IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 3, Ver. VI (May - June 2017), PP 33-41
www.iosrjournals.org

Speech Compression Using Wavelet Transform

Harshalata Petkar[1]
[1](*Centre for information and Language Engineering,Mahatma Gandhi International Hindi University,Wardha India)*

Abstract : *This paper applies wavelet analysis to speech compression. A mother or basis wavelet is first chosen for the compression. The signal is then decomposed to a set of scaled and translated versions of the mother wavelet. The resulting wavelet coefficients that are insignificant or close to zero are truncated achieving signal compression. Analysis of the compression process was performed by comparing the compressed-decompressed signal against the original. This was conducted to determine the effect of the choice of mother wavelet on the speech compression. The results however showed that regardless of bases wavelet used the compression ratio is relatively close to one another.*
Keywords: *Compression, Filter Bank, Lossless &Lossy Compression, Wavelet Transform, 1-D DWT.*

I. Introduction

Speech is a very basic way for humans to convey information to one another. With a bandwidth of only 4kHz, speech can convey information with the emotion of a human voice. People want to be able to hear someone's voice from anywhere in the world-as if the person would be in the same room. Speech can be defined such as the response of the vocal tract to one or more excitation signals. Compression of signals is based on removing the redundancy between neighboring samples and/or between the adjacent cycles. In data compression, it is desired to represent data by as small as possible number of coefficients within an acceptable loss of visual quality. Compression techniques can be classified into one of two main categories: lossless and lossy.

Compression methods can be classified into three functional categories:

1. Direct Methods: The samples of the signal are directly handled to provide compression.
2. Transformation Methods: such as Fourier Transform (FT), Wavelet Transform (WT), and Discrete Cosine Transform (DCT).
3. Parameter Extraction Methods: A preprocessor is employed to extract some features that are later used to reconstruct the signal.

Wavelet compression is a form of predictive compression where the amount of noise in the data set can be estimated relative to the predictive function. Most modern compression techniques use a two step process: First, a predictive compression function (such as wavelet transform) is applied. If the choice of the predictive compression function is good, the result will be a new set of data with smaller values and more repetition. Second, a coding compression step that will represent the data set in its minimal form (Huffman coding, run-length). The compression of speech signals has many practical applications. One example is in digital cellular technology where many users share the same frequency band-with. Compression allows more users to the system than otherwise possible.

II. Filter Bank

A filter bank is a set of filters, which split up the signal's frequency components into different signals, each with a subset of frequencies. The combined pass bands of the filter cover the entire frequency range, so the filters are complementary. A simple filter bank consists of one low pass filter and one high pass filter, both having a cut off frequency at half the frequency bandwidth. Applying this filter bank to signal results into two new signals, one with the lower half frequencies and one with the upper half frequencies. A block diagram of this filter bank is illustrated in Figure (1)..

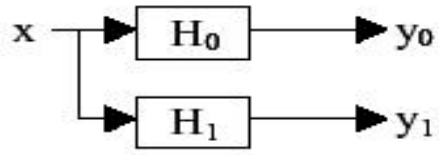

Fig. (1) Simple filter bank

Where x is input signal. H_0 and H_1 are low pass and high pass filters, respectively. And y_0, y_1 are output signals. To construct a filter bank with more than two frequency bands, y_0 could be filtered again by two filters, one low pass filter and again one high pass filter which divide the bands up again into two bands. The lengths of output signals (number of samples) have doubled. The solution is to downsample (or decimates). The *downsampling*operation, which is done in the analysis bank, shall save only the even-numbered components of the two outputs, where the odd-numbered components are removed as shown in Eq. (1) [4].

$$(\downarrow 2)\begin{bmatrix} . \\ v(-2) \\ v(-1) \\ v(0) \\ v(1) \\ v(2) \\ . \end{bmatrix} = \begin{bmatrix} . \\ v(-4) \\ v(-2) \\ v(0) \\ v(2) \\ v(4) \\ . \end{bmatrix} \qquad (1)$$

The downsampling operator is usually indicated by $\downarrow 2$. Decimating results in a signal with half the number of samples that represent the same time interval as the original signal. Thus, the sample rate is halved, too.

