Journal of Engineering and Applied Sciences 11 (4): 862-869, 2016
ISSN: 1816-949X
© Medwell Journals, 2016

Improving Speech Signal Encryption by Using Compression along with Encryption According to Wavelet Transform, Discrete Cosine Transform and Symmetric Permutation Methods

Vahid Hosseinzadeh and Jalil Shirazi
Department of Telecommunication Engineering, Islamic Azad University,
Gonabad Branch, Gonabad, Iran

Abstract: This study,aims at increasing the security level of the speech signal using both the encryption and compression of signals. Initially, it has done an overview of some encryption and compression methods of speech signal. Then, a new signal with an isolation property for high and low frequencies of speech has made by using discrete wavelet and discrete cosine transform and the signal is used for encryption. The new signal has compressed the coefficients on high and low indices; thereby, the possibility of compression is also provided. Results showed that the level of security in encrypted and compressed signal has reasonably increased compared to the previous methods and deciphering has become more difficult.

Key words: Asymmetric permutation, discrete cosine transform, discrete wavelet transform, spectrograph, compression

INTRODUCTION

Now a days, the concept of security has gone far beyond the scope of material health maintenance and is combined with other concepts. The emergence of communication technologies in human life has provided some requirements for new definitions in the field of security. Speech and its transferring tools are considered as requirements for maintaining human's security in the current era. Speech is the first and most simple way of communication for people and the need for security in it is of great importance for privacy protection issues.

So far, methods that are used to secure graph the speech signal are distinguishable in two sections of analog and digital and in two domains of time and frequency.

In this regard, Enache *et al.* (2015) have designed and discussed three methods of encryption in the frequency domain. These methods are reversing frequency coefficients, band shift and permutations of frequency bands. Methods of band shift and permutations of frequency bands are proposed based on a random pattern. In the results, also the method of reversing the frequency coefficients is suggested as the most insufficient method. However, the researchers have suggested the second and third methods as the methods appropriate for encryption. Ehsani and Borujeni (2002) have examined the permutation method of discrete Fourier transform samples. In this method, the sound first is entered into a 12 bit analog-to-digital converter and then it is stored. Framing is done and from each frame the discrete Fourier transform is obtained. The transformation coefficients are cluttered, then inverse transform is taken from each frame's transformation in order to obtain the encrypted signal in the time domain. Another article has proposed an innovative method for encryption in communications under the sea. This method is assigned a specific symbol to each frame (Peyvandi, 2011). First, the sound signal is encoded at low bit rates and then framing is done by indicating 7 bits for each frame. Ultimately, a symbol is assigned to each frame based on a specific table. With regard to the 7 bits frame, 128 symbols have been proposed for all possible states.

In the time domain, Borujeni (2002) has proposed permutation in the time coefficients of speech signal. The author has proposed two methods to carry out the project. The first method includes putting scrambling patterns in a read-only memory or ROM and the second is using a random number generator that acts like a shift register. In the first method, permutation sequence is stored in the memory and then it is applied. In the second method, the appropriate permutation of the random number is extracted and applied according to the random number which is obtained from the mentioned generator.

Encryption in the field of mathematical transformations is the most effective and best method of speech signal encryption which has been investigated and designed by researchers. In the conventional state of these methods, first the speech signal is carried to the transform domain and then the conversion coefficients are

Corresponding Author: Vahid Hosseinzadeh, Department of Telecommunication Engineering, Islamic Azad University, Gonabad Branch, Gonabad, Iran

cluttered. Finally, the signal is returned to the time domain once again and it is declared as the encrypted signal. In the same vein, a speech scrambling system is proposed based on the parallel wavelet conversions in an article (Sadkhan *et al*., 2007). This method is used for different types of wavelet in transformation. The pair-wavelets Db1 with Haar, b2 with Sym2 and b4 with Sym4 are proposed and the results are assessed through qualitative criteria.

