

# Speech Steganography System Using Lifting Wavelet Transform

Lecturer. Ahlam majead Kadum  
 Physics AL- Mustansiriyah University

Professor .Dr. Saad Najim Al-Saad  
 Computer sciences AL- Mustansiriyah University

## Abstract

This paper presents a new lossless speech steganography approach based on Integer-to-Integer Lifting Wavelet Transform (Int2Int LWT) and Least Significant Bits (LSBs) substitution. In order to increase the security level a simple encryption with chaotic key has been proposed. The proposed system has a high sensitivity in choosing keys because a small change in CKG causes a new secret key for transmitting. Speech steganography algorithm that based on (Int2IntLWT) can satisfy full recovery for the embedded secret messages in the receiver side.

**Keywords:**Speech steganography, information hiding, Int2Int LWT, (LSB) technique, XOR operation.

## 1 Introduction

Discrete wavelet transform (DWT) is widely used for analyzing signals, steganography art, compression and noise reducing. DWT implements multi resolution analysis for the signals that have an adjustable location in each of space (time) and frequency domains. Because of large amount of calculations required, and there have been many research efforts to improve DWT and give a new fast algorithms that are used for performance DWT. The major challenges in the buildings devices for 1-D DWT and 2-D DWT is the speed processing and the number of multiples, where the memory issue which dominate the hardware cost and complexity of the architecture<sup>[1]</sup>.

## 2 Lifting Wavelet Transform (LWT)

Lifting scheme is introduced to fast DWT, this easily achieved by the computer due to the great reduction in calculations. LWT scheme usually requires less mathematical operations compared with traditional approach convolution. LWT achievement does not require additional memory because of the in-place calculation features of the lifting. This is particularly suitable for the devices implementation of with a limited memory. LWT scheme submitted integer to integer transformation appropriate for lossless processing signal <sup>[2]</sup>. This approach is totally based on the spatial performance of the DWT. Basic concept of LWT is to exploit the correlation infrastructure that present in most real life signals to build a dispersed approximation. Correlation infrastructure is normally localization into space (time) and frequency; neighboring samples and frequencies are much more interconnected from that are in distant from. Figure (1) represents the forward transform scheme of three levels of three levels LWT with the three stages (split, production and update) <sup>[3, 4]</sup>:

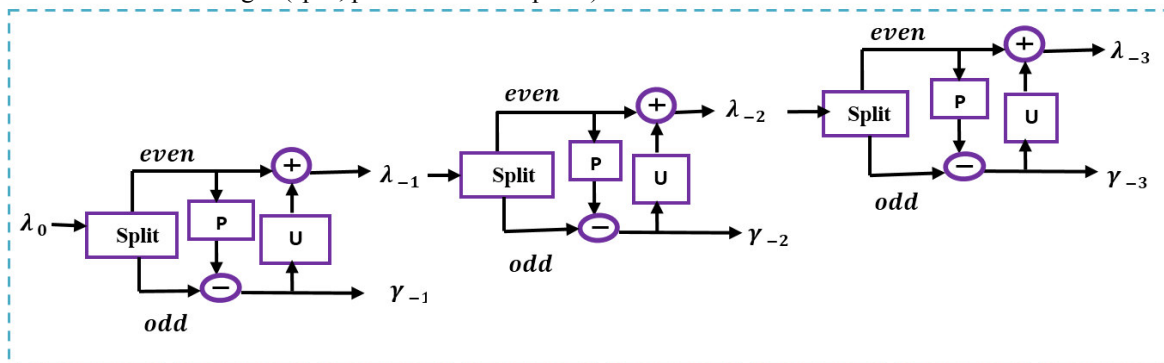


Figure (1) forward transforms operation for three levels of LWT

### 2.1 Split stage

This stage divides the set of signal into two frames:<sup>[4,5]</sup>

1. The first frame consists of even index samples such as  $(\lambda_{0,0}, \lambda_{0,2}, \lambda_{0,4}, \dots, \lambda_{0,2k})$ . We will call this frame coarser resolution signal or approximation.

$$\text{even} = \lambda_{0,2k} \dots (1)$$

2. The second frame consists of odd index samples such as  $(\lambda_{0,1}, \lambda_{0,3}, \lambda_{0,5}, \dots, \lambda_{0,2k+1})$ . We will call this frame as smoother resolution signal or detail.

$$\text{odd} = \lambda_{(0,2k+1)} \dots (2)$$

Each signal in first and second frames consists of  $\frac{N}{2}$  samples from the original signal samples. In split stage any mathematical operations are not performed. Splitting signal into two parts is called *lazy wavelets*

