Speech Scrambling based on Independent Component Analysis and Particle Swarm Optimization

Nidaa Abbas[1,2] and Jahanshah Kabudian[1]
[1]Computer Engineering and Information Technology Department, University of Razi, Iran
[2]College of Information Technology, University of Babylon, Iraq

Abstract: *The development of communication technologies and the use of computer networks has led that the data is vulnerable to the violation. For this reason this paper proposed scrambling algorithm based on the Independent Component Analysis (ICA), and the descrambling process was achieved on Particle Swarm Optimization (PSO) to resolve this problem. In the scrambling algorithm, the one speech signals segmented into two types, two and three. It then used the mixing process to result the scrambling of speech. In the descrambling process, we proposed the kurtosis and negative entropies as fitness function. The simulation results indicate that the scrambled speech has no residual intelligibility, and the descrambled speech quality is satisfactory. The performance of scrambling algorithm has been tested on four metrics Signal to Noise Ratio (SNR), Perceptual Evaluation of Speech Quality and Mean Opinion Score (PESQ-MOS), Linear Predictive Coding (LPC) and itakura-saito distance. Many input speech signal of sampling frequency 16 kHz was tested for two genders male and female.*

Keywords: *ICA, itakura-saito distance, LPC, PSO, speech scrambling, SNR*

Received July 6, 2015; accepted August 16, 2015

1. Introduction

With the rapid developments of information and communications technology, it became necessary multiple violations of media data protection in many of the real applications. To achieve these demands, it proposed a lot of scheme encryption to protect the multimedia data, including that of speech signals, images and videos.

The multimedia encryption scheme categorized into two types: analog and digital. There are different schemes were designed to analog encryption, e.g., permuting of the element, signal masking, frequency shuffling, all of them may be applied in transforming or time domain, or both [9].

The analog encryption is considered the one of the common encryption techniques has been used in speech communication [1]. The analog speech encryption (aka. speech scrambling), works on the samples of speech. The objective of speech scrambling algorithms is to convert the plain speech into unclear signal so that it is hard to decrypt it in the key absence [1].

Lin *et al.* [10, 11, 12, 13, 14, 15, 16]. Proposed Blind Source Separation (BSS) for the image and speech encryption. The main concept of works, they use in encryption process the underdetermined mixing matrix, and to get the unconditionally secure, the key signals were generated to fulfil a necessary condition for the proposed method.

Yang *et al.* [24] employing the idea of asynchronous of encryption for the plain text which are mixed mutually firstly and then mixed with the ciphers. Sheu *et al.* [21] proposed speech communication using BSS. The proposed system is utilized from the one time pad encryption avoiding the drawbacks in key interchange. Also, it is shown that the proposed system is resistant against classical attacks.

Guo and Lin [4] gives the analysis of the correlation of speech signals with key signals, and then achieve speech decryption by benefit from the calculation of the correlation.

The problem in Independent Component Analysis (ICA) algorithms is having a random behavior, i.e., the results of these algorithms are different depending on different initial conditions [19]. The finding of local optimal point for the algorithm of gradient descent optimization depends on the initial point where the search starts. These gradient descent algorithms rarely find the global optimal point, especially in high dimensional optimization problems.

To resolve the above mentioned problem, many researchers have been working on the ICA and Particle Swarm Optimization (PSO) to overcome this problem. In [3] employ the high and second order correlation coefficients to minimization of the cost function of independence of signals to separate the linear BSS using PSO. Xie and Jiang [23] proposed the ICA with PSO to overcome the convergence of local optimal solution in non-convex ICA optimization objective

function, such convergence led to the almost valuable ICs may be unreachable.

Nian *et al*. [18] suggested employing of improved PSO to resolve of problem of convergence to the classical searching scheme of ICA which is based on gradient algorithm. The algorithm of PSO uses variable inertia weight, which is based on evolution speed and fitness function

Igual *et al*. [7] addressed the weakening of contrast function which is responsible of measures the independence of the components. The authors used the stochastic global PSO algorithm to resolve the optimization problem. The authors in another work [6] are solving the optimization problem, through utilize from the ICA contrast function based on the MI employing the stochastic global PSO algorithm.

