HAL
archives-ouvertes.fr

Time–frequency ratio-based blind separation methods for attenuated and time-delayed sources

Matthieu Puigt, Yannick Deville

▶ **To cite this version:**

Matthieu Puigt, Yannick Deville. Time–frequency ratio-based blind separation methods for attenuated and time-delayed sources. Mechanical Systems and Signal Processing, Elsevier, 2005, 19 (6), pp.1348-1379. <10.1016/j.ymssp.2005.08.003>. <hal-00270877>

HAL Id: hal-00270877

https://hal.archives-ouvertes.fr/hal-00270877

Submitted on 7 Apr 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Authors' final version of a paper published in "Mechanical Systems and Signal Processing"

Paper reference:
M. Puigt, Y. Deville, "Time-frequency ratio-based blind separation methods for attenuated and time-delayed sources", Mechanical Systems and Signal Processing, vol. 19, pp. 1348-1379, 2005.

Note: a few "errors" in the version published by the editor were corrected in this authors' final version, i.e:

1. After Eq. (22): $k = \sigma(m)$ (instead of $m = \sigma(k)$) and in Appendix B: $k' = \sigma(m')$ (instead of $k' = \sigma^{-1}(m')$).
2. In and after Eq. (38): $q'_{im'}(n_{p'})$ (instead of $q'_{im}(n_{p'})$).
3. (and a few "errors in words").

Time-frequency ratio-based blind separation methods for attenuated and time-delayed sources

Matthieu PUIGT* and Yannick DEVILLE

Laboratoire d'Astrophysique de Toulouse-Tarbes
Observatoire Midi-Pyrénées - Université Paul Sabatier
14 Av. Edouard Belin, 31400 Toulouse, France

mpuigt@ast.obs-mip.fr, ydeville@ast.obs-mip.fr
Tel. : +33 5 61 33 28 24
Fax : +33 5 61 33 28 40

We propose two types of time-frequency (TF) blind source separation (BSS) methods suited to attenuated and delayed (AD) mixtures. These approaches, inspired from a method that we previously developed for linear instantaneous (LI) mixtures, almost only require each source to occur alone in a tiny TF zone, i.e. they set very limited constraints on the source sparsity and overlap, unlike various previously reported TF-BSS methods. Our approaches consist in identifying the columns of the (filtered permuted) mixing matrix in TF zones where these methods detect that a single source occurs, using TIme-Frequency Ratios Of Mixtures (hence their name TIFROM). We thus identify columns of scale coefficients and time shifts. The detection stage for time shifts uses regression lines associated to the above-mentioned TF ratios of mixtures. The detection stage for scale coefficients uses the variance of these TF ratios of mixtures, either in Constant-Time or in Constant-Frequency analysis zones. This yields two alternative BSS methods, which are resp. called AD-TIFROM-CT and AD-TIFROM-CF. These methods are especially suited to non-stationary sources. We derive their performance from many tests performed with AD mixtures of speech signals. This demonstrates that they yield major SNR improvements, i.e. about 45 dB with optimum parameters for time shifts ranging from 0 to 20 samples and above 18 dB for 200-sample time shifts.

Keywords:
attenuated and delayed mixtures,
blind source separation,
non-stationary sources,
sparsity,
time-frequency analysis,
variance analysis.

*Corresponding author.

1 Introduction

Blind source separation (BSS) consists in estimating a set of N unknown sources from a set of P observations resulting from mixtures of these sources through unknown propagation channels.
Most of the approaches that have been developed to this end are based on Independent Component Analysis (ICA) [1]. They assume that the sources are random stationary statistically independent signals, and that at most one source is Gaussian. They recombine the available observed signals so as to obtain statistically independent output signals. These output signals are then equal to the sources, up to some indeterminacies.
More recently, a few other concepts for BSS have also been considered. Especially, several methods based on a time-frequency (TF) analysis of the signals have been reported. Three main classes emerge from these TF-BSS methods. The first one directly results from classical BSS approaches, as it consists of TF adaptations of previously developed joint-diagonalization methods, with subsequent modifications [2]-[4]. The second class includes several methods based on ratios of TF transforms of the observed signals. Some of these methods, i.e. DUET and its modified versions, require the sources to have no overlap in the TF domain [5]-[9], which is quite restrictive. On the contrary, only slight differences in the TF representations of the sources are requested by the TIFROM method that we proposed in [10]-[14]. The third class is based on TF correlation [15]-[17] or TF coherence [14],[18] parameters. All these methods are restricted to linear instantaneous mixtures, except DUET, which applies to time-delayed mixtures but sets other restrictive conditions as stated above.
In this paper, we propose a novel TF-BSS method which is inspired by our linear instantaneous (LI) TIFROM method, but suited to more general mixtures, involving time shifts. We thus avoid the restriction of the DUET method concerning the sparsity of the sources in the TF domain.
The remainder of this paper is therefore organized as follows. In Section 2, we define the attenuated and delayed (AD) mixture configuration considered in this paper and the resulting goal of our investigations. We then present in Section 3 the LI method that we used as the starting point of our extension introduced in this paper. In Section 4, we describe a first AD extension of the LI-TIFROM method. We then propose an alternative version in Section 5. Section 6 reports on a detailed analysis of the experimental performance of both proposed approaches, applied to AD artificial mixtures of real speech sources. Section 7 presents the conclusions drawn from this paper and outlines extensions of the proposed methods. Specific topics are detailed in the appendices.

2 Problem statement

In this paper, we consider the configuration involving N sources and N observations. In the simplest class of mixtures which has been considered in the literature, the observations $x_i(n)$ are linear instantaneous (LI) mixtures of the sources $s_j(n)$, i.e. they read

$$x_i(n) = \sum_{j=1}^{N} a_{ij}\, s_j(n) \qquad i = 1 \ldots N \tag{1}$$

where a_{ij} are constant scale coefficients, which may represent attenuation during propagation from source j to observation i.
In this paper, we consider more general mixtures, which also take into account delays

during propagation, by means of integer-valued time shifts n_{ij}. The resulting extended model for attenuated and delayed (AD) mixtures then reads

$$x_i(n) = \sum_{j=1}^{N} a_{ij}\, s_j(n - n_{ij}) \qquad i = 1 \ldots N. \tag{2}$$

For the sake of simplicity, we assume that the coefficients a_{ij} are real-valued and strictly positive (this corresponds to direct propagation, i.e. without reflection) and that the sources are real-valued. The Fourier transform of (2) reads

$$X_i(\omega) = \sum_{j=1}^{N} a_{ij}\, e^{-j\omega n_{ij}}\, S_j(\omega) \qquad i = 1 \ldots N. \tag{3}$$

This yields in matrix form

$$\underline{X}(\omega) = A(\omega)\, \underline{S}(\omega) \tag{4}$$

with $\underline{S}(\omega) = [S_1(\omega) \cdots S_N(\omega)]^T$, $\underline{X}(\omega) = [X_1(\omega) \cdots X_N(\omega)]^T$ and

$$A(\omega) = \begin{bmatrix} a_{11}e^{-j\omega n_{11}} & \cdots & a_{1N}e^{-j\omega n_{1N}} \\ a_{21}e^{-j\omega n_{21}} & \cdots & a_{2N}e^{-j\omega n_{2N}} \\ \vdots & & \vdots \\ a_{N1}e^{-j\omega n_{N1}} & \cdots & a_{NN}e^{-j\omega n_{NN}} \end{bmatrix}. \tag{5}$$

BSS would ideally aim at estimating the above mixing matrix $A(\omega)$, which is assumed to be invertible at all frequencies. However, this can only be performed up to the well-known permutation and scale/filter indeterminacies inherent in the BSS problem. We here handle them by extending to AD mixtures an approach that we introduced in another type of LI BSS method, i.e. LI-TIFCORR [15],[16]. This approach may be defined as follows. We consider an arbitrary permutation function $\sigma(.)$, applied to the indices j of the source signals, which yields the permuted source signals $s_{\sigma(j)}(n)$. We then introduce scaled and time-shifted versions of the latter signals, equal to their contributions in the first mixed signal, i.e.

$$s'_j(n) = a_{1,\sigma(j)}\, s_{\sigma(j)}\left(n - n_{1,\sigma(j)}\right). \tag{6}$$

The mixing equation (2) may then be rewritten as

$$x_i(n) = \sum_{j=1}^{N} a_{i,\sigma(j)}\, s_{\sigma(j)}\left(n - n_{i,\sigma(j)}\right) \tag{7}$$

$$= \sum_{j=1}^{N} b_{ij}\, s'_j\left(n - \mu_{ij}\right) \tag{8}$$

with

$$\begin{cases} b_{ij} & = \frac{a_{i,\sigma(j)}}{a_{1,\sigma(j)}} \\ \mu_{ij} & = n_{i,\sigma(j)} - n_{1,\sigma(j)} \end{cases} \tag{9}$$

This mixing equation (8) has the same form as (2) except that the sources and parameters $s_j(n)$, a_{ij} and n_{ij} are here respectively replaced by $s'_j(n)$, b_{ij} and μ_{ij}. Using the same

approach as above, this yields the Fourier domain matrix expression

$$\underline{X}(\omega) = B(\omega)\,\underline{S'}(\omega) \tag{10}$$

where $\underline{S'}(\omega) = [S'_1(\omega) \cdots S'_N(\omega)]^T$ and

$$B(\omega) = \begin{bmatrix} b_{11}e^{-j\omega\mu_{11}} & \cdots & b_{1N}e^{-j\omega\mu_{1N}} \\ b_{21}e^{-j\omega\mu_{21}} & \cdots & b_{2N}e^{-j\omega\mu_{2N}} \\ \vdots & & \vdots \\ b_{N1}e^{-j\omega\mu_{N1}} & \cdots & b_{NN}e^{-j\omega\mu_{NN}} \end{bmatrix}. \tag{11}$$

We therefore aim at introducing a method for estimating $B(\omega)$. The output signals may then be obtained in the frequency domain by computing

$$\underline{Y}(\omega) = B^{-1}(\omega)\,\underline{X}(\omega) \tag{12}$$

where $\underline{Y}(\omega) = [Y_1(\omega) \cdots Y_N(\omega)]^T$ is the vector of Fourier transforms of the output signals. The time-domain versions of these signals are then obtained by applying an inverse Fourier transform to $\underline{Y}(\omega)$.

3 Summary of the TIFROM method for linear instantaneous mixtures

3.1 Time-frequency tool and assumptions

We recently proposed [10]-[14] a BSS method based on TIme-Frequency Ratios Of Mixtures, that we therefore called "TIFROM". More precisely, we call it LI-TIFROM hereafter, since it is restricted to linear instantaneous mixtures, unlike its extensions that we introduce in this paper. The TF transform of the signals considered in that approach is the Short-Time Fourier Transform [24] (STFT) defined as[1]:

$$U(n,\omega) = \sum_{n'=-\infty}^{+\infty} u(n')h(n'-n)e^{-j\omega n'} \tag{13}$$

where $h(n'-n)$ is a shifted windowing function, centered on time n. $U(n,\omega)$ is the contribution of the signal $u(n)$ in the TF window corresponding to the short time window centered on n and to the angular frequency ω.
We now introduce the only assumptions that we made in the LI-TIFROM approach with respect to the sources and the associated definitions.

Definition 1 *a source is said to "occur alone" in a TF area (which is composed of several adjacent above-defined TF windows) if only this source has a TF transform which is not equal to zero everywhere in this TF area.*

Definition 2 *a source is said to be "visible" in the TF domain if there exist at least one TF area where it occurs alone.*

[1] The LI-TIFROM method was defined in a continuous-time framework in [10]-[14]. We here consider its discrete-time version, which is better suited to its AD extension introduced in this paper.

Assumption 1 *each source is visible in the TF domain.*

Assumption 2 *there exist no TF areas where the TF transforms of all sources are equal to zero everywhere*[2].

Assumption 3 *when several sources occur in a given set of adjacent TF windows, they should vary so that at least one of the ratios* $\frac{X_i(n,\omega)}{X_1(n,\omega)}$ *of STFTs of observations, with* $i = 2 \dots N$*, does not take the same value in all these windows*[3]*. Especially, i) at least one of the sources must take significantly different TF values in these windows so that the variance of the ratio* $\frac{X_i(n,\omega)}{X_1(n,\omega)}$ *is non-negligible and ii) the sources should not vary proportionally.*

Assumption 4 *the mixing matrix* A *composed of the entries* a_{ij} *is such that* $a_{ij} \neq 0$*,* $\forall i, j$*.*

Assumption 4 implies that if a source occurs in one observation for a TF window (n, ω), then it also exists in all the other observations for this TF window.

3.2 Overall structure of LI-TIFROM

The problem statement that we presented in Section 2 for AD mixtures especially applies to LI mixtures, which correspond to the specific case when all time shifts μ_{ij} are zero. The matrix $B(\omega)$ defined in (11) is then restricted to the frequency-independent matrix B defined as

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1N} \\ b_{21} & \cdots & b_{2N} \\ \vdots & & \vdots \\ b_{N1} & \cdots & b_{NN} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ \frac{a_{2,\sigma(1)}}{a_{1,\sigma(1)}} & \cdots & \frac{a_{2,\sigma(N)}}{a_{1,\sigma(N)}} \\ \vdots & & \vdots \\ \frac{a_{N,\sigma(1)}}{a_{1,\sigma(1)}} & \cdots & \frac{a_{N,\sigma(N)}}{a_{1,\sigma(N)}} \end{bmatrix}. \tag{14}$$

The LI-TIFROM method aims at estimating this matrix, i.e. all the ratios[4] $\frac{a_{im}}{a_{1m}}$, with $i = 2 \dots N$ and $m = 1 \dots N$.
As explained in [10]-[14], this LI-TIFROM method is composed of 3 main stages, preceded by a pre-processing stage, i.e.:

1. The pre-processing stage consists in deriving the STFTs $X_i(n, \omega)$ of the mixed signals, according to (13).

2. In the detection stage, we consider constant-frequency TF analysis zones, composed of a few TF windows (n_p, ω_l) corresponding to adjacent n_p and the same ω_l. This detection stage aims at finding TF analysis zones where a source occurs alone. The principle of its basic version was explained in [10]-[14] and may be summarized as follows (this aspect of the method is explained in more detail in the description of

[2] This assumption is only made for the sake of simplicity: it may be removed in practice, thanks to the noise contained by real recordings, as explained in [13].