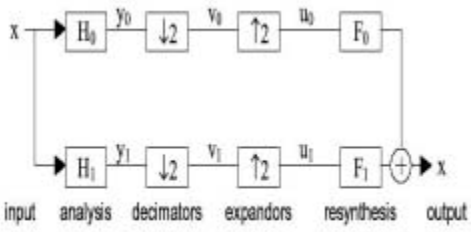

Fig. (2) Two-Channel Analysis/Re-Synthesis Filter Bank.

The decimated output can then be filtered again with the same filters to again split it up into lower and higher frequency contents. For reconstruction, *upsampling*(*expanding*) must be done in order to undo the decimation. Inserting a zero after each sample does this. Additionally, two re-synthesis filters F0 and F1 are needed to smooth out the zeros, reversing the analysis low pass and high pass filters. The resulting samples are obtained by adding the outputs of the re-synthesis filters. Figure (2) shows a two-channel filter bank analysis followed by re-synthesis. Applied to a half- length vector v, upsampling inserts zeros as in Eq (2), where $\uparrow 2$ indicates upsampling [4].

$$(\uparrow 2)\begin{bmatrix} . \\ v(0) \\ v(1) \\ v(2) \\ . \end{bmatrix} = \begin{bmatrix} . \\ v(0) \\ 0 \\ v(1) \\ 0 \\ v(2) \\ . \end{bmatrix} \qquad (2)$$

As discrete filters do not have an ideal cut off, the low pass and high pass filters' frequency responses overlap: the low pass lets through frequency components of the high pass band, conversely, the high pass filter lets through low frequencies -see Figure (3). This aspect, causes aliasing when downsampled. The solution for perfect reconstruction is to design the reconstruction filters F_0 and F_1 in such a way that they cancel out the aliasing of the analysis filters.

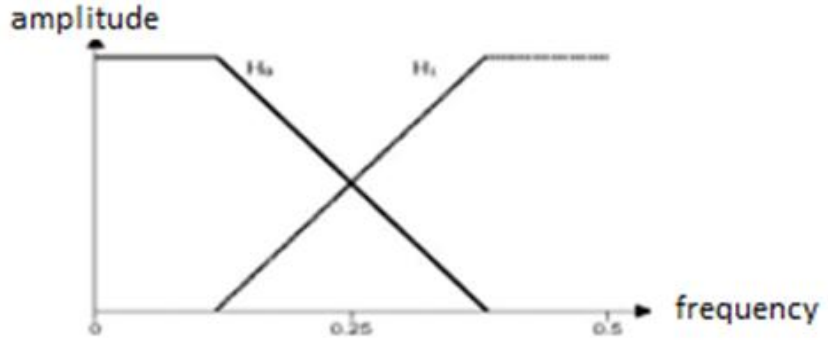

Fig (3) Overlapping Lowpass and Highpass Filter Responses

III. Wavelet Families

It is very important to briefly introduce wavelets' families, because they are the main tools. We can't say that one type of wavelet is better than the another, because every type has its own applications. So, a wavelet family may be good for one application, but not for another, this depends on the nature of the application. We will introduce, for each wavelet family, the scaling function **Φ(t)** in Eq. (3), the wavelet function **Ψ(t)** in Eq.(4) and its filters' values (for a two-channel filter back) in both of time-domain and frequency-domain representations.

$$\phi(t) = 2 \sum_{k=0}^{N} h_o(k)\phi(2t-k) \qquad (3)$$

$$\psi(t) = 2 \sum_{k=0}^{N} h_1(k)\phi(2t-k) \qquad (4)$$

IV. Examples Of Wavelet

The DWT is the most powerful and new signal compression technique which uses multi-resolution analysis for analyzing speech signals. Function of the DWT is a frequency scale adjustments and values shifting position

