Design and Evaluation of Transform – Based Speech Scramblers using different Wavelet Transformations

S. B. Sadkhan
Babylon University – Computer Science dept. – Iraq – Baghdad
E-mail: drengsattar@yahoo.com
N. H. Kaghed
Chief of Babylon University – Iraq – Babylon
L. M. AlSaidi
Babylon Univ. – Computer Science dept- Iraq – Babylon

***Abstract*- A Speech scramblers based on permutation of coefficients resulting from different wavelet transforms are designed and evaluated. The proposed systems offer two-dimensional scrambling process. The suitability of different wavelet transformations (Daubechies [db1, bd3], Symmlet [Sym2, Sym4] , and Coiflet [Coif1, Coif2]., each one is tested in different decomposition levels [level1, level2, level3]) for use in " transform – based speech scramblers " is investigated. Subjective as well as objective tests were conducted to investigate the residual intelligibility and the recovered speech quality under different channel conditions. The results indicates low residual intelligibility and good recovered speech quality.**

An evaluation for cryptanalytic strength of the designed scheme is also considered by applying some objective distance measures SNR, as well as the Spectral Segmental Signal – to – Noise Ratio.

The time domain representation, frequency domain spectrum, spectrogram and power density function for each frame of the input speech signal are calculated and compared with same group of the scrambled and recovered speech signals. Related analysis and simulation results indicate that the proposed scramblers are highly secure, and can be considered as a promising scrambling techniques if some other arrangements can be considered.

Key words: Speech scrambler, wavelet transforms, cryptanalytic evaluation, residual intelligibility

I. INTRODUCTION

In radio communications, including satellite communications, it is almost impossible to prevent unauthorized people from eavesdropping unless speech scramblers are used. Even in wire communications, speech scramblers may be needed to protect privacy. Among speech scramblers, analog scramblers are attractive due to their wide applicability. The conventional analog scramblers manipulate speech signal in the frequency or time domain. A typical frequency domain scrambler is the band splitting scrambler which breaks the speech signal into several sub-bands and permutes them. While a typical time domain scrambler is the time division scrambler which breaks the speech signal into short time segments and permutes them within a block of several segments. These conventional scramblers cannot provide sufficient security against cryptanalysis because the number of permutable elements in these scramblers is not large enough to provide an adequate number of different permutations due to hardware limitations and processing delays. To strengthen security, a two-dimensional scrambler which manipulates the speech signal both in the frequency domain and in the time domain was proposed [2]. Regarding other types of scramblers which can attain a high degree of security, the transform domain scrambler was proposed. Wyner proposed a method [3], in which an orthogonal transform called a Prolate Spheroidal transform (PST) is executed on a set of the sampled speech signal. The coefficients obtained from the PST are permuted and the time domain scrambled signal is obtained by the inverse PST on those permuted coefficients. In this way, the bandwidth of the analog signal can be kept unchanged after scrambling [4]. Other approaches similar to Wyner's but using a Fourier basis and scrambling the signal in the frequency domain, referred to as DFT scrambling here, were then proposed . Because of the fast Fourier transform (FFT) algorithms, the DFT approaches significantly simplifies the implementation complexity required in Wyner's method. The process of FFT scrambling is similar to PST scrambling except that FFT is used instead of PST. The performance of FFT scrambling is expected to be inferior to PST scrambling if the scrambled signal is transmitted over severely bandwidth limited channel.

Another frequency domain scrambling system was proposed using the short time Fourier analysis- synthesis techniques invented by Schafer and Rabiner and Portnoff. With these techniques , the original speech can be correctly recovered with the correct key even under very poor channel conditions, although with a relatively poor speech quality. This is a dramatic improvement in scrambling performance. However, this improved technique is still a frequency domain approach, using FFT algorithms to transform the sample values back and forth between time and frequency domains.

in 1993 paper [1], they claimed that:

- In time domain scrambling (for example, the hopping window time domain scramblers), there may be a sufficient information present in the cipher text (

scrambled speech) alone to reorder the speech segments and completely recover the original speech.
- In frequency domain scramblers (band splitting scrambling) the encryption permutation (or key) can be completely recovered from the cipher text alone (if this permutation is fixed).
- One family of analog scramblers that has shown a great deal of promise is the transform domain scrambler.
- A very little is known about the performance of other transforms (fast Fourier transform, discrete prolate transform) for speech encryption.
- They considered four discrete orthogonal transforms:- discrete Fourier transform (DFT), discrete Cosine Transform (DCT), Discrete Prolate Spheroidal transform (DPST), and Walsh –Hadmard Transform (WHT).
- Transform which decorrelate information in the transform domain are favored, since it is believed that the scrambled speech resulting from scheme using these transforms will have low residual intelligibility.
- It is desirable to be able to restrict the bandwidth of the encrypted speech resulting from a given transform based scrambler, and therefore transforms which facilitate this have to be chosen.
- Transforms that can be performed using a fast algorithm are also preferred.
- The transform – based analog speech scrambling are sensitive to transmission channel characteristics.
- A speech segment sampled at 8KHz and divided into 256 samples frames was used as input to each scrambler.

In this paper we consider different Wavelet Transform (WT) for application as a speech scrambler. As well as objective tests were conducted to compare the residual intelligibility and recovered speech quality of the transform-based scrambled schemes under channel conditions.

II. Wavelet Transformation

The wavelet transform is a new and promising set of tools and techniques for signal analysis. Wavelet analysis is different from the Fourier transform and Short Time Fourier Transform (STFT) — it is a windowing technique with variable-sized regions. In STFT analysis, new frequency-domain coefficients are computed with a FT (or DFT) every d*t*, and the complex amplitude values are a function of frequency and time. Note that in the STFT, the basis functions are sine and cosine waves, and the time or frequency analysis points are evenly spaced [6], [8].

Many families of wavelets have proven to be very useful in signal analysis[7], They share some common features:
1. They are localized in the time (space) domain. Instead of oscillating forever like a sine wave, they drop to zero after a time [10].
2. The family is derived by scaling and shifting a wavelet prototype function, called an “analyzing wavelet” or “mother wavelet”.
3. Continuous-time wavelets are linked to discrete-time filters through the limit of a logarithmic filter tree .
4. Scaling functions and wavelets inherit orthogonality, or biorthogonality, from the filter bank [11].

The Daubechies-N wavelet was developed to be compactly-supported and orthonormal , thus making discrete wavelet analysis practical.
General characteristics: Compactly supported wavelets with extremal phase and highest number (N) of vanishing moments for a given support width. Associated scaling filters are minimum-phase filters.

The Coiflets: is compactly supported wavelets with highest number of vanishing moments for both phi and psi for a given support width. Developed to be a more symmetrical version of the Daubechies wavelet.

The Symmlets are nearly symmetrical wavelets proposed as modification to the db family. The properties of the two wavelet families are similar. General characteristics: Compactly supported wavelets with least asymmetry and highest number of vanishing moments for a given support width. Associated scaling filters are near linear-phase filters [9].

III The Proposed Speech Scrambling System

The analog encryption process which employs a transformation of the input speech to facilitate encryption can best be described using matrix algebra. Consider the vector x which contains N speech time samples obtained from analog-to digit conversion process. Representing a frame of the original speech signal. Let this speech sample vector x be subject to an (N*N) orthogonal transformation matrix F such that:

$$U = Fx \qquad (1)$$

This transformation result in a new vector u made up of N transform coefficients. An N*N permutation matrix is applied to U, such that each transform coefficient is moved to a new position within the vector given by

$$V = PU \qquad (2)$$

A scrambled speech vector U is obtained by returning vector V to the time domain using the inverse transformation F^{-1} , where :

$$Y = F^{-1} V \qquad (3)$$

Decryption, or recovery of the original speech vector x' is achieved by first transforming Y back to the transform domain. The inverse permutation matrix P^{-1} is then used to return the transform coefficients to their original position. Finally, the resulting transform vector is returned to the time domain by multiplying by F^{-1}.

$$x' = F^{-1}P^{-1}FY. \qquad (4)$$

The transform domain scrambling process outline above requires the transform matrix F to have an inverse. One attempts to insure that the scrambling transformation $T=F^{-1}PF$ is orthogonal since orthogonal transformations are norm preserving. This property is useful sine any noise added to the scrambling signal during transmission will not be enhanced by the descrambling process .

The scrambled speech sequence is given by

$$Y = F^{-1}PFx = T x. \qquad (5)$$

Mostly the N elements are able to be permuted in the transform –based scrambling process. It is important to note that for a given sampling frequency, N will determine the delay introduced by the scrambling device. So a tradeoff between system delay and security .N is usually chosen to be equal to 256. The permutations must carefully screen to ensure that components will undergo a significant displacement from their original position in the vector. In addition, components which were adjacent in the original vector should be separated in the scrambled vector.