[3] When we further extend this approach to mixtures which involve time shifts, not only the ratio $\frac{X_i(n,\omega)}{X_1(n,\omega)}$ should thus vary, but also its modulus.

[4] We here consider the final form of LI-TIFROM proposed in [14]. On the contrary, its initial form [10]-[13] uses a matrix different from B and identifies the inverse ratios $\frac{a_{1m}}{a_{im}}$.

the AD extension of this approach provided below in the paper). We consider the ratios of STFTs of mixtures

$$\alpha_i(n,\omega) = \frac{X_i(n,\omega)}{X_1(n,\omega)} = \frac{\sum_{j=1}^{N} a_{ij}\, S_j(n,\omega)}{\sum_{j=1}^{N} a_{1j}\, S_j(j,\omega)} \tag{15}$$

If only source $S_k(n,\omega)$ occurs in the time-adjacent windows (n_p,ω_l) of an analysis zone, then Eq (15) shows that $\alpha_i(n_p,\omega_l)$ is constant (and equal to $\frac{a_{ik}}{a_{1k}}$) over these windows. On the contrary, it takes different values over these windows for at least one index i if several sources are present, due to Assumption 3. The TF analysis zones where the average over i of the variances of the ratios $\alpha_i(n_p,\omega_l)$ takes the lowest values are therefore considered as the "best" single-source zones. We select these zones by ordering TF analysis zones according to increasing values of the above averaged variance.
In [14], we proposed an improved version of this detection stage where we also consider the inverse ratios

$$\beta_i(n,\omega) = \frac{1}{\alpha_i(n,\omega)} = \frac{X_1(n,\omega)}{X_i(n,\omega)} \tag{16}$$

The averaged variance of β_i has lower or higher values than that of α_i, depending on the mixing coefficients. We therefore order the TF zones, according to increasing values of the minimum among the averaged variances of both ratios α_i and β_i.
In both versions of the detection stage, we eventually obtain an ordered list of single-source TF analysis zones which then allows us to estimate the columns of B in the identification stage, as will now be explained.

3. The identification stage identifies the columns of B in the above single-source analysis zones. This basically consists in successively using the first and subsequent single-source zones in the above ordered list for deriving a column of B in each of these zones. Denoting k the index of this column, all mixing coefficient ratios $\frac{a_{im}}{a_{1m}}$ with $i = 2 \cdots N$ and $m = \sigma(k)$ are resp. estimated as the mean values of the ratios $\alpha_i(n,\omega)$ of STFTs of observations over the considered single-source TF zone (the parameters $\beta_i(n,\omega)$ are also used in the improved version of this stage, as detailed below for the extension of this approach to AD mixtures)[5].

4. In the combination stage, we eventually left-multiply the vector of mixed signals by the inverse of the estimated mixing matrix B, in order to obtain the extracted source signals. This correspond to the time-domain version of (12) restricted to LI mixtures.

4 First extension of LI-TIFROM to attenuated and delayed (AD) mixtures

We now introduce a first approach for extending the above LI-TIFROM method so as to handle the attenuated and delayed (AD) mixtures that we defined in Section 2. The TF-BSS method thus introduced in this section uses constant-frequency (CF) TF zones

[5] The method then used for adequately gathering all these columns to form the overall matrix B may be found in [10]-[14] for LI-TIFROM and is detailed below in this paper for the extension of this approach to AD mixtures

for estimating some of the mixing parameters. It is therefore called AD-TIFROM-CF. Its core stages again consist in detecting single-source TF zones and then identifying inside these zones the parameters of the matrix $B(\omega)$ that we introduced in (11), i.e. the parameters b_{im} and μ_{im} which define the m-th column of $B(\omega)$. In this section, we successively describe:

- the methods that we propose for detecting (constant-frequency) single-source zones and identifying the parameters b_{im} inside them (see Subsection 4.1 below),
- the methods that we propose for detecting (constant-time) single-source zones and identifying the parameters $\mu_{im'}$ inside them (see Subsection 4.2 below),
- the procedure for coupling the above identified parameters b_{im} and $\mu_{im'}$ (see Subsection 4.3 below),
- and the overall BSS method which results from all these principles (see Subsection 4.4 below).

4.1 Basic detection and identification stages for the parameters b_{im}

As in the case of LI mixtures, the BSS method that we introduce here first includes a detection stage for finding (constant-frequency) single-source TF zones. Starting from the time-domain AD mixture equations (2) considered in this paper, we derived in Section 2 the corresponding frequency-domain mixture equations (3). The latter relationship between the observations and sources remains almost exact when expressed in the TF domain if the time shifts n_{ij} are small enough as compared to the temporal width of the windowing function $h(.)$ used in the STFT transform.
In this paper, we assume that this condition is met and therefore that the STFTs of the observations can be expressed with respect to the STFTs of the sources as

$$X_i(n,\omega) = \sum_{j=1}^{N} a_{ij}\, e^{-j\,\omega n_{ij}}\, S_j(n,\omega) \qquad i = 1 \dots N. \tag{17}$$

This should be compared to the case of LI mixtures, where the time-domain mixture equations (1) yield (without any approximation) the following TF-domain relationship:

$$X_i(n,\omega) = \sum_{j=1}^{N} a_{ij}\, S_j(n,\omega) \qquad i = 1 \dots N. \tag{18}$$

The relationship (17) for AD mixtures is therefore the same as the expression (18) for LI mixtures, except for the overall coefficient applied to each source. The property that we used in Section 3 in the detection stage of the approach for LI mixtures does not depend on the values of these coefficients and therefore still applies here. In others words, we again consider a TF analysis zone which consists of a set of M possibly overlapping TF windows corresponding to the same frequency ω_l and to adjacent time positions n_p. This set of M time points n_p, with $p = 1 \dots M$, is denoted as T hereafter and the corresponding TF zone is therefore denoted (T, ω_l).
Here again, we study the ratios $\alpha_i(n,\omega)$ and $\beta_i(n,\omega)$ defined by (15) and (16), which here

read

$$\alpha_i(n,\omega) = \frac{\sum_{j=1}^{N} a_{ij}\, e^{-j\omega n_{ij}}\, S_j(n,\omega)}{\sum_{j=1}^{N} a_{1j}\, e^{-j\omega n_{1j}}\, S_j(n,\omega)} \tag{19}$$

$$\beta_i(n,\omega) = \frac{\sum_{j=1}^{N} a_{1j}\, e^{-j\omega n_{1j}}\, S_j(n,\omega)}{\sum_{j=1}^{N} a_{ij}\, e^{-j\omega n_{ij}}\, S_j(n,\omega)} \tag{20}$$

If a source $S_k(n,\omega)$ occurs alone in the considered TF window (n_p,ω_l) then

$$\alpha_i(n_p,\omega_l) = \frac{a_{ik}}{a_{1k}} e^{-j\omega(n_{ik}-n_{1k})} \tag{21}$$

$$= b_{im} e^{-j\omega\mu_{im}} \tag{22}$$

with b_{im} and μ_{im} defined by (9) and $k=\sigma(m)$. Since we assumed all mixing coefficients a_{ik} to be real and positive, all resulting scale coefficients b_{im} are also real and positive. The modulus of the parameter value provided in (22) is therefore equal to

$$|\alpha_i(n_p,\omega_l)| = b_{im} \tag{23}$$

Therefore, if source $S_k(n,\omega)$ occurs alone over the TF analysis zone (T,ω_l), then $|\alpha_i(n_p,\omega_l)|$ is constant over this zone and its mean, defined by

$$\overline{|\alpha_i|}(T,\omega_l) = \frac{1}{M}\sum_{p=1}^{M} |\alpha_i(n_p,\omega_l)|\,, \tag{24}$$

then reads

$$\overline{|\alpha_i|}(T,\omega_l) = \frac{a_{ik}}{a_{1k}} = b_{im} \tag{25}$$

In the same way, $|\beta_i(n_p,\omega_l)|$ is constant over the analysis zone (T,ω_l) and $\overline{|\beta_i|}(T,\omega_l)$, defined by

$$\overline{|\beta_i|}(T,\omega_l) = \frac{1}{M}\sum_{p=1}^{M} |\beta_i(n_p,\omega_l)|\,, \tag{26}$$

then reads

$$\overline{|\beta_i|}(T,\omega_l) = \frac{a_{1k}}{a_{ik}} = \frac{1}{b_{im}}. \tag{27}$$

A natural approach may therefore be proposed for detecting the single-source zones and identifying the parameters b_{im}. It consists in using the same type of method as in the LI-TIFROM approach that we briefly described in the previous section, except that the parameters b_{im} are here defined by $|\alpha_i(n,\omega)|$ and $|\beta_i(n,\omega)|$. Thus, we compute the variances of $|\alpha_i(n,\omega)|$ and $|\beta_i(n,\omega)|$

$$var\left[|\alpha_i|\right](T,\omega_l) = \frac{1}{M}\sum_{p=1}^{M} \left||\alpha_i(n_p,\omega_l)| - \overline{|\alpha_i|}(T,\omega_l)\right|^2, \tag{28}$$

$$var\left[|\beta_i|\right](T,\omega_l) = \frac{1}{M}\sum_{p=1}^{M} \left||\beta_i(n_p,\omega_l)| - \overline{|\beta_i|}(T,\omega_l)\right|^2 \tag{29}$$

and the means over i of these variances

$$MVAR\left[|\alpha|\right](T,\omega_l) = \frac{1}{N-1}\sum_{i=2}^{N} var\left[|\alpha_i|\right](T,\omega_l), \tag{30}$$

$$MVAR\left[|\beta|\right](T,\omega_l) = \frac{1}{N-1}\sum_{i=2}^{N} var\left[|\beta_i|\right](T,\omega_l). \tag{31}$$

We then order the TF analysis zones according to increasing values of the parameter $min\left\{MVAR\left[|\alpha|\right](T,\omega_l), MVAR\left[|\beta|\right](T,\omega_l)\right\}$. The lowest values of this parameter are again obtained in the "best" single-source zones.
The parameters b_{im} are then identified by successively using as follows each of the first and subsequent TF analysis zones in the above ordered list. If the considered single-source analysis zone (T,ω_l) is such that

$$min\left\{MVAR\left[|\alpha|\right](T,\omega_l), MVAR\left[|\beta|\right](T,\omega_l)\right\} = MVAR\left[|\alpha|\right](T,\omega_l), \tag{32}$$

then, due to Eq (25), the identification parameters $\overline{|\alpha_i|}(T,\omega_l)$ yield an estimated column of the parameters b_{im} of $B(\omega)$. This column is kept only if its distance with respect to all previously identified columns is above a positive user-defined threshold ϵ_1, showing that this TF analysis zone does not contain the same source as previous zones in the ordered list. In the symmetrical case when

$$min\left\{MVAR\left[|\alpha|\right](T,\omega_l), MVAR\left[|\beta|\right](T,\omega_l)\right\} = MVAR\left[|\beta|\right](T,\omega_l), \tag{33}$$

in the considered analysis zone, due to Eq (27), the identification parameters $\overline{|\beta_i|}(T,\omega_l)$ yield (the inverses of) the parameters b_{im}, corresponding to a column of $B(\omega)$, which is kept according to the same criterion as above.
The identification procedure ends when the number of columns of $B(\omega)$ thus kept becomes equal to the specified number N of sources to be separated (this is theoretically guaranteed to occur because all sources are assumed to be visible in the considered data).
We present in Appendix A a modified version of the identification stage for the parameters b_{im}.

