Wavelet Based Speech Compression for VOIP Applications

***Noman Ahmed, ** A. Q. K. Rajput, *** Naveed Ahmed Junejo**
*Postgraduate Student, Institute of Information Technology, Mehran University of Engineering & Technology, Jamshoro.Sindh Pakistan
**Professor, Institute of Information Technology, Mehran University of Engineering & Technology, Jamshoro. Sindh Pakistan
***Systems Analyst, Fauji Fertilizer Company Ltd. Goth Machhi. Sadiqabad, Pakistan
junejonauman@hotmail.com, njunejo@hotmail.com
Tel # +92-221-783833, Fax # +92-221-771382

Abstract

With the growth of the multimedia technology over the past decades, the demand for digital information increases dramatically. This enormous demand poses difficulties for the current technology to handle. One approach to overcome this problem is to compress the information by removing the redundancies present in it. Speech compression is the technology of converting human speech into an efficiently encoded representation that can later be decoded to produce a close approximation of the original signal. Recently, compression techniques using Wavelet Transform(WT) have received great attention, because of their promising compression ratio, Signal to Noise ratio (SNR),and flexibility in representing speech signals. This paper explores the major issues concerning the compression of speech using Wavelet based speech coder and the choice of an optimal wavelet for speech signals, decomposition level in the Discrete Wavelet Transform (DWT), threshold criteria for coefficient truncation and efficient encoding of truncated coefficients. In this paper we present the results obtained by compressing speech signals using DWT Techniques. More specifically we see the speech signal and approximations at different scales, analyze the performance measure which includes signal to noise ratio (SNR), peak signal to noise ratio(PSNR), normalized root mean square error(NRMSE) and retained signal energy.

Keywords: Wavelet Transform, SNR, PSNR, NRMSE, Haar, Daubechies.

1. 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 voice from anywhere in the world.

As a result a greater emphasis is being placed on the design of new and efficient speech coders for voice communication and transmission. Today applications of speech coding and compression have become very numerous. Many applications involve the real time coding of speech signals, for use in mobile satellite communications, cellular telephony, and audio for videophones or video teleconferencing systems [1].

This paper looks at a new technique for analyzing and compressing speech signals using wavelets. Very simply wavelets are mathematical functions of finite duration with an average value of zero that are useful in representing data or other functions. Any signal can be represented by a set of scaled and translated versions of a basic function called the mother wavelet. This set of wavelet functions forms the wavelet coefficients at different scales and positions and results from taking the wavelet transform of the original signal[2].

The coefficients represent the signal in the wavelet domain and all data operations can be performed using just the corresponding wavelet coefficients. Speech is a non-stationary random process due to the time varying nature of the human speech production system. Non-stationary signals are characterized by numerous transitory drifts, trends and abrupt changes. The localization feature of wavelets, along with its time-frequency resolution properties makes them well suited for coding speech signals. In designing a wavelet based speech coder, the major issues explored in this paper are:

i. Choosing optimal wavelets for speech,
ii. Decomposition level in wavelet transforms,
iii. Threshold criteria for the truncation of coefficients,
iv. Efficiently representing zero valued coefficients and
v. Quantizing and digitally encoding the coefficients.

The performance of the wavelet compression scheme in coding speech signals and the quality of the reconstructed signals is also evaluated.

The wavelet analysis procedure is to adopt a wavelet prototype function, called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients [3]. And if you further choose the best wavelets adapted to your data, or truncate the coefficients below a threshold, your data is sparsely represented. This sparse coding makes wavelets an excellent tool in the field of data compression. Other applied fields that are making use of wavelets include astronomy, acoustics, nuclear engineering, sub-band coding, signal and image processing, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, turbulence, earthquake-prediction, radar, human vision, and pure mathematics applications such as solving partial differential equations[4].

The organization of the paper is as follows. Section 2, provides a brief overview of the related work on wavelet based speech compression and analysis. Section 3 presents the discrete wavelet transform. Section 4 demonstrates the results obtained by using Matlab

Toolbox for Wavelet Analysis and finally in Section 5, we provide concluding remarks followed by a discussion of opportunities for future research.

2. Related Work

A general overview of wavelet is given in [5-7].

Amara Graps[8] provides an overview of wavelets that cut up data into different frequencies components and than study each component with a resolution matched to its scale. The paper introduces wavelets to the interested technical person outside of the digital signal processing field. It describe the history of wavelets beginning with Fourier, compare wavelet transforms with Fourier transforms, state properties and other special aspects of wavelets, and finish with some interesting applications such as image compression, musical tones, and de-noising noisy data.