The incorporation of the ICA and PSO was proposed in real time noise cancellation for mobile radio system [2]. The author uses this combination to separate the speech and noise signals.

In this proposed work the scrambling process achieved by the ICA, and the descrambling process was carried out with the PSO. In the scrambling algorithm, the one speech signals segmented into two types, two and three. It then used the mixing process to result the scrambling of speech. In the descrambling process, the PSO algorithm used the kurtosis and negative entropies as fitness function. Many metrics were used to evaluate the performance of proposed work.

The remainder of this paper is organized as follows. Section 2 gives the overview of the background of theories which includes the fundamental of speech scrambling, ICA, and PSO. The details of the proposed scheme are reported in section 3. In section 4, the results of experiments to evaluate the proposed method are described. Finally, the conclusion of the proposed method is given in section 6.

2. Background of Theories

2.1. Speech Scrambling

Digital encryption at first employ digitization of the input speech signal, the output of this process is digitized signal which is compressed to make a bit stream at a proper bit rate. The bit stream is encrypted and transmitted over the channel using the modem. Typically the permutation is achieved into two types, permutation of speech segments usually employs in speech scrambling algorithm in all types of domain, time, frequency or time-frequency or permutation of transform coefficients of each speech block. Due to give the low residual intelligibility, the time-frequency algorithms considered as a much interest algorithms.

The combination of speech scrambling and descrambling in cryptography has acted as a controlling part in secure communications. The scrambling term has been used to represent the process of encryption to defend voice communications, whether is collected in analog or digital means. This process is done in time, frequency domain and two dimensional as well. The metric of fast execution of algorithms is a very interesting area for engineers and researchers over the last two decades beside the redundancy and security [17].

The degree of residual intelligibility is determining the level of security of the scrambled signal. The highest level of security is the lower the residual ineligibility and vice versa. The great importance of the one of the demanding attributes in the design of a good secure communication system is the quality of retrieved speech without losing information. Yet trade-off has to be made between security level and recovered speech signal quality.

2.2. Independent Component Analysis

ICA is an unsupervised statistical method used to separate a mixture of signals into independent sources. Without prior knowledge about the source signals and mixing matrix, the ICA can be used for signal separation only assuming that the signals are statistically independent. The basic linear model relates the unobservable source signal and the observed mixtures:

$$x(t)=As(t) \quad (1)$$

Where $s(t)=[s1(t),\ldots,s_m(t)]^T$ is a $m\times1$ column vector containing the source signals, similarly vector $x(t)$ grouping the m observed signals, A is a $m\times m$ matrix of unknown mixing coefficients, and t is the time index. The goal of ICA is to find the unmixing matrix W (the inverse of A) that will give y, the best possible approximation of s:

$$y(t)=Wx(t) \quad (2)$$

Where $y(t)=[y_1(t),\ldots,y_n(t)]^T$ is an estimate of $s(t)$ and W is the separating matrix with dimension of $n\times n$ Many ICA algorithms can be decomposed into two steps; in the first one, apply second order statistics for decoration (whitening step). The imposing of higher order statistics to estimate the orthogonal matrix that require the independence was achieved in the second step [19].

The result of the approximation of independence hypothesis is the estimation of sources which is transformed into an optimization problem described by the contrast function that is minimum when the estimated sources are as independent as possible. To overcome from the trapped in local optimization problems of gradient descent existing in traditional approaches, the PSO was employed to solve this ICA problem [7].