4.2 Detection and identification stages for the parameters μ_{im}

Thanks to expression (22) of the identification parameters $\alpha_i(n,\omega)$ in single-source analysis zones, we used above the moduli of these parameters to identify the scale factors b_{im}. Still considering this expression (22), a natural idea for estimating the time shifts μ_{im} then consists in trying to take advantage of the phase of $\alpha_i(n,\omega)$. What may be derived in practice is its "plain phase angle", i.e. the associated angle situated in the range $[-\pi,\pi]$, that we denote $\psi_i(n,\omega)$. Due to (22), in an analysis zone (T,ω_l) where only $S_k(n,\omega)$ occurs, we have (denoting $m=\sigma(k)$):

$$-\omega_l\mu_{im} = \psi_i(n_p,\omega_l) + 2q_{im}(n_p,\omega_l)\pi \tag{34}$$

where $q_{im}(n_p,\omega)$ is an unknown integer which yields an indeterminacy in the value of μ_{im}.
It should be remembered that the plain phase angles of FFT or STFT values suffer from the above type of indeterminacy when considered individually, but well-known techniques exist for introducing "coherence" between the phases of frequency-adjacent FFT or STFT

points, by unwraping these phases with respect to frequency. Combining this phase unwraping with the phase-based identification principle that we introduced above suggests a modified approach for identifying the parameters μ_{im}. This new approach is also based on the phase of $\alpha_i(n,\omega)$ considered in TF windows (n,ω). We consider independently each time position n associated to these TF windows and for each such position, we unwrap the phase of $\alpha_i(n,\omega)$ over all associated frequency-adjacent TF points. We will now show that this unwraped phase makes it possible to identify the parameters μ_{im}, provided it is considered over a connected subset of the above TF points where only one source occurs. This means that we should first detect single-source TF analysis zones, but it should be clear that we here have to consider another type of analysis zones as compared to those that we previously used for identifying the parameters b_{im}: due to the phase unwraping principle required here, we consider analysis zones which consist of TF windows corresponding to the same time position and to adjacent frequency positions. We therefore have to develop not only an identification method for the parameters μ_{im}, but also a method for detecting constant-time single-source analysis zones. Both methods may actually be derived from the above principles, as will now be shown.
Let us consider a constant-time analysis zone, i.e. a set of M' TF windows, corresponding to the same time position $n_{p'}$ and to adjacent angular frequencies $\omega_{l'}$, with $l' = 1 \dots M'$. This set of frequency points is denoted as Ω hereafter and the corresponding TF zone is therefore denoted $(n_{p'},\Omega)$. As stated above, it is to be contrasted with the analysis zones (T,ω_l) that we used above for the identification of the parameters b_{im} and which consist of time-adjacent windows. Let us focus on an analysis zone $(n_{p'},\Omega)$ where only source $S_{k'}(n,\omega)$ occurs. The expression (17) of the observed signals at any point of this analysis zone then reduces to:

$$X_i(n_{p'},\omega_{l'}) = a_{ik'}\, e^{-j\omega_{l'} n_{ik'}}\, S_{k'}(n_{p'},\omega_{l'}) \qquad i = 1 \dots N \tag{35}$$

The identification parameter (15) then reads:

$$\begin{aligned} \alpha_i(n_{p'},\omega_{l'}) &= \frac{a_{ik'}}{a_{1k'}} e^{-j\omega_{l'}(n_{ik'}-n_{1k'})} && (36)\\ &= b_{im'} e^{-j\omega_{l'}\mu_{im'}} \text{ with } m' = \sigma(k') && (37) \end{aligned}$$

The phase of $\alpha_i(n_{p'},\omega_{l'})$ in constant-time single-source analysis zones is therefore related to the parameters $\mu_{im'}$. More precisely, still assuming that $S_{k'}(n,\omega)$ occurs alone in an analysis zone $(n_{p'},\Omega)$, if we consider the unwraped phase $\phi'_i(n_{p'},\omega_{l'})$ of $\alpha_i(n_{p'},\omega_{l'})$ in this zone, we have:

$$-\omega_{l'}\mu_{im'} = \phi'_i(n_{p'},\omega_{l'}) + 2q'_{im'}(n_{p'})\pi, \tag{38}$$

where $q'_{im'}(n_{p'})$ is an unknown integer. The associated multiple of 2π in (38) again yields an indeterminacy. However, this indeterminacy is different from the one which occured in (34): thanks to the phase unwraping procedure that we applied, the integer $q'_{im'}(n_{p'})$ does not depend on frequency. Eq (38) shows that the curve associated to the variations of the phase $\phi'_i(n_{p'},\omega_{l'})$ with respect to $\omega_{l'}$ in a single-source analysis zone $(n_{p'},\Omega)$ is a line and that its slope does not depend on the value of $q'_{im'}(n_{p'})$ and is equal to $-\mu_{im'}$. This slope therefore provides a means for identifying $\mu_{im'}$, with no phase indeterminacy. As a result of the above analysis, our method for identifying the set of parameters $\mu_{im'}$ associated to a column of $B(\omega)$ operates as follows. In the selected constant-time single-source analysis zone, for each observed signal with index i, we consider the M' points which have two coordinates, resp. defined as the frequencies $\omega_{l'}$ and the corresponding values $\phi'_i(n_{p'},\omega_{l'})$ of the unwraped phase of the identification parameter. We determine

the least-mean square regression line associated to these points. The estimate of the parameter $\mu_{im'}$ is then set to the opposite of the slope of this regression line. More precisely, $\mu_{im'}$ is set to the integer which is the closest to the opposite of this slope, since $\mu_{im'}$ is defined as a time shift in a discrete-time representation and is therefore an integer. This approach may then be straightforwardly extended so as to also perform the detection of the constant-time single-source analysis zones which are required in this identification procedure. This extended version is detailed in Appendix B.
A final stage should now be added to our approach, in order to couple each column of parameters $\mu_{im'}$ to a column of parameters b_{im}.

4.3 Coupling the parameters b_{im} and $\mu_{im'}$

4.3.1 Alternative identification method for the parameters b_{im}

In Subsection 4.2, we introduced a method for detecting constant-time single-source analysis zones and we showed how the phase of the parameters $\alpha_i(n,\omega)$ in such zones may be used to identify the parameters $\mu_{im'}$. We here note that the moduli of these parameters in these zones also make it possible to identify the parameters $b_{im'}$: Eq (36) shows that, at any frequency $\omega_{l'}$ of such a zone, the modulus of $\alpha_i(n_{p'},\omega_{l'})$ is equal to $b_{im'}$. The latter parameter may therefore be identified as the mean value of the modulus of $\alpha_i(n,\omega)$ over a constant-time single-source analysis zone.
The value thus obtained is denoted $b'_{im'}$ below, in order to distinguish it from the value b_{im} provided by the methods that we introduced in Subsection 4.1. The alternative approach that we propose in this subsection is attractive because each considered analysis zone yields the parameters $b'_{im'}$ and $\mu_{im'}$ corresponding to the *same* source. It therefore inherently provides a solution to the coupling of these types of parameters. However, our experimental tests showed that the parameter value $b'_{im'}$ thus obtained estimate less accurately the actual mixture parameters than the values b_{im} that we described in Subsection 4.1. We therefore introduce a modified approach which takes advantage of both types of parameters hereafter.

4.3.2 Coupling the parameters b_{im} and $(b'_{im'},\mu_{im'})$

Taking advantage of all above-defined principles, we now introduce a method for eventually coupling the parameters b_{im} and $\mu_{im'}$. This method consists in:

1. determining the parameters b_{im} as explained in Subsection 4.1,
2. independently determining the couples $(b'_{im'},\mu_{im'})$ as explained in Subsections 4.2 and 4.3.1,
3. and then mapping the parameters $\mu_{im'}$ towards the parameters b_{im} thanks to the parameters $b'_{im'}$. This is achieved as follows. The above identification of the parameters b_{im} yields N columns of such parameters, each associated with a different source. In the detection of constant-time single-source analysis zones, we keep a number of zones significantly larger than N, by selecting all the zones where the mean-square error with respect to the associated regression line is below a user-defined threshold ϵ_3. For each such zone, we identify the two columns that resp. contain the parameters $b'_{im'}$ and $\mu_{im'}$ corresponding to that zone. We then consider the parameters $b'_{im'}$ and b_{im} resp. as coarse and accurate estimates of the scale

parameters associated to the mixing matrix and we map each column of $b'_{im'}$ towards the closest column[6] of b_{im}. Since the parameters $b'_{im'}$ were already coupled with the parameters $\mu_{im'}$, the latter parameters are thus mapped towards the N columns of parameters b_{im}. For each element (associated to the observation index i) in each such column of b_{im}, we should eventually keep only one parameter value $\mu_{im'}$. This is achieved as follows for each such element: among all the values $\mu_{im'}$ which were mapped above towards this element, we keep the value which has the highest number of occurences.

4.4 Overall AD-TIFROM-CF method

As a result of the above description, the overall AD-TIFROM-CF method that we first propose for AD mixtures contains the following stages:

1. The pre-processing stage consists in deriving the STFTs $X_i(n,\omega)$ of the mixed signals, according to (13).

2. We then only consider the parameters b_{im}. We first detect constant-frequency single-source analysis zones (T,ω_l) and we then identify the parameters b_{im} in these zones, using either the basic procedure described in Subsection 4.1 or its improved version introduced in Appendix A.

3. We then consider the parameters $b'_{im'}$ and $\mu_{im'}$. We first compute the regression lines and their associated mean-square errors over the constant-time TF analysis zones $(n_{p'},\Omega)$ defined in Subsection 4.2. We then detect single-sources zones and identify the parameters $b'_{im'}$ and $\mu_{im'}$ in these zones, using the procedures described in Subsections 4.2 and 4.3.1 and in Appendix B .

4. We then couple the parameters b_{im} and $\mu_{im'}$, as explained in Subsection 4.3.2. All the required parameters associated to the estimated matrix $B(\omega)$ are available at this stage.

5. Using the above parameters, we eventually derive the extracted sources from the observations, according to (12).

5 Second BSS approach for AD mixtures

5.1 Alternative detection and identification stages

In Subsection 4.1, we introduced a method which exploits constant-frequency TF zones (T,ω_l), based on the variances of the moduli of the ratios $\alpha_i(n,\omega)$ and $\beta_i(n,\omega)$, in order to detect single-source zones. We were thus able to compute very good estimates of b_{im} but we could not estimate corresponding time shifts μ_{im} in the same zones. In Subsection 4.2, we presented a method, using constant-time TF zones $(n_{p'},\Omega)$, and we introduced a new approach for finding single-source TF zones, based on the regression lines of the unwraped phases of $\alpha_i(n,\omega)$ and the associated regression errors. Constant-time zones are attractive because they inherently couple the two types of parameters identified inside them, i.e. $b'_{im'}$ and $\mu'_{im'}$. But we stated in Subsection 4.3.1 that the parameters $b'_{im'}$ obtained in zones selected according to regression errors estimate less accurately the mixtures than the

[6]Here again, a user-defined threshold ϵ_4 is used to ignore any column of parameters $b'_{im'}$ such that all its distances with respect to the N columns of parameters b_{im} are above that threshold.

values b_{im} above. We here aim at combining the advantages of the above two types of approaches, by considering constant-time zones in order to obtain coupled parameters $b'_{im'}$ and $\mu_{im'}$, and by detecting which of these zones contain a single source by means of the variances of the moduli of the ratios $\alpha_i(n,\omega)$ and $\beta_i(n,\omega)$. We therefore call this approach, which *only* uses constant-time (CT) analysis zones, AD-TIFROM-CT, as opposed to the AD-TIFROM-CF approach described in the previous section, which also uses constant-frequency analysis zones.
We now present in more detail the principles of this AD-TIFROM-CT method. If $S_{k'}(n,\omega)$ occurs alone in an analysis zone $(n_{p'},\Omega)$, then the moduli of $\alpha_i(n_{p'},\omega_{l'})$ and $\beta_i(n_{p'},\omega_{l'})$ resp. expressed in Eq. (19) and (20) are constant. Their means, resp. defined by

$$\overline{|\alpha_i|}(n_{p'},\Omega) = \frac{1}{M'}\sum_{l'=1}^{M'}|\alpha_i(n_{p'},\omega_{l'})| \tag{39}$$

and

$$\overline{|\beta_i|}(n_{p'},\Omega) = \frac{1}{M'}\sum_{l'=1}^{M'}|\beta_i(n_{p'},\omega_{l'})|\,, \tag{40}$$

then read

$$\overline{|\alpha_i|}(n_{p'},\Omega) = \frac{a_{ik'}}{a_{1k'}} = b_{im'} \tag{41}$$

$$\overline{|\beta_i|}(n_{p'},\Omega) = \frac{a_{1k'}}{a_{ik'}} = \frac{1}{b_{im'}} \tag{42}$$

with $m' = \sigma(k')$. We compute the variances of $|\alpha_i(n_{p'},\omega_{l'})|$ and $|\beta_i(n_{p'},\omega_{l'})|$:

$$var\left[|\alpha_i|\right](n_{p'},\Omega) = \frac{1}{M'}\sum_{l'=1}^{M'}\left||\alpha_i(n_{p'},\omega_{l'})| - \overline{|\alpha_i|}(n_{p'},\Omega)\right|, \tag{43}$$

$$var\left[|\beta_i|\right](n_{p'},\Omega) = \frac{1}{M}\sum_{l'=1}^{M'}\left||\beta_i(n_{p'},\omega_{l'})| - \overline{|\beta_i|}(n_{p'},\Omega)\right|^2 \tag{44}$$

and the mean over i of these variances

$$MVAR\left[|\alpha|\right](n_{p'},\Omega) = \frac{1}{N-1}\sum_{i=2}^{N} var\left[|\alpha_i|\right](n_{p'},\Omega), \tag{45}$$

$$MVAR\left[|\beta|\right](n_{p'},\Omega) = \frac{1}{N-1}\sum_{i=2}^{N} var\left[|\beta_i|\right](n_{p'},\Omega). \tag{46}$$

The mixing parameters $b'_{im'}$ are then identified by using the basic method defined in Subsection 4.1 or preferably its improved version depicted in Appendix A, except that the analysis zone (T,ω_l) is here replaced by the zone $(n_{p'},\Omega)$. Therefore:

1. The parameter used here for selecting single-source zones is

$$min\left\{MVAR\left[|\alpha|\right](n_{p'},\Omega), MVAR\left[|\beta|\right](n_{p'},\Omega)\right\} \tag{47}$$

instead of

$$min\left\{MVAR\left[|\alpha|\right](T,\omega_l), MVAR\left[|\beta|\right](T,\omega_l)\right\} \tag{48}$$

which previously appeared in (32), (33), (56). The condition (56) is thus replaced by:

$$min\left\{MVAR\left[|\alpha|\right](n_{p'},\Omega), MVAR\left[|\beta|\right](n_{p'},\Omega)\right\} \leq \epsilon_2. \quad (49)$$

2. The parameters used here for identifying the parameters $b'_{im'}$ are $\overline{|\alpha_i|}(n_{p'},\Omega)$ and $\overline{|\beta_i|}(n_{p'},\Omega)$ instead of $\overline{|\alpha_i|}(T,\omega_l)$ and $\overline{|\beta_i|}(T,\omega_l)$.

The identification of the parameters $\mu_{im'}$ is performed as follows. First consider the case when the basic method (adapted from Subsection 4.1) is used for identifying the parameters $b'_{im'}$. Then, in each selected single-source zone where a column of parameter $b'_{im'}$ is identified (and kept), the corresponding column of parameters $\mu_{im'}$ is simultaneously identified, by using again the method based on regression lines described in Subsection 4.2. Now, in the improved method, we first form N clusters by only considering the parameters $b'_{im'}$, using the approach adapted from Appendix A. For each point of these clusters, we also identify the associated parameters $\mu_{im'}$, again by means of regression lines. We then derive a single column of $\mu_{im'}$ for each cluster by using the same type of approaches as in Subsection 4.3.2, i.e.: for each index i independently among all values $\mu_{im'}$ corresponding to the considered cluster, we keep the value which has the highest number of occurrences.