Extraction of approximation and detail's coefficients has been applied by using two different wavelets and permutation of these coefficients and the encryption method. Further, the distance between the coefficients is suggested as the criteria for appropriateness of permutation which is cluttered with the original place of these coefficients in the original signal. Alternatively, Dawson (1991) has proposed that initially the fictitious coefficients should be added to the speech signal and then it is carried to the discrete cosine transform domain. After scrambling cosine transform coefficients of the new signal, it is returned to the time domain. Jameel *et al*. (2007) have discussed on another application of discrete cosine transform. First, signals are framed, the discrete cosine transform is obtained and then the coefficients of the transform are cluttered. In the next stage, the scrambled signal is multiplied with the second signal in the same transform domain and it is returned back to the time domain again. Given that multiplication of the second signal in the discrete cosine transform is equivalent to the convolution in the time domain, this operation has been described as a way to make the encryption more difficult because taking inverse convolution is more difficult than doing a simple multiplication. Also, a method is proposed for performing the mathematical transformation operations in a combined form. This method is based on the wavelet transform and discrete cosine transform along with the asymmetric permutation of transform coefficients and it was associated with improvement of speech signal encryption (Hosseinzadeh and Shirazi, 2016). The least level of comprehension in the encrypted signal as well as the lack of possibility of recognizing the encryption method based on the spectrograph is described as the advantages of the mentioned method.

Compression of the speech signal is another area of interest for researchers in the field of speech. This is suggested to facilitate and transform signal's data before doing any operation on it. In this regard, in a research which presenting wavelet transform as a powerful tool in the field of speech and image, a method is proposed on the basis of this transform for compression which is entitled wavelet packet decomposition method (Joseph and Anto, 2011). First, the signal is divided into two sections including, details and approximation by the wavelet transform and both sections again are decomposed into two other sections. This operation has been applied to the sub-sections once again. Then, a selected combination of a number of eightfold sub-sections was obtained which they have made the compressed signal. Again, Najih *et al*. (2003) have considered the wavelet transform as the basis of compression. In this study, after applying wavelet transform, it is utilized a threshold level dependent on the coefficient ratio for compression. The threshold level determines the coefficients that it will remain in the compressed signal. Various wavelets, including Haar, b2, b4, b7 and b10 have been applied for performing simulation. The compressed sound quality is considered as a function of the threshold level in this research.

Sushma and Sandeep (2011) have proposed compression on the basis of the wavelet transform. Sampling at the rate of 8 kHz, framing and applying pre-emphasis filter are the stages prior to the compression. At the stage of signal compression, the energy level of the samples has been the criteria of elimination. This time, the type of wavelet used in the compression is proposed as an important factor and the authors have considered the wavelet db20 as the most appropriate wavelet for compression. Rajesh *et al*. (2011) have benefited various techniques in transformation to achieve the goal of compression. Initially, compression has been predicted at three stages. Decomposition, thresholding and quantization and coding have been proposed respectively at the first, second and third stages. The first stage is conducted based on the wavelet transform and at the second stage the low energy coefficients have reached zero. Finally, Huffman coding is used. At the quantization stage, obtaining Δ has been considered as the criteria for operation. This parameter is defined as equivalent to the difference between the largest and the smallest coefficients divided by the number of coefficients. In simulations, the signal compression stages have been conducted based on the discrete cosine transform, discrete Fourier and discrete wavelet with five different types of wavelets. Results showed that the compression coefficient (obtained by dividing the length of the original signal into the compressed signal) has been around 3.7 in the discrete Fourier transform and an amount around 5.8 and 7.4, respectively in the discrete cosine and wavelet.

Moreno-Alvarado and Martinez-Garcia (2011) have used the discrete cosine transform to remove the unnecessary coefficients so, the number of samples has reached less than what the Nyquist rate requires us to consider it. Trueness of discrete cosine transform and energy compression are two properties that have been emphasized in this study. Chong and Kim (1997) have compared methods of using discrete cosine, wavelet and wavelet packet transform with each other. Results have shown that wavelet packet transform has allocated the best signal to noise ratio and wavelet transform has allocated shortest time for compression.

MATERIALS AND METHODS

A review of conventional methods of securing speech signal: There are two conventional methods for securing the speech signal; signal encryption and steganography. Signal encryption is based on the scrambling the signal or its permutation. Steganography is also hiding the speech signal in another host signal. These methods can be implemented in both frequency and time domains. Strategies that are applied in the domain of analog signal frequency are frequency inversion algorithms, frequency transmission, permutation (scrambling) of frequency bands, permutation of discrete Fourier transform samples and use of a wide spectrum. In the frequency inversion, the frequency samples are arranged from the end to the beginning. In the frequency transmission which can be performed along with a frequency inversion, the frequency spectrum is shifted using a modulation with a carrier frequency. Also, in frequency bands permutations, after framing, the spectrum is divided into sub-bands; then they are irregularly placed side by side. In the wide spectrum method, each sub-band is transmitted to a part of the wide spectrum with a different carrier frequency. The permutation of the Fourier transform samples as well as scrambling the samples are considered as one of the encryption methods in the frequency domain which is implemented after framing and obtaining the discrete Fourier transform by scrambling the transform coefficients.

