Comparison of Cubic and Fibonacci Activation Functions in Speech Signal Seperation using Independent Component Analysis

James P. Chibole, Heywood Ouma and Edward Ndungu

Abstract – The cubic activation function has been used extensively in audio signal separation algorithms. The intention of this paper is to measure the quality of separation for this activation function and compare its performance to that of the Fibonacci. The setup entails the use of the Natural Gradient Algorithm (NGA) to separate two mixed signals into their original components using the Independent Component Analysis (ICA). The NGA used is formulated using instantaneous Blind Signal Processing (BSP). The design uses a 2 x 2 Multiple Input Multiple Output (MIMO) system to accept the two speech signals, mix them and separate them to retain their original form or their filtered version. Two activation functions; the Cubic and Fibonacci are used interchangeably. The results are compared by analyzing the magnitude-squared coherence of the input and separated signals. The results show that Fibonacci is best suited to separate speech signals when applied in NGA than the cubic activation function.

Keywords - Blind Signal Processing (BSS), Independent Component Analysis (ICA), Natural Gradient Algorithm (NGA), Activation Functions, magnitude-Squared Coherence.

I. INTRODUCTION

IN signal processing, multiple signals are often mixed and later separated in various applications. In most of these applications, that include multiple biomedical, multiple antennas and multiple microphones, sensors have to be used to get the correct data [1]. Such systems are best described using Multiple-Input Multiple-Output (MIMO) system models.

In Blind Source Separation (BSS) the challenge is to recover the source signals from the observed mixed signals [2]. Blindness means that neither the sources nor the mixing system are known. Separation can be based on the theoretically limiting but practically feasible assumption that the sources are statistically independent [3]. The ICA algorithm used in this paper is the Natural Gradient Algorithm (NGA), which is applied to produce the best score function for separating the mixtures to solve the BSS problem.

James P. Chibole. Jomo Kenyatta University of Agriculture and Technology. Email:jpchibole@gmil.com
Heywood A. Ouma, University of Nairobi.
Email: habsaloms@gmail.com
Edward Ndungu, Jomo Kenyatta University of Agriculture and Technology.
Email: ndunguen@gmail.com

II. BSS MATHEMATICAL BACKGROUND

Suppose that there are N sources $s_1, s_2, \dots, s_N$ and receivers which each get some mixture of the input signals. Assuming also that each received signal x_j is a linear combination of the inputs through unknown coefficients h_{ij}, then the input would be [3]

$$x_j = \sum_{i=1}^{N} h_{ij} s_i \quad (1)$$

The simplest instant of this MIMO problem is when $N = 2$ and there are 2 system outputs. Equation (1) thus simplifies to

$$\begin{aligned} x_1 &= h_{11}s_1 + h_{12}s_2 \\ x_2 &= h_{21}s_1 + h_{22}s_2 \end{aligned} \quad (2)$$

From Equation (2), we can see that if the coefficients h_{ij} are known, then the problem simplifies to that of solving the system of linear algebraic equations. However in BSS, these coefficients are unknown. Further, increased complexity arises from the lack of knowledge of the input signals, thus *blind signals*. To find a solution to the foregoing problem, it is important to consider the formulation of BSS problem and also analyze the main conditions that should hold in order to efficiently separate the two signals. Equation (2) can be represented in matrix form as

$$X = HS \quad (3)$$

where S is the source vector, H is the mixing matrix and X is the mixed signal vector.

The mixing scheme is shown in Fig. 1

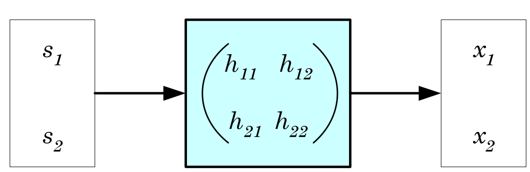


Fig. 1 Mixing Scheme for Two Input Signals.

In this problem, the signals are assumed to be statistically independent. The mixing of the signals is also assumed to be instantaneous, so that the obtained mixtures are the weighted sums of the input signals without any time delays.

In the scheme, an unknown mixing matrix H is applied to the input signal vector S to give rise to the mixed signal vector S. The task is then to determine the de-mixing matrix W that would result in the best estimate Y of S, the original signal vector. Mathematically we have that

$$Y = WX \tag{4}$$

The de-mixing matrix W was therefore first computed. Since $Y = WX = WHS$ mathematically speaking, obtaining perfect separation of the mixed signals requires that the de-mixing matrix be such that $W = H^{-1}$. The de-mixing scheme is illustrated in Fig.2.