### 2.2 Prediction stage (Dual lifting)

Predicting the odd coefficient basis of the linear combination of even samples and odd samples, this predicted stage is also referred to as the *dual lifting step*; the lost data are simply incorporated in odd coefficient [4, 6]. Predict the odd samples by using linear interpolation predict the odd coefficient based on a linear combination of even samples and odd samples (replace  $(\lambda_{0,2k+1})$  with  $\gamma_{-1,k}$ ) as follow:

$$\gamma_{-1,k} = \lambda_{0,2k+1} - P(\lambda_{-1,k}) \dots \dots \dots (3)$$

Odd value Predicted value

$$P(\lambda_{-1,k}) = \frac{1}{2} (\lambda_{0,2k} + \lambda_{0,2k+2}) \dots \dots \dots (4)$$

Substitute's equation (2) in (1) getting equation (3):

$$\gamma_{-1,k} = \lambda_{0,2k+1} - \frac{1}{2} (\lambda_{0,2k} + \lambda_{0,2k+2}) \dots \dots \dots (5)$$

### 2.3 Update stage (Primal lifting)

Update the even samples based on a linear combination of difference samples obtained from the predict stage. We require constructing update operator U for this lifting process [4, 6].

$$\lambda_{-1,k} = \lambda_{0,2k} + U(\gamma_{-1,k}) \dots \dots \dots (6)$$

$$U(\gamma_{-1,k}) = \frac{1}{4} (\gamma_{-1,k-1} + \gamma_{-1,k}) \dots \dots \dots (7)$$

$$\lambda_{-1,k} = \lambda_{0,2k} + \frac{1}{4} (\gamma_{-1,k-1} + \gamma_{-1,k}) \dots \dots \dots (8)$$

## 3 Chaotic Key Generation (CKG)

An important feature of chaos systems is their ability to produce very complicated patterns of behavior. This quality has made them especially advantageous for the application in a wide variety of disciplines, such as biology, economics, engineering, signal processing, secure communications, and compression the information and data encryption. In such applications, chaotic systems are used to produce the chaos, simulation, help or control of the various processes and improving their performance or provide more convenient output [7].

One of the simplest chaotic functions that have been studied recently for cryptography applications is the logistic map. The logistic map function is expressed as follows [8]:

$$x_{n+1} = r \cdot x_n \cdot (1 - x_n) \dots \dots \dots (9)$$

Where  $x_n$  takes value in the interval (0, 1), the parameter  $r$  is a positive constant taking values up to 4. Its value determines and explores the behavior of the logistic map.

## 4 The Proposed System Design

The basic design of the proposed steganography system consists of two phases embedding and extraction. In embedding Phase the sender side hide secret message inside a speech signal (male speaker, female speaker) each signal represented with bit resolution and frequency rate 16 bits /sample, 8000 Hz /sec respectively. The choice of speech signal should be suitable size and enough to embedding the message. The proposed system allows to choosing any speech cover and any secret message, there is no limitation for the cover size and message size. The proposed steganography system allows the user to hide any kind of electronic signals after converting the value of secret message data to binary digital system numbers. The user can hide a small message or large message under after comparing size cover with size message And make sure that the speech is enough to hide all message data within. The embedding phase contains two stages, they are as follows:

### 1. Preprocessing stage

The preprocessing stage is depicted in figure (2).The figure illustrates the steps of preprocessing stage.

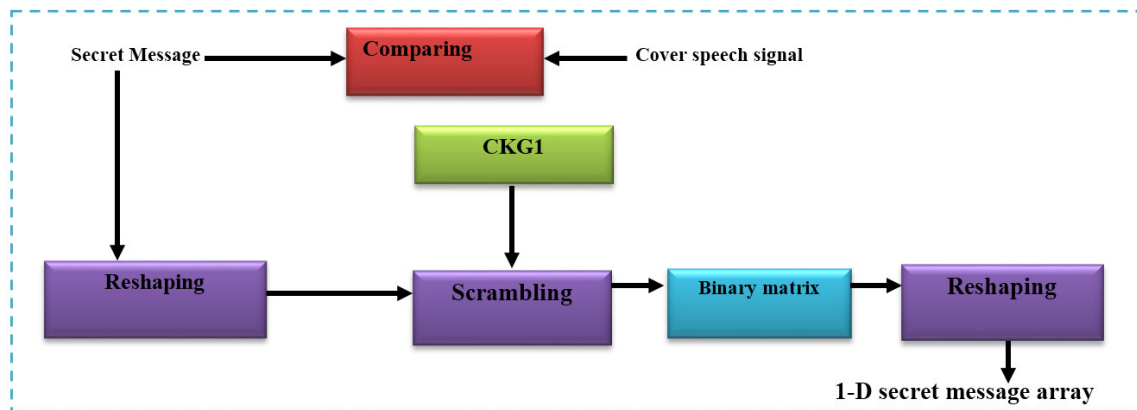


Figure (2) Block diagram for preprocessing stage

The inputs are secret message and CKG1 generated from algorithm (1).

