Authors' final version of a paper to published in "Signal Processing"

Paper reference:
F. Abrard, Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources", Signal Processing, Vol. 85, Issue 7, pp. 1389-1403, July 2005,
http://authors.elsevier.com/sd/article/S016516840500054X

A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources

Frédéric Abrard, Yannick Deville*

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

abrard@i2e.fr - ydeville@ast.obs-mip.fr

Abstract

In this paper, we propose a new blind source separation (BSS) method called TIFROM (for TIme-Frequency Ratio Of Mixtures) which uses time-frequency (TF) information to cancel source signal contributions from a set of linear instantaneous mixtures of these sources. Unlike previously reported TF BSS methods, the proposed approach only requires slight differences in the TF distributions of the considered signals: it mainly requests the sources to be cancelled to be "visible", i.e. to occur alone in a tiny area of the TF plane, while they may overlap in all the remainder of this plane. By using TF ratios of mixed signals, it automatically determines these single-source TF areas and identifies the corresponding parts of the mixing matrix. This approach sets no conditions on the stationarity, independence or non-Gaussianity of the sources, unlike classical Independent Component Analysis methods. It achieves complete or partial BSS, depending on the numbers N and P of sources and observations and on the number of visible sources. It is therefore of interest for underdetermined mixtures (i.e. $N > P$), which cannot be processed with classical methods. Detailed results concerning mixtures of speech and music signals are presented and show that this approach yields very good performance.

Keywords: blind source separation, gaussianity, non-stationary signals, partial separation, single-source area, statistically dependent signals, time-frequency analysis, short-time Fourier transform, sparsity, TIFROM, underdetermined mixtures.

*Corresponding author.

1 Introduction

Blind source separation (BSS) consists in estimating N unknown sources from P observations resulting from the mixture of these sources through unknown propagation channels. Denoting $\mathcal{A}$ the mixing operator, the relationship between the sources and observations reads $\underline{x} = \mathcal{A}\underline{s}$, where the vector $\underline{s} = [s_1,\ s_2,\ \ldots,\ s_N]^T$ contains the unknown sources while $\underline{x} = [x_1,\ x_2,\ \ldots,\ x_P]^T$ represents the observations. We here only consider linear instantaneous mixtures, so that the operator $\mathcal{A}$ corresponds to a scalar matrix. Our investigation includes the so-called "underdetermined case", i.e. the configuration involving more sources than observations ($N > P$).

Linear instantaneous mixtures have been extensively studied since the first papers by J. Herault and C. Jutten [12]--[14]. One can find a review of most classical methods in [5]. These approaches basically aim at separating the sources by combining the observations by means of a matrix adapted so that the output signals are independent. The fundamental assumption of these techniques, known as Independent Component Analysis (ICA), is that the sources must be independent. Moreover, most of these approaches can only separate stationary non-Gaussian signals. Due to these limitations, poor performance is often obtained when dealing with real sources, like audio signals, which do not match those requirements. Some authors [8],[9],[17],[23],[24] have proposed different approaches which take advantage of the non-stationarity of such sources in order to achieve better performance than classical methods for this type of signals. However, the approaches presented in [17],[23],[24] do not apply to the underdetermined case, as they then yield signals which are still mixtures of all source signals. To overcome the latter restriction, we proposed an original concept for the underdetermined case [8],[9]. This method is efficient but requires the sources to have specific stationarity properties. Audio signals, for example, are not well suited to this approach. [19] addresses the separation of several (speech) sources from a single observation, but requires prior knowledge about the sources in order to select basis functions.

We here propose a less specific method, which requires at least two observations however. We start from an approach based on the difference between sources in the time domain, and then extend this principle by exploiting the *time* and *frequency* variations of these signals. A few *time-frequency* (*TF*) BSS methods have previously been reported. The first one [3] provided limited performance despite its complexity. Good separation was obtained even for underdetermined mixtures [4],[27] with another type of methods, which require the sources to be sparse e.g. in the *TF* domain and which resynthesize them. A *TF* method called DUET was also developed for delay and attenuation mixtures [20] and tested with convolutive mixtures, using real-time computation [25]. This method ideally requires the sources to be W-disjoint orthogonal in the *TF* plane (i.e. only one source should occur in each TF window) which is quite restrictive, even if it has been shown that approximate W-disjoint orthogonality is sufficient for most speech signal [2],[26]. In the underdetermined case, this approach performs an inaccurate reconstruction of the sources, by applying an inverse transform to each source restricted to TF areas where it occurs alone, like the somewhat similar SAFIA algorithm [1].

Unlike the above approaches, our TF method applies to sources which almost fully overlap in the TF plane and it avoids artificial source reconstruction. More precisely, it almost only requires each source to occur alone in a tiny set of adjacent TF windows, while several sources may coexist everywhere else in the TF plane. It automatically determines such a single-source area and derives from it "cancelling coefficient" values which allow it to remove the contributions of this source from all observations. Note that recent works also show interest in the detection of single-source areas in the time frequency plane [10]. Our method leads to a complete BSS in a single step in simple configurations, or to a *partial source separation* [8] in more difficult cases, e.g. when $N > P$. This new approach also removes the main restrictions of classical ICA methods [5], as it applies to various dependent and/or Gaussian sources, which may be stationary or not.

This paper is organized as follows. In Section 2, we start from a temporal analysis of the simple configuration involving two mixtures of two sources to then introduce our *TIme Frequency Ratio Of Mixtures (TIFROM)* method. Section 3 shows that this approach applies to dependent and/or Gaussian signals, provided they match the assumptions introduced in Section 2. We then extend our method to the case of N sources and P observations in Section 4. The influence of background noise is studied in Section 5. We then provide several experimental results in Section 6 and draw various conclusions from this investigation in Section 7.

2 Basic case: Two mixtures of two sources

2.1 Preliminary approach: Temporal analysis

We here consider the following linear instantaneous mixture[1] of two real-valued sources:

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \end{cases} \tag{1}$$

where the coefficients a_{ij} of the mixing matrix A are real, constant and different from zero.

BSS may be seen as a method for finding an estimate $\hat{A}^{-1}$ of the inverse of A, so that the output vector $\underline{y} = \hat{A}^{-1}\underline{x}$ is equal to the source vector $\underline{s}$.

[1] The mixtures are assumed to be non-degenerate throughout this paper. In this section, this means $a_{11}/a_{21} \neq a_{12}/a_{22}$.

Due to classical indeterminacies, this separation can only be performed up to a scale factor and a permutation [5]. As an alternative, BSS may also be achieved by means of successive source cancellations by considering the linear combinations of the observations:

$$y(t) = x_1(t) - cx_2(t). \tag{2}$$

Combining (1) and (2) shows that the *cancelling coefficient values* c_1 and c_2 which respectively make it possible to cancel $s_1(t)$ or $s_2(t)$ and therefore to extract $s_2(t)$ or $s_1(t)$ are:

$$c_1 = \frac{a_{11}}{a_{21}}, \quad c_2 = \frac{a_{12}}{a_{22}}. \tag{3}$$

A simple BSS method may then be derived from the following principle: if we can find some locations in the time domain when $x_1(t)$ and $x_2(t)$ contain only the contribution of one source, the above cancelling coefficient values c_i are easy to compute. Considering for example a time t_n such that $s_1(t_n) \neq 0$ and $s_2(t_n) = 0$, (1) yields:

$$\begin{cases} x_1(t_n) = a_{11}s_1(t_n) \\ x_2(t_n) = a_{21}s_1(t_n) \end{cases} \tag{4}$$

By computing the ratio $\frac{x_1(t_n)}{x_2(t_n)} = \frac{a_{11}}{a_{21}}$, we directly obtain the value c_1 of equ. (3) which cancels source $s_1(t)$ using (2). This means that we theoretically only need a source to disappear at time t_n to find the corresponding cancelling coefficient value, i.e. to identify the associated information in the mixing matrix A. Note that if we can determine both c_1 and c_2, then we can easily invert A, achieving BSS up to a scale factor, by considering the inverse matrix:

$$\tilde{A}^{-1} = \begin{bmatrix} 1 & 1 \\ 1/c_1 & 1/c_2 \end{bmatrix}^{-1} \tag{5}$$

which yields

$$\underline{y}(t) = \tilde{A}^{-1}\underline{x}(t) = [a_{11}s_1(t), a_{12}s_2(t)]^T. \tag{6}$$

Unfortunately, this time t_n may hardly be determined in practical situations, since both sources are simultaneously active. Even if the situation when the value of one source crosses the zero frontier is really common for zero-mean signals, the time duration of such events is too short to allow one to detect them.