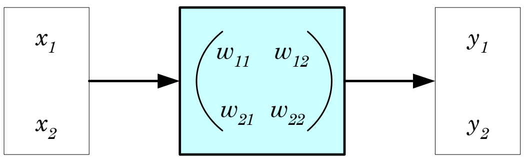


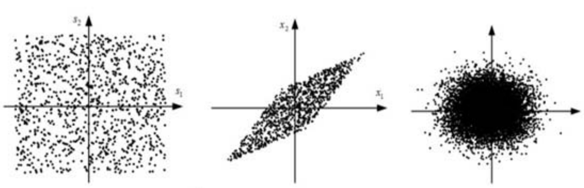
Fig. VII De-Mixing Scheme

Evidently, that the efficiency of BSS is highest when $W = H^{-1}$. However, the value of H is unknown, meaning the relationship cannot be used in the separation process. Nevertheless it is theoretically correct and any perfect separation achieved implies that the relationship has been achieved. This is predicated on the estimated signals being statistically independent, which means that the joint probability density function (PDF) of two estimated signals is equal to the product of the marginal PDFs (equation (5)).

$$p(y_1, y_2) = p(y_1)p(y_2) \tag{5}$$

III. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis (ICA) uses the assumption of the statistical independence of the input signals. This means that the value of any one input signal does not depend on the values of any other signal. Fig. 1 gives an illustration of the PDF transformation before and after mixing for uniform and normal distribution. Fig. 3(a) is the joint PDF from two uniform distributions. Fig. 3(b) is the joint PDF after application of the mixing matrix $A = \begin{pmatrix} 2 & 3 \\ 4 & 2 \end{pmatrix}$. Fig. 3(c) is the Normal distribution, which shows separation is impossible for such signals.



(a) (b) (c)

Fig. 3: Joint PDFs from Two Uniform Distributions

Apparently, after mixing the resulting signals became dependent. It can be seen from Fig. 3(b) that the values of x_1 and x_2 depend on each other.

The goal of the ICA is to attain independence, which means the transformation from joint PDF in Fig. 3(b) to joint PDF on Fig. 3(b). We can see that such transformation can be observed as scaling and rotation of the joint distribution. One of the most important thing to note here is that the ICA does not work with Gaussian signals whose distribution is given in Fig. 3(c). The joint PDF of multivariate normal distribution is symmetric, thus the rotation of the joint PDF cannot be estimated.

Before the application of the ICA, some preprocessing is needed. This involves centering and whitening or sphering of the data. Such data manipulations make two mixed signals uncorrelated, making the covariance matrix diagonal. Since the covariance matrix of white noise is diagonal with unit variances, normalizing the variances in the covariance matrix to unity through “whitening” achieves a semblance of white noise distribution. Centering and whitening are described by Equation (6) and Equation (7), respectively.

$$x_i = x_i - E(x_i) \tag{6}$$

$$X = \frac{X}{\sqrt{E(XX^T)}} \tag{7}$$

The preprocessing simplifies the ICA problem to the estimation of a single parameter; the rotation angle of the joint PDF. The joint PDF of the uniform distributed data before and after preprocessing is shown in Fig. 4.

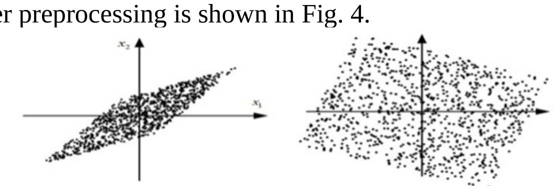


Fig. 4: Effect of Centering and Whitening on the PDF

ICA has two important components; the Objective Function and the Optimization Algorithm. The Objective Function determines the statistical independence of $\boldsymbol{y}$, and the robustness and infirmity of ICA depends on its selection. The Optimization Algorithm adjusts the Objective Function until it attains stability. The speed of convergence and stability of ICA depends on the algorithm chosen. Parameswaran *et al* **[1]** for example reported separation of three speech signals using ICA employing fastICA algorithm and the tanh activation function. ICA methods are best suited for signals and input systems that are termed as blind covered under blind signal separation (BSS).

IV. THE NATURAL GRADIENT ALGORITHM

One disadvantage of stochastic gradient optimization methods is that they suffer from slow convergence because of the inherent statistical correlation found in most real-world signals in the parameter updates. This traditional natural gradient algorithm adopts fixed-step-size; the choice of step size directly affects the convergence speed and steady state performance [2]. Gauss-Newton methods can be used to counter the performance drawbacks of these schemes [3], but they are expensive when it comes to computation and they often suffer from numerical problems if their implementation is poorly done. It is important to have an algorithm that is not only simply and robust of the stochastic gradient but which achieves convergence performance independent of any statistical dependencies [3]. To elaborate and as we shall see, the NGA does not necessarily depend on the mixing matrix, other than the initial conditions.

