A NEW SPEECH COMPRESSION MODEL USING DISCRETE WAVELET TRANSFORM

Sabah S. Alsheraidah
(Assistance Professor, Basrah Technical Institute,)
(E-mail: sabahhd@yahoo.com)

Abstract

The purpose of this research is to design and implement a new Discrete Wavelet Transform (DWT) encoder/decoder for speech compression. Three compression procedures are suggested these are, Wavelet Analysis-Synthesis, Wavelet Global Threshold and Wavelet By-Level Threshold. The major issues concerning the design of this Wavelet based speech coder are proposed mother wavelet for speech signal, decomposition level in DWT, threshold criteria for coefficient truncation and efficient encoding of truncated coefficients. Performances of these three procedures are compared and the encoded methods are evaluated based on compression ratio and signal to noise ratio factors.

KEYWORD: WAVELET, SPEECH, COMPRESSION, MODEL

نموذج جديد لضغط الكلام باستخدام التحويل ألمويجي المتقطع

صباح سليم الشريدة
(أستاذ مساعد/ المعهد التقني في البصرة)

الخلاصة

الهدف من هذا البحث هو تصميم وبناء مشفر/فك شفرة جديد باستخدام التحويل ألمويجي المتقطع لضغط الكلام. تم اقتراح ثلاث انهج ضاغطة باستخدام التحويل ألمـويجي هـم Wavelet Analysis-Synthesis, Wavelet Global Threshold and Wavelet By-Level Threshold . إن المسائل الرئيسية التي تتعلـق بتصميم نظام ضغط الكلام هي اختيار المويجة الأم لإشارة الكلام ومستوى التحليل للتحويل ألمـويجي المتقطـع واختيار العتبة المستخدمة وتشفير المعاملات. لقد تم مقارنة أداء (أنجازيه) هذه الطرق الثلاثة باسـتخدام عـاملي نسبة الضغط ونسبة أشارة إلى ضوضاء.

1- Introduction

There is an important need for speech signal compression either bandwidth-limited channel in areas such as internet, in wireless communications, in digital mobile phone technology where many users share the same frequency bandwidth or employ smaller amounts of memory to store signals such as in robot, in answering machine and other applications. For these applications, one wants to produce the smallest signal possible while accept a minimal loss in audible speech quality. In many real world applications, the signals are non-stationary such as speech signal. One solution for processing non-stationary signals is the wavelet transform. The reason of wavelet transform has gained a lot of popularity in the field of signal processing, is due to its capability of providing both time and frequency information simultaneously, hence giving a time-frequency representation of the signal.

Recent works that were suggested to solve the problem of speech compression had used different methods with different efficiencies. In 2000, C. Etemoglu [1] introduced a general approach to sinusoidal modeling of speech, where in a closed-loop Analysis-by-Synthesis technique sequentially extracts the parameters for each sinusoidal component. In 2001, W. Pereira [2] presented a system to improve the speech quality of algorithms that employ Linear Predictive Coding (LPC). In 2001, M. Nafornita [3] presented a new speech compression algorithm using orthogonal transform the Discrete Wavelet Packets Transform (DWPT). In 2002, G. Sin [4] investigated two compression algorithms which are second order statistics method and third order statistics method. The basic process is involving compression using autoregressive model. In 2004, K. Berglund [5] presented a speech compression and tone detection in a real–time system using a compression technique Adaptive Differential Pulse Code Modulation (ADPCM) for speech compression and tone detection. In 2004, S. Vakil [6] suggested coding scheme which works in a perceptual auditory domain. The input high dimensional frames of a speech are transformed to power spectral domain, using Discrete Fourier Transformer (DFT).

In this paper the implementations of a compression/decompression model using new wavelet techniques is proposed in order to use it on-line in the Internet applications or in the wireless communication. To achieve this result, three compression algorithms are suggested. These algorithms are wavelet analysis–synthesis method, wavelet global threshold method and wavelet by–level threshold method. The performances of the above three compression algorithms are comparing using compression ratio and signal-to-noise ratio factors.