George Tzanetakis and Perry cook[9] gives a brief account of their work i.e. to analyze the temporal and spectral properties of non stationary signals like audio using the Discrete Wavelet Transform, they describe some application of Discrete Wavelet Transform to the problem of extracting information from non-speech audio. More specifically automatic classification of various types of audio using the Discrete Wavelet Transform is described and compared with other traditional feature extractor proposed in the literature. In addition, a technique for detecting the beat attributes of music is presented.

M. L. Hilton and R. T . Ogden[10] propose a data adaptive scheme for selecting the threshold for wavelet shrinkage based noise removal. The method involves a statistical test of hypothesis based on a two dimensional cumulative sum of wavelet coefficients, which takes into account the coefficients magnitude and their positions. The amount of smoothing performed during the removal is controlled by the alpha attribute, which is the user supplied confidence level of the test. Simulated critical points for the statistical test are tabulated for a wide range of signals sizes and confidence levels.

Goran Kronquist and Henrik Storm[11], have considered a thresholding technique on the wavelet coefficients in order to reduce the background noise in the speech signal. The reduction of the coefficients is made with a variant of soft thresholding, especially adapted to speech signals. The noise levels are determined with the help of a training sequence and are adaptively changed. It shows that their way of thresholding is particularly suitable when denoising speech signal does not change the characteristics of background noise, it just decreases its amplitude.

3. Wavelet Analysis

The fundamental idea behind wavelets is to analyze according to scale. The wavelet analysis procedure is to adopt a wavelet prototype function called an *analyzing* wavelet or *mother* wavelet. Any signal can then be represented by translated and scaled versions of the mother wavelet. Wavelet analysis is capable of revealing aspects of data that other signal analysis techniques such as Fourier analysis miss, aspects like trends, breakdown points, discontinuities in higher derivatives, and self-similarity. Furthermore, because it affords a different view of data than those presented by traditional techniques, it can compress or de-noise a signal without appreciable degradation [8].

The Fourier Analysis is Defined by the equation

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \tag{1}$$

3.1 Continuous Wavelet Transform

In continuous wavelet transform (CWT), information about a signal is obtained by manipulating the wavelet functions along the time axis. A wavelet prototype function $\psi_{s,x}(x)$ at a scale *s* and a spatial displacement *u* is defined as:

$$\psi_{s,u}(x) = \sqrt{s}\,\psi\left[\frac{(x-u)}{s}\right] \tag{2}$$

Replacing the complex exponential in the Fourier Equation 1 with this function yields the ***continuous wavelet transform*** (CWT):

$$C(s,u) = \int_{-\infty}^{\infty} f(t) \sqrt{s}\,\psi\left[\frac{(x-u)}{s}\right] dt \tag{3}$$

which is the sum over all time of the signal multiplied by scaled and shifted versions of the wavelet function. The results of the CWT are many wavelet coefficients C, which are a function of scale and position. Multiplying each coefficient by the appropriately scaled and shifted wavelet yields the constituent wavelets of the original signal.

The basis functions in both Fourier and wavelet analysis are localized in frequency making mathematical tools such as power spectra (power in a frequency interval) useful at picking out frequencies and calculating power distributions.

The parameter scale used in wavelet transformation is similar to the scale used in the maps. At high scale, the wavelet seeks for global information or low frequencies information about the signal. At low scale, the wavelet seeks for detailed information or high frequencies information about the signal.

At high frequencies, this window will have a narrower width that corresponds to good time resolution and a longer height that corresponds to poor frequency resolution. At low frequencies, this window will have a wider width that corresponds to poor time resolution and a shorter height that corresponds to good frequency resolution.[12] Such analysis approach is suitable for most signals since most of the high frequencies occur for a small duration of time while low frequencies occur for long duration of time.

3.2 The Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) involves choosing scales and positions based on powers of two so called dyadic scales and positions. The mother wavelet is rescaled or dilated., by powers of two and translated by integers[13].

Specifically, a function *f(t)* $\mathcal{E}$.L^2 **(**R) (defines space of square integrable functions) can be represented as

$$f(t) = \sum_{j=1}^{L} \sum_{k=-\infty}^{\infty} d(j,k)\psi(2^{-j}t - k) + \sum_{k=-\infty}^{\infty} a(L,k)\phi(2^{-L}t - k) \tag{4}$$

The function ψ(t) is known as the mother wavelet, while Φ(t) is known as the scaling function. The set of functions

$$\{\sqrt{2^{-L}}\phi(2^{-L}t - k), \sqrt{2^{-j}}\psi(2^{-j}t - k) / j \le L, j,k,L \in Z\}, \tag{5}$$

where Z is the set of integers, is an orthonormal basis for L^2 **(**R).