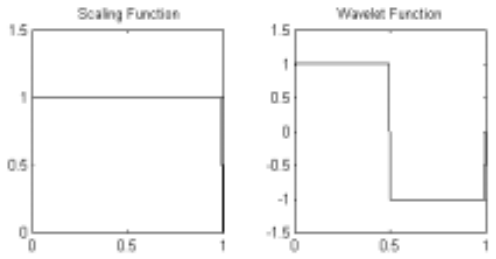

Fig (4.a) Haar wavelet & Scaling functions

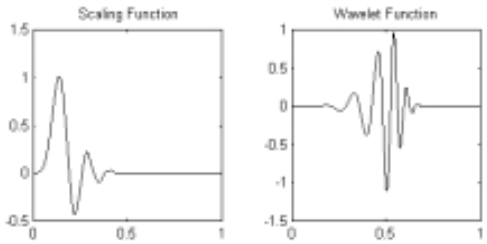

Fig (4.b) Daubechies (db2) Wavelet & scaling functions

The wavelet function needs to disappear towards - ∞ and ∞, and its integral is zero. Therefore, the wavelet needs to have at least one change of sign, making its shape looks like a small wave [5].The different families make tradeoff between how compactly the basis functions are localized in space and how smooth they are. Within each family of wavelets (such as the Daubechies family) are wavelet subclasses distinguished by the number of filter coefficients and the level of iteration. Wavelets are most often classified within a family by the number of *vanishing moments*. This is an extra set of mathematical relationships for the coefficients that must be satisfied. The extent of compactness of signals depends on the number of vanishing moments of the wavelet function used.

V. Discrete Wavelet Transform

The main idea is the same as is in the continuous wavelet transform (CWT). A time-scale representation of a digital signal is obtained using digital filtering techniques. The CWT was computed by changing the scale of the analysis window, shifting the window in time, multiplying by the signal, and integrating over all times. In the discrete case, filters of different cutoff frequencies are used to analyze the

signal at different scales [2]. The Discrete Wavelet Transform (DWT) is a special case of the WT that provides a compact representation of a signal in time and frequency that can be computed efficiently.

$$d_{jk} = \int x(t)\varphi_{jk}(t)dt = 2^{\frac{j}{2}}\int x(t)\varphi_{jk}(2^{j}t-k)dt$$

$$\varphi_{jk}(t) = 2^{\frac{j}{2}}\varphi_{jk}(2^{j}t-k), \quad j,k \in Z \qquad (5)$$

$$x(t) = \sum_{j}\sum_{k} d_{jk}\varphi_{jk}(t)$$

The DWT coefficient $d_{jk}$are defined by Eq (5). Where Ψ(t) is the mother wavelet and x(t) is the time signal. For many signals, the low-frequency content is the most important part. It gives the identity (*approximations*) of the signal. The high frequency on the other hand covered the *details* of the signal.

5.1 APPROXIMATIONS AND DETAILS

Consider the human voice. If the high frequency components are removed, the voice sounds become different, but we can still tell what is being said. However, if enough of the low-frequency components are removed, we hear gibberish. It is for this reason that, in wavelet analysis, we often speak of approximations and details [3].
Figure (5) shows the effect of the low and high frequencies in a signal.

5.2 VANISHING MOMENTS

"The number of vanishing moments of a wavelet indicates the smoothness of the wavelet function as well as the flatness of the frequency response of the wavelet filters (filters used to compute the DWT). In fact, the higher the number of vanishing moments, the faster the decay rate of wavelet coefficients [6].

$$\int_{R} t^{j}\varnothing(t)\,dt = 0 \qquad (6)$$

As in [3], typically a wavelet with (k+1) vanishing moments satisfies Eq (6) A higher number of vanishing moments lead to a faster decay rate of wavelet coefficients which will lead to a more compact signal representation and are hence useful in coding applications. However, in general, the length of the filters increases with the number of vanishing moments and the complexity of computing the DWT coefficients increases with the size of the wavelet filters.

VI. The Fast Wavelet Transform Algorithm

The Discrete Wavelet Transform (DWT) coefficients can be computed by using Malta's Fast Wavelet Transform algorithm. This algorithm is sometimes referred to as the two-*channel sub-band coder* and involves filtering the input signal based on the wavelet function used. The signal is passed through a series of highpass and lowpass filters to analyze the high frequencies and low frequencies respectively, followed by downsampling operation i.e. Filter Bank [3].

