Given $j \in \{1, 2, \ldots, M\}$, let $Y^j_{(1)}, Y^j_{(2)}, \ldots, Y^j_{(L)}$ be the L values $|\mathcal{S}_{x_j}(t, f)|$ sorted by ascending order.

1) **[Search interval]**:

   a) Choose some positive real value Q less than or equal $1 - \frac{L}{4(L/2-1)^2}$.

   b) Set $h = 1/\sqrt{4L(1-Q)}$

   c) Compute $\min = L/2 - hL$. According to Bienaymé-Chebyshev's inequality and since the probabilities of presence of the signals are assumed to be less than or equal to one half, the probability that the number of observations due to noise alone is above $k_{\min}$ is larger than or equal to Q. In the experimental results presented below, Q was set to 0.95 for the computation of $k_{\min}$.

2) **[Existence]**:

   IF there exists a smallest integer k in $\{k_{\min}, \ldots, L\}$ such that

$$|Y^j_{(k)}| \leqslant \Big(\mu_j(k)/\kappa\Big)\xi\left(\sqrt{\log(2ML)}\right) < |Y^j_{(k+1)}| \tag{8}$$

   with

$$\mu_j(k) = \begin{cases} \frac{1}{k}\sum_{r=1}^{k} |Y^j_{(r)}| & \text{if } k \neq 0 \\ 0 & \text{if } k = 0, \end{cases} \tag{9}$$

   set $k^* = k$.

   ELSE, set $k^* = k_{\min}$.

3) **[Value]**: The estimate σ^*_j of the noise standard deviation on the j^{th} sensor is then

$$\widehat{\sigma_j} = \mu_j(k^*)/\kappa, \tag{10}$$

The final estimate $\widehat{\sigma}$ of the noise standard deviation is then obtained by averaging the values $\widehat{\sigma_j}$ so that $\widehat{\sigma} = (1/M)\sum_{j=1}^{M} \widehat{\sigma_j}$.

B. Signal source detection for mixing matrix estimation (autosource selection)

In this section, we propose a test for selecting the time-frequency points where one signal source is probably present alone. To perform this selection, we make the distinction between signals with either low or high overlapping rate in the time-frequency domain. Chirp signals (resp. audio signals) are typical examples of signals with low (resp. high) overlapping rate. It is worth noticing that the estimation procedures proposed below for each class have reasonable computational costs.

1) The case of signals with low overlapping rate: Since the sources have low overlapping rate, we suppose that the observations detected by the thresholding test of Section III-A mostly pertain to one signal source. In other words, we neglect the effect on the matrix estimation performance of the few points where sources may overlap, inasmuch as the impact of such time-frequency points is further reduced by the averaging effect inherent to any mixing matrix estimation method.

2) The case of signals with high overlapping rate: When signals overlap significantly in the time-frequency domain, the time-frequency detection of Section III-A is now inappropriate. Indeed, the statistical procedure of Section III-A is aimed at detecting time-frequency points where signal sources are present, whatever the number of these sources, whereas it is now required to discriminate points where one single source is present from points where multiple sources occur. We assume that in case of different sources present at time-frequency point (t, f), they are uncorrelated and incoherently combined. The resulting energy at (t, f) is thus supposed to be smaller than the energy attained at the time-frequency points where one single source is present only.

Our purpose is thus to detect the time-frequency points where the signal energy is big enough in presence of noise. Basically, this problem amounts to deciding whether $|\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t, f)|$ is above some value τ or not. The value τ^2 thus represents the minimum energy level above which we consider that the signal energy is big enough to assume that one single source is actually present at (t, f). For any $\lambda \in (0, \infty)$, it follows from [23, Lemma 4, statement (iii)] that

$$\mathbb{P}\Big[\, |\mathcal{S}_{\boldsymbol{x}}(t, f)| > \lambda \,\Big|\, |\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t, f)| < \tau \,\Big] \leqslant 1 - F_{\chi^2_{2M}(\tau^2/\sigma^2)}\big(\lambda^2/\sigma^2\big)\,, \tag{11}$$

where $F_{\chi^2_d(\delta)}(\cdot)$ stands for the cumulative distribution function of the non-centered chi-2 distribution with d degrees of freedom and non-centrality parameter δ. The degree of freedom in (11) is $2M$ since each $\mathcal{S}_{\boldsymbol{x}}(t, f)$ is an M-dimensional complex random vector and, thus, a $2M$-dimensional real random vector. Given some level $\gamma \in (0, 1)$, it then suffices to choose

$$\lambda = \lambda(\tau, \gamma) = \sigma \sqrt{F^{-1}_{\chi^2_{2M}(\tau^2/\sigma^2)}(1 - \gamma)}. \tag{12}$$

to guarantee a "false alarm probability" $\mathbb{P}\Big[\, |\mathcal{S}_{\boldsymbol{x}}(t, f)| > \lambda \,\Big|\, |\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t, f)| < \tau \,\Big]$ less than or equal to γ.

Therefore, for a given time-frequency point (t, f), the decision is that $|\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t, f)| < \tau$ if $|\mathcal{S}_{\boldsymbol{x}}(t, f)| < \lambda(\tau, \gamma)$ and that $|\boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t, f)| \geqslant \tau$ if $|\mathcal{S}_{\boldsymbol{x}}(t, f)| \geqslant \lambda(\tau, \gamma)$. For mixing matrix estimation, we then keep the time-frequency points (t, f) such that $|\mathcal{S}_{\boldsymbol{x}}(t, f)| \geqslant \lambda(\tau, \gamma)$, which are considered as to time-frequency points pertaining to one single source. In practice, since the actual value of σ is unknown, we replace this true value by its estimate $\widehat{\sigma}$ provided by the DATE.

Although the two parameters γ and τ must be fixed, there is no need to choose them for each signal to noise ratio. Parameter τ, which is independent of the noise level, can be fixed via a small noiseless database. Similarly, level γ can be determined via a few preliminary test on a small representative database.

IV. Simulation results

In most of the following simulations, the mixing matrix is chosen according to [14, Eq. (38)] so as to model N sources arriving at the sensor array at different angles $\theta_1, \theta_2, \ldots, \theta_M$. The entries of matrix $\boldsymbol{A}$ are therefore $a_{j,k} = e^{\mathrm{i}\pi(j-1)\sin(\theta_k)}$ for $j \in \{1, \cdots, M\}$ and $k \in \{1, \cdots, N\}$. In the sequel, we proceed by choosing four sources ($N = 4$), three sensors ($M = 3$), $\theta_1 = 15\,^\circ$, $\theta_2 = 30\,^\circ$, $\theta_3 = 45\,^\circ$ and $\theta_4 = 75\,^\circ$.

