WAVELET BASED INDEPENDENT COMPONENT ANALYSIS FOR MULTI-CHANNEL SOURCE SEPARATION

Rachid Moussaoui, Jean Rouat, Roch Lefebvre

Département de génie électrique et de génie informatique
Université de Sherbrooke, Québec, Canada

ABSTRACT

We consider the problem of separating instantaneous mixtures of different sound sources in multi-channel audio signals. Several methods have been developed to solve this problem. Independent component analysis (ICA) is certainly the most known method and the most used. ICA exploits the non-Gaussianity of the sources in the mixtures. In this study, we propose an improved signal separation algorithm where simultaneously we increase the non-Gaussian nature of signals and we initiate the preliminary separation. For this, the observations are transformed into an adequate representation using the wavelet packets decomposition. In this study, we consider the instantaneous mixture of two sources using two sensors. We validate our approach by using synthetic and recorded audio signals. Preliminary results show a strong improvement when compared to conventional ICA (FastICA), with specific signals.

1. INTRODUCTION

Source separation is an interesting problem in signal processing. Its objective is to extract meaningful signals from a mixture of many signals with minimum a priori information on the mixture process. In the instantaneous mixture case, the source separation problem can be solved by many approaches using the ICA algorithm. One approach estimates the unmixing matrix by minimizing the mutual information between the separated sources [1]. Others exploit the non-Gaussianity of the source signals and perform separation by maximizing this non-Gaussianity [2]. For example Tanaka *et al.* [3] have developed a method using a subband decomposition in combination with ICA. In [4], Kisilev *et al* have used geometric algorithms to separate the mixed signals. In this paper, we propose an algorithm which uses the idea of applying a preprocessing in the transformed domain but the separation is performed in the time domain.

1.1. Restrictions of ICA

The standard formulation of ICA requires at least as many sensors as sources. Thus, in this paper, we assume that the number of sources is equal to the number of sensors and we ignore additive noise. In the instantaneous mixture case, the sources are not observed directly but as a linear combination such that :

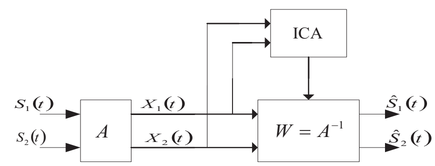

Fig. 1. *Principle of ICA*

$$x_i(t) = \sum_{j=1}^{N} a_{ij} s_j(t), \tag{1}$$

where s are source signals, x are observed signals and $A = [a_{i,j}]$ is an unknown full rank mixing matrix.

Figure 1 shows the principle of ICA. In practice, the goal of ICA is to find the inverse of A, which is the unmixing matrix $W = A^{-1}$. To estimate W, we have to make certain assumptions and impose some restrictions [2] : the individual components $s_i(t)$ are assumed to be statistically independent over the observation time and the individual components must have non-Gaussian distributions. In comparison to previous work, the novelty of our approach resides in the preprocessing implementation before the source separation process in to : i) relax the previous restrictions by increasing the non-Gaussianity that is a pre-requirement for ICA, ii) initiate a preliminary separation by decreasing the mutual information between the resultant signals from the preprocessing.

The preprocessing transforms the observed signals to find an adequate representation where the signals distributions are non-Gaussian. For this, the wavelet transform is used to emphasize the non-Gaussian nature of the observed signals. Once the inverse matrix W is found with the wavelet packets based ICA then, the separation is performed in the time domain.

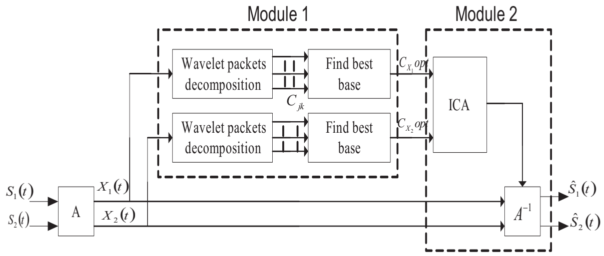

Fig. 2. *Overview of the proposed system*

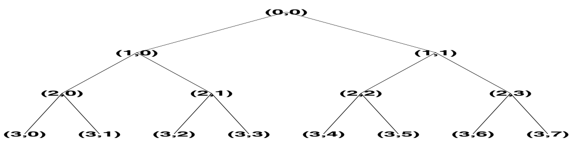

Fig. 3. *The tree of wavelet packets decomposition. Each doublet (j,k) represents the depth (j) and the number (k) of the nodes*

Section 2 describes the proposed method. Results are given in section 3. Finally, section 4 is the conclusion.