**Algorithm (1): Preprocessing stage**

**Input:** Message // Secret message  
 Cover // Cover sound file  
 X // Number LSB replaced for each coefficient of LWT  
 W // Length each frame  
 CKG 1 // First chaotic Key Generation

**Output:** Scrambled message, Capacity (CP), Length message (Len)

**Began:**

**Step1:** Reade secret message and calculate its size

$msg \leftarrow imread(message)$

$[c1\ c2\ c3] \leftarrow size(msg)$

**Step2:** Reade sound cover and finding its bits resolution (nbits) and rate samples( $F_s$ ) and its size

$[Y, F_s, nbits] \leftarrow wavered(Cover)$  // Store sound cover in matrix Y

$[c4, c5] \leftarrow size(Y)$

**Step3:** Calculate size secret message and size of cover file

$Len \leftarrow c1 * c2 * c3 * 8$      $L \leftarrow c4 * c6 * nbits$

**Step4:** Calculate total number of frames in sound cover ( $Frm_{cov}$ ) and total frames that will be needed to hide message ( $Frm_{msg}$ ) and residue set bits (Q1)

$Frm_{cov} \leftarrow \frac{L}{W}$

$Frm_{msg} \leftarrow \text{fix}[\frac{Len}{X * (\frac{W}{2} + \frac{W}{4})}]$

$Q \leftarrow \text{mod}[\frac{Len}{X * (\frac{W}{2} + \frac{W}{4})}]$

**Step5:** Compering the cover size with message size

If  $Frm_{msg} > Frm_{Cov}$

Error  $\leftarrow$  Message Box (Cover is small to hide this message)

Break

End if

**Step6:** Calculating capacity of cover and compering the size cover with size message in bits

$Capacity \leftarrow \frac{Len * 100}{L}$

**Step7:** Scrambling secret message data in random locations

$msg2 \leftarrow msg1(CKG1)$

**Step8:** Save Scrambled message

$msg22 \leftarrow reshapeing(msg2, c1, c2, c3)$

$secrambl\ message \leftarrow save(msg22)$

**Step9:** Converted ( $msg2$ ) from decimal to binary with 8 bites to get data values

$msg3 \leftarrow decimal\ to\ binary(msg2)$

**Step10:** reshaping ( $msg3$ ) from matrix 8 column to matrix 1 column and calculate its length

$Len \leftarrow length(msg3)$

## 2. Embedding stage

The speech signal is framed to the number of frames, each frame contain 512 samples. Int2IntLWT has been used to convert speech signal to the frequency domain. First and second levels of Int2IntLWT were implemented for each frame where the results were four sub bands matrixes. High frequency for sub-band1 and sub-band2 were used for embedding operation. The secret message data were scrambled by using first chaotic key generation (CKG1), and then the message was divided in a number of blocks bits. The block of the message was divided into two bits sets. The two sets were embedded within each frame of cover by replacing LSBs method for each coefficient of two high frequency sub-band. The embedding stage is depicted in figure (3).