This approach based on temporal analysis is therefore restricted to the very special case when each source occurs alone in large enough time intervals. Similar BSS methods, which also consider temporally sparse sources, have by the way already been reported, e.g. in [16],[21]. Beyond this preliminary temporal approach, this paper mainly aims at introducing our *TF* extension of the above method, thus yielding much less restrictive sparsity requirements.

2.2 Time-frequency analysis

2.2.1 Definition of the time-frequency tool

During the last fifty years, many powerful *TF* methods have been developed and applied to various fields. One can find most of them with detailed references in [6],[7],[11][15]. Among these methods, we here restrict ourselves to the simple *Short-Time Fourier Transform (STFT)* [7] which avoids interference terms and benefits from powerful FFT algorithms. So, we first multiply each mixed signal $x_i(\tau)$ by a shifted real-valued window function $h(\tau - t)$, centered at time t, which produces the windowed signal $x_i(\tau)h(\tau - t)$. The latter signal depends on two time variables, i.e. the selected time t when the local spectrum of $x_i(\tau)$ is analyzed, and the running time τ. The STFT of x_i is then given by:

$$X_i(t, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x_i(\tau)h(\tau - t)e^{-j\omega\tau} d\tau. \qquad (7)$$

$X_i(t, \omega)$ is the contribution of signal x_i in the short time and frequency windows respectively centered on t and ω.

It should be noted that the STFT is initially defined for deterministic signals and is indeed applied in such a framework in this paper: even if the considered sources and observations are random processes, the STFTs used hereafter only concern a single, and therefore deterministic, realization of these signals (which is requested to satisfy the assumptions defined below).

2.2.2 Exploiting time-frequency information

We will now show that extending the idea presented in Subsection 2.1 to the *TF* plane allows one to identify the cancelling coefficient values corresponding to the sources. We will then propose an automatic method for finding the appropriate *TF* areas. To this end, we request the following assumptions to find all the cancelling coefficient values and to achieve a complete source separation:

Assumption 1

- *The mixing matrix A is such that $a_{ij} \neq 0$, $\forall\, i, j$.*
- *The power of each source is non negligible at least at some times t.*

The first item of this assumption implies that if a source occurs in one observation for a *TF* window (t_j, ω_k), i.e. if its *TF* transform in non-negligible in this window, then it also occurs in all the other observations for this window. The second item is implicitly required in most BSS methods to prevent them from failing: it avoids the situation when some sources have very low dynamics as compared to other ones.

Assumption 2 *For each source s_i, there exist some adjacent* TF *windows (t_j, ω_k) centered on time t_j and angular frequency ω_k where only s_i occurs, i.e. where*[2]*: $S_l(t_j, \omega_k) \ll S_i(t_j, \omega_k)$, $\forall\, l \neq i$.*

Our method is then based on the complex ratio of mixtures independently used in the DUET [20] or SAFIA [1] algorithms:

$$\alpha(t_j, \omega_k) = \frac{X_1(t_j, \omega_k)}{X_2(t_j, \omega_k)}, \tag{8}$$

which is computed for each *TF* window. Taking into account Equ. (1) and (7), this ratio may be written as:

$$\alpha(t_j, \omega_k) = \frac{a_{11} S_1(t_j, \omega_k) + a_{12} S_2(t_j, \omega_k)}{a_{21} S_1(t_j, \omega_k) + a_{22} S_2(t_j, \omega_k)}. \tag{9}$$

Therefore, if one source occurs alone in the *TF* window (t_j, ω_k), then $\alpha(t_j, \omega_k)$ is equal to the cancelling coefficient value, among c_1 and c_2 defined in (3), which makes it possible to cancel this source and thus to extract the other one. This situation when sources only disappear in some areas of the *TF* plane is much more frequent than the case when they disappear at all frequencies during a whole time period. The *TF* BSS method that we thus introduce therefore applies to a much wider class of signals than the preliminary temporal approach that we described in Subsection 2.1. Note that the latter approach is a specific case of this *TF* BSS method, obtained by using a one-sample discrete-time STFT window.

Now the remaining question is: how can we define a method which automatically finds these single-source *TF* areas ? The following assumption is required to this end:

Assumption 3 *When several sources occur in a given set of adjacent* TF *windows they should vary so that $\alpha(t, \omega)$ does not take the same value in all these windows. Especially, i) at least one of the sources must take significantly different* TF *values in these windows*[3] *and ii) the sources should not vary proportionally.*

If only source $s_i(t)$ occurs in several time-adjacent windows[4] (t_j, ω_k), then $\alpha(t_j, \omega_k)$ is constant and equal to c_i over these successive windows,

[2] This is e.g. common for speech or music: some formants of speakers or instruments are located in *TF* areas which do not overlap completely.

[3] Due to statistical fluctuations, each realization of theoretically stationary (e.g. white noise) signals satisfies this condition for short time windows in practice.

[4] The same concept may be applied to frequency-adjacent windows.

whereas it takes different values over these windows if both sources are present and if Assumption 3 is met. To exploit this phenomenon, we compute the sample variance of the complex ratio $\alpha(t,\omega)$ on series Γ_q of M short half-overlapping time windows corresponding to adjacent t_j, applying this approach to each frequency ω_k. Figure 1 shows how we build two such (also half-overlapping) series Γ_q and Γ_{q+1} in the TF plane. We respectively define the sample mean and variance of $\alpha(t,\omega)$ on Γ_q and ω_k by:

$$\overline{\alpha}(\Gamma_q, \omega_k) = \frac{1}{M}\sum_{j=1}^{M} \alpha(t_j, \omega_k) \tag{10}$$

$$var[\alpha](\Gamma_q, \omega_k) = \frac{1}{M}\sum_{j=1}^{M} |\alpha(t_j, \omega_k) - \overline{\alpha}(\Gamma_q, \omega_k)|^2 \tag{11}$$

If e.g. $S_2(t_j, \omega_k) = 0$ for these M windows, then (9) shows that $\alpha(t_j, \omega_k)$ is constant over them, so that its variance $var[\alpha](\Gamma_q, \omega_k)$ is equal to zero. Conversely, under Assumption 3, if both $S_1(t_j, \omega_k)$ and $S_2(t_j, \omega_k)$ are different from zero, then $var[\alpha](\Gamma_q, \omega_k)$ is significantly different from zero.

This shows the importance of Assumption 3, which guarantees that (in the noiseless case considered at this stage) $var[\alpha](\Gamma_q, \omega_k) = 0$ if and only if only one source occurs, which then makes it possible to take advantage of the fact that the ratio $\alpha(t_j, \omega_k)$ is then exactly equal to one of the cancelling coefficients values. On the contrary, if Assumption 3 was not met, obtaining $var[\alpha](\Gamma_q, \omega_k) = 0$ would not guarantee that only one source occurs in the considered area (Γ_q, ω_k). To illustrate this phenomenon, consider the case when the two sources $S_1(t_j, \omega_k)$ and $S_2(t_j, \omega_k)$ would be proportional over (Γ_q, ω_k), i.e. such that $S_2(t_j, \omega_k) = \beta S_1(t_j, \omega_k)$. The ratio $\alpha(t_j, \omega_k)$ would then be constant, as shown by (9), and its variance would be zero, although two sources would occur in this area.

So, under Assumption 3, by searching the lowest value of $var[\alpha](\Gamma_q, \omega_k)$ vs all the available series of windows (Γ_q, ω_k), we directly find a TF domain (Γ_q, ω_k) with only one source. The corresponding value c_i which cancels this source is then estimated by the mean $\overline{\alpha}(\Gamma_q, \omega_k)$. We find the second cancelling coefficient value c_i by searching the next lowest value of $var[\alpha](\Gamma_q, \omega_k)$ vs (Γ_q, ω_k) associated to a significantly different value of $\overline{\alpha}(\Gamma_q, \omega_k)$, using a threshold set to the minimum difference that we request between the two values in (3). We thus obtain estimates of the two cancelling coefficient values defined in (3). The separated signals are then derived from these values by using i) successive source cancellations (2) or ii) the global matrix inversion (6).