Unless specified otherwise, the source signals are speech signals randomly chosen in the TI-digits database [36] . This large speech database collected in a quiet environment is commonly used in speech processing. In this paper, the chosen speech signals were downsampled to 8 kHz. All signals involve 8192 samples. In Figure 4, the left four subplots (a)-(d) show the time-domain representations of the original source signals and the right four subplots (e)-(h) represent their corresponding spectrograms. Figure 5 displays a spectrogram of a mixture of these speech signals when the mixing matrix $\boldsymbol{A}$ is applied to them at SNR $= 10$dB. The spectrograms of the other mixtures are not presented because the differences between any two of them are not visually noticeable since the mixing matrix $\boldsymbol{A}$ involves no null entry.

The two parameters required for the mixing matrix estimation are then fixed to $\tau = 4$ and $\gamma = 10^{-3}$. The source separation performance is measured by the normalized mean square error (NMSE):

$$\text{NMSE} = \min_{i,j} \left\{ 10 \log_{10} \left(1 - \left(\frac{\langle \hat{s}_i, s_j \rangle}{\|\hat{s}_i\| \cdot \|s_j\|} \right)^2 \right) \right\}. \tag{13}$$

Throughout this section, NMSEs are calculated over 100 Monte-Carlo runs.

A. SUBSS method

The modified SUBSS algorithm is obtained by using both the DATE and SNT for source recovery and mixing matrix estimation by SNT, respectively, as explained in Section III. It is used to separate the four source signals from the noisy mixed signals observed by the three sensors.

The waveforms of the recovered source signals by the modified SUBSS algorithm are represented in Figure 6. The left four subplots (a)-(d) show the time-domain representations of the recovered source signals in the noiseless case (input SNR = 45dB), and the right four subplots (e)-(h) represent time-domain representations of the recovered source signals with input SNR = 10dB.

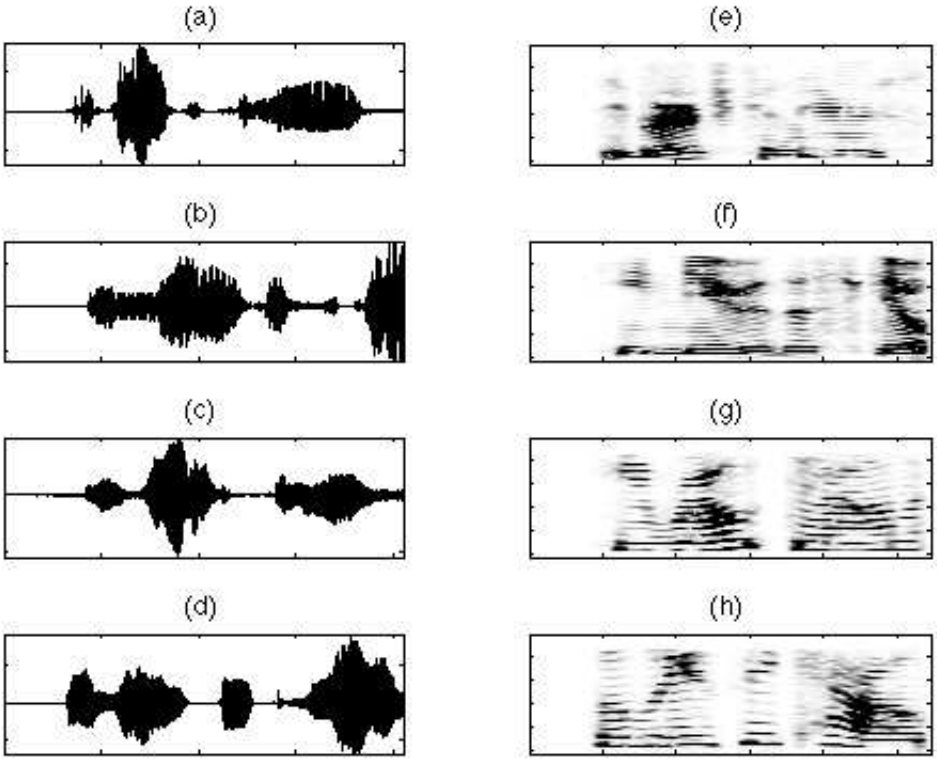


Fig. 4. (a)-(d) show the waveforms of the original source signals in the time domain, (e)-(h) display the spectrograms of these source signals in the time-frequency domain.

In Figure 7, the performance of the modified SUBSS algorithm, with and without denoising, is compared to that obtained by the originally SUBSS algorithm of [15]. The denoising mentioned above is described in appendix A as a standard linear estimation.

The modified SUBSS algorithm outperforms the original SUBSS algorithm [15], which relies on thresholds that are manually chosen for each input SNR. Moreover, modified SUBSS without denoising yields performance measurements that do not significantly depart from those attained by the original subspace-based UBSS algorithm. In addition, Figure 7 displays the NMSEs obtained by using the MAD estimator instead of the DATE in the modified SUBSS algorithm without denoising. The use of the MAD instead of the DATE induces a significant performance loss, which illustrates the relevance of the DATE and the weak-sparseness model. In Figures 8 and 9, we present the NMSEs obtained by the modified

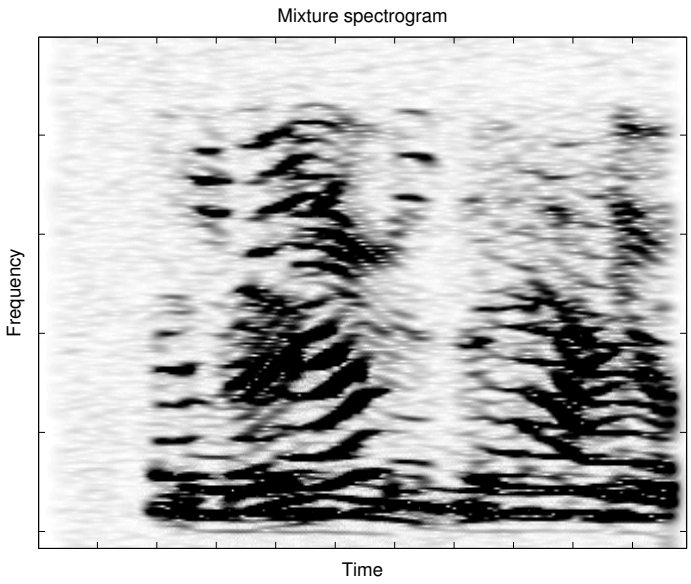


Fig. 5. Speech mixture spectrogram when mixing matrix $\boldsymbol{A}$ is applied to the four sources of Figure 4 (SNR $=$ 10dB).