2.2.1. Preprocessing of ICA

- *Centering:* is the subtracting the mean of the

observation vector

$$x' = x - E[x] \tag{3}$$

Subsequently the mean vector can be added to the estimates of the sources

$$s = s' + A^{-1}E[x] \tag{4}$$

- *Whitening:* For the sake of computational efficiency, the mixed signals *x* is whitened first. To obtain the components of the observations uncorrelated and have unit variance, the linear transform was applied as in Equation 5:

$$\tilde{x} = Wx \Rightarrow E[\tilde{x}\tilde{x}^T] = I \tag{5}$$

This can be fulfilled through principal components

$$\tilde{x} = \Lambda D^{-1/2}\Lambda^T x \tag{6}$$

Where the columns of Λ and the diagonal of *D* are eigenvector and eigen values of $E[xx^T]$, respectively. The benefit of whitening makes the mixing matrix orthogonal. Which has the advantage of halving the number of parameters that need to be estimated, since an orthogonal matrix only has *n* (*n-1)/2* free parameters [19].

2.2.2. Contrast Functions in ICA

- *Negative Entropy:* The entropy of a variable is a measure of randomness. The entropy for a discrete-valued variable is defined, as:

$$H(Y) = -\sum_i P(Y = a_i)\log P(Y = a_i) \tag{7}$$

Where the *H* is the entropy of mixed signals and is estimates of source speech signals. A uniform and Gaussian variable has the largest entropy among discrete and continuous valued variables respectively. To measure the non-Gaussianity, the negentropy was used, which is the differential measurement of entropy relative to a Gaussian.

$$J(y) = H(y_G) - H(y) \tag{8}$$

Here y_G is a Gaussian variable with same variance as *y*. That *J*(*y*) is always non negative, and equal to zero for a Gaussian. Negentropy is characterized as statistically robust, but computationally intensive, since it requires density estimation, possibly non-parametric. Since the estimation of negentropy is difficult, one typically uses approximations proposed in [10], which have the form

$$J(y) \propto [E\{G(y)\} - E\{G(v)\}]^2 \tag{9}$$

Where *v* and *y* are a Gaussian random vectors with zero mean and unit variance and *G* is a no quadratic function. Several choices of *G* have been shown to work well, including [10].

$$\left.\begin{aligned} G_1(u) &= \frac{1}{a_1}\log \cosh(a_1 u) \quad 1 \le a_1 \le 2 \\ G_2(u) &= -\exp(-u^2/2) \end{aligned}\right\} \tag{10}$$

- *Kurtosis*

Kurtosis is the measure of non-Gaussianity, which is defined as the fourth order cummulant

$$kurt(y) = E[y^4] - 3(E[y^2])^2 \tag{11}$$

Kurtosis can be positive, negative or zero. When the kurtosis is zero, the variable is Gaussian, and positive the variable is supergaussian, and finally if the kurtosis is negative, the variable is subgaussian. The advantage of using Kurtosis being computationally cheap

2.3. Particle Swarm Optimization

Kennedy and Eberhart [8] developed the PSO after the study the behavior of bird flocking. This optimization is related to evolution-inspired problem solving techniques such as genetic algorithms.

As aforementioned, PSO imitated the bird flocking behaviours. Assume the following script: There is only one piece of food in searched area, and a group of birds is randomly search food in this area. What's the best strategy to find the food? Although, all the birds do not know where the food is. The influential one is to go after the bird which is nearest to the food.

The single solution in the search space is named particle. These particles were evaluated by the fitness function to be optimized; the flying of particles is controlled by velocities. The particles fly through the problem space by according to the current optimum particles.

The PSO algorithm starts by initialize the random particles (solutions) and then updated the generations (swarm) looking for the optimal position according to its own experience as well as to the experience of its neighbourhood. Two factors describe a particular status in the n-dimensional search space: Its velocity and position which are updated according to the following equations at the nth iteration:

$$v_i(t+1) = wv_i(t) + c_1 r_1(t)(pbest_i(t) - x_i(t)) + c_2 r_2(t)(gbest_i(t) - x_i(t)) \tag{12}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{13}$$

Where the v is the particle velocity, x is represents the position vector of the particle, *pbest* and *gbest* are the personal and present best position of particle consecutively, *w* is the inertia weight, c_1 and c_2 are two acceleration constants, called cognitive and social parameters, respectively, and r_1 and r_2 are two random functions in the range [0, 1] [5, 22].