Methods that are used in the time domain are time inversion algorithm, permutation algorithm in this domain and performing the amplitude coverage. The inversion of time coefficients is obtained by setting time coefficients from the first to the end in each frame. Permutation is also achieved by three methods. Its easiest mode is displacement of the coefficients of a frame. The other mode is achieved by dividing the time frame into several sub-sections which are called segments. In one of these methods, the segments are displaced and the interior elements are permuted in the other one.

Sometimes, time and frequency methods are used simultaneously to increase the security. This kind of usage is called the two-dimensional algorithms of encryption. Methods of this domain are inverting frequency-permutation in the time domain and another is permutation of frequency bands with time scrambling, which are explained above.

Some methods have also been studied in the digital domain for encryption which can mention to algorithm of digital amplitude covering and encryption by voiced coders. In the first method, the amplitude of the digitized signal is added with a random signal. Of course, this seemingly random signal for encrypting has previously been determined. In the second method, coefficients are permutated after being coded.

One of the most widely used encryption methods is the encryption in the domain of transformations. This is the best and safest method of interlacing the signal, which is implemented in discrete transform of the speech signal. Discrete cosine transform is one of the most common transformations. Data compression capabilities and trueness of transform coefficients are considered as the most important advantages of this transformation.

Discrete wavelet transform is also one of the most important mathematical transformations in the signal processing domain. Considering the multi-resolution analysis nature, this transform has reached its position in many processing applications and it is sometimes used as the most powerful tool (Sayyadi, 2008).

Encryption on the basis of wavelet transform, discrete cosine and asymmetric permutation along with compression: In this method of encryption, the speech signal is transformed to the domain of one-level discrete wavelet transform. At the next stage, each component is returned separately to the time domain and then the discrete cosine transform is obtained by the two components (approximation and details) and in this way, permutation of coefficients is providing. Figure 1 shows the stages of performing this method of speech encryption.

The final obtained signal has two properties in itself; first, it has been in the discrete cosine domain and it is real; second, its low and high frequency coefficients in the cosine domain have had more strength in the low and high indices, respectively. In other words, approximation coefficients have been compressed in low indices as a result of discrete cosine transform and the details have also been compressed in high indices (Hosseinzadeh and Shirazi, 2016). Therefore, the resulted signal has coefficients proportional to the low frequencies at the low indices and high frequencies at the high indices. Figure 2 shows that this result is obtained by combining above transformations.

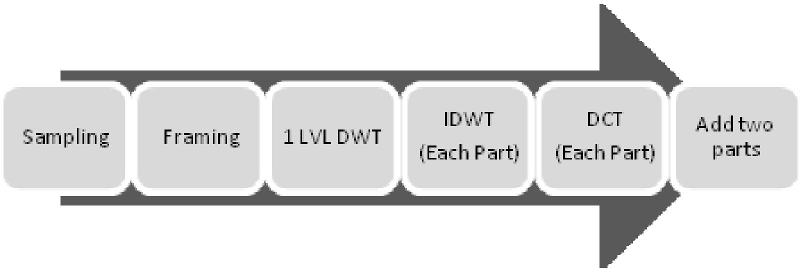

Fig. 1: Stages of making a new signal based on the DWT and DCT (Hosseinzadeh and Shirazi, 2016)

