

Blind Signal Separation Using an Extended Infomax Algorithm

Samira Ashouri

*Noushirvani University of Technology,
Faculty of Computer & Electrical Engineering, Babol, Iran
E-mail: samiraashouri@yahoo.com*

Abstract

The Infomax algorithm is a popular method in blind source separation problem. In this article an extension of the Infomax algorithm is proposed that is able to separate mixed signals with any sub- or super-Gaussian distributions. This ability is the results of using two different nonlinear functions and new coefficients in the learning rule. In this paper we show how we can use the distribution of observation vectors for selecting suitable coefficients for our algorithm. Hence, the proposed algorithm is suitable for real applications in which the distribution of source signals might be unknown. It is also shown in this paper that the extended Infomax algorithm is able to separate 23 sources with a variety of distributions. Incidentally, we use a performance criterion for the evaluation of our results, based on the comparison of Kurtosis of the original signals and estimated signals.

Keywords: *Blind source separation, distribution function, kurtosis.*

1. Introduction

Blind source separation (BSS) is a powerful technique in signal processing. This method is applied to a mixture of several sources to separate them. Blind in this context emphasizes that the source signals are unobservable and no prior information exists about them or the mixing process is unknown. It is assumed that the sources are independent and they have been mixed linearly [1][2]. One common approach employed for solving this problem is the use of Independent Component Analysis (ICA) technique [1]. The ICA has applications in telecommunications, medical science, speech recognition and speech denoising [3].

Within blind source separation research, instantaneous BSS and convolutive BSS are two important problems that are generally considered. The difference between these two is based on the nature of the signal mixing process. In contrast to convolutive BSS, time delay is not considered in instantaneous BSS. This paper, however, concerns with the instantaneous case.

Assume that we have N statistically independent zero mean signals, $s_i(t), i = 1, \dots, N$, observed using M sensors. Hence, a set of M observation signals $s_i(t), i = 1, \dots, M$, is obtained that are mixture of the sources. By illustrating the time delay, the observed signals can be modeled using the following instantaneous model [1]:

$$x(t) = As(t), \tag{1}$$

where $A \in \mathfrak{R}^{M \times N}$ ($N \leq M$) is an unknown matrix called the mixing matrix. The objective is to recover the original signals from only the observation vector. It is shown that an unmixing matrix $W \in \mathfrak{R}^{M \times N}$ can be found to estimate the source signals $y_i(t), i = 1, \dots, N$ using the following equation:

$$x(t) = Wx(t), \quad (2)$$

In this paper we consider the number of sensors equal to the number of sources ($N = M$) and the observations are zero-meaned signals.

Since the original sources, $s_i(t)$ were assumed to be independent, by using ICA to reduce the statistical dependence between components and makes them as independent as possible, one can be assured that y_1 would actually be one of the original sources [1]. Most of the existing algorithms use this concept to solve the BSS problem, but they can be distinguished by the criterion employed to measure the dependence between the signals.

A generic approach to solve the basic BSS problem consists of [13]:

- data preprocessing, that usually includes data centering and data whitening,
- considering a criterion to measure non-Gaussianity or independency, and
- optimizing an objective function following the criterion measuring the output dependency.

Various algorithms have been proposed in the literature to solve the BSS problem. Some of these algorithms include Infomax [4], EASI algorithm [5], the natural gradient [6], and FastICA algorithm [7]. In general, these algorithms can be classified in two groups: the algorithm based on statistical analysis [8] and those use neural network [9]. The neural network based methods are considered to be more computationally efficient compared to the statistical methods [10], but they might be slow in convergence. Hence, in this article we intend to improve the Infomax algorithm, which is one of the successful neural network-based approaches in separating super-Gaussian signals. One of the limitations of the Infomax algorithm is that it cannot adopt itself to inputs with a variety of distributions. In other words, the learning rule defined by Bell and Sejnowski assumes that there is only one nonlinear function for the mapping network. Consequently, it can only separate signals with the same distribution [11]. This persuaded other researchers to modify the type of learning rule in Infomax algorithm [12]. In this paper we propose to use two different functions with adaptive coefficients. When signals are fed to the system in the proposed approach, the coefficients can be changed to adjust with the observation signals. Hence, there is no need to have prior information about the distribution of the received signals.

