ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

A time-scale correlation-based blind separation method applicable to correlated sources

Yannick Deville, Dass Bissessur, Matthieu Puigt, Shahram Hosseini, Hervé Carfantan

Observatoire Midi-Pyrénées - Université Paul Sabatier Toulouse 3
Laboratoire d'Astrophysique de Toulouse-Tarbes
14 Av. Edouard Belin, 31400 Toulouse, France

Abstract. We first propose a correlation-based blind source separation (BSS) method based on time-scale (TS) representations of the observed signals. This approach consists in identifying the columns of the (permuted scaled) mixing matrix in TS zones where this method detects that a single source is active. It thus sets very limited constraints on the sparsity of the sources in the TS domain. Both the detection and identification stages of this approach use local correlation parameters of the TS transforms of the observed signals. This BSS method, called TISCORR (for TIme-Scale CORRelation-based BSS), is an extension of our previous two temporal and time-frequency versions of this class of methods. Our second contribution in this paper consists in proving that all three approaches apply if the (transformed) source signals are linearly independent, thus allowing them to be correlated. This extends our previous demonstration, which only guaranteed our previous two approaches to be applicable to uncorrelated sources. Experimental tests show that our TISCORR method achieves good separation for linear instantaneous mixtures of real, correlated or uncorrelated, speech signals (output SIRs are above 40 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]. 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).

Some other BSS approaches, especially based on time-frequency analysis have also been reported (see e.g. [2]-[7] and references therein). Especially, we proposed in [7] two CORRelation-based BSS methods, i.e. a TEMPoral version (TEMPCORR) of this method and an extension (TIFCORR) which operates in the TIme-Frequency plane. Our contributions in this paper are related to these two correlation-based methods and are twofold. On the one hand, we introduce a modified version of this type of approaches, called TISCORR since it is based on a TIme-Scale signal representation (such representations were also used in a few reported BSS methods, e.g. [4],[8]). On the other hand, we demonstrate that this TISCORR approach (and our previous temporal and time-frequency versions) applies to a wider class of source signals than those which were shown in [7] to be separable by our TEMPCORR and TIFCORR methods.

2 Problem statement

We assume that N unknown, possibly complex-valued, source signals $s_j(t)$ are mixed in a linear instantaneous way, thus providing a set of N 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 $x(t) = [x_1(t) \dots x_N(t)]^T$ and where A is a NxN unknown, supposedly constant and invertible, mixing matrix. Its coefficients a_{ij} may be complex-valued and are assumed to be non-zero hereafter. BSS would ideally consist in deriving an estimate of the matrix A. 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 outputs of BSS systems. We showed in [7] that these indeterminacies make it possible to reformulate the BSS problem as follows. Let us denote

$$s'_j(t) = a_{1,\sigma(j)} s_{\sigma(j)}(t) \tag{2}$$

$$b_{ij} = \frac{a_{i,\sigma(j)}}{a_{1,\sigma(j)}} \tag{3}$$

where $\sigma(.)$ is a permutation, $s'_j(t)$ are the permuted scaled[1] source signals and b_{ij} are the corresponding permuted scaled mixing coefficients. The set of mixing equations (1) may then be rewritten as

$$x(t) = Bs'(t) \tag{4}$$

where $s'(t) = [s'_1(t) \dots s'_N(t)]^T$ and B contains the coefficients b_{ij} (note that, due to (3), the first row of B always consists of 1). Assume that we succeed in deriving an estimate $\hat{B}$ of B. Then, by computing the output vector of a BSS system defined as

$$y'(t) = \hat{B}^{-1} x(t) \tag{5}$$

$$= \hat{B}^{-1} B s'(t) \tag{6}$$

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

[1] This "scaling" 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. The same principle may of course may applied to any other mixed signal instead.