As suggested in Section 1, this method for estimating the cancelling coefficient values does not set the same constraints as the approach independently proposed in the DUET algorithm: we just need a source to occur alone in a single tiny set of M adjacent *TF* windows in order to cancel its contributions, whereas DUET requires the sources to be W-disjoint orthogonal (or at least approximate W-disjoint orthogonal) which is a much stronger assumption. On the contrary, we use *adjacent* single-source *TF* windows, but this is not a major constraint since we need very few of them, as illustrated in Section 6. Also, since our approach first focuses on *TF* areas where only one source occurs and then computes the mean $\overline{\alpha}(\Gamma_q, \omega_k)$ in such supposedly single-source areas, $\overline{\alpha}(\Gamma_q, \omega_k)$ is a suitable estimator of c_i. This should be contrasted with DUET, which operates in the complete *TF* plane, so that plain histogram-based methods yield limitations and more sophisticated versions should be preferred, as shown in [2] (an alternative approach may also be found in [22]).

3 Dependent and/or Gaussian signals

As stated above, our method applies to various dependent and/or Gaussian sources, unlike classical ICA-based BSS approaches. The latter methods are statistical approaches, which require the sources to be statistically independent and which consist in forcing the output signals to become independent, so that they get equal to the sources. Our approach is totally different, as it is based on the sample statistics of a single signal realization, which allows it to determine some domains in the *TF* plane where a single source occurs. It therefore almost only requires such domains to exist and applies to (realizations of) various dependent and/or Gaussian sources which meet this condition. This is illustrated hereafter by means of two typical examples. It should be noted that other *TF* BSS methods yield similar properties, as suggested e.g. in [27].

3.1 Dependent signals

Consider the two source signals $s_1(t) = u(t) + v(t)$ and $s_2(t) = v(t) + w(t)$, where $u(t)$, $v(t)$ and $w(t)$ are three stationary independent zero-mean signals and where:

a) $v(t)$ only has components in the frequency band $[f_1, f_2]$, and $u(t)$ and/or $w(t)$ also have components at the frequencies where $v(t)$ occurs,

b) $u(t)$ only has components in the frequency band $[0, f_2]$,

c) $w(t)$ only has components above f_1.

The cross-correlation of $s_1(t)$ and $s_2(t)$ is non-zero, due to their common component $v(t)$. These two source signals are therefore dependent. However, it may be checked easily that they match all the assumptions required in our method. We can then separate (realizations of) these signals with our approach, despite their dependence, thanks to the differences in their *TF* representations. This is an important case as traditional BSS methods, like kurtosis maximization, cannot separate this kind of signals.

3.2 Gaussian signals

We now consider two zero-mean Gaussian *i.i.d* signals $v_1(t)$ and $v_2(t)$ that we transfer through two stop-band filters $h_1(t)$ and $h_2(t)$ which have disjoint stop-bands. The resulting source signals $s_1(t) = h_1(t) * v_1(t)$ and $s_2(t) = h_2(t) * v_2(t)$ are also Gaussian but (their realizations) can be separated with our method in the same way as above.

The approaches of Subsections 3.1 and 3.2 may also be combined so as to define sources which are both dependent and Gaussian but which may be separated by our method.

4 Extension to P mixtures of N sources

4.1 $N > 2$ sources, $P = 2$ observations

4.1.1 Definition of the time-frequency method

The observations here become:

$$\begin{cases} x_1(t) = \sum_{m=1}^{N} a_{1m} s_m(t) \\ x_2(t) = \sum_{m=1}^{N} a_{2m} s_m(t) \end{cases} \tag{12}$$

The complex ratio $\alpha(t, \omega)$ of (8) then reads:

$$\alpha(t, \omega) = \frac{\sum_{m=1}^{N} a_{1m} S_m(t, \omega)}{\sum_{m=1}^{N} a_{2m} S_m(t, \omega)}. \tag{13}$$

Under the same assumptions as above, consider a *TF* window (t_j, ω_k) where only source s_i occurs. The complex ratio in (13) then becomes:

$$\alpha(t_j, \omega_k) = \frac{a_{1i}}{a_{2i}}. \tag{14}$$

This is exactly the value c_i of the coefficient c of (2) required for cancelling the contribution of source $s_i(t)$ from the observations by using (2). The BSS method defined in Subsection 2.2 is therefore straightforwardly extended to the current case, but then leads to a *partial* separation, i.e. to the cancellation of only one of the existing sources in each output signal. This is of high practical interest in signal enhancement applications anyway, as this method gives an efficient solution for removing the contribution of an undesirable source.

This method is also useful in karaoke-like applications. In usual audio recordings, all instruments are recorded one by one and then artificially mixed in studios using linear instantaneous mixing devices. Using the stereo observation of such recorded songs, we are able under Assumptions 1 to 3 to cancel the contribution of any singer or instrument[5].

[5]Stereo filtering, like stereo reverberation, is sometimes added in such recordings. The mixture then becomes convolutive. However, when no delays are added, a major part of the

4.1.2 Estimation of the number of sources

The above method may also be used to derive an estimate $\hat{N}$ of the number of sources contained in the observations: we set $\hat{N}$ to the number of significantly different values of $\overline{\alpha}(\Gamma_q, \omega_k)$ obtained in areas (Γ_q, ω_k) where the variance of $\alpha(t, \omega)$ is low. Under Assumption 2, $\hat{N}$ is equal to N.

For the sake of generality, we also consider the case when Assumption 2 is not satisfied in the remainder of this paper. Some sources are then *hidden* by the other ones in the *TF* plane, i.e. there is no *TF* window where they occur alone in the observations and thus no associated value of $\overline{\alpha}(\Gamma_q, \omega_k)$ in low-variance areas. By computing $\hat{N}$ as explained above, we then only get the number Q of *visible*, i.e. not hidden, sources.

4.2 General case: N sources, P observations

4.2.1 Coherence of the time-frequency maps

As already suggested above, due to Assumption 1, the areas (Γ_q, ω_k) where a given source appears alone in observations are the same for all observations. We call this phenomenon the "*coherence of the* TF *maps*" and illustrate it in Figure 2, where the time-frequency windows corresponding to only one source are the same in both observations. Thanks to this *coherence*, single-source areas may be detected for most mixing matrices by analyzing the variance of the ratio $\alpha(t, \omega) = X_i(t, \omega)/X_j(t, \omega)$ associated to only one arbitrary pair of observations: here again, this variance is low (only) in single-source areas under Assumption 3.

An exception to this principle appears when $P > 2$ however: in areas (Γ_q, ω_k) where several sources are active, $\alpha(t, \omega)$ may have a low variance for some pairs of observations, because the corresponding subset of mixing coefficients results in proportional observations in these areas. However, for a given area this phenonemon may not occur for all pairs of observations, otherwise the mixing matrix would be degenerate. This case which only concerns very specific mixing matrices is therefore handled by performing variance analyses for all pairs of observations $(x_1(t), x_j(t))$. We skip this specific case hereafter and therefore only consider a single variance analysis, thus introducing a fast BSS method.

4.2.2 Fast BSS method for $N = P$

We now describe i) a global algorithm, that simultaneously provides several

interfering sources corresponds to their instantaneous contributions, and our experimental tests show that our method succeeds in cancelling these major contributions.

output signals and ii) another approach, which creates each output signal by successively cancelling visible sources. The latter approach is of special interest when only specific sources should be removed or when the number Q of visible sources is equal to 1.

i) Global matrix inversion:

We first perform a single variance analysis with two observations, as explained in Subsection 4.2.1, and we again denote Q (with $Q \leq N$) the number of visible sources thus obtained. If this yields $Q \geq 2$, we then apply the procedure described here, whereas if $Q = 1$ the successive source cancellation approach defined further in this section should preferably be used as suggested above (and if $Q = 0$ no separation may be performed with our approaches).

