International Journal of Computing and Network Technology
ISSN (2210-1519)
Int. J. Com. Net. Tech. 7, No. 2 (May-2019)

http://dx-doi.org/10.12785/ijcnt/070204

Front End to Back End Speech Scrambler

Taha A. Alsabbagh[1]

[1] *Dept. of Electronic Engineering, College of Electronics Engineering, Ninevah University, Mosul, Iraq*

Received: 10 Jan. 2019, Revised: 20 April. 2019, Accepted: 26 April. 2019, Published: (1 May 2019)

Abstract: In this paper, a modified approach for frequency permutation of speech signal has been designed implemented and verified based on MATLAB Simulink. A separate permutation of real and imaginary parts of the frequency component has been used. A Short Time Objective Intelligibility has been used for checking the performance and residual intelligibility of this method. The result shows a positive indication as compared with classical permutation.

Keywords: Analog Scramble; Speech Encryption; Speech Coding, Secure Voice; Residual Intelligibility.

1. Introduction

In communications, it is almost impossible to prevent people from eavesdropping. Thus, many authorities including military, police, and companies need to protect their radio communications, even outlaw need privacy to evade prosecution. The problem of providing some forms of privacy with a good level of security is becoming increasingly essential. Mainly there are two categories of protecting the privacy of speech; scrambling which deals with the original analog data, and encryption that usually refers to operations executed in the digital domain. Each of these categories has different procedures [1-5].

Most scrambling algorithms used permutation in the time or frequency domain. The main objective of these techniques reserved the bandwidth so that the transmitted signal could use the ordinary phone channel [6].

A full-duplex speech scrambling system is proposed for mobile communication systems. This algorithm used the rearrangement of the fast Fourier transform coefficients guided by adaptive dummy spectrum and compounding operation. The simulation results indicate that the scrambled signal has no residual intelligibility with a satisfied descrambled speech quality [7]. A frequency voices scrambler with less residual intelligibility is simulated in Matlab, two techniques are used, Fast Hartley Transform (FHT) and Orthogonal Frequency Division Multiplexing (OFDM) [8].

A paper of puzzle-solving problem, treat each piece of the spectrogram as a rectangular-shaped grayscaled image puzzle. In the paper two different methods are proposed; the first method is based on human heuristics, while the second method is based on the Ant Colony System (ACS) algorithm [9].

A monaural intelligibility prediction algorithm is proposed by J. Tensen and C.H.Tall [10], the algorithm shows similarities to the Short Time Objective Intelligibility (STOI) algorithm. As a consequence, Extended STOI is able to precisely predict the intelligibility of an impure speech of temporally highly modulated noise sources as well as noisy signals processed with time-frequency weighting [11-14].

This paper is dedicated to studying front end to back end speech scrambler carried out on a personal computer so that the system does not require any modems. The first step in this work is to sample the continuous speech signal and rehash it using Analog to Digital Converter (ADC) which implemented through the built-in microphone of the computer, passed through an anti-aliasing filter which is the first stage of the sound card. The block diagram which summarized the whole system is shown in Fig. 1.

It is clear from the block diagram in Fig. 1 that the Fast Fourier Transform (FFT) coefficients are scrambled. These scrambled data must be converted again into the time domain with an Inverse Fast Fourier Transform (IFFT) before transmitted them via the phone channel. The received part of the system has the descrambling process that works according to the type chosen for the scrambling process in the transmitted part.

The remaining of this paper is as follows. Section two describes the scrambling design. Section three introduces the suggested approach of real and imaginary

E-mail address: taha.alsabbagh@uoninevah.edu.iq

permutation. Section four is dedicated to simulation and results. Finally, section five concludes the work.