The natural [4] [5] (or relative [6]) gradient methods have proved to be useful techniques in solving the problem of iterative estimations. In its simplest form, the natural gradient algorithm is a modified gradient search whereby its original gradient search direction changes according to the local Riemannian structure. The Riemannian metric tensor for the values of the parameters is used to measure the degree to which the small variations in the parameters in different directions in the parameters space changes in accordance to the statistical characteristics of the parameters [3]. The gradient search direction is adjusted by the inverse of the Riemannian metric tensor, which results in the natural search direction. Evidence that the natural gradient gives asymptotically efficient performance in the Fisher sense for linear estimation, multilayer perception learning and blind source separation found in [4].

A) Signal Separation using Spatial Independence Algorithms

In this category of BSS-ICA algorithms, each $s_i(k)$ is assumed to be statistically independent of $s_j(k)$ for $1 \leq i \leq 0 < j \leq n$. In addition, the marginal probability density function (mPDF) of at least $(m-1)$ of the sources must be non-Gaussian. Taken together, these two assumptions imply that the joint PDF (jPDF) of $\boldsymbol{s}(k)$ is of the form

$$p_s(s_1, \dots, s_m) = p_{s_1}(s_1)\, x\, p_{s_2}(s_2)\, x \dots p_{s_m}(s_m) \qquad (8)$$

where $p_{s_i}(s_i)\,x$ – is the mPDF of $s_i(k)$, cannot be the Gaussian kernel for more than one value of i. Evidently, that the efficiency of BSS is highest when $W = A^{-1}$. Such condition holds as long as the estimated signals are statistically independent, which means that the joint probability density function (jPDF) of two estimated signals is equal to the product of their marginal PDFs.

$$p(s_1, s_2) = p(s_1)\,p(s_2) \qquad (9)$$

Equations (8) and Equations (9) are the same, only difference is that the second uses only two signals with simplified variables.

The non-Gaussianity of the sources is because of certain identifiability conditions that need to be satisfied for any BSS-ICA formulation to work properly. BSS methods that use this formulation depend on some knowledge of the lower-order or higher-order amplitude statistics of the source signal to perform the separation [7]. Algorithms for BSS-ICA that use spatial independence are classified into two: (i) those that use density matching of the sources, and (ii) those that use contrast function optimization. This paper explores the Natural Gradient Algorithm that uses density matching of the sources.

The basic idea behind density matching can be stated as follows: adjust $\boldsymbol{W}(k)$ so that the jPDF of $\boldsymbol{y}k$), denoted as $p_{\boldsymbol{y}}(\boldsymbol{y})$is as close as possible to some model distribution $\bar{p}_{\boldsymbol{y}}(\boldsymbol{y})$. Although several formulations to this type of density-match approach can be developed [8] [4]. Cardoso has proved that all these formulations can be unified using the Kullback-Leibler divergence measure [9]:

$$KL(p_y \parallel \bar{p}_y) = \int p_y(y) log\left(\frac{p_y(y)}{\bar{p}_y(y)}\right) dy \qquad (10)$$

here $p_y(y)$ and $\bar{p}_y(y)$ are the actual and model distribution, respectively of the output signal vector. Evidently, that $KL(p_y \parallel \bar{p}_y) \geq 0$ and is equal to zero only if $p(y) = \bar{p}(y)$. Therefore, it can be used as objective function. Equation (13) measures the "distance" between $p_y(y)$ and $\bar{p}_y(y)$, although this measure is asymmetric. The choice of $\bar{p}_y(y)$ is governed by the assumption on, and a priori knowledge of $\boldsymbol{s}(k)$. If all $s_i(k)$ are identically distributed, a reasonable choice is

$$\bar{p}_y(y) = C \prod_{i=1}^{m} p_s(y_i) \qquad (11)$$

where C is an integration constant chosen such that $\bar{p}_y(y)$ integrates to unity. This choice of density model yields a maximum-likelihood (ML) estimates of the de-mixing matrix $\boldsymbol{W}(k)$ for the given signal statistics. Alternatively, by estimating the mPDF $\bar{p}_{\boldsymbol{y}_i}(y_i)$ for the current $\boldsymbol{W}(k)$ and setting

$$\bar{p}_y(y) = C \prod_{i=1}^{m} \bar{p}_{\boldsymbol{y}_i}(y_{\boldsymbol{i}}) \qquad (12)$$

One obtains a minimum mutual information approach for BSS [9].

