Audio Hiding based on Wavelet Transform and Linear Predictive Coding

Jamal Nasir Hasoon

Computer sciences

University of Al Mustansiriyah

Dr. Saad Najim Al-Saad

Computer sciences

University of Al Mustansiriyah

Abstract- **In this work an efficient method for hiding a speech in audio is proposed. The features of secret speech is extracted with LPC (Linear Predictive Coding), and these parameters embedded in audio in chaotic order. Discrete Wavelet Transform (DWT) is applied on audio frames to split the signal in high and low frequencies. The embedding parameters are embedded in high frequency. The stego audio is perceptually indistinguishable from the equivalent cover audio. The proposed method allows hiding a same duration of speech (secret) and audio (cover). The stego audio is subjected to objective tests such signal to noise ratio (SNR), signal to noise ratio segmental (SNRseg), Segmental Spectral SNR, Log Likelihood Ratio (LLR) and Correlation (Rxy) to determine the similarity with original audio**.

Index Term - Steganography, Linear Predictive Coding (*LPC*), and Wavelet Transform.

I. Introduction

The development of the digital multimedia technology and the widespread popularity of the Internet have brought about convenience. The various shortcomings and application constraints of conventional cryptography method, information hiding, as a new technology and method in information security domain, has drawn more and more attention from both research communities and application groups. Information hiding technology includes digital watermarking and steganography, applied to the copyright protection of the digital multimedia arts and the covert communication for the secret message. Several steganography techniques were used to send message secretly. Many techniques have been developed for hiding secret signals into other cover signals [1, 2]. The general structure for audio steganography is depicted in Figure (1).The structure contains two phases: embedding (to hide the secret speech in cover audio) and extraction (to extract the secret speech from the cover audio).

The popularity of speech files and its ability to convey secret information make many researchers investigate how speech signals and speech properties can be used in the domain of information hiding [3]. Several approaches were conceived, the most popular ones are: Least Significant Bit [4], Echo hiding [5], Hiding in Silence Interval [6], Phase Coding [7], Amplitude Coding [8], Spread Spectrum [9], and Discrete Wave Transform [10]. The objective of this work is to develop a high performance audio steganography system.

II. Discrete Wavelet Transform

The Wavelet Transform provides a time-frequency representation of the signal. It was developed to overcome the shortcoming of the Short Time Fourier Transform (STFT), which can also be used to analyze non- stationary signals [11, 12]. While STFT gives a constant resolution at all frequencies, the Wavelet Transform uses multi-resolution technique by which different frequencies are analyzed with different resolutions. In Wavelet Transform, the width of the wavelet function changes with each spectral component. The Wavelet Transform, at high frequencies, gives good time resolution and poor frequency resolution, while at low frequencies; the Wavelet Transform gives good frequency resolution and poor time resolution. The Discrete Wavelet Transform (*DWT*), which is based on sub-band coding, is found to yield a fast computation of Wavelet Transform. It is easy to implement and reduce the computation time and resources required. The general form of an L-level DWT is written in terms of L detail sequences as shown equation (1),

$$d (k) j \text{ for } j=1,2,....,L,$$

and the *L*-the level approximation sequence,

$c (k) L$ as follows :

$$f(t) = \sum_k c_l(k)\Phi_l(t) + \sum_{j=1}^{l} \sum_k d_j(k)\Psi_j(t) \cdots (1)$$

Where:$\Phi_l(t)$ is the L-the level scaling function and $\Psi_j(t)$ for j=1,2,…,L are wavelet function

www.uoitc.edu.iq

sequences for L different levels. In order to work directly with the wavelet transform coefficients, the relationship between the detailed coefficients at a given level in terms of those at previous level is used. In general, the discrete signal assumes the highest achievable approximation sequence, referred to as 0-th level scaling coefficients. The approximation and detailed sequences at level j are given by equation (2 and 3):

$$c_{j+1}(k) = \sum_m h_o(m - 2k)c_j(m) \qquad (2)$$

$$d_{j+1}(k) = \sum_m h_1(m - 2k)c_j(m) \qquad (3)$$

Equations (2) and (3) state that approximation sequence at higher scale (lower level index), with the wavelet and scaling filters, *ho (t)* and *h1 (t)* respectively, can be used to calculate the detail and approximation sequences (or discrete wavelet transform coefficients) at lower scales.