The numbers *a(L, k)* are known as the approximation coefficients at scale L, while *d(j,k)* are known as the detail coefficients at scale j. The approximation and detail coefficients can be expressed as:

$$a(L,k) = \frac{1}{\sqrt{2^L}} \int_{-\infty}^{\infty} f(t)\phi(2^{-L}t - k)dt \tag{6}$$

$$d(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t)\psi(2^{-j}t - k)dt \tag{7}$$

To provide some understanding of the above coefficients consider a projection $f_l(t)$ of the function *f(t)* that provides the best approximation (in the sense of minimum error energy) to *f(t)* at a scale l[14]. This projection can be constructed from the coefficients *a(L, k)*, using the equation

$$f_l(t) = \sum_{k=-\infty}^{\infty} a(l,k)\phi(2^{-l}t - k). \tag{8}$$

As the scale l decreases, the approximation becomes finer, converging to $f(t)$ as l ->˜ 0. The difference between the approximation at scale $l + 1$ and that at l, $f_{l+1}(t) - f_l(t)$, is completely described by the coefficients $d(j, k)$ using the equation

$$f_{l+1}(t) - f_l(t) = \sum_{k=-\infty}^{\infty} d(l,k)\psi(2^{-l}t - k).$$

(9)

Using these relations, given by $a(L, k)$ and $\{d(j, k) \mid j < L\}$, it is clear that we can build the approximation at any scale. Hence, the wavelet transform breaks the signal up into a coarse approximation $f_L(t)$ (given $a(L, k)$) and a number of layers of detail $\{f_{j+1}(t)-f_j(t) \mid j < L\}$ (given by $\{d(j, k) \mid j\ L\}$). As each layer of detail is added, the approximation at the next finer scale is achieved

3.3 The Fast Wavelet Transform Algorithm

The Discrete Wavelet Transform (DWT) coefficients can be computed by using Mallat 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[5].

3.3.1 Implementation Using Filters

To explain the implementation of the Fast Wavelet Transform algorithm consider the following equations:

$$\phi(t) = \sum_k c(k)\phi(2t - k)$$

$$\psi(t) = \sum_k (-1)^k c(1 - k)\phi(2t - k)$$

$$\sum_k c_k c_{k-2m} = 2\delta_{0,m}$$

.(10)

The first equation is known as the *twin-scale relation* (or the dilation equation) and defines the scaling function $\Phi(t)$. The next equation expresses the wavelet ψ. in terms of the scaling function $\Phi(t)$. The third equation is the condition required for the wavelet to be orthogonal to the scaling function and its translates

The coefficients c(k) or $\{c_0, \ldots, c_{2N-1}\}$ in the above equations represent the impulse response coefficients for a low pass filter of length 2N, with a sum of 1

The high pass filter is obtained from the low pass filter using the relationship $g_k = (-1)^k c(1-k)$, where k varies over the range (1-(2N -1)) to 1.

Top

Equation 10 shows that the scaling function is essentially a low pass filter and is used to define the approximations. The wavelet function defined by equation 10 is a high pass filter and defines the details. Starting with a discrete input signal vector s, the first stage of the FWT algorithm decomposes the signal into two sets of coefficients. These are the approximation coefficients cA1 (low frequency information) and the detail coefficients cD1 (high frequency information)[15], as shown in the figure 1 below.