Once a cost function has been chosen, any locally convergent optimization procedure can be used to adjust the elements of $\boldsymbol{W}(k)$. The cost function is actually the activation function that directs the algorithm to attain optimization levels.

V. ACTIVATION FUNCTIONS

For effective convergence and separation to be realized, the algorithm used must employ an activation (nonlinear) function that best approximates the type of input signals. This means that for a mixture of sub-Gaussian mixtures then an activation that gives negative (-1) PDF must be used and a mixture that is made up of super-Gaussian signals, an activation function that gives positive (1) PDF must be used.

Theoretically, the form of activation function $\varphi(y)$ plays an important role in the success of the algorithm. The ideal form for $\varphi(y)$ is the commulative density function (cdf) of the distribution of the independent sources. In practice, however, if $\varphi(y)$ is a sigmoid function, the learning rule reduces to that proposed in [10]. The algorithm is then limited to separating sources with super-Gaussian distribution. Girolami and FyFe [11] used $\varphi(y) = +\tan(y) - y$ for sub-Gaussian signals and

$\varphi(y) = +\tan(y) - y$ for super-Gaussian signals. Then they simply provide a means of calculating the kurtosis of signals and use either of the two if the signals are sub-Gaussian or super-Gaussian. However, they cannot effectively separate signals that are a combination of super- and sub-Gaussian mixtures.

This paper uses the *Cubic Activation Function*

$$\varphi(y) = y^3 \tag{13}$$

and the *Fibonacci Activation Function*

$$\varphi(y) = \frac{\sqrt{5-1}}{2}y^3 + \frac{3-\sqrt{3}}{2}y^5 \tag{14}$$

each separately with the NGA for the mixture of the speech signals.

VI. SUPER- AND SUB-GAUSSIAN SIGNALS

There are two types of non-Gaussian signals. The two types are known by various names such as super-Gaussian and sub-Gaussian or "planty kurtotic" and "lepto kurtotic" respectively [12]. Super- and sub-Gaussian densities are represented in a number of forms.

A) Sub-Gaussian Sources

For sub-Gaussian density, a symmetrical form is adopted as follows

$$p(\mu_i) = \frac{1}{2}\left(N(\mu, \sigma^2) + N(-\mu, \sigma^2)\right) \tag{15}$$

where $N(\mu, \sigma^2)$ is the normal density with mean μ and variance σ^2 [12]. Signals with sub-Gaussian PDF have a wide distributed function as illustrated in Fig. 5.

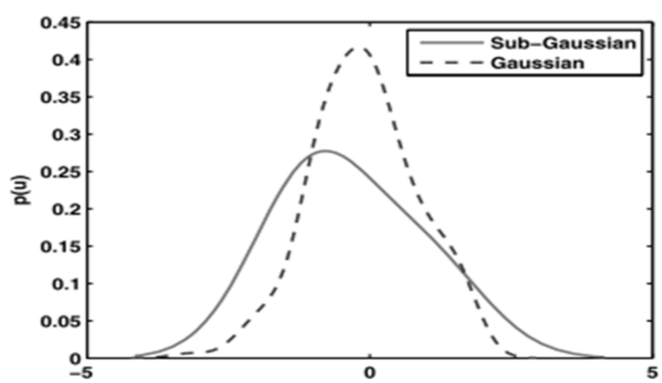


Fig. 5: PDF of a sub-Gaussian signal **[12]**

A saw-tooth signal, music signal and white noise signal are typical sub-Gaussian sources [13]. The sub-Gaussian signals have PDF that are not as peaky as those of Gaussian signal counterparts are.

B) Super-Gaussian Sources

For a super-Gaussian density of speech signal, the Laplacian density is used and represented as:

$$p(\mu_i) = \frac{1}{\sqrt{2}\sigma}e^{-\frac{\sqrt{2}|\mu_i|}{\sigma}} \tag{16}$$

If the PDF of the sub- and super-Gaussian functions are determined as of (2.32) and (2.33) respectively. The values of PDF for a super-Gaussian signal are clustered around zero. A speech signal is a typical example for a super-Gaussian source [12]. Fig. 2.9 is a representation of a typical super-Gaussian (speech) signal.

