SOURCES SEPARATION OF INSTANTANEOUS MIXTURES USING A LINEAR TIME-FREQUENCY REPRESENTATION AND VECTORS CLUSTERING

Braham Barkat[1], *Farook Sattar*[2], *and Karim Abed-Meraim*[3]

[1]Electrical Engineering Program, Academic Directorate, The Petroleum Institute,
PO Box 2533, Abu Dhabi, U.A.E., Email: bbarkat@pi.ac.ae
[2]School of Electrical& Electronic Engineering, Nanyang Technological University,
Nanyang Avenue, Singapore 639798, Email:efsattar@ntu.edu.sg
[3]Telecom-Paris, Signal and Image Processing Department, 46 Rue Barrault,
75634 Paris, France, Email:abed@tsi.enst.fr

ABSTRACT

In this paper, we address the problem of separating N unknown sources using as many observed mixtures. The sources considered here are assumed to be of a non-stationary nature, i.e., their spectral contents are assumed to be time-varying. Using linear time-frequency (TF) representations of the mixtures along with a classification procedure based on vector clustering yield an effective way to separate the sources. Compared to other existing TF based separation methods, the proposed one is characterized by its simplicity and ease of implementation. Moreover, it can be applied in situations where others cannot. Specifically, the algorithm can handle monocomponent as well as multicomponent sources and its assumptions about the mixing matrix are more relaxed than other existing algorithms. Example is presented to prove the validity and efficiency of the proposed algorithms.

1. INTRODUCTION

Blind source separation (BSS) deals with the problem of recovering unknown signals from several observed mixtures. Usually, these mixtures are obtained as the output of a set of sensors, whereby, each sensor collects a different combination of the source signals. The separation procedure is referred to as blind because, at the sensors, the source signals are not known and there is no information about the mixing process [1].

Various methods have been introduced for BSS. These methods use different approaches, such as the probabilistic approach, the spectral/time-coherence approach and the time-frequency (TF) approach [2, 6]. The TF based approach is the most effective in dealing with nonstationary source signals, i.e., signals whose spectral contents vary with time.

In the literature, we find many BSS studies based on TF analysis. A pioneering TF based method is reported in [3]. The technique developed in [3] introduces the so-called space time-frequency distribution (STFD). Essentially, the STFD based algorithm uses a whitening procedure of the STFD matrix followed by a joint block diagonalization procedure of spatial quadratic TF matrices evaluated at some particularly well selected TF points. To ensure the diagonalization structure of the STFD, which is necessary for the block diagonalization procedure, each of the selected TF points must belong to an auto-term of one of the sources only. This condition was later relaxed to allow utilization of TF points from both auto-term as well as cross-term regions [4]. In the STFD based algorithm, one can use any TF representation. However, the best results are obtained for a reduced or cross-terms free TF representation, such as the short-time Fourier transform (STFT). We observe that the STFD based algorithm

- is a block-based technique and, consequently, necessitates somewhat sophisticated and expensive processing
- requires full knowledge of the auto-terms and cross-terms regions
- requires the source signals to be of a monocomponent nature
- needs the user to attribute a particular TF point to its corresponding source.

To avoid the above limitations, we propose in this paper a novel and simpler TF based method. The proposed method can use an arbitrary linear TF representation. Below, we choose the STFT to be such a TF tool. The use of a linear TF representation stems from the fact that it yields a simple mathematical relationship between the mixtures vector and the sources vector in the TF domain. In addition, it completely alleviates the problem of the usually undesirable cross-terms. Contrary to the algorithm in [3], the proposed algorithm automatically attributes a given TF point to its corresponding source. This is performed by applying either of two proposed classification schemes. The elements of a

1-4244-0469-X/06/$20.00 ©2006 IEEE

particular class are vector estimates of a particular vector column of the mixing matrix. Subsequently, the mean or the centroid of the vectors in a given class are used as an estimate of a particular vector column of the mixing matrix. Once the sources are separated in the TF domain, we use an inverse TF transformation to recover each of these sources in the time-domain.

2. PROBLEM STATEMENT

In this section, we assume the existence of N independent source signals $s_1(t), \ldots, s_N(t)$ and the observation of as many mixtures $x_1(t), \ldots, x_N(t)$. The mixtures are assumed linear and instantaneous, i.e., $x_i(t) = \sum_{n=1}^{N} a_{in} s_n(t)$ for $i = 1, \ldots, N$. In matrix form, the considered BSS model can be written as

$$\mathbf{x}(t) = A\,\mathbf{s}(t) \tag{1}$$

where $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^T$ represents the unknown nonstationary sources, $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ represents the mixtures, and A represents the $N \times N$ unknown mixing matrix. In the sequel, we assume the mixing matrix entries to be arbitrary and real, and its columns to be linearly independent. Because of the inherent ambiguities in the BSS problem, the sources separation is only possible up to an unknown scaling and an unknown permutation [3]. That is, the estimated signals may not be recovered in an orderly manner and their amplitudes are multiplied by some constant scalars.