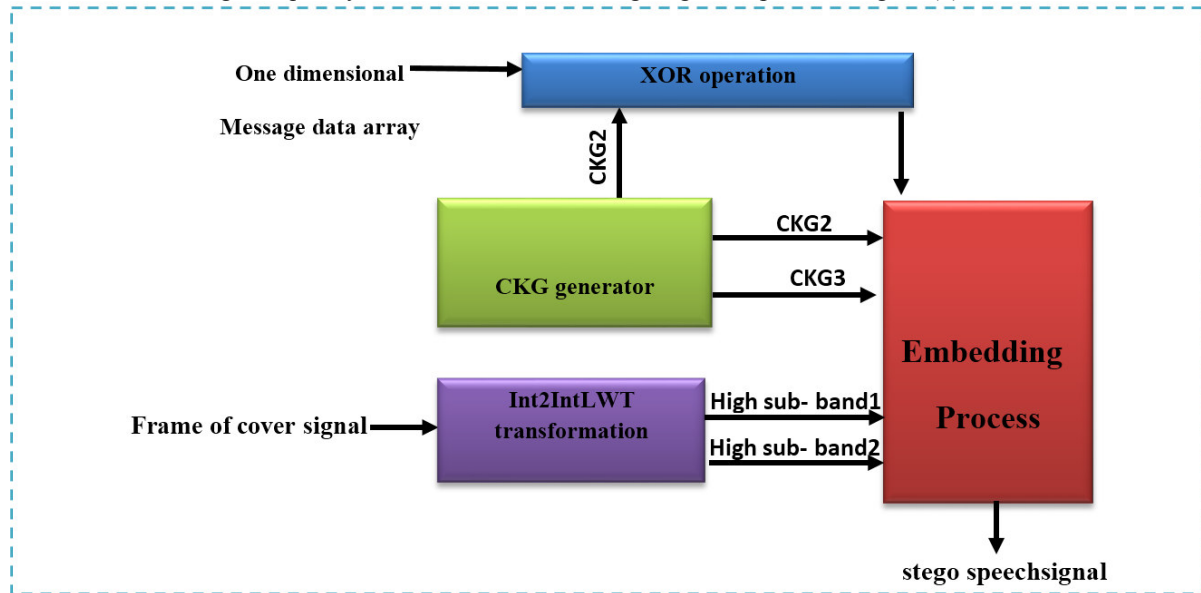


Figure (3) Block diagram for embedding stage

Two chaotic keys (CKG1, CKG2) also were used to select the chaotic indexes of coefficients that their LSBs are replaced with message bits. XOR operation was performed between the block message and the chaotic numbers resulted from CKG1. The embedding stage algorithm is listed in algorithm (2):

### Algorithm (2): Embedding secret message

**Input:**  $y$  : Matrix of Cover sound file  
 $msg3$  : Binary matrix of secret message  
 $X$  : Number LSBs replaced for each coefficient of Int2IntLWT  
 $W$  : Length each frame  
 $Frm_{msg}$  : **Number frames that needed to embed secret message**  
 $U$  : Index bit of the real number of CKG2 for the purpose of XOR operation  
 CKG2, CKG3: chaotic keys using to select coefficients positions.

**Output:** Stego : Stego speech signal

**Began:**

**Step1:** Beginning the hiding process within all frames of cover signal that are needed to hide message

$F1 \leftarrow 0$

$F2 \leftarrow 1$

**For**  $i \leftarrow 1$  **to**  $Frm_{msg}$

$Frm \leftarrow Y(F1 * W: F2 * W)$

**Step2:** implemented 2 levels Int2IntLWT for each frame

$[low1 (1 : \frac{W}{2}), high1 (1 : \frac{W}{2})] \leftarrow Int2IntLWT(Frm)$

$[low2 (1 : \frac{W}{4}), high2 (1 : \frac{W}{4})] \leftarrow Int2IntLWT(low1)$

**Step3:** Cutting set of bit from real number of CKG2 and put it in new matrix

$V1 \leftarrow gettingbits(CKG2((1 : \frac{W}{2}), U))$

$V2 \leftarrow V1(1 : \frac{W}{4})$

**Step4:** Implemented XOR operation and replace LSB of matrixes coefficients high sub-band1

$a1 \leftarrow 0$

$a2 \leftarrow 1$

**For**  $j \leftarrow 2$  **to**  $X$

```

msg4(a1 *  $\frac{W}{2}$ : a2 *  $\frac{W}{2}$ ) ← XOR(msg3(a1 *  $\frac{W}{2}$ : a2 *  $\frac{W}{2}$ ), V1)
high11(CKG2(a1 *  $\frac{W}{2}$ : a2 *  $\frac{W}{2}$ )) ← replace(high1(CKG2(a1 *  $\frac{W}{2}$ : a2 *  $\frac{W}{2}$ )), j, msg4(a1 *  $\frac{W}{2}$ : a2 *  $\frac{W}{2}$ ))
[] ← msg4(a1 *  $\frac{W}{2}$ : a2 *  $\frac{W}{2}$ )
a1 ← a1 + 1
a2 ← a2 + 1
End for // j
Step5: Implemented XOR operation and replace LSB of matrixes coefficients high sub-band2
b1 ← 0
b2 ← 1
For k ← 2 to X
msg4(0:  $\frac{W}{4}$ ) ← XOR(msg3(0:  $\frac{W}{4}$ ), V2)
high22(CKG3(b1 *  $\frac{W}{4}$  + b2 *  $\frac{W}{4}$ )) ← replace(high2(CKG3(b1 *  $\frac{W}{4}$ : b2 *  $\frac{W}{4}$ )), k, msg4(0:  $\frac{W}{4}$ ))
[] ← msg4(0:  $\frac{W}{4}$ )
b1 ← b1 + 1
b2 ← b2 + 1
End for // k
Step5: Implemented invers Int2IntLWT for two levels
high111 ← InversInt2It LWT[low2, high22]
D1 ← InversInt2It LWT[low1, high111]
Stego(F1 * W: F2 * W) ← D1
Step6: Ending hiding process for one frame
F1 ← F1 + 1
F2 ← F2 + 1
End for // i
End
    
```

Figure (4) shows the steps of extraction stage. It is implemented as the same way of embedding stage but in reverse form.