There are a number of basis functions that can be used as the mother wavelet for Wavelet Transformation [13]. Since the mother wavelet produces all wavelet functions used in the transformation through translation and scaling, it determines the characteristics of the resulting Wavelet Transform. Therefore, the details of the particular application should be taken into account and the appropriate mother wavelet should be chosen in order to use the Wavelet Transform effectively.

III. Linear Predictive Coding

Linear Predictive Coding (LPC) is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate. It provides accurate estimates of speech parameters, and is relatively efficient for computation [14]. LPC starts with the assumption that a dynamic speech signal can be viewed as a stationary waveform for short periods of time.

LPC analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of

the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal is called the residue. The numbers that describe the formants and the residue can be stored or transmitted somewhere else. LPC synthesizes the speech signal by reversing the process: using the residue to create a source signal, using the formants to create an all-pole filter, and running the source through the filter, result in speech [15].

IV. Chaotic theory (Logistic Map)

Chaos theory is a field of mathematics that studies the behavior of dynamical systems that are highly sensitive to initial conditions. The logistic map is one of the chaos function. The idea behind using logistic map is to generate a similar sequence number from the initial values called (initial parameters and control parameters). Equation (4) illustrates the behavior of logistic map.

$$x_{new} = R * x_{old}(1 - x_{old}) \qquad (4)$$

Where x_{old} is the initial value, R is control parameters, and x_{new} is the generated value. The logistic map used for selection the location of the speech parameters that will be embedded in the cover audio by applying equation (4) to generate random numbers. These numbers are sorted ascendingly. The first fifteen numbers are chosen as the location for embedding.
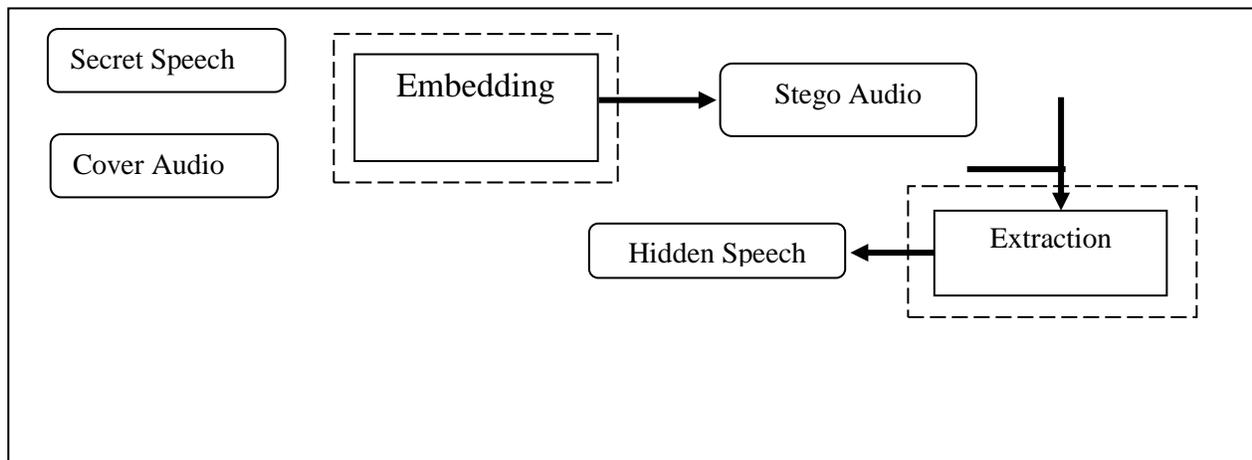

