4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

TEMPORAL AND TIME-FREQUENCY CORRELATION-BASED BLIND SOURCE SEPARATION METHODS

Yannick DEVILLE

Université Paul Sabatier
Laboratoire d'Acoustique, Métrologie, Instrumentation
Bât. 3R1B2, 118 Route de Narbonne, 31062 Toulouse Cedex, FRANCE
ydeville@cict.fr

ABSTRACT

In this paper, we propose two versions of a correlation-based blind source separation (BSS) method. Whereas its basic version operates in the time domain, its extended form is based on the time-frequency (TF) representations of the observed signals and thus applies to much more general conditions. The latter approach consists in identifying the columns of the (permuted scaled) mixing matrix in TF areas where this method detects that a single source occurs. Both the detection and identification stages of this approach use local correlation parameters of the TF transforms of the observed signals. This BSS method, called TIFCORR (for TIme-Frequency CORRelation-based BSS), is shown to yield very accurate separation for linear instantaneous mixtures of real speech signals (output SNR's are above 60 dB).

1. INTRODUCTION

Blind source separation (BSS) methods aim at restoring a set of unknown source signals from a set of observed signals which are mixtures of these source signals [1]-[2]. Most of the approaches that have been developed to this end are based on Independent Component Analysis (ICA). They assume the sources to be random stationary statistically independent signals, and they recombine the available observed signals so as to obtain statistically independent output signals. The latter signals are then equal to the sources, up to some indeterminacies and under some conditions (especially, at most one source may be Gaussian for such methods to be applicable if no additional constraints are set on the sources).

Whereas many ICA methods have been proposed in the last 15 years, a few other concepts for BSS have also been considered. Especially, several investigations based on time-frequency (TF) analysis have been reported [3]-[11]. Two main classes emerge from these TF-BSS methods. The first one is still significantly related to classical BSS approaches, as it consists of TF adaptations of previously developed joint-diagonalization methods, with subsequent modifications [3]-[5]. The other class includes several methods based on the ratio of the TF transforms of the observed signals [6]-[10]. Some of these methods require the sources to have no overlap in the TF domain [6]-[8]. On the contrary, only slight differences in the TF representations of the sources are requested by the methods that we proposed in [9]-[10].

It should be noted that, unlike ICA-based BSS methods, TF-based BSS approaches are intrinsically well-suited to non-stationary signals and set no restrictions on the gaussianity of the sources. They are therefore e.g. especially attractive for speech signals.

This paper mainly introduces a TF-BSS approach based on a different type of parameter than in the previously reported TF-BSS methods, i.e. on the local correlation of the observed signals in the TF domain. Before describing it, we present a purely temporal version of this approach. This aims at first introducing the proposed concept in a simpler framework, but also yields an additional temporal BSS method. The latter method only applies to more restrictive conditions than its subsequent TF extension, as shown below.

The remainder of this paper is therefore organized as follows. We define the proposed temporal BSS method in Section 2 and then introduce its TF extension in Section 3. In Section 4, we describe tests performed with artificial mixtures of real speech sources. Conclusions are drawn from this investigation in Section 5 and more specific topics are addressed in the Appendix.

2. PROPOSED TEMPORAL APPROACH

2.1. Assumptions and definitions

In this section, devoted to the proposed statistical temporal BSS approach, we assume that N unknown, random, possibly complex-valued, source signals $s_j(t)$ are mixed in a linear instantaneous way, thus providing a set of P observed signals $x_i(t)$. This reads in matrix form:

$$x(t) = As(t), \tag{1}$$

where $s(t) = [s_1(t) \dots s_N(t)]^T$ and $x(t) = [x_1(t) \dots x_P(n)]^T$ and where the matrix A consists of all possibly complex-valued mixing coefficients a_{ij}, which is denoted $A = [a_{ij}]$ below. We now introduce the only assumptions that we make in this section with respect to the sources and mixing stage, and the associated definitions.

Definition 1: a source is said to "occur alone" in a time area if the variance of only this source is not equal to zero in this time area[1].

The considered time area may be restricted to a single time t from a theoretical point of view, as we use a statistical approach in this section. However, in practice, all signal means are estimated over time windows and the considered time area then consists of such a window.