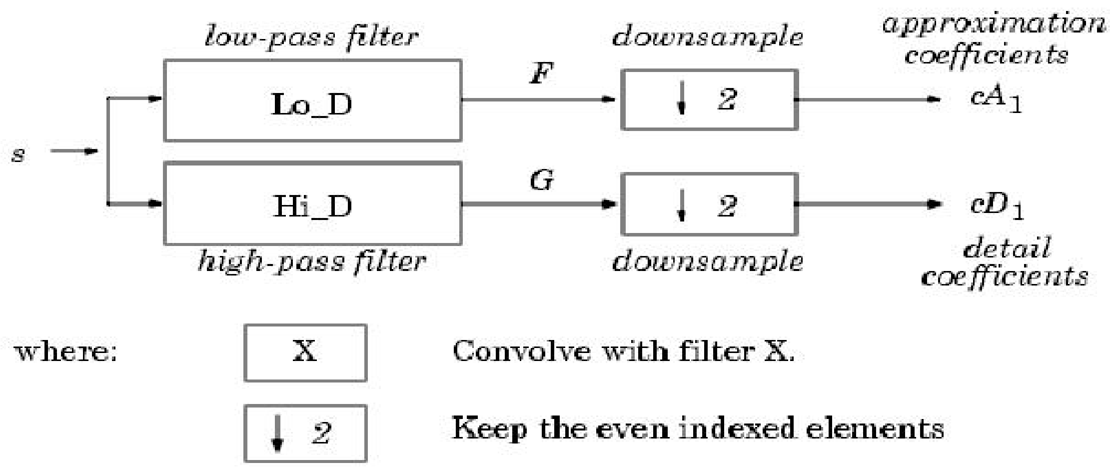

Figure 1 Filtering Operation of the DWT

The coefficient vectors are obtained by convolving s with the low-pass filter Lo_D for approximation and with the high-pass filter Hi_D for details. This filtering operation is then followed by dyadic decimation or down sampling by a factor of 2.

Mathematically the two-channel filtering of the discrete signal s is represented by the expressions:

$$cA_1 = \sum_k c_k s_{2i-k} , \qquad cD_1 = \sum_k g_k s_{2i-k} \tag{11}$$

These equations implement a convolution plus down sampling by a factor 2 and give the forward fast wavelet transform. If the length of each filter is equal to 2N and the length of the original signal s is equal to n, then the corresponding lengths of the coefficients cA1 and cD1 are given by the formula:

$$\text{floor}\left(\frac{n-1}{2}\right) + N \tag{12}$$

This shows that the total length of the wavelet coefficients is always slightly greater than the length of the original signal due to the filtering process used.

3.3.2 Multilevel Decomposition

Top

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 shown in figure 2 below.

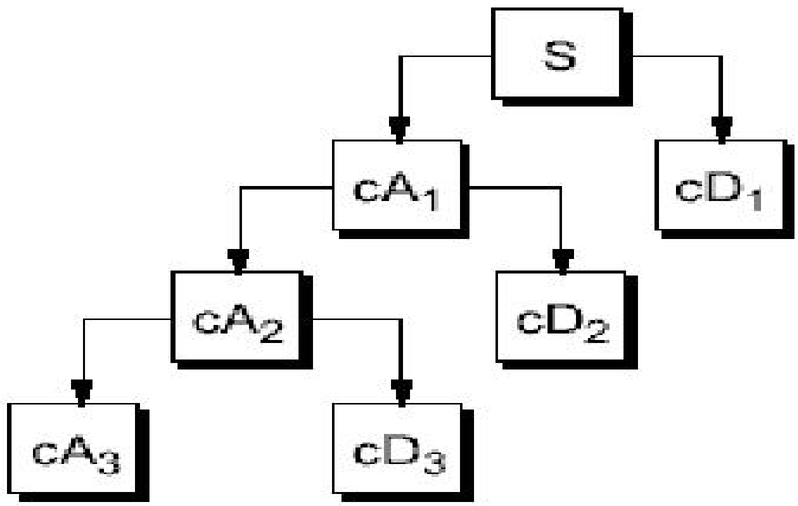

Figure 2. Decomposition of DWT Coefficients

The wavelet decomposition of the signal s analyzed at level j has the following structure $[cA_j, cD_j, –.., cD_1]$.

Looking at a signals wavelet decomposition tree can reveal valuable information. The figure 3 below shows the wavelet decomposition to level 3 of a sample signal S.