3 Proposed time-scale BSS method

3.1 Time-scale tool

The Time-Scale (TS) representation of the signals[2] considered in this paper is obtained by computing their Continuous Wavelet Transform (CWT) [9],[10]. The wavelets involved in the CWT are defined as shifted and scaled versions of a mother wavelet $\psi(t)$ and read

$$\psi_{\tau,d}(t) = \frac{1}{\sqrt{d}}\psi\left(\frac{t-\tau}{d}\right) \tag{7}$$

where τ is a shift parameter and d is a scale (i.e. *dilation*/contraction) factor. The CWT of a complex-valued signal $v(t)$ is then defined by the inner products of that signal with the wavelets $\psi_{\tau,d}(t)$, i.e.

$$W_v(\tau, d) = \int_{-\infty}^{+\infty} v(t)\psi^*_{\tau,d}(t)dt = \int_{-\infty}^{+\infty} v(t)\frac{1}{\sqrt{d}}\psi^*\left(\frac{t-\tau}{d}\right)dt. \tag{8}$$

This makes it possible to map the considered signal $v(t)$ into the time-scale (TS) plane defined by the parameters τ and d. Each wavelet coefficient $W_v(\tau, d)$ then defines the local behavior of the considered signal $v(t)$ around time τ, at scale d (with an associated frequency proportional to $1/d$).

We here use the dyadic version of this transform, which consists in computing (8) only for scales defined as $d = 2^j$, where j are integers. Moreover, when applying this transform to discrete-time signals, the set of possible values of τ consists of integers. In other words, τ is then varied with a step equal to 1, for each considered value of d.

It should be noted that the CWT is an atom-based, linear, transform. Therefore, it does not introduce interference terms, unlike energy-based, time-frequency or time-scale, transforms. It thus keeps the linear "instantaneous" mixing structure when applied to the observed signals (1) considered in this paper.

3.2 Assumptions and definitions

Each value of a TS transform corresponds to a single "TS point", associated to: i) the selected time position τ and ii) the selected scale d (or the associated frequency). The BSS method that we propose below uses means associated to these TS transforms, computed over "analysis zones" which consist of adjacent TS points. An analysis zone may have any shape in the TS domain. We here focus on the case when it forms a "scale line" (which defines an associated "frequency line"), i.e. when all its points correspond to the same time position τ and to a discrete set of L adjacent scales d_p, with $p = 1 \ldots L$. This set is

[2] As in our TIFCORR method in [7], we here use a deterministic framework, i.e. either the original source signals are deterministic, or they are random processes, but in the latter case we only consider a single realization of these processes (this is what is actually available in practice). The following description then concerns this single, deterministic, realization.

denoted D hereafter. Each analysis zone is then specified in terms of the couple (τ, D), which completely defines the part of the TS domain associated to this analysis zone.

Thanks to the transform applied to the signals and to the shape of the analysis zones, the associated BSS approach exploits the local behavior of the source signals around a given time position, at a set of scales situated on a fixed (i.e. "automated") and geometric grid. This decomposition is e.g. very well-suited to the structure of speech signals. This approach is therefore attractive, e.g. as compared to our previous method based on Short-Time Fourier Transforms (STFTs), which uses a linear frequency grid and requires the user to select the STFT parameters.

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

$$\overline{W}_v(\tau, D) = \frac{1}{L} \sum_{p=1}^{L} W_v(\tau, d_p). \tag{9}$$

Similarly, for any couple of signals $v_1(t)$ and $v_2(t)$, whose TS transforms are denoted $W_{v_1}(\tau, d)$ and $W_{v_2}(\tau, d)$, the cross-correlation of the centered versions of the TS transforms of these signals over the considered analysis zone is measured: i) either by the TS local non-normalized covariance parameter

$$C_{v_1 v_2}(\tau, D) = \frac{1}{L} \sum_{p=1}^{L} [W_{v_1}(\tau, d_p) - \overline{W}_{v_1}(\tau, D)].[W_{v_2}(\tau, d_p) - \overline{W}_{v_2}(\tau, D)]^* \tag{10}$$

or ii) by the corresponding covariance coefficient

$$c_{v_1 v_2}(\tau, D) = \frac{C_{v_1 v_2}(\tau, D)}{\sqrt{C_{v_1 v_1}(\tau, D) C_{v_2 v_2}(\tau, D)}}. \tag{11}$$

For each analysis zone (τ, D), the centered CWT values of any source signal $s_j(t)$ are equal to $W_{s_j}(\tau, d_p) - \overline{W}_{s_j}(\tau, D)$. The vector consisting of these values is denoted $V_{s_j}(\tau, D)$ hereafter.

