ISSN:2231-0711
www.ijcset.net

Frequency Domain Compression of Speech Signals

N. Siddiah[1], T.Srikanth[2] and Y. Satish Kumar[3]

[1] *Dept. of E.C.E., Mekapati Raja Mohana Reddy Institute of Technology & Science, Udyagiri, A.P., India.*
[2] *Dept. of E.C.E., Kallam Haranadha Reddy Institute of Technology, Guntur, A.P., India.*
[3] *Dept. of E.C.E., Chalapathi Institute of Technology, Guntur, A.P., India.*
siddun89@ymail.com

***Abstract*- In communication systems, sound files and also disk space for storage, compression of these files has become a necessity. 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. In this paper we are using one of the familiar wavelet transform for compressing speech signal in frequency domain.**
Wavelet analysis is the breaking up of a signal into a set of scaled and translated versions of an original wavelet. Taking the wavelet transform of a signal decomposes the original signal into wavelets coefficients at different scales and positions. These coefficients represent the signal in the wavelet domain and all data operations can be performed using the just corresponding wavelet coefficients. The major issues concerning the design of this Wavelet based speech coder are choosing optimal wavelets for speech signals, decomposition level in the wavelet transform, thresholding criteria for coefficient truncation and efficient encoding of truncated coefficients. The performance of the wavelet compression scheme on recorded speech was calculated. A significant advantage of using wavelets for speech compression ratio can easily be varied, while most of the other techniques have fixed compression ratios.

Keywords- **Speech Signal, Speech Compression, Wavelet transform, Frequency domain.**

I. INTRODUCTION

Human language in its original form is an acoustic signal. For the purpose of communication and storage, it is necessary to convert it into an electric signal. This is achieved with the help of certain instruments called transducers. With the advent of digital computing machines, it was proposed to exploit the power of the same for processing voice signals. This requires a digital representation of voice to achieve this, the analog signal is sampled at some frequency and then quantized at discrete levels .

Thus, parameters of digital speech are

1. Sampling rate
2. Bits per second
3. Number of channels.

The sound files can be stored and played in digital computers.

In recent years, the transfer of information on a large scale remote computing and the development of mass storage and retrieval systems have witnessed tremendous growth. To cope with the growth in size of the databases, additional storage devices need to be installed and modems and multiplexers have to be constantly updated to allow large quantities of data transfer between computers and remote terminals.. This leads to an increase in the cost as well as equipment. One solution to these problems is "COMPRESSION" where the database and the transmission sequence can be encoded efficiently.

The primary objective of this paper is to represent the wavelet based method for the compression of speech. The algorithm presented here was implemented in MATLAB. This paper is an application of wavelets, it was natural to study the basics of wavelets in detail.. However, the wavelet itself is an engrossing field, and a comprehensive study was beyond the scope of our graduate level. Hence, attempt is made only to explain the very basics which are indispensable from the compression point of view. This approach led to the elimination of many of the mammoth sized equations and vector analysis inherent in the study of wavelets.

II. THE WAVELET TRANSFORM

The fundamental idea behind wavelets is to analyse according to scale. Indeed, some researchers in the wavelet field feel that, by using wavelets, one is adopting a whole new mindset or perspective in processing data [16].

Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions. This idea is not new. Approximation using superposition of functions has existed since the early 1800's, when Joseph Fourier discovered that he could superpose sines and cosines to represent other functions. However, in wavelet analysis, the *scale* that we use to look at data plays a special role. Wavelet algorithms process data at different *scales* or *resolutions.* If we look at a signal with a large "window", we would notice gross features. Similarly, if we look at a signal with a small "window", we would notice small features. The result in wavelet analysis is to see both the forest *and* the trees, so to speak [1].

This makes wavelets interesting and useful. For many decades, scientists have wanted more appropriate functions than the sines and cosines which comprise the bases of Fourier analysis, to approximate choppy signals [2]. By their definition, these functions are non-local (and stretch out to infinity). But with wavelet analysis, we can use approximating functions that are contained neatly infinite domains. Wavelets are well-suited for approximating data with sharp discontinuities [3]-[5].

The wavelet analysis procedure is to adopt a wavelet prototype function, called an analysing 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. 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 [8], [15].

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 [7].