2. The Extended Infomax Algorithm

There are three different types of distributions for signals: super-Gaussian, Gaussian, and sub-Gaussian signals. The difference between PDF of sub-Gaussian and super-Gaussian distributions are shown in Figure 1. One of the limitations of the existing BSS algorithms is that they cannot adopt themselves to a variety of inputs distributions.

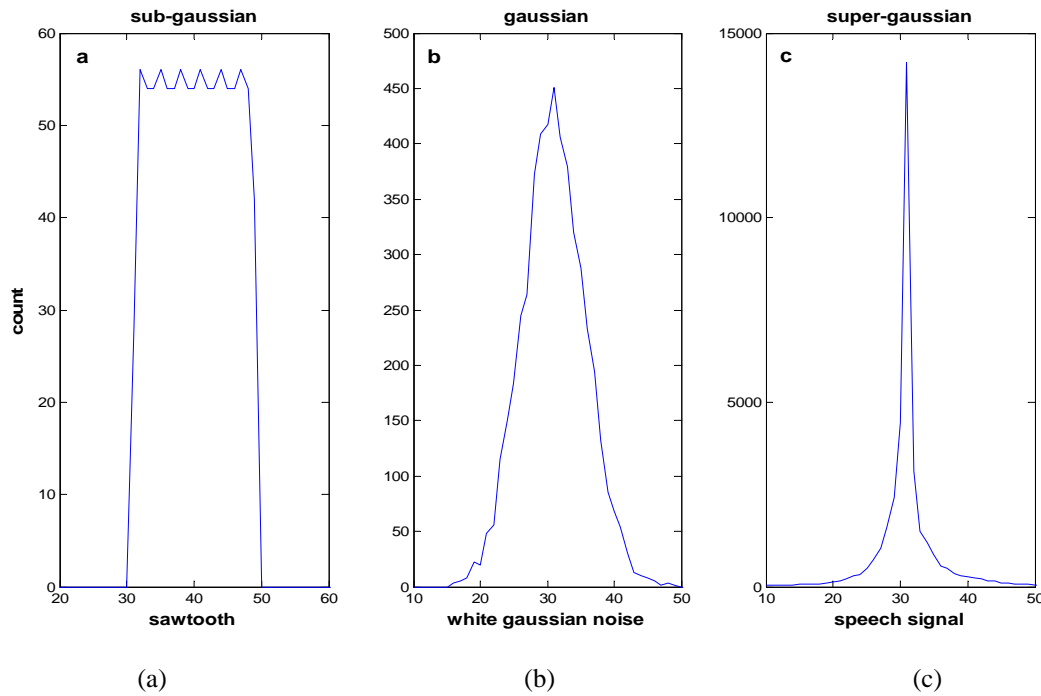


Figure 1: pdf of signals with different distribution: (a) a sub-Gaussian signal, (b) a Gaussian signal, and (c) a super-Gaussian signal.

Kurtosis, the fourth-order cumulant, can be considered as a measure indicating non-Gaussianity of a random variable. For a zero-mean variable, the normalized Kurtosis can be expressed as [15]:

$$kurt(y) = \frac{E\{y^4\}}{(E\{y^2\})^2} - 3 \quad (3)$$

where $E\{\cdot\}$ is the expectation operator. For a Gaussian random variable, Kurtosis is zero. For distributions with heavy tails and a peak at zero, it is greater than zero. Thus, they are called super-Gaussian random variables. For flatter densities with lighter tails, Kurtosis is negative. Thus, they are called sub-Gaussian random variables [15]. Kurtosis may also be defined as below[1][15]:

$$kurt(y) = \frac{E\{y^4\}}{(E\{y^2\})^2} \quad (4)$$

Following equation (18), the signal's Kurtosis for Gaussian, super-Gaussian, and sub-Gaussian distribution would respectively be 3, greater than 3, and below 3.

The original Infomax algorithm is effective in separating super-Gaussian sources[16]. This is due to using only one nonlinear function in the learning rule of the neural network. Hence one of the objectives of extended Infomax algorithm is to provide a simple learning rule that can separate sources with a variety of distributions [16].

The authors in [16] have improved the Infomax algorithm to consider input signals with super- or sub-Gaussian distribution. In this approach, they have used the following learning rule [16]:

$$\Delta w \propto \frac{[I - \tanh(u)u^T - uu^T] \times W : \text{super-Gaussian}}{[I + \tanh(u)u^T - uu^T] \times W : \text{sub-Gaussian}} \quad (5)$$

The above leaning rule can be rewritten as below:

$$\Delta W \propto [I - K \tanh(u)u^T - uu^T] \times W, \quad (6)$$

where K represents the sign of Kurtosis of the signal. In (20), K is either 1 or -1 for respectively super-Gaussian and sub-Gaussian signal.

