Comparison of Fast ICA and Gradient Algorithms of Independent Component Analysis for Separation of Speech Signals

K. Mohanaprasad [#1], P. Arulmozhivarman [#2]

[#] School of Electronics Engineering, VIT University
Vellore, Tamil Nadu, India
[1] kmohanaprasad@vit.ac.in
[2] parulmozhivarman@vit.ac.in

Abstract — **Voice plays a vital role in distant communications like video conferencing, teleconferencing and hands free mobile conversion etc. Here, the quality of speech is degraded by the Cocktail party problem. Cocktail party problem is described as combination of various sources of speech signal received by a microphone. Solution for the above problem can be obtained by using Independent component Analysis (ICA), which has the ability to separate multiple speech signals into individual ones. This paper deals with application of principle of negentropy from maximization of non-gaussianity technique of ICA using Gradient and Fast ICA algorithm. The results in Matlab show that Fast ICA provides better execution time compared with gradient with minimum number of iteration.**

Keyword - ICA, Negentropy, Fast ICA, Gradient, Maximization of non-gaussianity.

I. Introduction

Imagine that two people speaking simultaneously are recorded using two microphones placed in different positions of the room. The microphones give two recorded signals $x_1(t)$ and $x_2(t)$ with x_1 and x_2 amplitudes, and t the time index. Each of the recorded signal is the linear combination of the two speech signals emitted by the speaker is denoted by $s_1(t)$ and $s_2(t)$ [1]. So we could express this linear equation as

$$x_1(t) = a_{11}s_1+a_{12}s_2 \quad (1)$$
$$x_2(t) = a_{21}s_1+a_{22}s_2 \quad (2)$$

Where a_{11}, a_{12}, a_{21}, a_{22} are the parameters that depend on the distances of the microphones from the speakers. Here the source signals s_1 and s_2 is estimated from the mixed signals x_1 and x_2 using Independent component analysis. This is known as blind source separation. In this process the mixed signals are obtained from statistically independent and non-Gaussian source signals. For simplicity we assume the unknown mixing matrix A, as the square matrix. The estimated source signals could be obtained up to their permutation, sign, and amplitude only that is their order and variance cannot be obtained with independent component analysis.

In recent years, Researchers had proposed many criterions, Minimization of Mutual information have been used to estimate source signals using Independent component analysis. In those maximization of non-Guassianity gives the better performance. There are two techniques in maximizing non-guasianity, they are using kurtosis and negentropy. In which negentropy is more reliable as kurtosis is most sensitive to outliers and computationally robust process.

In this paper, we estimated the source signals using Independent component analysis [2] by maximizing negentropy. The maximization of negentropy can be done using two algorithms (Fast ICA and gradient). To estimate the source signals the demixing matrix is estimated. The fundamental restriction in ICA [3] is that the independent components are non-guassian in nature. To see why gaussian variables make ICA impossible, assume that the signals are Gaussian and mixing matrix is orthogonal. Then x_1 and x_2 are Gaussian, uncorrelated, and of unit variance. Their joint density is given by

$$p(x_1, x_2) = \frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right) \quad (3)$$

This distribution is illustrated in Fig 1. The Figure 1 shows that the density is completely symmetric. So, it does not contain any information on the directions of the mixing matrix.

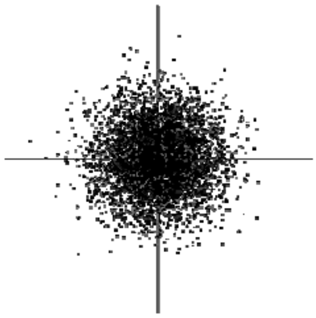

Fig. 1. Multivariate distribution of two independent Gaussian variables

To estimate one of the independent components, consider a linear combination of the x_i let us denote this by

$$y = w^T x = \sum_t w_t x_t \qquad (4)$$

where *w* is a vector to be determined. If *w* were one of the rows of the inverse of A, then the linear combination will equal one of the independent components. So determine such a *w* (i.e inverse of A) without knowledge of A matrix is not practical, but we can find an estimator that gives good approximation. To see how this leads to the basic principle of ICA estimation, let us make change of variables, defining