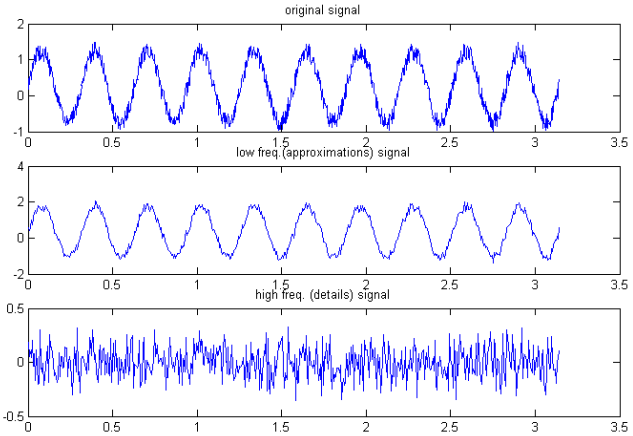

Fig (5) Approximations and Details of A Signal

Eqs (7) & (8) can express the two filtering and downsampling operations.

$$y_{high} = \sum_{n} x[n].g[2k - n] \qquad (7)$$

$$y_{low} = \sum_{n} x[n].h[2k - n] \qquad (8)$$

The low pass filter output is then filtered once again for further decomposition. The DWT of the original signal is then obtained by concatenating all coefficients starting from the last level of decomposition. The DWT will then have the same number of coefficients as the original signal. The decomposition 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. Figure (6) shows the wavelet decomposition to level 3 of a sampled signal.

Since the analysis process is iterative, in theory it can be continued indefinitely. In reality, the decomposition can only proceed until the vector consists of a single sample. Normally, however there is little or no advantage gained in decomposing a signal beyond a certain level. The selection of the optimal decomposition level in the hierarchy depends on the nature of the signal being analyzed or some other suitable criterion, such as the low-pass filter cut-off. The reconstruction in this can be obtained by following the above procedure in reverse order. The signals at every level are upsampled by two, passed through the synthesis filters g'[n], and h'[n] (highpass and lowpass, respectively), and then added. The interesting point here is that the analysis and synthesis filters are identical to each other, except for a time reversal. Therefore, the reconstruction formula becomes (for each level) as Eq (9)

$$x[n] = \sum_{k=-\infty}^{\infty} \left(\begin{pmatrix} y_{high}[k].g[-n + 2k] \\ + \left(y_{low}[k].h[-n + 2k]\right) \end{pmatrix} \right) \qquad (9)$$

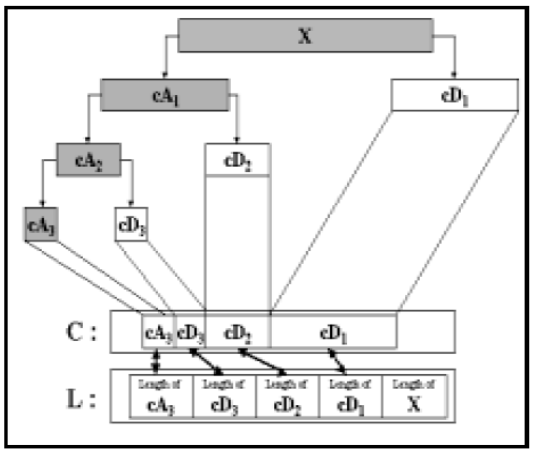

Fig (6) Level 3 Decomposition of a Sampled Signal{2}

VII. Wavelet Speech Compression Techniques

The idea behind signal compression using wavelets is primarily linked to the relative scarceness of the wavelet domain representation for the signal. Wavelets concentrate speech information (energy and perception) into a few neighboring coefficients [7]. Therefore as a result of taking the wavelet transform of a signal, many coefficients will either be zeros or have negligible magnitudes. Data compression is then achieved by treating small valued coefficients as insignificant data and thus discarding them. The process of compressing a speech signal using wavelets involves a number of different stages, each of which are discussed below.