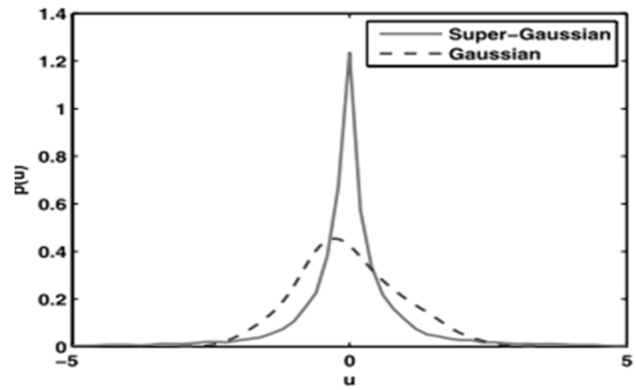


Fig. 6: PDF of a super-Gaussian (speech) signal **[12]**

From Fig. 6, it is also clear that the super-Gaussian signals contain PDFs that are more peaky than that of Gaussian signals.

The non-linear function $\varphi(\mu_i)$ that can be used in confirming the sub-Gaussian and super-Gaussian signals can be represented as:

$$\varphi(\mu_i; k_i) = \begin{cases} \sqrt{2}sign(\mu_i) & for\ k_i = 1;\ super - Gaussian \\ \mu_i - \tanh(\mu_i) & for\ k_i = -1;\ \ sub - Gaussian \end{cases} \tag{17}$$

where sign (μ_i) gives 1 when μ_i is positive and -1 when μ_i is negative [12]. Therefore, the nonlinearity (activation)function is represented as $\varphi(\mu_i; k_i)$ where k_i is 1 for super-Gaussian function and -1 for sub-Gaussian function. The switching condition for k_i between the sub- and super-Gaussian distribution is determined according to the sign of the kurtosis of estimated source $(\mu_i; k_i) = 1$ for positive kurtosis and $k_i =$ -1 for negative kurtosis [12].

It is important to note that the paper does not investigate whether the mixed signals are super- or sub-Gaussian. Intensive studies already show that speech signals are super-Gaussian while music signals are sub-Gaussian. Research dedicated to testing whether the non-Gaussian signals are super- or sub-Gaussian can be found at [12] [14].

VII. MAGNITUDE-SQUARED COHERENCE

The magnitude-squared coherence between two signals $s(n)$ and $y(n)$ is measured using Welch's averaged modified periodogram method. The magnitude-squared coherence estimate is a function of the frequency and the quotient output is a real value between 0 and 1, which indicate how close y resembles s at each frequency ω [15]. The magnitude-squared coherence is a function of the power spectral densities $P_{ss}(\omega)$ and $P_{yy}(\omega)$, of s and y, and the cross power spectral density $P_{sy}(\omega)$ of s and y. The magnitude-squared coherence equation is shown below (Equation 10)

$$C_{sy}\omega = \frac{|P_{sy}(\omega)|^2}{P_{ss}(\omega)P_{yy}(\omega)} \qquad (18)$$

where s and y must be of the same length [15]. For real s and y, the MATLAB function for measuring magnitude-squared coherence, *mscohere* returns a one-sided coherence estimate. For complex s and y, it returns a two-sided estimate. An example of a MATLAB generated magnitude-squared coherence between an input signal s and and its corresponding output y is shown in Fig. 7:

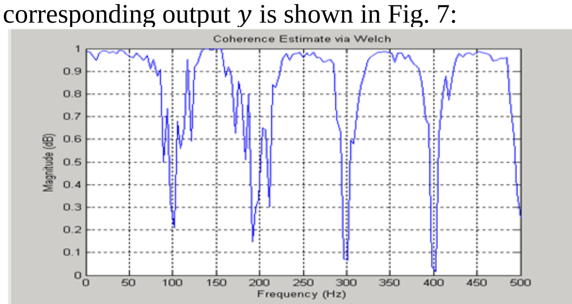


Fig. 2: Measure of magnitude-squared coherence **[15]**.

If the *mscohere* is such that it operates on only a single record, the functions returns all ones at all frequencies indicating a straight line and revealing that the signals are similar – ideal situation.

VIII. METHODOLOGY

In this paper, the problem is formulated to depict the mixing and separation of two speakers as illustrated in Fig..5.

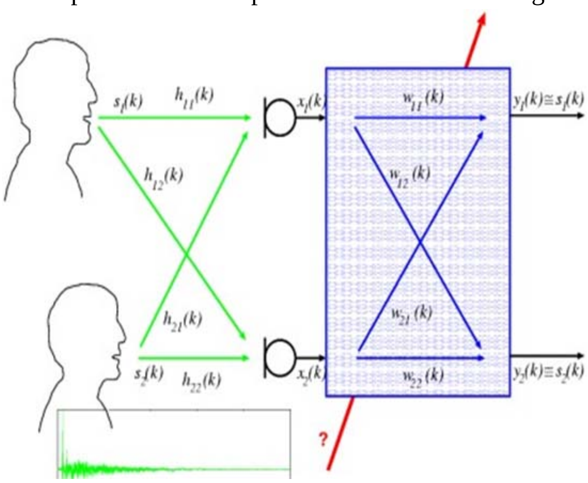
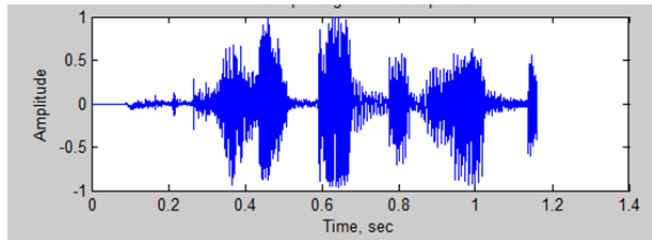