3. Proposed Scheme

To overcome of the multiple violations of media data protection in many of the real applications, it proposed a lot of scheme encryption to protect the multimedia data one of them our proposed system.

In this section, we describe a speech scrambling system based on ICA and PSO which is implemented using Matlab. Algorithm below and the block diagram of mathematical model as shown in Figure 1 describe the proposed system.

1. Read source speech signal.
2. Split the signal into two types of segments (2 or 3) where each segment has an equal sample, and is independent from each other which assigned to s.
3. Change in the mean and variance of standard normal distribution of mixing matrix *A* to get more unintelligibility.
4. Scrambling process results from (1).
5. Preprocessing process for *x*
   a. Centering from (3)
   b. Whitening from (6)
6. Initialization the parameters of PSO, population, velocity, position of the particle, c_1,c_2, r_1 and r_2.
7. Centers the separated signal to the individual, and the separated signals and whitening.
8. Individual fitness value calculated accordance with the (11) and (9) save the global optimum.
9. Follow the (12) and (13) to update the position and velocity of the particle.
10. Recalculate fitness values
11. If the maximum number of iterations, the algorithm terminates; otherwise, turn go to step 7.

Evaluate the results depends on some evaluation metrics (SNR, LPC, PESQ-MOS and Itakura-Saito distance).

3.1. Measurements Criteria

To evaluate the validity of speech scrambling system, many metrics are used like SNR, LPC distance, PESQ-MOS, and Itakura-Saito distance (IS).

- *Signal to Noise Ratio (SNR):* is computed as

$$SNR = 10 log_{10} \frac{\sum_{n=-\infty}^{\infty} s^2(n)}{\sum_{n=-\infty}^{\infty} (s(n) - \hat{s}(n))^2} (dB) \quad (14)$$

