BLIND SEPARATION OF DEPENDENT SOURCES USING THE "TIME-FREQUENCY RATIO OF MIXTURES" APPROACH

Frédéric Abrard, Yannick Deville

LAMI, Université Paul Sabatier,118 route de Narbonne, 31062 Toulouse Cedex, FRANCE

abrard@i2e.fr - ydeville@cict.fr

ABSTRACT

In this paper, we first briefly recall the principles of the "TIme-Frequency Ratio Of Mixtures" (TIFROM) approach that we recently proposed. We then show that, unlike Independent Component Analysis (ICA) methods, our approach can separate dependent signals, provided there exist some areas in the time-frequency plane where only one source occurs. We achieve this attractive property because, whereas ICA methods aim at creating independent output signals, we use another concept, i.e. we directly estimate the mixing matrix by using the time-frequency information contained in the observations. Detailed results concerning mixtures of voice and music signals are presented and show that this approach yields very good performance for signals which cannot be separated with traditional ICA methods.

1. INTRODUCTION

Blind source separation (BSS) consists in estimating a set of N unknown sources from P observations resulting from the mixture of these sources through unknown propagation channels.

Denoting the mixing operator by $\mathcal{A}$, 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.

Traditional Independent Component Analysis (ICA) approaches basically aim at separating the sources by combining the observations so that the output signals are independent [1] which means that the fundamental assumption of ICA techniques is that the sources must be independent. Moreover, most of these approaches can only separate stationary non-Gaussian signals. Because of these limitations, poor performance is often obtained when dealing with real sources, like audio signals, which do not match those requirements. Some authors [2]-[6] have proposed different approaches which take advantage of the non-stationarity of such sources but they still require their independence or uncorrelation.

Using a different concept, we recently introduced the TIFROM approach [7], [8], which is based on a Time-Frequency (*TF*) analysis of the observed mixed signals. We showed that this method does not require the same assumptions as traditional BSS approaches. Especially, when required conditions are satisfied, this new method applies to underdetermined mixtures (i.e. $N > P$) for which it achieves a partial BSS.

This paper aims at showing that this TIFROM approach is in addition able to separate dependent signals, which is a very attractive advantage over classical BSS methods. In Section 2, we recall the basics of the TIFROM approach. In Section 3 we show that this method can be applied to dependent signals. We then provide several experimental results in Section 4 and draw various conclusions in Section 5. For simplicity we consider throughout this paper the basic case of two sources and two observations. However, we emphasize the fact that this approach is not restricted to this case, as shown in [8] and in a future paper (for N sources and P observations).

2. PRINCIPLE OF THE "TIFROM" APPROACH

2.1. Model

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

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

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

The separation of the sources s_i can classically only be performed up to a scale factor and a permutation [1] and BSS may thus be seen as a method for finding an estimate of $\tilde{A}^{-1} = \Lambda P A^{-1}$, where Λ and P are resp. arbitrary diagonal and permutation matrices. Inside this class of matrices, we here focus on:

$$\tilde{A}^{-1} = \begin{bmatrix} 1 & 1 \\ 1/c_1 & 1/c_2 \end{bmatrix}^{-1} \quad (2)$$

[1] The mixtures are assumed to be non-degenerate throughout this paper.

where

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

which yields :

$$\underline{y}(n) = \tilde{A}^{-1}\underline{x}(n) = [a_{11}s_1(n), a_{12}s_2(n)]^T. \tag{4}$$

2.2. Time-frequency approach

The TIFROM approach is based on a simple and efficient way to automatically determine the above coefficients c_i using the *TF* information included in the observations.

To this end we compute the *short-time Fourier transforms (*STFT*)* [9]-[11] of the observations, denoted $X_i(n,\omega)$, which represent their contributions in the short time and frequency windows resp. centered on n and ω.
We require the following assumptions :

Assumption 1 *The mixing matrix A is such that $a_{ij} \neq 0$, $\forall\, i,\ j$ and the power of each source is non negligible at least at some times n.*

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

The TIFROM method is then based on the complex ratio:

$$\alpha(n_j,\omega_k) = \frac{X_1(n_j,\omega_k)}{X_2(n_j,\omega_k)}, \tag{5}$$

which is computed for each *TF* window. The linearity of the *STFT* operator leads to:

$$\alpha(n_j,\omega_k) = \frac{a_{11}S_1(n_j,\omega_k) + a_{12}S_2(n_j,\omega_k)}{a_{21}S_1(n_j,\omega_k) + a_{22}S_2(n_j,\omega_k)}. \tag{6}$$