In some of the BSS algorithms such as HJ [14] and EASY [14] algorithms, it is common to use various functions in order to separate signals with sub-Gaussian or super-Gaussian distribution. Therefore, this induced us to employ two different functions with adaptive coefficients for the Infomax learning rule, in order to solve the problem of the Infomax algorithm in separating sub-Gaussian or super-Gaussian signals. For new inputs with different distributions, these coefficients can change and adjust themselves to the new observation input. Therefore, there is no need to have a prior information about distributions of the original signals.

3. The New Extended Infomax Algorithm

3.1. Two Mixed Sources

To illustrate the proposed extended Infomax algorithm, assume two signals with any sub- or super-Gaussian distributions mixed together in two different ways to make the observed signals x_1 and x_2 . The observed signals are passed through a prewhitening stage to decorrelate them.

The prewhitening operation often makes the algorithm to be simpler and to speeds up their convergence [7]. For convenience, from here on, we assume that x is the prewhitened observed signal.

As we mentioned above, we want to use two different functions instead of a single function in the original Infomax algorithm, so that in different situation, – as described later on – based on the Kurtosis of observed signals, both of the functions can be used in separation process. In the HJ algorithm, functions such as t , t_3 , $sgn(t)$, and $\tanh(t)$ have been used to separate signals. Note that t and t_3 are ascending function, but $sgn(t)$ and $\tanh(t)$ are descending function. Hence, we intend to use both groups of functions for our new algorithm.

Suppose s as the original signal, x as the observed signals and $y=Bx$ as the estimated signals, the following rule is suggested:

$$B_{new} = B_{old} + \mu (I - C_{1.2}.g(y).y^T - C_{1.2}.h(y).y^T).B_{old} \quad (7)$$

where B is the separating matrix. We have chosen $g(y)=\tanh(y)$ and $h(y)=t^3$, and the μ is a learning factor. C_1 and C_2 coefficients have been considered in order to determine the effect of each of the separating functions in the separation process. Of course, this should be performed by considering the distribution of observed signals. In fact, we suppose that we have no information about the distribution of the original signals. Therefore, our main concern is to choose C_1 and C_2 properly. Hence, the first step in this algorithm is to determine the Kurtosis of observed signals. But, in order to find the relationship between the Kurtosis of observed signals with the Kurtosis of original signals, first of all, a series of experiments are conducted. These experiments may have any of the following conditions:

- a. The mixture of two super-Gaussian signals became super-Gaussian too.
- b. The mixture of two sub-Gaussian signals became sub-Gaussian too.
- c. The distribution of the mixture of two signals, a super-Gaussian with a sub-Gaussian, would be unpredictable. Indeed, the distribution depends on the mixing matrix and coefficients of the matrix. As further illustration, assume that there are two signals with Kurtosis of 1.5053 and 3.8885. We mix them by different mixing matrix and the results are illustrated in Table 1.

Table 1. A mixture of two signals.

The mixing matrix	Observed signals' Kurtosis	
0.3 0.7	2.6453	3.5808
0.7 0.6	sub-Gaussian	super-Gaussian
0.8 0.1	1.5485	1.6398
0.9 0.2	sub-Gaussian	sub-Gaussian
0.1 0.8	3.7927	3.8546
0.2 0.9	super-Gaussian	super-Gaussian

According to Table 1, in the first row, we get one sub-Gaussian and one super-Gaussian signal as the result of mixing these two signals. With the change in the coefficient of mixing matrix, in the second and third rows, respectively, the results are two sub-Gaussian signals and two super-Gaussian signals.

The above results are consistent with the central limit theorem.

According to the central limit theorem, the distribution of a mixture of statistically independent signals will have a distribution closer to the Gaussian distribution compared to the distributions of the original signals [17][18]. For example, the sum of two sub-Gaussian signals has a distribution that is closer to the Gaussian distribution than the distribution associated with each of the two original signals (see Figure 2).