A. ***Discrete Wavelet Transform***

Calculating wavelet coefficients at every possible scale (for continuous WT) is a fair amount of work, and it generates an awful lot of data. It turns out, rather remarkably, that if we choose scales and positions based on powers of two -- so-called dyadic scales and positions -- then our analysis will be much more efficient and just as accurate. We obtain such an analysis from the discrete wavelet transform (DWT). An efficient way to implement this scheme using filters was developed in 1988 by Mallat. The Mallat algorithm is in fact a classical scheme known in the signal processing community as a two-channel subband coder. This very practical filtering algorithm yields a fast wavelet transform -- a box into which signal passes, and out of which wavelet coefficients quickly emerge. A discussion of MRA (Multi-resolution analysis or approximation) bridges the gap between wavelets and the filter-bank implementation of DWT explained in this section.

We directly begin our discussion with the formula of DWT and then veer towards the decomposition of signal into approximation and detail coefficients. The filter banks used to achieve this are also discussed. The reverse process, i.e. reconstruction of signal from the coefficients is described later. Examples of haar, and db10 are used to demonstrate the filter coefficients, frequency response of the low and high pass decomposition and reconstruction filters. This chapter forms the basis for the next chapter, which discusses compression.

The Discrete Wavelet Transform (DWT) involves choosing scales and positions based on powers of two-the so called dyadic scales and positions [10]. The mother wavelet is rescaled or "dilated" by powers of two and translated by integers. Specifically, a function *f(t) Î* L2 **(**R) (defines space of square inferable functions) can be represented as

$$f(t) = \sum_{J=1}^{L} \sum_{K=-\infty}^{\infty} d(j,k).\Psi(2^{-j}t - k) + \sum_{K=-\infty}^{\infty} a(L,k).\varphi(2^{-L}t - k)$$

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

$$\{\sqrt{2^{-L}}\varphi(2^{-L}t - k), \sqrt{2^{-j}}\Psi(2^{-j}t - k)..|j \le L; j, k, L \in Z\}$$

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.

These approximation and detail coefficients can be expressed as

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

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

The above two equations give a mathematical relationship to compute the approximation and detail coefficients.

This procedure is seldom adopted. A more practical approach is to use Mallat's Fast Wavelet Transform algorithm. The Mallat algorithm for discrete wavelet transform (DWT) is, in fact, a classical scheme in the signal processing community, known as a *two channel subband coder* using conjugate quadrature filters or quadrature mirror filters (QMF).

B.One-Stage Filtering

For many signals, the low-frequency content is the most important part. It is what gives the signal its identity. The high-frequency content, on the other hand, imparts flavor or nuance. Consider the human voice. If we remove the high-frequency components, the voice sounds different, but we can still tell what's being said. However, if we remove enough of the low-frequency components, we hear gibberish In wavelet analysis, we often speak of approximations and details. The approximations are the high-scale, low-frequency components of the signal. The details are the low-scale, high-frequency components. The filtering process, at its most basic level, looks like this:

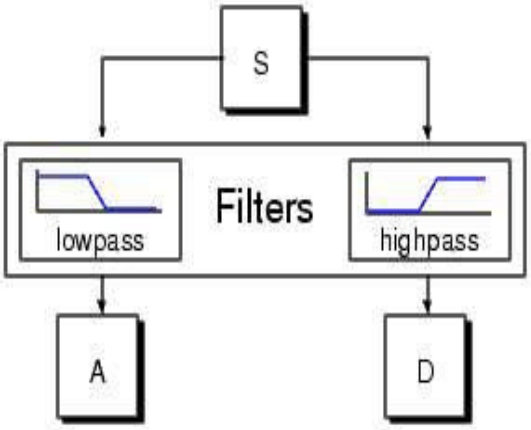

Fig.1: One stage filtering scheme producing the approximation and detail components of the signal.

The original signal, S, passes through two complementary filters and emerges as two signals. Unfortunately, if we actually perform this operation on a real digital signal, we wind up with twice as much data

as we started with. Suppose, for instance, that the original signal S consists of 1000 samples of data. Then the resulting signals will each have 1000 samples, for a total of 2000.

These signals A and D are interesting, but we get 2000 values instead of the 1000 we had. There exists a more subtle way to perform the decomposition using wavelets. By looking carefully at the computation, we may keep only one point out of two in each of the two 2000-length samples to get the complete information. This is the notion of ***down sampling***. We produce two sequences called cA and cD

The process on the right, which includes down sampling, produces DWT coefficients. To gain a better appreciation of this process, let's perform a one-stage discrete wavelet transform of a signal. Our signal will be a pure sinusoid with high-frequency noise added to it.