$$z = A^T w \qquad (5)$$

$$y = w^T As = z^T s$$

y is thus a linear combination of s_i, with weights given by z_i. Since a sum of even two independent random variables is more gaussian than the original variables, $z^T s$ is more gaussian than any of the s_i and becomes least gaussian when it in fact equals one of the s_i. In this case, only one of the elements z_i of z is nonzero. Therefore, we could take as *w* a vector that maximizes the non gaussianity of $w^T x$ [4]. Such a vector would necessarily correspond to a z which has only one nonzero component. This means that $w^T x = z^T s$ equals one of the independent components. Maximizing the non gaussianity of $w^T x$ thus gives us one of the independent components. To find several independent components, we need to find all the local maxima. Its not difficult, because different independent components are uncorrelated. This corresponds to orthogonalization in a suitably transformed (i.e. whitened) space.

II. Evaluation of Independent Components by Maximizing a Quantitative Measure of Non-Gaussianity

Two quantitative measures of non-gaussianity are used in ICA estimation are kurtosis and negentropy.

A. *Negentropy*

Negentropy is based on the information-theoretic quantity [5] of differential entropy, which we here call simply entropy. The more “random”, i.e., unpredictable and unstructured the variable is, the larger its entropy. The (differential) entropy H of a random vector y with density $p_y(\eta)$ is defined as

$$H(y) = -\int p_y(\eta) \log p_y(\eta) d\eta \qquad (6)$$

Gaussian variable has the largest entropy among all random variables. This means, entropy could be used as a measure of nongaussianity. Negentropy J is defined as follows

$$J(y) = H(y_{gauss}) - H(y) \qquad (7)$$

Where y_{gauss} is a Gaussian random vector of the same covariance matrix as ‘y’. Negentropy, or negative normalized entropy, is always non-negative, and is zero if and only if ‘y’ has a Gaussian distribution.

Negentropy can be done using two algorithms as stated above. To make computation easy we center the data to make mean zero and then we go for whitening process to make uncorrelated data with variance one. The whitening process is done by eigen value decomposition method.

$$\tilde{x} = ED^{-1/2}E^T x \qquad (8)$$

The estimation of negentropy is difficult, as mentioned above, and therefore this contrast function remains mainly a theoretical one. The classical method of approximating negentropy is using higher-order moments,

$$J(y) \approx \frac{1}{12}E\{y^3\}^2 + \frac{1}{48}kurt(y)^2 \qquad (9)$$

The random variable y is assumed to be of zero mean and unit variance. In particular, these approximations suffer from the no robustness encountered with kurtosis. To avoid the problems encountered with the preceding approximations, new approximations were developed. These approximations were based on the maximum-entropy principle. In general we obtain the following approximation

$$J(y) \approx \sum_{i=1}^{p} k_i[E\{G_i(y)\} - E\{G_i(v)]^2 \qquad (10)$$

Where k_i are positive constants, and v is a gaussian variable of zero mean and unit variance

B. Negentropy based fixed point algorithm

A much faster method for maximizing negentropy [6] is done using fixed-point algorithm. The resulting FastICA algorithm finds a direction, i.e., a unit vector w, such that the projection $w^T z$ maximizes non-gaussianity. Non-gaussianity is here measured by the approximation of negentropy.

FastICA is based on a fixed-point iteration for finding a maximum of the nongaussianity of $w^T z$. The FastICA algorithm using negentropy combines preferable statistical properties due to negentropy. The fixed point iteration can be approximated as follows:

$$w \leftarrow E\{zg(w^T z)\} \qquad (11)$$

The above iteration does not have the good convergence properties of the FastICA using kurtosis, because the non polynomial moments do not have the same nice algebraic properties as real cumulants like kurtosis. So the modified iteration process can be as below,

$$w \leftarrow E\{zg(w^T z)\} \Leftrightarrow (1+\alpha)w = E\{zg(w^T z)\} + \alpha w \qquad (12)$$