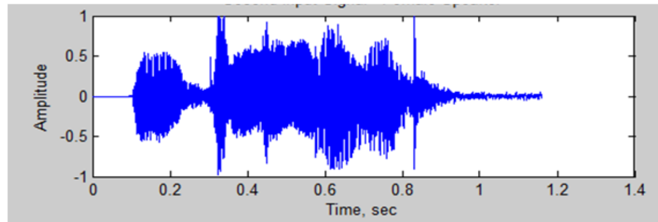
Fig. 8 A 2×2 MIMO System for Two Speech Signals

Two human voices were recorded using the internal microphones of the computer and audio files stored as .wav files in the same fold that the MATLAB program is kept. The male speech speaks the words "I study Matlab"; while the second voice is of a female speaking the words "I am Junior". The pair is then used as input to the system. The male speech represents the source signal s_1 while the female speech represents s_2.

The waveforms of the two input speeches is shown in Fig.



Male Speech



Female Speech

Fig. 9: Waveforms of the input signals

The mixing matrix H is applied to the two input signals resulting in two mixtures x_1 and x_2. The dimensions of the mixing matrix H are:

$$A = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \qquad (3.1)$$

The mixture is then normalized and then whitened. The NGA is applied to separate the two mixtures using the Cubic and Fibonacci activation functions, separately. After the separation the input signal is compared to its corresponding estimated signals using the measure of Magnitude Squared Coherence.

IX. RESULTS

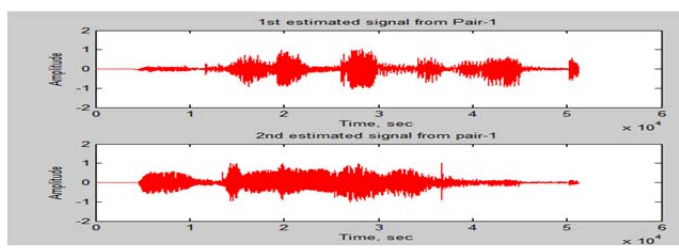


Fig. 11 Separated signal waveforms using cubic activation

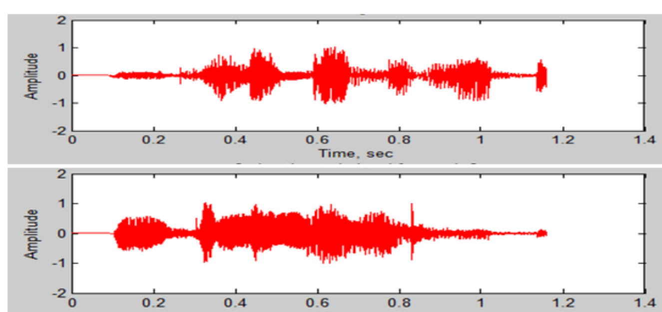


Fig. 11 Separated signals using FAF

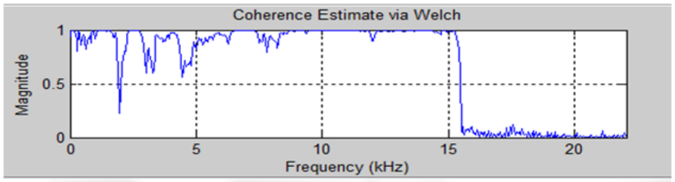


Fig. 12 Magnitude-Squared Coherence for the male speech using Cubic Activation Function.