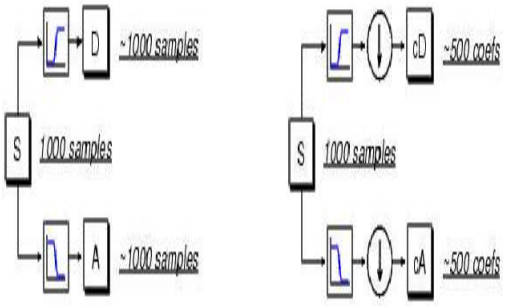

Fig 2: Producing approximation and detail coefficients at the first level

Here is our schematic diagram with real signals inserted into it: The detail coefficients cD are small and consist mainly of a high-frequency noise, while the approximation coefficients cA contain much less noise than does the original signal.

The actual lengths of the detail and approximation coefficient vectors are slightly more than half the length of the original signal. This has to do with the filtering process, which is implemented by convolving the signal with a filter. The convolution "smears" the signal, introducing several extra samples into the result. In this section, we considered only one-stage decomposition of the signal into cA and cD coefficient. This process can be repeated to get multiple-level decomposition.

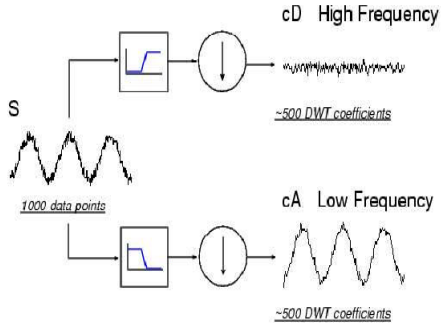

Fig. 3: Demonstration of one-stage filtering scheme for producing approximation and detail coefficient.

C. Reconstructing Approximations and Detail Coefficients

It is possible to reconstruct our original signal from the coefficients of the approximations and details. As an example, let's consider how we would reconstruct the first-level approximation A1 from the coefficient vector cA1 We pass the coefficient vector cA1 through the same process we used to reconstruct the original signal. However, instead of combining it with the level-one detail cD1, we feed in a vector of zeros in place of the detail coefficients vector:

The process yields a reconstructed approximation A1, which has the same length as the original signal S and which is a real approximation of it.

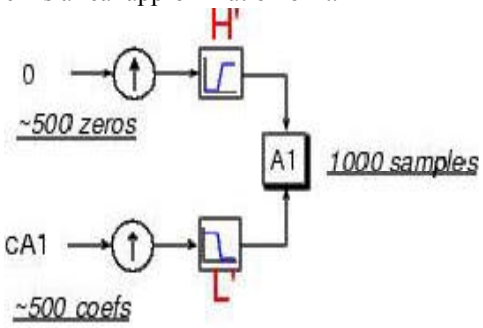

Fig 4: Obtaining the first level approximation of the signal.

Similarly, we can reconstruct the first-level detail D1, using the analogous process:

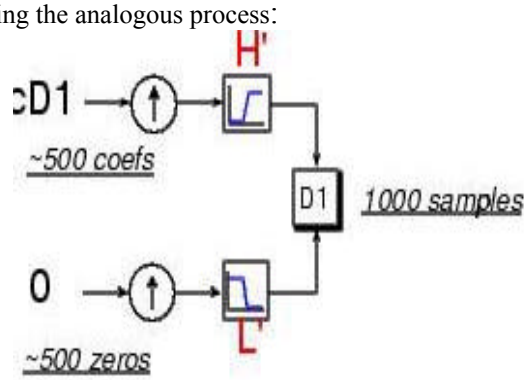

Fig 5: Obtaining the first level detail of the signal

The reconstructed details and approximations are true constituents of the original signal. In fact, we find when we combine them that:

$$A1 + D1 = S$$

Note that the coefficient vectors cA1 and cD1 -- because they were produced by down sampling and are only half the length of the original signal -- cannot directly be combined to reproduce the signal. It is necessary to reconstruct the approximations and details before combining them.

D. Multiple-Level Decomposition:

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*

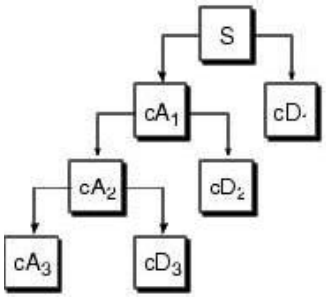

Fig 6: Multiple level decomposition trees.