[1] This definition corresponds to the theoretical point of view. From a practical point of view, this means that the ranges of values taken by all other sources are negligible with respect to those of the source that occurs alone.

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

Assumption 1: i) each source is visible in the time domain and ii) there exist no time areas where the variances of all sources are equal to zero.

It should be noted that the considered sources are therefore non-stationary[2].

Assumption 2: the sources are mutually uncorrelated.

Note that we do not require the complete independence of the sources, as the approach introduced below only uses their second-order statistics.

Assumption 3: the number P of observations is equal to the number N of sources.

The matrix A is then a square matrix.

Assumption 4: i) A is invertible and ii) none of its elements is equal to zero.

For the sake of simplicity, the notations $s(t)$ and $x(t)$ introduced above directly refer to the centered version of the signals in this section (whereas they correspond to the original, possibly non-centered, signals in Section 3).

2.2. Problem statement and solution

BSS would ideally consist in deriving an estimate $\hat{A}$ of A, so as to then determine the vector

$$y(t) = \hat{A}^{-1}x(t) \qquad (2)$$

$$= \hat{A}^{-1}As(t). \qquad (3)$$

Each component $y_j(t)$ of the vector $y(t)$ would then be equal to the source signal having the same index, i.e. to $s_j(t)$ (up to estimation errors). It is well known however that this can only be achieved up to two types of indeterminacies, which resp. concern the scale factors and order with which the source signals appear in the output vector $y(t)$. We detail this phenomenon hereafter, as we will use this discussion in the BSS method proposed below. Any of the mixed signals corresponding to the matrix form (1) reads exlicitly:

$$x_i(t) = \sum_{j=1}^{N} a_{ij}s_j(t) \qquad i = 1 \dots P. \qquad (4)$$

However, it may also be expressed in a different way, by applying two transforms to it. The first one consists in changing the order in which the terms (resp. associated to the sources $s_1(t) \dots s_N(t)$) appear in the sum in (4). This is achieved by applying a permutation $\sigma(.)$ to the indices j in this sum. The above mixed signal then reads:

$$x_i(t) = \sum_{j=1}^{N} a_{i,\sigma(j)}s_{\sigma(j)}(t) \qquad i = 1 \dots P. \qquad (5)$$

The second transform consists in normalizing the scales of the contributions of each source signal $s_{\sigma(j)}(t)$ with respect to the contribution of this signal in the first mixed signal[3]. The latter contribution is equal to $a_{1,\sigma(j)}s_{\sigma(j)}(t)$, so that we express $x_i(t)$ as:

$$x_i(t) = \sum_{j=1}^{N} \frac{a_{i,\sigma(j)}}{a_{1,\sigma(j)}} a_{1,\sigma(j)}s_{\sigma(j)}(t) \qquad i = 1 \dots P. \qquad (6)$$

We therefore introduce the notations:

$$s'_j(t) = a_{1,\sigma(j)}s_{\sigma(j)}(t) \qquad (7)$$

$$b_{ij} = \frac{a_{i,\sigma(j)}}{a_{1,\sigma(j)}}, \qquad (8)$$

where $s'_j(t)$ are the permuted scaled source signals and b_{ij} are the corresponding permuted scaled mixing coefficients. Eq. (6) then reads:

$$x_i(t) = \sum_{j=1}^{N} b_{ij}s'_j(t) \qquad i = 1 \dots P \qquad (9)$$

or in matrix form:

$$x(t) = Bs'(t), \qquad (10)$$

with $s'(t) = [s'_1(t) \dots s'_N(t)]^T$ and $B = [b_{ij}]$. We thus get back to the initial mixture expression (1), except that the mixed signals are now expressed with respect to the permuted scaled source signals $s'_j(t)$. The discussion at the beginning of this subsection may then be reinterpreted as follows: assume that we succeed in deriving an estimate $\hat{B}$ of B. Then, by computing the output vector

$$y'(t) = \hat{B}^{-1}x(t) \qquad (11)$$

$$= \hat{B}^{-1}Bs'(t), \qquad (12)$$

all components $y'_j(t)$ of this vector are resp. equal to $s'_j(t)$ (up to estimation errors), i.e. to the contributions of the (possibly) permuted sources in the first mixed signal.