5.2 Overall AD-TIFROM-CT method

As a result of the above description, the overall AD-TIFROM-CT method that we also propose for AD mixtures contains the following stages:

1. The pre-processing stage consists in deriving the STFTs $X_i(n,\omega)$ of the mixed signals, according to (13).

2. We then detect constant-time single-source analysis zones $(n_{p'},\Omega)$ and we identify the parameters $b'_{im'}$ and $\mu_{im'}$ in these zones, using the procedure introduced in Subsection 5.1. All the required parameters associated to the estimated matrix $B(\omega)$ are available at this stage.

3. Using the above parameters, we eventually derive the extracted sources from the observations, according to (12).

6 Experimental results

6.1 Test conditions and performance criteria

In this section, we present a large number of tests performed with the two methods proposed in this paper, i.e. TIFROM-CF and TIFROM-CT (in this section, we omit "AD-" in the names of these methods in order to improve readability and because this section only concerns AD mixtures of signals and AD BSS methods). These tests were carried out in the following conditions.

- We use English speech signals sampled at 20 kHz. For the sake of simplicity, we only performed these tests for two sets of sources. Each set consists of speech from different male speakers. The sources in Set 1 consist of 2.5 s of continuous speech, while those in Set 2 last 5 s and contain silence. The signals in Set 1 consist of a part of those in Set 2. All these sources were first centered and scaled so that their highest absolute value is equal to 1.

- We derive various artificial AD mixtures of these sources.
- We process these mixed signals with the TIFROM-CF and TIFROM-CT methods resp. defined in Subsections 4.4 and 5.2. For both methods, we here use the improved identification method described in Appendix A for the parameters b_{im} (or its adapted version).

The performance achieved in each test is first measured by the overall Signal to Noise Ratio (SNR) associated to the outputs of the considered BSS system (denoted SNR^{out} hereafter) and/or by the SNR Improvement achieved by this system (denoted $SNRI$ below). These parameters are defined in Appendix C, together with the input SNR associated to the processed mixed signals (denoted SNR^{in} hereafter). Moreover, Appendix C shows that the two performance parameters, i.e. SNR^{out} and $SNRI$, are linked by the following relationship:

$$(SNRI)_{dB} = (SNR^{out})_{dB} - (SNR^{in})_{dB}. \tag{50}$$

These performance parameters are measured over all tests when the considered methods succeed in identifying N columns of $B(\omega)$, where N is the number of considered sources. As shown below, the few cases when they fail to do so correspond to situations when quite large STFT windows are used, so that some sources are no more visible in the corresponding TF representations.
It should also be remembered that the actual and estimated $\mu_{im'}$ are integers. This makes it possible to check whether all estimates of parameters $\mu_{im'}$ are exactly equal to the actual ones. The percentage of cases when this condition is met, among all tests (i.e. including the tests when some columns of $B(\omega)$ were not identified), is used as an additional performance criterion and called the "success rate" hereafter. It should be clear however that even when the considered BSS method does not "succeed" in identifying all $\mu_{im'}$ *exactly*, it may still be able to identify all columns of $B(\omega)$ and yield acceptable performance in terms of $SNRI$, provided the estimates of $\mu_{im'}$ are not too far from their actual values. The above-defined "success rate" is therefore a more pessimistic and partial performance criterion than the overall $SNRI$.
In this section, we only consider the case when $N = 2$, i.e. the configuration involving two mixtures of two sources. All our tests aimed at estimating the influence of the scale coefficients and time shifts on the performance of the proposed methods. We therefore used symmetrical mixing matrices defined as

$$A(\omega) = \begin{bmatrix} 1 & \lambda\, e^{-j\omega\eta} \\ \lambda\, e^{-j\omega\eta} & 1 \end{bmatrix}, \tag{51}$$

where we successively considered different real values for the cross-coupling scale factors λ, in order to vary the mixture ratio and therefore the SNR^{in} associated to the observed signals $x_i(n)$. Similarly, the influence of time shifts was investigated by varying the integer-valued parameter η.
The values that we considered for λ and η are

$$\begin{cases} \lambda = 0.5, 0.9 \\ \eta = 0, 10, 20, 200 \end{cases} \tag{52}$$

The SNR^{in} are equal to 6.0 and 0.9 dB respectively for $\lambda = 0.5$ and 0.9.
As explained in Subsections 4.4 and 5.2, the proposed methods use TF representations of the observed signals $x_i(n)$, obtained by computing their STFTs $X_i(n, \omega)$. More precisely, this type of representation is used twice in the TIFROM-CF method, i.e. first when

considering constant-frequency analysis zones used for estimating the parameters b_{im}, and then when considering constant-time analysis zones used for estimating the parameters $b'_{im'}$ and $\mu_{im'}$. These two types of analysis zones may lead to different optimum values as for the parameters of STFTs and numbers of STFT windows per analysis zone. Therefore, we independently considered two sets of such parameters, resp. associated to the above two types of analysis zones in the TIFROM-CF method, i.e.:

- we here denote d (resp. d') the number of samples of observed signals $x_i(n)$ in each time window of the STFTs used in constant-frequency (resp. constant-time) analysis zones,
- as stated above in Section 4, M (resp. M') is the number of adjacent windows in constant-frequency (resp. constant-time) analysis zones,
- we here denote ρ (resp. ρ') the temporal overlap between the time windows in the STFTs used in constant-frequency (resp. constant-time) analysis zones.

On the contrary, the TIFROM-CT method only uses one type of TF representation and therefore a single set of parameters, i.e. d', M' and ρ'.

6.2 Tests with moderate time shifts

6.2.1 Additional test conditions

As stated in Subsection 4.1, when developing both AD-TIFROM methods, we assumed the time shifts to be small as compared to the size of the windowing function $h(.)$. Therefore, in the tests reported below, we can only state beforehand that performance should be good when $\eta \ll min\{d, d'\}$ for TIFROM-CF and $\eta \ll d'$ for TIFROM-CT. As a first step, we therefore only consider situations such that this condition is roughly met in this subsection, i.e. so that we typically have $\frac{\eta}{min\{d,d'\}} \leq \frac{1}{10}$. To this end, we here only consider the cases $\eta = 0, 10$ or 20, while the parameters of the proposed BSS methods are selected as follows.
Both types of STFTs use a Hanning windowing function $h(.)$. The influence of the parameters of constant-frequency analysis zones was investigated in detail for LI mixtures when testing the performance of the LI-TIFROM method in [14]. For the sake of simplicity, we only consider the optimum values found in [14] in the tests of the TIFROM-CF method reported here, i.e.:

- $d = 256$,
- $M = 10$,
- $\rho = 75\%$.

On the contrary, we analyze in detail the influence of the parameters associated to constant-time analysis zones that we introduce in the current paper for AD mixtures. This is especially motivated by the fact that the parameters of STFT windows influence time shift estimates according to [19]. We varied this second set of parameters as follows:

- The number d' of samples per STFT was geometrically varied from 512 to 16384 samples.

- The number M' of windows per analysis zone was successively set to 4, 8, or 16 when $d' = 512$. This range of values of M' was then increased geometrically with d'. Thus, the widths of the continuous-time frequency bands associated to the frequency domains Ω of the analysis zones $(n_{p'}, \Omega)$ took the same values whatever d'. In these tests, these values were 156.25 Hz, 312.5 Hz and 625 Hz.
- The temporal overlap ρ' was successively set to 50%, 75% and 90%. This parameter was varied independently from d' and M'.

The other parameters of the methods were constant in these tests, i.e.:

- The distance between two identified columns b_{im} of the matrix $B(\omega)$ was measured by the highest absolute difference between elements of these vectors which have the same index. The distance threshold ϵ_1, defined in Appendix A, was set to 0.15 (considering the mixing coefficient values).
- The single-source analysis zones (T, ω_l) (resp. $(n_{p'}, \Omega)$) were detected as the zones where the means of the variances are below a threshold ϵ_2, as shown in (56) (resp. in (49)). This threshold was set to 1.5e-2.
- The threshold ϵ_3 which defines which single-source analysis zones $(n_{p'}, \Omega)$ are kept in the TIFROM-CF method, as explained in Subsection 4.3.2, was set to 0.1.
- In the TIFROM-CF method, the distance between two identified columns b_{im} and $b'_{im'}$, which coupled the parameters b_{im} and $\mu_{im'}$ as explained in Subsection 4.3.2, was measured by the highest absolute difference between elements of these vectors which have the same index. The associated distance threshold ϵ_4 was set to 0.05.

In order to compare the performance of our methods, we studied TIFROM-CT under same the conditions.
The set of combinations of parameters associated with the estimation of the time shifts thus allows us to carry out 54 experiments per BSS method, for a fixed couple of signals and a fixed mixing matrix. By considering the above-defined two couples of signals and the 6 mixing matrices corresponding to $\eta = 0, 10$ or 20 in (52), 648 tests were performed for each method.

6.2.2 Performance of TIFROM-CF

6.2.2.1 Estimation of parameters b_{im}

In the TIFROM-CF method, the estimates of b_{im} are independent from the parameters d', ρ' and M' varied in these tests. Table 1 provides the Frobenius norm of the difference between the estimated and actual matrices of parameters b_{im}. This norm is quite low, showing that the proposed method always succeeds in identifying the parameters b_{im} very accurately.
We expected Set 2 of sources to yield better performance than Set 1, because it contains silence phases in addition, thus making the sources more visible. This is confirmed by Table 1. When $\lambda = 0.9$, both sets yield low estimation errors and the difference between them is quite small.

6.2.2.2 Estimation of parameters $\mu_{im'}$ and global performance

The overall performance of the TIFROM-CF method is shown in Table 2. The mean

$SNRI$ over all BSS method parameters ranges from 16.5 to 66.3 dB, depending on the sources and mixing conditions, with a global value over all tests equal to 43.6 dB. This table shows that this mean $SNRI$ significantly decreases when the delays η increase. However, the above-defined "success rate" (with respect to the *exact* estimation of the time shifts $\mu_{im'}$) then remains quite high, with a global value over all tests equal to 85.2%. We expected such good results because all the assumptions made in our method are here met. Note that in each of these tests, TIFROM-CF found all sources to be visible.
The influence of the scale parameter λ on performance should be analyzed with care: when λ is changed, the input signals are no more mixed to the same extent, so that SNR^{in} is modified. Eq (50) shows that the correspondence between $SNRI$ and SNR^{out} is then modified, so that these two performance criteria may not have the same variations with respect to λ. This is reflected in Table 2, where SNR^{out} has a lower sensitivity to λ than $SNRI$. For the sake of brevity, we only consider $SNRI$ hereafter.
Let us now analyze in more detail the influence of the considered set of sources on performance. In the same way as for the parameters b_{im}, one may expect Set 2 to yield better estimates of the parameters $\mu_{im'}$ than Set 1, because it contains silence phases. The success rates in Table 2 confirm that expectation. Moreover, the mean values of $SNRI$ in this table show the same phenomenon (except in one case: $\lambda = 0.9$ and $\eta = 20$).
Tables 3 to 5 provide a more detailed analysis of some aspects of the above tests: they only contain the overall values of the considered performance criteria when considering all the values of η (i.e. $\eta = 0, 10$ or 20), but each of these tables details the variations of these criteria vs one of the parameters of the BSS method (while averaging over the others). Table 3 first shows that the mean $SNRI$ and success rate have a low sensitivity to the overlap ρ' of STFT windows. Based on these results, we preferably choose $\rho' = 75\%$.
Table 4 contains the variations of the mean $SNRI$ and success rate with respect to the number d' of points in the STFT windows. This table shows that the success rate significantly depends on d'. Better results are obtained for intermediate values of d', i.e. $d' = 1024$ or $d' = 2048$, which yield success rates almost always equal to 100%, and always above 89%. This phenomenon may be explained as follows. If we compute the STFTs with too few samples, then the frequency accuracy of our TF representation is low and we are not able to estimate accurately the time shifts $\mu_{im'}$, which is in agreement with [19]. On the contrary, if the number of samples is too high, it is more difficult to find single-source TF zones in Set 1, because the time width of the STFTs is too large. Note that when $d' = 1024$ or $d' = 2048$, the mean $SNRI$s are always over 24 dB.
As explained above, the number M' of windows in constant-time analysis zones $(n_{p'}, \Omega)$ defines the frequency width of these zones. Table 5 shows that it should be set to the intermediate tested value, i.e. $M' = 32$ for $d' = 2048$.
To summarize, in the considered conditions, the TIFROM-CF method should preferably be operated with $d' = 2048$ samples (or 1024: both values yield almost the same performance) and $\rho' = 75\%$ of overlap in the STFT computations. One may then set the width of constant-time analysis zones to $M' = 32$ (or $M' = 16$ if $d' = 1024$). The mean $SNRI$s with both sets of parameters are then equal to 47 and 44.3 dB.
As an example, we detail the application of the proposed method with the above chosen parameters, i.e. $d' = 2048$, $\rho' = 75\%$ and $M' = 32$. The source signals are those of Set 2 which last 5 s , as you can see in Fig. 1. The mixing parameters are set to

$$\begin{cases} \lambda &= 0.5 \\ \eta &= 20 \end{cases} \tag{53}$$

The corresponding matrix $B(\omega)$ defined in (11) is therefore equal to

$$\begin{bmatrix} 1 & 1 \\ 0.5\,e^{-j\omega.20} & 2\,e^{j\omega.20} \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 1 \\ 2\,e^{j\omega.20} & 0.5\,e^{-j\omega.20} \end{bmatrix}, \tag{54}$$

depending whether it corresponds to a non-permuted or permuted version of the source signals.
Although the TF transforms of these source signals have significant differences (see Fig. 2 and 3), the TF transforms of the resulting mixed signals are quite similar (see Fig. 4 and 5). Nevertheless, the TIFROM-CF method succeeds in separating these signals with a high accuracy, as will now be shown.
The estimated matrix $\hat{B}(\omega)$ thus obtained is equal to

$$\hat{B}(\omega) = \begin{bmatrix} 1 & 1 \\ 2.0064\,e^{j\omega.20} & 0.4914\,e^{-j\omega.20} \end{bmatrix}, \tag{55}$$

which is very close to the second expression in (54). Consequently, the estimated output signals are almost equal to the (scaled and permuted) sources, as confirmed by Fig 6, 7 and 8. The $SNRI$ is thus equal to 38.6 dB.