3. PROPOSED BSS METHOD

In this section, we present the theoretical derivations of the proposed technique, some of its implementational aspects and an example to prove its validity.

3.1. Derivations

For simplicity, let us consider the noise-free model given by Eq. (1), namely,

$$\underbrace{\begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix}}_{\mathbf{x}(t)} = \underbrace{\begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N,1} & a_{N,2} & \ldots & a_{N,N} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} s_1(t) \\ \vdots \\ s_N(t) \end{bmatrix}}_{\mathbf{s}(t)} \tag{2}$$

where $\mathbf{x}(t)$, A and $\mathbf{s}(t)$ represent the same quantities defined earlier. Now, let us take the STFT of each mixture $x_i(t)$, $i = 1, \ldots, N$. This operation yields the following result

$$\begin{aligned} X_1(t,f) &= [a_{1,1}\ a_{1,2}\ \ldots\ a_{1,N}] \begin{bmatrix} S_1(t,f) \\ \vdots \\ S_N(t,f) \end{bmatrix} \\ &\vdots \\ X_N(t,f) &= [a_{N,1}\ a_{N,2}\ \ldots\ a_{N,N}] \begin{bmatrix} S_1(t,f) \\ \vdots \\ S_N(t,f) \end{bmatrix} \end{aligned} \tag{3}$$

where $S_i(t,f)$, $i = 1, \ldots, N$ is the STFT of the corresponding source signal $s_i(t)$, $i = 1, \ldots, N$. In a more compact form, the above result can be re-written as

$$\begin{bmatrix} X_1(t,f) \\ \vdots \\ X_N(t,f) \end{bmatrix} = \begin{bmatrix} a_{1,1}S_1(t,f) + \ldots + a_{1,N}S_N(t,f) \\ \vdots \qquad\qquad \vdots \\ a_{N,1}S_1(t,f) + \ldots + a_{N,N}S_N(t,f) \end{bmatrix} \tag{4}$$

Therefore, for an arbitrary TF point, say (t_1, f_1), where only the source signal $s_i(t)$ exists, the result in (4) reduces to

$$\begin{bmatrix} X_1(t_1,f_1) \\ \vdots \\ X_N(t_1,f_1) \end{bmatrix} = S_i(t_1,f_1) \begin{bmatrix} a_{1,i} \\ \vdots \\ a_{N,i} \end{bmatrix} \tag{5}$$

which is just a complex scalar value $S_i(t_1, f_1)$ multiplied by the i^{th} column vector of the mixing matrix A.

3.2. Classification

The previous result indicates that if we can select N points in the TF plane such that each point belongs to only one source, then, we can estimate all the N column vectors of the mixing matrix A. In what follows, we will present a classification method to automatically select such TF points. Moreover, this classification method does not select only one TF point for a given source but will select a set of points for each source. This, in turn, will yield better estimates of the colum vectors of the matrix A.

Now, how to decide that two arbitrary TF points belong to the same source or not? To answer this, let us consider two different TF points (t_1, f_1) and (t_2, f_2). If these two points belong to the same source, say $s_i(t)$, we can write

$$X(t_1,f_1) = \begin{bmatrix} X_1(t_1,f_1) \\ \vdots \\ X_N(t_1,f_1) \end{bmatrix} = S_i(t_1,f_1) \begin{bmatrix} a_{1,i} \\ \vdots \\ a_{N,i} \end{bmatrix}$$

and

$$X(t_2,f_2) = \begin{bmatrix} X_1(t_2,f_2) \\ \vdots \\ X_N(t_2,f_2) \end{bmatrix} = S_i(t_2,f_2) \begin{bmatrix} a_{1,i} \\ \vdots \\ a_{N,i} \end{bmatrix}.$$

This implies that the real (or imaginary) parts of the vectors $X(t_1, f_1)$ and $X(t_2, f_2)$ must be co-linear. Thus, we attribute a set of TF points to a particular class if their corresponding mixture vectors $X(t, f)$ have co-linear real (or imaginary) parts. Based on this, the general steps of the proposed algorithm are stated in Table I.

If the sources overlap, or if there is noise, in the TF domain we may have vectors $X(t, f)$ whose real (or imaginary) parts are not co-linear to any of the vectors of the N classes discussed above. Thus, these vectors cannot be classified in any of the N classes associated with the sources.