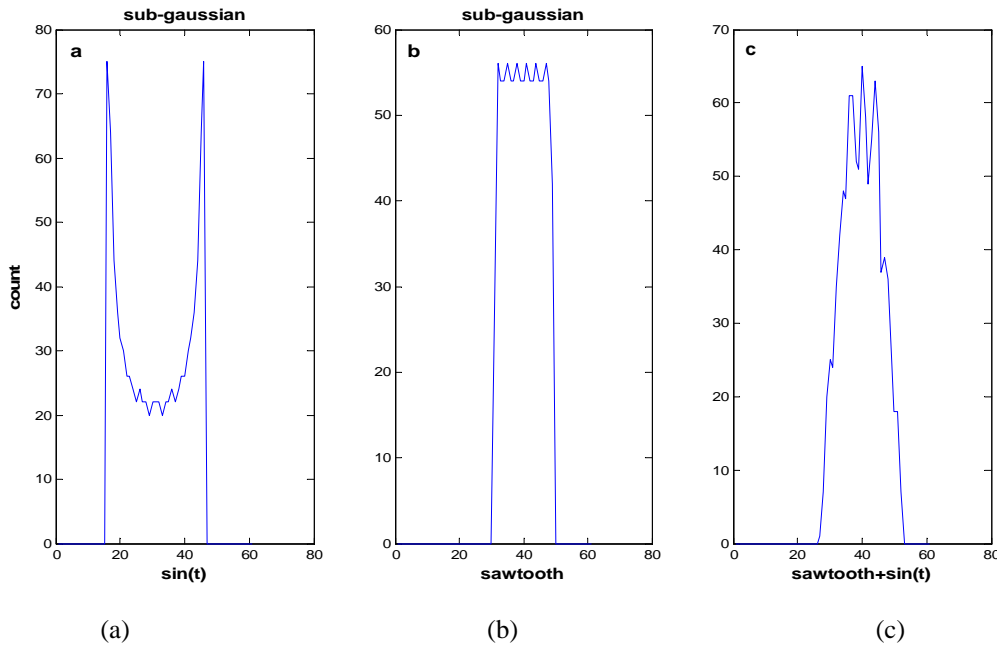


Figure 2. Comparing the distributions of two sub-Gaussian signals, (a) and (b), with that of the result of adding the two signals (c).

Following the above results if we determine the Kurtosis of the observed signals, we will have one of the three situations as follows:

I. If the observed signals Kurtosis is greater than 3^1 (see line 3 in table 1), it does not mean that the original signals were necessarily super-Gaussian, but rather, we can say either both of them were super-Gaussian or, if one of them was sub-Gaussian, the mixing matrix has been so that the effect of Super-Gaussian signal has been more than the effect of sub-Gaussian one. So, in this condition the algorithm by selecting $C_1 > 0$ and $C_2 > 0$, chooses the function $g(y) = \tanh(y)$ as an active function. In our simulations we have experientially selected 1 for positive position of each coefficient.

II. If the observed signals Kurtosis was lower than 3 (see line 2 in Table 1, neither does it mean that the original signals were necessarily sub-Gaussian, but rather, it is possible that either both of them were sub-Gaussian or if one of them was super-Gaussian, the mixing matrix played an essential role. So, in this condition $C_1 > 0$ and $C_1 = 0$ are $C_2 > 0$ selected and the active function became $h(y) = t^3$.

III. The mixed signal may contain super-Gaussian and sub-Gaussian signals. In this case, there has to be an additional condition in order to do the separation process. In fact, in these circumstances, first, the algorithm considers one of the coefficients (C_1 or C_2) to be zero and so, one of the functions becomes the active function, hence we get the primary estimated signals. At the next stage, the Kurtosis of these estimated signals will be compared with the Kurtosis of primary observed signals. If they are in contrast to the central limit theorem, the algorithm changes the active function by changing the coefficients.

1. We use equation (4) for calculating the signal's kurtosis.

3.2. More Than Two Mixed Sources

With the increase of the number of sources, we will have one of these three situations:

- all the observed signals are Sub-Gaussian
- all the observed signals are super-Gaussian
- the observed signals are a mixture of sub-Gaussian and super-Gaussian signals

For the first two situations, we act as described in the previous section. For the third situations the separation process is required to be performed in two stages. In fact, since the super-Gaussian signals have a high Kurtosis¹, therefore, in the first step, we can separate them from each other and of course from the mixed sub-Gaussian signals. Then, in the next stage, the remaining signals, which are a mixture of sub-Gaussian signals, are separated from each other.