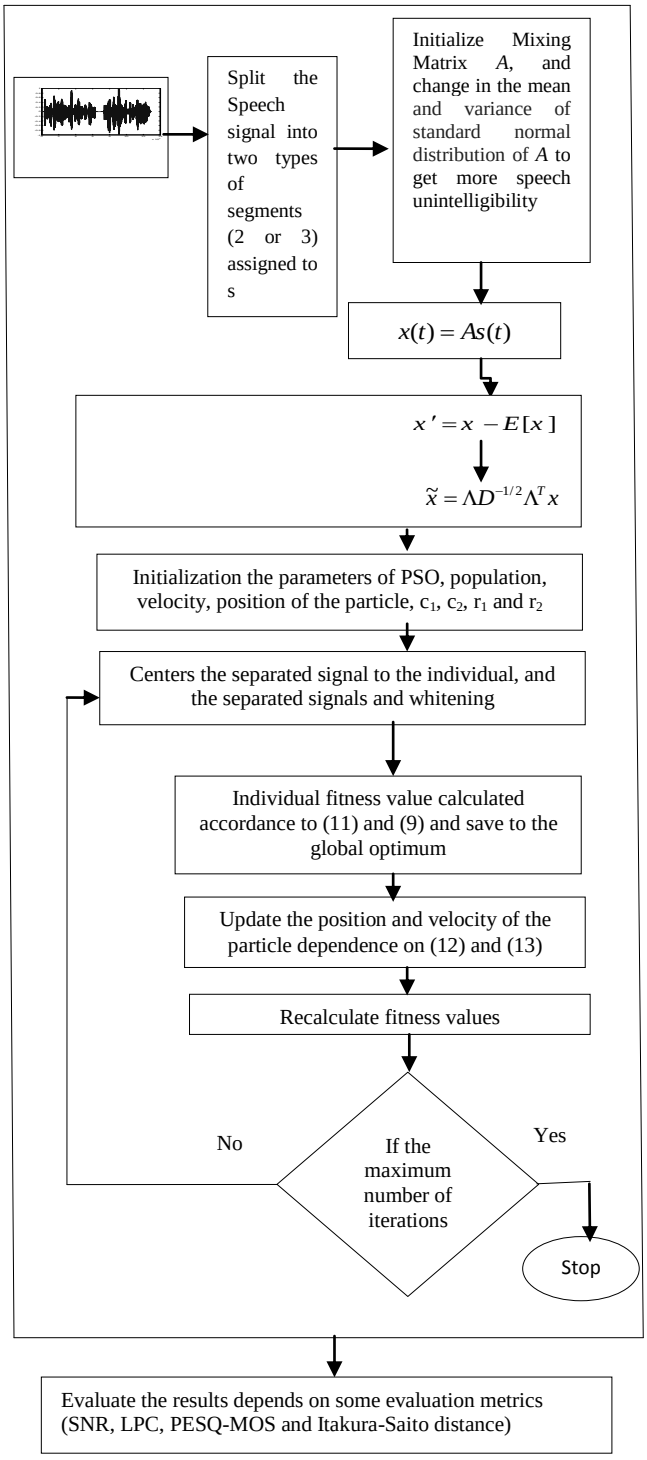


Figure 1. Block diagram mathematical model of the proposed system.

Here n is the number of samples, $s(n)$ is the amplitude of the input speech signal and $\hat{s}(n)$ is the amplitude of recovered speech signal. A high SNR ($\gg$ 1) indicates high precision data, while a low SNR indicates noise contaminated data.

- *Linear Predictive Coding Distance (LPC):* is defined as [20]

$$d_{lpc}(a_e, a_c) = log\left(\frac{a_e R_c a_e^T}{a_c R_c a_c^T}\right) \quad (15)$$

Where R_c is the autocorrelation matrix of the clear speech signal, a_c is the LPC vector of the original

speech signal frame and a_e is the LPC vector of the estimated (descrambled) speech signal frame.

- *Itakura-Saito Distance (IS):* is defined as [20]

$$d_{IS}(a_e, a_c) = \frac{\sigma_c^2}{\sigma_e^2}\left(\frac{a_e R_c a_e^T}{a_c R_c a_c^T}\right) + log\left(\frac{\sigma_c^2}{\sigma_e^2}\right) - 1 \quad (16)$$

Where σ_c^2 and σ_e^2 are the LPC gains of the clean and estimated (descrambled) signals, respectively. If the IS distance is low, then the algorithm performs well and otherwise is bad.

- *Perceptual Evaluation of Speech Quality (PESQ-MOS):* The ranges of PESQ-MOS which define by ITU recommendation P.862 from 1.0 (worst) up to 4.5 (best). PESQ simulates a listening test and is optimized to reproduce the average result of all listeners.

3. Simulation Results and Analysis

The experiments on sentences "This is an example of the AT&T natural voice speech engine, it is the most human sounding text to speech engine in the world", was achieved to test the validity of the proposed system. We examine the speech signal, with the variety of samples and genders. Some results that shown in Tables 1, 2, 3, and 4. According to the quantitative evaluation in table 1, the results of SNR reveal that the signal more than the noise for all genders. Also the LPC shows the distance between the original and descrambled speech is very low. The resultant of PESQ-MOS in a range of best as confirmed that the algorithm of scrambling based on PSO is acceptable. The distance of IS show that is low in descrambled speech while is high in scrambled speech. The resultant of Table 2 is satisfied with comparison the results of Table 1.

Table 1. Results of measurement criteria for two segments.