Definition 1: a source $s_j(t)$ is said to be "isolated" in an analysis zone if only this source is such that its centered TS transform is not equal to zero everywhere in this analysis zone, i.e. if only this source is such that $V_{s_j}(\tau, D) \neq 0$.

Definition 2: a source is said to be "visible" in the TS domain if there exist at least one analysis zone where it is isolated.

Assumption 1: i) each source is visible in the TS domain and ii) there exist no analysis zones where the centered TS transforms of all sources are equal to zero everywhere[3].

Assumption 2: for each analysis zone (τ, D), the non-zero vectors $V_{s_j}(\tau, D)$ are linearly independent (if there exist at least two such vectors in this zone).

[3] *Assumption 1-ii* is only introduced for the clarity of the proof in the Appendix but is not restrictive: in practice, the TS transforms of observed signals containing source and noise contributions are not *strictly* zero.

ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

3.3 The TISCORR method

The BSS method that we introduce in this paper mainly takes advantage of the above *Assumption 1-i*, i.e. of the fact that there exist analysis zones where each source is isolated. These single-source zones should first be detected, so as to operate inside them. As all observed signals have proportional CWTs in any such zone, an appealing approach for detecting these zones consists in checking the TS covariance coefficients $c_{x_1x_i}(\tau, D)$ of the observed signals $x_1(t)$ and $x_i(t)$, defined by (11). More precisely, a necessary and sufficient condition for a source to be isolated in the TS analysis zone (τ, D) is

$$|c_{x_1x_i}(\tau, D)| = 1 \qquad \forall\, i, \quad 2 \leq i \leq N \tag{12}$$

as shown in the Appendix. It should be stressed that the approach used in this appendix is more general than the proof that we provided in [7] for the TEMPCORR and TIFCORR methods, and could also be applied to them in order to extend the class of sources for which those methods are guaranteed to be suitable: whereas we assumed the sources to be uncorrelated in [7], we here only suppose them to be linearly independent (in the TS plane, as defined in *Assumption 2*), thus allowing them to be correlated.

Now consider an analysis zone where a source is isolated, say $s_k(t)$. The observed signals then become restricted to

$$x_i(t) = a_{ik}s_k(t) \qquad i = 1 \ldots N. \tag{13}$$

Again using correlation parameters associated to the CWTs of these observed signals then makes it possible to identify part of the matrix B. More precisely, when (13) is met, one derives easily

$$\frac{C_{x_ix_1}(\tau, D)}{C_{x_1x_1}(\tau, D)} = \frac{a_{ik}}{a_{1k}} \qquad i = 2 \ldots N. \tag{14}$$

The set of values thus obtained for all observations indexed by i identifies one of the columns of B, as shown by (3). By repeatedly performing such column identifications for analysis zones associated to all sources, we eventually identify the overall matrix B, which completes the proposed approach. The structure of the BSS method thus introduced may be summarized as follows.

Step 1 The pre-processing stage consists in deriving the CWTs $W_{x_i}(\tau, d)$ of the observed signals.

Step 2 The detection stage consists in detecting the TS analysis zones where a source is isolated, i.e. where condition (12) is met. To this end, we first consider the analysis zones (τ, D) corresponding to all values of τ and to all half-overlapping sets D. For each such zone, we compute the mean of $|c_{x_1x_i}(\tau, D)|$ over i, with $2 \leq i \leq N$. We then order all analysis zones according to decreasing values of this mean of $|c_{x_1x_i}(\tau, D)|$. The first zones in this list are then considered as the "best" single-source zones.