7.1 Choice of Wavelet

The choice of the mother-wavelet function used in designing high quality speech coders is of prime importance. Choosing a wavelet that has compact support in both time and frequency in addition to a significant number of vanishing moments is essential for an optimum wavelet speech compressor. Several different criteria can be used in selecting an optimal wavelet function. The objective is to minimize reconstructed error variance and maximize signal to noise ratio (SNR). In general optimum wavelets can be selected based on the energy conservation properties in the approximation part of the wavelet coefficients. In [8] it was shown that the Battle-Lemarie wavelet concentrates more than 97.5% of the signal energy in the approximation part of the

coefficients. This is followed very closely by the Daubechies D20, D12, D10 or D8 wavelets, all concentrating more than 96% of the signal energy in the Level 1 approximation coefficients. Wavelets with more vanishing moments provide better reconstruction quality, as they introduce less distortion into the processed speech and concentrate more signal energy in a few neighboring coefficients. However the computational complexity of the DWT increases with the number of vanishing moments and hence, for real time applications it is not practical to use wavelets with an arbitrarily high number of vanishing moments [6].

7.2 Wavelet Decomposition

Wavelets work by decomposing a signal into different resolutions or frequency bands, and choosing the wavelet function and computing the Discrete Wavelet Transform (DWT) carries out this task [9]. 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 or some other suitable criterion such as entropy. For the processing of speech signals decomposition up to scale 5 is adequate [8]**,** with no further advantage gained in processing beyond scale 5.

7.3 Truncation Of Coefficients

After calculating the wavelet transform of the speech signal, compression involves truncating wavelet coefficients below a threshold. An experiment conducted on a male spoken sentence [7]**,** 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 [7]. Thus the small valued coefficients can be truncated or zeroed and then be used to reconstruct the signal. This compression scheme provided a segmental signal-to-noise ratio (SEGSNR) of 20 dB, with only 10% of the coefficients. Two different approaches are available for calculating thresholds. The first, known as Global Thresholding involves taking the wavelet expansion of the signal and keeping the largest absolute value coefficients. In this case you can manually set a global threshold, a compression performance or a relative square norm recovery performance. Thus only a single parameter needs to be selected. The second approach known as By Level Thresholding consists of applying visually determined level dependent thresholds to each decomposition level in the wavelet transform.

7.4 Encoding Coefficients

Signal compression is achieved by first truncating small-valued coefficients and then efficiently encoding them. One way of representing the high-magnitude coefficients is to store the coefficients along with their respective positions in the wavelet transform vector [9]. Another approach to compression is to encode consecutive zero valued coefficient [7]**,** 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 compactness a suitable bit-encoding format, can be used to quantize and transmit the data at low bit rates. A low bit rate representation can be achieved by using an entropy coder like Huffman coding or arithmetic coding.

VIII. Software Implementation

The design of the wavelet transform speech coder is based on the concepts covered in the "Wavelet Speech Compression Techniques". Figure (7) below illustrates the different processes involved in coding speech signals using wavelets. The MATLAB's Wavelet Toolbox incorporates many different wavelet families, from the "Wavelet Speech Compression Techniques Section (8.1)". It was decided to use the Haar and Daubechies wavelets for coding speech signals. We introduce the wavelet-based speech coder, as some functions (i.e. *Compress, Decompress, Encode, Decode, Pefcal, Playingsound, Results* and *Plotresult*) which are mainly called from the function *Main.*

8.1 Calculating Thresholds

For the truncation of small-valued transform coefficients, two different thresholding techniques are used, Global Thresholding and By-Level Thresholding.