SUBSS and the original SUBSS when the number of sources increases and for SNR = 10dB and SNR = 20dB. In both figures, the NMSEs degrade, because an increase of the source interference invalidates assumption 1.

We now consider the case of complex chirp signals. These ones were generated by slightly modifying the MATLAB routine *MakeSignal.m* of the WAVELAB toolbox, so as to obtain complex chirp signals. The 4 chirp signals we use as sources are $s_1(t) = \sqrt{t(1-t)}e^{\mathrm{i}\frac{\pi \mathrm{T}}{2}t^2}$, $s_2(t) = \sqrt{t(1-t)}e^{-\mathrm{i}\frac{\pi \mathrm{T}}{4}t^2}$, $s_3(t) = e^{-\mathrm{i}\pi \mathrm{T}t^2}$ and $s_4(t) = e^{\mathrm{i}\frac{2}{3}\pi \mathrm{T}t}$, where $t \in [0,1]$ and $\mathrm{T} = 8192$ is the number of samples for each signal. Two of these chirp signals are LFM ones and one is a pure sine. Figure 10 then displays the spectrograms of the four chirp signals under consideration, whereas Figure 11 presents the spectrogram of a mixture of these sources when matrix $\boldsymbol{A}$ is applied and SNR$=$ 10dB. The spectrograms of the other mixtures are not displayed for the same reasons as those given previously for the speech signal mixtures.

The experimental procedure for assessing the modified SUBSS in comparison to the original SUBSS method is then the same as above. As specified in Section III-B1, the thresholds used for the mixing matrix estimation are the detection ones. Therefore, no additional parameter is needed. The results obtained in Figure 12 show the relevance of this choice for the thresholds, explained by the fact that chirp signals present very few overlapping time-frequency components.

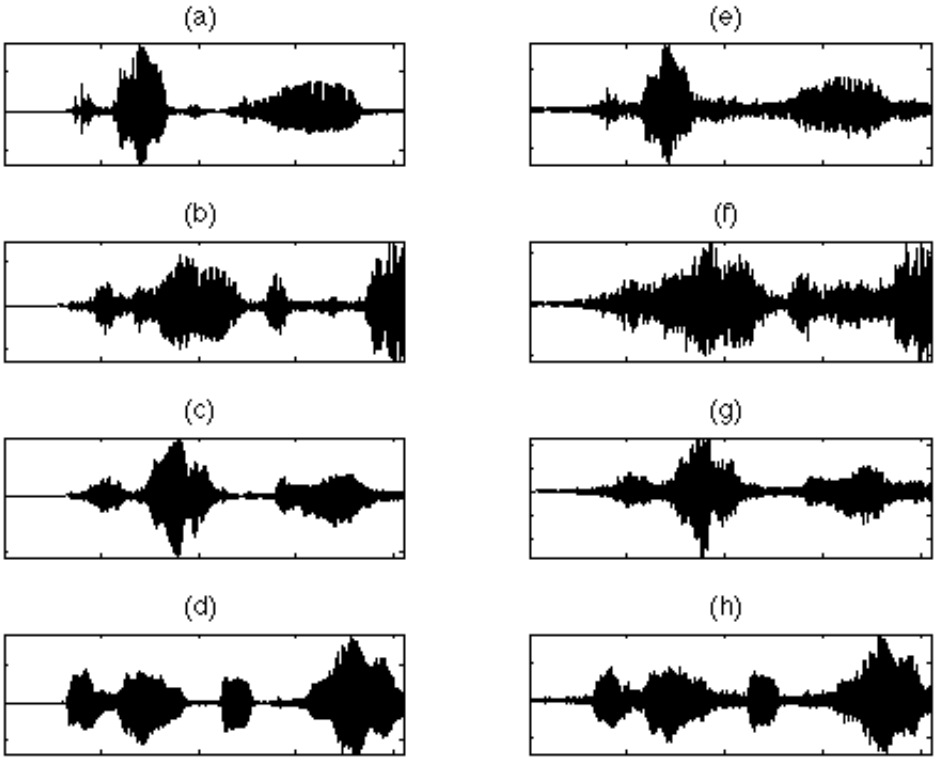


Fig. 6. Simulation results: (a)-(d) show the waveforms of the source signals recovered by modified SUBSS with input SNR=45dB, (e)-(h) show the waveforms of the source signals recovered by modified SUBSS with input SNR=10dB

B. Other methods

As described in Sections III-A and III-B, The DATE and SNT can be used to perform multisource and autosource selections, respectively. Said otherwise, the statistical tests of the aforementioned sections make it possible to obtain the time-frequency points where noisy mixtures are present and the set of time-frequency points where only one single source exists. In this subsection, we comment the results we obtain by so proceeding with respect to the several UBSS methods addressed in Section II and other than SUBSS.

In the underdetermined case, TIFROM achieves partial source separation only. Therefore, to better assess the contribution of our statistical tests to TIFROM, we consider the determined case where four source signals from four speakers are mixed. The mixing matrix is now 4×4 with independent Gaussian

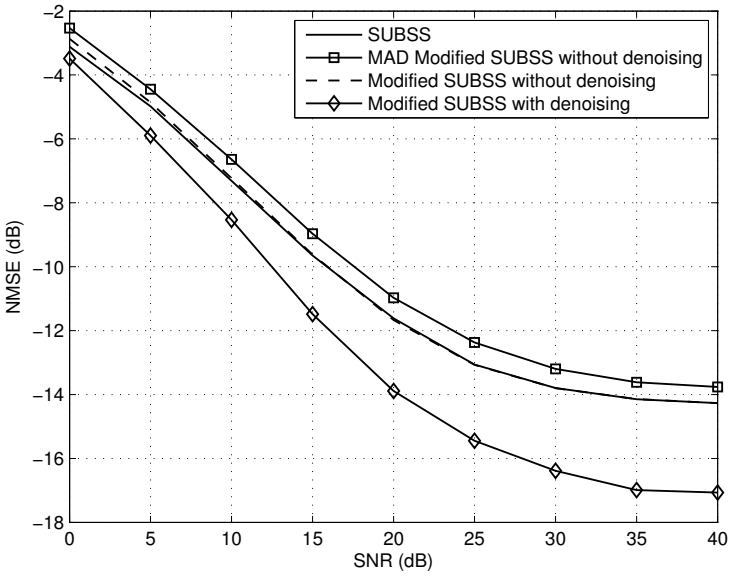


Fig. 7. Comparison between SUBSS, modified SUBSS with and without denoising, modified SUBSS with MAD estimate instead of DATE and without denoising: NMSE versus SNR.