The two BSS approaches proposed in this paper precisely aim at estimating this matrix B, then providing the corresponding vector $y'(t)$ of separated source signals.

In the framework of this section, we propose a method which takes advantage of the above *Assumption 1-i*, i.e. of the fact that there exist time areas where each source occurs alone. Such areas should first be detected, so as to operate inside them. As all observed signals are proportional in any such area, an appealing approach for detecting these areas consists in checking the cross-correlation coefficients $\rho_{x_1x_i}(t)$ between the observed signals $x_1(t)$ and $x_i(t)$, defined as:

$$\rho_{x_1x_i}(t) = \frac{E\{x_1(t)x_i^*(t)\}}{\sqrt{E\{x_1(t)x_1^*(t)\}E\{x_i(t)x_i^*(t)\}}}, \qquad (13)$$

where $E\{.\}$ stands for expectation. More precisely, we show in the Appendix that a necessary and sufficient condition for a source to occur alone at time t is:

$$|\rho_{x_1x_i}(t)| = 1 \qquad \forall\, i, \quad 2 \leq i \leq P. \qquad (14)$$

Now consider such an area where a single source occurs, say $s_k(t)$. The observed signals (4) then become restricted to:

$$x_i(t) = a_{ik}s_k(t) \qquad i = 1 \dots P. \qquad (15)$$

Again using correlation parameters associated to these observed signals then makes it possible to identify part of the matrix B. More precisely, when (15) is met,

$$\frac{E\{x_i(t)x_1^*(t)\}}{E\{x_1(t)x_1^*(t)\}} = \frac{a_{ik}}{a_{1k}} \qquad i = 2 \dots P. \qquad (16)$$

[2] More precisely, they are long-term non-stationary, but they should be short-term stationary in practice in order to make it possible to estimate the above-mentioned signal means over short time windows.

[3] The same principle may of course may applied to any other mixed signal instead.

The set of values thus obtained for all observations indexed by i identifies one of the columns of B, as shown by (8) (which also indicates that the first row of B consists of 1). By repeatedly performing such a column identification for time areas associated to all sources, we eventually identify the overall matrix B, which completes the proposed approach.

The overall structure of the BSS method thus introduced may be summarized as follows. This approach consists of 3 stages, i.e:

1. The detection stage consists in detecting the time areas where a single source occurs. This is achieved by finding the times t (or practical time windows) where $|\rho_{x_1 x_i}(t)|$ is very close to 1 for all $i, \quad 2 \leq i \leq P$.
2. The identification stage consists in identifying the columns of B. This is achieved by successively considering each of the single-source areas detected in the first stage as follows. The correlation parameters on the left-hand side of (16) yield a column of B. This column is kept only if its distance with respect to all previously identified columns is above a user-defined threshold, showing that the considered time area does not contain the same source as the previous ones. The identification procedure ends when the number of columns of B thus kept becomes equal to the number of sources (this is guaranteed to occur because all sources are assumed to be visible in the considered data)[4].
3. The combination stage consists in recombining the mixed signals according to (11), in order to obtain the extracted source signals.

3. TIME-FREQUENCY EXTENSION OF THE PROPOSED APPROACH

3.1. Motivations and basic principles

The approach that we introduced in the previous section has the advantage of being simple. It may be considered to be of limited practical applicability however, because it assumes all sources to occur alone in associated time areas, which is a restrictive condition. But it opens the way to a much more powerful method if we now think in terms of the time-frequency (TF) distributions of the signals, instead of their plain time distributions considered up to this point. Indeed, the TF counterpart of the above approach may be briefly defined as follows: assume that each source occurs alone in a TF area, then it may be expected from the above presentation that corresponding columns of B may be identified in such areas, thus allowing one to eventually perform BSS.

The remainder of this section presents this approach in a more formal way. It should be stressed again that the TF version of the proposed approach that we will thus introduce has a much broader scope of application that its above temporal version. This is due to the fact that it only requests each source to occur alone in a small bounded TF area, which is a much less restrictive condition that having it alone in a time area, i.e. at *all* frequencies for that time area. For instance, mixtures of continuous speech signals meet the assumptions considered here, because most of the energy of these signals is concentrated in a few bounded time-varying frequency regions, corresponding to formants.