Step 3 The identification stage consists in identifying the columns of B. This is

wavelet	L		
	4	8	12
Morlet	66.3	71.4	76.2
Gaussian	48.9	69.0	41.8
Gaussian derivative	41.9	56.8	55.3
Mexican hat	50.1	60.1	73.2

Table 1: Output Signal/Interference Ratio, depending on mother wavelet and number L of time-scale points in analysis zones, for uncorrelated speech signals.

achieved by successively using as follows each of the first and subsequent single-source analysis zones in the above ordered list. The correlation parameters on the left-hand side of (14) yield an estimated 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 analysis zone 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).

Step 4 The combination stage consists in recombining the mixed signals according to (5), in order to obtain the extracted source signals.

4 Test results

We first present tests performed with two artificial linear instantaneous mixtures of two real continuous speech signals. These signals correspond to different sentences uttered by different male speakers and are therefore uncorrelated (the temporal sample covariance coefficient of these overall time series is - 0.0017). These signals last 2.5 seconds. They were sampled at 20 kHz and rescaled so that their maximum absolute values are equal to unity. The mixing matrix was

$$A = \begin{bmatrix} 1 & 0.9 \\ 0.8 & 1 \end{bmatrix}. \tag{15}$$

Tests were performed for various mother wavelets, and numbers L of TS points in analysis zones. The CWTs were computed with the Wavelab 802 package available at http://www-stat.stanford.edu/˜ wavelab/ using default parameter values. The resulting output Signal/Interference Ratios (SIRs) are shown in Table 1. They are higher than 40 dB whatever the parameter values of our method. This demonstrates the good separation capability of this approach and shows that it may also be "automated" in the sense that the above parameters do not require user tuning to achieve good performance.

We then tested this method with two correlated sources, created as follows. We used an additional speech signal, from a female speaker. We rescaled it so that its absolute maximum value is equal to 0.2, and we added it to each of the previous two male speech signals. The two signals thus obtained were considered as the sources in this second series of tests and were again mixed according to

wavelet	L		
	4	8	12
Morlet	53.9	43.0	65.8
Gaussian	1.4	55.3	43.9
Gaussian derivative	41.7	55.3	45.9
Mexican hat	0.4	55.4	49.5

Table 2: Output Signal/Interference Ratio, depending on mother wavelet and number L of time-scale points in analysis zones, for correlated speech signals.

(15). Unlike classical BSS approaches, our TISCORR method is supposed to be applicable to these correlated source signals (the temporal sample covariance coefficient of these overall time series is 0.090). However, due to the addition of the female speech signal to each male signal, each source considered here may fill the TS plane to a larger extent than in our first series of tests. This may reduce the amount and quality of single-source analysis zones and may therefore somewhat degrade performance as compared to Table 1. This analysis is confirmed by our test results (see Table 2): the output SIRs achieved here tend to be lower than in the previous series of tests, but they are still higher than 40 dB (except in two cases with $L = 4$, so that such small analysis zones should preferably not be used).

5 Conclusion

In this paper, we introduced a CORRelation-based BSS approach, which operates in the TIme-Scale domain and is therefore called TISCORR. This approach consists in identifying the columns of the (permuted scaled) mixing matrix in TS analysis zones where this method detects that a single source is active. It should be noted that, unlike ICA-based BSS methods, our TISCORR approach is intrinsically well-suited to (realizations of) non-stationary and/or cross-correlated sources and sets no restrictions on their gaussianity. We experimentally showed that it yields good performance for linear instantaneous mixtures of real, correlated or uncorrelated, speech sources. Our future investigations will especially aim at creating more robust versions of this TISCORR approach, by using clustering methods for selecting which columns of the (permuted scaled) mixing matrix are kept, as suggested in [7]. We will also check the performance of a slighly simpler version of this approach, which uses the non-centered version of the TS correlation parameters involved in the detection and identification stages.