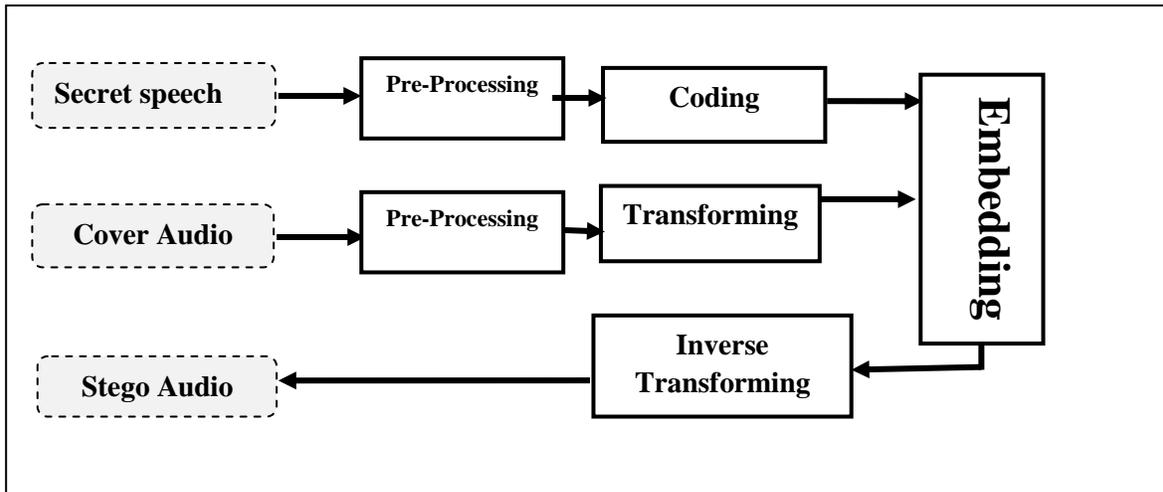Fig. (1) The block diagram of general structure of Audio steganography system

31

Fig. (2) The block diagram of embedding phase

V. The proposed method

The proposed method lies in embedding and extraction phases. The embedding is consist in turn of sub phases: secret speech coding, cover audio processing, cover audio transforming, embedding of secret speech and cover audio inverse transforming. While extraction consist of stego-audio processing, transforming, extraction, secret speech synthesis and secret speech reconstruction.

A. Embedding Phase

Figure (2) represents the block diagram of embedding phases. The inputs are secret speech and cover audio while the output is stego audio. The pre-processing consists of framing, overlapping and windowing figure (3). The number of sample (NoS) is calculated by:

$$NoS = Frame\ time\ *10^{-3}*Frequency\ \ ..... 3.1$$

Where *Frequency* is in Hz.
The frames are overlapped with each 15 ms and windowing with Hamming window.

- Secret Speech Coding

The secret speech frame is processed to extract features with least possible parameters to be embedded in cover audio file as shown in Figure (4). The Linear Predictive Coding (LPC) is applied to each frame to obtain the LPC's parameters (P=12),

error signal, gain, pitch and voiced| unvoiced bit. These parameters (P+3) represent the feature extraction of secret speech.

- Cover Audio Processing

The framing of cover file is similar to that of secret one but without overlapping and windowing. The length of frame is equal to or greater than the length of secret file's frame.

- Cover Audio Transforming

The cover audio file frame is transformed with discreet wavelet transform (DWT) (two channels) to get wavelet domain. The DWT is used in one level to separate the high frequency from low frequency as shown in Figure (5). The high and low frequency component are half (*NoS=120*) of frame (*NoS=240*). The wavelet domain contains exactly the same information as time domain, but in a different form.