Denoting $s_1, \ldots s_Q$ the visible sources for ease of presentation, we consider Q TF areas where each of these visible source occurs without the others in all the observations. We then adapt the approach of Subsection 4.1 to each pair of observations $(x_1(t), x_j(t))$. We thus compute the mean of the ratio $X_1(t,\omega)/X_j(t,\omega)$ in one area where only s_i exists, with $1 \leq i \leq Q$. This yields the value $c_{ij} = a_{1i}/a_{ji}$ of the cancelling coefficient for removing this source s_i. This overall set of coefficients provides an estimate of the part of the mixing matrix corresponding to the visible sources (again up to a scale factor), i.e:

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & \cdots & 1 \\ 1/c_{12} & \cdots & 1/c_{Q2} \\ \vdots & & \vdots \\ 1/c_{1P} & \cdots & 1/c_{QP} \end{bmatrix}. \tag{15}$$

If $Q = N$, we use the inverse of this square matrix to achieve a global inversion up to a scale factor, i.e: $\underline{y}(t) = \tilde{A}^{-1}\underline{x}(t) = [a_{11}s_1(t), \ldots, a_{1N}s_N(t)]^T$. This efficient method therefore leads to a complete BSS in one step when no sources are initially hidden, i.e. when all the c_{ij} may be derived directly from the observations.

On the other hand, when $Q < N$ this non-square matrix $\tilde{\mathbf{A}}$ misses the coefficients c_{ij} associated to the hidden sources and then cannot be used directly to separate all sources by means of its inverse. However, we can derive $Q \times Q$ square sub-matrices from $\tilde{\mathbf{A}}$ by keeping its first line and $Q-1$ arbitrary other lines. If we now multiply the inverse of any such sub-matrix by the vector containing the mixed signals $x_i(t)$ which correspond to the Q lines kept from $\tilde{\mathbf{A}}$, we get a vector of Q "recombined signals". Each such signal with index j only consists of contributions from the visible source with the same index j and from all $N-Q$ initially hidden sources. At this preliminary stage, we thus perform a partial BSS [8], restricted to the visible sources.

We can repeat this procedure by keeping different sets of lines from

$\tilde{\mathbf{A}}$ and deriving corresponding sub-matrices and recombined signal vectors. For any given index j, all recombined signals with index j thus obtained are mixtures of source j and all initially hidden sources, as explained above. Moreover, $N - Q + 1$ linear independent recombined signals may thus be created for any given index j, depending on which lines are kept from $\tilde{\mathbf{A}}$.

We thus reduced the initial BSS problem to Q independent BSS sub-problems, each associated to a specific index j, and each only involving $N - Q + 1$ linearly independent mixtures of the same $N - Q + 1$ sources. Moreover, additional sources may now be visible in these new mixtures, because they are no more hidden by other sources, which have been cancelled when deriving these new mixtures. Each such sub-problem is then addressed independently, by recursively calling again the approaches proposed in this paper for the subset of sources considered in this sub-problem, in order to further separate visible sources. A detailed pseudo-code corresponding to this complete algorithm is provided in the appendix of this paper.

Note that i) if the sources which form the subset considered in one such sub-problem are all visible, they are totally separated by our method and ii) if all N sources eventually become visible in sub-problems thus introduced, we achieve a complete separation of the overall set of N sources, although some of these sources may be initially hidden.

ii) Successive source cancellations
Instead of (2), the basic step of the approach described here creates "new observations" $x_1(t) - c_{ij}x_j(t)$, where c_{ij} are again the coefficient values which cancel the contributions of the selected source s_i in these new observations. This basic step is recursively applied to these new observations which replace the initial observations, for various selected sources. If $Q = N$, applying this step $N - 1$ times yields a signal in which $N - 1$ selected sources s_i are cancelled and the N^{th} source is therefore extracted (this is then repeated so as to extract each source). If $Q < N$, performing this step Q times first cancels all initially visible sources. Other sources may thus become visible in the resulting new observations (another variance analysis is performed to detect it). This procedure is then applied again recursively so as to cancel other sources.

4.2.3 Fast BSS method for $N > P$

The above-defined two methods can easily be extended to the underdetermined case: the previous algorithms then just end when there remains only one new observation or when as many sources as possible have been cancelled (note that the basic version of this case has been detailed in Subsection 4.1.1). Whatever algorithm we choose, we thus succeed in achieving *partial source separation* [8]. On the contrary, classical methods fail for underdetermined mixtures, i.e. their outputs are still mixtures of all sources.

5 Influence of background noise

In the previous sections, we considered noise-free mixtures. But in real cases, the observations $x_i(t)$ contain additive wide-band background noise signals $n_i(t)$. The basic configuration then becomes:

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + n_1(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + n_2(t) \end{cases} \tag{16}$$

The complex ratio (8) then reads:

$$\alpha(t,\omega) = \frac{a_{11}S_1(t,\omega) + a_{12}S_2(t,\omega) + N_1(t,\omega)}{a_{21}S_1(t,\omega) + a_{22}S_2(t,\omega) + N_2(t,\omega)}, \tag{17}$$

with $N_i(t,\omega) \neq 0$ for all considered windows (t_j, ω_k). Thanks to this background noise, the special case when both $S_1(t,\omega)$ and $S_2(t,\omega)$ are locally equal to zero does not cause any singularity, i.e. it does not result in a $0/0$ indeterminacy in the ratio $\alpha(t,\omega)$ over a complete domain (Γ_q, ω_k). Moreover, due to independent statistical fluctuations of $N_1(t,\omega)$ and $N_2(t,\omega)$, the value $\frac{N_1(t,\omega)}{N_2(t,\omega)}$ is not constant in *TF* windows (t,ω). This implies that the variance of $\alpha(t,\omega)$ over areas where only N_1 and N_2 occur is significant, so that our algorithm does not mistakenly focus on such areas when looking for single-source areas.

Assuming that the mixing matrix A does not have very low entries, a reasonable condition to estimate accurately the cancelling coefficient values for source s_i is:

Assumption 4 *There exists a* TF *area* (Γ_q, ω_k) *where only source* s_i *occurs and,* $\forall\ (t_j, \omega_k) \in (\Gamma_q, \omega_k)$, $|N_1(t_j, \omega_k)| \ll |S_i(t_j, \omega_k)|$ *and* $|N_2(t_j, \omega_k)| \ll |S_i(t_j, \omega_k)|$.

It should be noted that this is not a stringent condition, since it only requires the noise level to be low with respect to the level of the considered source in *one* tiny TF area.

6 Experimental results

6.1 Configuration with two mixtures of two sources

The first test was performed with source signals from two speakers, sampled at 22 kHz, and with the mixing matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0.9 \\ 0.8 & 1 \end{bmatrix}. \tag{18}$$

The two theoretical cancelling coefficient values are, according to Equ. (3): $c_1 = 1.25$ and $c_2 = 0.9$. Combining the source powers and the above mixing matrix, we found respective Signal-to-Interference Ratios (SIRs) of 2.38 dB and -2.38 dB for sources s_1 and s_2 in the first observation and -0.47 dB and 0.47 dB in the second observation. Figures 7 and 8 show the SIRs obtained for s_1 and s_2 in the output signals y_1 and y_2 depending on the size N_{STFT} of the STFT windows[6] and the number M of windows used to estimate the variance of $\alpha(t, \omega)$.

This shows that these SIRs are very good, i.e. almost always higher than 45 dB, whatever N_{STFT} ranging from 32 to 512 and M from 4 to 12.

As an example, we detail the case when $N_{STFT} = 128$ and $M = 8$. In order to cancel a source, our method then only requests it to occur alone in a single frequency window during 8 (half-overlapping) time windows. This represents both a very short time period, i.e. 26.2 ms, and a tiny part of the *TF* plane, i.e. 0.01 % of this plane for the considered 2.7 s speech signals. This illustrates that the proposed approach applies to almost fully overlapping sources, whereas various others methods cited above need each source to occur alone in a very much larger number of *TF* windows to achieve an acceptable degree of signal separation. The two coefficient values provided by our method for extracting s_2 and s_1 are then $c_1 = 1.2516$ and $c_2 = 0.9004$, and the associated output SIRs are 49 dB and 58 dB. On listening to output signals, the difference between the original and estimated sources is not perceptible. Figures 3 and 4 show that, although both sources are active everywhere in the time domain (which makes the temporal algorithm of Section 2.1 fail), their *TF* representations are somewhat different, which confirms the relevance of our approach for such signals and is in agreement with [26]. For better legibility, we plotted the inverse of the variance of $\alpha(t, \omega)$ in Figure 5. This representation enhances the domains where the variance is low and in the same range. This shows in which *TF* domains the cancelling coefficient values are obtained.