At the transmitter, speech signal is sampled and segmented into frames, and then transformed into wavelet space, after that the wavelet coefficients are permuted by using a good permutations, for example, P is considered as set of permutations, then P should satisfy the following requirements:

- Any permutation in P must not produce an intelligible scrambled speech signal.
- Any inverse permutation in P^{-1} must not produce an intelligible descrambled speech signal if the inverse permutation dose not corresponds to the permutation used in the transmitter.

After doing a permutation, inverse wavelet transform is applied, yielding the scrambled speech. At the receiver, the scrambled signal is transform back into the wavelet space, and then into inverse permutation and finally transformed into inverse wavelet transform, resulting in , the original speech signal.

In the proposed WT scrambling system, the measure of the difference between the original speech signal and the scrambled speech signal, and between the original speech signal and the descrambled speech signal was performed, also many wavelet functions such as Haar, Daubechies , Coiflets , Symlets, are used in this proposed system and the performance between them was compared [5].

IV Results and Discussion

In the proposed WT scrambling system, some justifiable assumptions have been made, which are, any nonlinear phase characteristics introduced b the channel is able to be equalized by the receiver (descrambling part), any time jitter in the sampling introduced at the receiver, doesn't affect the performance of descrambling., a narrow band transmission channel was used, the scrambled speech signal is transformed into analog by passing through inverse wavelet transformation, and a speech segment sampled at 8KHz and divided into 256 sample frames was used as input to each scrambler. The message in Arabic; this message may be spoken by a man or woman.

At the transmitter, the sample speech signal is converted into frames with each frame containing 256 samples and then the Wavelet Transformation is performed on each frame. After that, the transform coefficients are permuted before the inverse transform is applied. The resulting scrambled speech signal is saved in a wave file.

At the receiver, frame by frame of length 256 samples are descrambled and saved in wave file. The proposed system investigates four types of wavelets: (Haar, db3, sym2 and sym4), each one with three different levels. Two types of tests have been used to examine the performance of the simulation, these are:

a) Subjective Test: in which the scrambled speech files have been played back to a number of listeners to measure the residual intelligibility, subjectively. For all cases, the judge is that; the listened files contain noise only, which means that the residual intelligibility is very low. The analog recovered speech files have been tested in a similar way to measure the quality of the recovered speech files; the judge is that the files are exactly the same as the original copies.

b) Objective Test: As mentioned earlier, the objective test is a valuable measure to the residual intelligibility of the scrambled speech, and the quality of the recovered speech.

- The distance measures indicate the perceptual similarity of the speech recovered following decryption and the original speech; they are also used to quantify the difference between scrambled speech and original speech.
- The signal to noise ratio (SNR) and the segmental signal to noise ratio (SEGSNR) distance measures have been chosen to test the residual intelligibility of the scrambled speech, and the quality of the recovered speech for all files. The segmental signal to noise ratio measure (SEGSNR) is an improved version measure of the (SNR).

Generally, these distance measures for all the scrambled speech files are very low which means that the residual intelligibility is very low, and the distance measures for all the recovered speech files are very high (large positive value) which means that the quality of the recovered speech is very high.

We use the relation between estimated PSD (dB/Hz) in relation with frequency of the used speech signals in two cases, as follows:

- To compare the original and scrambled speech.
- To compare the original and descrambled speech.

The proposed speech scrambling system have been tested under two states of the simulation, these are:

1- Free Channel Simulation; simulation results of typical experiments with the scrambler, and the descrambler for an Arabic word spoken by women's voice 'evening '' are shown in figures (1) to (4), and Tables (1) to (2), using different wavelets and different decomposition levels.

Case Study ; Using (Haar) Wavelet and (db3) wavelet each one will be considered with three different levels for the Arabic word " evening ". Figure (1) shows the waveform, spectrum, and spectrogram of a sample original clear speech signal that represents an Arabic word "evening ".

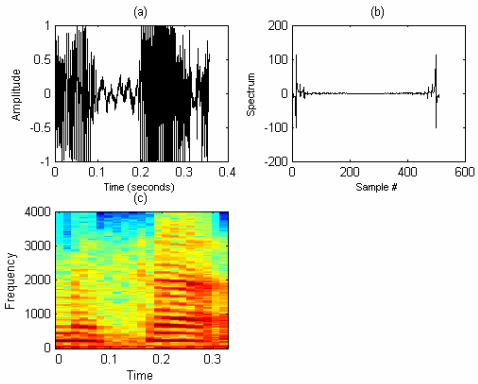

Figure 1 Original Speech Signal; a) Waveform. (b) Spectrum. (c) Spectrogram.

