Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey, May 27-29, 2006 (pp140-144)

An Adaptive Algorithm for Speech Source Separation in Overcomplete Cases Using Wavelet Packets

Behzad Mozaffary[1] Mohammad A. Tinati[1] Ali Aghagolzadeh[1] Abbas Erfanian[2]
1: Faculty of Electrical and Computer Engineering, University of Tabriz, Iran
2: Department of Electrical Engineering, Iran University of Science & Technology

Abstract-- Speech process has benefited a great deal from the wavelet transforms. Wavelet packets decompose signals in to broader components using linear spectral bisecting. In this paper, mixtures of speech signals are decomposed using wavelet packets, the phase difference between the two mixtures are investigated in wavelet domain. A geometrical model is defined and an adaptive algorithm is proposed based on phase difference parameter in wavelet domain. Speech signals are separated from mixtures in an overcomplete-case.

Keywords: Wavelet Packets, Blind Source Separation, Speech Processing, Overcomplete, Scatter Plots, BSS

1. Introduction

Blind source separation of speech signals has been a topic of research investigation in the field of signal processing in recent years. This problem involves recovering unknown sources only by observing some mixed signals of data [1]. Generally, it is assumed that sources are statistically independent from each other and at most one of them could be a Gaussian signal [2].

Several algorithms have been proposed in literature addressing the overcomplete source separation problem recently. Lewicki [3] provided a complete Baysian approach assuming Laplacian source prior to estimating both the mixing matrix and the source in time domain. Clustering solutions were introduced by Hyvarinen [4] and Bofill-Zibulesky [5]. Davies and Miltianoudis [6] employed modified discrete cosine transform (MDCT) to obtain a sparse representation.

In this paper, two-sensor source separation with no additive noise is explored where the source separation problem becomes a one-dimensional optimal detection problem. A geometrical model in wavelet domain is defined and separation of sources is accomplished in accordance with their mixing order. We propose an algorithm based on phase difference between the data obtained from the sensors. Wavelet packets are obtained from decomposition of the two mixtures.

2. Background Material

Field of signal processing has benefited a great deal from wavelet transforms. Wavelets are transform methods that has been used a lot as a tool over the past decade. Wavelet transform is a time-scale representation that decomposes signals into basis functions of time and scale, which makes it useful in applications such as signal denoising, wave detection, data compression, feature extraction, etc.

There are many techniques based on wavelet theory, such as wavelet packets, wavelet approximation and decomposition, discrete and continuous wavelet transform, etc.

Backbone of the wavelets theory is the following two equations:

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^{j}t-k) \tag{1}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^{j}t-k) \tag{2}$$

Where $\varphi(t)$ and $\psi(t)$ are basic scaling function and mother wavelet function respectively [7].

Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey, May 27-29, 2006 (pp140-144)

A wavelet system is a set of building blocks to construct or represent a signal or a function. It is a two dimensional expansion set. A linear expansion would be as:

$$f(t)=\sum_{k=-\infty}^{+\infty} c_k \varphi(t-k)+\sum_{k=-\infty}^{+\infty}\sum_{j=0}^{+\infty} d_{j,k}\psi(2^j t-k) \qquad (3)$$

Most of the results of wavelet theory are developed using filter banks. In applications one never has to deal directly with the scaling functions or wavelets, only the coefficients of the filters in the filter banks are needed. A full wavelet packet decomposition binary tree for tree scale wavelet packet transform is shown in figure (1).

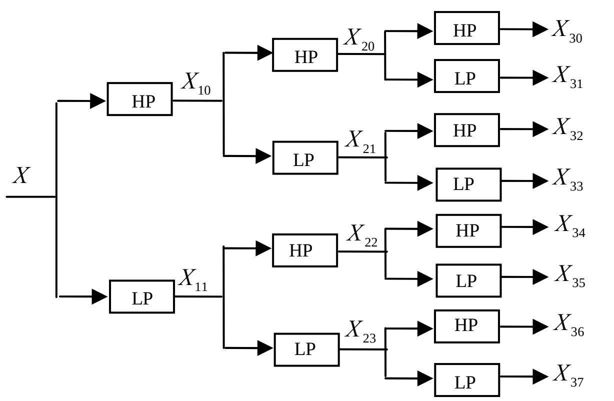

Figure (1) Full wavelet packet decomposition by filter banks

3. Scattering plots

Assume a set of M sensors and N source signals expressed as the following vectors: $\boldsymbol{X}(t)=[x_1(t),x_2(t),x_3(t),\ldots,x_M(t)]^T$ where $x_i(t)$ is the output of the i^{th} sensor.
$\boldsymbol{S}(t)=[s_1(t),s_2(t),s_3(t),\ldots,s_N(t)]^T$ where again $s_i(t)$ is the i^{th} source signal.