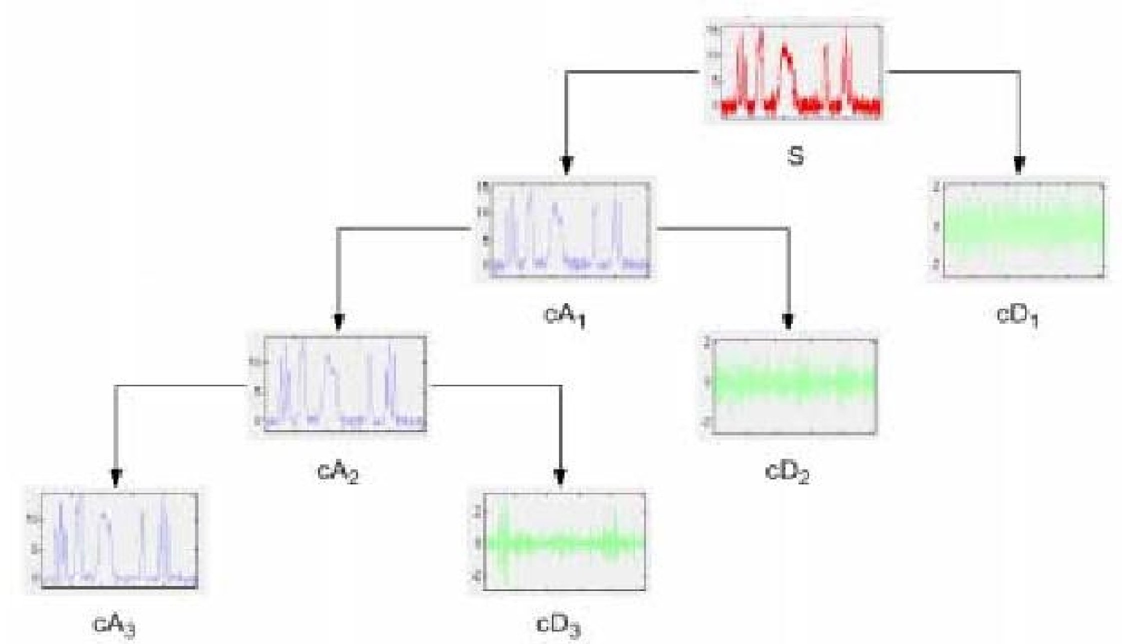

Figure 3. Level 3 Decomposition of Sample Signal S

Top

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.

3.4 Signal Reconstruction

The original signal can be reconstructed or synthesized using the inverse discrete wavelet transform (IDWT). The synthesis starts with the approximation and detail coefficients cA_j and cD_j, and then reconstructs cA_{j-1} by up sampling and filtering with the reconstruction filters shown in figure 4 below.

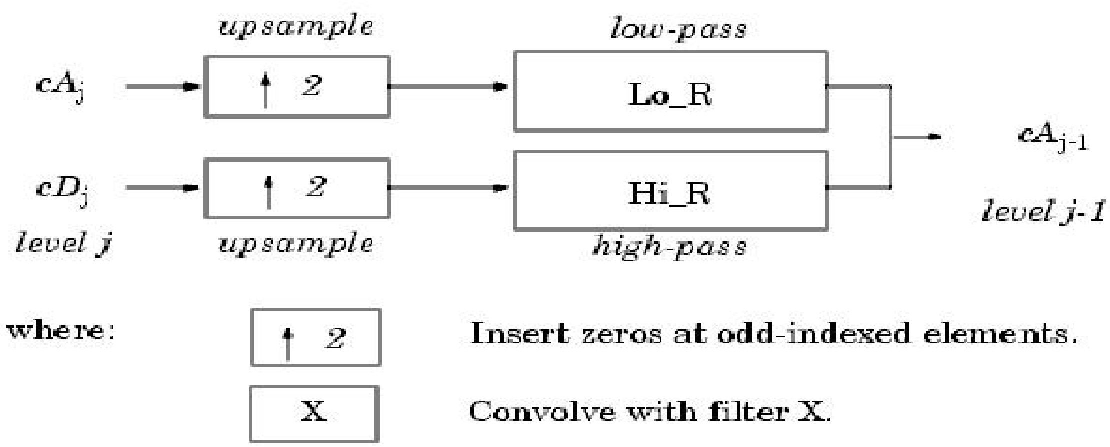

Figure 4. Wavelets Reconstruction

The reconstruction filters are designed in such a way to cancel out the effects of aliasing introduced in the wavelet decomposition phase. The reconstruction filters (Lo_R and Hi_R) together with the low and high pass decomposition filters, forms a system known as *quadrature 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.

4. Results and Discussion

The figure 5 below illustrates the different processes involved in compressing speech signals using wavelets. In the compression software these different stages were designed and coded in software, using Matlab version 5.3[16], with the exception of the last two processes.