2. Discrete Wavelet Transform (DWT)

The DWT which is based on sub-band coding is a special case of the WT that provides a compact representation of a signal in time and frequency that can be computed efficiently. It is easy to implement and reduces the computation time and resources required. The DWT is computed by successive high pass and low pass filtering of the discrete time-domain signal and is defined by the following equations [7]:

$$cD = \sum_n x[n]\, H[2k-n] \qquad \text{... (1)}$$

$$cA = \sum_n x[n]\, G[2k-n] \qquad \text{... (2)}$$

Where $x(n)$ is the input signal, cD and cA are the outputs of the high pass (H) and low pass (G) filters, respectively after sub sampling by 2, this is called the Mallat algorithm or Mallat-tree decomposition. Figure (1) explains this operation. The low pass filter is denoted by G_0 while the high pass filters is denoted by H_0. The high pass filter produces detail information; $cD(n)$, while the low pass filter associated with scaling function produces approximations, *cA* (*n*). The approximations are the high-scale, low-frequency components of the signal. The details are the low-scale, high-frequency components.

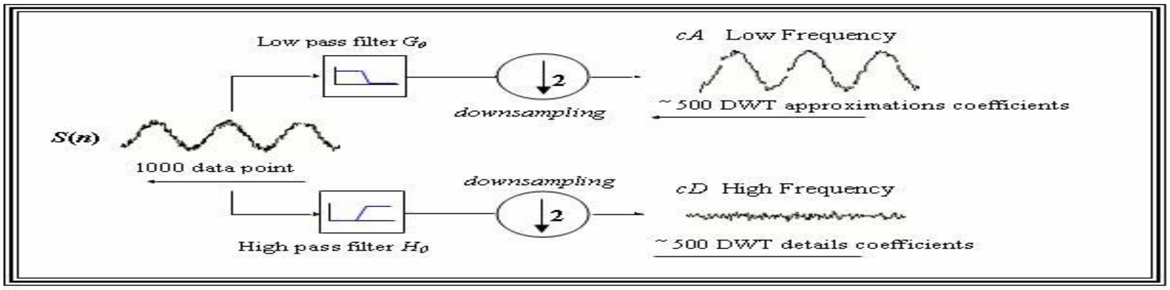

Figure (1): One-stage filtering operation of DWT.

The decomposition or analyze process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. This is called the wavelet decomposition tree as shown in Figure (2).

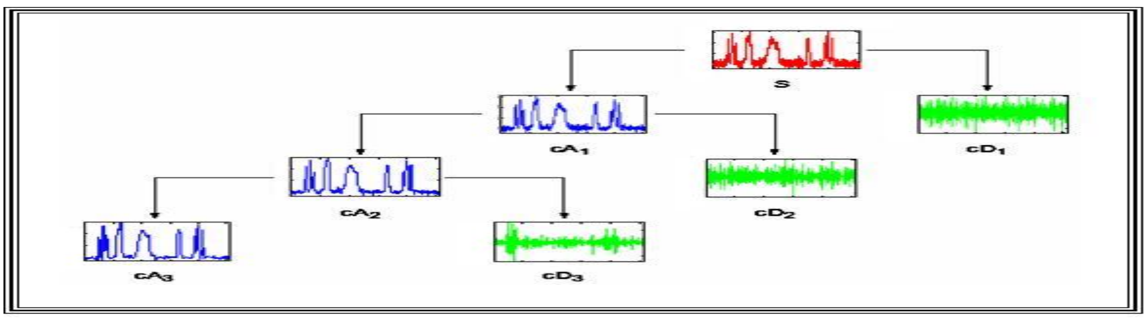

Figure (2): Three levels decomposition of sample signal S.

The original signal can be reconstructed or synthesized using the inverse discrete wavelet transform (IDWT). The synthesis starts with the approximation and detail coefficients *cAj* and *cDj*, and then reconstructs *cAj*-1 by up sampling and filtering with low pass and high pass reconstruction (synthesis) filters and then added. This process is continued through the same number of levels as in the decomposition process to obtain of the original signal. The Mallat algorithm works equally well if the analysis filters, G_0 and H_0, are exchanged with the synthesis filters. The reconstruction filters (*Lo_R* and *Hi_R*) together with the low and high pass reconstruction filters, forms a system known as quadrate mirror filters (QMF). For a multilevel analysis, the reconstruction process can itself be iterated producing successive approximations at finer resolutions and finally synthesizing the original signal as shown in Figure (3).