entries. In Figure 13, we present the NMSEs obtained by the TIFROM, SNT-TIFROM and Modified SNT-TIFROM. Specifically, SNT-TIFROM uses SNT to select times frequency points where a source exists alone. SNT-TIFROM, as TIFROM, performs no multisource selection for source recovery. In contrast, the modified SNT-TIFROM performs multisource selection and forces to zero the unselected time-frequency points. These results show that SNT makes it possible to actually select the autosource time-frequency points, with no performance loss and without resorting to the empirical threshold required by the original TIFROM. The performance yielded by the modified SNT-TIFROM further emphasizes that the detection threshold adjusted with the DATE selects appropriate multisource time-frequency points for source recovery. The gain for low SNRs is explained by the fact that this selection can be regarded as a non-linear denoising. The gain brought by this denoising effect decays when the SNR increases.

Another contribution of our statistical approach to sparseness-based methods is the estimation of the noise standard deviation. Indeed, some methods need an estimate or the true value of the noise standard deviation. For instance, Bofill and Zibulevsky, in [18], use the ℓ_1-norm minimization to recover the sources. In the noisy case, they propose to solve the optimization problem:

$$\min_{\mathcal{S}_{\boldsymbol{s}}(t,f)} \frac{1}{2\sigma^2} \left\| \mathcal{S}_{\boldsymbol{x}}(t,f) - \boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f) \right\|_2^2 + \left\| \mathcal{S}_{\boldsymbol{s}}(t,f) \right\|_1 .$$

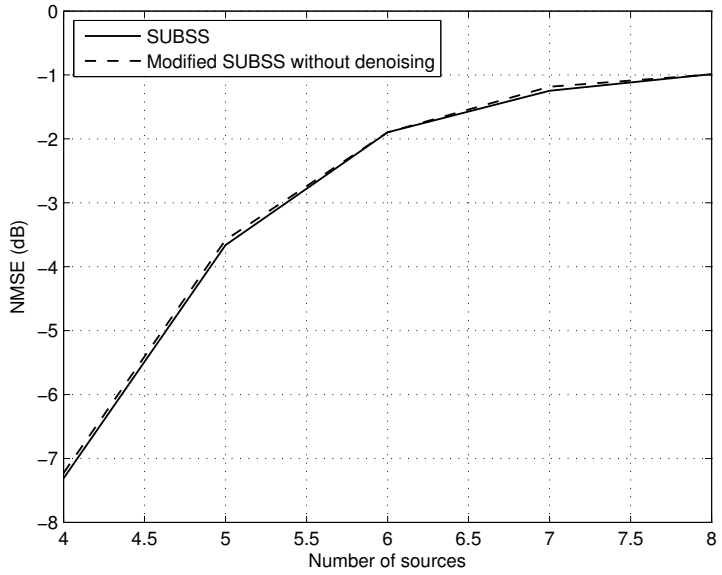


Fig. 8. Comparison between SUBSS and modified SUBSS without denoising when input SNR = 10dB: NMSE versus number of sources.

Because of the weakly sparseness of the sources in noise, we hereafter prefer following [37] dedicated to stable recovery of not exactly sparse signals. We therefore solve the optimization problem

$$\min_{\mathcal{S}_{\boldsymbol{s}}(t,f)} \|\mathcal{S}_{\boldsymbol{s}}(t,f)\|_1 \quad \text{subject to} \quad \|\mathcal{S}_{\boldsymbol{x}}(t,f) - \boldsymbol{A}\mathcal{S}_{\boldsymbol{s}}(t,f)\|_2 \leq \sigma^2(M + 2\sqrt{2M}). \tag{14}$$

This approach can then be improved in two ways. First, by solving this optimization problem on only the time-frequency points selected by the multisource procedure propounded in Section III-A. Second, by replacing the unknown true value of the noise standard deviation by its estimate provided by the DATE. In this respect, Figure 14 displays the performance measurements obtained by the original method based on the ℓ_1-criterion of Eq. (4) (L1 Minimization) in comparison to the modified ℓ_1-criterion of Eq. (14) applied to the outcome of the the multisource selection when the noise standard deviation is estimated by the DATE (Modified L1 minimization). As expected, the gain brought by multisource selection and Eq. (14), both adjusted by the noise standard deviation estimate provided by the DATE, is significant. It is also worth noticing that the DATE estimation error does not impact significantly the separation performance in comparison to the case where the noise standard deviation is perfectly known. This can also be seen in Figure 14, where the performance measurements are given when the multisoure selection and ℓ_1-criterion of Eq. (14) are both adjusted with the actual value of the noise standard deviation (Oracle Modified L1

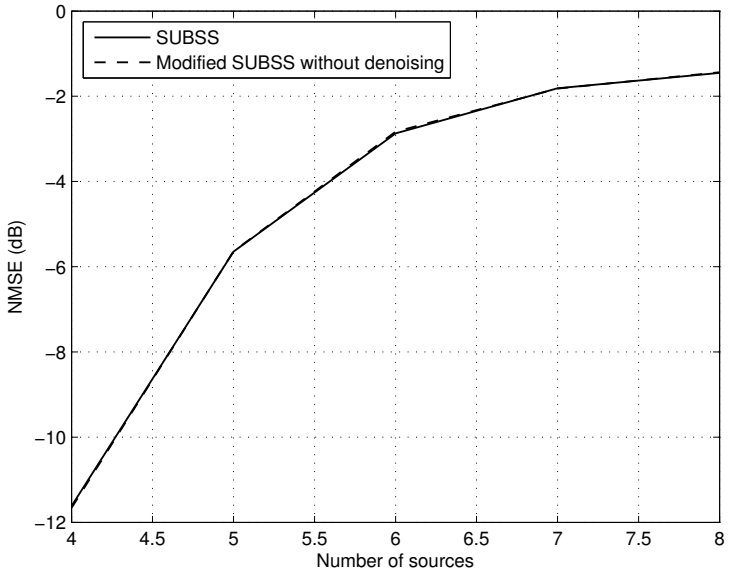


Fig. 9. Comparison between SUBSS and modified SUBSS without denoising when input SNR = 20dB: NMSE versus number of sources.