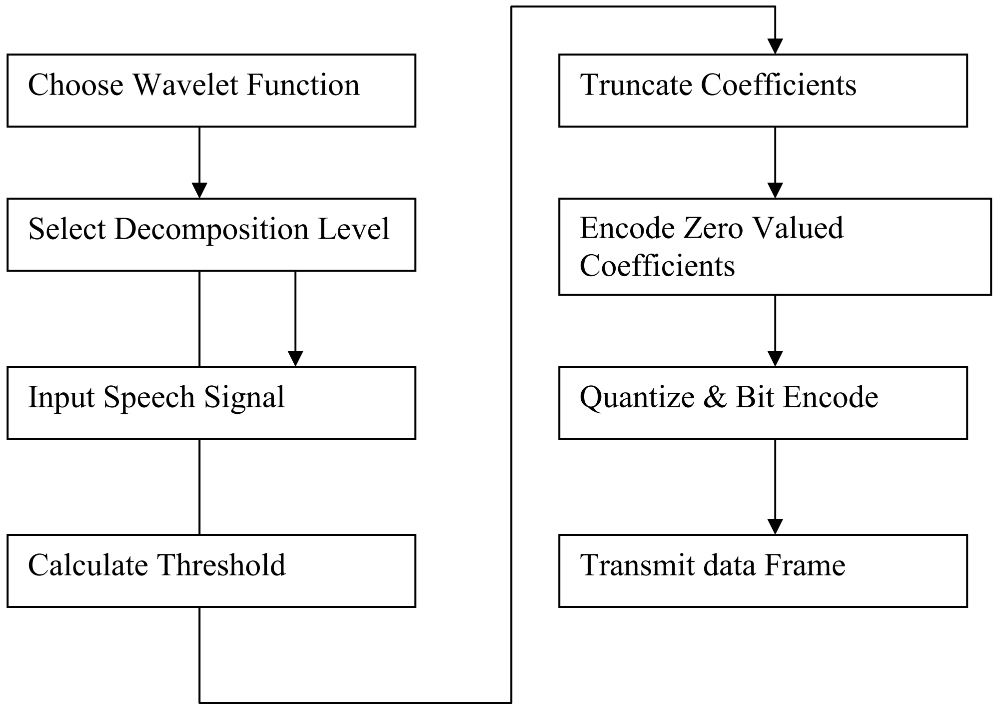

Figure 5. Design Flow of Wavelet Based Speech Coder

4.1 Performance Measures

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

- Signal to Noise Ratio (SNR),
- Peak Signal to Noise Ratio (PSNR),
- Normalized Root Mean Square Error (NRMSE),
- Retained Signal Energy

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

4.1.1 Signal to Noise Ratio

$$SNR = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \tag{13}$$

σ_x^2 is the mean square of the speech signal and σ_e^2 is the mean square difference between the original and reconstructed signals[17].

Top

4.1.2 Peak Signal to Noise Ratio

$$PSNR = 10 log_{10} \frac{NX^2}{\|x-r\|^2} \tag{14}$$

N is the length of the reconstructed signal, X is the maximum absolute square value of the signal x and || *x-r*||˜ is the energy of the difference between the original and reconstructed signals[17].

4.1.3 Normalised Root Mean Square Error

$$NRMSE = \sqrt{\frac{\overline{(x(n)-r(n))^2}}{\overline{(x(n)-\mu_x(n))^2}}} \tag{15}$$

x(n) is the speech signal, *r(n)* is the reconstructed signal, and *µx(n)* is the mean of the speech signal[17]

4.1.4 Retained Signal Energy

$$RSE = \frac{100 * \|x(n)\|^2}{\|r(n)\|^2} \tag{16}$$

|| *x(n)*|| is the norm of the original signal and *) r(n* is the norm of the reconstructed signal. For one-dimensional orthogonal wavelets the retained energy is equal to the L_2-norm recovery performance.

4.2 Performance Measure Results

The original speech signal which was used to obtain the performance measure results is shown in figure 6. This speech was recorded using windows sound recorder for testing of speech coding software.

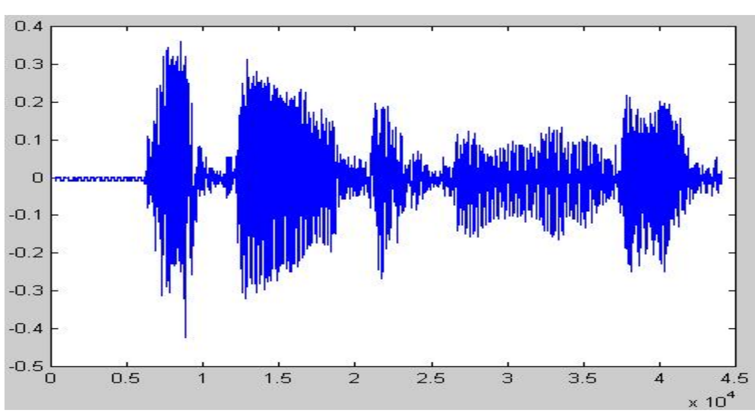

Figure 6. Original Speech Signal

Top

A male spoken speech signals was decomposed at scale 3 and level dependent thresholds were applied. Since the speech files were of short duration, the entire signal was decomposed at once without framing. The performance measure results are summarized in the following table for the different wavelets used.