[4] A possibly more robust version of this approach consists in keeping the identified columns corresponding to all available single-source areas, and then clustering them, so as to eventually derive each column of B as an average of all its identified occurences.

It should also be noted that the considered time-frequency tools, which are presented in Subsection 3.2, are originally defined in a deterministic framework. This paper uses them in such a framework and concerns two cases i.e: i) either the considered original sources are deterministic or ii) they are random processes, but in the latter case only a single realization of these processes is considered (this is what is actually available in practice). The following discussion then concerns this single, deterministic, realization and the tools and properties that we use only require such a realization.

3.2. Time-frequency tools

Many TF representations have been proposed over the last fifty years and are e.g. presented in [12]-[13]. We here use the short-time Fourier transform (STFT), esp. because it yields low computional load thanks to FFT algorithms and it does not introduce interference terms, thus keeping the linear instantaneous mixing structure when applied to the considered observed signals.

The STFT of a signal $u(t)$ is obtained by first multiplying the signal $u(t')$ by a shifted windowing function $h^*(t'-t)$, centered at time t. This yields the windowed signal $u(t')h^*(t'-t)$. This signal depends on two time variables, i.e. the selected time t where the local spectrum of $u(t')$ is analyzed and the running time t'. The STFT of $u(t)$ is then defined as the Fourier transform of the above windowed signal, i.e:

$$U(t,\omega) = \int_{-\infty}^{+\infty} u(t')h^*(t'-t)e^{-j\omega t'}dt'. \tag{17}$$

$U(t,\omega)$ is then the contribution of the considered signal $u(t)$ in the TF area corresponding to the short time window associated with t and to the angular frequency ω.

3.3. Assumptions and definitions

The same assumptions and definitions as in Subsection 2.1 are considered hereafter, except that i) we eventually use a deterministic framework, as explained above, and ii) the temporal concepts used in Subsection 2.1 are here replaced by their TF counterparts, i.e:

Definition 1-TF: a source is said to "occur alone" in a TF area if only this source is such that its centered TF transform is not equal to zero everywhere in this TF area.

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

Assumption 1-TF: i) each source is visible in the TF domain and ii) there exist no TF areas where the centered TF transforms of all sources are equal to zero everywhere.

More precisely, each value of a TF transform corresponds to a single "TF point", associated to a single angular frequency ω and to a complete time window defined by the selected analysis time t and considered finite-length windowing function $h^*(.)$, as shown in Subsection 3.2. The BSS method that we propose below then uses means associated to these TF transforms, computed over "analysis zones" which consist of adjacent TF points. An analysis zone may have any shape in the TF domain. We here focus on the case when it forms a "temporal line", i.e. when all its points correspond to the same frequency ω and to adjacent (half-overlapping) time windows. The latter windows resp. correspond to a discrete set of L time positions t_p, with $p = 1 \ldots L$. This set is denoted as T hereafter. Each analysis zone is then specified in terms of the couple (T,ω), which completely defines the part of the TF domain associated to this analysis zone.

The proposed BSS method then uses the following parameters, associated to the above-defined analysis zones. For any signal $v(t)$, whose TF transform is denoted $V(t, \omega)$, the mean of its TF transform over the considered analysis zone is:

$$\overline{V}(T, \omega) = \frac{1}{L} \sum_{p=1}^{L} V(t_p, \omega). \tag{18}$$

Similarly, for any couple of signals $v_1(t)$ and $v_2(t)$, whose TF transforms are denoted $V_1(t, \omega)$ and $V_2(t, \omega)$, the cross-correlation of the (centered versions of the) TF transforms of these signals over the considered analysis zone is measured either by

$$\begin{aligned} C_{v_1 v_2}(T, \omega) \quad = \quad & \frac{1}{L} \sum_{p=1}^{L} \{ [V_1(t_p, \omega) - \overline{V_1}(T, \omega)] \\ & * [V_2(t_p, \omega) - \overline{V_2}(T, \omega)]^* \} \end{aligned} \tag{19}$$

or by the corresponding TF local cross-correlation coefficient:

$$r_{v_1 v_2}(T, \omega) = \frac{C_{v_1 v_2}(T, \omega)}{\sqrt{C_{v_1 v_1}(T, \omega) C_{v_2 v_2}(T, \omega)}}. \tag{20}$$

The source uncorrelation assumption of Section 2 is then replaced by:

Assumption 2-TF: over each analysis zone (T, ω), the (centered) TF transforms of the sources are uncorrelated, i.e: $C_{s_i s_j}(T, \omega) = 0, \ \forall \, i \neq j$.