Gender	No. of Samples	SNR in dB	LPC	PESQ-MOS	Itakura-Saito Distance	
					Original to Estimated	Original to Scrambled
F1	129546	1.1839	0.0499	3.8696	3.1568	6.7357
F2	129128	1.2582	0.0555	4.2272	3.0224	6.5215
F3	131328	1.0822	0.0467	3.5532	3.3096	6.4774
M1	127296	1.4330	0.0263	4.2698	2.7864	7.0396
M2	122112	1.3160	0.0957	4.1630	2.9444	6.6754
M3	122026	1.3096	0.0415	4.1018	2.9568	6.4736

Table 2. Results of measurement criteria for three segments.

Gender	No. of Samples	SNR in dB	LPC	PESQ-MOS	Itakura-Saito Distance	
					Original to Estimated	Original to Scrambled
F1	129546	1.1830	0.0489	3.1326	3.1370	6.4786
F2	129128	1.2522	0.0582	3.4559	3.0719	6.9149
F3	131328	1.0771	0.0468	3.2047	3.3158	6.9227
M1	127296	1.4305	0.0267	3.7032	2.8055	6.8658
M2	122112	1.3008	0.0997	3.3176	2.9869	6.8717
M3	122026	1.2897	0.0444	3.1487	3.0311	6.4810

Tables 3 and 4 shows the plot of the original, scrambled and descrambled speech. It is evident from these tables that the scrambled speech of three segments gave the more unintelligibility

Table 3. Plot the original, scrambled and recover speech for two segments.

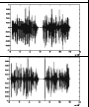
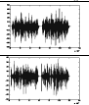
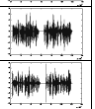
Gender	Original Speech	Scrambled Speech	Recover Speech
F1			
F2			
F3			
M1			
M2			
M3			

Table 4. Plot the original, scrambled and recover speech for three segments.

Gender	Original Speech	Scrambled Speech	Recover Speech
F1			
F2			
F3			
M1			
M2			
M3			

3. Conclusions

This paper presented the speech scrambling based on ICA and PSO. We benefit from PSO in side of separation process using the kurtosis and negative entropies as fitness function. In the scrambling process, the speech signal split into two types two and three segments. We can conclude that the use of three segments of speech signals gives more unintelligibility than two segments. The testing results of the performance of the proposed system using four criteria SNR, LPC, PESQ-MOS and IS distance also show that our proposed system is promising. The results of simulations reveal the no residual intelligibility in scrambled speech, good quality of recovered descrambled speech employing many input speech signal of sampling frequency 16 kHz for two genders male and female.

References

[1] Beker H. and Piper F., *Secure Speech Communications*, Academic Press, 1985.

[2] Bor R., "Real-time Noise Cancellation Using ICA-PSO-PE," M.S. Thesis, Bilkent University,

2012.

[3] Gao Y. and Xie S., "A Blind Source Separation Algorithm Using Particle Swarm Optimization," *in Proceeding of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication*, Shanghai, pp. 297-300, 2004.

[4] Guo D. and Lin Q., "Fast Decryption Utilizing Correlation Calculation for BSS-based Speech Encryption System," *in Proceeding of 6th International Conference on Natural Computation*, Yantai, pp. 1428-1432, 2010.

[5] Hussain I., Khanum A., Abbasi A., and Javed M., "A Novel Approach for Software Architecture Recovery using Particle Swarm Optimization," *The International Arab Journal of Information Technology*, vol. 12, no.1, pp. 32-41, 2015.

[6] Igual J., Ababneh J., Llinares R., and Igual C., "Using Particle Swarm Optimization For Minimizing Mutual Information In Independent Component Analysis," *in Proceeding of 11th International Work-Conference on Artificial Neural Networks*, Torremolinos, pp. 484-491, 2011.

[7] Igual J., Ababneh J., Llinares R., Miro-Borras J., and Zarzoso V., "Solving Independent Component Analysis Contrast Functions with Particle Swarm Optimization," *in Proceeding of International Conference on Artificial Neural Networks*, Thessaloniki, pp. 519-524, 2010.

[8] Kennedy J. and Eberhart R., "Particle Swarm Optimization," *in Proceeding of the IEEE International Conference on Neural Networks*, Perth, pp. 1942-1948, 1995.