- Embedding of Secret Speech

The embedding step is achieved by replacing the specific location of high frequency (result in from transformation step) of cover audio frame with vector of secret speech parameters which were extracted before (in coding step). The secret speech parameters scaled before embedding to avoid the noise in the signal as possible.
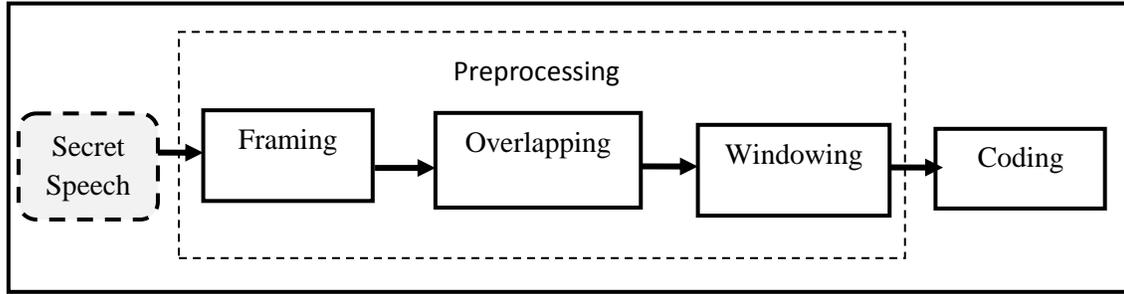
32
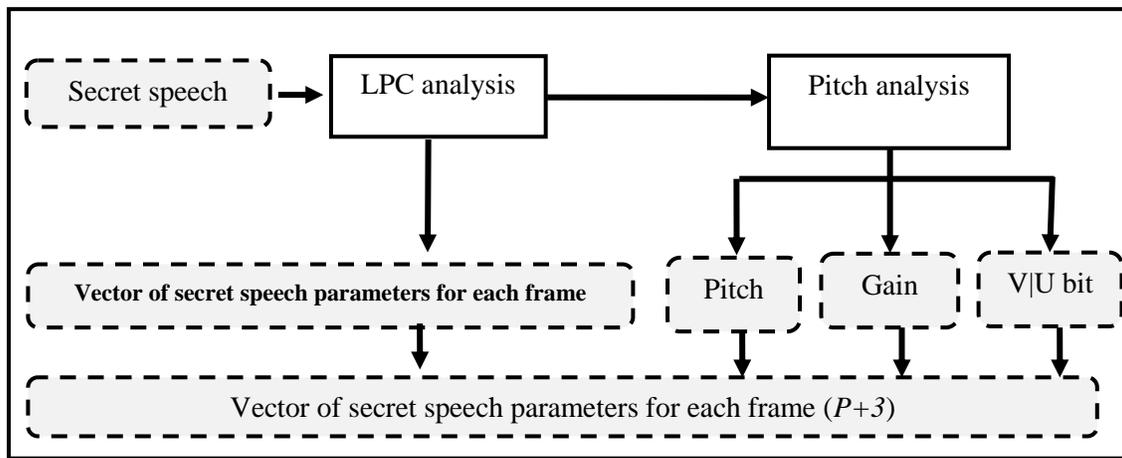
Fig. (3) Secret speech files pre-processing

Fig. (4) secret speech coding

• The Inverse Transforms of Cover Audio

The inverse transforms, Inverse Discrete Wavelet Transform (IDWT) are applied to cover audio frame to produce stego audio frame. The time domain is returned to form the wavelet domain by applying Inverse Discrete Wavelet Transform. The produced frames are concatenated to construct stego audio file.

B. The Extraction Phase
The process for this task is depicted in Figure (6). There are framing, and DWT to split the stego file into high and low frequencies. The keys in the embedding are the same used in extraction to determine the same location used in each frames in embedding this task accomplished by extraction process the output of extraction is a vector of LPC parameters and the three values of pitch, gain and voiced |unvoiced bit. The synthesis process is illustrated in Figure(7), the figure is block diagram contains the details of the step in. The stego audio is partitioned into frames with the same length of cover audio's frame length, and submitted for transforming by (DWT) to split to high and low frequencies. Figure (6) shows disciplined extraction of hidden speech. The framing and Chaotic Key Generation in extraction phase are the same procedures in embedding phase.