A Appendix

Taking the CWTs of the scalar mixing equations in (1) and centering them over the considered analysis zone eventually yields in vector form

$$V_{x_i}(\tau, D) = \sum_{j=1}^{N} a_{ij} V_{s_j}(\tau, D) \qquad i = 1 \ldots N. \tag{16}$$

Besides, the TS covariance coefficients $c_{x_1x_i}(\tau, D)$ of observed signals, defined according to (11), may then be expressed as

$$c_{x_1x_i}(\tau, D) = \frac{< V_{x_1}(\tau, D), V_{x_i}(\tau, D) >}{||V_{x_1}(\tau, D)||.||V_{x_i}(\tau, D)||} \tag{17}$$

where the notations $< .,. >$ and $||.||$ resp. stand for the inner product and vector norm. Applying the Cauchy-Schwarz inequality to (17) then shows that

$$|c_{x_1x_i}(\tau, D)| \leq 1 \qquad \forall\ i, \quad 1 \leq i \leq N \tag{18}$$

with equality if and only if $V_{x_1}(\tau, D)$ and $V_{x_i}(\tau, D)$ are linearly dependent.

Let us now analyze this condition in a given analysis zone (τ, D), depending on the number of non-zero vectors $V_{s_j}(\tau, D)$, i.e. on the number of sources which are active in the considered analysis zone. Due to *Assumption 1-ii*, at least one of these vectors $V_{s_j}(\tau, D)$ is not equal to zero. If only one of them is not equal to zero, since all mixing coefficients a_{ij} are assumed to be non-zero, (16) shows that all vectors $V_{x_i}(\tau, D)$, with $1 \leq i \leq N$, are non-zero and colinear. Therefore, equality holds whatever i in (18) and the detection condition (12) is fulfilled.

The only case that remains to be considered is then the situation when at least two vectors $V_{s_j}(\tau, D)$ are non-zero. It may then be shown easily that if $V_{x_1}(\tau, D)$ and $V_{x_i}(\tau, D)$ were linearly dependent for all i, with $2 \leq i \leq N$, then, due to *Assumption 2*, all the columns of the mixing matrix A with indices equal to the indices j of the non-zero vectors $V_{s_j}(\tau, D)$ would be colinear. This is not true, since A is assumed to be invertible. Therefore, in the considered case, at least one pair of vectors $(V_{x_1}(\tau, D), V_{x_i}(\tau, D))$ does not consist of linearly dependent vectors, so that $|c_{x_1x_i}(\tau, D)| < 1$ and the detection condition (12) is not fulfilled.

As an overall result, condition (12) is fulfilled if and only if exactly one of the vectors $V_{s_j}(\tau, D)$ is not equal to zero in the considered analysis zone, i.e. if a source is isolated in that zone, which yields the detection criterion of the proposed BSS method.

References

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

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

[3] O. Yilmaz, S. Rickard, Blind separation of speech mixtures via time-frequency masking, *IEEE Transactions on Signal Processing*, 52:1830-1847, 2004.

[4] M. Zibulevsky, B.A. Pearlmutter, Blind source separation by sparse decomposition in a signal dictionary, *Neural Computation*, vol. 13, no. 4, 2001.

[5] A. Belouchrani, M.G. Amin, Blind source separation based on time-frequency signal representations, *IEEE Transactions on Signal Processing*, 46:2888-2897, 1998.

[6] F. Abrard, Y. Deville, A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources, *Signal Processing*, 85:1389-1403, Elsevier, 2005.

[7] Y. Deville, Temporal and time-frequency correlation-based blind source separation methods, *proceedings of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation* (ICA2003), pages 1059-1064, April 1-4, Nara (Japan), 2003.

[8] M.G. Jafari, J.A. Chambers, Time-scale approach to blind source separation, *proceedings of the* 11^{th} *European Signal Processing Conference* (EUSIPCO 2002), pages 569-572, Sept 3-6, Toulouse (France), 2002.

[9] S. Mallat. *A wavelet tour of signal processing*, Academic Press, San Diego, 1999.

[10] O. Rioul, M. Vetterli, Wavelets and signal processing, *IEEE Signal Processing Magazine*, pages 14-38, Oct. 1991.