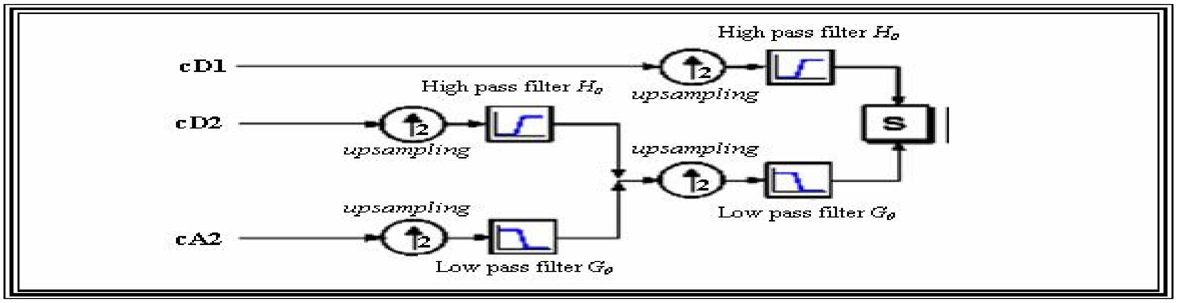

Figure (3): Two levels reconstruction of sample signal S.

There are a number of basis functions that can be used as the mother wavelet for wavelet transformation. 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. Figure (4) shows some of the commonly used wavelet functions.

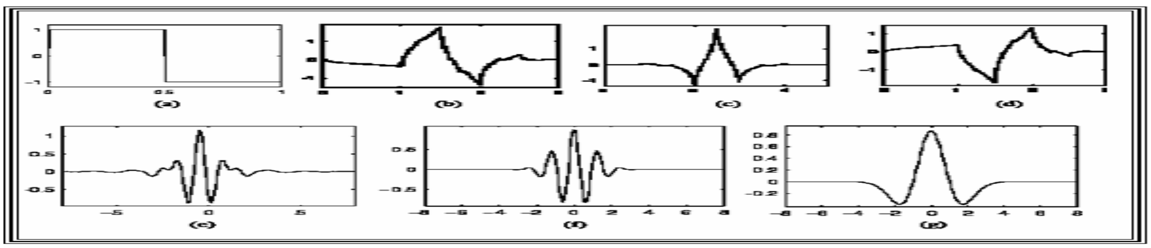

Figure (4): Wavelet families (a) Haar (b) Daubechies4 (c) Coiflet1 (d) Symlet3 (e) Meyer (f) Morlet (g) Mexican Hat.

3. <u>Speech Compression Model</u>

The general model for speech compression is presented in Figure (5), which consists of preprocessing, coder represented by compression program, storage device and decoder represented by decompression program. The coder and decoder work in similar manner but in inverse way.

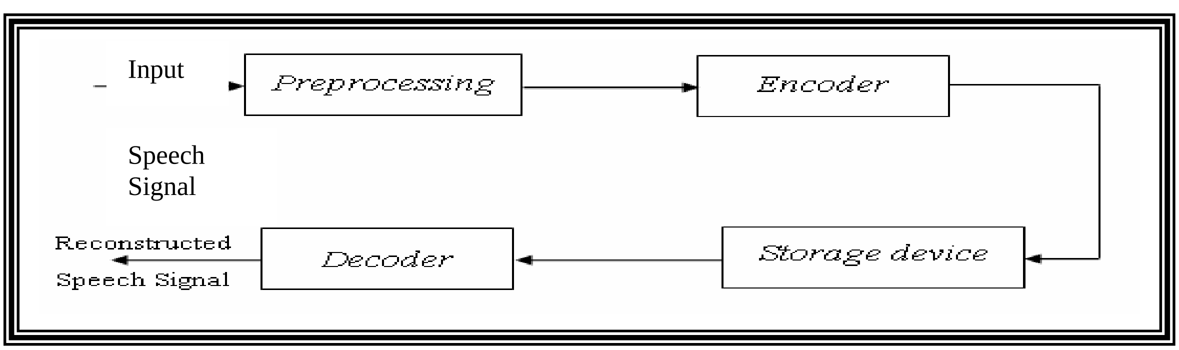

Figure (5): Speech compression model.

The speech preprocessing includes many operations; they are sampling, quantization, filtration and normalization.

