ON THE USE OF TIME-FREQUENCY REPRESENTATION IN MULTICOMPONENT SIGNAL SEPARATION

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 unknown multicomponent signals from their instantaneous mixtures. Using linear time-frequency (TF) representation of the mixtures along with vectors classification scheme provide us a simple and efficient technique to separate multicomponent signals. The proposed algorithm can handle monocomponent as well as multicomponent sources and its assumptions about the mixing matrix are more relaxed compared to other existing TF based algorithms. The source separation results for the mixed synthetic signals as well as mixed real audio signals, such as mixture of speech and music, are shown to illustrate the validity and efficiency of the proposed scheme.

1. INTRODUCTION

Source separation deals with the problem of recovering unknown signals from several observed mixtures using a number of sensors. This source separation has to be done considering that the source signals as well as their mixing process are not known. Source separation has various applications, including the separation of individual speech signals from a mixture of simultaneous speakers, the elimination of cross-talk between horizontally and vertically polarized microwave communications transmissions, and the separation of multiple telephone signals at a base station. Other applications include radar, acoustics and biomedical engineering [1].

A number of techniques have been introduced for source separation, such as the probabilistic technique, the spectral/time-coherence technique and the time-frequency (TF) technique [1, 2]. The TF based technique is very effective in dealing with the separation of multicomponent source signals, which are usually non-stationary, i.e., the spectral contents of the signals vary with time. This is because TF analysis is an ideal tool for the analysis of such signals. In addition to the spatial diversity that other methods exploit, TF based techniques use the joint time-frequency characteristics of the sources in order to separate them.

The early TF based source separation method is presented in [3], which has used the so-called space time-frequency distribution (STFD). This method is based on the block diagonalization procedure of the STFD, which is necessarily a sophisticated and expensive processing and requires full knowledge of the auto-terms and cross-terms regions as well as the source signals to be of monocomponent nature. Moeover, it requires the selected TF points belong to auto-terms of one of the sources only. Later on this STFD technique has been improved by allowing the utilization of TF points from both auto-term as well as cross-term regions [4]. The proposed linear TF based method does not have the limitations as shown in [3, 4].

The best separation results can be obtained for a reduced or cross-terms free TF representation, such as the short-time Fourier transform (STFT). That is why we choose the STFT as a linear TF tool. Among the other STFT based source separation methods which have appeared, we can cite [5] and [6]. The method in [5] is based on the ratio of the STFTs of the mixtures at every non-zero TF point and is used to separate/unmix two speech signals. In fact, this method requires some conditions to meet, i.e., all the coefficients of the mixing matrix must be of the same sign, must be strictly different from zero and all their pairwise ratios are to be different. In [6], the authors also use the ratio of the STFTs of the mixtures at every non-zero TF point. However, their classification procedure is based on the variance and the mean of the ratio evaluated for a set of TF points. In [6], it is assumed that all the mixing matrix coefficients must be non-zero and if there is an overlap of the sources in adjacent TF windows, they should vary such that the ratios of their STFTs do not take the same value in all these windows. In this algorithm, a thresholding procedure is also necessary to segregate two different sources or matrix components. Our proposed method does not require such limitations and is different in the way the sources are separated. Contrary to the above methods, the proposed method automatically attributes a given TF point to its corresponding source by a applying classification scheme. This will be well detailed in the sections developed later.

The paper is organized as follows. In Section 2, we state the problem. In Section 3, we present the proposed algorithm along with various examples to validate the proposed source separation method. Section 4 concludes the paper.

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 source separation model can be written as

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

where $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^T$ represents the unknown sources, $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ represents the mixtures, and A repre-

1-4244-0367-7/06/$20.00 ©2006 IEEE

sents 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 this blind source separation problem, the source 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 SOURCE SEPARATION METHOD

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

3.1. Derivations

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

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

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

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

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

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

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.

If the sources overlap, or if there is noise, in the TF plane 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. 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).

We present the algorithm in Table I. As mentioned earlier, the above algorithm is different from those proposed in [5, 6]. The difference is not only in the way the sources are classified and separated

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 colinearity 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. I: The proposed algorithm.