Figure (2) shows the waveform, spectrum, spectrogram and the comparison of the scrambled speech signal, that resulted from applying a wavelet transform of type (Haar) with a specified level (level 1). While figure 3 shows the comparison between original speech and scrambled speech using PSD measure. Figure 4 shows the same group of illustrations for the recovered speech signal after descrambling.

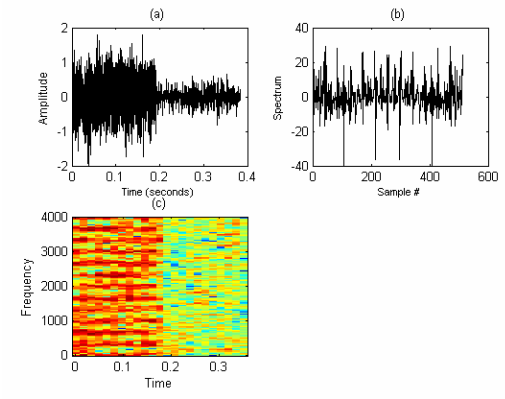

Figure (2) Scrambled Speech Signal Using Haar Wavelet With Level 1; (a) Waveform. (b) Spectrum. (c) Spectrogram**.**

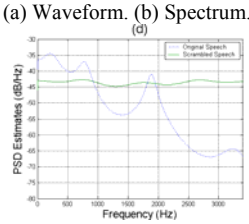

Fig. (3) The comparison between original speech and scrambled Speech using PSD Estimates

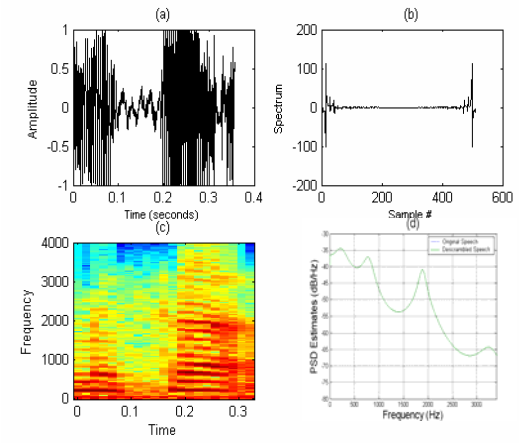

Figure (4) Descrambled Speech Signal Using Haar Wavelet With Level 1; (a) Waveform. (b) Spectrum. (c) Spectrogram. (d) The Comparison Between Original Speech and Descrambled Speech.

Table (1) shows distance measure (SEGSNRs) for the scrambled speech, while Table (2) shows the (SEGSNRd) distance measure for the descrambled speech for different Wavelets and different decomposition levels.

2- Noisy Channel Simulation; An evaluation of the proposed speech scrambling system with different signal to noise ratios from (5 dB up to 25 dB) was tested [5]. The results from such tests are shown in Tables (3) and Table (4). Table (3) shows the SEGSNRs distance measure for the scrambled speech, and Table (4) shows the $SEGSNR_d$ distance measure for the descrambled speech, with SNR = 5 dB.

V Conclusion:

The performance of the proposed speech scrambling system based on wavelet transform was examined on actual " **Arabic Speech Signals** " , and the results showed that there was not any residual intelligibility in the scrambled speech signal, since the listeners hearing anoisy signal, and descrambled speech signal at receiver was exactly identical to the original . Hence it provides the high security scrambled speech system and the reconstructed signal was perfect. Some interesting points can be mentioned here:

a. is clear from **(**SNR_s, $SEGSNR_s$**)** and (SNR_d, $SEGSNR_d$) in the measurements, that (SNRs & $SEGSNR_s$) give us small values at any decomposition level, while (SNR_d & $SEGSNR_d$), indicate large values. As the level decreases the system performs better. The absolute low values of distance measures does not necessarily mean a perceptually poor assessments. The distance measures (SNR and SEGSNR) for scrambled/descrambled speech, can in some cases, be used for design purposes as a relative number of intelligibility loss or speech quality.

b. The spectrogram is used because it is a powerful tool that allows us to see what's happening in the frequency and time domain all at once. Thus you can easily see the theory at work here by observing the original signal, it's scrambled version, and the descrambled version. Note that on the scrambled plot it is observed that the order of the frequencies has changed. And, as expected the descrambled version has been correctly decoded to its original form.