In this paper we will assume noise-free instantaneous mixing model i.e. $\boldsymbol{X}(t)=\boldsymbol{A}.\boldsymbol{S}(t)$, where $\boldsymbol{A}$ denotes the mixing matrix. The source separation problems consist of estimating the original sources $\boldsymbol{S}(t)$, given the observed signals $\boldsymbol{X}(t)$. In the case of equal number of sources and sensors $(N=M)$, a number of robust approaches using independent component analysis (ICA) have been proposed by Mitianoudis [8]. In the overcomplete source separation case $(M<N)$, we need to estimate *i)* mixing matrix A and *ii)* source signals $\boldsymbol{S}(t)$.

The scatter plots are plots of two or more signals in a single coordinate axes. In figure (2) we have shown a typical scatter plot of two sensor signals, that is, two mixtures of three speech signals are plotted. As the scatter plot shows each source signal is aligned in a particular direction. The most important parameter here is the angle θ which we refer to as the phase difference of two observed signals and is calculated as:

$$\theta_t = arctg[\frac{P_i(x_2)}{P_i(x_1)}] \qquad (4)$$

Where $P_i(x_j)$ is the i^{th} packet wavelet of j^{th} observation signal. By using scatter plot, two dimensional BSS problem is mapped into a one dimensional case with θ_t being the parameter of concern only.

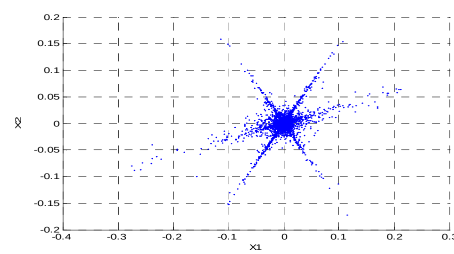

Figure (2) scatter plot of $x_2(t)$ respect to $x_1(t)$ in wavelet Domain

4. Separation algorithm

In this section we propose an algorithm to separate speech signals from mixtures whenever there are more sources than sensors. The separation criteria, is formulated using estimated mixing matrix.

For instance, if there are two sensors and three sources then the mixing matrix is expressed as the following equations:

$$\begin{cases} x_1(t)=a_{11}s_1(t)+a_{12}s_2(t)+a_{13}s_3(t) \\ x_2(t)=a_{21}s_1(t)+a_{22}s_2(t)+a_{23}s_3(t) \end{cases} \qquad (5)$$

In the matrix form:

$$\boldsymbol{X}=\boldsymbol{A}.\boldsymbol{S} \qquad (6)$$

For simplicity we can set all the coefficients in one of the above equations to one. The reason is only the ratios of the signals are important to us. Equation (5) can be rewritten as:

$$\boldsymbol{X}(t) = \vec{\boldsymbol{b}}_1 s_1(t) + \vec{\boldsymbol{b}}_2 s_2(t) + \vec{\boldsymbol{b}}_3 s_3(t) \qquad (7)$$

Equation (7) suggests that each source signal in the scatter plots will be in the $\vec{\boldsymbol{b}}_j$ direction. For instance if $s_2(t)$ and $s_3(t)$ were both zero then the scatter plot will be in the direction of a_{11}/a_{21}. For assuming $a_{11}=1$ then the direction will be determined by a_{21}. A typical figure showing the directions of signals of equation (7) are shown in figure (3).

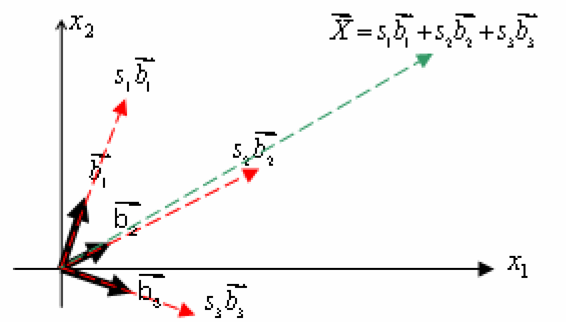

Figure (3) mixture vectors and source vectors in mixture space

In order to increase the sparsity of signals we use the wavelet packet decomposition (WPD) on the observed signals [9]-[11], and wavelet packet coefficients will be used to plot the scatter-representation.