We then compared the performance of our approach and of several classical methods, using the Matlab toolbox *ICALAB 2.2* available at [18]. One can find links towards references concerning the considered algorithms in the *Help* included in this package. Figure 9 shows the output SIRs obtained for the above-defined source signals and mixing matrix. Case A shows results derived by applying these methods to the entire signals (2.7 s). Case B and Case C correspond to results respectively obtained with two short windows (45 ms). These windows were manually selected because the corresponding source signal parts exhibit good stationarity and independence (the magnitudes of their zero-lag cross-correlation coefficients $|E[s_1 s_2]|/\sqrt{E[s_1^2]E[s_2^2]}$ are equal to 0.0153 in window B and 0.0155 in window C). These signal parts are thus expected to be better suited to classical algorithms than the

[6]No zero-padding is used when computing FFTs.

configuration considered in Case A.

Several of these classical methods yield good output SIRs, i.e. up to about 40 dB. However, the SIRs obtained in configurations A, B and C are often quite different for a given method, and some of them are very low. This means that the user should choose very carefully a correct signal part to obtain good separation. On the contrary, our method first automatically determines the best area associated to the above-defined criterion and then computes the cancelling coefficient values. It thus yields significantly better SIRs than classical approaches, i.e. about 50 to 70 dB for most considered sizes and numbers of STFT windows.

We performed additional tests (not detailed in this paper) using highly correlated sources which match the assumptions required by our method, for example guitar and voice playing the same score, i.e. D chord (D F# A) for the guitar and D for the voice. Our experimental results show that such sources can easily be separated by our method, with comparable SIRs even if their zero-lag cross-correlation coefficient is around 0.9.

6.2 Configuration with two mixtures of three sources

We recorded a stereo song with two guitars s_1 and s_2 playing nearly the same instrumental part and continuous voice s_3. Guitar s_1 is hidden by s_2 which contains more high-frequency harmonics. We here aim at showing the ability of the proposed approach to cancel the voice from the mixtures, although the guitars are continuously playing. All these sources were recorded one by one on a 4-track tape recorder with a SNR around 60 dB and sampled at 44100 kHz. Like in most commercial songs, we chose to put the voice in the middle of the stereo mix, whereas guitars s_1 and s_2 are respectively situated more on the left and right sides using the following mixing matrix:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.4 & 0.8 \\ 0.3 & 0.8 & 0.8 \end{bmatrix}. \tag{19}$$

The theoretical cancelling coefficient values, defined by (14), are equal to $c_1 = 2.33$ for the first guitar, $c_2 = 0.5$ for the second guitar and $c_3 = 1$ for the voice.

As this configuration involving 3 sources for 2 observations is underdetermined, we obtain *partial source separation* on the 2 outputs. Let us consider the output signals given by $y(t) = x_1(t) - c_i x_2(t)$ with c_i corresponding to the cancelling coefficient value associated with source s_i. The contribution of each source s_j in y is $(a_{1j} - c_i a_{2j})s_j$ whereas its contribution in observation x_1 is $a_{1j}s_j$. We therefore define the (opposite) attenuation *Att* of the source s_j in the obtained output as compared to its contribution in reference x_1 by:

$$Att = 10log_{10}\frac{E\{(a_{1j}s_j)^2\}}{E\{((a_{1j} - c_i a_{2j})s_j)^2\}} = 20log_{10}\left|\frac{a_{1j}}{a_{1j} - c_i a_{2j}}\right| \tag{20}$$

This parameter takes a very high value when the considered source is almost cancelled in the considered output. The results in Figure 10 therefore show that we are able to cancel the voice s_3 on one output, then giving a nearly perfect karaoke playback. Similar results are obtained on the other output for the guitar s_2. Guitar s_1 is hidden by s_2, so that it cannot be cancelled with these two observations. It should be noted however that with 3 observations its contribution may be cancelled by using the procedure defined in Subsection 4.2.2. We can see on the inverse variance graph in Fig. 6 that most low-variance points are in the frequency band between 7 and 15 kHz, where only voice is present. No low-variance points exist for frequencies higher than 15 kHz, i.e. where only noise occurs.
We here showed that *TF* information allows one to achieve nearly perfect source cancellation.

7 Conclusion

In this paper, we proposed a simple and efficient method for solving the linear instantaneous BSS problem with N sources and P observations. This approach is based on the TIme-Frequency version of Ratios Of Mixtures of source signals, and is therefore called "TIFROM". It mainly relies on the assumption that a source is "visible", i.e. that it occurs alone (as opposed to the other sources) in at least one tiny area in the *TF* plane. It automatically determines such an area and then derives coefficients which e.g. allow one to cancel the contributions of this source from the observed signals. This makes it possible to separate all sources in various situations, esp. when $N = P$ and all sources are visible. This method is still of interest in other cases, esp. for underdetermined mixtures: it then separates part of the sources, whereas the methods proposed in [8],[9] set more restrictive constraints on the sources to achieve such partial BSS and classical methods then yield still completely mixed signals.

Some previously reported *TF* BSS methods also apply to underdetermined mixtures, but they set much more restrictive constraints on the *TF* distributions of the sources, i.e. they require them to be (approximately) W-disjoint orthogonal or sparse, while we allow them to almost fully overlap (and therefore to be multicomponent). Our approach also provides other major advantages over classical ICA-based methods, i.e. it applies to stationary and non-stationary signals and to various dependent and/or Gaussian sources, provided their *TF* representations satisfy the above-defined assumptions.

Some restrictions may appear when applying this method to a large number of sources, as their chance to be visible then decreases. However, experimental tests performed with commercial CDs showed that we are e.g. able to cancel some instruments from only 2 observations of more than 5 sources. Our future investigations will especially concern this case involving many sources and the extension of the proposed approach to convolutive mixtures, which is a significantly more complex problem.

Acknowledgments

The authors would like to thank Paul White, from the SPCG group of the *Institute of Sound and Vibration Research (ISVR)* of the University of Southampton, for his help in the first stages of this work, and the reviewers for their very helpful comments.

A Pseudo-code of the proposed TF BSS algorithm

We here provide the pseudo-code of the algorithm based on global matrix inversion that we introduced in Subsection 4.2.2 for the case when $N = P$ and when some sources may be initially hidden.

- Step 1: compute the STFT $X_i(t_j, \omega_k)$ of each source $x_i(t)$, $i = 1...N$.
- Step 2: compute the ratio $\alpha(t_j, \omega_k) = X_1(t_j, \omega_k)/X_i(t_j, \omega_k)$ for a single couple (X_1, X_i), with $i \neq 1$, for example $i = 2$.
- Step 3: select the number M of successive time windows included in the series of windows Γ_q and then compute $var[\alpha](\Gamma_q, \omega_k)$ for all the available series of windows (Γ_q, ω_k).
- Step 4: sort the values $var[\alpha](\Gamma_q, \omega_k)$ in ascending order. Only use the TF areas (Γ_q, ω_k) associated to the first values in this ordered list, i.e. the areas such that $var[\alpha](\Gamma_q, \omega_k)$ is below a user-defined threshold. The first and subsequent areas in this beginning of the list are successively used as follows. Each considered area is kept, as the j-th area, only if all distances between i) the column of values $c_{i,j}$, with $i = 1, ..., N$ (defined in Subsection 4.2.2) corresponding to this TF area and ii) the columns of values corresponding to the previously kept TF areas in this procedure are higher than a user-defined threshold. This yields Q columns of values, with $Q \leq N$, where Q is the number of initially visible sources. These columns are gathered in the matrix defined as

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & \cdots & 1 \\ 1/c_{12} & \cdots & 1/c_{Q2} \\ \vdots & & \vdots \\ 1/c_{1N} & \cdots & 1/c_{QN} \end{bmatrix}. \tag{21}$$

- Step 5:

– if $Q = N$, use the inverse of this square matrix $\tilde{\mathbf{A}}$ to achieve a global inversion up to a scale factor, i.e: $\underline{y}(t) = \tilde{A}^{-1}\underline{x}(t) = [a_{11}s_1(t), \ldots, a_{1N}s_N(t)]^T$.
– if $Q < N$ then derive the $N - Q + 1$ square sub-matrices with size $Q \times Q$:

$$\tilde{\mathbf{A}}_{\mathbf{p}} = \begin{bmatrix} 1 & \cdots & 1 \\ 1/c_{12} & \cdots & 1/c_{Q2} \\ \vdots & & \vdots \\ 1/c_{1,Q-1} & \cdots & 1/c_{Q,Q-1} \\ 1/c_{1,p} & \cdots & 1/c_{Q,p} \end{bmatrix} \tag{22}$$

with $p = Q, ..., N$ and compute the "recombined signals":

$$\underline{y}^{(p)}(t) = \tilde{A}_p^{-1}[x_1(t), \ldots x_{Q-1}(t), x_p(t)]^T. \tag{23}$$

Then create Q new independent subsystems $\underline{x}^{(i)}(t)$, $i = 1, ..., Q$. Each of them consists of $N - Q + 1$ linearly independent mixtures of the same $N - Q + 1$ sources :
subsystem 1: $\underline{x}^{(1)}(t) = [y_1^{(Q)}(t), y_1^{(Q+1)}, ..., y_1^{(N)}]^T$, where $y_1^{(p)}$ represents the first element of vector $\underline{y}^{(p)}(t)$. All the signals in the vector $\underline{x}^{(1)}(t)$ only contain sources $s_1, s_{Q+1}, ..., s_N$ (still denoting $s_1, \ldots s_Q$ the Q initially visible sources).
subsystem 2: $\underline{x}^{(2)}(t) = [y_2^{(Q)}(t), y_2^{(Q+1)}, ..., y_2^{(N)}]^T$, where all the signals in this vector only contain sources $s_2, s_{Q+1}, ..., s_N$.
...
subsystem Q: $\underline{x}^{(Q)}(t) = [y_Q^{(Q)}(t), y_Q^{(Q+1)}, ..., y_Q^{(N)}]^T$, where all the signals in this vector only contain sources $s_Q, s_{Q+1}, ..., s_N$.
Then recursively apply the whole process (from Step 1) independently to each of the Q subsystems thus introduced ... and so on, until complete (or maximum) source separation.

References

[1] M. Aoki and al. Sound source segregation based on estimating incident angle of each frequency component of input signals acquired by multiple microphones. In *Acoust. Sci. and Tech, 22, 2*, 2001.

[2] R. Balan and J. Rosca. Statistical properties of STFT ratios for two channel systems and applications to blind source separation. In *Proc. of ICA*, pages 429–434, Finland, 2000.

[3] A. Belouchrani and M. G. Amin. Blind source separation based on time-frequency signal representations. *IEEE Transactions on Signal Processing*, 46(11):2888–2897, November 1998.

[4] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representation. *Signal Processing*, 81(11):2353 – 2362, 2001.

[5] J. F. Cardoso. Blind signal separation: statistical principles. In *Proceedings of the IEEE*, volume 86, number 10, pages 2009–2025, October 1998.

[6] L. Cohen. Time-frequency distributions - a review. In *Proceedings of the IEEE*, volume 77, No. 7, pages 941–979, July 1989.

[7] L. Cohen. *Time-frequency analysis.* Prentice hall PTR, Englewood Cliffs, New Jersey, 1995.

[8] Y. Deville and M. Benali. Differential source separation: concept and application to a criterion based on differential normalized kurtosis. In *Proceedings of EUSIPCO*, Tampere, Finland, September, 4-8, 2000.

[9] Y. Deville and S. Savoldelli. A second-order differential approach for underdetermined convolutive source separation. In *Proceedings of ICASSP 2001*, Salt Lake City, USA, May 7-11 2001.

[10] C. Févotte and C. Doncarli. Two contributions to blind source separation using time-frequency distributions. In *IEEE Signal Processing Letters, Vol 11, No. 3*, March 2004.

[11] J. K. Hammond and P. R. White. The analysis of non-stationary signals using time-frequency methods. *Journal of sound and vibrations*, pages 419–447, 1996.

[12] J. Herault and C. Jutten. Space or time adaptive signal processing by neural network models. In *International Conference on Neural Networks for computing*, Snowbird, USA, 1986.

[13] J. Herault and C. Jutten. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, July 1991.

[14] J. Herault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite pour une architecture de calcul neuromimétique en apprentissage non supervisé. In *GRETSI*, pages 1017–1022, Nice, France, May 1985.

[15] F. Hlawatsch and G. F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9:21–67, April 1992.

[16] M. V. Hulle. Clustering approach to square and non-square blind source separation. In *NNSP*, pages 315–323, Madison, Wisconsi, August 23-25 1999.

[17] A. Hyvarinen. Blind source separation by nonstationarity of variance: a cumulant-based approach. *IEEE Trans. on Neural Networks*, 12(6):1471–1474, November 2001.

[18] ICALAB. *http://www.bsp.brain.riken.jp/ICALAB/*.

[19] G.J. Jang, T.W. Lee, and Y.H. Oh. Blind separation of single channel mixture using ica basis functions. In *ICA*, San Diego, CA, December 2001.

[20] A. Jourjine, S. Rickard, and . Yilmaz. Blind separation of disjoint orthogonal signals: demixing n sources from 2 mixtures. In *Proceedings of ICASSP 2000*, volume 6, pages 2986–2988, Turkey, June, 6-9, 2000.

[21] T. W. Lee, M. Lewicki, M. Girolami, and T. J. Sejnowski. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Proc. Letters*, 6(4):87–90, April 1999.

[22] L. T. Nguyen, A. Belouchrani, K. Abed-Meraim, and B. Boashash. Separating more sources than sensors using time-frequency distributions. In *ISSPA*, pages 583–586, Malaysia, August 2001.

[23] L. Parra and C. Spence. Convolutive blind separation of non-stationary sources. *IEEE Transaction on Speech and Audio Processing*, 8(3):320–327, May 2000.

[24] D. T. Pham and J. F. Cardoso. Blind separation of instantaneous mixtures of non-stationary sources. *IEEE Trans. on Signal Processing*, October 2000.

[25] S. Rickard, R. Balan, and J. Rosca. Real-time time-frequency based blind source separation. In *Proceedings of ICA 2001*, San Diego, CA, December, 9-13 2001.

[26] S. Rickard and O. Yilmaz. On the approximate w-disjoint orthogonality of speech. In *ICASSP*, Orlando, Florida, May, 13-17 2002.

[27] M. Zibulevsky and B. A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary, in *Independent component analysis: Principles and practice.* Robert S. J. and Everson R. M. editors, Cambridge University Press, 2000.

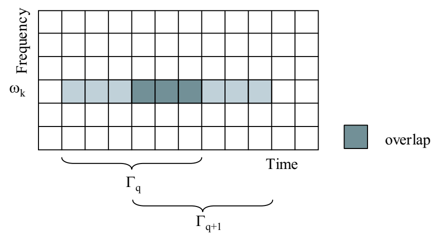

Figure 1: Definition of two series of time-frequency windows.

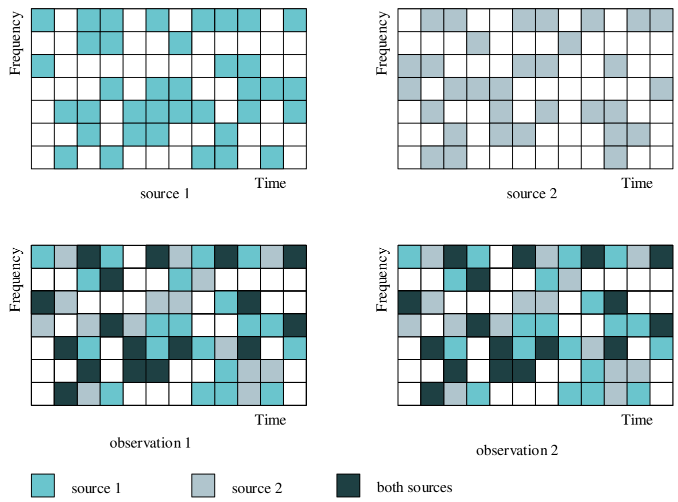

Figure 2: Coherence of the time-frequency maps.