- **Sampling and Quantization**

Since the speech file is import to handle by Matlab, the sampling and quantization stage does not need to be implemented; that is, the sampled and quantized speech signals are directly read from files. Where speech signal is processed in PCM (Pulse Code Modulation) format, and it sampled at 8000 Hz and quantized at 8 bit.

- **Filtration**

The speech signal may contain noise because it is recorded in different situations, so to increase the performance of the compression system, the noise must be reduced as possible. Infinite Impulse Response (IIR) band pass digital filter is used to perform this task. The general form of the digital system function is [8]:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{r=0}^{M} b_r z^{-r}}{1 + \sum_{j=1}^{N} a_j z^{-j}} \quad \ldots (3)$$

Where,

H(z): the system function of the digital filter,

X(z) and Y(z): the Z-transform of the input and output signals respectively, and r, j, M, N: are

integers with ($M \le N$); N is the order of the filter, and

a, b: the coefficients of filter.

From equation (3), a band pass IIR digital filter of six orders is used to reduce the noise which is distributed in speech signal. A band pass between 100 and 4000 Hz is chosen [7]. From Matlab the parameter b_r and a_j are founded by using Butterworth function. These parameters are as follow:

b_0 = 0.8545; b_1 = 0.0000; b_2 = -2.5635; b_3 = 0.0000; b_4 = 2.5635; b_5 = 0.0000; b_6 = -0.8545.

a_0 = 1.0000; a_1 = 5.5511; a_2 = -2.6862; a_3 = -1.1102; a_4 = 2.4197; a_5= 6.6613; a_6 = -0.7301.

The filter order is chosen equal to six because many different order values (3 –12) are taken and are tested. Then it is found that order six is the best one.

. **Normalization**

Normalization is a process of scaling the numbers in a data set to improve the accuracy of the subsequent numeric computations. A way to normalize signal is to center it at zero mean (average) and scale it to unit standard deviation according to the following equation [9]:

$$Y = \frac{\sum_{i=1}^{n} x_i - a}{s} \quad \ldots (4)$$

Where,

Y: is the output signal,

x_i: is the input signal,

a: is average of signal,

n: is the number of elements in the signal, and

s: unit standard deviation defines by the following equation.

$$s = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - x')^2 \right)^{\frac{1}{2}} \qquad \dots (5)$$

Where, $$x' = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \dots (6)$$

The importance of normalization step is appear from the necessary of put the speech signal wanted to be compress in the same level of intensities, since the incoming speech signal usually have different intensities due the speaker loudness or speaker distance from microphone.

4. Implementation of DWT

Speech compression with wavelet–based techniques is a relatively new field, and stills many suggestions to speech coder schemes are presented in the related press. The maximization of the compression ratio can be done, if a good selection of the parameters of that used in DWT. The parameters must be selected when the DWT is computed, taking into account the fact that the signal to be processed is a speech signal. These parameters are:

a) Selecting the best mother wavelet:

Because of the variant and numerousness of mother wavelets, for that the choice of the mother-wavelet function used in designing high quality speech coders is of the prime importance. The compression ratio and SNR are different for same speech signal when using different mother wavelet. Several different criteria can be used in selecting a mother-wavelet function. Several experiments are performed with various speech files, with many wavelets are given below and the result are listed in the Table (1). From Table (1) it is found that, the db15 gives better result than other wavelets.

Table (1): DWT analyzes using different mother wavelets.

Wavelet name	SNR	Compression ratio %
Haar	18.4354	73.3160
Symlet	26.1318	72.8299
Coiflet	26.6507	72.6215
db3	29.8455	73.1076
Biorthogonal	30.2505	72.9861
db15	33.5673	72.9514

b) Numbers of wavelet analysis-decomposition levels

Signal compression is based on the concept that selecting a small number of approximation coefficients (at a suitably chosen level) and some of the detail coefficients can accurately represent regular signal components. Choosing a decomposition level for the DWT usually depends on the type of signal being analyzed. Numbers of wavelet decomposition levels play important role in determine the compression ratio. From many testing, it's found that the compression of speech signals decomposition up to level 3 gives good result, with no further advantage gained in compression ratio beyond level 3 as shown in Table (2).

Table (2): DWT analyzes using decomposition levels.