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 must be different. In [6], it is assumed that all the mixing matrix coefficients must be non-zero and if there is an overlap of the sources in adjacent TF windows, they should vary such that the ratios of their STFTs do not take the same value in all these windows. In our proposed method, we do not require such limitations. In fact, as shown in the following example, the coefficients can be 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) belong to the individual sources.

Note that $C(t,f) = \sum_{i=1}^{N} |X_i(t,f)|$ used in Step 1 of our 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 generate the vectors $X(t,f)$.

To show the efficiency of the proposed algorithm, let us consider the following illustration. In this example, we consider two sources which are highly non-stationary. The first source consists of a real-life signal emitted by a bat. The other source consists of a decreasing linear FM signal. The mixing matrix chosen here is given by $A = [-1.2\ 0.4; 1.2\ 0.3]$ and the SNR is fixed to 10 dB. Observe that the linear FM source overlaps with the second component of the bat signal. Using the proposed algorithm, the two sources are successfully separated, as shown in Figure 1. Once again, we observe the efficiency of the proposed algorithm in separating and extracting the sources. For a closer comparison, we display in Figure 2, the normalized amplitude of the original (top plot) and extracted (bottom plot) bat signals, respectively.

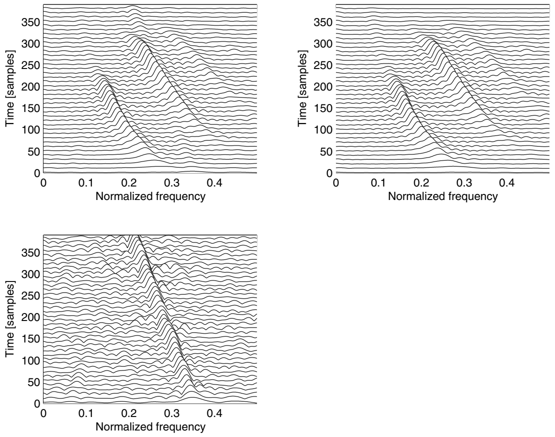

Fig. 1. The TF representation $C(t,f)$ (top left plot) of a mixture consisting of two sources: a multicomponent real-life bat signal and a linear FM. The remaining plots display the TFDs of the separated sources.

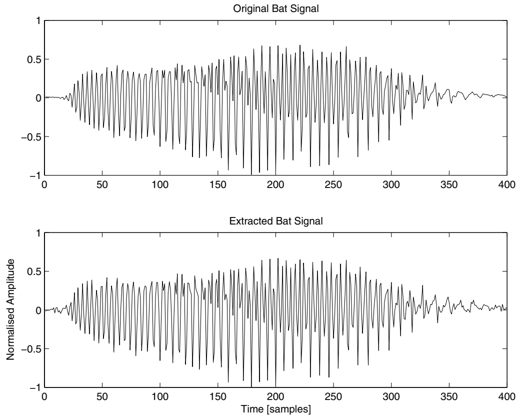

Fig. 2. Normalized amplitudes of the original (top plot) and extracted (bottom plot) real-life bat signal.

Note that even-though the mixing matrix entries are chosen to be of different signs, the proposed algorithm is able to separate the sources. However, the algorithm in [5] cannot be applied in this example because the assumptions (concerning A) are violated.

3.3. Alternative Classification Based on Vectors Clustering

Another classification approach based on a statistical optimization is proposed here when there is a significant amount of TF overlap between the sources. In this procedure, which is known as vectors clustering [7, Chap. 6], the real parts of the L vectors $X(t,f)$ (obtained in Step 4, Table I) are considered to be spatial points in a multidimensional space. Starting from an initial set of N points (called centroids) arbitrarily chosen among the L ones, this classification scheme 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. A similar classification procedure, along with a wavelet packets transformation, has been used in [2] in the context of analysis of sparse representation.