Minimization). In contrast, there is significant performance loss when the multisource selection and Eq. (4) are calculated by using the MAD instead of the DATE (MAD Modified L1 Minimization). The reason still relates to the fact that the DATE is more robust to weak-sparseness than the MAD.

The multisource selection based on the detection threshold adjusted by the estimate provided by the DATE can be further exploited by the DUET reconstruction, as illustrated in Figure 15. In this simulation, the input signals are the chirp signals considered above, so that the W-disjoint orthogonality assumption is satisfied. Moreover, the mixing matrix $\boldsymbol{A}$ is now assumed to be known. On the one hand, we perform the DUET source recovery by considering the whole time-frequency plane. On the other hand, we consider the modified DUET, that is, the DUET source recovery applied to the selected multisource time-frequency points only. The results are similar to those obtained above by TIFROM and its modified versions. Here, the gain brought by the multisource selection, which acts as a denoising, is bigger on a wider SNR range because the time-frequency representation of chirp signals is sparser than that of audio signals.

V. Discussion

A. Assessment

The algorithms we propose are very general. They are not dedicated to a given sparseness-based BSS

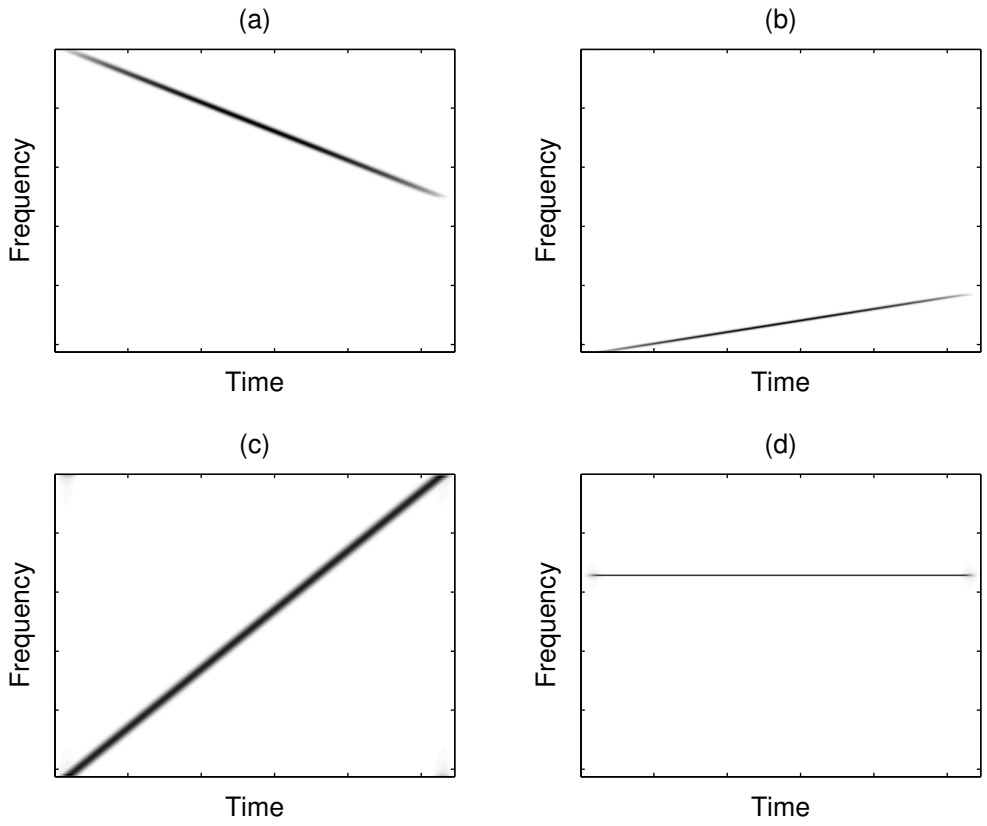


Fig. 10. Spectra of 4 chirp signals used as sources.

method. They are simple to apply without any adjustment. From the results of Section IV, our procedures can therefore be used to improve, simplify or bring robustness to the standard sparseness-based BSS methods considered in the paper.

More specifically, the weak-sparseness-based time-frequency detection procedure of Section III-A can be used as an automatized pre-processing for multisource selection. For example, the time-frequency detection in [15] requires one threshold value for each instrumented SNR. The detection procedure of Section III-A then makes it possible to avoid this empirical parameter choice, which brings robustness and significant simplification. Used as a pre-processing for TIFROM [16], which basically involves no selection of time-frequency points, the multisource selection we propound can improve the separation performance.

For mixing matrix estimation, our approach described in Section III-B relies on no weak-sparseness

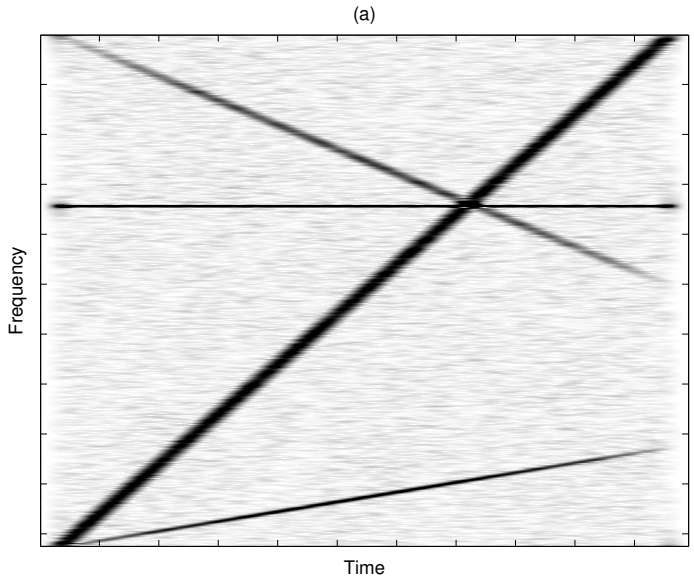


Fig. 11. Chirp signal mixture spectrogram when mixing matrix $\boldsymbol{A}$ is applied to the chirp signals of Figure 10 (SNR= 10dB).