Decomposition level	SNR	Compression ratio %
1	53.8604	19.3142
2	41.0017	51.5885
3	33.5673	72.9514
4	23.6961	85.4253
5	11.6062	91.8056
6	5.1847	95.0347

c) Choose Threshold

Threshold is the main method to make a resulting of wavelet coefficients most suitable for compression and wavelet transform are threshold such that the error due to threshold is inaudible to our ears. There are two kinds of threshold that are hard threshold and soft threshold. Equations 8 and 9 present hard and soft threshold signals.

- Hard Threshold $$T_H(x_i) = \begin{cases} 0 & |x_i| \le T \\ x_i & |x_i| > T \end{cases} \qquad \ldots (7)$$
- Soft Threshold $$T_S(x_i) = \begin{cases} 0 & |x_i| \le T \\ \operatorname{sgn}(x_i)(|x_i| - T) & |x_i| > T \end{cases} \qquad \ldots (8)$$

The results of many experiments are performed on methods of speech compression using wavelet are declared that the hard threshold gives better SNR values than soft threshold and the compression ratio values are the same for both threshold as shown in Table (3).

Table (3): DWT analyzes using soft and hard thresholds.

File Name	Threshold	SNR	Compression ratio %
Male 1	hard	33.5673	72.9514
	soft	24.2243	72.9514
Male 2	hard	25.1368	73.4174
	soft	16.2232	73.4174
Female 1	hard	24.1717	72.8798
	Soft	12.4175	72.8798
Female 2	hard	24.726	73.5657
	Soft	14.2167	73.5657

d) Truncation of Coefficients

After calculating the wavelet transform of the speech signal, compression involves truncating wavelet coefficients below a threshold. An experiment behavior on a male spoken sentence shows that most of the coefficients have small magnitudes. More than 90% of the wavelet coefficients have less than 5% of the maximum value. This means that most of the speech energy is in the high-valued coefficients, which are few. Thus the small valued coefficients can be truncated or zeroed and then be used to reconstruct the signal.

e) Coding Coefficients

Signal compression is achieved by first truncating small-valued coefficients and then efficiently encoding them. Coding approach to compression is to encode consecutive zero valued coefficients, with two bytes. One byte to indicate a sequence of zeros in the wavelet transforms vector and the second byte representing the number of consecutive zeros. For further data compaction a suitable bit encoding format, can be used to quantize and transmit the data at low bit rates.

The whole implementation software system of the wavelet technique is shown in Figure (6).

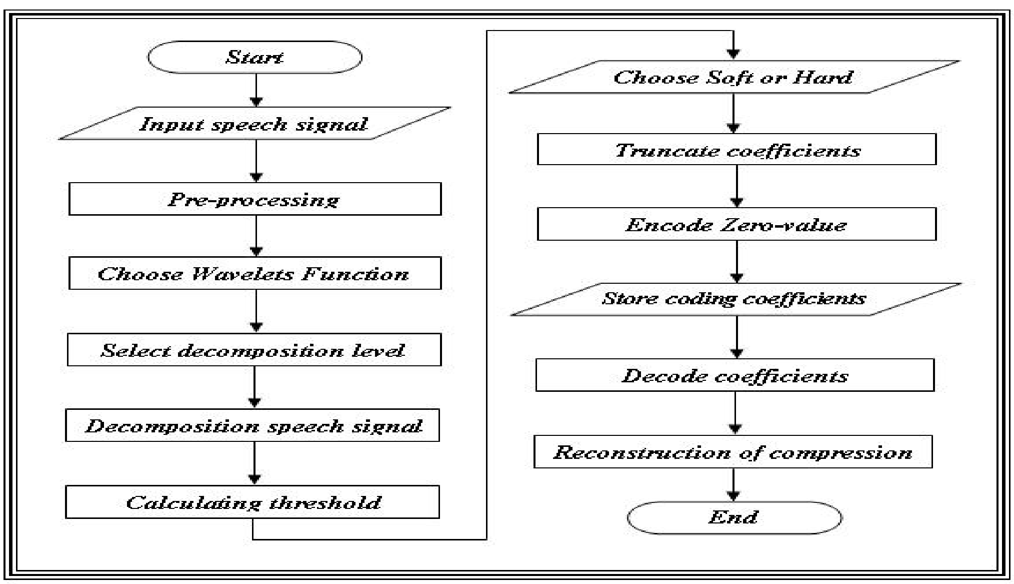

Figure (6): Flowchart of the wavelet based speech coder.