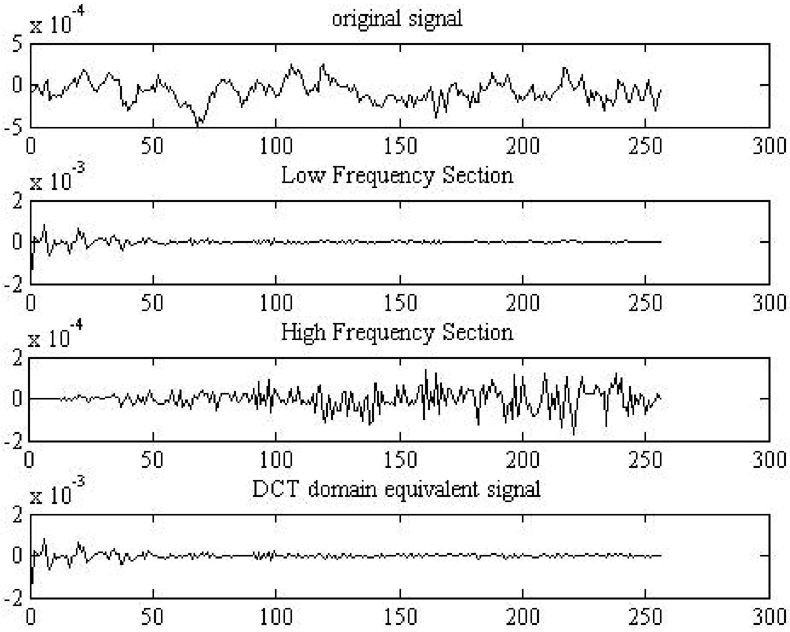

Fig. 2: Stages of making a new signal for a sample frame; the above figure is the original frame; the second and third figure contains the components, respectively low and high frequencies and the figure in the bottom is the sum of the second and third sections (the equivalent signal) (Hosseinzadeh and Shirazi, 2016)

In the encryption method of this paper is proposed that the cosine coefficients be transmitted as the speech signal due to their trueness property and then the permutation of coefficients be performed in the 256 amples frame, between the first 64 samples of the frame related to the low frequency section and the last 64 samples of the frame related to the high frequency section. This issue has made the recognition of the encrypted signal more difficult and it also has reduced the necessary permutations. Figure 3 shows the encryption according to this method. Considering that encryption is done only in sections with high-value coefficients and the original signal data is embedded in this section, it would be possible to eliminate the undervalued sections which are the middle sections of the encrypted signal.

According to Fig. 1, the speech signal which is sampled at a rate of 8 kHz and it has 256 samples in each frame is transformed to the discrete wavelet domain, the wavelet inversion for each component and then the discrete cosine and finally it is encrypted. Its first and last 64 samples which contains signal's valuable data, are kept and its middle 128 samples are eliminated. Two important issues can be fulfilled by doing so. First, the size the encrypted signal is reduced by half; second, it becomes more difficult to guess the encryption method. Figure 4 shows applying the above method on a frame of speech signal in which the numbers are counted from 0-4 in Persian.

Also, Fig. 5 shows the entire speech signal which is encrypted and compressed by this method. Specifically,

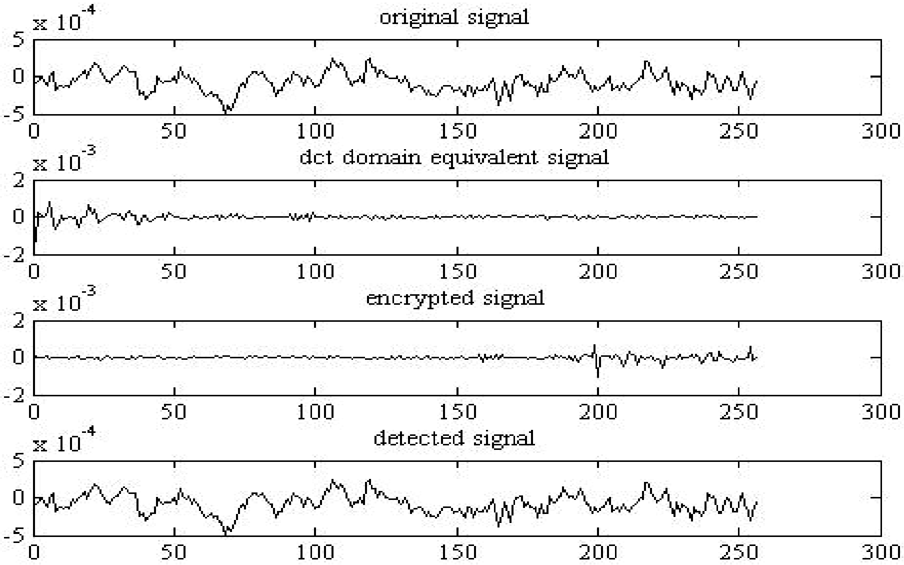

Fig. 3: Performing encryption for a sample frame; the above figure is the original frame, the second is the equivalent frame, the third is the encrypted equivalent frame and the fourth is the revealed frame (Hosseinzadeh and Shirazi, 2016)