Looking at a signal's wavelet decomposition tree can yield valuable Information

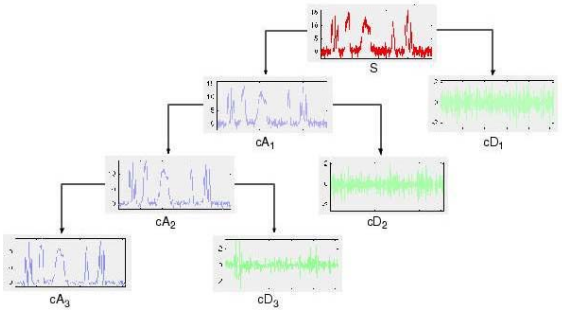

Fig 7: Multiple level decomposition of a signal.

*E. **Number of Levels:***

Since the analysis process is iterative, in theory it can be continued indefinitely. In reality, the decomposition can proceed only until the individual details consist of a single sample or pixel. In practice, you'll select a suitable number of levels based on the nature of the signal, or on a suitable criterion such as entropy The next step splits the approximation coefficients cA_1 in two parts using the same scheme, replacing s by cA_1 and producing cA_2 and cD_2, and soon. Now that we have seen the decomposition of a signal into wavelet (approximation and detail) coefficient s, it is natural to ask whether the reverse is possible, i.e., is it possible to generate the original signal back from the coefficients, and if yes, how to achieve this Fortunately, there does exist a method to do it, and it is very similar to the one used for decomposition.

*F. **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 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 D10, D8, D6 or D4 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.

*G. **Wavelet Decomposition***

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

III RESULTS AND PERFORMANCE MEASURE

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),
- Normalised Root Mean Square Error (NRMSE),
- Retained Signal Energy
- Compression Ratios
- Retained signal energy
- Compression factor

1. Signal to noise ratio (SNR): This value gives the quality of reconstructed signal.

Higher the value, better. It is given by:

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

where σ_x^2 and σ_e^2 are respectively the mean square of the speech signal and the mean square difference between the original and reconstructed signals.

2. Peak Signal to Noise Ratio [5]

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

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

3. Normalised Root Mean Square Error [5]

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

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

4. Retained Signal Energy

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 L2-norm recovery performance.

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

5. Compression Ratio

cWC is the length of the compressed wavelet transform vector.

$$C = \frac{length(x(n))}{length(cWC)}$$

6. Compression factor: It is the ratio of the original signal to the compressed signal. Of course, for the compressed signal we have to take into account all the

values that would be needed to completely represent the signal. As has been explained in the previous section, this project implements encoding using a modification of RLE wherein 2 vectors are produced, we must take into account the combined length of these 2 vectors.

7. Retained signal energy: This indicates the amount of energy retained in the compressed signal as a percentage of the energy of original signal. When compressing using orthogonal wavelets, the Retained energy in percentage is defined by: The amount of energy concentrated in level one approximation coefficients.

$$Retained\ energy = \frac{100 * (vector\ norm(coeffs\ of\ the\ current\ decomposition, 2))^2}{(vector\ norm(original\ signal, 2))^2}$$

Optimal Decomposition Level in Wavelet Transforms

The figure below shows a sample speech signal and approximations of the signal, at five different scales. These approximations are reconstructed from the coarse low frequency coefficients in the wavelet transform vector.

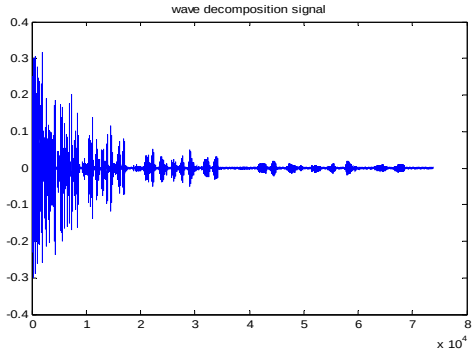

Fig 8: Wave decomposed signal.