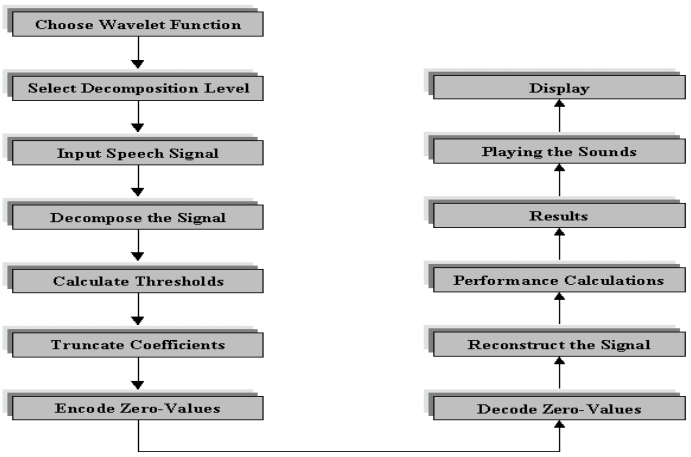

Fig (7) Design Flow of Wavelet Based Speech Coder

The aim of Global Thresholding is to retain the largest absolute value coefficients, regardless of the scale in the wavelet decomposition tree. Global thresholds are calculated by setting the % of coefficients to be truncated.Level dependent thresholds are calculated using the Birge-Massart strategy [3]. This thresholding scheme is based on an approximation result from Birge and Massart and is well suited for signal compression.

8.2 Encoding Zero-Valued Coefficients

After zeroing wavelet coefficients with negligible values based on either calculating threshold values or simply selecting a truncation percentage, 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 string of zeros and the second byte keeps track of the number of successive zeros. Due to the scarcity of the wavelet representation of the speech signal, this encoding method leads to a higher compression ratios than storing the non-zero coefficients along with their respective positions in the wavelet transform vector, as suggested in the "Wavelet Speech Compression Techniques (Section 8.4)". This encoding scheme is the primary means of achieving signal compression.

IX. Performance Measurement Tools

A number of quantitative parameters can be used to evaluate the performance of the coder, in terms of reconstructed signal quality after compression scores. The following parameters are compared:

1. Signal to Noise Ratio (SNR)
2. Peak Signal to Noise Ratio (PSNR)
3. Normalized Root Mean Square Error (NRMSE)
4. Retained Signal Energy(RSE)
5. Compression Ratios (CR)

The results obtained for the above quantities are calculated using the following formulas:

SIGNAL TO NOISE RATIO (SNR)

$$SNR = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \qquad (10)$$

$\sigma^2 x$is the mean square of the speech signal and $\sigma^2 e$ is the mean square difference between the original and reconstructed signals.

PEAK SIGNAL TO NOISE RATIO (PSNR)

$$PNSR = 10\log_{10}\frac{NX^2}{\|x-r\|^2} \qquad (11)$$

N is the length of the reconstructed signal, X is the maximum absolute square value of the signal x and $\| x-r\|^2$ is the energy of the difference between the original and reconstructed signals.

NORMALIZED ROOT MEAN SQUARE ERROR (NRMSE)

$$NRMSE = \sqrt{\frac{(x(n) - r(n))^2}{(x(n) - \mu_x(n))^2}} \qquad (12)$$

x(n) is the speech signal, *r(n)* is the reconstructed signal, and $\mu_x(n)$ is the mean of the speech signal.

Table (1) A Male Speech Signal Decomposed at Different Levels

Scale	CR	Zeros(%)	RSE (%)	SNR	PSNR	NRMSE
1	1.6263	44.5647	99.9799	36.9691	52.0111	0.015357
2	2.6527	67.8053	99.9298	31.5338	46.5733	0.028721
3	3.5387	76.7924	99.8626	28.6213	43.6615	0.040159
4	4.1757	80.9681	99.7852	26.6790	41.7224	0.050205
5	4.6689	83.3044	99.7011	25.2453	40.2903	0.059204

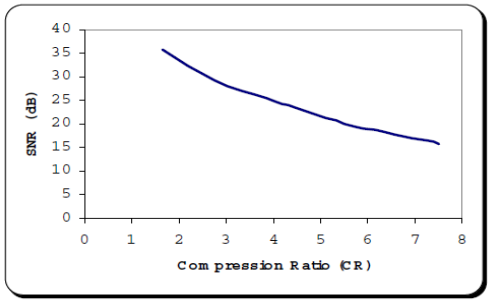

Fig (8) Compression Performance of Speech Signal vs. SNR

RETAINED SIGNAL ENERGY (RSE)

$$RSE(\%) = \frac{\| x(n) \|^2}{\| r(n) \|^2} * 100 \qquad (13)$$

||x(n)|| is the norm of the original signal and ||r(n)|| is the norm of the reconstructed signal.