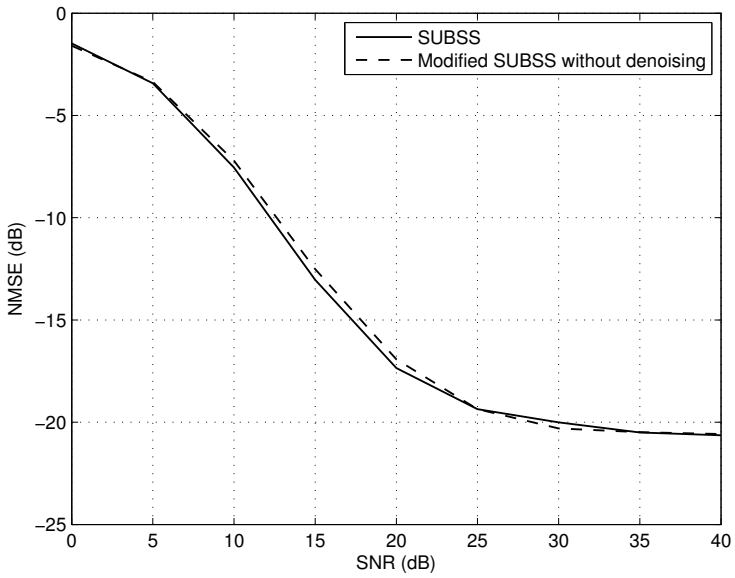


Fig. 12. Comparison of performance between SUBSS and modified SUBSS without denoising for chirp signals: NMSE versus SNR.

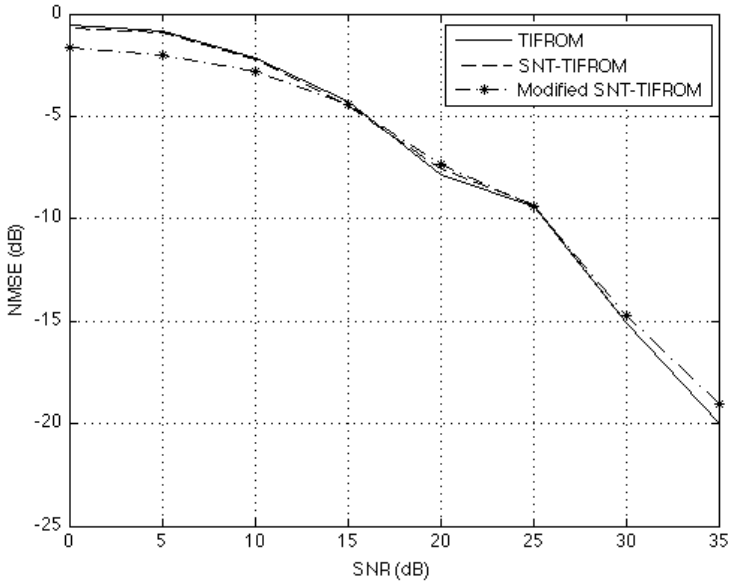


Fig. 13. Comparison of performance between TIFROM, SNT-TIFROM and Modified TIFROM: NMSE versus SNR.

assumption and involves two parameters only, that is, the tolerance and the false-alarm probability. These parameters are valid over the signal-to-noise ratio (SNR) range, in contrast to [15] for instance. Furthermore, the assumptions made by TIFROM can be relaxed by using the autosource selection of Section III-B. It is also worth noticing that the two parameters we need for mixing matrix estimation have a physical meaning, which is not the case for some standard sparseness-based BSS methods.

B. Convolutive mixture case

There exists a great variety of possible strategies for dealing with the convolutive mixture case, which is more realistic than the instantaneous one. In the convolutive mixture case, exhibiting a well-established family of methods such as that considered above in the instantaneous mixture one is hardly feasible. However, despite this variety, the statistical framework proposed in this paper can be expected to be used in the convolutive mixture case, at least, for methods based on time-frequency representations for which, separating time-frequency points of noise alone from those of noisy signals can be helpful. For instance, this detection procedure for multisource selection can be used straightforwardly to detect the time-frequency points required by the convolutive SUBSS presented in [38]. The modified convolutive SUBSS thus obtained discards the empirical threshold required in [38] for multisource selection. This

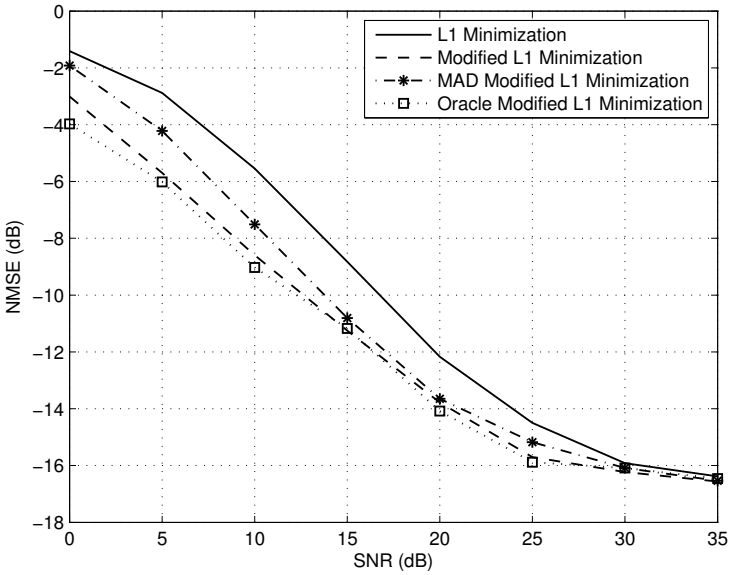


Fig. 14. Comparison of performance (NMSE versus SNR) between the original Bofill and Zibulevsky's method based on the ℓ_1-criterion of Eq. (4) (L1 Minimization), the modified ℓ_1-criterion of Eq. (14) after multisource selection when: the noise standard deviation is known (Oracle Modified L1 Minimization) or estimated via either the DATE (Modified L1 Minimization) or the MAD (MAD Modified L1 Minimization).

entails no significant performance loss, as illustrated by Figure 16. Studying the added-value brought by SNT in the convolutive mixture case requires further analysis that could be achieved in some forthcoming work.