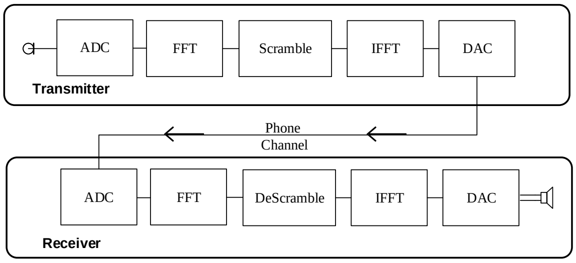


Figure 1. Overview of the scrambler descrambler system

2. SCRAMBLER DESIGN

A paper written by Sridharan et al. [15] describes the encryption of the speech by permuting the discrete Fourier transform (DFT) coefficients and considers the DFT of a vector *x* of range *N* defining one frame of speech in the time domain. The output coefficients of the FFT unit are in complex form is:

$$X = F\,x \qquad (1)$$

Accordingly, all the elements of *X* such as X_i and X_{N-I} for *i* =0, 1, ... (*N*/2) −1 are complex conjugates. The permutation is fulfilled by *N*×*N*. A permutation matrix *P* is applied to the speech DFT vector *X* to produce a vector *V*:

$$V = P\ X \qquad (2)$$

As *X* is complex conjugate so the elements of *V* should satisfy the symmetry property:

$$V_i = V^*_{N-1} \qquad \text{for } i = 0,1,\ldots..N-1 \qquad (3)$$

Where * indicates complex conjugate.

Equation (3) could be satisfied if the matrix *P* is selected to have the form:

$$P = \begin{bmatrix} P_1 & 0 \\ 0 & P_1' \end{bmatrix} \qquad (4)$$

Where P_1 and P'_1 are of order *N*/2.

Since the bandwidth of the telephone channel is restricted in a band of 300-3400Hz, so there is no need to scramble all coefficients in the whole frame but, limit them to specific coefficients to avoid increasing bandwidth. P_1 and P'_1 in (4) are selected to have the form:

$$P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad , \quad P_1' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & P_2' & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (5)$$

Where P_2 and P'_2 are of *L*×*L* permutation matrices, and *L* is the number of DFT coefficients within the speech bandwidth. In this work, a 16 kHz sampling frequency is used, with a selection of *N*=2048 samples in each analysis frame. These characteristics would imply that *L*=348 spectral coefficients should be used to restrict the bandwidth to the phone channel [1][16]. The transmitted output scrambled signal *y* through the phone channel is in the time domain and can be expressed as:

$$y = F^{-1}\,V \qquad (6)$$

The descrambling process is performed in the receiver unit in the same procedures except that the permutation matrix *P* is in inverse form.

3. REAL AND IMAGINARY PERMUTATION (RIP)

The suggested approach is used to pursue a good scattering of the speech data in a frequency domain, as these data are in complex form. The permutation is achieved separately for the real and the imaginary parts, by changing their sequences for the real part in its frame, while the imaginary part of the same sample is sent to another sequence. According to this approach (2 and 3) are modified and rewritten as in the following equations:

$$\mathrm{Re}(V) + j\,\mathrm{Im}(V) = P\ \ \mathrm{Re}(X) + j\,P\ \ \mathrm{Im}(X) \qquad (7)$$

Where:

$$\mathrm{Re}(V_1) + j\,\mathrm{Im}(V_1) = \mathrm{Re}(V_{N\text{-}i}) - j\,\mathrm{Im}(V_{N\text{-}i}) \qquad (8)$$

for *i*= 0,1,…. *N* -1

Since both the real part and imaginary part of matrix *V* have the same construction so two matrices *P* and *G* are used to execute the scrambling permutation, they are symmetric in size but with different set of random keys.

$$V = P\,(\,\mathrm{Re}(\,X\,)\,) + j\,G\,(\,\mathrm{Im}\,(\,X\,)\,) \qquad (9)$$

From the above equations it is clear that a certain algorithm permutation could be executed on the real matrix *P Re(X)* of size *L*×*L* and another algorithm permutation could be executed on the imaginary matrix *G Im(X)* of size *L*×*L* separately. Equation (6) can be formulated in the following manner.

$$y = F^{-1} \begin{bmatrix} P\,\mathrm{Re}\,(X) + j\,G\,\mathrm{Im}(X) & 0 \\ 0 & P\,\mathrm{Re}\,(X) - j\,G\,\mathrm{Im}\,(X) \end{bmatrix} \qquad (10)$$

Based on the algorithm mentioned in (10), the process of scrambling the data are scattered completely in a new unintelligible frame. These new coefficients are converted to the time domain and be ready for transmitting via the phone channel

A descrambling process is performed at the back end of the system, in the same manner, to acquire the original speech signal *x* in the time domain, except that P^{-1} and G^{-1} are used to return the coefficients in their original sequences.