The block diagram of the proposed algorithm is shown in figures (4) and (5).

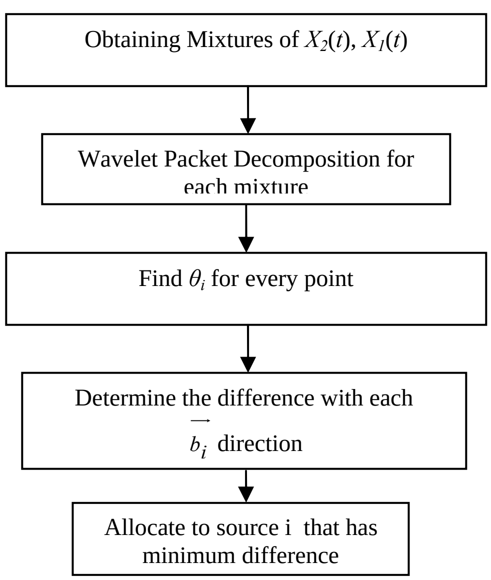

Figure (4) block diagram of separation algorithm

In figure (3) the mixing vector is also shown. Our purpose is to decompose the mixing vector in the direction of unit vectors $\vec{b}_1, \vec{b}_2, \vec{b}_3$ in such that the optimum value for every source signal is obtained. Note that there could be many different vector combinations of $s_i\vec{b}_i$ that will produce $\vec{X}$. This is because we have less degree of freedom (overcomplete-case).

In the scatter plots every source signal could be partitioned into two groups of points. First group are the geometrical focus points that are aligned in the direction of their own unit vectors, that is in the direction of $\vec{b}_i$. In order to find these focus points the phase angles of all the points in the vector space are calculated and compared with the direction of every source in the scatter plot. A minimum angle difference is assumed. Those points with lower than the threshold level are therefore allocated to a specific $\vec{b}_i$ direction. This is shown in figure (4).

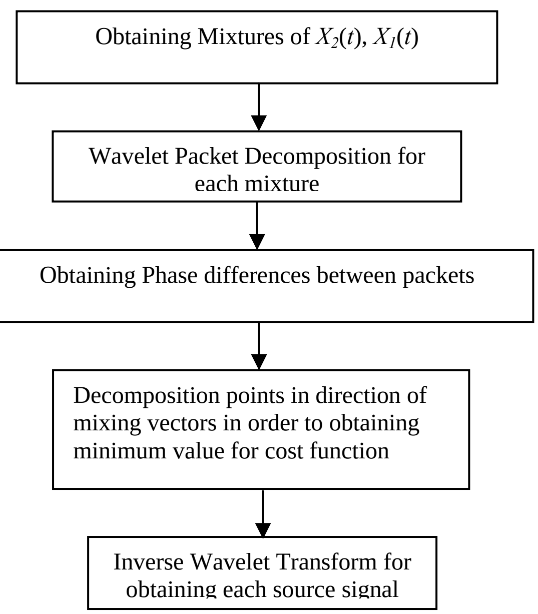

Figure (5) block diagram of separation algorithm

The second group are those points that do not pass from the threshold for phase angle and therefore we need to set criteria in order to allocate them to specific direction. This is explained as in figure (5).

As the block diagram of figure (5) shows, first wavelet packet of each mixture is calculated. Then using equation (4) the phase angle between the mixtures is obtained. The next step is to use a cost function as:

$$J = \frac{1}{2}\sum_{j=1}^{3}\sum_{i=1}^{3}\left|s_i s_j\right| \quad for\ all \quad i \neq j$$
$$= \left|s_1 s_2\right| + \left|s_1 s_3\right| + \left|s_2 s_3\right| \tag{8}$$

To obtain minimum correlation between any two decomposed signals. Minimization is based on steepest decent algorithm:

$$\frac{\partial J}{\partial s_k} = 0 \tag{9}$$

Minimization is accomplished in an adaptive fashion.

5. Simulation results

We have tested our algorithm in two different cases, which are *i)* two sources and *ii)* three sources. Figures (6-a) and (6-b) show a music and speech signals respectively. These signals are mixed and two mixtures obtained which are shown in figures (6-c) and (6-d). Then we applied our algorithm to these mixtures and results are plotted in figures (6-e) and (6-f). In figures (6-g) and (6-h) we have also shown errors between original source signals and those estimated using the proposed algorithm.