5. Wavelet Analysis–Synthesis Method

The frequencies that are most prominent in the original signal will appear as high amplitudes (low frequency) in that region of the DWT signal (approximation coefficient) which includes those particular frequencies. The frequency bands that are not very prominent in the original signal will have very low amplitudes (high frequency) and that part of the DWT signal (detail coefficient) can be discarded without any major loss of information, allowing data compression. This idea is used in the wavelet analysis-synthesis method for compression of speech signals. After the speech signal, has been filtered and normalized from the filtration and normalization stage, this speech signal will be sent to the DWT to process it with this compression scheme.

The results of computing wavelet analysis-synthesis methods on 10 speech signals are presented in Table (4). The quality of reconstructed speech signal is very good and similar to the original speech signal. It is clearly seen that about half of signal is used and is lead to value of compression ratio being nearly 50 %.

Table (4): The results are of the wavelet analysis-synthesis method.

Original file size in byte	File size after compression in byte	SNR (db)	Compression ratio %
11943	5986	29.7456	49.8786
50408	25218	27.5760	49.9722
130560	65294	29.7119	49.9893
172800	86414	19.2904	49.9919
185620	92824	28.1298	49.9925
253440	126734	29.6023	49.9945
272661	136345	26.4564	49.9947
434787	217408	26.2727	49.9967
3708	1868	29.4526	49.6224
4114	2071	28.4927	49.6597

Figure (7) shows the waveforms of original signal and the reconstructed speech signal using wavelet analysis–synthesis method.

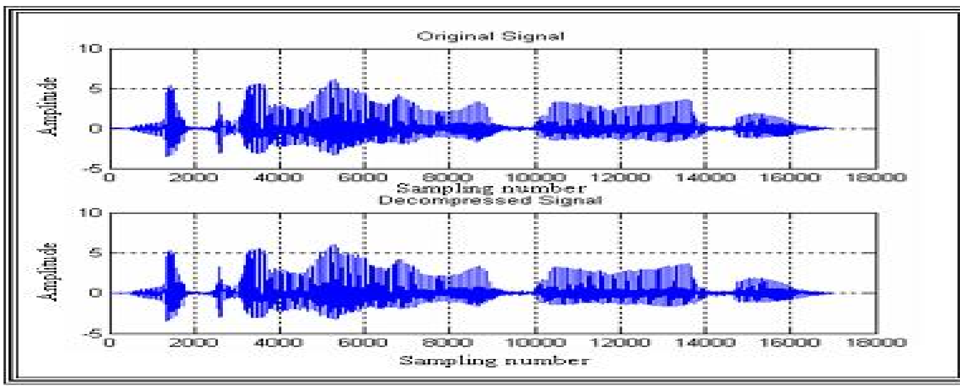

Figure (7): The original and reconstructed signals using wavelet analysis–synthesis method.

6. Wavelet Global Threshold Method

The aim of Global Threshold is to retain the largest absolute value coefficients, not considering of the numbers of the wavelet decomposition levels. Global thresholds are calculated by computing the largest value of detail coefficient at level one and then multiplied by *a* as shown in the next equation.

$$Thr_{glb} = a * max\ (cD1) \qquad \ldots (9)$$

Where,

Thr_{glb}: is the threshold value using global threshold methods,
a: is a compression factor and its value is $1 < a < 2$, and
$cD1$: is the detail vector at level 1.

After calculating the threshold *Thr,* the approximation coefficient keeps without any change and the threshold is applied to detail coefficient and use the same *Thr* for each level (i.e. there is one threshold value). If a certain detail coefficient lies below this threshold *Thr*, it is set to zero, if not; it is left by using hard threshold.

Another type is soft threshold, which, like all the previous ones, by setting all the elements below *Thr* to zero, but make all the others equal to sign(x)*x-Thr. The resulting coefficient vector after applying a threshold method with the threshold *Thr* contains many sequence of zeros value. After zeroing wavelet coefficients with negligible values based on calculating threshold values, the transform vector needs to be compressed. In this implementation consecutive, zero valued coefficients are encoded with two bytes. One byte is used to specify a starting siring of zeros and the second byte continue the number of successive zeros. Due to the scarcity of the wavelet representation of the speech signal, this encoding method leads to higher compression ratios than storing the non-zero coefficients along with their respective positions in the wavelet transform vector. This encoding scheme is the primary means of achieving signal compression.