[9] Li S., Li C., Lo K., and Chen G., "Cryptanalyzing an Encryption Scheme Based on Blind Source Separation," *IEEE Transactions on Circuits and Systems*, vol. 55, no. 4, pp. 1055-1063, 2008.

[10] Lin Q. and Yin F., "Blind Source Separation Applied To Image Cryptosystems with Dual Encryption," *Electronics Letter*, vol. 38, no. 19, pp. 1092-1094, 2002.

[11] Lin Q. and Yin F., "Image Cryptosystems Based On Blind Source Separation," *in Proceeding of International Conference on Neural Networks and Signal Processing*, Nanjing, pp. 1366-1369, 2003.

[12] Lin Q., Yin F., and Liang H., "A Fast Decryption Algorithm For BSS-Based Image Encryption," *in Proceeding of International Symposium on Neural Networks*, Chengdu, pp. 318-325, 2006.

[13] Lin Q., Yin F., and Liang H., "Blind Source Separation-Based Encryption Of Images And Speeches," *in Proceeding of International Symposium on Neural Networks*, Chongqing, pp. 544-549, 2005.

[14] Lin Q., Yin F., and Zheng Y., "Secure Image Communication Using Blind Source Separation," *in Proceeding of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication*, Shanghai, pp. 261-264, 2004.

[15] Lin Q., Yin F., Mei T., and Liang H., "A Blind Source Separation Based Method for Speech Encryption," *IEEE Transactions on Circuits and Systems*, vol. 53, no. 6, pp. 1320-1328, 2006.

[16] Lin Q., Yin F., Mei T., and Liang H., "A Speech Encryption Algorithm Based On Blind Source Separation," *in Proceeding of International Conference on Communications, Circuits and Systems*, Chengdu, pp. 1013-1017, 2004.

[17] Matsunaga A., Koga K., and Ohkawa M., "An Analog Speech Scrambling System Using The FFT Technique With High-Level Security," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 4, pp. 540-547, 1989.

[18] Nian F., Li W., Sun X., and Li M., "An Improved Particle Swarm Optimization Application to Independent Component Analysis," *Information Engineering and Computer Science*, Wuhan, pp. 1-4, 2009.

[19] Oja E., Hyvarinen A., and Karhunen J., *Independent Component Analysis*, John Wiley and Sons, 2001.

[20] Quackenbush S., Barnwell T., and Clements M., *Objective Measures of Speech Quality*, Prentice-Hall, 1988.

[21] Sheu L., Chiou H., and Chen W., "Semi- One Time Pad Using Blind Source Separation for Speech Encryption," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 5, no. 8, pp. 803-806, 2011.

[22] Shi Y. and Eberhart R., "A Modified Particle Swarm Optimizer," *in Proceeding of the IEEE Congress on Evolutionary Computation*, Anchorage, pp. 69-73, 1998.

[23] Xie L. and Jiang L., "Global Optimal ICA and its Application in Brain MEG Data Analysis," *in Proceeding of Conference Neural Networks and Brain*, Beijing, pp. 353-357, 2005.

[24] Yang Z., Zhou G., Wu Z., and Zhang J., "New Method For Signal Encryption Using Blind Source Separation Based on Subband Decomposition," *Progress in Natural Science*, vol. 18, no. 6, pp.751-755, 2008.

Nidaa Abbas Completed her Doctoral degree from Computer Science Dept. in University of Technology, Iraq. She is a faculty member in the department of Software, IT faculty, University of Babylon. Her research areas include image processing, statistical signal processing, speech scrambling.

Jahanshah Kabudian Completed his doctoral degree from AmirKabir University of Technology, (Tehran PolyTechnic), Iran. He is presently working as Ass. Professor, Department of Computer Engineering and Information Technology, Razi University, Iran. His areas of interest include Digital Signal Processing (DSP), Audio, Speech and Language Processing.