Therefore, if only one source occurs in the *TF* window (n_j, ω_k), then $\alpha(n_j,\omega_k)$ is equal to the corresponding coefficient value, among c_1 and c_2 defined in (3). Note that in practical situations there always exists a small amount of noise in the observations so that $X_2(n_j,\omega_k)$ is always different from zero and $\alpha(n_j,\omega_k)$ is always defined, for each j and k. We add the following assumption:

Assumption 3 *When several sources occur in a given set of adjacent* TF *windows they should vary so that $\alpha(n,\omega)$ does not take the same value in all these windows.*

It may be shown easily that if only source $s_i(n)$ is present in several time-adjacent windows[3] (n_j,ω_k), then $\alpha(n_j,\omega_k)$ is constant and equal to c_i over these successive windows. On the contrary, it takes different values over these windows if both sources are present and if Assumption 3 is met.

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

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

To exploit this property, we proposed to analyze, for each frequency ω_k, the sample variance of the complex ratio $\alpha(n_j,\omega_k)$ on series Γ_q of M short half-overlapping time windows corresponding to adjacent n_j : $var[\alpha](\Gamma_q,\omega_k) = \frac{1}{M}\sum_{j=1}^{M} |\alpha(n_j,\omega_k) - \overline{\alpha}(\Gamma_q,\omega_k)|^2$, where the sample mean is defined as: $\overline{\alpha}(\Gamma_q,\omega_k) = \frac{1}{M}\sum_{j=1}^{M}\alpha(n_j,\omega_k)$.

If e.g. $S_2(n_j,\omega_k) = 0$ for these M windows, then (6) shows that $\alpha(n_j,\omega_k)$ is constant over them, so that its variance $var[\alpha](\Gamma_q,\omega_k)$ is equal to zero. Conversely, under Assumption 3, if both $S_1(n_j,\omega_k)$ and $S_2(n_j,\omega_k)$ are different from zero then $var[\alpha](\Gamma_q,\omega_k)$ is significantly different from zero.

So, by searching for 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 is then given by $\overline{\alpha}(\Gamma_q,\omega_k)$. We find the second coefficient value c_i by searching for 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 coefficient values defined in (3). The separated signals are then derived from these values by using i) either the original version of the TIFROM approach based on individual source extractions that we proposed in [7], [8] or ii) its new version that we introduced in this paper, which is based on the matrix (2). If the lowest value of the ratio variance is obtained when s_2 is zero this yields (3) and (4). Otherwise a permutation occurs in (3) and (4).

3. DEPENDENT SIGNALS

As stated above, ICA 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. The TIFROM approach is totally different, as it uses sample statistics of a single signal realization to determine some domains in the TF plane where a single source occurs. It therefore only requires such domains to exist and applies to (realizations of) various dependent sources which meet this condition.

To illustrate this capability, consider for example the two source signals $s_1(n) = u(n) + v(n)$ and $s_2(n) = v(n) + w(n)$, where $u(n)$, $v(n)$ and $w(n)$ are three stationary independent zero-mean signals and where:
a) $v(n)$ only has components in the frequency band $[f_1, f_2]$, and $u(n)$ and/or $w(n)$ also have components at the frequencies where $v(n)$ occurs,
b) $u(n)$ only has components in the frequency band $[0, f_2]$,
c) $w(n)$ only has components above f_1.
The cross-correlation of $s_1(n)$ and $s_2(n)$ is non-zero, due to their common component $v(n)$. 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 the TIFROM approach, despite their dependence, thanks to the differences in their *TF* representations.
A similar situation occurs with musical instruments, where each one has his own time properties (attack, decay, sustain, release) and frequency components which make it sound differently from another one. Now two different instruments playing in the same tone have common frequencies which make their signals correlated and thus dependent. Moreover, thanks to their own properties, they usually do not vary in a coherent way over time-adjacent *TF* windows and assumption (3) of the TIFROM approach therefore holds to them. This is an important case as traditional BSS methods, like kurtosis maximization cannot separate this kind of signals.