VI. Conclusion and perspectives

The algorithms presented in this paper contribute to blind source separation in the underdetermined mixture case, by avoiding empirical choices of parameters present for the so-called family of weak-sparseness based methods. Our first algorithm aimed at selecting the suitable time-frequency points for source recovery is full automatic. The second, dedicated to mixing matrix estimation, requires fixing two parameters only, regardless of the instrumented SNRs.

The question is now to what extent the statistical tests used above in the instantaneous mixture case can possibly be exploited in the convolutive mixture case, especially in complement to the results discussed in Section V-B. It can also be wondered whether these tests can be extended so as to deal with colored noise. Work on this topic is under progress.

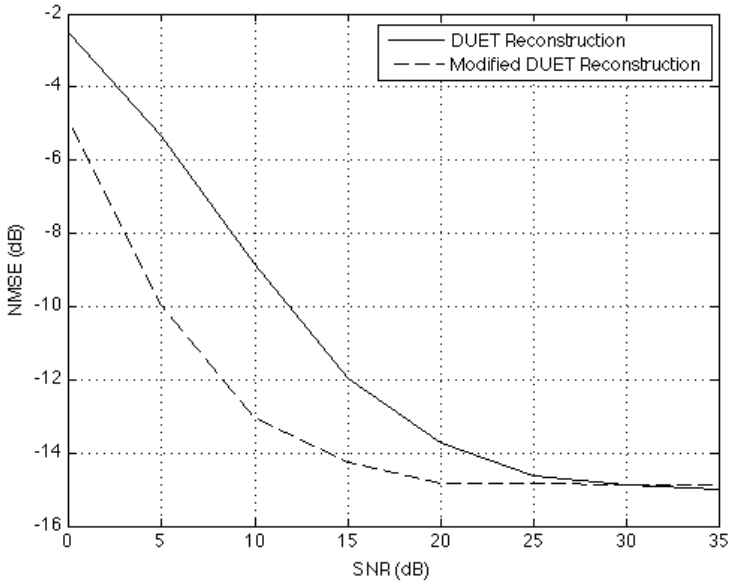


Fig. 15. Comparison of performance between DUET reconstruction and Modified DUET reconstruction on chirp signals

The theoretical and experimental results of this paper pinpoint that the subfunctions of the source separation methods considered above, completed with the statistical tests we have proposed, can be regarded as elementary components that can be interchanged and associated to provide new algorithms for source separation in different applicative contexts. This opens new practical prospects. For instance, it would be desirable to construct a toolbox involving all these elementary components for further developments and studies. Such a toolbox would also make it possible to carry out exhaustive experimental assessments on large databases of signals via the BSSEval toolbox, downloadable from [39].

APPENDIX A
DENOISING-BASED SOURCE RECOVERY

The SUBSS method presented in [15] estimates the index set of the sources present at a given time-frequency point (t, f). Let us denote by J this set of indexes. Then, equation (2) reduces to:

$$\mathcal{S}_{\boldsymbol{x}}(t,f) = \boldsymbol{A}_J \mathcal{S}_{\boldsymbol{s}_J}(t,f) + \mathcal{S}_{\boldsymbol{n}}(t,f) \tag{15}$$

and the STFT coefficients of these active sources can be recovered using:

$$\mathcal{S}_{\boldsymbol{s}_J}(t,f) \approx \boldsymbol{A}_J^{\#} \mathcal{S}_{\boldsymbol{x}}(t,f), \tag{16}$$

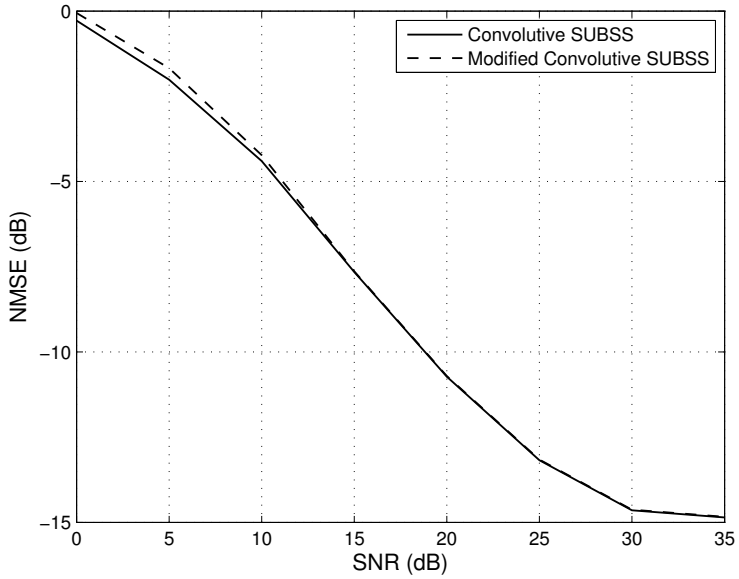


Fig. 16. Comparison of performance between standard convolutive SUBSS and modified convolutive SUBSS: the signals used are same audio one as those considered in Simulation Section. Each mixture is a sum of filtered source signal where each filter is randomly chosen RIF with order 4.

where $\boldsymbol{A}_J^{\#} = (\boldsymbol{A}_J^H \boldsymbol{A}_J)^{-1} \boldsymbol{A}_J^H$ is the Moore-Penrose pseudoinverse of $\boldsymbol{A}_J$.

We propose to use the noise standard deviation estimate provided by the DATE to jointly denoise and separate the sources on the basis of the time-frequency points selected by the statistical test of Section III-A. So, instead of performing the source separation as specified by Eq. (16), the source separation is now carried out by computing

$$\widehat{\mathcal{S}}_{\boldsymbol{s}_J}(t,f) = \boldsymbol{R}_{\boldsymbol{s}_J} \boldsymbol{A}_J^H (\boldsymbol{A}_J \boldsymbol{R}_{\boldsymbol{s}_J} \boldsymbol{A}_J^H + \widehat{\sigma}^2 \boldsymbol{I}_M)^{-1} \mathcal{S}_{\boldsymbol{x}}(t,f) \tag{17}$$