3.4. Proposed approach

As we progressively introduced most required concepts in the previous sections and subsections, the resulting proposed TF-BSS method may now be directly presented, as the TF adaptation of the overall temporal approach that we described at the end of Section 2. It therefore consists of the same 3 stages as the latter approach, adapted to the TF context however, and therefore preceded by a pre-processing stage, i.e:

1. The pre-processing stage consists in deriving the STFT transforms $X_i(t, \omega)$ of the mixed signals, according to (17).
2. The detection stage is based on the following property, shown in the Appendix: a necessary and sufficient condition for a source to occur alone in the TF analysis zone (T, ω), is:

   $$|r_{x_1 x_i}(T, \omega)| = 1 \qquad \forall \, i, \quad 2 \leq i \leq P. \tag{21}$$

   In this stage, we therefore detect the analysis zones where a single source occurs by checking in which zones $|r_{x_1 x_i}(T, \omega)|$ is very close to 1 for all $i, \quad 2 \leq i \leq P$.
3. In the identification stage, we then identify the columns of B in single-source analysis zones. This is based on the same approach as in the temporal BSS method, except that the parameter in (16) is here replaced by[5]:

   $$\frac{C_{x_i x_1}(T, \omega)}{C_{x_1 x_1}(T, \omega)}, \tag{22}$$

   which is equal to a_{ik}/a_{1k} when only source $s_k(t)$ occurs in the considered TF analysis zone.
4. In the combination stage, we eventually recombine the mixed signals according to (11), in order to obtain the extracted source signals.

It should be noted that the overall structure of this approach has some similarities with the TF approach that we proposed in [9]-[10]. However, these two approaches use completely different detection and identification parameters. As the detection and identification stages of any such approach are independent one from the other, we may also derive mixed approaches by using the detection method of one of the two approaches resp. proposed in this paper and in [9]-[10]. and the identification stage of the other approach.

4. TEST RESULTS

We now present tests performed with two artificial linear instantaneous mixtures of two real speech signals, sampled at 22050 Hz. The mixing matrix is set to:

$$\begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}. \tag{23}$$

The corresponding matrix B is therefore equal to

$$\begin{bmatrix} 1 & 1 \\ 0.9 & 1.1111 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & 1 \\ 1.1111 & 0.9 \end{bmatrix}, \tag{24}$$

depending whether it corresponds to a non-permuted or permuted version of the source signals.

The considered sources signals are shown in Fig. 1. 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), due to the considered hard mixing conditions. Nevertheless, the TF BSS method proposed in this paper succeeds in separating these signals with a high accuracy, as will now be shown.

As an example, the proposed approach has been applied with 256 samples per STFT window and 8 such windows per analysis zone. The estimated matrix $\hat{B}$ thus obtained is equal to

$$\begin{bmatrix} 1.0000 & 1.0000 \\ 0.8999 & 1.1110 \end{bmatrix}, \tag{25}$$

which is extremely close to the first expression in (24). Consequently, the estimated output signals are almost equal to the (scaled and non-permuted) sources, as confirmed by Fig. 6. More precisely, the Signal/Noise Ratios (SNR's) obtained for these two estimated sources are resp. equal to 68.9 and 61.2 dB, while the SNR's in the observed mixed signals are equal to 3.6 and - 3.6 dB. This clearly demonstrates the high separation capability of the proposed approach.