In the global threshold method, the result of compressing speech signal is also good, while the reconstructed speech signal is very well and the value of SNR is high. The main problem is that the compression ratio is not stable (not fixed), where its value differ form speech signal to another as shown in Table (5). Figure (8) shows the original and reconstructed speech signals for this method.

Table (5): The results are of the wavelet global threshold method.

Original file size in byte	File size after compression in byte	SNR (db)	Compression ratio %
11943	3672	25.5015	69.2540
50408	16472	29.2905	67.3226
130560	54890	30.5903	57.9580
172800	56492	26.1017	67.3079
185620	57108	28.6374	69.2339
253440	135179	37.3742	46.6623
272661	103481	30.1032	62.0477
434787	168771	30.2826	61.1831
3708	1030	24.7812	72.2222
4114	1330	23.9165	67.6714

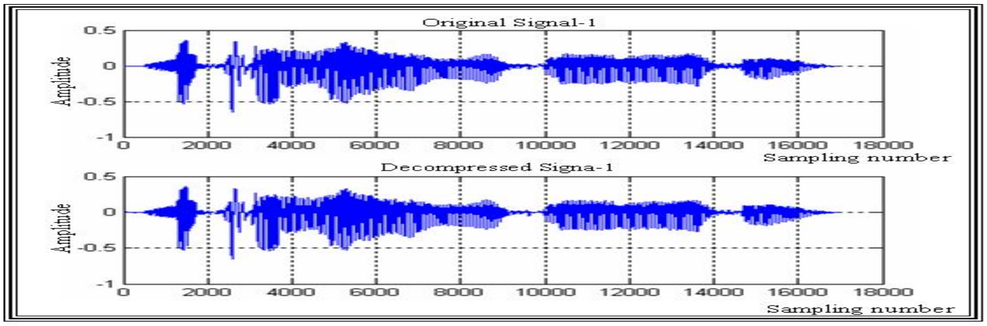

Figure (8): The original and reconstructed speech signals using wavelet global threshold method.

7. Wavelet By-Level Threshold Method

The by-level threshold technique is applied after the coefficients vector is obtained. Level dependent thresholds are calculated using the Birge-Massart strategy [7]. The approximation coefficients are kept at the level of decomposition J and computed the number of detail coefficients at level i. Starting from i to J according to the following formula:

$$n_j = \frac{M}{(J + 2 - j)^a} \quad \text{… (10)}$$

where,

n_j: is the er of largest detail coefficients are kept at level j,
J: number of the decomposition levels,
j: is the index of decomposition level, and
a: is a compression parameter and its value is $1 < a < 2$.

The value of M denotes how scarcely distributed the wavelet coefficients are in the transform vector. If L denotes the length of the coarsest approximation coefficients then M takes as shown below, depending on the signal being analyzed. Three different choices for M are proposed:
Scarce high: $M = L$, Scarce medium: $M = 1.5*L$, Scarce low: $M = 2*L$ .And the threshold value is given by the following formula.

$$Thr_{lvl} = \max\left(cD_j\right) * \frac{\log\left(n_j\right)}{\sqrt{n_j}} \quad \text{… (11)}$$

Where,

Thr_{lvl}: is the threshold value using By-level methods,
cD_J: is the detail coefficient at level j, and
n_j: is the number of detail coefficients to be kept at level j .

Thus this approach will, select the highest absolute valued coefficients at each level. The results of three different threshold values are found, one for each decomposition level. Then the threshold is applied on the detail coefficients using either hard threshold or soft threshold. Now the modify detail coefficient contains successive zeros. In order to compress the coefficient, the consecutive zero valued coefficients are encoded with two bytes as explained before. This method provide higher compression ratio with conservatism of reconstructed speech quality further that the SNR is good as shown in Table(6). The waveform results are shown in Figure (9).

Table (6): The results are of the wavelet by-level threshold method.