Measure of Magnitude-Squared Coherence for the Female Speech using Cubic Activation Function

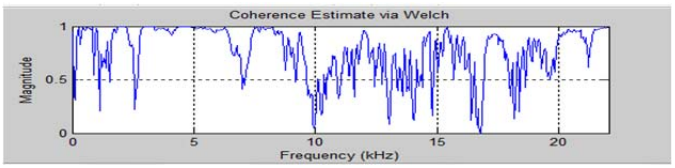


Fig. 13 Magnitude-Squared Coherence for the female speech using Cubic Activation Function

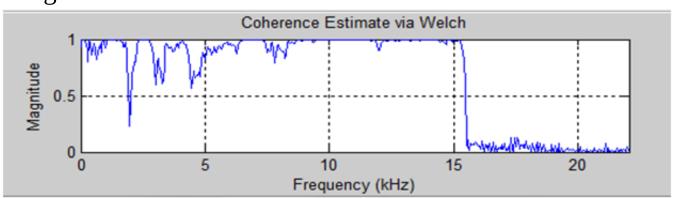


Fig. 14 Magnitude-Squared Coherence for the male speech using cubic activation function

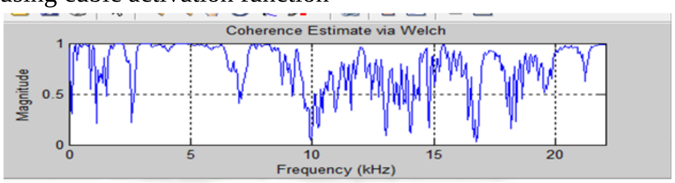


Fig. 15 Magnitude-Squared Coherence for the male speech using FAF

X. DATA ANALYSIS

The sound generated by the separated signals approximates well the input signals, with minimal interference between the two original signals being observed. However, the waveforms of the separated male and female speech using Fibonacci Activation function are wider than those of the Cubic Activation function. The sound of the generated signals is louder in the FAF case than in the Cubic Activation Function. Although the measure of magnitude-squared coherence, show almost similar results for the two activation, the combined output of the waveforms, and sound generated show promising results for the Fibonacci case.

XI. CONCLUSION

The effectiveness of the ICA-NGA using the Fibonacci Activation Function (FAF) has been demonstrated. Good quality separation of two speech signals was achieved. The quality was both to the ear and statistically, taking into account the fact that the precise amplitudes of the blind original signals cannot be reconstructed. In evaluating the results it must be borne in mind that the estimated signals y_1 and y_2 may represent s_2 and s_1 respectively or vice versa, and so to avoid the inherent confusion, the separated signals as simply termed as "separated signals" without invoking the symbols signals $y_1$and y_2, as the order is not known. The results of algorithm performance have proven the high efficiency of ICA-NGA with FAF.

REFERENCES

[1] S. Parameswaran, K. C. Finitha and A. A. Sama, "Blind Source Separation for Speech Signals," *IIST*, pp. 4 - 9, 2010.

[2] J. Y. Peng and Y. Yang, "Blind Source Separation Based on Improved Natural Gradient Algorithm," in *Proceedings of the 8th World Congress on Intelligent Control and Automation, July 6-9*, Jinan, China, 2010.

[3] S.-i. Amari, S. C. Douglas, A. Cichocki and Y. H. Howard, "Multichannel Blind Deconvolution and Equalization using the Natural Gradient," in *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Paris, 1997.

[4] S. Amari, A. Cichocki and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 752-763, 1995.

[5] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, p. 251–276, 1998.

[6] J. -F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3017-30130, 1996.

[7] D. C. Scott, "7 Blind Signal Seperation and Blind Deconvolution," New York, CRC Press LLC, 2002.

[8] A. J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129-1159, 1995.

[9] J. F. Cardoso, "Blind signal separation: statistical principles," in *Proceedings of the IEEE*, Cambridge, MA, 1998.

[10] A. Bell and T. Sejnowski, "An Information Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, vol. 7, pp. 1129-1159, 1995.

[11] M. Girolami and C. FyFe, "Extraction of Independent Signal Sources using a Deflationary Exploratory Projection Pursuit Network with Lateral Inhibition," *I.E.E Proceedings on Vision, Image and Signal Processing Journal.*.

[12] G. R. Naik, "Measure of Quality of Sound Seperation for Sub and Super Gaussian Audio Mixtures," *INFORMATICA*, vol. 23, no. 4, pp. 581-599, 2012.

[13] R. Cristescu, T. Ristaniemi, J. Joutsensalo and J. Karhunem, "Cdma delay estimation using fast ica algorithm.," in *Procceedings of IEEE International Symposium on Personl, Indoor and Mobile Communication*, California, 2000.

[14] C. S. Rai and S. Yogesh, "Source distribution models for blind source seperation.," *Neurocomputing*, vol. 57, pp. 501-502, 2004.

[15] Mathsworks, "Spectral Analysis," MATLAB, 2014. [Online].

Available: http://www.mathworks.com/help/signal/ug/spectral-analysis.html. [Accessed 10 July 2014].

[16] S. Amari and A. Cichocki, "Adaptive Blind Signal Processing – Neural network Approaches," *Proceeding of the IEEE,* vol. 86, no. 10, 1998.