The sentence spoken was “ok this time we will do it”

Wavelet	**Zeroes(%)**	**Retained Energy(%)**	**SNR**	**PSNR**	**NRMSE**
Haar	42.5067	99.9885	58.6709	66.4207	0.0012
DB4	53.1152	99.9999	39.3795	47.1321	0.0108
DB6	52.8592	99.9886	39.4174	47.1699	0.0107
DB8	55.1680	99.9864	38.6610	46.4136	0.0117
DB10	55.0480	99.9865	38.7040	46.4602	0.0116

Table Performance Measure Results

The graphical comparison of performance measures is shown in the following Bar chart graph of figure 7.. The graph clearly demonstrates that the HAAR wavelet have a relatively high SNR as compared to the DAUBECHIE family

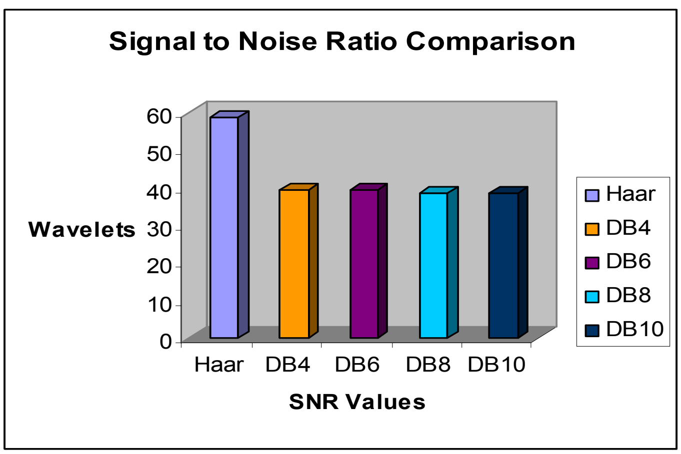

Figure 7 .Signal to Noise Ratio Comparison Results

Figure 8 shows the Peak Signal to Noise Ratio Performance Measure Results, and once again HAAR wavelet have a relatively high PSNR as compared to the DAUBECHIE family.

Top

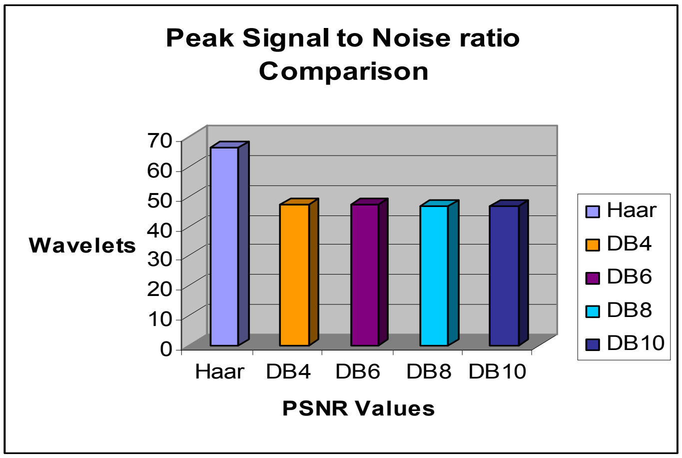

Figure 8 Peak Signal to Noise Ratio Comparison Results

Figure 9 shows the Normalized Root Mean Square Error Performance Measure Results, and once again HAAR wavelet have a relatively low NRMSE values as compared to the DAUBECHIE family