To further illustrate this situation, consider the signals provided in Figure 3. In this figure, four signals, two sub-Gaussian and two super-Gaussian are mixed together and the observed signals are shown in Figure 4. The signal separation process is performed in two stages as below:

Stage 1: By choosing $C_1=1$ and $C_2=0$ and activating the $\tanh(y)$ function in (7), the super-Gaussian signals are estimated (see Figure 5). As can be seen, the Kurtosis of the estimated super-Gaussian signals is very close to the Kurtosis of the observed signals. But, the resulted sub-Gaussian signals are still a mixture of each other. Therefore, in the second stage, we only need to separate the sub-Gaussian signals. So, the input signals of the second stage are the two sub-Gaussian signals shown in Figure 6.

Stage 2: Now, by choosing coefficients $C_1=0$ and $C_2=1$, and activating t^3 function in (7), by considering the signals of Figure 5 as the input of the network, the estimated signals will be as follows in Figure 6.

All the above stages are done automatically by the program itself. Finally, we get the results shown in Figure 8.

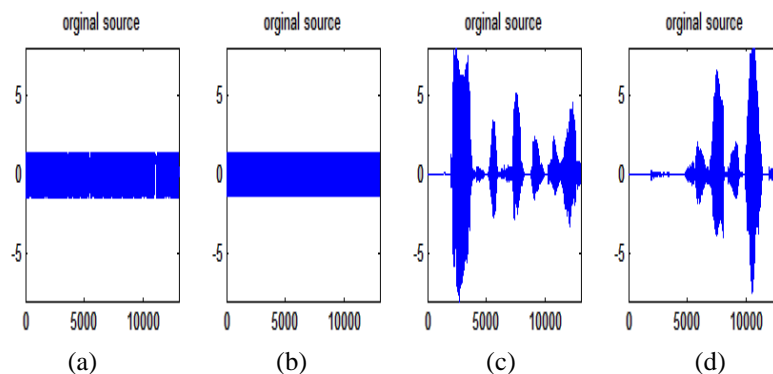


Figure 3. Original signals which include (a,b) two sub-Gaussian signals with the Kurtosis of 1.5006 and 1.5000 and (c,d) two speech signals as the super-Gaussian signals with the Kurtosis of 17.0620 and 19.6150.

1 . The kurtosis of super-Gaussian signals used in the simulatons, are above of 10.

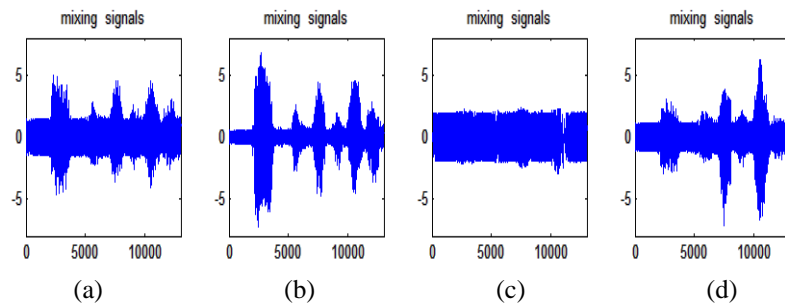


Figure 4. Observed signals which are a mixture of the above signals include the signals with Kurtosis of 8.1229, 9.6392, 2.3111 and 19.6150.

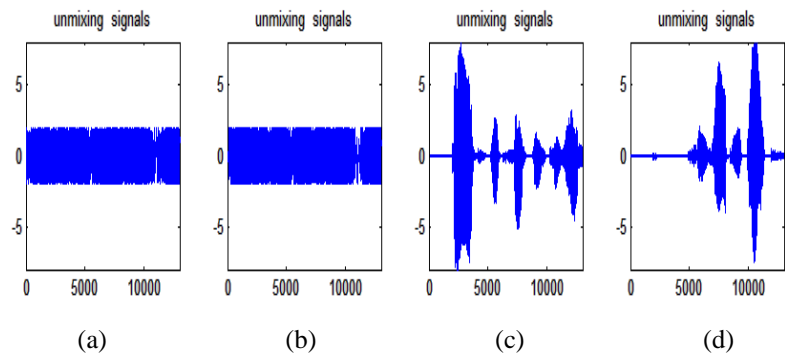


Figure 5. Estimated signals from stage 1 with the Kurtosis of 2.2414, 2.2827, 17.0625 and 19.6148.