[17] P. Lecumberri, M. Gómez and A. Carlosena, "Multichannel Blind Deconvolution of Impulsive Signals," in *Proceedings of the 13 European Signal Processing Conference*, Antalya, Turkey, 2005.

[18] S. C. Douglas and M. Gupta, Convolutive blind source separation for audio signals, S. Maino, P. Okuyayi, J. Tawwoi, N. OscN, n. JiouN, p. Juz and K. Yusufkeukeu, Eds., New York: New York Springer, 2007.

[19] F. Vrins and M. Verleysen, "On the entropy minimization of a linear mixture of variables for source separation," *UCL, Electrical Engineering Department, Microelectronics Laboratory. Signal Processing* , vol. 85, pp. 1029 -1044, 2005.

[20] A. Cichocki and S.-i. Amari, Adaptive Blind Signal and Image Processing: learning Algorithms and Applications, New York: John Wiley & Sons, 2002.

[22] S. Amari, A. Cichocki and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems,* pp. 752-763, 1995.

[23] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation.*

[24] P. Lecumberri, M. Gómez and A. Carlosena, "Multichannel Blind Deconvolution of Impulsive Signals," in *Proceedings of the 13 European Signal Processing Conference*, Antalya, 2005.

[25] N. Kofidisa, A. Margarisa, K. Diamantarasb and M. Roumeliotis, "Blind system identification: instantaneous mixtures of n sources," *International Journal of Computer Mathematics,* vol. 85, no. 9, p. 1333–1340, 2008.

[26] A. Jalobeanu, B. L. Feraud and J. Zerubia, "An adaptive Gaussian model for satellite image deblurring.," *IEEE Trans. Image Processing,* vol. 13, pp. 613-621, 2004.

[27] Y. Hujun and I. Hussain, "Independent Component Analysis and non gaussianity for blind image deconvolution and bluring," *Integrated Computer-Aided Engineering,* vol. 15, pp. 219-228, 2008.

[28] C. D. Meyer, Matrix Analysis and Applied Linear Algebra, Cambridge, UK: Cambridge University Press, 2000.

[29] G. R. Naik and D. K. Kumar, "An overview of Independent Component Analysis and its Application," *Informatica,* vol. 35, pp. 63-81, 2011.

[30] Daqarta for DOS, "Fourier Transform Theory," [Online]. Available:http://www.daqarta.com/0t0ifft1.htm. [Accessed 27 August 2013].

[31] H. Mathis, T. P. von Hoff and M. Joho, "Blind Separation of Signals with Mixed Kurtosis Signs Using Threshold Activation Functions," *IEEE Trans. on Neural Networks,* vol. 12, no. 3, pp. 618-624, 2001.

[32] L. Q. Zhang, A. Cichoki and S. Amari, "Self-adaptive blind source separation based on activation functions adaptation," *IEEE Trans. Neural Networks,* vol. 15, no. 2, pp. 233-244, 2004.

[33] L. Zhang, S. Amari and A. Cichocki, "Equi-convergence Algorithm for blind separation of sources with arbitrary distributions," in *6th Int'l Work Conf. on Artificial and Natural Neural Networks*, 2001.

[34] S. Bose and B. Friedlander, "On the direct estimation of the Wiener beamformer using second and fourth-order cumulants," in *Proceedings of the 31st Asilomar Conference on Signals, Systems, and Computing*, Pacific Grove, CA, 1997.

[35] S. C. Douglas, "Combined subspace tracking, prewhitening, and contrast optimization for noisy blind signal separation," in *Proceedings of the IEEE International Workshop Indep. Compon. Anal. Signal Sep.*, Helsinki, Finland, 2000.

[36] K. Matsuoka, M. Ohya and M. Kawamoto, "A neural net for blind separation of nonstationary signal sources," *Neural Networks,* vol. 8, no. 3, pp. 411-419, 1995.

[37] M. K. Tsatsanis and C. Kweon, "Blind source separation of non-stationary sources using second-order statistics," in *Proceedings of the 32th Asilomar Conference on Signals, Systems, and Computing, Pacific*, Pacific Grove, CA, 1998.

[38] T. M. Cover and J. A. Thomas, , Elements of Information Theory,, New York: JohnWiley & Sons , 1991.

[39] S. Amari, T. P. Chen and A. Cichocki, "Stability analysis of learning algorithms for blind source separation," *Neural Networks,* vol. 10, no. 8, p. 1345–1351, 1997.

[40] D. G. Luenberger, Linear and Nonlinear Programming, 2nd. ed, Reading, MA: Addison-Wesley, 1984.