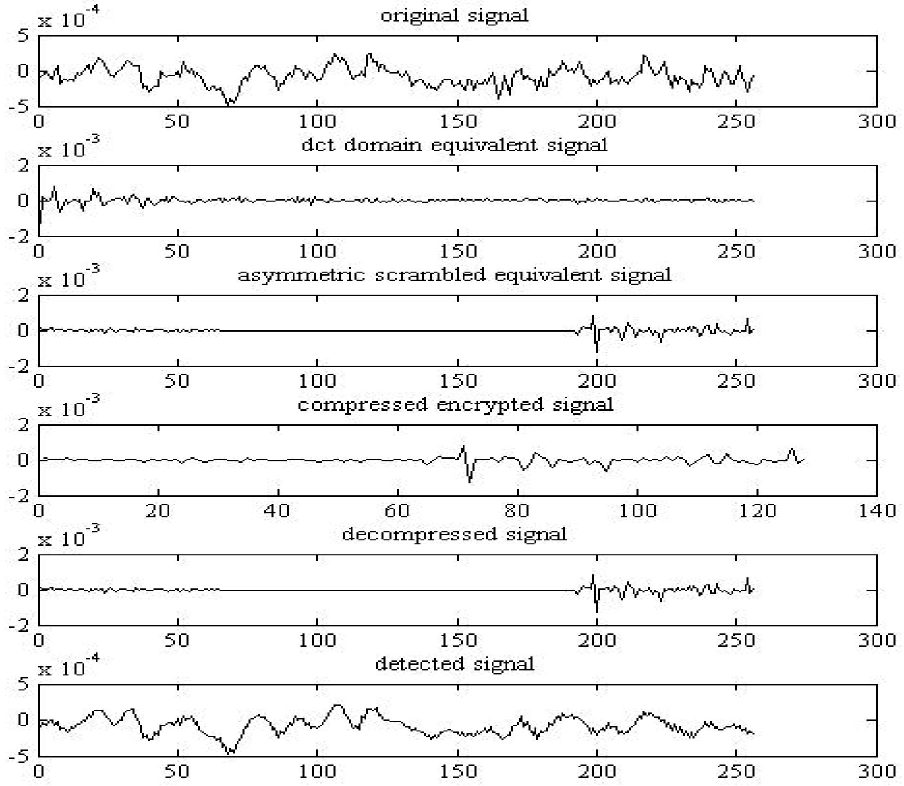

Fig. 4: Stages of compression for a sample frame, the above figure is the original frame, the second is the equivalent frame, the third is the encrypted equivalent frame, the fourth is the frame with the eliminated (compressed) elements, the fifth is the signal retrieved from the compressed signal and the sixth is the signal retrieved in the time domain

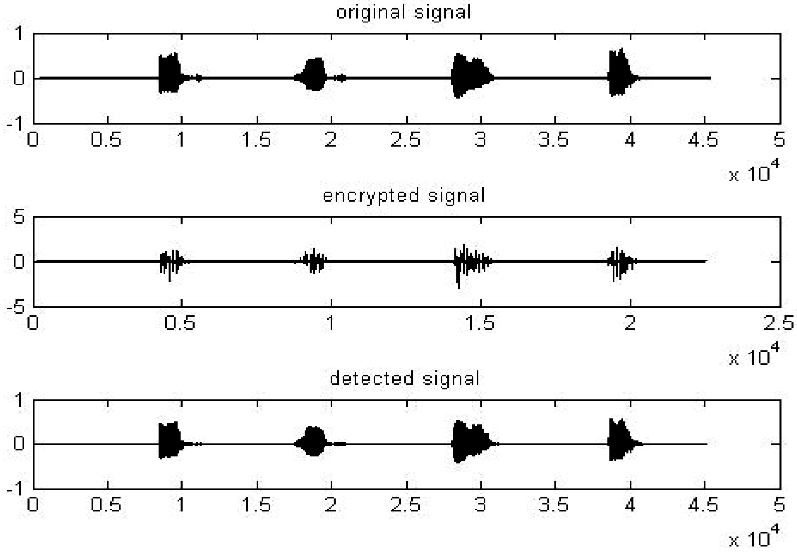

Fig. 5: Stages of performing encryption and compression for the entire sample signal, the above figure is the original signal, the second figure is the encrypted and compressed signal and the third figure is the revealed signal