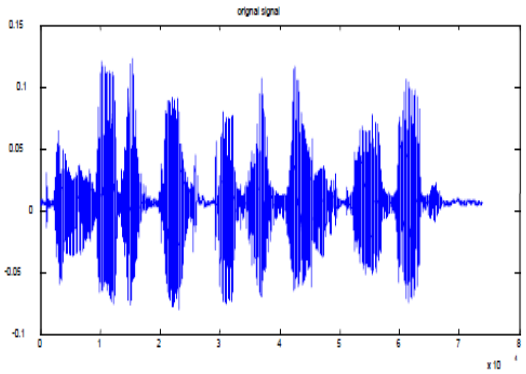

Fig 9:Original signal.

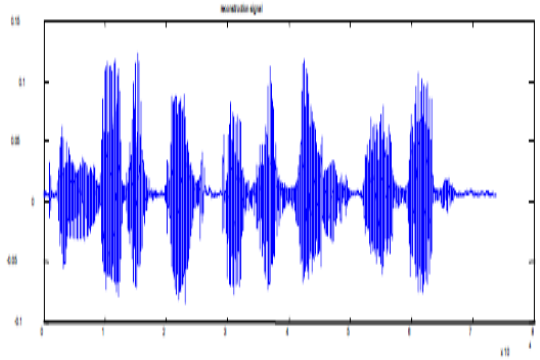

Fig 10: Output signal at decomposition level 1

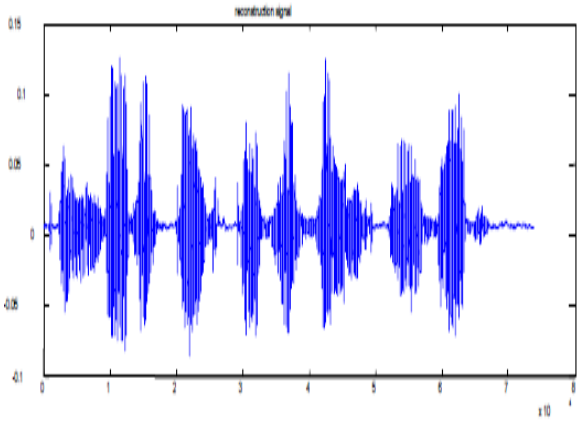

Fig 11: Output signal at decomposition level 2

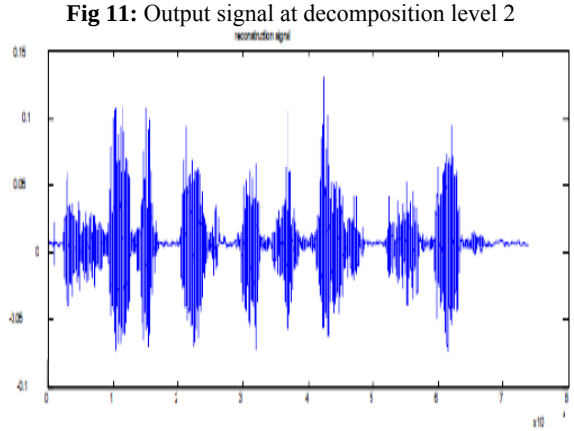

Fig 12: Output signal at decomposition level 3

A spoken speech signals were decomposed at scale 5 and level dependent thresholds were applied using the Birge-Massart strategy. Since the speech files were of short duration, the entire signal was decomposed at once without framing. A summary of the performance is given below for the different wavelets used.

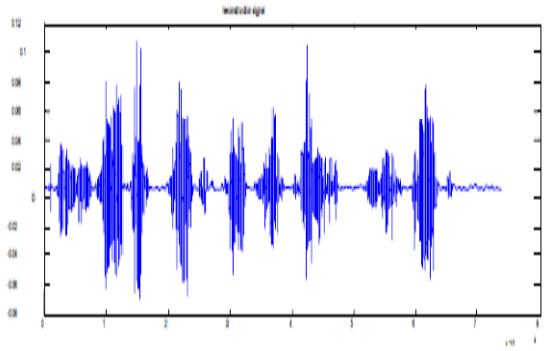

Fig 13:Output signal at decomposition level 4

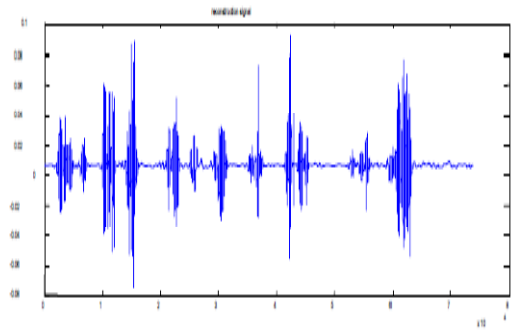

Fig 14: Output signal at decomposition level 5

A spoken speech signals were decomposed at scale 5 and level dependent thresholds were applied using the Birge-Massart strategy. Since the speech files were of short duration, the entire signal was decomposed at once without framing. A summary of the performance is given below for the different wavelets used.