where $\widehat{\sigma}$ is the noise standard estimate returned by the DATE and $\boldsymbol{R}_{\boldsymbol{s}_J} = \mathbb{E}[\mathcal{S}_{\boldsymbol{s}_J}(t,f) \mathcal{S}_{\boldsymbol{s}_J}^H(t,f)]$. The derivation of the optimal linear estimator of (17) is standard. It involves minimizing the risk $\mathbb{E}\left[\|\mathcal{S}_{\boldsymbol{s}_J}(t,f) - \boldsymbol{D}\mathcal{S}_{\boldsymbol{x}}(t,f)\|^2\right]$ when $\boldsymbol{D}$ ranges over the space of the $\mathrm{card}(J) \times M$ matrices and under the assumption that the sources are spatially decorrelated. In practice, matrix $\boldsymbol{R}_{\boldsymbol{s}_J}$ is unknown and must be estimated. We then proceeded as follows. On the one hand, we have $\boldsymbol{R}_{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_M$. On the other hand, $\boldsymbol{R}_{\boldsymbol{x}}$ can be estimated by $\widehat{\boldsymbol{R}_{\boldsymbol{x}}} = \frac{1}{\#t} \sum_t \mathcal{S}_{\boldsymbol{x}}(t,f) \mathcal{S}_{\boldsymbol{x}}(t,f)^H$, where $\#t$ stands for the number of time windows on which the STFT is calculated. Since estimates of $\boldsymbol{A}$ and σ are known, we derive from the expressions of $\boldsymbol{R}_{\boldsymbol{x}}$ and $\widehat{\boldsymbol{R}_{\boldsymbol{x}}}$ an estimate $\widehat{\boldsymbol{R}_{\boldsymbol{s}}}$ of $\boldsymbol{R}_{\boldsymbol{s}}$. An estimate of $\boldsymbol{R}_{\boldsymbol{s}_J}$ follows by picking the

appropriate columns in $\widehat{\boldsymbol{R}_{\boldsymbol{s}}}$.

REFERENCES

[1] V. Varajarajan and J. Krolik, "Multichannel system identification methods for sensor array calibration in uncertain multipath environments," in *IEEE Signal Processing Workshop on Statistical Signal Processing (SSP)*, Singapore, October 2001, pp. 297–300.

[2] A. Rouxel, D. L. Guennec, and O. Macchi, "Unsupervised adaptive separation of impulse signals applied to EEG analysis," in *IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP)*, vol. 1, Istanbul, Turkey, June 2000, pp. 420–423.

[3] K. Abed-Meraim, S. Attallah, T. Lim, and M. Damen, "A blind interference canceller in DS-CDMA," in *IEEE International Symposium on Spread Spectrum Techniques and Applications*, Parsippany, September 2000, pp. 358–362.

[4] I. Durán-Díaz and S. A. Cruces-Alvarez, "A joint optimization criterion for blind DS-CDMA detection," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 79248, pp. 1–11, 2007.

[5] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, "Underdetermined blind audio source separation using modal decomposition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, no. 85438, pp. 1–15, 2007.

[6] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation: Independent Component Analysis and Blind Deconvolution*. Academic Press, February 2010.

[7] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proc. of the IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.

[8] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, February 1997.

[9] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Transactions on Signal Processing*, vol. 46, no. 11, pp. 2888–2897, November 1998.

[10] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source separation using second order cyclostationary statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 694–701, April 2001.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal Of The Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

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

[13] T. Melia and S. Rickard, "Underdetermined blind source separation in echoic environments using DESPRIT," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 86484, pp. 1–19, 2007.

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

[15] A. Aïssa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier, "Underdetermined blind separation of nondisjoint sources in the time-frequency domain," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 897–907, March 2007.

[16] F. Abrard and Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Signal Processing*, vol. 85, no. 7, pp. 1389–1403, July 2005.

[17] S. Arberet, R. Gribonval, and F. Bimbot, "A robust method to count and locate audio sources in a multichannel underdetermined mixture," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 121–133, January 2010.

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

[19] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 87, no. 8, pp. 1833–1847, March 2007.

[20] S. Araki, T. Nakatani, H. Sawada, and S. Makino, "Stereo source separation and source counting with MAP estimation with dirichlet prior considering spatial aliasing problem," in *Independent Component Analysis and Signal Separation (ICA)*, ser. LNCS, vol. 5441. Springer, 2009, pp. 742–750.

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

[22] D. Pastor and F.-X. Socheleau, "Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1545–1555, April 2012.

[23] D. Pastor, "Signal norm testing in additive and independant standard Gaussian noise," available at http://www.telecom-bretagne.eu/publications/publication.php?idpublication=10706, Institut Mines-Télécom; Télécom Bretagne, UEB, Lab-STICC UMR CNRS 3192, Tech. Rep., 2011.

[24] S. M. Aziz-Sbaï, A. Aïssa-El-Bey, and D. Pastor, "Robust underdetermined blind audio source separation of sparse signals in the time-frequency domain," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3716–3719.

[25] D. Pastor, R. Gay, and A. Gronenboom, "A sharp upper bound for the probability of error of likelihood ratio test for detecting signals in white gaussian noise," *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 228–238, January 2002.

[26] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, Istanbul, Turkey, June 2000, pp. 2985–2988.

[27] Y. Deville and M. Puigt, "Temporal and time-frequency correlation-based blind source separation methods. part I: Determined and underdetermined linear instantaneous mixtures," *Signal Processing*, vol. 87, no. 3, pp. 374–407, March 2007.

[28] F. Theis and E. Lang, "Formalization of the two-step approach to overcomplete BSS," in *Signal and Image Processing (SIP)*, Kauai, USA, August 2002, pp. 207–212.

[29] A. M. Atto, D. Pastor, and G. Mercier, "Detection thresholds for non-parametric estimation," *Signal, Image and Video processing*, vol. 2, no. 3, pp. 207–223, February 2008.

[30] S. M. Berman, *Sojourns and extremes of stochastic processes*. Wadsworth, Reading, MA, January 1992.

[31] S. Mallat, *A wavelet tour of signal processing, second edition*. Academic Press, 1999.

[32] R. J. Serfling, *Approximations theorems of mathematical statistics*. Wiley, 1980.

[33] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, August 1994.

[34] F. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, June 1974.

[35] P. Huber and E. Ronchetti, *Robust Statistics, second edition*. John Wiley and Sons, 2009.

[36] R. Leonard, "A database for speaker-independent digit recognition," in *IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP)*, vol. 9, San Diego, California, USA, March 1984, pp. 328–331.

[37] J. R. E. J. Candès and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, August 2006.

[38] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, "Blind separation of underdetermined convolutive mixtures using their time-frequency representation," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 15, no. 5, pp. 1540–1550, July 2007.

[39] C. Févotte, R. Gribonval, and E. Vincent, "A toolbox for performance measurement in (blind) source separation," available at http://bass-db.gforge.inria.fr/bss_eval/.