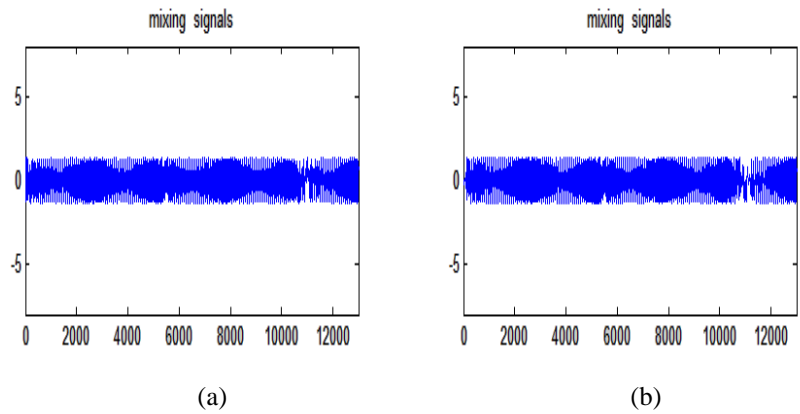


Figure 6. The input signals of the second stage with the Kurtosis of 2.2414 and 2.2827.

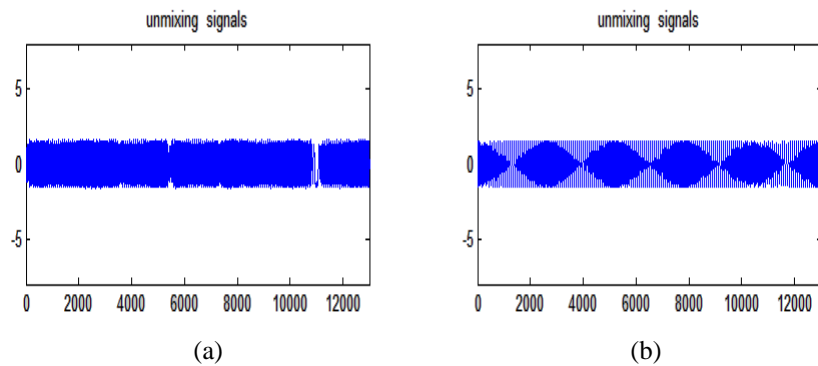


Figure 7. The output signals from the second stage with the Kurtosis of 1.5317 and 1.5480.

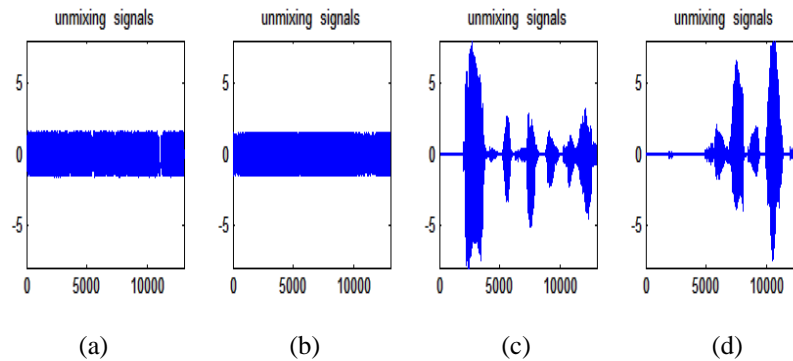


Figure 8. The results of the proposed method when the signals in Figure 4 are fed to the system.

In Table 2, we can compare the Kurtosis of the original signals with the estimated signals.

Table 2. The value of signals' Kurtosis

Original signals' Kurtosis	1.500.	1.5006	17.062	19.6150
Observed signals' Kurtosis	2.3111	3.9578	8.1229	9.6392
Estimated signals' Kurtosis	1.5317	1.5480	17.0625	19.6148

4. Experimental Results

Our experimental results in this paper include 3 Cases as described below. In all of the cases, we have mixed original signals by a matrix with the values randomly vary between 0 and 1. In Case 1, we show the ability of new extended Infomax algorithm to separate two sub-Gaussian signals. In the case 2, original signals have sub- and super-Gaussian distribution. In Case 3, the number of sources is increased up to 32.

Case 1

Suppose that two original signals are sub-Gaussian. One of them is a chirp wave and the other one is a Sinosoide signal. The results of mixing the signal are shown in Figure 9. Also, the value of signals Kurtosis have been provided in Table.3.