TABLE 1: PERFORMANCR RECORDED SPEECH CODING

Wavelet	*%of zeros*	*Retained energy*	*SNR*	*PSNR*	*NRMSE*
Haar	94.280	97.340	17.250	7.720	0.313
Db4	94.270	96.870	19.020	10.420	0.229
Db6	94.267	96.862	19.191	10.660	0.222
Db8	94.257	96.852	19.363	10.901	0.216
Db10	94.249	96.851	19.363	10.908	0.215

IV. CONCLUSION

Speech coding is currently an active topic for research in the areas of Very Large Scale Integrated (VLSI) circuit technologies and Digital Signal Processing (DSP). The Discrete Wavelet Transform performs very well in the compression of recorded speech signals. For speech processing however, its performance is not as good. Therefore for speech coding it is recommended to use a wavelet with a small number of vanishing moments at level 5 decomposition or less.

The wavelet based compression software designed reaches a signal to noise ratio of 19.36 db at a compression ratio of 3.88 using the Daubechies 10 wavelet. The performance of the wavelet scheme in terms of compression scores and signal quality is good. In addition, using wavelets the compression ratio can be easily varied, while most other compression techniques have fixed compression ratio.

REFERENCES:

[1] Chan, Y.T., Wavelet Basics, Kluwer Academic Publisher, Boston, 1995.
[2] Cooley, J.W. and Tukey, J.W., An Algorithm For The Machine Computation Of Complex Fourier series, Mathematics of Computation, Vol. 19. pp: 297-301, 1965.
[3] Daubechies, I., The Wavelet Transform, Time Frequency Localization and Signal Analysis, IEEE Transaction on Information Theory, Vol. 36, No.5 pp: 961-1005, 1990.
[4] Feng, Yanhui,Thanagasundram, Schlindwein, S., Soares, F., Discrete wavelet-based thresholding study on acoustic emission signals to detect bearing defect on a rotating machine, Thirteenth International Congress on Sound and Vibration, Vienna, Austria July 2-6, 2006.
[5] Graps, A., An Introduction to Wavelets, IEEE Computational Sciences and Engineering, Volume 2, Number 2, pp: 50-61, Summer 1995.
[6] Karam, J.R., Phillips, W.J. and Robertson, W., Optimal Feature Vector for Speech Recognition Of Unequally Segmented Spoken Digits, IEEE, proceedings of CCECE, Halifax, pp:327-330 May, 2000.
[7] Karam, J., A Global Threshold Wavelet-Based Scheme for Speech Recognition, Third International conference on Computer Science, Software Engineering Information Technology, E-Business and Applications, Cairo, Egypt, Dec. 27-29 2004.
[8] Karam, J., Saad, R., The Effect of Different Compression Schemes on Speech Signals, International Journal of Biomedical Sciences, Vol. 1 No. 4, pp: 230 234, 2006.
[9] Oppenheim, A.V. and Schafer, R.W., Discrete-Time Signal Processing, Prentice Hall, Englewood Cliffs, New Jersey, 1989.
[10] Rabiner, L., Digital Formant Synthesizer For Speech Synthesis Studies, J. Acoust. Soc. Am., Vol, 43, No. 2, pp: 822-828, April 1968.
[11] Rabiner, L. Juang, B., Fundamental of Speech Recognition, Prentice Hall, New Jersey, 1993.
[12] Rabiner, L.R. and Schafer, R.W., Digital Processing of Speech Signals, Prentice Hall, New Jersey, 1978.
[13] Picone, J.W., Signal Modeling Techniques in Speech Recognition, IEEE, Vol.81, No.9, September 1993.
[14] Strang, G. and Nguyen, T., Wavelets and Filter Banks, Wellesley MA, Wellesley–Cambridge Press, Wellesley, MA, 1996.
[15] Taswell, C., Speech Compression with Cosine and Wavelet packet near best bases, IEEE International Conference on Acoustic, Speech, and Signal Processing, p.p 566-568 Vol. 1, May 1996.
[16] Young, R.K., Wavelet Theory and its Applications, Kluwer Academic Publishers, Lancaster, USA 1995.