• The Secret Speech Synthesis

33

www.uoitc.edu.iq

The hidden speech synthesis is the last and most important step in extraction phase. The classification of frames is important for synthesizing. The voiced or unvoiced parameter controls the generation of signal that replaces the excitation signal (error signal). The reconstruction of voiced frame needs Impulse Train Generator (ITG), while Random Noise Generator (RNG) is needed for unvoiced frame. The length of ITG and RNG is equal to secret speech frame length. These two kinds of frame are the input to filter. The LPC's parameters are the filter coefficients that are dynamic with frames to synthesize speech. The pitch is used to determine the pitch period. The gain is used to scale the volume of frame generation.
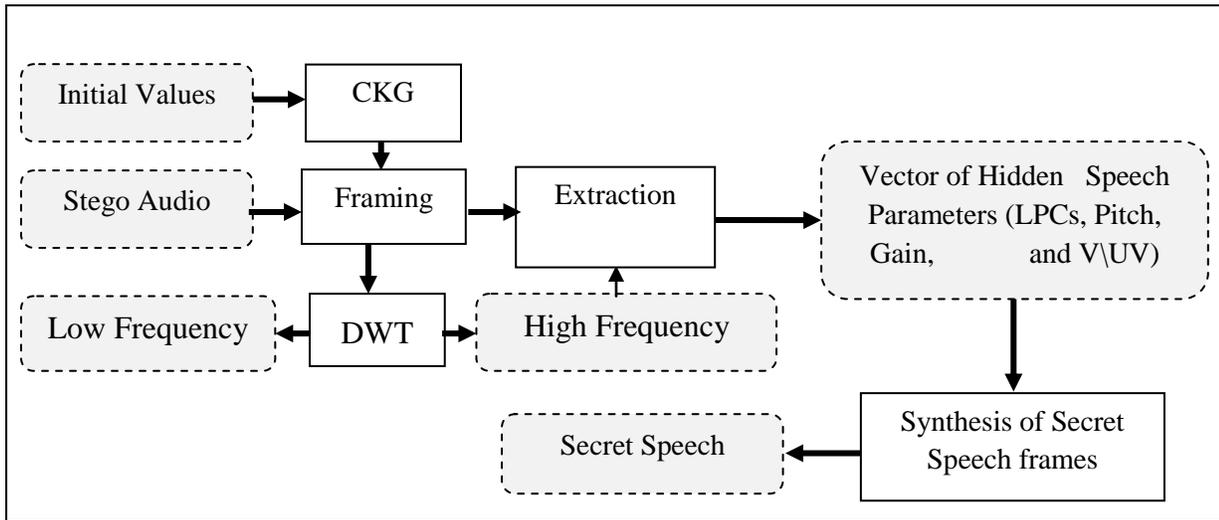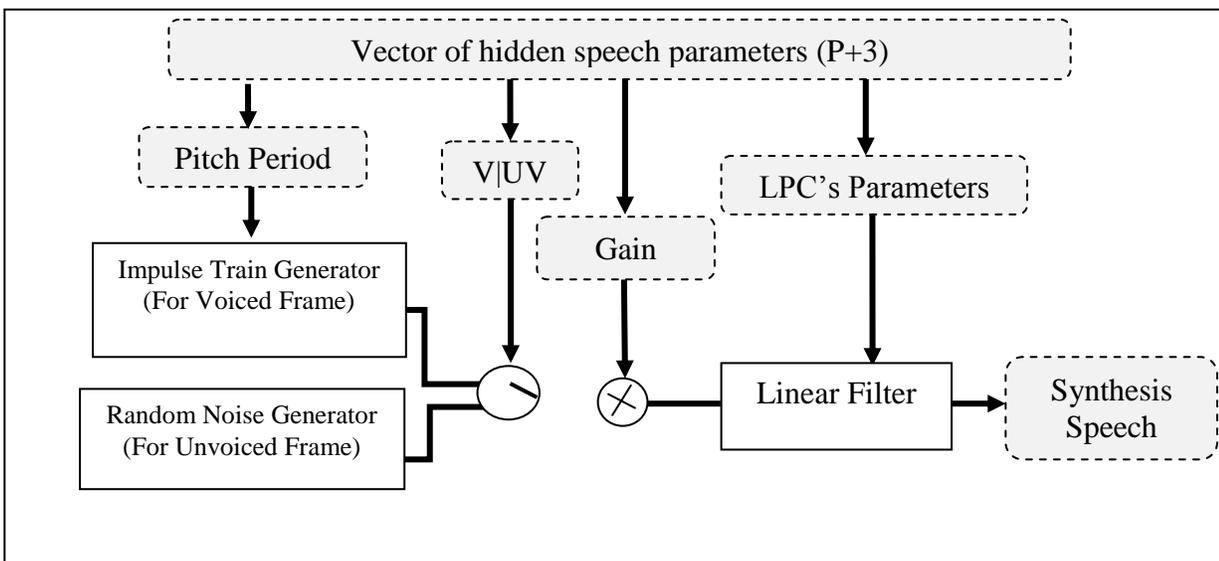
Fig. (5) extraction phase (extraction of hidden speech)

Fig. (6) principle of hidden speech synthesis

34

www.uoitc.edu.iq

Networks Security & Distrubuted Systems
(NSDS'**2015**)

The Annual
Conference On
Networks
Security
&
Distributed
Systems
(NSDS'2015)

25 - 26 November , Baghdad IRAQ

VI. Results

Audio quality assessment falls into one of two quality measures categories: subjective and objective. Objective measures do not require human listeners, and so is less expensive and less time-consuming than subjective measures. However, there are examples where the quality estimations from objective measures and subjective measures do not match. Thus, subjective quality measures are still the most conclusive way to measure the perceived quality. Recent objective measures are good estimators of subjective quality, which can be used to get a rough quality estimate, which can be followed by subjective quality assessment for selected conditions to confirm the perceived quality. In the proposed system, the selected language is Arabic. Males and females speech are selected to embed (secret) in cover audio. There are many objective measures applied to performance Audio quality test as shown in table (1) such as signal to noise ratio

(SNR) and signal to noise ratio segmental (SNRseg) base time domain, Segmental Spectral SNR based frequency domain, and Log Likelihood Ratio (LLR) and Correlation (Rxy) base LPC domain.