4. EXPERIMENTAL RESULTS

To illustrate our ability to separate dependent signals, we consider the case of musical instruments. Source s_1 is a guitar playing a D chord, which consists in D, $F\#$, A. Source s_2 is a D from a singer. These sources are dependent as we can see on Fig. 1 which shows the absolute value of zero-lag cross-correlation coefficients $|E[s_1 s_2]|/\sqrt{E[s_1^2]E[s_2^2]}$, computed for each considered time window. We recorded these two sources using CD quality (16 bits, 44,1 kHz) and then mixed them using the matrix :

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0.9 \\ 0.8 & 1 \end{bmatrix} \quad (7)$$

giving for s_1 resp. SNR's of 1.6 dB and -1.3 dB on x_1 and x_2. Spectrograms of the sources (Fig. 2 and 3), with $N_{STFT} = 256$ samples per STFT window, clearly show that there exist some differences in the TF plane between these signals. As an example, we analyzed the variance of $\alpha(n_j, \omega_k)$ for $M = 8$ on 1.13 s of signal (50000 samples), which took approximately 1 s with matlab code on a 1GHz PIII, and plotted in Fig. 4 and 5 the results $-log_{10}(var[\alpha](\Gamma_q, \omega_k))$ and $\frac{1}{var[\alpha](\Gamma_q,\omega_k)}$. One can easily see that there exist some areas with low variance (bright areas in Fig. 4 and peaks in Fig. 5), corresponding to windows where only one source occurs. These settings give an output SNR of 34.2 dB for s_1 and 71.3 dB for s_2, which are quite good values for such dependent signals. Note that we have been unable to separate these sources with the classical kurtosis maximization method, due to the dependence of the sources. We give some additional SNR results in Table 1 for different STFT and variance analysis window sizes. As we can see, the separation is always achieved with good SNR's. Experimental results show that the selected areas should have a variance below 10^{-3} to provide good results for "normalized" signals.

Table 1: Output SNR vs N_{STFT} and M for each output.

		N_{STFT}		
M		64	128	256
4	s_1	25.1	44.2	34.0
	s_2	50.4	82.7	67.8
6	s_1	25.4	26.0	28.0
	s_2	34.0	57.0	54.1
8	s_1	30.4	34.2	34.2
	s_2	34.6	49.0	71.3
10	s_1	24.9	29.7	27.8
	s_2	36.7	50.7	61.6
12	s_1	38.0	31.0	29.8
	s_2	42.0	67.3	52.7

5. CONCLUSION

In this paper, we recalled the basics of the TIFROM approach that we recently introduced in [7], [8]. We then proved and illustrated its ability to separate dependent signals. Unlike classical ICA methods which separate the sources by combining the observations so that the output signals are independent our approach 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 local area in the *TF* plane. Then it automatically determines such an area and derives coefficients which e.g. allow one to directly build an inverse mixing matrix in the case we considered here. This makes it possible to separate classes of signals for which classical methods fail, e.g. dependent signals, provided there exist some areas in the time frequency plane where only one source occurs. As an example we recorded audio signals from a guitar and a singer playing in the same tone, giving two dependent signals. We then showed that we can successfully separate them using the TIFROM approach. For the sake of clarity we presented the simple case of 2 sources and 2 observations but this approach is easily extended for Nsources and P observations, giving source separation if $N \leq P$ or partial source separation otherwise, as will be shown in a future paper.

6. REFERENCES

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

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

[3] A. Hyvarinen, "Blind source separation by nonstation-

arity of variance: a cumulant-based approach," *IEEE Trans. on Neural Networks*, vol. 12, no. 6, pp. 1471–1474, November 2001.

[4] 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.

[5] Y. Deville, F. Abrard, and M. Benali, "A new source separation concept and its validation on a preliminary speech enhancement configuration," in *Proceedings of CFA2000*, Lausanne, Switzerland, September 3-6, 2000, pp. 610–613.

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

[7] F. Abrard, Y. Deville, and P. R. White, "A new source separation approach for instantaneous mixtures based on time-frequency analysis," in *Proceedings of ECM^{2}S*, Toulouse, France, May 2001.

[8] ——, "From blind source separation to blind source cancellation in the underdetermined case: a new approach based on time-frequency analysis," in *Proceedings of ICA 2001*, San Diego, CA, Dec., 9-13 2001.

[9] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE SP Magazine*, vol. 9, pp. 21–67, April 1992.