This approach is carried out by a Simulink model as depicted in Fig. 2. In the above paragraphs, it is stated that the coefficients to be processed are only 346 out of 2048. Therefore, a filter is used to limit these coefficients and make the rest equal to zero.

A. *Swapping Permutation*

The first proposed algorithm of permutation studied in this paper is to divide the FFT coefficients of each frame, which are ranked from low to high frequency, into groups, and make an exchange between them. Any coefficient of the frequency domain *X* in (11) will have a new value according to the new real and imaginary parts.

The swapping process was achieved in different ways. The first trial is performed on the set *X* of data that represents the FFT coefficients of the frame. The coefficients are divided into two groups, "Group A" which contains X_0 to $X_{N/2-1}$ and "Group B" which contains $X_{N/2}$ to X_{N-1}. As it is known that the first coefficient in the *X* set represents the value of the lowest frequency appeared in the frame while the last value represents the highest good results as shown in Fig. 3

The second trial is to divide each group in the previous step into two subgroups then execute the swapping operation between them. Fig. 4 depicts these operations.

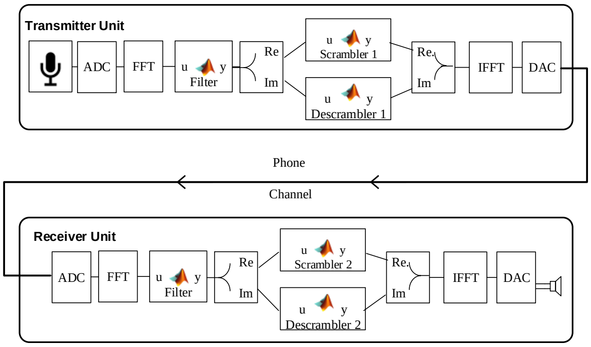


Figure 2. Simulink model for the modified approach of real and imaginary permutation

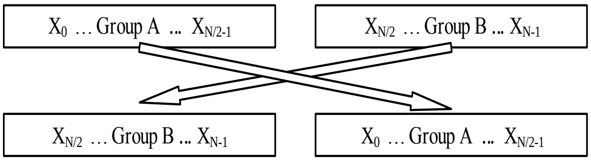


Figure 3. Dividing the FFT coefficients Into two groups and perform the swapping permutation beteen them.

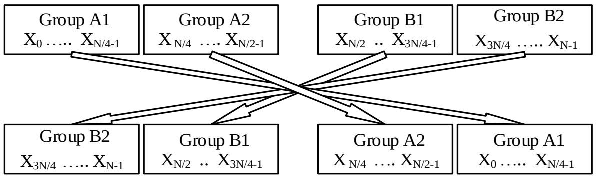


Figure 4. Dividing the two groups into two subgroups and perform the swapping permutation

Repeating the second trial more than one time, in the same manner, does not give better results and still, the first trial gives the best results comparing with others. That is because when these operations repeated much more, the places of the FFT coefficients returned back close to each other.

According to the new proposed RIP algorithm, the real parts are divided into a number of groups while the imaginary parts are divided into another number of groups. So more noises are obtained and more security is acquired.