6.2.3 Performance of TIFROM-CT and comparison with TIFROM-CF

In this subsection, we analyze the performance of the TIFROM-CT method described in Section 5. Table 6 shows that its mean $SNRI$ ranges from 27.4 to 61.3 dB, with a global mean over all tests equal to 39.6 dB. The success rate over all these tests is 88.1%. The method failed to identify at least one column of $B(\omega)$ in 45 of the 648 considered tests. These 45 tests represent 58.4% of the cases when the proposed method could not identify exactly all parameters $\mu_{im'}$. The TIFROM-CT method therefore yields somewhat lower overall performance than TIFROM-CF. Comparing Tables 2 and 6 shows that, depending on the considered sources and mixing conditions, the best performance is achieved either by TIFROM-CT or by TIFROM-CF.
Tables 7 to 9 provide a more detailed analysis of some aspects of the above tests, using the same approach as in Tables 3 to 5. Table 7 first shows that the mean $SNRI$ and success rate have a low sensitivity to the overlap ρ' of STFT windows, with slightly better performance for $\rho' = 90\%$.
Table 8 contains the variations of the mean $SNRI$ and success rate with respect to the number d' of points in the STFTs. It shows that the optimum value of d' depends on η. d' should preferably be set to 1024 when $\eta = 0$. The mean $SNRI$ then ranges from 60.3 to 66.2 dB and the success rate is always equal to 100%. When $\eta = 10$ or 20, the optimum values of d' range from 2048 to 8192, depending on the value of λ and the considered set of sources. The mean $SNRI$ then ranges from 36.0 to 48.1 dB, while the success rates vary from 44.4 to 100% and they are equal to 100% in most cases. By disregarding the case $\eta = 0$, a good compromise is obtained by choosing $d' = 4096$, which always yields a mean $SNRI$ above 36 dB. By comparing these results with the performance of TIFROM-CF provided in Table 4, TIFROM-CT is more sensitive to the choice of d' than TIFROM-CT. Comparing the columns of these tables corresponding to the preferred values of d' shows that, in these cases when TIFROM-CT always identifies all columns of $B(\omega)$: i) TIFROM-CT yields better performance than TIFROM-CF for Set 1 and ii) for Set 2, i.e. when the sources contain silence phases, the best method depends on the mixing conditions, i.e. on λ and η.
Table 9 shows that M' should be set to the intermediate tested value, i.e. $M' = 64$ for

$d' = 4096$.
As a result, a trade-off between these parameters may be obtained, for $\eta \neq 0$, by using $d' = 4096$ samples and $\rho' = 90\%$ of overlap in the STFT computations. One may then set the width of constant-time analysis zones to $M' = 64$. The mean $SNRI$ of TIFROM-CT with the selected parameters is equal to 45.2 dB. We thus see that both methods yield almost the same mean $SNRI$.
We then investigated the performance of both methods for much larger time shifts, i.e. $\eta = 200$. The results thus obtained are detailed in Appendix D and the main conclusions that may be drawn from these tests are presented hereafter.

7 Conclusions and extensions

Most reported TF BSS methods were developed for LI mixtures. The rare methods which were proposed for AD mixtures are very restrictive, i.e. they require the sources to be (approximately) W-disjoint orthogonal. In this paper, we avoid all these restrictions as follows. We were inspired by the TF BSS approach that we previously developed for LI mixtures, which is based on the TIme Frequency versions of Ratios Of Mixtures of source signals. We here introduced two extensions of this approach to AD mixtures, called "AD-TIFROM-CT" and "AD-TIFROM-CF", depending whether they only use Constant-Time analysis zones or also Constant-Frequency zones. The proposed methods consist in identifying the columns of the (filtered permuted) mixing matrix by first finding TF zones where only one source occurs and then independently estimating the scale coefficients and the time shifts in these single-source zones, again using ratios of mixtures. Thanks to this principle, these approaches apply to non-stationary sources, such as speech signals, but also to stationary and/or dependent sources [13], provided there exist at least a tiny TF zone per source where this source occurs alone.
We presented various aspects of the experimental performance of these approaches, derived from a large number of tests. Our first series of tests showed that our methods yield very good performance when the assumptions that we made for developing these methods are satisfied, i.e. when the time shifts η involved in the mixing stage are significantly lower than the temporal widths of the STFTs used in our BSS methods. The mean $SNRI$s in that case are above 44 dB with both methods and optimum parameters. More precisely, the mean $SNRI$s over BSS parameters decrease from more than 60 dB with LI mixtures (i.e. no time shifts) to more than 27 dB when the time shifts increased up to 20 samples. Moreover, our second series of tests performed with $\eta = 200$ showed that our methods still yield attractive $SNRI$s when time shifts become similar to STFT temporal widths: the mean $SNRI$s of both methods are then almost always above 18 dB when adapting the parameters of these methods to these larger time shifts (except that AD-TIFROM-CT yields low performance when $\eta = 200$ and scale coefficients are set to $\lambda = 0.9$). It should be mentioned that AD-TIFROM-CF yields better global performance than AD-TIFROM-CT.
Our future investigations will concern the underdetermined case, i.e. the situation when the number P of observations is lower than the number N of sources, where the approach that we introduced [13] for the LI-TIFROM method may be straightforwardly extended. We will also aim at extending the proposed approaches to general convolutive mixtures.

A Improved identification stage for the parameters b_{im}

The basic type of procedure for estimating the scale coefficients b_{im} which was presented in Subsection 4.1 was initially introduced in the case of LI mixtures. It yields very good experimental performance for such mixtures, with various numbers of sources, as shown in [14]. When applied to more complex mixtures involving delays, it was also satisfactory for mixtures of $N = 2$ sources involving moderate delays (i.e. $n_{ij} \leq 20$ for the speech sources considered in Section 6). However, for $N > 2$ or when delays were very large in the case $N = 2$, it turned out to yield false results in a significant number of experimental tests. More precisely, the parameters b_{im} of the columns of $B(\omega)$ were thus identified in analysis zones which were selected because they were at the beginning of the ordered list created in the detection stage, but these identified columns did not correspond to the actual (scaled permuted) mixing matrix, so that the outputs of our BSS system did not provide well separated sources. This problem can be solved thanks to clustering techniques, because only a few occurrences are obtained for each false column value, so that such columns may then be discarded by clustering methods. We now detail such an approach, that we developed and successfully tested for the considered type of mixtures, as shown in Section 6. In this approach, we form clusters of "points" where each point consists of a tentative column of parameters b_{im}. To this end, we only consider the analysis zones which are such that

$$min\left\{MVAR\left[|\alpha|\right](T,\omega_l), MVAR\left[|\beta|\right](T,\omega_l)\right\} \leq \epsilon_2, \tag{56}$$

where ϵ_2 is a small positive user-defined threshold. We thus only keep single-source zones, which correspond to the beginning of the ordered list created in the detection stage. We successively consider each of the first and subsequent analysis zones in this beginning of the ordered list and we use them in a slightly different way than in the basic identification procedure that we described above.

Here again, for each considered analysis zone, the estimates of the parameters b_{im} associated to a column of $B(\omega)$ are set to the values of $\overline{|\alpha_i|}(T,\omega_l)$ or $\frac{1}{\overline{|\beta_i|}(T,\omega_l)}$, depending on which of the parameters $MVAR\left[|\alpha|\right]$ and $MVAR\left[|\beta|\right]$ takes the lowest value in this zone. The estimated column associated to the first zone in the ordered list is kept as the first point in the first cluster. Each subsequently estimated column is then used as follows. We compute its distances with respect to all clusters created up to this stage, where the distance with respect to a cluster is defined as the distance with respect to the first point which was included in it. If such a distance is below a user-defined threshold ϵ_1, this new column is inserted as a new point in the corresponding cluster. Otherwise, this new column is kept as the first point of a new cluster. This is repeated for all analysis zones which fulfill condition (56). If the threshold ϵ_1 is low enough, the number of clusters thus created is at least equal to the specified number N of sources to be extracted. We then keep the N clusters which contain the highest numbers of points. For each cluster, we eventually derive a representative, by selecting its point which corresponds to the lowest value of $min\left\{MVAR\left[|\alpha|\right](T,\omega_l), MVAR\left[|\beta|\right](T,\omega_l)\right\}$ and thus presumably to the best single-source zone[7].

This yields the N columns of estimates of parameters b_{im}. Let us stress again that this may be considered as a particular, simple method for obtaining more robust estimates of these parameters by clustering techniques, and that many other related methods may also be developed, some of them being more powerful but also more complex.

[7] Other representatives may be used instead, such as the center of gravity of the points in the considered cluster.

B Detection of constant-time single-source analysis zones and identification of the parameters μ_{im}

We presented in Subsection 4.2 the main principles of the method for identifying the time shifts $\mu_{im'}$. That approach may then be straightforwardly extended so as to also perform the detection of the constant-time single-source analysis zones which are required in this identification procedure, as will now be shown. The overall detection and identification method that we propose for the parameters $\mu_{im'}$ then operates as follows. We successively consider all constant-time analysis zones $(n_{p'}, \Omega)$ that may be associated to the available TF points, i.e. the zones corresponding to all values of $n_{p'}$ and to adjacent frequency areas Ω which each contain M' adjacent frequency points $\omega_{l'}$. In each such zone $(n_{p'}, \Omega)$, we determine the above-defined regression line and the mean-square error of the points $(\omega_{l'}, \phi'_i(n_{p'}, \omega_{l'}))$ with respect to their associated regression line. The best single-source zones are those which yield the lowest mean-square error[8]. These zones may be then be used in various ways for eventually identifying the parameters $\mu_{im'}$, e.g. by ordering these zones according to increasing values of their mean-square error or by using clustering techniques, in the same way as in the identification of the parameters b_{im} that we described above. We do not detail these procedures for the following reason. They eventually yield a set of column vectors. Each of these vectors contains the values $\mu_{im'}$, which correspond to all observations with indices i and to the source with index $k' = \sigma(m')$ which occurs in the considered analysis zone. These vectors are obtained in an arbitrary order, depending on which source occurs in which of the considered time-constant analysis zones (this relates to the permutation issue in BSS). The same comment applies to the identification of the parameters b_{im} that we previously performed with the methods described in Subsection 4.1. It should be remembered that we eventually aim at identifying the parameters which define each column of $B(\omega)$, so that we should determine which column of parameters b_{im} goes with which column of parameters $\mu_{im'}$, i.e. corresponds to the same source. This is not yet defined at this stage of our discussion, since these two types of parameters were obtained independently and each of them in an arbitrary order as mentioned above. A final stage should therefore be added to our approach, in order to couple each column of parameters $\mu_{im'}$ to a column of parameters b_{im}. This stage is described in Subsection 4.3.

C SNR of mixed signals and performance criteria of BSS methods

We here define the Signal to Noise Ratio (SNR) associated to the mixed signals which are processed by our BSS methods and the main parameters used to measure the performance of these methods in the tests reported in Section 6.
First consider a single source, with a given index[9] k. We define the input SNR of our BSS system associated to source k by using the following two-stage approach. As a first stage, we consider a single input with index i of the BSS system, which receives the mixed signal

[8] Standard phase unwraping procedures sometimes have a spurious effect, i.e. they keep a phase discontinuity equal to a multiple of 2π at some (rare) frequencies. The above description shows that this is not a major problem in our approach: if such a discontinuity occurs inside a single-source analysis zone, it will result in a high regression error in this zone. A very small set of single-source zones may thus be missed in the detection of single-source zones. This does not make the proposed method fail however, because practical signals yield enough single-source zones and even if a few of them are missed, others will be used instead.

[9] The index of each source signal is known when testing our BSS methods with given source signals.

$x_i(n)$. This signal consists of:

1. A contribution from the source with index k. This contribution is considered as the signal of interest contained by input i of the BSS system and is equal to $a_{ik}s_k(n-n_{ik})$.

2. Contributions from all others sources with indices $j \neq k$. These sources are considered as interfering signals or "noise" contained by input i of the BSS system. Their overall contribution in $x_i(n)$ is equal to $x_i(n) - a_{ik}s_k(n-n_{ik})$.

The elementary input SNR of any of the proposed BSS systems, associated to its input i and to source k, is then defined as the ratio of the powers of the above signal and noise contributions[10], i.e.

$$SNR_k^{in}(i) = \frac{E\left\{|a_{ik}s_k(n-n_{ik})|^2\right\}}{E\left\{|x_i(n) - a_{ik}s_k(n-n_{ik})|^2\right\}}. \tag{57}$$

As a second stage, we define the overall input SNR of the considered BSS system associated to source k, i.e. when taking account all observed signals $x_i(n)$ used as the inputs of this system. This overall SNR is defined as

$$SNR_k^{in} = \max_{i=1\ldots N}\left(SNR_k^{in}(i)\right), \tag{58}$$

i.e., for the considered source k, we take into account the observed signal where this source has the highest SNR.
We then use the same approach for defining the output SNR of the considered BSS system. Therefore, we first consider source k and output i of the BSS system, which provides the signal $y_i(n)$. This signal consists of:

1. The useful contribution associated to output i of the BSS system. This contribution is defined as the ideal value of output $y_i(n)$ when the source extracted on that output is source k. Due to the principle of the considered BSS methods which were presented in Sections 4 and 5, this ideal output is equal to the contribution of source k in the first mixed signal, i.e. $a_{1k}s_k(n-n_{1k})$ as shown in (6).