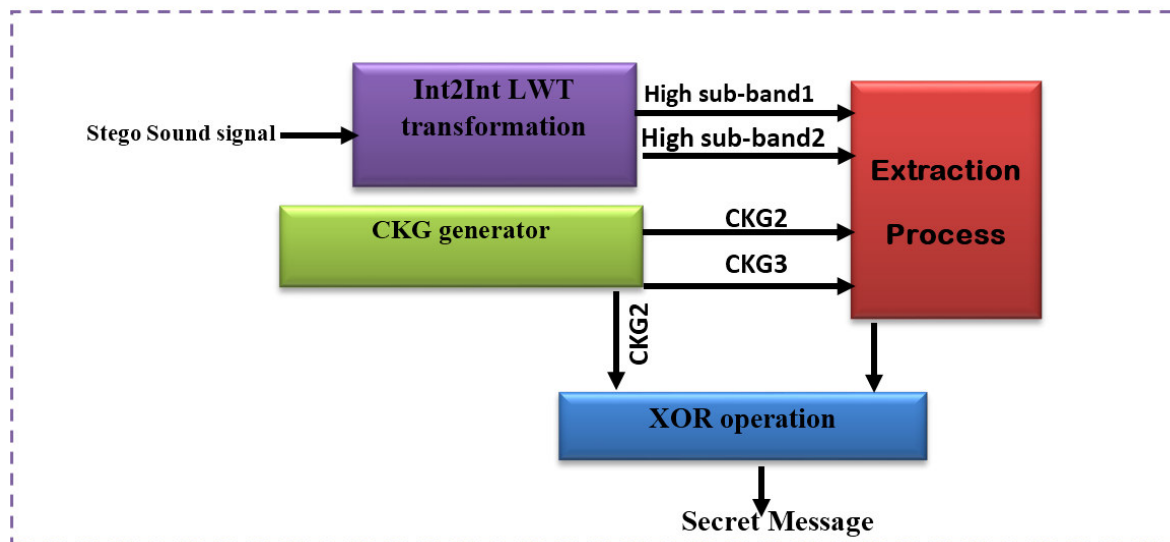


Figure (4) Block diagram for extraction phase

## 5 Experimental results

The proposed algorithm show efficiency of hiding in terms of security level, stego signal properties are unchanged as a result of hiding secret message. The experimental result points to that the stego file is undetectable and imperceptible by the HAS. Figure (5) shows waveform of stego speech file and it's original.

Several testing measurements for the quality of stego signals are presented, and three types of secret messages have been embedded within speech signals:

- Color image message of size (512\*512).
- Sound speech message of time (2 minutes) with bit resolution (8 bits/sample) and frequency (8000

Hz/sec).

- Text message with length 15206 character.

Figures (6, 7 and 8) show the three type of message in three cases of original, encrypted and extracted message. Tables (1, 2 and 3) illustrate the objective quality measurements using four files of speech covers. The tables pointing that the quality measurements signal to noise ratio (SNR), signal to noise ratio segmental ( $SNR_{seg}$ ) and signal to noise ratio spectral ( $SNR_{spec}$ ) are decreasing with increasing the replaced LSBs numbers. The measure MSE increased with increasing the replaced LSBs numbers, because of increasing error in host signal. Finally the parameter runtime and correlation test ( $R_{xy}$ ) are decreasing with increasing the replaced LSBs numbers.