1. Evaluate the STFT, $X_i(t, f)$, of each mixture signal $x_i(t),\ i = 1, \ldots, N$.
2. For each TF point, (t, f), form the vector $X(t, f) = [X_1(t, f), X_2(t, f), \ldots, X_N(t, f)]^T$.
3. Classify these vectors into N classes using the co-linearity rule explained earlier.
4. For each class, use its vectors mean as an estimate of a column vector of the mixing matrix A.
5. Invert the estimate of the matrix A and multiply it by the mixtures vector $\mathbf{x}(t)$ to obtain estimates of the original sources $\mathbf{s}(t)$.

Table. I: Basic BSS algorithm for no TF overlap of the sources.

Consequently, the classification procedure will result in more than the N classes associated to the N sources. Therefore, in the classification procedure the initial number of classes is chosen equal to L where $L > N$. The initial number of classes L can be chosen in many ways. A simple way is to choose L equal to the number of TF points in the STFT. That is, in the implementation, we start by assuming that we have as many classes as there are vectors $X(t, f)$. Then, using the co-linearity rule, this number is decreased each time two vectors are found to be co-linear. Obviously, going through all the TF points of the STFT might be computationally demanding. To avoid this, a better alternative is proposed below.

First, let us consider that in the TF domain the signals are characterized by high peaks around their instantaneous frequencies. Second, we observe that in the classification procedure there is no need to consider $X(t, f)$ for all TF points but only for some points where the sources exist. Therefore, selecting only the highest peaks of the TF representation will certainly result in TF points belonging only to the sources or their possible overlaps. A possible way to select the highest peaks of the TF representation is to select the peaks of each of its slices. In the procedure below, we choose to select only N peaks from each slice of the TF representation. This is because, at most, we can have N sources for each time instant t. In this way, the initial number of classes used in the classification reduces to only $L = N \cdot T$ compared to $L = F \cdot T$, the total number of points in the TF matrix (T and F represent the number of discrete-time and discrete-frequency bins used in the implementation of the STFT, respectively).

Thus, we state in Table II an improved version of the algorithm given in Table I. As mentioned earlier, the above algorithm is different from those proposed in [5]. The difference is not only in the way the sources are classified and separated in the TF domain but also in the assumptions made in the respective methods. In [5], all the coefficients of the mixing matrix must be of the same sign, must be strictly different from zero and all their pairwise ratios different (refer to (22-24)). In our proposed method, we do not require such limitations. In fact, as shown in the coming examples, the coefficients can be zero and of arbitrary signs. However, our proposed method assumes the vector columns of the mixing matrix to be linearly independent and the majority of the selected TF points (collected in Steps 2 & 3) to belong to the individual sources.

1. Evaluate the STFT, $X_i(t, f)$, of each mixture signal $x_i(t),\ i = 1, \ldots, N$ as well as the matrix $C(t, f) = \sum_{i=1}^{N} |X_i(t, f)|$.
2. For the starting time instant t_1, find the frequencies corresponding to N highest peaks of the slice $C(t_1, f)$. Save these TF points.
3. Repeat the above step for all the other time instants.
4. Evaluate $X(t, f) = [X_1(t, f), \ldots, X_N(t, f)]^T$ for all the L time-frequency pairs (t, f) collected in the previous two steps.
5. Classify these L vectors into classes using the co-linearity rule explained earlier. Keep only the N largest classes (in terms of number of vectors in them).
6. For each class, use its vectors mean as an estimate of a column vector of the mixing matrix A.
7. Invert the estimate of the matrix A and multiply it by the mixtures vector $\mathbf{x}(t)$ to obtain estimates of the original sources $\mathbf{s}(t)$.

Table. II: Improved version of the algorithm given in Table I.

Note that $C(t, f) = \sum_{i=1}^{N} |X_i(t, f)|$ used in Step 1 of the improved algorithm has been used only to localize the peaks of the sources in the TF domain, consequently, other reduced interference TF representations can also be used instead. However, once the peaks (or their corresponding TF points) have been selected, it is the STFT that we use in order to build the vectors $X(t, f)$.

3.3. Example

To prove the efficiency of the proposed algorithm, let us consider the following example. In this example, we consider three sources, which are highly non-stationary. One of these sources is chosen to be a sinusoidally frequency modulated(FM) signal. The second source is chosen to be a multicomponent signal consisting of two linearly FM components, one of an increasing nature and the other of a deceasing nature. The third source is also a multicomponent signal consisting of a quadratically FM component and a sinusoid. The sources are mixed using A given by $A = [0.9\ 0.4\ 0.1; 0.0\ 0.9\ 0.3; -0.4\ 0.0\ 0.8]$. Then, white Gaussian noise is added to each mixture signal. The signal-to-noise ratio (SNR), evaluated as the power of the noise over the power of the weakest source, is chosen equal to 10 dB. The TF representation $C(t, f)$ is displayed in Figure 1 (top left plot) along with the successfully separated sources. As we can see, the proposed algorithm performs an efficient separation procedure.