2. PROPOSED METHOD

The proposed system is presented in Figure 2. It comprises two modules shown in dotted boxes. The first module (preprocessing) extracts appropriate signals from the observed signals to facilitate the source separation. For this, the observed signals are projected on suitable bases, more precisely on one of the wavelet packets bases. The second module (separation) performs the source separation using standard ICA [1]. The input of this module is the extracted signals from module 1 and the observed signals.

2.1. Preprocessing module

2.1.1. Description

In module 1 (figure 2) the observed signals are decomposed with a generalization of Mallat's algorithm [5]. This algorithm is known as the wavelet packets decomposition. It offers interesting properties : i) the wavelet packets coefficients emphasize the non-Gaussian distribution, ii) the coefficients produced by the wavelet packets decomposition are sparse, iii) the dependencies between the wavelet coefficients of different resolutions are very weak. We use Daubechies wavelets with a tree of wavelet packets decomposition as shown in figure 3. In this case, the wavelet decomposition level is equal to 3.

The wavelet packets decomposition of the observed signals leads to many possible choices of bases (see figure 3). Each base can be considered as an appropriate signal. Thus, it is necessary to find the best basis. The notion of the best basis of wavelet packets was introduced by Coifman *et al.* [6]. The basis selection can be done by using Shannon entropy criterion. In our case, we select only one node (instead complete basis) for which the information is concentrated. The selection of this node is done as follows:

i) decompose the observed signals into wavelet packets;

ii) compute the entropy value at each node;

iii) select the node that has the lowest entropy.

The Shannon entropy is defined for each node (j,k) as :

$$H(j,k) = -\sum p_i \log(p_i), \qquad (2)$$

with $p_i = \frac{||C_{j,k}(i)||^2}{||x||^2}$, where $C_{j,k}$ are the wavelet coefficients and x is the observed signal.

2.1.2. Impact of preprocessing on non-Gaussianity

From the wavelet packets decomposition properties, only a few wavelet coefficients of the selected node are significant. Thus, the histogram of these coefficients will have a large peak around zero. Figure 4 shows a typical histogram example of the observed signal and its wavelet packets coefficients. The observed signal is a mixture of two distinct sources (male speech and music) with the mixing matrix $A = [2\ 1\,;1\ 1]$. We note in figure 4 that the observed signal distribution is more Gaussian than that of the coefficients.

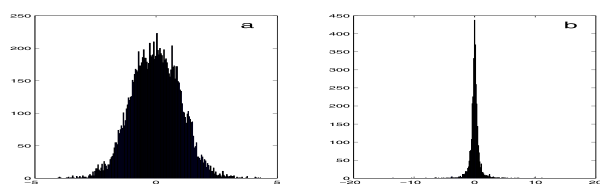

Fig. 4. *Normalized histogram of (a) observed signal* ($normalizedKurtosis = 0.22$)*, and (b) wavelet coefficients from the selected node* ($normalizedKurtosis = 22.85$)

To find and quantify the non-Gaussian signature present in the signals, one can use statistical tests. We use the normalized fourth order moment (Kurtosis) of the distribution that equals zero for a Gaussian distribution. The normalized Kurtosis value of the wavelet coefficients is significantly larger than the normalized Kurtosis of the observed signals (figure 4). Thus, the coefficients issued from the wavelet packets decomposition decrease the Gaussian nature of the distributions. Consequently, during the application of ICA, we will have a significant gain. Let us recall that one of the ICA approaches is the exploitation of non-Gaussianity.