Due to the subsequent normalization of *w* to unit norm, the latter equation gives a fixed-point iteration that has the same fixed points. So choice of α is more useful, it may be possible to obtain an algorithm that converges as fast as the fixed-point algorithm using kurtosis. So the algorithm can be further simplified as

$$w \leftarrow E\{zg(w^T z)\} - E\{g'(w^T z)\}w \qquad (13)$$

Step wise procedure for Fast ICA Negentropy [7]

1. Center the data to make its mean zero.
2. Whiten the data to give z.
3. Choose an initial vector w of unit norm.
4. Let $w \leftarrow E\{zg(w^T z)\} - E\{g'(w^T z)\}w$, where g is defined as
   $g_1(y) = \tanh(y)$
   $g_2(y) = y\exp(-y^2/2)$
5. Let $w \leftarrow w / \|w\|$
6. If not converged, go back to step 4.

C. Negentropy based gradient algorithm[8]

A simple gradient algorithm can be derived as, Taking the gradient of the approximation of negentropy with respect to w, and taking the normalization $E\{(w^T z)^2\} = \|w\|^2 = I$ into account, we can obtain the following algorithm,

$$\Delta w \propto \gamma E\{zg(w^T z)\} \qquad (14)$$

$$w \leftarrow w / \|w\| \qquad (15)$$

Where $\gamma = [E\{G(w^T z)\} - E\{G(v)\}]$, v being any Gaussian random variable with zero mean and unit varience. The normalization is necessary to project *w* on the unit sphere to keep the variance of $w^T z$ constant.

The parameter γ, which gives the algorithm a kind of "self-adaptation" quality, can be easily estimated as follows

$$\Delta\gamma \propto [E\{G(w^T z)\} - E\{G(v)\}] - \gamma \qquad (16)$$

Step wise procedure for gradient negentropy:

1. Center the data to make its mean zero.
2. Whiten the data to give z.
3. Choose an initial vector *w* of unit norm, and an initial value for γ.
4. Update $\Delta w \propto \gamma z g(w^T z)$, where g is defined as in above algorithm.
5. Normalize $w \leftarrow w / \|w\|$
6. If the sign of γ is not known a priori, update $\Delta\gamma \propto [E\{G(w^T z)\} - E\{G(v)\}] - \gamma$.

If not converged, go back to step 4.

D. Deflamationary orthogonalisation

By the above process we will estimate only one independent component [9] and to estimate all the components we have to run the process several times which is not reliable one so we use an algorithm known as deflamationary orthogonalisation which works on the property of orthogonalisation. Orthogonality is described as Non overlapping or uncorrelated. So, by this property we will find out the orthogonal demixing matrices and with these matrices we will estimate the corresponding independent components.

Deflamationary orthogonalisataion means finding *w* matrix which are orthogonal to each other. After estimating the *w* matrix using one unit algorithm for the first time, we have to run the whole one unit algorithm for estimating the other *w* matrix which is orthogonal to the first estimated *w* matrix.

III. SIMULATION

A. Results

In this simulation two source signals Male and Female voices which are recorded from external sources are used. Then the signal 'S' is produced by adding the two source signals. Now this signal is multiplied with random matrix to get a mixed signal 'X'. The whitened signal is obtained when the mixed signal is done through the whitening process. The sample length of mixed signal X and estimated independent components are both of same order in the simulation.

Fig. 2 and Fig. 3 are the two source signals (male and female voices respectively). Fig. 4 is the mixed signal X, Fig. 5 is the whitened signal. Now the demixing matrix is found by using any one of the one unit algorithm as explained above.

After the completion of the one unit algorithm we get one of the source signal as separated signal.