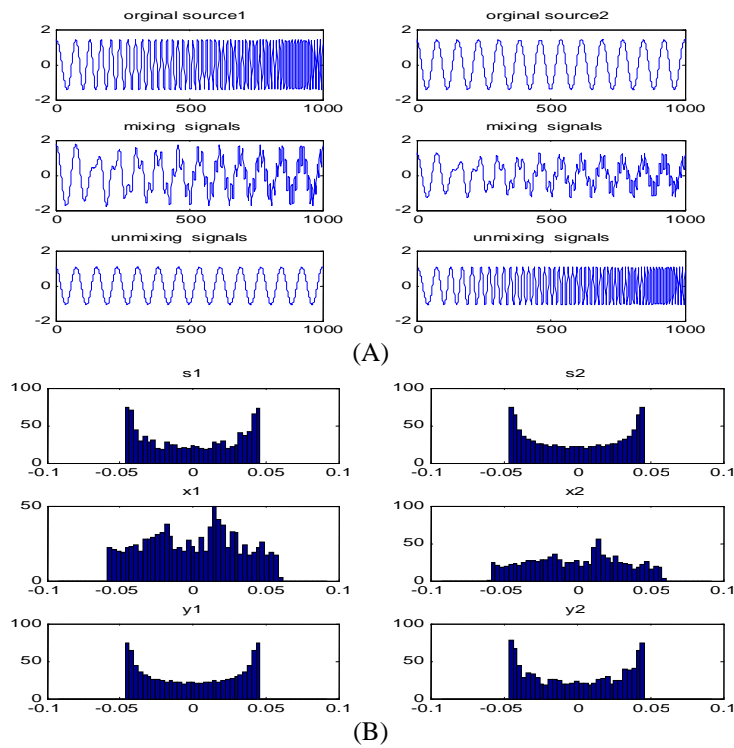


Figure 9. A: Respectively source, mixed, and estimated signals. B: Histogram of signals for respectively the source mixed, and estimated signals.

Table 3. The values of signals' Kurtosis.

Original signals' Kurtosis	1.5000	1.4996
Observed signals' Kurtosis	1.9624	1.9336
Estimated signals' Kurtosis	1.4999	1.4996

Case 2

Suppose five original signals: one of them is sub-Gaussian, and the rest of them are speech signals with a super-Gaussian distribution. The results of mixing them are shown in Figure 10. Incidentally, the values of signals' Kurtosis have been listed in Table 4.

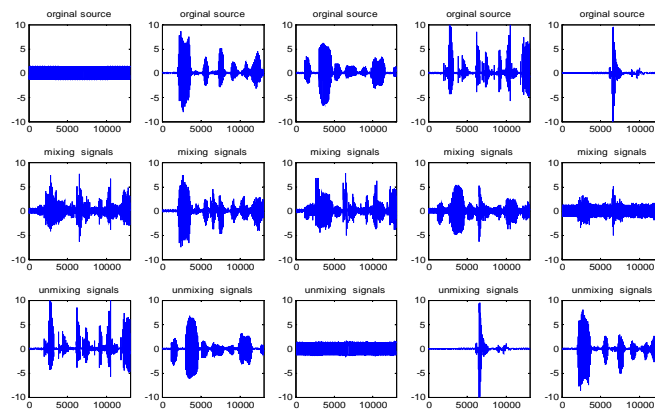


Figure 10. First row: source signals, medial row: mixed signals, final row: estimated signals.

Table 4. The values of signals' Kurtosis

Original signals' Kurtosis	1.500.	14.377	17.062	17.089	54.573
Observed signals' Kurtosis	4.120	8.603	8.978	9.193	11.872
Estimated signals' Kurtosis	1.503	14.341	17.062	17.081	54.573

Case 3

As a general example, we used 32 sources in which four of them are sub-Gaussian and the rest are speech signals with Kurtosis above 10. First we mixed two of them and recorded error of the system. Then, we increased the number of sources up to 32 and recorded errors for each experiment too. Finally, we plotted the errors in Figure 11. As can be seen, with the number of sources above 23 the results of the separation process are unsatisfactory.