It is noted that the proposed TF based source separation algorithm when implemented using this effective classification scheme as discussed above, can have the same steps as in Table I except for Step 5, which now includes as

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

To illustrate the effectiveness of this classification scheme, let us consider two sources consisting of a speech signal and a music signal mixed together using $A = [-0.5\ 0.4; 0.5\ -0.8]$. These two sources drastically overlap in the time-frequency domain, as shown in Figure 3.

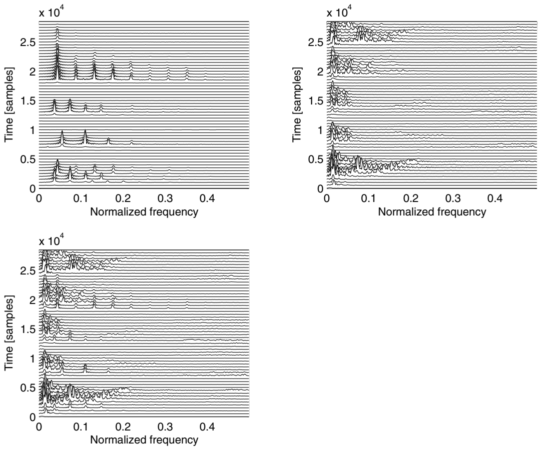

Fig. 3. TF representations of a music signal (top left plot), a speech signal (top right plot) and one of their mixtures (bottom plot).

Some white Gaussian noise is added to each mixture (SNR is fixed to 25 dB). For a comparison purpose, we use both classifications to separate the sources. It turns out that the previous classification based on the largest number of vectors in the classes cannot separate the signals even at large SNR. However, using this clustering based classification, we are able to separate the signals, as shown in Figure 4. (see at http://www.ntu.edu.sg/home/efsattar/web/bsstf.htm for listening test).

4. CONCLUSION

In this paper, we have studied the problem of separating the unknown multicomponent source signals from their observed mixtures. Therefore, we proposed a simple and effective algorithms based on the linear TF representations of the mixtures and vectors classification schemes. Two different classification schemes have been presented. The first one can be applied when the sources do not drastically overlap in the TF domain; whereas, the second one is more robust to the

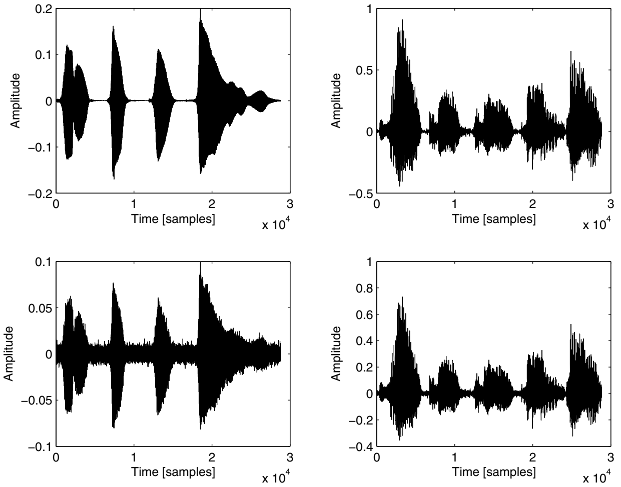

Fig. 4. The original music (top left plot) and speech signals (top right plot) and their respective estimates (bottom) for an SNR equal to 25 dB.

sources TF overlap, provided there exist regions in the TF domain where only one source occurs. In comparison to other existing TF based source separation methods, we have shown that the proposed scheme is characterized by their simplicity and ease of implementation. It is also shown that, for the proposed algorithms, the assumptions about the mixing matrix are more relaxed than other existing algorithms. Ilustrative examples, using synthetic as well as real-life data, are presented to prove the validity and efficiency of the proposed schemes. In the full version of the paper, the modified version of this TF based method for the under-determined case (when the number of sources is larger than the number of sensors) together with more illustrative results, will be presented.

5. REFERENCES

[1] 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.

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

[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] F. Abrard and Y. Deville, "Blind separation of dependent sources using the time-frequency ratio of mixtues approach," *Proc. Int. Symp. Signal Process. & Its Appl.*, pages 81–84, 1-4 July 2003.

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