[10] L. Cohen, "Time-frequency distributions - a review," in *Proceedings of the IEEE*, vol. 77, No. 7, July 1989, pp. 941–979.

[11] ——, *Time-frequency analysis*. Englewood Cliffs, New Jersey: Prentice hall PTR, 1995.

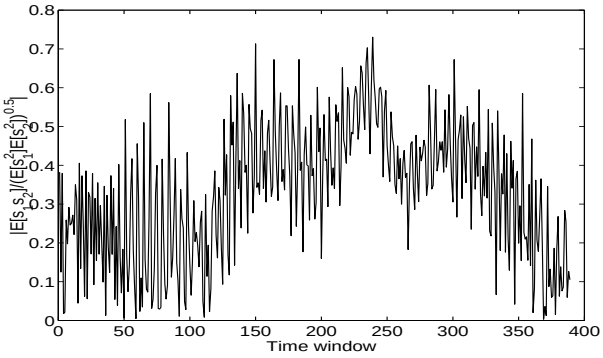

Figure 1: Absolute value of cross-correlation coefficient $|E[s_1 s_2]|/\sqrt{E[s_1^2]E[s_2^2]}$ for each 256-sample window.

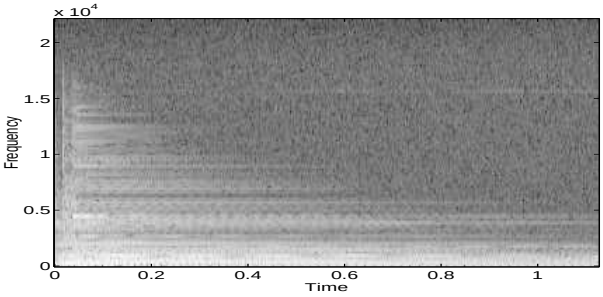

Figure 2: Spectrogram of guitar s_1.

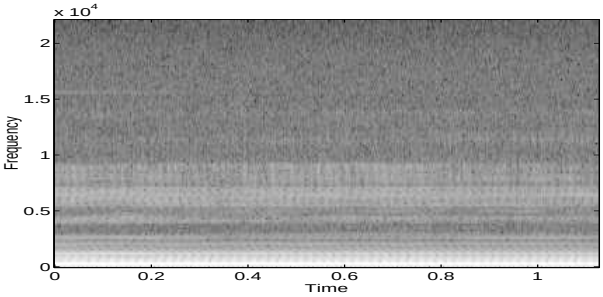

Figure 3: Spectrogram of voice s_2.

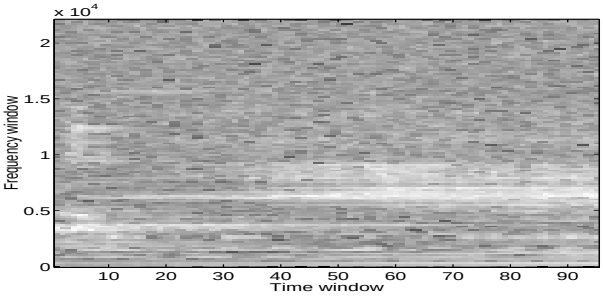

Figure 4: Time-Frequency representation of $-log_{10}(var[\alpha](\Gamma_q, \omega_k))$. Axes units : Time window indices, corresponding to [0 s, 1.13 s]. Frequency window indices, corresponding to [0 Hz, 22.05 kHz].

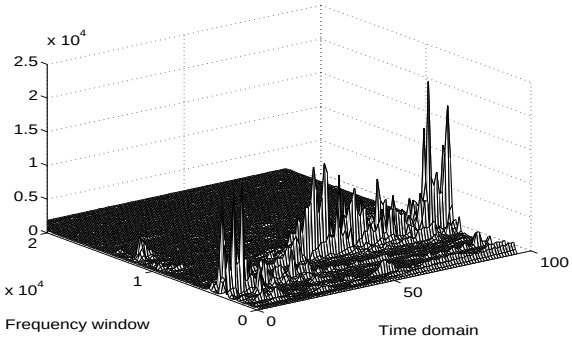

Figure 5: Time-Frequency representation of $\frac{1}{var[\alpha](\Gamma_q, \omega_k)}$. Axes units : Time window indices, corresponding to [0 s, 1.13 s]. Frequency window indices, corresponding to [0 Hz, 22.05 kHz].