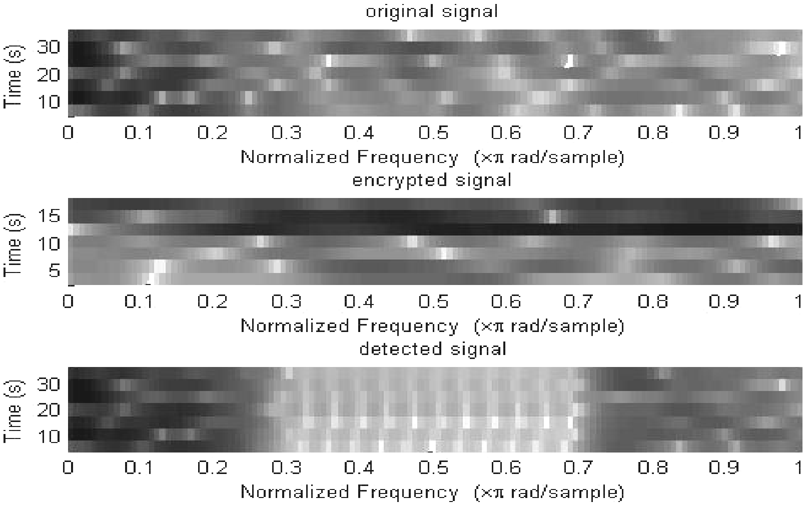

Fig. 6: The above figure is the original signal, the second figure is the encrypted and compressed signal and the third figure is the retrieved signal

in addition to disturbance of the signal's appearance, its time has been changed due to the compression (it has been halved) that it would become the detection more complicated. Observing the spectrum of the encrypted signal in the spectrograph is considered as one of strategies to discover the encryption method. Displacement of strong frequencies and discontinuities within the boundaries of the spectrum can help to guess the method used for encryption. It is evident that scrambling the spectrum irregularly will enhance the quality of encryption and it will make the detection of the applied method more difficult so that it would not be possible to guess the encryption method (Nichols and Lekkas, 2002). Scrambling the speech signal spectrum is another advantage of the present method so that it has made more difficult to guess the encryption method. Figure 6 shows the spectrum of the sample frame. Displacement of strong frequencies in the second image, does not express a clear concept of the applied method. Also, the third image shows that it is possible to extract major strong frequencies from the encrypted signal in the retrieval process.

Table 1: MOS criteria for counting

Language	Comprehension -quality (orig. signal)	Comprehension -quality (encr. signal)	Comprehension (retrieved signal)	Quality (retrieved signal)
P	5	1	4.3	3.7
E	5	1	4.2	3.6

RESULTS AND DISCUSSION

According to the method used for obtaining the new signal and the existence of cosine and wavelet transforms simultaneously in the process of accessing it, the level of signal security has increased and it has made difficult to access the method of making and encrypting signal as well as process inversion (Hosseinzadeh and Shirazi, 2016).

Moreover, adding the compression, i.e., elimination of the unnecessary coefficients; thereby colliding the interval of broadcasting the encrypted sound has increased the difficulty of detecting the sound. Also, for evaluating the remaining comprehension in the signal, considering that comprehension is intuitive and qualitative, a qualitative criterion of MOS (Mean Opinion Score) is used to evaluate the encryption method.

At this stage, 40 speech signal is expressed by 40 persons (men and women) in the Persian including, counting zero to three to evaluate the proposed method. Then, they are encrypted and are tested. Also, the test is repeated for 40 similar speech signals in English. Afterwards, the encrypted signals were heard by 30 persons and they have allocated a score between 1-5 to them for intelligibility and naturalness of the signal. These scores are given in Table 1. According to the mean, encryption has been successful compared to the time and time-frequency methods (Jayant *et al.*, 1983). The comprehension was at a good level and quality number has also obtained an acceptable value.

Also, the above test has been conducted for continuous sentences in English and Persian which were expressed by both a man and a woman. The results were similar to the previous results.