- The Reconstructed Speech

The quality of reconstruction of hidden speech depend on the recovering the features of speech. The LPC is lossy method and the synthesis of speech signal differs from original signal. The reconstruction of secret speech should be as near as possible to the coded speech. Figures (7 and 8) show the LPC speech signal before embedding and after extraction for male and female speech respectively. It has been shown clear similarity.

Table (1)   audio objective measurements

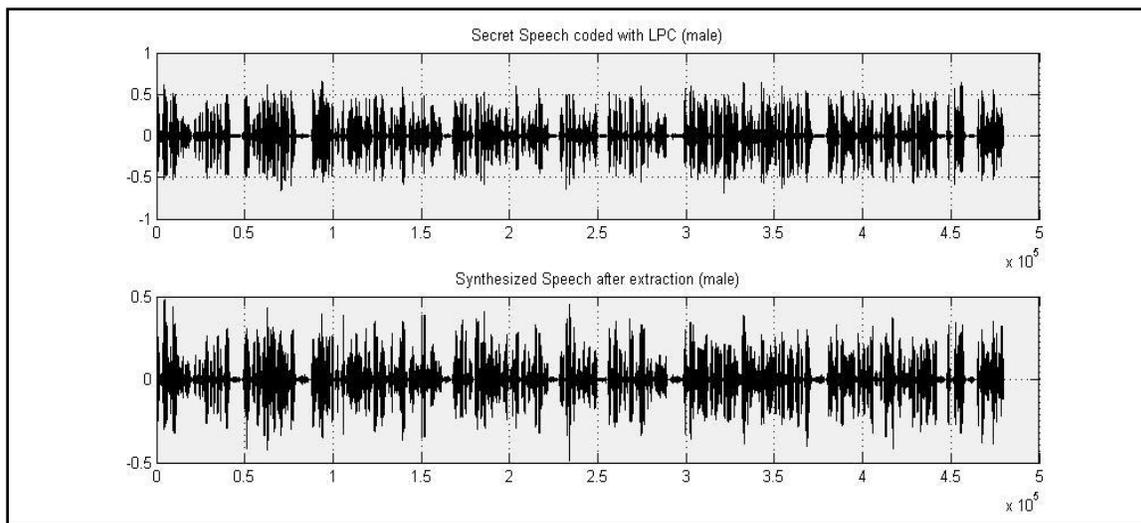| Secret Speech	Cover Audio	SNR	SNRseg	SSNRseg	LLR	R(x,y)
Female1	Audio1	21.8259	13.9936	54.4527	0.4321	0.8978
Female2	Audio2	24.1384	18.5832	59.4269	0.2777	0.9366
Male1	Audio3	23.2528	16.0748	56.6289	0.3796	0.9085
Male2	Audio4	18.9430	10.7251	51.1296	0.4984	0.8776
Audio Single Chanel	Audio5	22.8080	15.9720	53.2911	0.4162	0.8966
Average		22.1936	15.0697	54.9859	0.4008	0.9034