2.1.3. *Impact of preprocessing on mutual information*

The minimization of mutual information is used as a criterion to estimate the sources in some ICA algorithms. Thus, with the preprocessing module, we initiate the preliminary separation because the preprocessing minimizes the mutual information between the outputs. To confirm this, we conducted some experiments with different sources and different mixing matrices. In the case of two sources and two sensors, we computed the mutual information between the signals of the two channels. The mutual information value between the two channels strongly decreases after the preprocessing module compared with the one before preprocessing. For example, if the observed signals are issued from two sources (male speech and music) with the mixing matrix $A = [2\ 1\,;1\ 1]$, the mutual information between the sources (s_1, s_2), the observed signals (x_1, x_2) and the selected wavelet packets coefficients (c_{x_1}, c_{x_2}) is respectively 0.01, 0.65 and 0.12.

A more obvious illustration of the proposed preprocessing advantage is shown in figure 5. This Figure shows the joint distribution of the observed mixtures (x_1, x_2) and the joint distribution of the wavelet coefficients issued from the selected node. The observed signals are generated by two distinct sources (speech signal and music signal). We can see on the wavelet coefficients joint distribution (figure 5-b) the main lines which can be regarded as a representation of each source separately. This is in clear contrast with the joint distribution of the observed signals (figure 5-a).

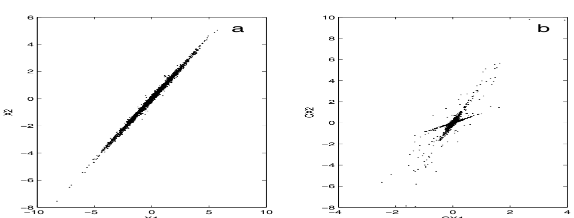

Fig. 5. *The joint distribution of (a) observed signals, and (b) wavelet coefficients from the selected node*

2.2. Separation Module

Module 2 (figure 2) consists in determining the mixing matrix. The aim is to obtain the inverse of the mixing matrix A^{-1} using the independent component analysis (ICA). The ICA algorithm inputs are the wavelet coefficients from the node that minimizing the entropy. Thus, we replace the mixtures x_1 and x_2 by the wavelet coefficients. The second module output is the inverse mixing matrix. Finally the estimated sources are obtained by taking into account the original observations.

3. RESULTS AND DISCUSSION

3.1. Simulation results

To test the performance of the source separation algorithm described in section 2, the proposed method and the FastICA algorithm [7] have been applied to a synthetic signal. Specifically, the first source is a chirp sound and the second source is a sinusoid. Figure 6 shows the synthetic sources and the observed signals resulting from the mixtures of both synthetic sources with the mixing matrix $A = [2\ 5\,;6\ 9]$. Similar results are obtained with different mixing matrices on those same signals.

The results of the separation are shown in figure 7. In this particular case where the simulated signal and their frequencies were selected arbitrarily, the ICA was not able to separate the two sources. On the other hand, the proposed method allowed a good synthetic sources separation (see figure 7).

3.2. Experimental results

We have also tested the performance of the two methods on multi-channel (two sources and two sensors) mixtures of different sound types and different mixing matrices. We chose recorded audio signals that represent real life situations like mixture of musical instruments, speech and music from the RWC music database [8]. For the experiment of this paper, we used four different

mixtures with the same mixing matrix $A = [2\ 1\,;1\ 1]$. An evaluation of the quality of the source separation algorithms was made using different evaluation criteria. We used the subjective criterion called perceptual evaluation of audio quality ($PEAQ$) [9] and the objective criterion called the log spectral distortion (LSD) which is defined as :

$$LSD = \frac{1}{L}\sum_{l=0}^{L-1}[\frac{1}{K}\sum_{k=0}^{K-1}(20log_{10}\frac{|S_{k,l}+\epsilon|}{\left|\hat{S}_{k,l}+\epsilon\right|})^2]^{\frac{1}{2}}, \qquad (3)$$

where L is the number of frames, K is the number of frequency bins, and ϵ is meant to prevent extreme values.