Original file size in byte	File size after compression in byte	SNR (db)	Compression ratio %
11943	3465	21.4364	70.9872
50408	13325	24.7260	73.5657
130560	34689	24.7606	73.4306
172800	44568	21.3841	74.2083
185620	49378	23.7961	73.3983
253440	67041	28.6548	73.5476
272661	72119	24.0770	73.5499
434787	114410	24.1135	73.6860
3708	1128	22.1048	69.5793
4114	1255	20.9063	69.4944

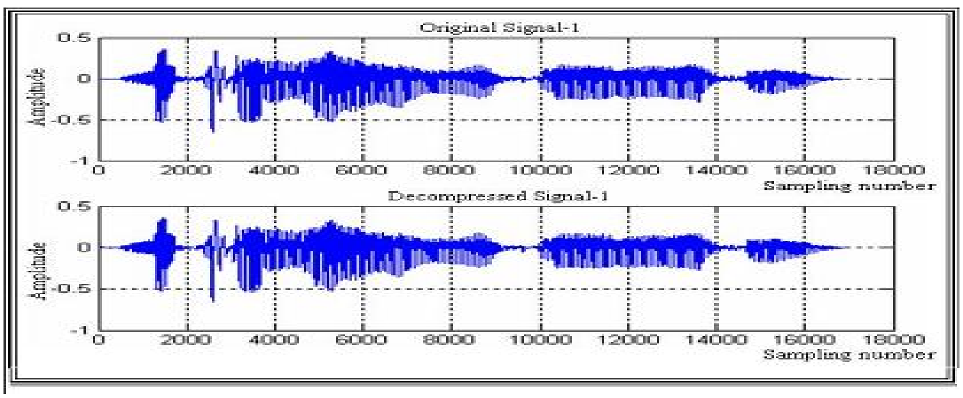

Figure (9): The original and reconstructed speech signals using wavelet by-level threshold method.

8. Conclusions

After implementing and testing the proposed speech compression methods the following points can be concluded from the results:

- Wavelet by-level threshold presents large compression ration with saving of sound quality where SNR value obtained is also large. There is a trade off between compression ratio and signal quality.
- The threshold technique, which is used with wavelet global threshold and wavelet by-level threshold methods, serves to increase the compression ratio.
- The more interested feature of the wavelet by-level threshold method is capability to obtain different values of compression ratio for the same signal, while most other compression techniques have fixed compression ratio. Since this method has several parameters which can be adjusted to obtain the desire compression ratio.
- The signal quality is more important to the listener. It is found that wavelet by-level method gave better quality in reconstructed speech signal with high compression

ratio. Although the quality of reconstructed signal in wavelet analysis-synthesis method is excellent (similar to original speech signal), but it is provide a middle value of compression ratio nearly 50 %. Wavelet global threshold method is similar to wavelet analysis-synthesis method. The reconstructed signal is good but the values of compression ratio are varied.

References

1- C. Etemoglu and V. Cuperman, "Speech Coding with an Analysis-By-Synthesis Sinusoidal Model", IEEE ICASSP'2000, Istanbul, Turkey, June 2000.
2- W. Pereira, "Modifying LPC Parameter Dynamics to Improve Speech Coder Efficiency", M. Sc. Thesis, Electrical and Computer Engineering Dept., University of Montreal, September 2001.
3- M. Nafornita, A. Isar, and D. Isar, "A New Speech Compression Method", Electrical Engineering Dept., University of Facta, Nis, France, IEEE, Vol. 17, pp. 391-404, December 2001.
4- G. Sin, "Speech Compression Using Higher Order Statistics", School of Information Technology and Electrical Engineering, University of Queensland, October 2002.
5- K. Berglund, "Speech Compression and Tone Detection in a Real – Time System", M. Sc. Thesis, Computer science and Elec. Engineering Dept., Lulea University of Technology, 2004.
6- S. Vakil, "Gaussian Mixture Model Based Coding of Speech and Audio", Electrical and Computer Engineering Dept., University of McGill, Canada, October 2004.
7- Yi Hu and P. C. Loizou, "Speech Enhancement Based on Wavelet Threshold the Multistage Spectrum", IEEE Transactions on Speech and Audio Processing, Vol. 12, No. 1, January 2004.
8- D. Joseph Krusienski, "Enhanced Structured Stochastic Global Optimization Algorithms for IIR and Nonlinear Adaptive Filtering", Ph. D. Thesis, Electrical Engineering Dept., University of the Pennsylvania State, August 2006.
9- X. Yuan, "Auditory Model-Based Bionic Wavelet Transform For Speech Enhancement", Electrical and Computer Engineering Dept, University of Marquette, May 2003.