COMPRESSION RATIO (CR)

$$CR = \frac{Length(x(n))}{Length(r(n))} \qquad (14)$$

x(n) is the original signal and r(n) is the reconstructed signal.

X. Result

As shown in Table (1), a speech file, in spoken English language spoken, is recorded by male with size of 71.7 Kbytes is tested with DWT coder using "Db20" wavelet. The recorded words are as follow:
"Providence is always on the side of the strongest battalions" There are many factors affect the wavelet-based speech coders' performance, mainly what compression ratio could be achieved at suitable SNR value with low value of NRMSE. To improve the compression ratio of wavelet based coder, we have to consider that it is highly speaker dependent and varies with his age and gender. The speed at which the speaker speaks is another factor, which affects the compression ratio. That is low speaking speed cause high compression ratio with high value of SNR. Increasing the scale value in wavelet-based speech coder gives higher compression ratios, but this cost decreasing in quality of the speech signal, which can be cleared by Figure (8) which show the SNR variation of relative to compression ratios for Tables (1).

XI. Conclusion

Data compression is the technology of representing information with lowest number of bits (minimum size). This technology is needed in the field of speech to satisfy transfer requirements of huge speech signals via communication companies and Internet, decreasing storage equipment is another need. The limited number of channels available to the Palestinian cellular company, Jawwal, and the high demand for mobile telephone services, put on a lot of pressure on the company to find a way to solve this problem. This paper proposed a method of using wavelet compression to give more room for more users to access Jawwal networks. A simple lossy compression algorithm for one-dimensional signals (as speech signal) based on wavelet transform coding is developed. It compacts as much of the signal energy into as few coefficients as possible. These

coefficients are preserved and the other coefficients are discarded with little loss in signal quality. Performance of the wavelet coder is tested by male speech signal. Results illustrate that wavelet-based coder achieved high compression ratio and quality.

References

[1]. Graps, A. "An Introduction to Wavelets", 1997. http://www.amara.com/current/wavelet.html
[2]. Polikar, R. "The Wavelet Tutorial" 1996. http://engineering.rowan.edu/~polikar/WAVELETS/WTpart4.html
[3]. Misiti, M., Misiti, Y., Oppenheim G. and Poggi, J. "Wavelet Toolbox User's Guide", Mathworks, 1997.
[4]. Strang, G. and Nguyen, T. "Wavelets and Filter Banks", Weslley-Cambridge Press, USA, 1996.
[5]. Bomers, F. "Wavelets in real time digital audio processing", 2000 http://www.daimi.au.dk/~fungus/DSP/Litteratur/Wavelets
[6]. Viswanathan, V., Anderson, W., Rowlands, J., Ali, M. and Tewfik, A. "Real-Time Implementation of a Wavelet-Based Audio Coder on the T1 TMS320C31 DSP Chip", 5th International Conference on Signal Processing Applications &Technology (ICSPAT), Dallas, TX, Oct.1994. http://citeseer.nj.nec.com/rd/28405181
[7]. Kinsner, W. and Langi, A. "Speech and Image Signal Compression with Wavelets", IEEE Wescanex Conference Proceedings, IEEE, New York, NY, 1993, pp. 368-375.
[8]. Agbinya, J.I. "Discrete Wavelet Transform Techniques in Speech Processing", IEEE Tencon Digital Signal Processing Applications Proceedings, IEEE, New York, NY, 1996, pp 514-519.
[9]. Fgee, E.B., Phillips, W.J. and Robertson, W."Comparing Audio Compression using Wavelets with other Audio Compression Schemes", IEEE Canadian Conference on Electrical and Computer Engineering, IEEE, Edmonton, Canada, 1999, pp. 698-701.