2. In the same way as in input signals, the noise contribution in output i is then defined as the remainder of $y_i(n)$, i.e. it is equal to $y_i(n) - a_{1k}s_k(n-n_{1k})$.

The elementary output SNR of the considered BSS system, associated to its output i and to source k, is then defined as the ratio of the powers of the above signal and noise contributions, i.e.

$$SNR_k^{out}(i) = \frac{E\left\{|a_{1k}s_k(n-n_{1k})|^2\right\}}{E\left\{|y_i(n) - a_{1k}s_k(n-n_{1k})|^2\right\}}. \tag{59}$$

We then define the overall output SNR of the considered BSS system associated to source k as

$$SNR_k^{out} = \max_{i=1\ldots N}\left(SNR_k^{out}(i)\right), \tag{60}$$

[10] In our tests, we first centered each overall time series defining one source signal.

We then define the SNR Improvement (SNRI) achieved by the BSS system with respect to source k as

$$SNRI_k = \frac{SNRI_k^{out}}{SNRI_k^{in}}. \tag{61}$$

All above parameters only refer to a single source. The corresponding overall features of the BSS system are eventually defined as the geometrical means of each considered parameter over all sources k, i.e. as the arithmetic means of this parameter expressed in dB. This yields explicitly

$$SNR^{in} = \Big(\prod_{k=1}^{N} SNR_k^{in} \Big)^{\frac{1}{N}} \quad \text{and} \quad (SNR^{in})_{dB} = \frac{1}{N} \sum_{k=1}^{N} (SNR_k^{in})_{dB} \tag{62}$$

$$SNR^{out} = \Big(\prod_{k=1}^{N} SNR_k^{out} \Big)^{\frac{1}{N}} \quad \text{and} \quad (SNR^{out})_{dB} = \frac{1}{N} \sum_{k=1}^{N} (SNR_k^{out})_{dB} \tag{63}$$

$$SNRI = \Big(\prod_{k=1}^{N} SNRI_k \Big)^{\frac{1}{N}} \quad \text{and} \quad (SNRI)_{dB} = \frac{1}{N} \sum_{k=1}^{N} (SNRI_k)_{dB} \tag{64}$$

Note that this also entails

$$\frac{SNR^{out}}{SNR^{in}} = \Big(\prod_{k=1}^{N} \left[\frac{SNR_k^{out}}{SNR_k^{in}} \right] \Big)^{\frac{1}{N}} = \Big(\prod_{k=1}^{N} SNRI_k \Big)^{\frac{1}{N}} = SNRI \tag{65}$$

and therefore

$$(SNRI)_{dB} = (SNR^{out})_{dB} - (SNR^{in})_{dB}. \tag{66}$$

The parameters SNR^{out} and/or $SNRI$ are used to measure the performance of the considered BSS system, whereas SNR^{in} indicates to which extent the signals processed by this system are mixed.

D Tests with large time shifts

In this appendix, we study the behaviour of our methods for a much larger η than in Section 6, i.e. $\eta = 200$, still considering the same sources and values of λ as in Section 6. We again perform tests in conditions such that $\frac{\eta}{min\{d,d'\}} \leq \frac{1}{10}$ but we also consider cases when η and d are similar, in order to analyze the resulting performance of TIFROM-CF. More precisely, the number d of samples in STFTs of constant-frequency analysis zones was geometrically varied from 256 to 2048 samples. The other parameters of these zones were fixed to the same values as above for the sake of simplicity, i.e.

- $M = 10$,
- $\rho = 75\%$.

Similarly, when we computed the STFTs used in constant-time zones, we decided to fix some parameters for the sake of simplicity. Thus, the temporal overlap ρ' was fixed to 75% for TIFROM-CF and 90% for TIFROM-CT. Note that these values are the optima that we selected in Subsection 6.2. The other parameters of STFTs were varied as follows:

- The number d' of samples per STFT was geometrically varied from 4096 to 16384

samples[11].

- The number M' of windows per analysis zone was successively set to 32, 64 or 132 when $d' = 4096$. This range of values of M' was then increased geometrically with d'.

The other parameters of the methods were fixed and equal to the values used in Subsection 6.2.

D.1 Performance of TIFROM-CF

In these tests, the estimates of b_{im} only depend on the number d of samples in the STFTs. Table 10 provides the Frobenius norm of the difference between the estimated and actual matrices of parameters b_{im}. The proposed method succeeds in identifying the parameters b_{im}, except in two cases: with Set 1, $d = 256$ and $\lambda = 0.5$, we do not estimate the parameters b_{im} correctly, while with Set 2, $d = 2048$ and $\lambda = 0.9$, one source is not visible. These two cases illustrate the potential limitations of the proposed approach. On the one hand, performance degradation is expected when $\frac{\eta}{d}$ becomes too large (here $\frac{\eta}{d} \simeq 1$ when $d = 256$). On the other hand, using large analysis zones, i.e. a high d, makes some sources invisible. Optimum performance is therefore expected for intermediate values of d'. This is confirmed by Table 10, where the best global performance is obtained for $d = 1024$ (or 512). Note that this results in $\frac{\eta}{d} \simeq 0.2$ (or 0.4), i.e. slightly larger values than initially expected, due to the above-mentioned limitation set by invisible sources.
Table 11 shows the overall performance of TIFROM-CF. The mean $SNRI$ thus varies from -17.1 to 30.3 dB and the method succeeds in 45.6% of the tests. Let us stress again that this includes cases resulting in poor performance because we intentionally considered situations when the proposed method does not apply since the condition $\frac{\eta}{d} \ll 1$ is not met at all. This table shows that the best $SNRI$ and success rates are obtained for $d = 512$ or 1024, which is in agreement with the results obtained above for the parameters b_{im}.
Let us now study the influence of the other parameters of our method. In table 12, we provide the variations of the mean $SNRI$ and success rate of this method with respect to the values of d' and d. The best results are almost always obtained with $d' = 8192$. One could be surprised that the optimal value of d' is significantly larger than the optimum of d. It should be remembered, however, that d' is involved in constant-time analysis zones, and that the temporal width of these zones is exactly d' samples. On the contrary, d is involved in constant-frequency analysis zones, which here consist of $M = 10$ time-adjacent d-sample overlapping windows, with an overlap $\rho = 75\%$. The optimum values of d and d' are therefore quite different but they result in relatively similar temporal widths for the corresponding two types of analysis zones.
Table 13 proves that M' should be preferably set to the intermediate or highest tested values, i.e. $M' = 128$ or $M' = 256$ when $d' = 8192$.
As a result, a trade-off between these parameters may be obtained by using 1024 samples (resp. 8192) in the STFT computations for the identification of the scale coefficients b_{im} (resp. the time shifts $\mu_{im'}$). One may fix the width of constant-time analysis zones to $M' = 128$ (or $M' = 256$). The mean $SNRI$s of TIFROM-CF with both sets of parameters are thus equal to 20.2 and 18.8 dB. Note that for $\lambda = 0.5$, each of these mean $SNRI$s is equal to 24.3 dB, while they are resp. equal to 16.1 and 13.2 dB when $\lambda = 0.9$.

[11] The case $d' = 32768$ was also tested but, in each test, our methods did not find all the columns of $B(\omega)$. This case is therefore not considered hereafter.

D.2 Performance of TIFROM-CT and comparison to TIFROM-CF

Table 14 shows that the TIFROM-CT method yields mean $SNRI$s varying from -16.8 to 25.4 dB. The method succeeds in 44.4% of the tests and hardly identifies the mixing matrix $B(\omega)$ when $\lambda = 0.9$, unlike TIFROM-CF. On the contrary, TIFROM-CT yields somewhat better optimum results than TIFROM-CF when $\lambda = 0.5$ (see Tables 11 and 14).
Table 14 also shows that $d' = 4096$ yields the best performance with continuous speech, while Set 2 allows us to use higher values, i.e $d' = 8192$ (or $d' = 16384$). As TIFROM-CT almost always yields bad performance for $\lambda = 0.9$, we here only consider $\lambda = 0.5$ for selecting d', so that we preferably set $d' = 4096$.
The influence of M' is shown in Table 15. The optimum value of M' depends on the considered conditions, so that, by again disregarding the case $\lambda = 0.9$, a trade-off would consist in choosing its intermediate tested value, i.e. $M' = 64$ when $d' = 4096$.
To summarize, a trade-off between the parameters of TIFROM-CT may be obtained by using $d' = 4096$ samples in the STFT computations. One may then fix the width of constant-time analysis zones to $M' = 64$. The mean $SNRI$ of TIFROM-CT is then equal to 18.7 dB when $\lambda = 0.5$. This result is significantly lower than the mean $SNRI$ obtained with TIFROM-CF in that configuration (24.3 dB).

References

[1] A. Hyvärinen, J. Karhunen, E. Oja: Independent Component Analysis, Wiley-Interscience, New York, 2001.

[2] A. Belouchrani, M. Amin: Blind source separation based on time-frequency signal representations, IEEE Transactions on Signal Processing, vol. 46, no. 11, pp. 2888-2897, November 1998.

[3] A. Holobar, C. Févotte, C. Doncarli, D. Zazula: Single autoterms selection for blind source separation in time-frequency plane, Proceedings of the 11th European Signal Processing Conference (EUSIPCO 2002), Toulouse, France, September 3-6, 2002.

[4] C. Févotte, C. Doncarli: Two contributions to blind source separation using time-frequency distributions, IEEE Signal Processing Letters, vol. 11, no. 3, pp. 386-389, March 2004

[5] A. Jourjine, S. Rickard, O. Yilmaz : Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures, Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2000), IEEE Press, Istanbul, Turkey, June 5-9, 2000, vol. 5, pp. 2985-2988.

[6] S. Rickard, R. Balan, J. Rosca: Real-time time-frequency based blind source separation, Proceedings of the 3rd International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2001), December 9-13, 2001, San Diego, USA, pp. 651-656.

[7] S. Rickard, O. Yilmaz: On the approximate W-disjoint orthogonality of speech, Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002), Orlando, USA, May 13-17, 2002.

[8] R. Balan, J. Rosca, S. Rickard: Non-square blind source separation under coherent noise by beamforming and time-frequency masking, Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003), Nara, Japan, April 2003, pp. 313-318.

[9] M. Baeck, U. Zölder: Performance analysis of a source separation algorithm, Proceedings of the 5th Digital Audio Effects (DAFx-02), Hamburg, Germany, September 26-28 2002, pp. 207-210.

[10] F. Abrard, Y. Deville, P. White: A new source separation approach based on time-frequency analysis for instantaneous mixtures, Proceedings of the 5th International Worshop on Electronics, Control, Modelling, Measurement and Signals (ECM2S 2001), pp. 259-267, Toulouse, France, May 30 - June 1, 2001.

[11] F. Abrard: Méthodes de séparation aveugle de sources et applications : des statistiques d'ordre supérieur à l'analyse temps-fréquence, Ph.D, Université Paul Sabatier, Toulouse, France, 2003.

[12] F. Abrard, Y. Deville, P. White: From blind source separation to blind source cancellation in the underdetermined case: a new approach based on time-frequency analysis, Proceedings of the 3rd International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2001), San Diego, California, USA, December 9-13, 2001.

[13] F. Abrard, Y. Deville: A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources, to appear in Signal Processing.

[14] Y. Deville, M. Puigt, B. Albouy: Time-frequency blind signal separation: extended methods, performance evaluation for speech sources, Proceedings of the International Joint Conference on Neural Networks (IJCNN 2004), IEEE Catalog Number: 04CH37541C, ISBN: 0-7803-8360-5, pp. 255-260, Budapest, Hungary, July 25-29, 2004.

[15] Y. Deville: Temporal and time-frequency correlation-based blind source separation methods, Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003), pp. 1059-1064, Nara, Japan, April 1-4, 2003.

[16] Y. Deville, M. Puigt: Temporal and time-frequency correlation-based blind source separation methods. Part I: linear instantaneous mixtures, submitted to Signal Processing.

[17] D. Smith, J. Lukasiak and I. Burnett: Two Channel, block adaptative audio separation using the cross correlation of time frequency information, Proceedings on the 5th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2004), Granada, Spain, September 22-24, 2004.

[18] B. Albouy, Y. Deville: A time-frequency blind source separation method based on segmented coherence function, Proc. of the 7th International Work-conference on Artificial and Natural Neural Networks (IWANN 2003), special session, vol. 2, pp. 289-296, J. Mira and J. R. Alvarez eds (Springer), Mao, Menorca, Spain, June 3-6, 2003.

[19] R. Balan, J. Rosca, S. Rickard, and J. O'Ruanaidh: The Influence of Windowing on Time Delay Estimation, Proceedings of the 35th Annual Conference on Information Sciences and Systems (CISS 2000), Volume 1, Pages WP1(15-17), Princeton, NJ, March 2000.

[20] Y. Zhang, M. Amin: Blind separation of sources based on their time-frequency signatures, Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2000), Istanbul, Turkey, June 5-9, 2000, IEEE press, vol.5, pp.3065-3068.

[21] H. Wu, J. Principe, D. Xu: Exploring the time-frequency microstructure of speech for blind source separation, Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1998), IEEE Press, volume 2, pp. 1145-1149, paper no. 2161, Seattle, USA, May 12-15, 1998.

[22] S. Ikeda, N. Murata: An Approach to Blind Source Separation of Speech Signals, Proceedings of International Conference on Neural Networks (IJCNN 1998), Skövde, Sweden, September 2-4, 1998.

[23] N. Murata, S. Ikeda: An On-line Algorithm for Blind Source Separation on Speech Signals, Proceedings of the International Symposium on Non Linear Theory and its Applications (NOLTA'98), Crans-Montana, Switzeland, September 14-17, 1998.