5. CONCLUSION

In this paper, we introduced two versions of a correlation-based BSS approach. Whereas its first version operates in the time domain, its second version is based on TF analysis and thus applies to much more general conditions. The latter approach, called TIFCORR for TIme-Frequency CORRelation-based BSS, consists in identifying the columns of the (permuted scaled) mixing matrix in TF areas where this method detects that a single source occurs. We experimentally showed that it yields very good performance

[5]In the identification stage, the non-centered version of the cross-correlation parameters $C_{..}(T, \omega)$ may be used instead and is simpler.

for linear instantaneous mixtures of real speech sources. Our future investigations will first consist in a more detailed characterization of the experimental performance of the two approaches that we proposed in this paper. We will also compare, both from the theoretical and experimental points of view, the TIFCORR method to the TF ratio-based approach that we introduced in [9]-[10], which has the same structure by uses completely different parameters, as explained in Section 3. We will also aim at extending the approach proposed in this paper to convolutive mixtures.

6. REFERENCES

[1] A. Hyvarinen, J. Karhunen, E. Oja, ”Independent Component Analysis”, Wiley, New York, 2001.

[2] J.-F. Cardoso, ”Blind signal separation: statistical principles”, Proceedings of the IEEE, vol. 86, no. 10, pp. 2009-2025, Oct. 1998.

[3] A. Belouchrani, M.G. Amin, ”Blind source separation using time-frequency distributions: algorithm and asymptotic performance”, Proc. ICASSP’97, pp. 3469-3472, Munich, Germany, April 21-24, 1997.

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

[5] A. Holobar, C. Févotte, C. Doncarli, D. Zazula, ”Single autoterms selection for blind source separation in time-frequency plane”, Proc. EUSIPCO’2002, Toulouse, France, Sept 3-6, 2002.

[6] A. Jourjine, S. Rickard, O. Yilmaz, ”Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures”, Proc. ICASSP’2000, vol. 5, pp. 2985-2988, Istanbul, Turkey, June 5-9, 2000.

[7] R. Balan, J. Rosca, ”Statistical properties of STFT ratios for two channel systems and applications to blind source separation”, Proc.ICA’2000, Helsinki, Finland, June 18-22, 2000.

[8] S. Rickard, R. Balan, J. Rosca, ”Real-time time-frequency based blind source separation”, Proc.ICA’2001, pp. 651-656, San Diego, California, Dec. 9-13, 2001.

[9] F. Abrard, Y. Deville, P. White, ”A new source separation approach based on time-frequency analysis for instantaneous mixtures”, Proc. ECM2S’2001, pp. 259-267, Toulouse, France, May 30 - June 1, 2001.

[10] 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”, Proc. ICA’2001, pp. 734-739, San Diego, California, Dec. 9-13, 2001.

[11] L. Giulieri, N. Thirion-Moreau, P.-Y. Arquès, ”Blind sources separation using bilinear and quadratic time-frequency representations”, Proc. ICA’2001, pp. 486-491, San Diego, California, Dec. 9-13, 2001.

[12] F. Hlawatsch, G.F. Boudreaux-Bartels, ”Linear and quadratic time-frequency signal representations”, IEEE Signal Processing Magazine, pp. 21- 67, April 1992.

[13] L. Cohen, ”Time-frequency distributions - a review”, Proceedings of the IEEE, vol. 77, no. 7, pp. 941-981, July 1989.

A. APPENDIX

We here show the validity of the detection criteria (14) and (21) resp. used in the temporal and TF versions of the proposed BSS approach.

In the frame of the temporal version of this approach described in Section 2, by using the corresponding assumptions it may be shown that, for the mixed signals defined in (4), the cross-correlation coefficient defined in (13) may be expressed as:

$$\rho_{x_1 x_i}(t) = \frac{< V_1(t), V_i(t) >}{||V_1(t)||.||V_i(t)||}, \qquad (26)$$

where the notations $< .,. >$ and $||.||$ resp. stand for the inner product and vector norm, and where the j-th component of each N-dimensional vector $V_i(t)$ is equal to $a_{ij}\sqrt{\lambda_j(t)}$, with:

$$\lambda_j(t) = E\{s_j(t) s_j^*(t)\}. \qquad (27)$$

Applying Schwarz inequality to (26) then shows that

$$|\rho_{x_1 x_i}(t)| \leq 1 \qquad \forall\, i, \quad 1 \leq i \leq P, \qquad (28)$$

with equality if and only if $V_1(t)$ and $V_i(t)$ are dependent.