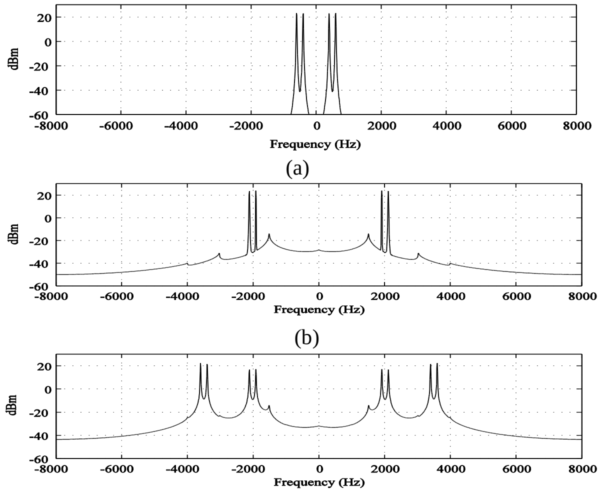
B. *Random Permutation*

The conventional random permutation algorithm is based on scrambling the complex coefficients randomly. Some papers focus on the random speech scrambling process [17][18]. The new RIP algorithm is to scramble the real and imaginary parts in a different manner from the other. This means that the real parts are shifted according to a random set of numbers while the imaginary parts are shifted with other random sets of numbers. The descrambling process works on the same set random numbers used for real while the imaginary works with its set. The scrambled signal in random permutation gives more security compared to the swapping permutation algorithm.

4. SIMULATION RESULTS

A signal consists of two tones 400 Hz and 600 Hz within the speech bandwidth is used as an input test signal, the spectrum of this signal is shown in Fig. 5-a. A swapping permutation is fulfilled on the frequency domain of this signal. Accordingly, a new scrambled signal of two strange tones 1916Hz, 2113Hz are appeared and ready to be transmitted via the phone channel as in Fig. 5-b, while on using the new RIP algorithm, the real and imaginary coefficients are scrambled in two different algorithms. The real coefficients are divided into two groups then swapped between them, while the imaginary coefficients are divided into four groups, then make the swapping process between them. The spectrum of the scrambled signal in the RIP algorithm has four strange tones as shown in Fig. 5-c.

The spectrogram in Fig. 6 shows that the Short Time Fourier Transform STFT of the output speech signal after descrambling from the receiver unit is identical to the original speech signal. While its spectrograms of the scrambled speech signal in both the conventional and RIP algorithms that shown in Fig. 7 are quite different from the original speech signal.



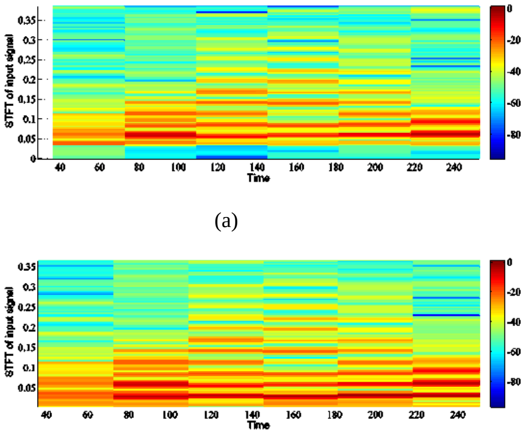
(a)

(b)

(c)

Figure 5. The spectrum for a signal of two tones (a) The Original signal spectrum (b) The scrambled signal in conventional swapping algorithm (c) The scrambled signal in RIP algorithm

A good indication could be obtained by observing the cross-correlation between the original speech signal and the received signal at the back end of the system as in Fig.8. The cross-correlation in Fig. 9-a depicts that the conventional random algorithm has a good result. While Fig. 9-b depicts the new proposed RIP algorithm which has better results compared to the conventional random algorithm.



(a)

(b)

Figure 6. Comparing continuity of STFT (a) The original speech signal (b) The received speech signal.

The software of Extended Short-Time Objective Intelligibility (ESTOI) algorithm [10] is used to calculate a scalar output of the intelligibility index for various scrambling models that are studied in this paper, so it could be said that it is a temporal average of the intelligibility indices.

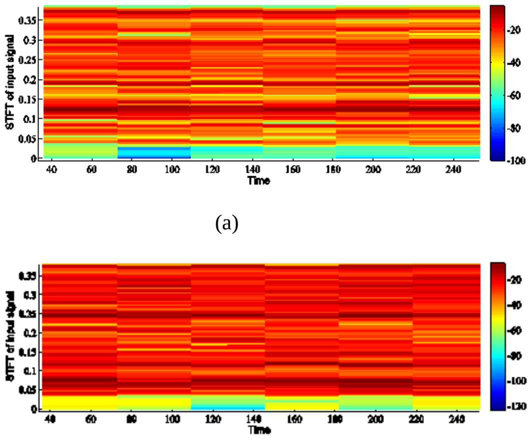
$$T = \frac{1}{M}\sum_{m=1}^{M} T_m \qquad (11)$$

Where: T is intelligibility index

M is the number of time segments in the produced scrambled signal.