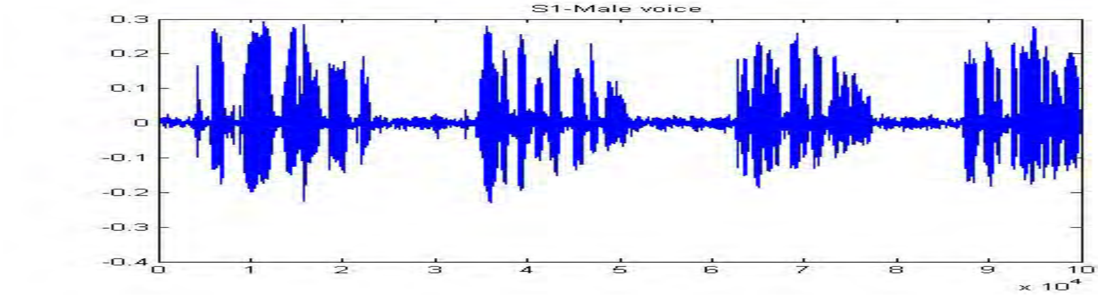

Fig. 2. S1-Male voice signal

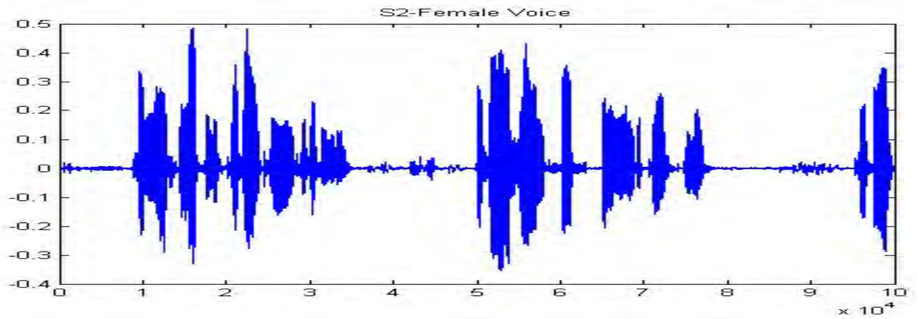

Fig. 3. S2-Female voice signal

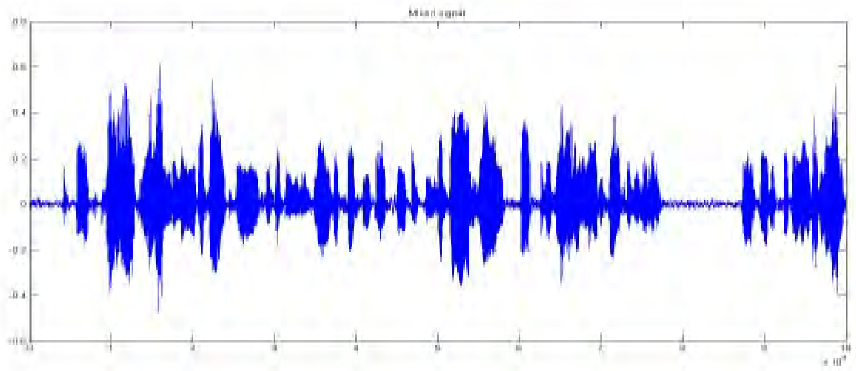

Fig. 4. S- Mixed voice signal

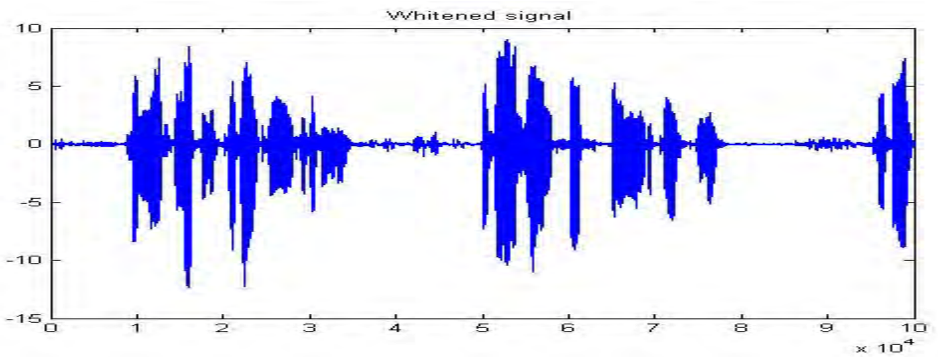

Fig. 5. Y- Whitened voice signal

2

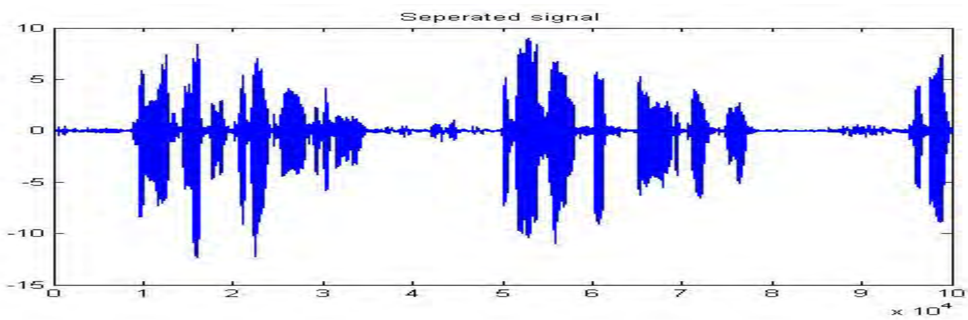

Fig. 6. S2 Separated Female voice signal from Mixed voice signal

By using deflamationary orthogonalisation S1 separated male voice signals are estimated after one unit algorithm

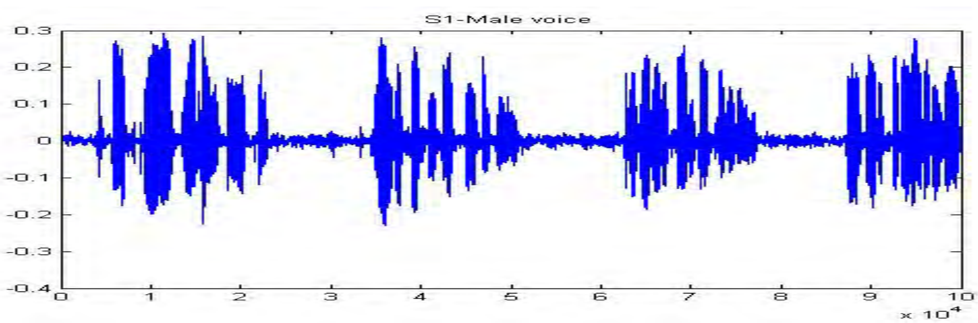

Fig. 7. Separated signal male voice signal from mixed voice signal

In the next simulation two standard signals are used as source signals, such as chirp and gong. Then the signal ‘S’ is produced by adding the two source signals. Now this signal is multiplied with random matrix to get a mixed signal ‘X’. The whitened signal is obtained when the mixed signal is done through the whitening process. The sample length of mixed signal X and estimated independent components are both of same order in the simulation.

Fig.8 and Fig. 9, are the two source signals (chirp and gong signals respectively). Fig.10 is the mixed signal X. Now the demixing matrix is found by using any one of the one unit algorithm as explained above. After the completion of the one unit algorithm we get one of the source signal as separated signal which as shown in Fig.11 and Fig.12.