Let us now analyze this condition at a given time t, depending on the values of the source parameters $\lambda_j(t),\ j = 1, \ldots N$. Due to *Assumption 1-ii*, at least one of the values $\lambda_j(t)$ is not equal to zero. If only one is not equal to zero, then all vectors $V_i(t)$ are clearly dependent, so that equality holds for all of them in (28) and therefore the detection condition (14) is fulfilled.

The only case that remains to be considered is then the situation when at least two values $\lambda_j(t)$ and $\lambda_k(t)$ are not equal to zero. It may then be shown that if $V_1(t)$ and $V_i(t)$ were dependent for all $i, \quad 2 \leq i \leq P$, then the columns with indices j and k of the mixing matrix A would be dependent, which is not true due to *Assumption 4-i*. Therefore, in the considered case, at least one pair of vectors $(V_1(t), V_i(t))$ does not consist of dependent vectors, so that $|\rho_{x_1 x_i}(t)| < 1$ and the detection condition (14) is not fulfilled.

As an overall result, this condition is fulfilled if and only if exactly one of the values $\lambda_j(t)$ is not equal to zero at the considered time t, i.e. if only one source occurs at that time.

Now, as for the TF version of this approach described in Section 3, by using the corresponding assumptions it may be shown that, for the mixed signals defined in (4), the cross-correlation coefficient over an analysis zone defined in (20) may again be expressed according to the right-hand side of (26) except that the vectors in (26) here depend on the considered analysis zone (T, ω), and with the same notations except that $\lambda_j(t)$ is replaced by

$$\lambda_j(T, \omega) = C_{s_j s_j}(T, \omega), \qquad (29)$$

with $C_{s_j s_j}(T, \omega)$ defined according to (19). This leads to the same discussion as above, except that the considered time t is replaced by the considered analysis zone (T, ω). This eventually yields the detection criterion (21).

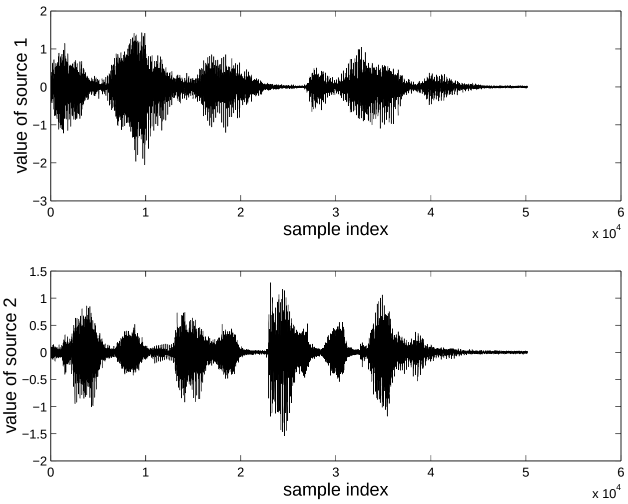

Fig. 1. Sample values of both sources.

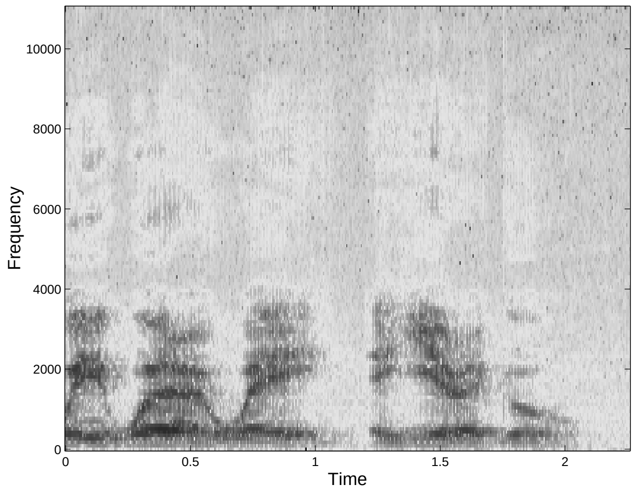

Fig. 2. Time-frequency transform of first source.