Since $-1 \leqslant T_m \leqslant 1$, it follows that $-1 \leqslant T \leqslant 1$

Having a look at Table (1) showing that the better results for the intelligibility index are obtained when using the RIP algorithm with both algorithms under test, the swapping and random.



(a)

Figure 7. Comparing continuity of STFT (a) The scrambled speech signal with conventional random algorithms (b) The scrambled speech signal with RIP algorithm.

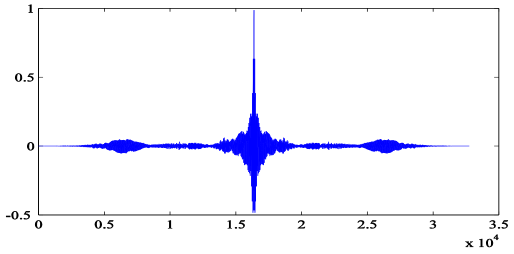
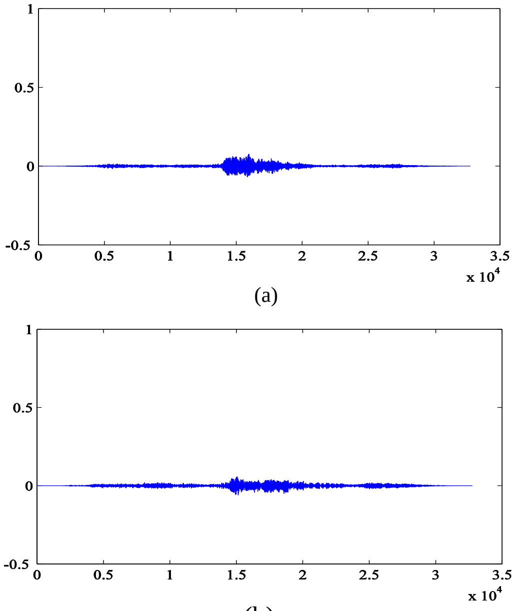


Figure 8. A normalized cross-correlation between the transmitted speech signal and the received speech signal



(a)

(b)

Figure 9. A normalized Cross-Correlation; (a) A cross-Correlation between the original speech signal and scrambled signal in conventional random algorithm (b) A cross-Correlation between the original speech signal and scrambled speech signal in RIP algorithm.

TABLE 1. A COMPARISON BETWEEN DIFFERENT MODELS OF THE INTELLIGIBILITY INDEX WHICH ARE EVALUATED IN THIS PAPER.

Model Type	intelligibility index
Received signal	1
Conventional Random Permutation	0.1984
Random with RIP Permutation	0.1531
Swapping Permutation	-0.1589
Swapping with RIP Permutation	-0.0546

CONCLUSION

A speech scrambling system is simulated in MATLAB Simulink. A scrambling process of speech signal has been executed in the frequency domain with two conventional scrambling algorithm; swapping which achieved by different methods, and random algorithms. The intelligibility index illustrates that the swapping algorithm gives better results than the random algorithm. This is because all the FFT coefficients that represent frequencies in the swapping algorithm are completely shifted into new frequencies, whilst at random, some of these coefficients may appear in such places near the original frequencies. On the other hand, the swapping method is easier to detect and bring to light when compared with the random method. Therefore the two methods should be taken into account. Finally, the RIP algorithm is proposed to process the real and imaginary parts of each sample in the frame separately. The results obtained using the RIP algorithm from the spectrogram, the Cross-Correlation and the intelligibility index promise to deep investigation in time and frequency domain simultaneously. This would create more security in the scrambled signal which is transmitted via the phone channel. Accordingly, it gives an estrangement signal that makes the transmitted signal more secured.

REFERENCES

[1] S. E. Borujeni, "Speech encryption based on fast Fourier transform permutation," In ICECS 2000. 7th IEEE International Conference on Electronics, Circuits and Systems, Vol. 1, pp. 290–293, Dec. 2000.