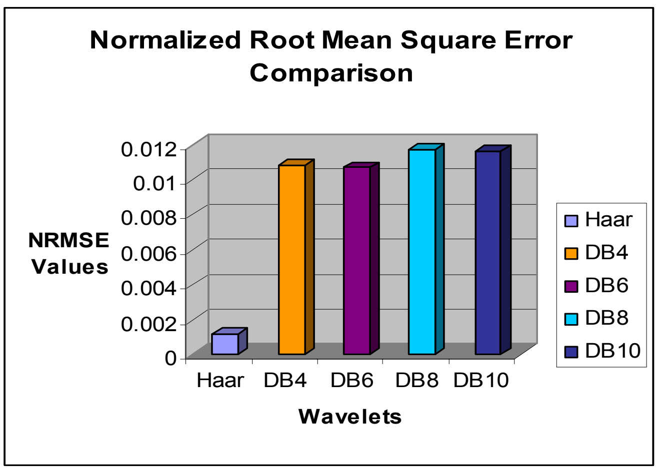

Figure 9 Normalized Root Mean Squared Error Comparison

5. Conclusion

Speech coding is currently an active topic for research in the areas of VOIP and Mobile Communication Systems. The Discrete Wavelet Transform performs very well in the compression of recorded speech signals. The analysis was carried out on a recorded speech signal and performance measure results were obtained using the Haar and Daubechies wavelet families. More specifically, the results demonstrates that the Haar wavelet best suits the compression of speech signals due to their high SNR values of 58.6709 and low NRMSE values of 0.0012 as compared with the other families of Daubechies wavelet. These parameter values are very significant in design of efficient wavelet based speech compression software for multimedia and mobile applications. In addition, using wavelets the compression ratio can be easily varied, while most other compression techniques have fixed compression ratios.

Acknowledgement

The authors acknowledge the support provide by Mehran University of Engineering and Technology, Jamshoro, Pakistan for providing us with the lab facilities and software tools.

Reference

[1] R.E. Blahut, *Digital Transmission of Informatio*n, Addison-Wesley, New York, 1990.

[2] R. Polikar, ”The Wavelet Tutorial,”, available at http://engineering.rowan.edu/~polikar/WAVELETS/WTtutorial.html, March 20, 2001

[3] A. Chen, N. Shehad, A. Virani and E. Welsh, "Discrete Wavelet Transform for Audio Compression," http://is.rice.edu/~welsh/elec431/wavelets.html,July. 16, 2001.

[4] M. Vetterli and J. Kovacevic, *Wavelets and Subband Codin*g, Prentice Hall PTR, Englewood Cliffs, N.J, 1995.

[5] Y. Nievergelt, *Wavelets made eas**y,*** Birkhäuser, Boston, 1999.

[6] Ingrid Daubechies. "The Wavelet Transform, time frequency localization and signal analysis", IEEE transactions on information theory, 36(5):961-1005, September 1990.

[7] Alfred Haar. "Zur Theorie der orthogonalen funktionensysteme", Mathematische Annalen, 69:331-371, 1910.

[8] A. Graps, "An Introduction to Wavelets" *IEEE Computational Sciences and Engineerin*g, http://www.amara.com/IEEEwave/IEEEwavelet.html, March.15, 2001).

[9] George Tzanetakis and Perry cook "Audio analysis using the Discrete Wavelet transform", Proc International Conference on Acoustic Speech and Signal processing (ICASS), 1980.

[10] M. L. Hilton and R. T . Ogden "data Analysis Wavelet Threshold Selection in 2D Signal De-noising", IEEE Transaction on Pattern Analysis and Machine Learning , vol 11 PP 674-693, 1999.

[11] Goran Kronquist and Henrik Storm, " Denoising a Speech Signal using Wavelets", IEEE Transaction on Communication vol 28, 1990 pp84-95.

[12] V. Viswanathan, W. Anderson, J. Rowlands, M. Ali and A. Tewfik, ieReal-Time Implementation of a Wavelet-Based Audio Coder on the T1 TMS320C31 DSP Chip,lC *5*

th International Conference on Signal Processing Applications &Technology (ICSPAT), Dallas, TX, Oct. 1994.

[13] B. Lin, B. Nguyen and E.T. Olsen, ilOrthogonal Wavelets and Signal Processing,l. *Signal Processing Methods for Audio, Images and Telecommunication*s, .M.Clarkson and H.Stark, ed., Academic Press, London, 1995, pp. 1-70.

[14] S. Mallat, *A Wavelet Tour of Signal Processin*g, Academic Press, San Diego,Calif., 1998.

[15] J. Ooi and V. Viswanathan, iiApplications of Wavelets to Speech Processing,ln *Modern Methods of Speech Processin*g, R.P. Ramachandran and R. Mammone, ed., Kluwer Academic Publishers, Boston, 1995, pp. 449-464.

[16] The Wavelet Toolbox User Guide ver 1.2 , MathWorks Inc., Matlab, Version 5.3

[17] E.B. Fgee, W.J. Phillips, W. Robertson, ioComparing Audio Compression using Wavelets with other Audio Compression Schemes,lo *IEEE Canadian Conference on Electrical and Computer Engineering,* IEEE, Edmonton, Canada, 1999, pp.698-701.

Top