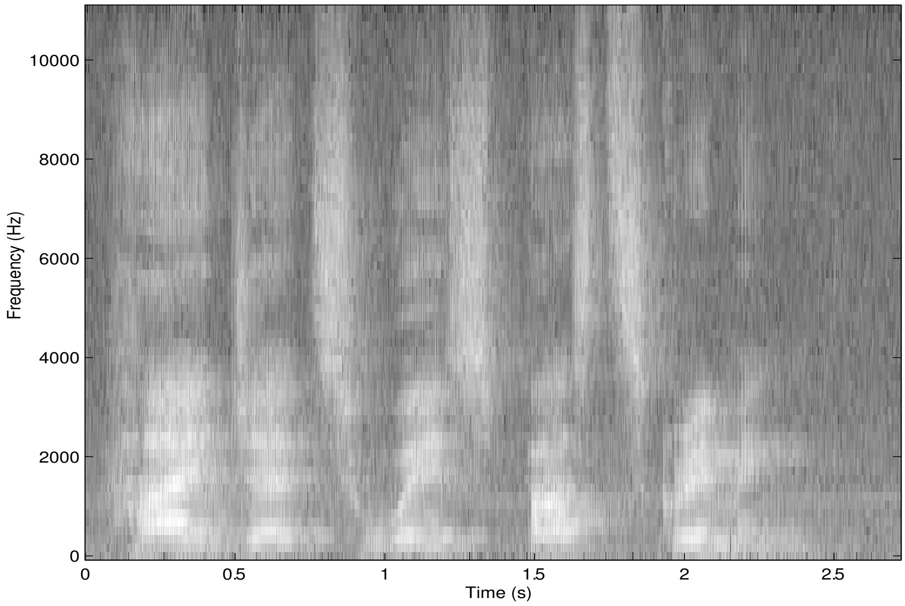

Figure 3: Time-Frequency map of source s_1.

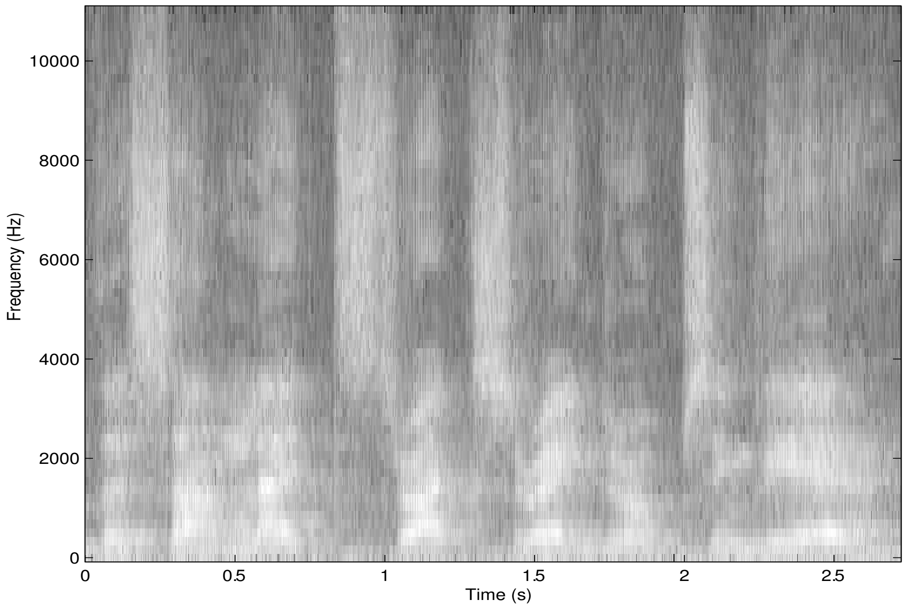

Figure 4: Time-Frequency map of source s_2.

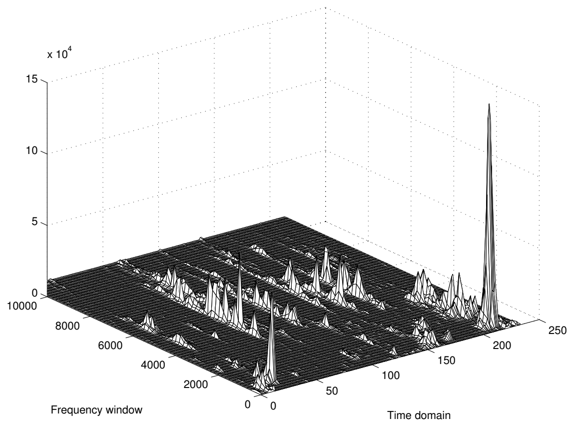

Figure 5: Time-Frequency representation of $\frac{1}{var[\alpha(\Gamma,\omega)]}$. Axes units: Time window indices, corresponding to [0 s, 2.7 s]. Frequency window indices, corresponding to [0 Hz, 11050 Hz].

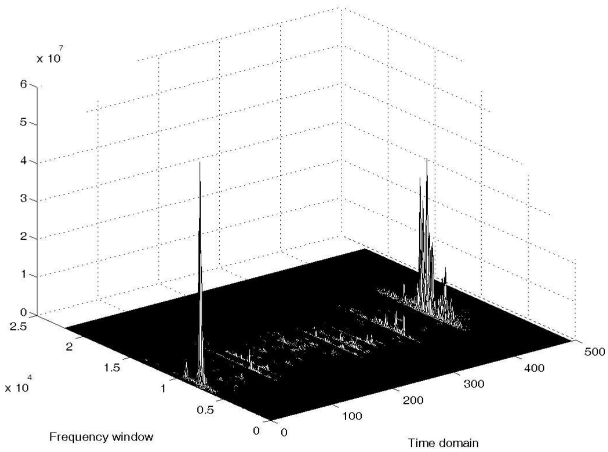

Figure 6: Time-Frequency representation of $\frac{1}{var[\alpha(\Gamma,\omega)]}$. Axes units: Time window indices, corresponding to [0 s, 4.3 s]. Frequency window indices, corresponding to [0 Hz, 22.05 kHz].

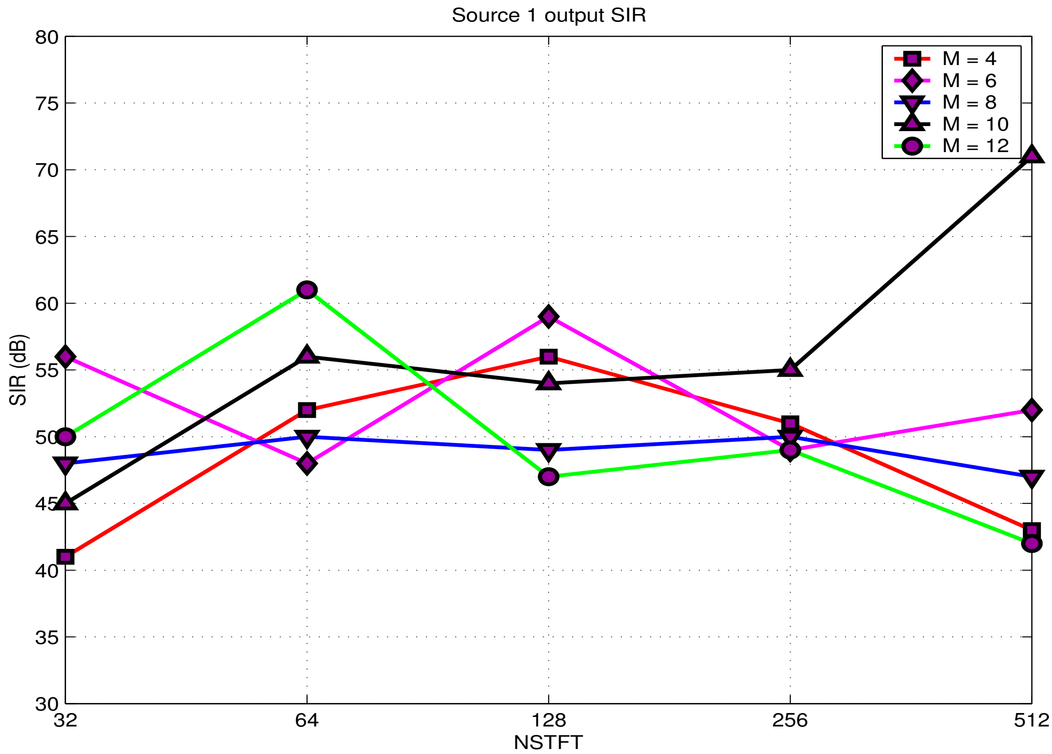

Figure 7: Output SIRs (in dB) for source s_1 vs number N_{STFT} of samples in STFT windows and number M of windows in domains Γ_q.