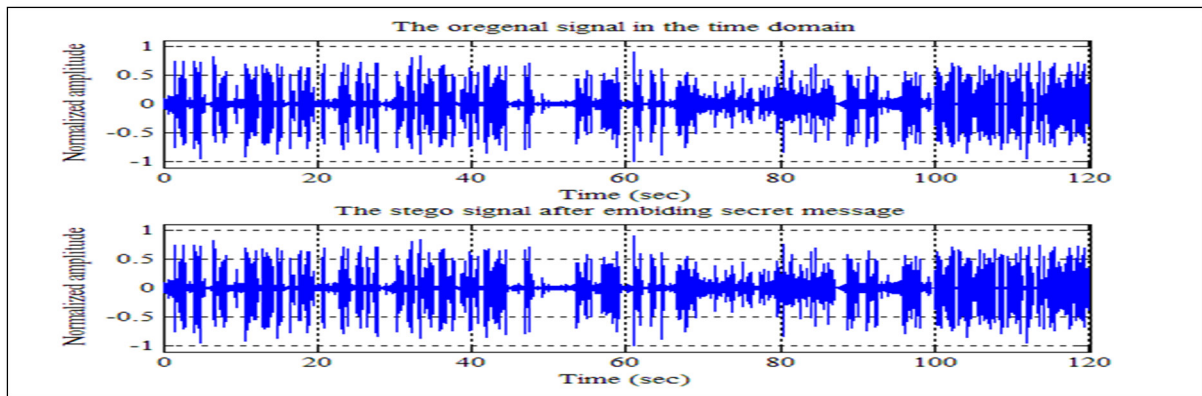


Figure (5) Waveform for the original and stego speech signal

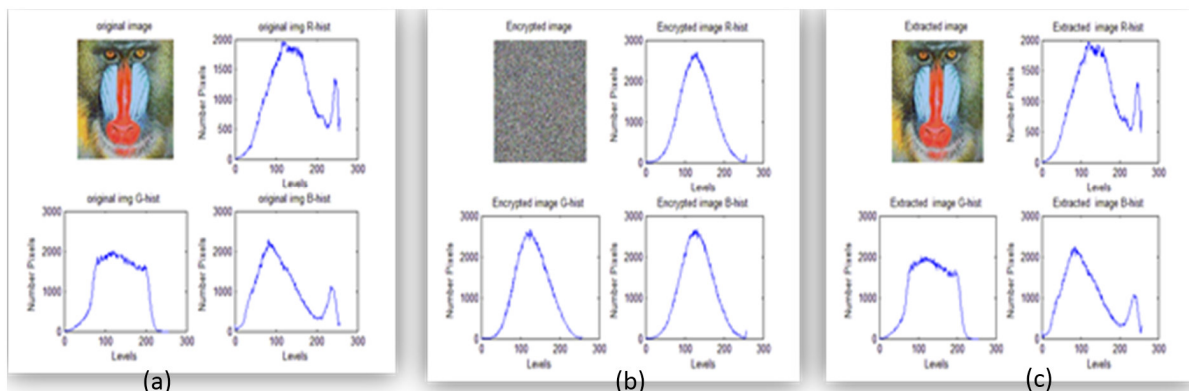


Figure (6) Histogram of components (RGB) (a) original image (b) scrambled image (c) Extracted image

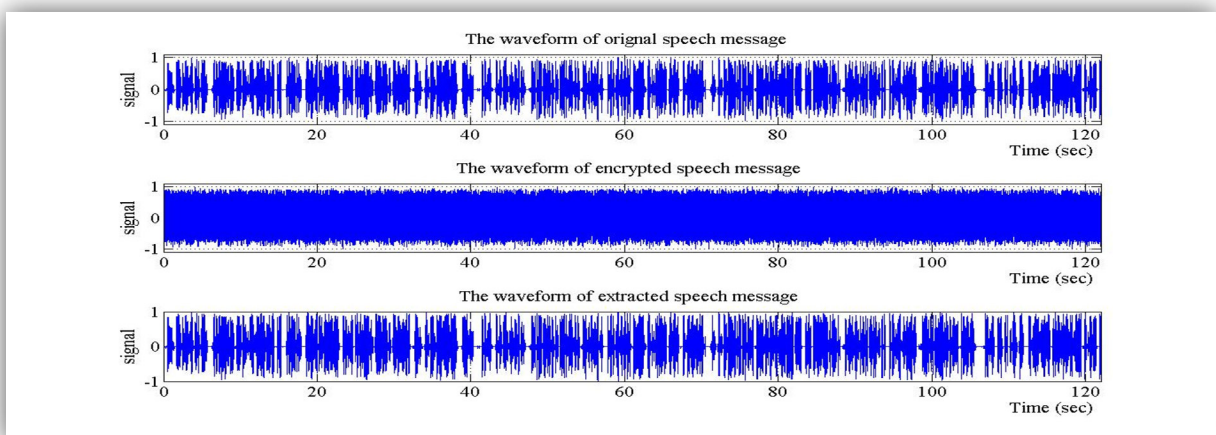


Figure (7) the waveform of secret speech message before and after embedding operation.