[24] F. Hlawatsch, G.F. Boudreaux-Bartels: Linear and Quadratic Time-Frequency Signal Representations, IEEE SP Magazine, April 1992, pp. 21-67.

Matthieu Puigt was born in Perpignan, France, in 1980. He first studied mathematics at the University of Perpignan where he received in 2002 the Master's degree in mathematical engineering. He then studied physics at the University Paul Sabatier of Toulouse (France) where he received in 2003 the D.E.A degree in acoustics. He is now a Ph.D student in signal processing and blind source separation.

Yannick Deville was born in Lyon, France, in 1964. He graduated from the Ecole Nationale Supérieure des Télécommunications de Bretagne (Brest, France) in 1986. He received the D.E.A and Ph.D degrees, both in Microelectronics, from the University of Grenoble (France), in 1986 and 1989 respectively. From 1986 to 1997, he was a Research Scientist at Philips Research Labs (Limeil, France). His investigations during this period concerned various fields, including GaAs integrated microwave RC active filters, VLSI cache memory architectures and replacement algorithms, neural network algorithms and applications, and nonlinear systems. Since 1997, he has been a Professor at the University of Toulouse (France). From 1997 to Oct. 2004, he was with the Acoustics lab of that University. Since Oct. 2004, he has been with the Astrophysics lab in Toulouse, which is part of the University but also of the French National Center for Scientific Research (CNRS) and of the Midi-Pyrénées Observatory. Yannick Deville's current major research interests include signal processing, higher-order statistics, time-frequency analysis, and especially blind source separation methods and their applications to Astrophysics, Acoustics and communication/electromagnetic signals.

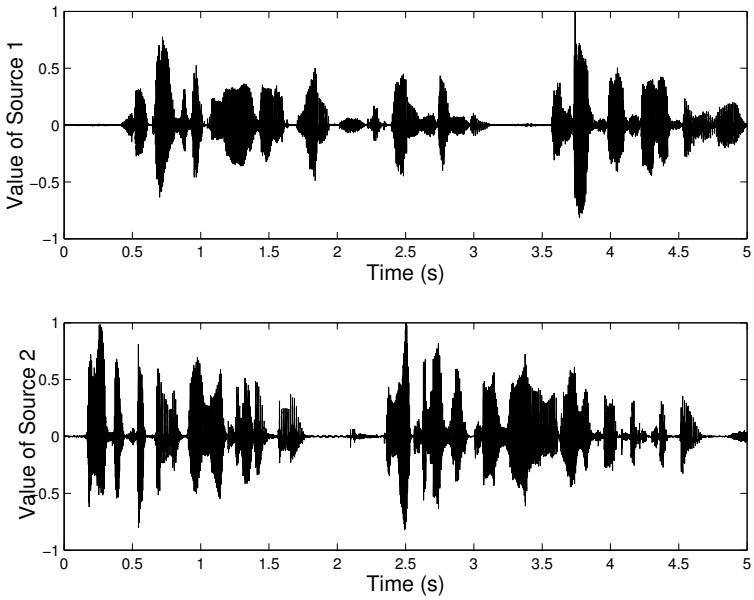

Figure 1: Temporal representations of sources.

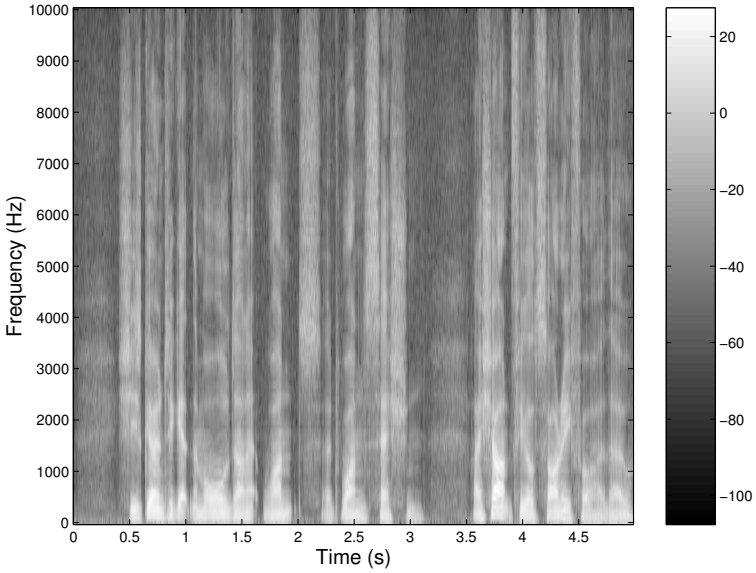

Figure 2: Time-frequency transform of Source 1.

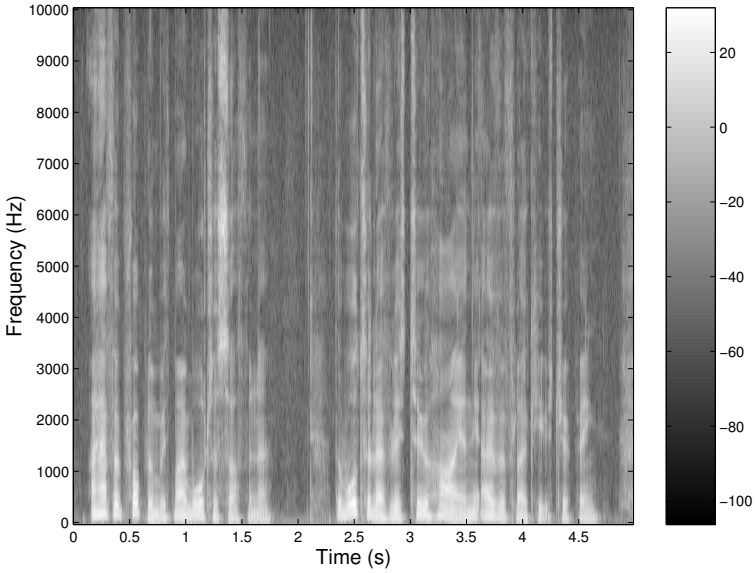

Figure 3: Time-frequency transform of Source 2.

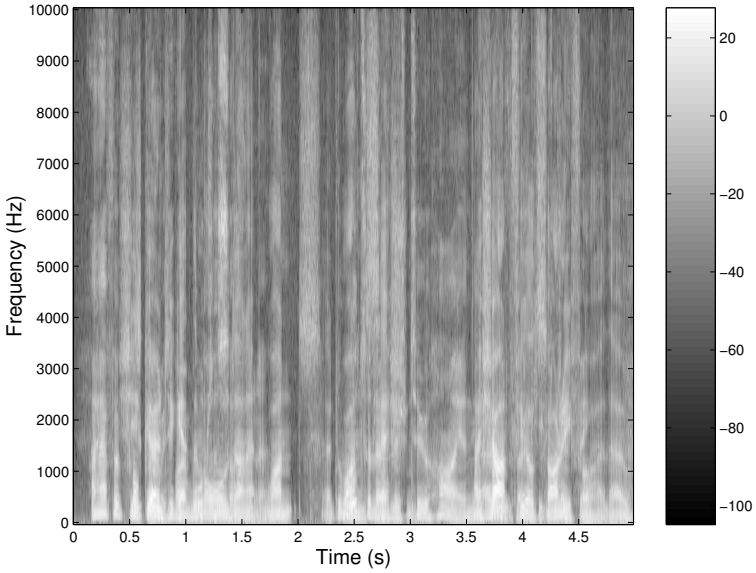

Figure 4: Time-frequency transform of Observation 1.

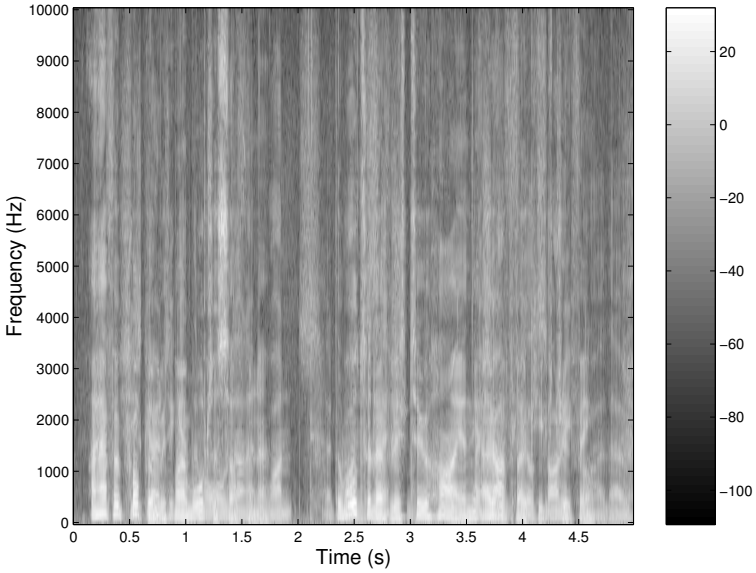

Figure 5: Time-frequency transform of Observation 2.

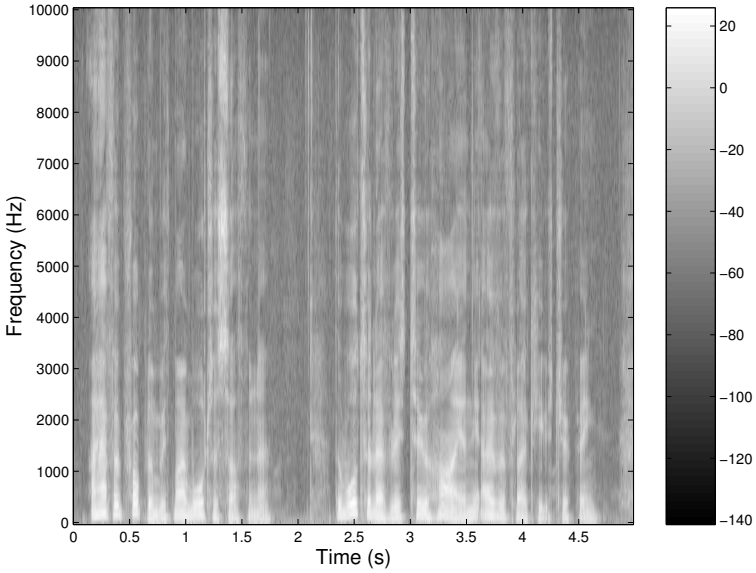

Figure 6: Time-frequency transform of Output signal 1.

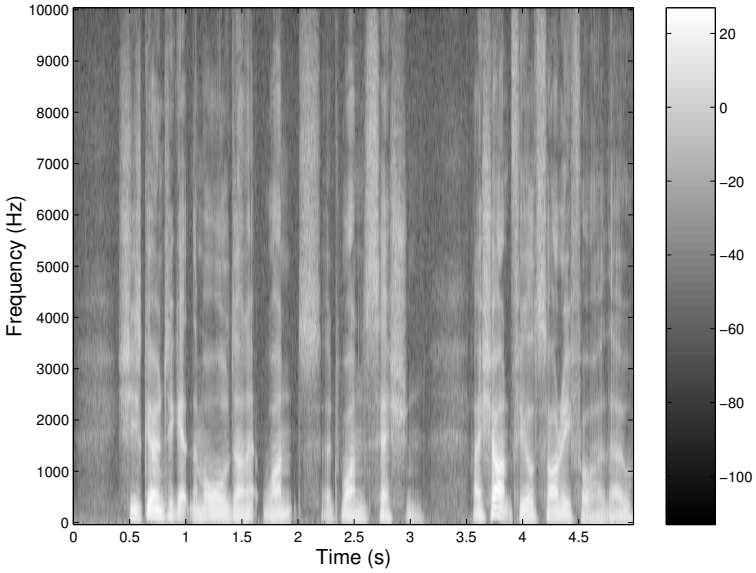

Figure 7: Time-frequency transform of Output signal 2.

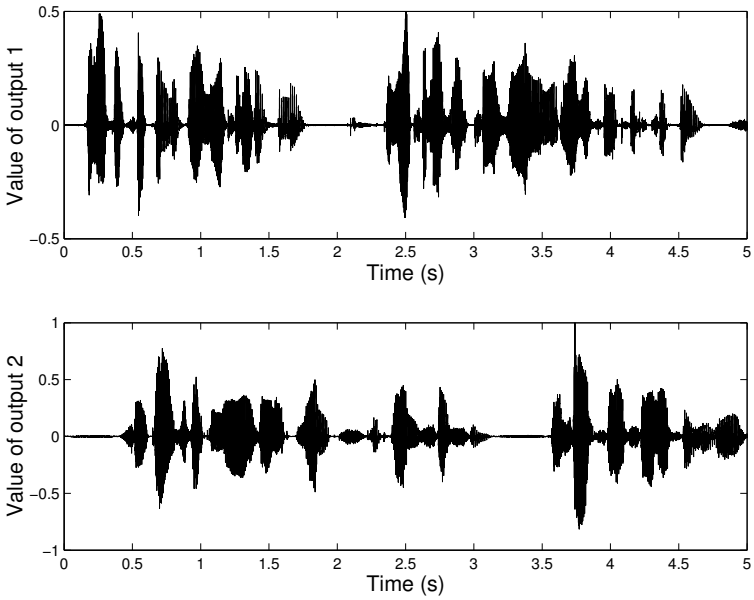

Figure 8: Temporal representations of output signals.

λ	Set	η		
		0	10	20
0.5	1	0.0050	0.0803	0.0204
	2	0.0005	0.0040	0.0107
0.9	1	0.0001	0.0162	0.0190
	2	0.0001	0.0019	0.0116

Table 1: Frobenius norm of the difference between the actual matrices of parameters b_{im} and their estimates provided by TIFROM-CF, vs set of sources and parameters λ and η of the mixing matrix defined in (51).