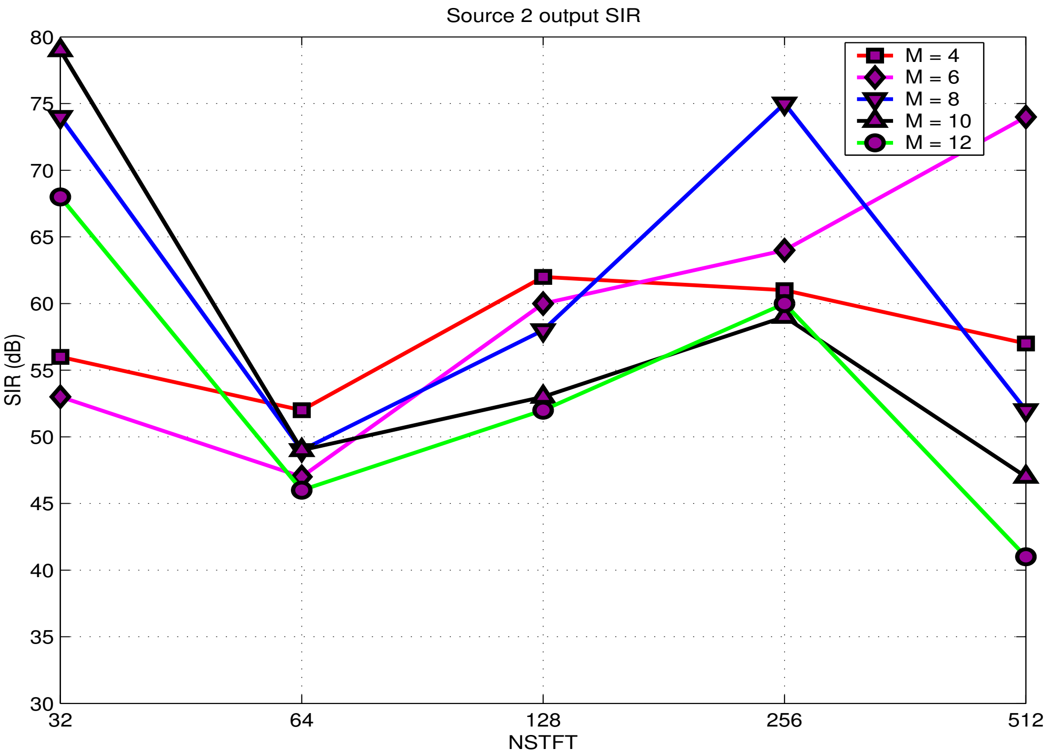

Figure 8: Output SIRs (in dB) for source s_2 vs number N_{STFT} of samples in STFT windows and number M of windows in domains Γ_q.

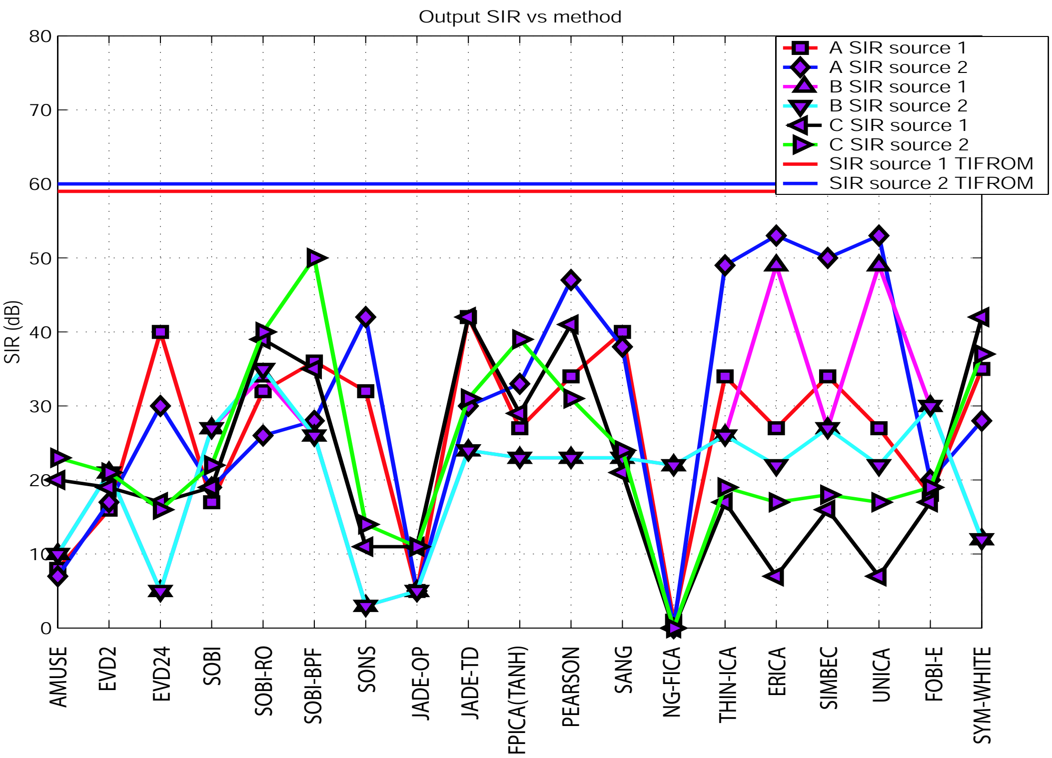

Figure 9: Output SIRs (in dB) for sources s_1 and s_2 obtained with classical algorithms and with our TIFROM method.

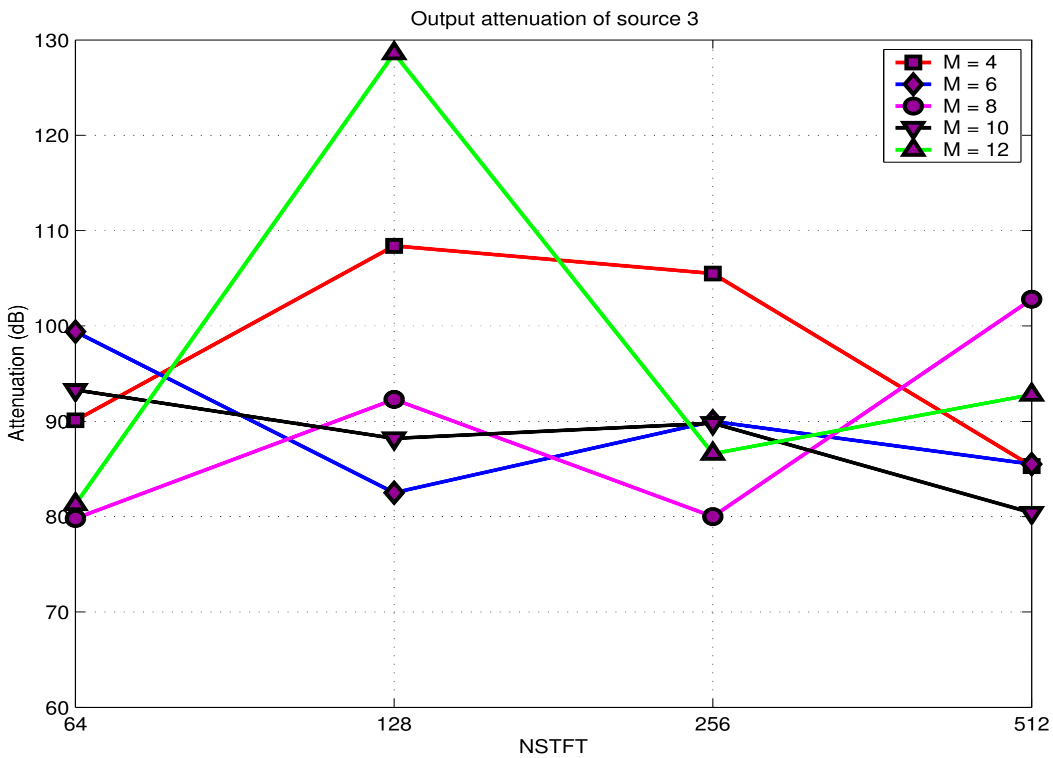

Figure 10: Attenuation (dB) of source s_3 in output y_1 vs number N_{STFT} of samples in STFT windows and number M of windows in domains Γ_q.