CONCLUSION

In addition to the advantage was mentioned in the previous section compared with the time and time-frequency methods; another advantage of combining encryption method and compression is the uncertainty of the encryption method on the basis of signal spectrum which has increased the complexity of encryption. Compared with the frequency domain methods in which the whole or part of the method is discovered using spectrogram, this is considered as a strong point. Putting this item alongside the combination of wavelet and cosine transforms, it makes the retrieval of the encrypted signal more difficult for a person unfamiliar with the algorithm used in it. On the other hand, the method used in coefficients permutation has reduced the encryption time as well. Moreover, colliding the time of expressing the signal, considering the compression has complicated the speech comprehension more than before.

REFERENCES

Borujeni, S.E., 2002. Cryptography by pseudo random number generator. Proceedings of the 1st International IEEE Symposium on Intelligent Systems, Volume 1, September 10-12, 2002, Iran, pp: 244-247.

Chong, W. and J. Kim, 1997. Speech and image compressions by DCT, wavelet and wavelet packet. Proceedings of the International Conference on Information, Communications and Signal Processing, Volume 3, September 9-12, 1997, Singapore, pp: 1353-1357.

Dawson, E., 1991. Design of a discrete cosine transform based speech scrambler. Electr. Lett., 27: 613-614.

Ehsani, M.S. and S.E. Borujeni, 2002. Fast Fourier transform speech scrambler. Proceedings of the 1st International IEEE Symposium on Intelligent Systems, Volume 1, September 10-12, 2002, Iran, pp: 248-251.

Enache, F., D. Deparateanu, T. Oroian, F. Popescu and I. Vizitiu, 2015. Theoretical and practical implementation of scrambling algorithms for speech signals. Proceedings of the 7th International Conference on Electronics, Computers and Artificial Intelligence, June 25-27, 2015, Bucharest, pp: 49-52.

Hosseinzadeh, V. and J. Shirazi, 2016. Improving the speech signal encryption by using wavelet and discrete cosine transforms and asymmetric permutation. Proceedings of the International Conference on New Researches in Engineering Sciences, June 1-2, 2016, Tehran.

Jameel, A., M.Y. Siyal and N. Ahmed, 2007. Transform-domain and DSP based secure speech communication system. Microprocessors Microsystems, 31: 335-346.

Jayant, N.S., R.V. Cox, B.J. McDermott and A.M. Quinn, 1983. Analog scramblers for speech based on sequential permutations in time and frequency. Bell Syst. Technical J., 62: 25-46.

Joseph, S.M. and P.B. Anto, 2011. Speech compression using wavelet transform. Proceedings of the International Conference on Recent Trends in Information Technology, June 3-5, 2011, Chennai, pp: 754-758.

Moreno-Alvarado, R.G. and M. Martinez-Garcia, 2011. DCT-compressive sampling applied to speech signals. Proceedings of the 21st International Conference on Electrical Communications and Computers, February 28-March 2, 2011, San Andres Cholula, pp: 55-59.

Najih, A.M.M.A., A.R.B. Ramli, V. Prakash and A.R. Syed, 2003. Speech compression using discrete wavelet transform. Proceedings of the IEEE 4th National Conference on Telecommunication Technology Proceeding, January 14-15, 2003, Shah Alam, Malaysia.

Nichols, R.K. and P. Lekkas, 2002. Wireless Security Models, Threats and Solutions. McGraw-Hill, New York, pp: 317-320.

Peyvandi, H., 2011. Security in data communication and privacy in conversations for underwater wireless networks using scrambled speech scheme. Proceedings of the Oceans'11 MTS/IEEE Kona, September 19-22, 2011, Waikoloa, HI., pp: 1-3.

Rajesh, G., A. Kumar and K. Ranjeet, 2011. Speech compression using different transform techniques. Proceedings of the 2nd International Conference on Computer and Communication Technology, September 15-17, 2011, Allahabad, pp: 146-151.

Sadkhan, S.B., N. Abdulmuhsen and N.F. Al-Tahan, 2007. A proposed analog speech scrambler based on parallel structure of wavelet transforms. Proceedings of the 24th National Radio Science Conference, March 13-15, 2007, Egypt, pp: 1-12.

Sayyadi, A., 2008. A basic introduction to the wavelet. Faculty of Electrical Engineering, Sharif Industrial University, Tehran.

Sushma, G.S. and D.R. Sandeep, 2011. Compression and enhancement of speech signals. Proceedings of the International Conference on Sustainable Energy and Intelligent Systems, July 20-22, 2011, Chennai, pp: 774-779.