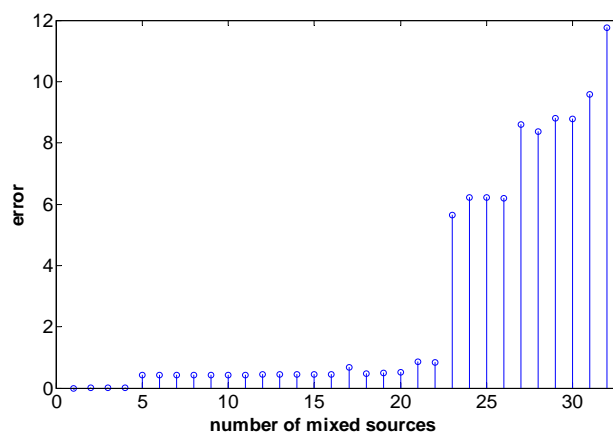


Figure 11. The error rate of the proposed method in separating sources from the mixed signals. The errors represent the distance between the Kurtosis of estimated signals with the original ones.

5. Conclusions

In this paper, an extension of the Infomax algorithm was proposed. One of the advantages of our algorithm is that we use two different nonlinear functions and new coefficients in the learning rule to separate the mixed signals with any sub- or super-Gaussian distributions. In fact, we have used the distribution of observation vectors for estimating the source signal. In the proposed approach, the signals are separated in two stages. In first stage, the super-Gaussian signal are separated. Then in the second stage, the sub-Gaussian signals are estimated. Incidentally, we have shown a performance measurement for the evaluation of BSS methods, which is based on the comparison of Kurtosis of the original signals and the estimated signals.

References

1. A. Hyvarinen, and E. Oja, "Independent Component Analysis: Algorithms and Applications", *Neural Networks*, 411-430, 2000.
2. Sriram Srinivasan, "Knowledge-Based Speech Enhancement", *Thesis for the degree of Doctor of Philosophy*, 2005.
3. S. Srinivasan, "Knowledge-Based Speech Enhancement", *Thesis for the degree of Doctor of Philosophy*, 2005.
4. A. J. Bell, and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution", *Neural Computation*, 7: 1129-1159, 1995.
5. J.-F. Cardoso, B. Laheld, "Equivalent adaptive source separation", *IEEE Trans. Signal Process.* 44 (12), 3017-3030, 1996.
6. S.-I. Amari, A. Cichocki, H. Yang, "A new learning algorithm for blind signal separation", in: D. S. Touretzky, M. C. Mozer, M. E. Hasselmo (Eds.), *Advances in Neural Information Processing Systems*, Vol. 8, pp. 757-763, 1996.
7. Hyvarinen, E. Oja, "A fast fixed-point algorithm for independent component analysis", *Neural Comput.* 9 (7), 1483-1492, 1997.
8. S. Amari and A. Cichocki, "Adaptive blind signal separation: Neural network approaches", *Proc. of IEEE*, Vol. 86, No. 10, pp. 2026-2048, October 1998.
9. J.-F. Cardoso, "Blind signal separation: Statistical principles", *Proc. of IEEE*, Vol. 86, No. 10, pp. 2009-2047, October 1998.
10. Xi, J, Chicharo, JF, Tsoi, AC & Siu, WC, "On the INFOMAX algorithm for blind signal separation", *WCCC-ICSP 2000. 5th International Conference on Signal Processing Proceedings*, vol 1, 425-428, 21-25 August 2000.
11. P. Comon, "Independent component analysis, A new concept?" *Signal Processing*, 36: 287-314, 1994.
12. A. Bell, and T. Sejnowski, "Fast blind separation based on information theory", *Neural Computation*, 7: 1129-1159, 1995.
13. V. David Sanchez A, "Frontiers of research in BSS/ICA", *Neurocomputing* 49, pp: 7 - 23, 2002.
14. J. Karhunen, "Neural approach to independent component analysis and source separation", *Proceedings of the 4th European Symposium on Artificial Neural Networks (ESANN '96)*, 1996.
15. K. R. Krishnamachari and Robert E. Yantorno and Jereme M. Lovekin and Daniel S. Benincasa and Stanley J. Wenzel, "Use of local Kurtosis measure for spotting usable speech segments in co-channel speech", *IEEE ICASSP*, 649-652, 2001.
16. T.-W. Lee, M. Girolami, T. Sejnowski, "Independent component analysis using an extended Infomax algorithm for mixed sub-gaussian and super-gaussian sources", *Neural Comput.* 11, 417-441, 1999.
17. J. G. Proakis "Digital Communications, 4th Edition" McGRAWHILL, 2000.
18. R. Prasad, "Does mixing of speech signals comply with central limit theorem?" *Int J Electron Commun (AEU)*, 2007.