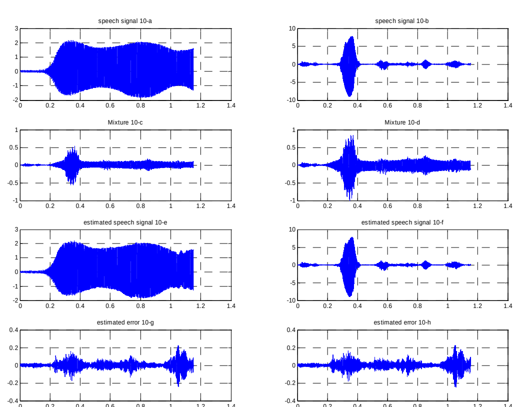

Figure (6) source signals (a,b), mixture signals (c,d), estimated signals (e,f), error signals (g,h)

To evaluate the efficiency of our algorithm, we computed signal to noise ratio (SNR) as:

$$SNR = 10\log\left(\frac{\left|s(t)\right|^2}{\left|s(t)-\hat{s}(t)\right|^2}\right) \tag{10}$$

where $\hat{s}(t)$ is the estimated signal? Results are shown in table below.

Table: SNR of estimated source signals

-	SNR1(db)	SNR2(db)	SNR3(db)
2 speech & music	32.56	32.34	-
3 speech	29.8	28.65	25.63

Figures (7) show the simulation results for three speech signals.

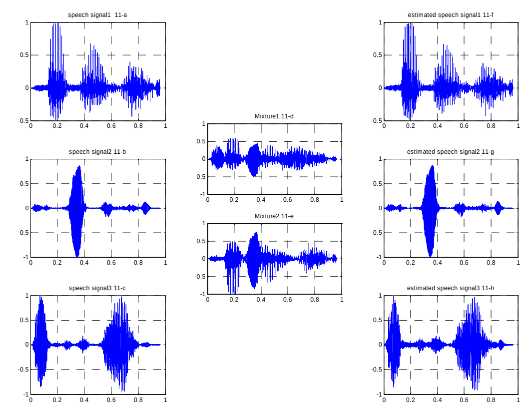

Figure (7) source signals (a,b,c), mixture signals (d,e), estimated signals (f,g,h)

6. Conclusion

In this investigation we have shown that one can use the coherent phase information between wavelet packets to estimate mixing matrix in a speech mixtures.

We mapped two dimensional problem to one dimensional (phase differences between two packets in wavelet domain.) and then we get more accurate estimation of the source signals. Two examples with two and three source components in the mixture were undertaken for simulations. Results indicate that we have been able to estimate the mixing matrix with a high degree of accuracy. Finally we must add that if high resolutions are used in wavelet packet domain, we will be able to obtain better estimations of source signals.

7. References

[1] A. Mansour, A. Kardec Barros, and N.Ohnishi, "Blind separation of sources: Methods, assumptions and applications," IEICE Transaction on Fundamentals of Electronics, Communications and Computer Sciences, vol. E83-A. No. 8, pp. 1498-1512, August 2000

[2] P. Comon, "Independent component analysis: A new concept," Signal Processing, vol.36, no.3, pp.287-314, April 1994

[3] M. Lewicki and T.J. Sejnowski, "Learning over complete representations networks," Neural Compute., vol. 12, pp.337-365,2000

[4] A. Hyvarinen, "Independent component analysis in the presence of Gassian noise by maximizing joint likelihood networks," Neural Compute., vol. 22, pp.49-67,1998

[5] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparce representation networks," Signal Processing, vol. 81, no. 11, pp. 2353-2362, 2001

[6] M. Davies and N. Mitianoudis, "A simple mixture model for sparce overcomplete ICA networks," Proc. Inst. Elec. Eng. Vision, Image,Signal Process., vol. 151, no. 1, pp. 35-43 , 2004.

[7] C. S. Burrus, R. A. Gopinath, H. Guo, "Introduction to Wavelets and Wavelet Transforms, a primer" Prentice Hall New jersey, 1998.

[8] N. Mitianoudis, "Audio source separation using independent component analysis," Ph.D. dissertation, Queen Mary, London, U.K, 2004

[9] M.A. Tinati, B. Mozaffari, "Comparison of Time-frequency and Time-scale analysis of speech signals using STFT and DWT, " WSEAS Transaction on Signal Processing, Issue 1, Vol. 1, pp. 11-16, Oct. 2005

[10] M.A. Tinati, B. Mozaffari, "A Novel Method for Noise Cancellation of Speech Signals Using Wavelet Packets"