c. An evaluation of the proposed speech scrambling system with different power levels of the additive white Gaussian noise was tested. The results proved that as the signal to noise ratio increases, the correspondence between original and descrambled speech increases. Hence, it can be concluded that, the WT algorithm can be implemented to scramble and descramble speech with high efficiency.

d. For real time speech scrambling it is recommended to use a wavelet with a small number of order at a reasonable decomposition level (level **3** decomposition or less), because the number of coefficients required to represent a given signal increases with the level of decomposition (higher wavelet decompositions requires more computation time, which should be minimized for real time speech scrambling) and with the large number of order.

e. Related analysis and simulation results indicate that the proposed scramblers are highly secure, and can be considered as a promising scrambling technique if some other arrangements can be considered.

References:

[1]. B.Goldburg, S.Sridharan , and E. Dawson, "Design and cryptanalysis of transform- based analog speech sdramblers", IEEE Journal of Selected Areas on Communication, vol. 11, pp. 735-743,1993.

[2]. K. Sakurai, K. Koga , and T. Miratani,"A speech scrambler using the fast Fourier transform technique", IEEE J. Select. Areas Communication,vol.SAC-2, no.3,pp434-442, May 1984.

[3]. A.Matsuanaga.K. Koga ,and M.Ohkawa, "An Analog Speech Scrambling system using using FFT technique with high level security", IEEE J.Select.A Areas Communication, vol. 7, pp. 540-547,1989.

[4]. S.Sidharan, E.Dawson, and B.Goldberug, " Fast Fourier transform based encryption system", IEEE Proc.I, communication, Speech,Vision,vol.138, no. 3,pp215-223,1991.
[5]. B. Sattar, H. Nabel, H. Lamis," A Proposed Speech Scrambling Based on Wavelet Transform and Permutation", Third International Conference on Systems, signals and Devices, SSD'05, March 21-24 , 2005, Sousse, Tunisia.
[6]. A. Graps, .An Introduction to Wavelets,. *IEEE Computational Sciences and Engineering*, http://www.amara.com/IEEEwave/IEEEwavelet.html
[7]. I.Dubechies,'Ten Lectures On Wavelets', Rutgter University and AT&T.Bell Laboratories,SIAM,Philphia,PA,1992.
[8]. B. Lin, B. Nguyen and E.T. Olsen, "Orthogonal Wavelets and Signal Processing" , *Signal Processing Methods for Audio, Images and Telecommunications*, P.M.Clarkson and H.Stark, ed., Academic Press, London, 1995, pp. 1-70.
[9]. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif., 1998.
[10]. Y. Nievergelt, *Wavelets made easy***,** Birkhäuser, Boston, 1999.
[11]. J. Ooi and V. Viswanathan, .Applications of Wavelets to Speech Processing,. *Modern Methods of Speech Processing*, R.P. Ramachandran and R. Mammone, ed., Kluwer Academic Publishers, Boston, 1995, pp. 449-464.

Table (1) SEGSNRs (dB) for the scrambled speech, for each wavelet with a specific level.

Levels of decomposition / Wavelet Type	1	2	3
Haar	-4.8732	-4.0857	-4.0907
db3	-4.7673	-3.7751	-3.8147
sym2	-4.8064	-3.7710	-3.6743
sym4	-4.6620	-3.7125	-3.9071

Table (2) SEGSNRd (dB) for the recovered speech, for each wavelet with a specific level.

Levels of decomposition / Wavelet Type	1	2	3
Haar	310.2651	305.6744	303.1264
db3	15.8569	17.4644	9.5916
sym2	112.9616	18.1980	13.9168
sym4	12.8815	10.0766	13.0374

Table (3) SEGSNRs (dB) for the scrambled speech, for each wavelet with a specific level, with SNR = 5 dB.

Level / Wavelet Type	1	2	3
Haar	-5.8671	-5.3838	-5.2932
db3	-5.7119	-5.1464	-5.2938
sym2	-5.7810	-5.0943	-5.0174
sym4	-5.7231	-5.2204	-5.3608

Table (4) SEGSNRd (dB) for the recovered speech, for each wavelet with a specific level, with SNR = 5 dB.

Level / Wavelet Type	1	2	3
Haar	3.1951	2.8529	2.9930
db3	2.5307	2.0152	1.4078
sym2	2.9114	2.4818	2.1516
sym4	2.2625	1.9958	1.2045