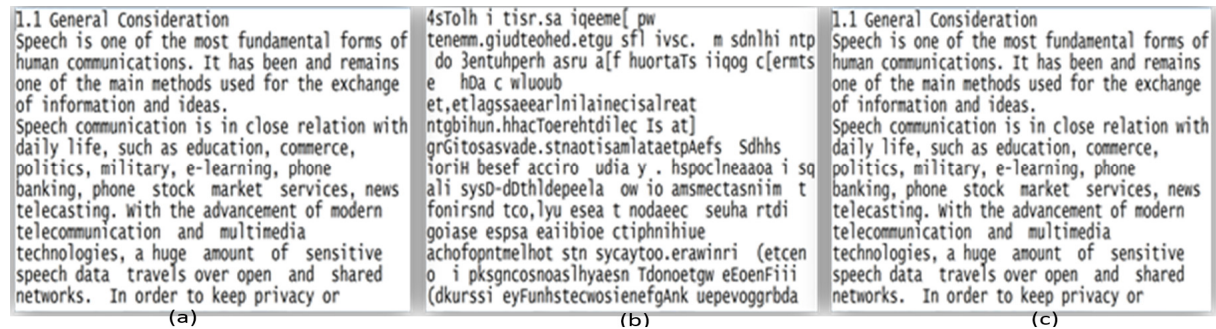


Figure (8) the form of a text message (a) original text, (b) scrambled text and (c) extracted text  
 Table (1) objective measurements of hiding Baboon image (256\*256) within speech file (4 minutes)

cover name	LSB replaced	Run time	SNR db	SNRseg db	SNRspc db	MSE	R <sub>xy</sub>
female1	2	36.327394	59.4319	57.0692	60.1042	2.5397e-09	0.9950
	4	24.230065	50.1294	45.1449	48.1344	2.1629e-08	0.9816
	6	20.363995	39.6915	33.1455	36.1945	2.3923e-07	0.9445
	8	19.084399	28.2820	28.5295	31.4593	3.3096e-06	0.8720
	10	17.336183	16.5483	21.2981	24.3148	4.9333e-05	0.7574
fremale2	2	36.115240	58.2740	54.7049	57.7147	2.5393e-09	0.9946
	4	24.355155	48.9569	40.9612	43.9742	2.1699e-08	0.9711
	6	20.378122	38.1407	34.9588	37.9030	2.6185e-07	0.8656
	8	18.549485	26.6096	17.4292	20.4468	3.7253e-06	0.7487
	10	17.384655	14.9634	15.1575	18.1441	5.4423e-05	0.6385
male1	2	33.525607	58.1688	56.5986	59.6201	2.5490e-09	0.9952
	4	22.832595	48.8850	44.5108	47.5558	2.1615e-08	0.9842
	6	19.068775	38.4122	33.1276	36.1871	2.4101e-07	0.9362
	8	17.411953	27.1528	36.0960	39.0693	3.2209e-06	0.8681
	10	16.210631	15.3513	22.0099	25.0180	4.8767e-05	0.7528
male2	2	33.550752	57.1845	54.8371	57.8703	2.5346e-09	0.9956
	4	23.590319	47.8666	41.9465	44.9641	2.1662e-08	0.9907
	6	19.592131	37.5255	31.2427	34.4402	2.3432e-07	0.9792
	8	17.513022	26.2267	30.8793	32.8737	3.1600e-06	0.9323
	10	16.162748	13.9088	17.5552	20.2395	5.3885e-05	0.7117

It is Possible to hide more than one secret message inside the same speech cover under the condition of suitable size cover with the required number messages that hiding in it. It is possible to use speech signal as channel to send any secret message between two persons for the purpose of secret communication with good security.

Table (2) objective measurements of hiding speech message (1 minute) within speech file (4 minutes)

cover name	LSB replaced	Run time sec	SNR db	SNRseg db	SNRspc db	MSE	R <sub>xy</sub>
female1	4	68.686385	45.5489	44.2016	47.1463	6.2097e-08	0.9875
	6	47.528121	35.0978	32.6359	35.6429	6.8896e-07	0.9444
	8	38.115784	23.7502	29.5613	32.4736	9.3961e-06	0.8578
	10	38.134364	12.0505	22.8450	25.7964	1.3897e-04	0.7250
fremale2	4	69.339644	44.3984	42.1503	45.1076	6.1984e-08	0.9780
	6	49.710716	33.6956	29.1762	32.1951	7.2873e-07	0.8690
	8	40.494508	22.1310	23.3020	26.4488	1.0448e-05	0.7131
	10	33.894926	10.4819	11.6270	13.5941	1.5273e-04	0.5688
male1	4	71.873835	44.2985	43.6775	46.6317	6.2144e-08	0.9899
	6	49.636455	33.8555	31.7001	34.7058	6.8818e-07	0.9372
	8	39.845036	22.5674	26.2030	29.0627	9.2580e-06	0.8479
	10	34.785053	10.8187	20.5055	23.2936	1.3848e-04	0.6984
male2	4	69.719489	43.2972	41.6885	44.6671	6.2034e-08	0.9958
	6	49.816237	32.9721	30.2385	33.2578	6.6856e-07	0.9876
	8	39.609890	21.8051	21.8447	24.5496	8.7467e-06	0.9431
	10	34.026424	9.6500	19.1549	22.0956	1.4366e-04	0.7064