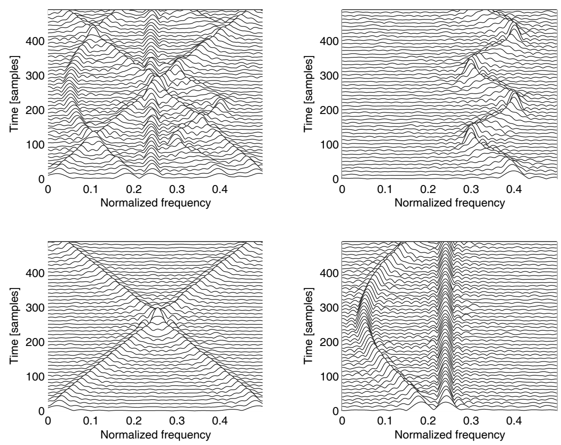

Fig. 1. The TF representation $C(t, f)$ (top left plot) along with the separated sources (remaining plots).

Note that even-though the mixing matrix entries are chosen to be of different signs and some of them have zero value, the proposed algorithm is able to separate the sources. However, the algorithms in [5] cannot be applied in both examples because their assumptions (concerning A) are violated.

3.4. Alternative Classification

We propose to use an alternative classification procedure based on a statistical optimization when there is a significant amount of TF overlap between the sources. In this procedure, known as vectors clustering [7, Chap. 6], the real parts of the L vectors $X(t, f)$ (obtained in Step 4, Table II) are considered to be spatial points in a multi-dimensional space. Starting from an initial set of N points (called centroids) arbitrarily chosen among the L ones, the classification procedure tries to statistically classify the real-parts of the selected L vectors $X(t, f)$, based on their distances to the centroids, into N classes (called clusters). The algorithm, then, updates its centroids and re-evaluates the distances to yield a new set of clusters. This adaptive procedure stops when it finds the optimal clusters. The optimality is in terms of minimizing the sum, over all clusters, of the within-cluster sums of point-to-centroid distances. In this classification, and without loss of generality, we need to set the norms of the real-parts of the selected vectors to unity.

Note that the proposed TF based BSS algorithm when implemented using the improved classification procedure, discussed above, will have exactly the same steps as in Table II except for Step 5, which now reads as

5	Classify the L vectors into N classes using vector clustering method.

4. CONCLUSION

In this paper, we studied the problem of separating N unknown non-stationary source signals using N observed mixtures. For that, we proposed a simple and efficient algorithm based on the linear TF representations of the mixtures and vectors classification. Two different classification procedures were presented for the sources having less or drastic TF overlap, respectivly. In comparison to other existing TF based separation methods, we showed that the proposed ones are characterized by their simplicity and ease of implementation. We also showed that the proposed algorithms can handle monocomponent as well as multicomponent sources and its assumptions about the mixing matrix are more relaxed than other existing algorithms. Ilustrative results were presented to prove the validity and efficiency of the proposed algorithms. Note that a modified version of the proposed algorithm for the under-determined case (when the number of sources is larger than the number of sensors) as well as illustrative results for real data (such as, mixture of speech and music) are not included here due to space constraint.

5. REFERENCES

[1] J.F. Cardoso, "Blind source separation: statistical principles," *Proc. of the IEEE*, 86(10):2009–2025, Oct. 1998.

[2] N.L. Trung, A. Belouchrani, K. Abed-Meraim, and B. Boashash, "Separating more sources than sensors using time-frequency distributions, *EURASIP Journal on Applied Signal Processing*, 2005(17):2818–2847, Sept. 2005.

[3] A. Belouchrani and M.G. Amin,"Blind source separation based on time-frequency representations,"*IEEE Trans. on Signal Process.*,46(11):2888–2897, Nov. 1997.

[4] A. Belouchrani, K. Abed-Meraim, M.G. Amin, and A.M. Zoubir, "Blind separation of nonstationary sources," *IEEE Signal Process. Letters*, Vol. 11:605–608, July 2004.

[5] ö. Yilmaz and S. Rickard,"Blind separation of speech mixtures via time-frequency masking, *IEEE Trans. on Signal Process.*, 52(7):1830–1847, July 2004.

[6] Y. Li, A. Cichocki, and S. Amari,"Analysis of sparse representation and blind source separation,"*Neural Computation*, 16(6):1193-1234, June 2004.

[7] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*, Wiley Inter-Science, 1998.