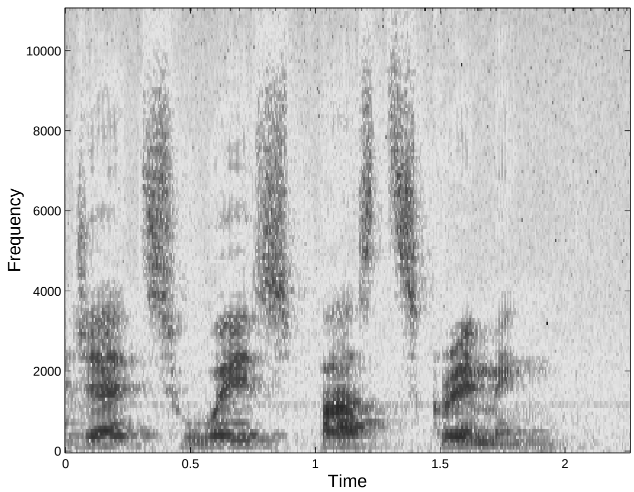

Fig. 3. Time-frequency transform of second source.

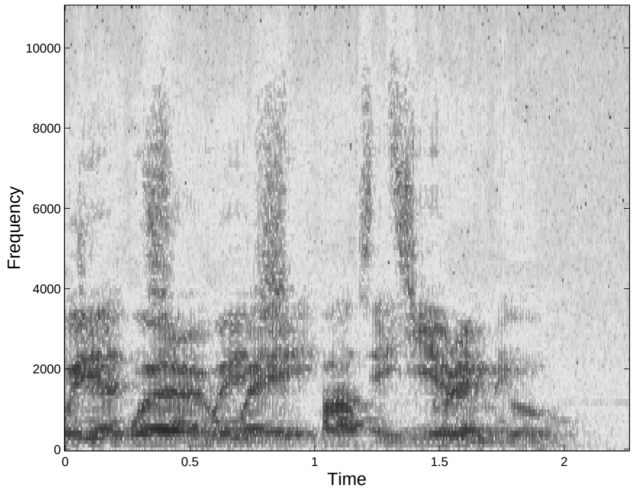

Fig. 4. Time-frequency transform of first mixed signal.

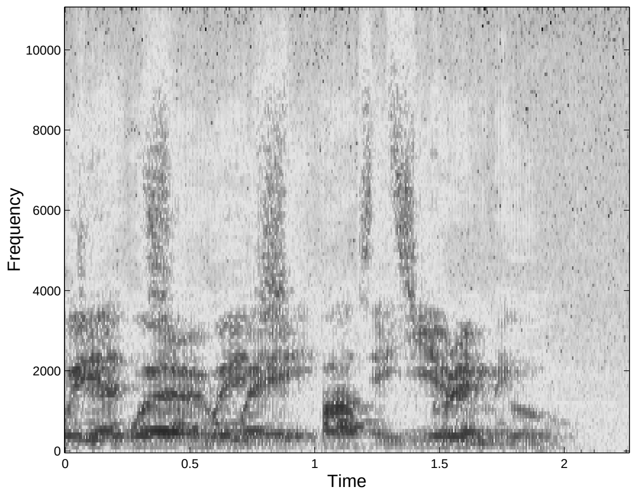

Fig. 5. Time-frequency transform of second mixed signal.

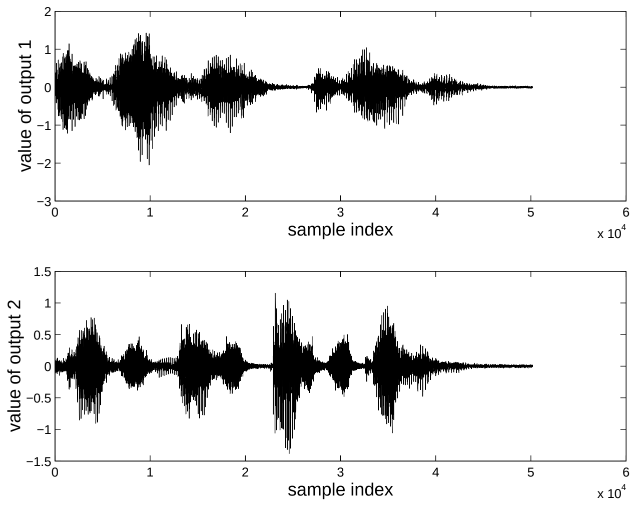

Fig. 6. Sample values of both estimated source signals.