Table 1 summarizes the evaluations results obtained from the two source separation algorithms. In the majority of cases, we can see that the proposed method gives the best results, especially when the two sources have the same characteristics and come from the same musical instrument. For example with the PEAQ, the subjective impairment scale varies between -4 and 0.2 with the same parameter values than in [9]. Grade -4 corresponds to a very annoying impairment and grade 0.2 corresponds to an imperceptible impairment. Informal listening tests confirm these PEAQ values, i.e. that the separated sounds are perceptually very close to the original sounds prior to mixing when using the

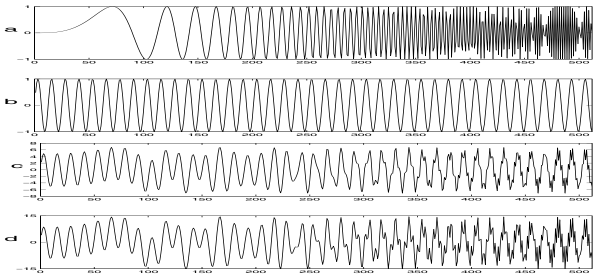

Fig. 6. *The simulated sources. (a) and (b) sources s_1, s_2. (c) and (d) mixtures x_1, x_2*

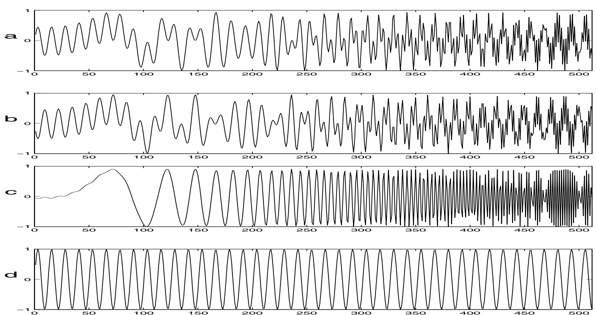

Fig. 7. *The estimated sources $\hat{s}_1$, $\hat{s}_2$ obtained using the two methods. (a) and (b) with conventional ICA [1]. (c) and (d) with the proposed method*

proposed algorithm, but that the quality using the FastICA alone is very poor in the case of flute and percussion instruments.

4. CONCLUSION

We proposed to combine ICA with the wavelet packets decomposition to separate two sources. We used the wavelet packets decomposition to increase the non-Gaussian nature of the signals to be further processed by ICA. Once the inverse matrix W is found with the wavelet packets based ICA, it is used to perform the separation in the time domain. The proposed method has been applied to audio signals and an evaluation (objective and subjective) of the separation quality has been made. In the instantaneous mixture case, the current results of the separation are promising. However, more investigations are then needed to separate the convolutive mixtures with the proposed method.

Acknowledgments

Many thanks to the reviewers for constructive comments and pointing out reference [10].

Sources	PEAQ		LSD	
	FastICA	PM	FastICA	PM
Male speech	-0.1	0.0	0.9	0.1
Music	0.1	0.2	0.6	0.5
Castanets instrument	0.0	-0.9	0.2	0.3
Gong instrument	-2.7	0.1	0.6	0.1
Flute1	-3.5	0.0	6.1	0.5
Flute2	-3.5	0.1	4.9	0.3
Male speech	-0.3	0.1	0.4	0.2
Female speech	-1.6	-1.6	2.4	2.3

Table 1. Comparison of the obtained results with FastICA and proposed method (PM)

5. REFERENCES

[1] A.J. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1004–1034, 1995.

[2] A Hyvarinen, J. Karhunen, and E. Oja, "Independent component analysis," *Wiley and Sons*, 2001.

[3] T. Tanaka and A. Cichocki, "Subband decomposition independent component analysis and new performance criteria," *ICASSP 2004*, pp. 541–544.

[4] P. Kisilev and M. Zibulevsky, "Blind source separation using multinode sparse representation," *ICIP'01, Thessalonicki, Greece*, 2001.

[5] S. Mallat, "A theory for multiresolution signal decomposition," *IEEE Tr on PAMI*, vol. 11, pp. 674–693, 1989.

[6] R. Coifman and M. Wickerhausser, "Entropy-based algorithms for best-basis selection," *IEEE Tr on IT*, vol. 38, pp. 713–718, 1992.

[7] A. Hyvrinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Tr on NN*, vol. 10, pp. 626–634, 1999.

[8] M. Goto, "Development of the rwc music database," *ICA2004*, pp. 553–556.

[9] P. Kabal, "An examination and interpretation of itu-r bs.1387: Perceptual evaluation of audio quality," *TSP Lab Technical Report*, 2003.

[10] A. Cichocki Y. Li and S. Amari, "Sparse component analysis for blind source separation with less sensors than sources," *in ICA2003*, pp. 89–94, 2003.