[2] R. Bernardini, G.M. Cortelazzo, and G.A. Mian, "A general scrambling rule for multidimensional fft algorithms," IEEE Transaction On Signal Processing, Vol. 42, No. 1 , pp. 1786-1794, Jul. 1994.

[3] M. S. Ehsani and S. E. Borujeni, "Fast Fourier transform speech scrambler," Proceedings First International IEEE Symposium Intelligent Systems, Varna, Bulgaria, Vol.1, pp. 248-251, 2002.

[4] Dhanya G and Dr. J. Jayakumari, "Permutation based speech scrambling for next generation mobile communication," International Journal of Engineering and Technology (IJET), Vol 8, No 2, pp. 707-713, Apr.-May 2016.

[5] K. Sakurai, K. Koga, and T. Muratani, "A speech scrambler using the fast fourier.transform technique," IEEE Journal on selected areas in communications, Vol. SAC-2, No. 3, pp. 434-442, May 1984.

[6] J. F. de Andrade, M. L. R. de Campos, and J. A. Apolinario, "Speech privacy for modern mobile communication systems," In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1777–1780, Mar. 2008.

[7] A. Matsunaga, K. Koga, and M. Ohkawa, "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, May 1989.

[8] DS Anjana and M. Kuriakose, "Frequency speech scrambler based on hartley transform and OFDM algorithm," International Journal of Computer Applications, Vol.61, No. 8, pp. 36–40, 2013.

[9] Y. Zhao, M. Su, Z. L. Chou, and J. Lee, "A puzzle solver and its application in speech descrambling," In WSEAS Int. Conf. Computer Engineering and Applications, pp. 171–176, 2007.

[10] J. Jensen and C. H. Taal, "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, No. 11, pp. 2009–2022, Nov. 2016.

[11] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," IEEE Transaction on Audio, Speech, and Language Processing, Vol. 19, No. 7, pp. 2025-2136, Sep. 2011.

[12] J. Jensen and C. H. Taal, "Speech intelligibility prediction based on mutual information," IEEE/ACM Transactions on audio, speech, and language processing, Vol. 22, No. 2, pp. 430-440, Feb 2014.

[13] D. C. Tseng and J. H. Chiu, "An OFDM speech scrambler without residual intelligibility," TENCON 2007 - 2007 IEEE Region 10 Conference, pp. 1-4, Nov. 2007.

[14] A. Srinivasan and P. A. Selvan, "A review of analog audio scrambling methods for residual intelligibility," Innovative Systems Design and Engineering, Vol. 3, No. 7, pp. 22-38, 2012.

[15] S. Sridharan, E. Dawson, and B. Goldburg, "Fast Fourier transform based speech encryption system," IEE Proceedings I - Communications, Speech and Vision, Vol.138, No. 3, pp. 215– 223, Jun. 1991.

[16] F. A. de O. Nascimento and R. G. Toscano "Frequency speech scrambler based on the hartley transform and the insertion of random frequency components," The International Journal Of Forensic Computer Science, Vol. 7, No. 1, pp. 8-15, 2012.

[17] Dhanya G and Dr. J. Jayakumari, "An efficient voice scrambling technique for next generation communication systems," International Journal of Engineering and Technology, Vol. 8, No. 1, pp. 293-299, Feb.-Mar. 2016.

[18] Z. Zhu, K. Yamamoto, M. Unoki and N. Aoki, "Study on scrambling method for speech signal by using random-bit shift of quantization," journal of signal processing, Vol 18, No. 6, pp. 303-307, Nov. 2014.

Taha Al-Sabbagh received his BSc. in electrical engineering from Mosul University in Iraq 1983 Next, awarded the MSc. Degree In electronics and communications engineering from Mosul University 1989. , He joined the High Technical Institute of Al-Gubba, Libya as a teacher in 1997-2001, He worked as a teacher in the High Teacher Institute of Al-Gubba (part-time), Libya 1999-2001. Currently, he is working in Dept. of Electronic Engineering, College of Electronics Engineering, Ninevah University, Mosul, Iraq 2004- till now.