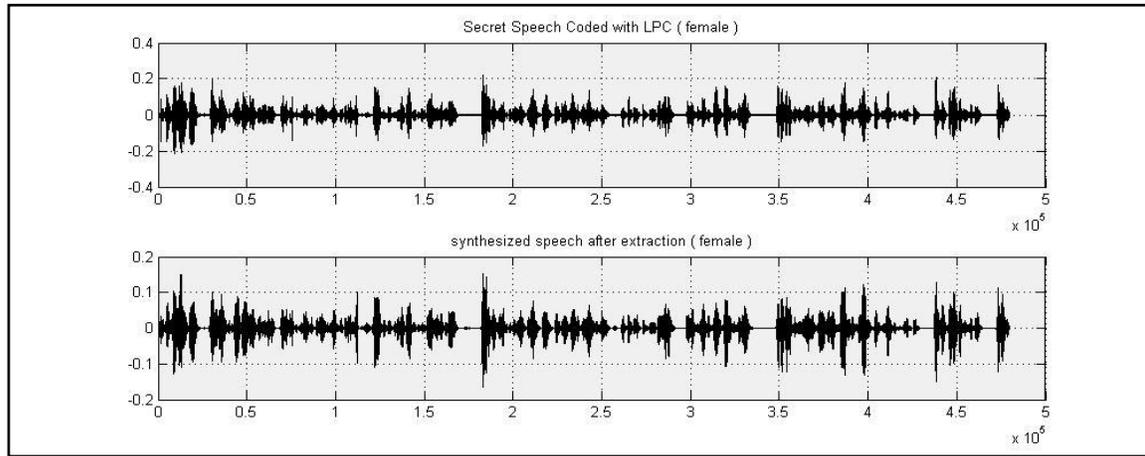Fig. (7) the original and stego signal of embedding male in audio

35

Fig. (8) original and stego signal of embedding female in audio

VII. Conclusion

This work presents a speech steganography algorithm base on:

1. LPC as good tools for speech coding. It gives a low bit rate (about 2.4 kbps).

2. Wavelet transform as a good tool for splitting low frequency from high frequency and ability to extract new properties not found in time domain.

3. Chaotic map to generate a random numbers. It is used as a tool for selection the samples of each frame to store the important features of secret speech.

4. Audio files as good choice for file rather than speech file.

The future work of this approach is to find the algorithm to hide compression speech in another speech signal that also compressed that will be more efficient to sending and receiving over transmission media or stego speech signal also compressed in the sender side and decompressed in receiver side.

References

[1] Fatiha Djebbar, Beghdad Ayad, Karim Abed Meraim and Habib Hamam "*Comparative study of digital audio steganography techniques*" EURASIP Journal on Audio, Speech, and Music Processing, Springer, 2012.

[2] F. Djebbar, B. Ayad , K. Abed-Meraim and H. Hamam , *"A view on latest audio steganography"*, 7th IEEE International Conference on Innovations in Information Technology, Abu Dhabi, UAE, 2011.

[3] W. Bender, D. Gruhl, N. Morimoto, and A. Lu *"Techniques for data hiding"* IBM System Journal, Volume 35, NOS 3&4, 1996, USA.

[4] Hiba Mohammed Saleem AL-Kawaz "*Low Rate Hiding in Audio Data Using Phase Domain*" M.Sc. thesis, College of Science of AL-Nahrain University, 2007.

[5] Wai C. Chu "*Speech Coding Algorithms Foundation and Evolution of Standardized Coders*" Wiley Interscience, 2003.

[6] Amol R. Madane, Zalak Shah, Raina Shah, and Sanket Thakur, "*Speech Compression Using Linear Predictive Coding*" Machine Intelligent Research, 2009.

[7] Wissam Abdulhadi "*Design and Implementation of High Speech Quality*"

36

Vocoder System at 2400 bps Using MELP Model" MSc. Thesis, Electrical Engineering University of Baghdad, 2002.

[8]　Visala Namburu "*Speech Coder using Line Spectral Frequencies of Cascaded Second Order Predictors*" MSc. Thesis, The Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, 2001.

[9]　Brigham, E. Oran **"The fast Fourier transform and its applications"** Englewood Cliffs, N.J. Prentice Hall, 1998.

[10]　Mallat, S. G. "*A Theory for Multi Resolution Signal Decomposition: The Wavelet Representation*" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, 1989.

[11]　Steven W. Smith *"The scientist and Engineers Guides to Digital Signal Processing"* 2nd edition, California Technical publishing, USA, 1999.

[12]　David Arroyo, Gonzalo Alvarez, and Veronica Fernandez "*On the inadequacy of the logistic map for cryptographic applications*" Instituto de FısicaAplicada, Consejo Superior de Investigations Cient´ıficas, Serrano Madrid, Spain, 2008.

[13]　Osama S. Faragallah **"An Efficient Block Encryption Cipher Based on Chaotic Maps for Secure Multimedia Applications"** Information Security Journal: A Global Perspective, Vol. 20, No. 3, 2011.

37

　　　　　　　　　　　　　　www.uoitc.edu.iq