Table (3) objective measurements of hiding text1 (letters) within speech file (4 minutes)

cover name	LSB replaced	Run time sec	SNR db	SNRseg db	SNRspc db	MSE	R <sub>xy</sub>
female1	2	7.612382	66.6470	56.6180	59.1459	4.8225e-10	0.9996
	4	7.335659	56.7899	45.1078	47.3353	4.6664e-09	0.9959
	6	6.668128	46.2650	30.7271	33.3725	5.2659e-08	0.9643
	8	6.465787	35.0710	42.8864	43.7286	6.9321e-07	0.8787
	10	6.445262	23.4215	39.0612	39.8034	1.0135e-05	0.6877
fremale2	2	7.282281	65.4896	52.7026	55.3838	4.8213e-10	0.9978
	4	7.173080	55.6893	38.1206	40.9617	4.6047e-09	0.9725
	6	6.930713	45.1200	29.7603	32.8658	5.2496e-08	0.9577
	8	6.811351	34.0262	38.6654	39.6550	6.7531e-07	0.8850
	10	6.774821	22.3195	10.0244	11.2429	1.0004e-05	0.6480
male1	2	7.240987	65.3927	56.6536	59.2034	4.8304e-10	0.9995
	4	6.918434	55.5616	45.4700	47.7950	4.6461e-09	0.9931
	6	6.533973	44.9949	44.1163	45.0095	5.2937e-08	0.9691
	8	6.513125	33.9674	35.3275	36.9053	6.7067e-07	0.9133
	10	6.487728	22.2459	21.9312	23.4187	9.9693e-06	0.7419
male2	2	7.146733	64.3663	52.5130	55.2585	4.8498e-10	0.9996
	4	6.746366	54.5640	41.1123	43.7542	4.6340e-09	0.9989
	6	6.519577	44.0086	39.8647	40.8738	5.2662e-08	0.9939
	8	6.459934	32.7326	29.2785	30.1296	7.0647e-07	0.9089
	10	6.427240	21.0165	21.0183	21.9476	1.0488e-05	0.9011

## References

- [1] Sayed Ahmad Salehi, Rasoul Amirfattahi, "VLSI Architectures of Lifting-Based Discrete Wavelet Transform", Isfahan University of Technology, Iran 2011.
- [2] Tinku Acharya, Chaitali Chakraabati, "A Survey on Lifting-based Discrete Wavelet Transform Architectures", Springer, Volume 42, Issue 3, pg. (321-339), March 2006
- [3] Ingrid Daubechies, Wim Sweldens, "FACTORING WAVELET TRANSFORMS INTO LIFTING STEPS", NSF (grant DMS-9401785), AFOSR (grant F49620-95-1-0290), ONR (grant N00014-96-1-0367) Princeton University, Princeton, New Jersey, 1996.
- [4] "Lifting Scheme of Wavelet Transform"  
[http://shodhganga.inflibnet.ac.in/bitstream/10603/4341/7/07\\_chapter%203](http://shodhganga.inflibnet.ac.in/bitstream/10603/4341/7/07_chapter%203).
- [5] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, "Wavelet Toolbox™ User's Guide", Book, COPYRIGHT 1997–2014 by The Math Works, Inc, 2014.
- [6] GEERT UYTTERHOEVEN "Integer Wavelet Transforms using the Lifting Scheme", Katholieke Universiteit Leuven, IEEE, IMA CS, OTE, p(6253-6257), 1999.
- [7] Eman Hato Hashim, "Speech Signal Encryption Using Chaotic Maps", M.Sc. Thesis, Al Mustansiriyah University, Computer Science department, September 2013.
- [8] Jamal Nasir Hasoon, "Speech Hiding Using Vector Quantization", M.SC. Thesis, AL-Mustansiriyah University, January 2014.