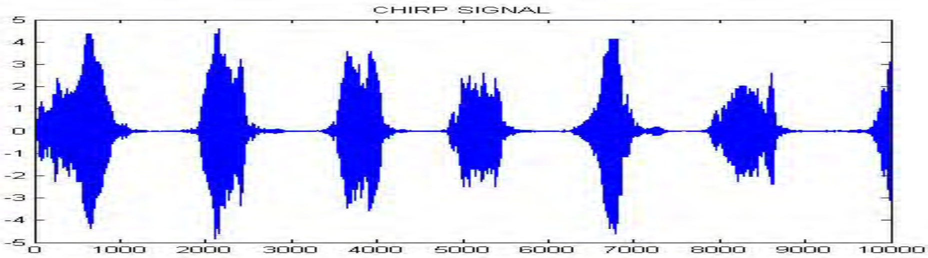

Fig .8 P1-Standard Chirp signal

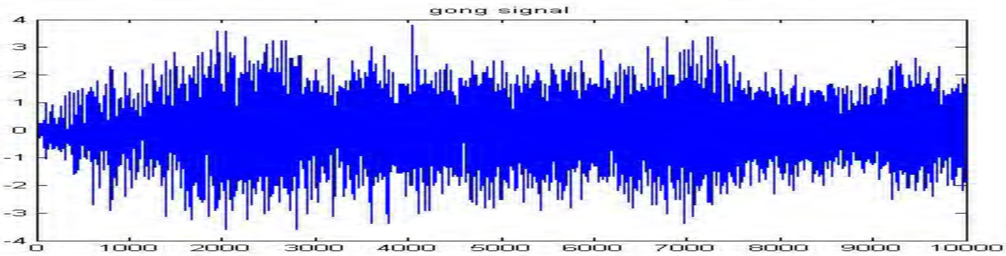

Fig. 9 P2-Standard Gong signal

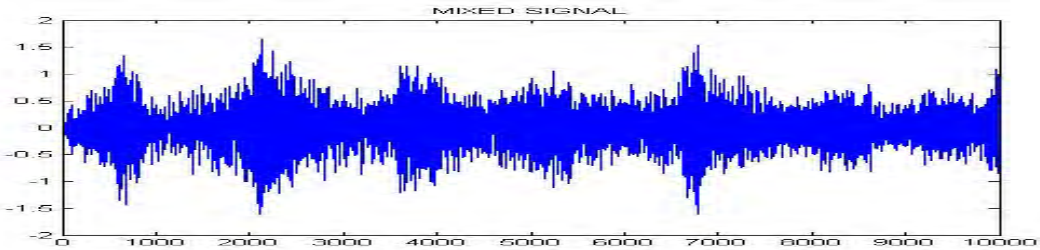

Fig. 10. P- Mixed of Chirp and Gong signal

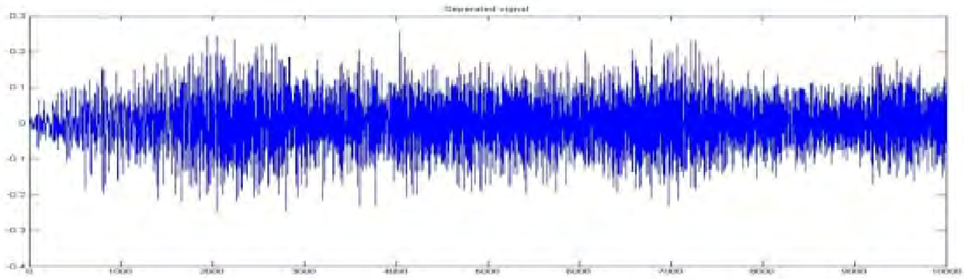

Fig. 11. P2 Separated Gong signal from Mixed signal P

By deflamationary orthogonalisation other signals are estimated after one unit algorithm.

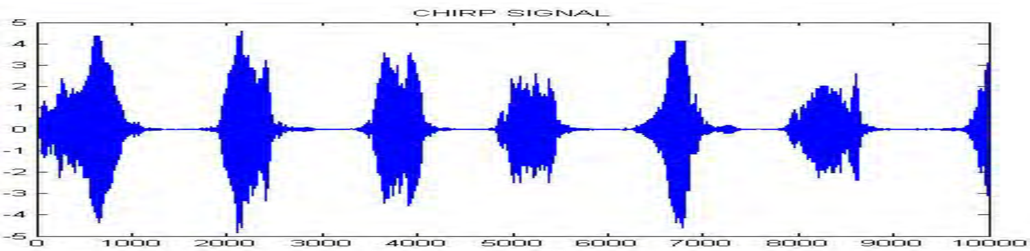

Fig. 12. P1 Separated Chirp signal from Mixed signal P

B. Comparison between Fast ICA and Gradient Negentropy

Here we have calculated the execution time and the amount of error signal in terms of correlation coefficient present in the separated signal (i.e., the amount of the other source signal present in the 'separated signal' when one source signal is separated) for many number of observations and took an average of the observations and compared for Gradient and Fast ICA algorithms [10]