λ	Set	perf. criterion	η			
			0	10	20	0-20
0.5	1	$SNRI$	52.7	16.5	27.1	32.1
		SNR^{out}	58.7	22.5	33.1	38.1
		% of success	83.3	53.7	64.8	67.3
0.5	2	$SNRI$	61.9	45.2	35.2	47.4
		SNR^{out}	67.9	51.2	41.2	53.4
		% of success	98.1	88.9	88.9	92.0
0.9	1	$SNRI$	66.0	27.4	40.2	44.5
		SNR^{out}	66.9	28.5	41.1	45.4
		% of success	100	74.1	81.5	85.2
0.9	2	$SNRI$	66.3	49.1	34.7	50.3
		SNR^{out}	67.2	50.0	35.6	51.2
		% of success	100	92.6	96.3	96.3

Table 2: Performance of TIFROM-CF, for each set of sources, each value of λ, $\eta = 0, 10$ or 20 and global performance for η = 0-20. Performance criteria: i) mean values of $SNRI$ and SNR^{out} (in dB) and ii) success rate over all parameter values of BSS method.

λ	perf. criterion	ρ'		
		50	75	90
0.5	$SNRI$	37.8	41.0	40.2
	% of success	73.1	81.5	84.3
0.9	$SNRI$	44.7	45.8	44.7
	% of success	88.9	95.6	90.7

Table 3: Global performance of TIFROM-CF for both sets of sources, η = 0-20 and for each value of λ, vs overlap ρ' between STFT windows. Performance criteria: mean value of $SNRI$ (in dB) and success rate.

η	λ	Set	perf. criterion	d'					
				512	1024	2048	4096	8192	16384
0	0.5	1	$SNRI$	55.8	62.1	62.1	62.1	44.0	30.0
			% of success	88.9	100	100	100	66.7	44.4
		2	$SNRI$	56.5	62.9	62.9	62.9	62.9	62.9
			% of success	88.9	100	100	100	100	100
	0.9	1	$SNRI$	66.0	66.0	66.0	66.0	66.0	66.0
			% of success	100	100	100	100	100	100
		2	$SNRI$	66.2	66.2	66.2	66.2	66.2	66.2
			% of success	100	100	100	100	100	100
10	0.5	1	$SNRI$	20.3	24.6	24.6	17.7	6.8	5.0
			% of success	66.7	100	100	44.4	11.1	0
		2	$SNRI$	35.1	45.9	49.6	49.6	49.6	41.7
			% of success	66.7	88.9	100	100	100	77.8
	0.9	1	$SNRI$	29.2	35.4	35.4	30.0	32.1	2.2
			% of success	77.8	100	100	77.8	88.9	0
		2	$SNRI$	41.2	52.6	52.6	52.6	52.6	43.1
			% of success	77.8	100	100	100	100	77.8
20	0.5	1	$SNRI$	25.7	34.4	34.4	34.4	26.8	7.2
			% of success	55.6	88.9	88.9	88.9	66.7	0
		2	$SNRI$	25.0	38.6	38.6	38.6	38.6	31.5
			% of success	55.6	100	100	100	100	77.8
	0.9	1	$SNRI$	22.8	33.0	33.0	33.0	29.9	11.9
			% of success	66.7	100	100	100	88.9	33.3
		2	$SNRI$	27.2	36.1	36.1	36.1	36.1	36.1
			% of success	77.8	100	100	100	100	100
0-20	0.5	1-2	$SNRI$	36.4	29.8	45.4	44.2	38.1	30.3
			% of success	70.4	96.3	98.2	88.9	74.1	50
	0.9	1-2	$SNRI$	42.1	48.3	48.3	47.5	47.2	37.6
			% of success	83.4	100	100	96.3	96.3	68.6

Table 4: Performance of TIFROM-CF, for each set of sources, each value of λ, $\eta = 0, 10$ or 20 and global performance for both sets of sources and η = 0-20, vs STFT window size d' (in samples). Performance criteria: same as previous table.

λ	perf. criterion	frequency width		
		156.25	312.5	625
0.5	$SNRI$	36.1	42.0	41.3
	% of success	67.6	85.2	86.1
0.9	$SNRI$	40.0	46.7	45.7
	% of success	83.3	95.4	93.5

Table 5: Global performance of TIFROM-CF for both sets of sources, η = 0-20 and for each value of λ, vs width of constant-time analysis zones (in Hz).

λ	Set	perf. criterion	η			
			0	10	20	0-20
0.5	1	$SNRI$	46.5	31.7	28.0	35.4
		SNR^{out}	52.5	37.7	34.0	41.4
		% of success	87.0	74.1	66.7	75.9
0.5	2	$SNRI$	53.9	38.8	33.4	42.0
		SNR^{out}	59.9	44.8	39.4	48.0
		% of success	100	96.3	92.6	96.3
0.9	1	$SNRI$	50.2	34.7	27.4	37.4
		SNR^{out}	52.1	35.6	28.3	38.3
		% of success	100	79.6	72.2	84.0
0.9	2	$SNRI$	61.3	38.3	31.4	43.7
		SNR^{out}	62.2	39.2	32.3	44.6
		% of success	100	100	88.9	96.3

Table 6: Performance of TIFROM-CT, for each set of sources, each value of λ, $\eta = 0, 10$ or 20 and global performance for η = 0-20.

λ	perf. criterion	ρ'		
		50	75	90
0.5	$SNRI$	37.6	38.9	39.7
	% of success	79.6	87.0	88.0
0.9	$SNRI$	39.0	40.7	41.9
	% of success	88.0	91.7	90.8

Table 7: Global performance of TIFROM-CT for both sets of sources, η = 0-20 and for each value of λ, vs overlap ρ' between STFT windows.

η	λ	Set	perf. criterion	d'					
				512	1024	2048	4096	8192	16384
0	0.5	1	$SNRI$	58.7	61.2	46.7	41.6	30.2	24.1
			% of success	100	100	100	100	88.9	33.3
		2	$SNRI$	56.3	60.3	54.1	57.7	58.6	41.1
			% of success	100	100	100	100	100	100
	0.9	1	$SNRI$	64.1	66.0	53.6	50.1	37.9	29.3
			% of success	100	100	100	100	100	100
		2	$SNRI$	66.3	66.2	61.0	59.8	62.7	54.7
			% of success	100	100	100	100	100	100
10	0.5	1	$SNRI$	24.6	32.8	41.4	36.0	29.5	12.5
			% of success	77.8	100	100	88.9	77.8	0
		2	$SNRI$	28.7	29.0	38.5	48.1	44.9	44.1
			% of success	100	100	100	100	100	77.8
	0.9	1	$SNRI$	26.9	31.2	38.2	40.8	45.9	11.0
			% of success	88.9	100	100	100	66.7	0
		2	$SNRI$	29.9	29.6	40.7	41.4	44.9	43.6
			% of success	100	100	100	100	100	100
20	0.5	1	$SNRI$	20.8	30.5	35.1	36.0	19.4	12.6
			% of success	66.7	88.9	100	100	44.4	0
		2	$SNRI$	26.0	23.6	31.2	40.6	42.3	38.4
			% of success	100	88.9	100	100	100	66.7
	0.9	1	$SNRI$	16.2	24.6	31.7	37.1	34.7	9.7
			% of success	66.7	100	100	100	66.7	0
		2	$SNRI$	20.3	22.9	30.2	41.7	38.4	36.3
			% of success	66.7	100	100	100	100	66.7
0-20	0.5	1-2	$SNRI$	35.9	39.6	41.2	43.3	37.5	28.8
			% of success	90.8	96.3	100	98.2	85.2	46.3
	0.9	1-2	$SNRI$	37.3	40.1	42.6	45.2	44.1	30.8
			% of success	87.1	100	100	100	88.9	66.1

Table 8: Performance of TIFROM-CT, for each set of sources, each value of λ, $\eta = 0, 10$ or 20 and global performance for both sets of sources and η = 0-20 , vs STFT window size d' (in samples).

λ	perf. criterion	frequency width		
		156.25	312.5	625
0.5	$SNRI$	36.8	40.5	39.4
	% of success	86.1	89.8	80.6
0.9	$SNRI$	39.6	41.1	36.3
	% of success	87.0	93.5	89.8

Table 9: Global performance of TIFROM-CT for both sets of sources, η = 0-20 and for each value of λ, vs width of constant-time analysis zones (in Hz).

λ	Set	d			
		256	512	1024	2048
0.5	1	1.7605	0.0481	0.1263	0.1415
	2	0.0645	0.0859	0.0142	0.2025
0.9	1	0.0516	0.1759	0.1128	invisible
	2	0.0513	0.0511	0.0195	0.2556

Table 10: Frobenius norm of the difference between the actual matrices of parameters b_{im} and their estimates provided by TIFROM-CF in the case $\eta = 200$, vs set of sources and λ. This norm cannot be computed when one source is invisible.

λ	Set	perf. criterion	d			
			256	512	1024	2048
0.5	1	$SNRI$	-17.1	17.3	10.3	7.9
		% of success	0	55.6	33.3	22.2
	2	$SNRI$	19.3	22.4	10.3	14.5
		% of success	66.7	88.9	77.8	77.8
0.9	1	$SNRI$	6.8	7.2	10.3	invisible
		% of success	44.4	33.3	22.2	0
	2	$SNRI$	16.9	22.0	30.3	-14.3
		% of success	77.8	100	100	0

Table 11: Performance of TIFROM-CF, for each set of sources, each value of λ and $\eta = 200$, vs STFT window size d (in samples).

λ	Set	d	perf. criterion	d'		
				4096	8192	16384
0.5	1	256	$SNRI$	-17.1	-17.1	-17.1
			% of success	0	0	0
		512	$SNRI$	16.7	25.4	9.8
			% of success	33.3	100	33.3
		1024	$SNRI$	17.0	13.9	-0.0
			% of success	66.7	33.3	0
		2048	$SNRI$	12.4	11.4	-0.2
			% of success	33.3	33.3	0
	2	256	$SNRI$	21.5	24.4	12.0
			% of success	66.7	100	33.3
		512	$SNRI$	20.8	23.2	23.2
			% of success	66.7	100	100
		1024	$SNRI$	30.8	38.0	29.2
			% of success	66.7	100	66.7
		2048	$SNRI$	14.8	15.6	13.2
			% of success	66.7	100	66.7
0.9	1	256	$SNRI$	8.5	17.2	-5.1
			% of success	66.7	66.7	0
		512	$SNRI$	11.4	10.4	-0.4
			% of success	33.3	66.7	0
		1024	$SNRI$	14.7	-0.0	3.9
			% of success	66.7	0	0
		2048	$SNRI$	invisible	invisible	invisible
			% of success	0	0	0
	2	256	$SNRI$	17.4	21.3	12.0
			% of success	66.7	100	66.7
		512	$SNRI$	22.0	22.0	22.0
			% of success	100	100	100
		1024	$SNRI$	30.3	30.3	30.3
			% of success	100	100	100
		2048	$SNRI$	-15.8	-15.8	-11.3
			% of success	0	0	0

Table 12: Performance of TIFROM-CF, for each set of sources, each value of λ and $\eta = 200$, vs STFT window sizes d and d' (in samples).

λ	Set	d	perf. criterion	frequency width		
				156.25 Hz	312.5 Hz	625 Hz
0.5	1	256	$SNRI$	-17.1	-17.1	-17.1
			% of success	0	0	0
		512	$SNRI$	18.9	21.1	11.9
			% of success	66.7	66.7	33.3
		1024	$SNRI$	11.5	9.7	9.7
			% of success	33.3	33.3	33.3
		2048	$SNRI$	3.5	7.2	13.0
			% of success	0	0	66.7
	2	256	$SNRI$	17.2	24.4	16.4
			% of success	33.3	100	66.7
		512	$SNRI$	20.8	23.2	23.2
			% of success	66.7	100	100
		1024	$SNRI$	22.0	38.0	38.0
			% of success	33.3	100	100
		2048	$SNRI$	12.3	15.6	15.6
			% of success	33.3	100	100
0.9	1	256	$SNRI$	0.7	8.5	11.3
			% of success	0	66.7	66.7
		512	$SNRI$	9.6	4.6	7.2
			% of success	33.3	0	66.7
		1024	$SNRI$	6.1	9.7	2.7
			% of success	0	66.7	33.3
		2048	$SNRI$	invisible	invisible	invisible
			% of success	0	0	0
	2	256	$SNRI$	21.3	17.4	12.1
			% of success	100	66.7	66.7
		512	$SNRI$	22.0	22.0	22.0
			% of success	100	100	100
		1024	$SNRI$	30.3	30.3	30.3
			% of success	100	100	100
		2048	$SNRI$	-15.2	-15.2	-12.4
			% of success	0	0	0

Table 13: Performance of TIFROM-CF, for each set of sources, each value of λ and $\eta = 200$, vs STFT window size d and width of constant-time analysis zones (in Hz).

λ	Set	perf. criterion	d'			
			4096	8192	16384	global
0.5	1	$SNRI$	19.6	17.9	5.9	16.8
		% of success	66.7	33.3	0	33.3
	2	$SNRI$	14.6	25.4	15.2	18.4
		% of success	100	100	66.7	88.9
0.9	1	$SNRI$	-3.9	-5.7	invisible	-4.6
		% of success	33.3	0	0	11.1
	2	$SNRI$	-2.6	-16.8	19.4	-0.0
		% of success	33.3	0	100	44.4

Table 14: Performance of TIFROM-CT, for each set of sources, each value of λ and $\eta = 200$, vs STFT window size d' (in samples).

λ	Set	perf. criterion	frequency width		
			156.25	312.5	625
0.5	1	$SNRI$	14.3	15.5	26.5
		% of success	33.3	33.3	33.3
	2	$SNRI$	22.9	21.6	10.7
		% of success	100	100	66.7
0.9	1	$SNRI$	1.0	-3.7	-17.7
		% of success	33.3	0	0
	2	$SNRI$	5.7	-2.4	3.3
		% of success	66.7	33.3	33.3

Table 15: Performance of TIFROM-CT, for each set of sources, each value of λ and $\eta = 200$, vs width of constant-time analysis zones (in Hz).