Correlation coefficient can be calculated using the equation

$$Corrcoef\,(X,Y) = Co\,\mathrm{var}\,iance(X,Y)\,/\,\sqrt{Co\,\mathrm{var}\,iance(X,X)\,Co\,\mathrm{var}\,iance(Y,Y)} \qquad (17)$$

When the correlation coefficient reaches 1 then the two signals are highly correlated. When the value of correlation coefficient reaches nearly zero then there is no correlation between the two signals.

Algorithm for negentropy	**Male voice (S1) separated from the mixture S**			**Female voice (S2) separated from the mixture S**		
	Correlation Coefficients between S1&S2	**Average Execution time (sec)**	**No of iterations**	**Correlation Coefficients between S1&S2**	**Average Execution time (sec)**	**No of iterations**
FAST ICA	0.0064	0.245	7	0.0069	0.252	6
GRADIENT	0.0009	0.673	12	0.0011	0.661	11

Table I. Performance of Male and Female voice separation in Fast ICA and Gradient Algorithm

Algorithm for negentropy	**Chirp signal (P1) separated from the mixture P**			**Gong signal (P2) separated from the mixture P**		
	Correlation Coefficients between P1&P2	**Average Execution time (sec)**	**No of iterations**	**Correlation Coefficients between P1&P2**	**Average Execution time (sec)**	**No of iterations**
FAST ICA	0.0039	0.142	6	0.0042	0.165	5
GRADIENT	0.0011	0.673	10	0.0021	0.595	10

Table II. Performance of Chirp and Gong signal separation in Fast ICA and Gradient Algorithm

From the above Table I and Table II we observed the Fast ICA provides better execution time compared to gradient with minimum no of iteration. Gradient ICA provides lesser values for correlation coefficients, which indicates that there is no correlation between separated signal with other signal

IV. CONCLUSION

This paper shows that Gradient based negentropy algorithm provides higher efficiency in separating speech signals compared with Fast ICA based negentropy algorithm. Fast ICA needs less execution time as compared to gradient based negentropy with minimum number of iterations.

REFERENCES

[1] Haykin, Simon, and Zhe Chen. "The cocktail party problem." Neural computation 17.9 (2005): 1875-1902.
[2] Hyvärinen, A., J. Karhunen and E. Oja," Independent Component Analysis." New York: Wiley,(2001) pp: 165-202.
[3] P. Comon, "Independent Component Analysis-A new concept?" Signal Processing, vol. 36, pp. 287-314, 1994.
[4] Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." Neural networks 13.4 (2000): 411-430.
[5] Hyvarinen, Aapo. "Fast and robust fixed-point algorithms for independent component analysis." IEEE Transactions on Neural Networks 10.3 (1999): 626-634.
[6] T. Chien and B.C. Chen, "A new independent component analysis for speech recognition and separation." IEEE Trans. Speech Audio Process. 14. 4 (2006): 1245-1254.
[7] K. Mohanaprasad and P. Arulmozhivarman. "Comparison of Independent component analysis techniques for Acoustic Echo Cancellation during Double Talk scenario." Australian Journal of Basic and Applied Sciences. 7. 4 (2013): 108-113.
[8] R. Ganesh, and K. Dinesh , " An overview of independent component analysis and its application" Informatica., 2011: pp: 63-81.
[9] H.M.M. Joho and G. Moschytz, " Combined blind / non blind source separation based on the natural gradient," IEEE signal process. Lett. 8. 8 (2001): 236-238.
[10] S. Miyabe, T. Takatani, H. Saruwatari, K.Shikano, and Y. Tatekura. "Barge-in and noise-free spoken dialogue interface based on sound field control and semi-blind source separation." In proc. Eur. Signal Process. Conf., 2007:232-236.