Speech Compression using Different TransformTechniques

G. Rajesh[1], A. Kumar[1] and K. Ranjeet[1]
[1]Indian Institute of Information Technology, Design & Manufacturing Jabalpur
Jabalpur (M.P.)-482005, INDIA
E-mail: rajesh1315@gmail.com, anilkdee@gmail.com, ranjeet281@gmail.com

***Abstract*— Speech Compression is a field of digital signal processing that focuses on reducing bit-rate of speech signals to enhance transmission speed and storage requirement of fast developing multimedia. This paper explores a transform based methodology for compression of the speech signal. In this methodology, different transforms such as Discrete Wavelet Transform (DWT), fast Fourier Transform (FFT) and Discrete Cosine Transform (DCT) are exploited. A comparative study of performance of different transforms is made in terms of Signal-to-noise ratio (*SNR*), Peak signal-to-noise ratio (*PSNR*) and Normalized root-mean square error (*NRMSE*). The simulation results included illustrate the effectiveness of these transforms in the field of data compression. When compared, Discrete Wavelet Transform gives higher compression respect to Discrete Cosine Transform and Fast Fourier Transform in terms of compression ratio, and DWT as well as good fidelity parameters also.**

Keywords- Speech compression, FFT, DCT, DWT, Huffman encoding.

I. INTRODUCTION

Speech may be defined as the response of the vocal tract to one or more excitation signal. Speech is a very basic way for humans to convey our information with emotion whenever in the same room or all over the world to each other with 4 kHz bandwidth [1, 2]. Speech signal can be store as a digital data for further processing. A huge amount of data is a big issue for storage or transmission. Therefore it is necessary to compress speech signal. A major objective of speech compression is to represent, a speech signal with less or few bits as possible with level of quality [1, 2]. Compression of signal is based on removing non-essential information of the signal and it is also desired that the data is represented by as small as possible number of coefficients so that the quality of reconstructed data is within an acceptable limit [3 4]. Speech coding and compression are extensively used in many applications such as cellular telephony, audio for videophones or video teleconferencing systems and mobile satellite communications. Other applications include the storage of speech for speech synthesis and playback, or for the transmission of voice at a later time such as voice mail systems, voice memo wristwatches, voice logging recorders and interactive PC software [5, 6, 7]. All techniques are based efficient speech compression techniques.

In early stage of research, several data compression techniques were developed. Generally these compression techniques were classified into two categories [8]: dedicated techniques and general techniques. In dedicated techniques, they give minimum distortion such as AZTEC, FAN, CORTES, and CCSP. While in second category, such as differential pulse code modulation, subband coding, and vector quantization come and they have sound mathematical foundation. The detailed discussion of these techniques is presented in [8].

In the last two decades, substantial progress has been made in the field of speech compression. So far, several efficient speech compression techniques have been reported in literature such as Linear Predictive Coding (LPC) [9-12], Waveform coding [13-15] and Subband coding [16-19].

In past, marked researches have made in the many transform methods such as Discrete Cosine Transform (DCT), Fast Fourier Transform (FFT) and Discrete Wavelet Transform (DWT) which are extensively used in several applications such speech processing, image processing and electrocardiogram (ECG) processing. The fast Fourier transform (FFT) is a discrete Fourier transform (DFT) algorithm which reduces the number of computations needed for N points from $2N^2$ to $2N\log N$, where log is the base-2 logarithm. A DFT is used in Fourier analysis of signal. It transforms the signal into frequency domain. DFT decompose a set of values into components of different frequency values. The detailed discussion on FFT is given in [20-23].

The DCT has extensively been used for real-life data compression. Both one-dimensional DCT1 and 2-dimensional DCT2 transforms are used for data compression. For speech compression [24], one-dimensional DCT is exploited. DCT can be calculated using the FFT algorithm as it is closely related to Fourier transform; however, DCT gives real coefficients for real signals. Both DCT1 and DCT2 give more weight to low-pass coefficients than to high-pass coefficients. DCT gives nearly optimal performance in the typical signal having high correlations in adjacent samples [25]. The detailed review on DCT is available in [24, 25].

In the last decade, Wavelet Transform, and more particularly, Discrete Wavelet Transform (DWT) has emerged as powerful and robust tool for analyzing and extracting information from speech signal due to the time varying nature of the human speech production system [2]. Speech signals are non-stationary signals which are characterized by numerous abrupt changes, transitory drifts, and trends. Wavelet has localization feature along with its time-frequency resolution properties which makes it suitable for analyzing non-stationary signals such as speech and ECG signals [26]. Recently, several other methods [27-30] have been developed based on wavelet or wavelet packets for compressing speech signal.

In above context, this paper presents some new results on speech compression using DCT, FFT and DWT. The paper is organized as follow. A brief introduction on speech compression techniques is presented in Section 1. While, Section 2 discusses these transforms such as DCT, FFT and DWT. Section 3 describe methodology based on these transforms. Finally, Section 4 presents the simulation results, followed by concluding remarks in Section 4.

II. OVERWIEV OF TRANSFORM TECHNIQUES

In this paper, three transform such DCT, FFT and DWT is employed for speech compression.

A. Fast Fourier Transform

A signal having periodic function of time can be analyzed or synthesized as a number of harmonically related to sine and cosine signals. A periodic signal $f(t)$ with period T_0 can be represented by Fourier series:

$$f(t) = A_0 + \sum_{n-1}^{\infty} a_n \sin\left(2\Pi nt/T_0\right) + \sum_{n-1}^{\infty} b_n \cos\left(2\Pi nt/T_0\right) \quad (1)$$

where, A_0 is the average, or mean value of signal $f(t)$. While a_n and b_n are the Fourier series coefficients. t and n is the time and coefficient index, respectively.

The above Fourier series coefficients are found by FFT.

$$\begin{aligned} f(t) &= A_0 + \frac{1}{2}\sum_{k-1}^{\infty}(a_n - jb_n)e^{j2\pi n/T_0} \\ &= \sum_{k-1}^{\infty} \alpha_n e^{j2\pi n/T_0} \end{aligned} \quad (2)$$

where, α_n are complex coefficients, and expressed as:

$$\alpha_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} f(t)e^{-j2\pi n/T_0}dt \qquad \text{n}=0,\pm1,\pm2,..... \quad (3)$$

For the sampled periodic speech signal, the discrete-time complex coefficients of the series are:

$$\alpha_n = \frac{1}{N}\sum_{K}^{N-1} f(k)e^{j2\pi kn/N} \quad (4)$$

and
$$f(k) = \sum_{k}^{N-1}\alpha_n e^{-j2\pi kn/N} \quad (5)$$

Where, k is discrete time index. N is the number of speech signal samples. From equations (1), (2) and (4) give the Fourier series coefficients of equation (5) calculated using FFT technique. Eqn. (4) must be performed on each voiced signal segment. Fourier series coefficients used to synthesize the original signal using Eqn. (5).

B. Descrete Cosine Transform

DCT method is widely used for the compression because being a transformation method. The energy of speech signal is concentrated in a few transform coefficients which yields good compression.

Discrete Cosine Transform (DCT) of a 1-D sequence $b[n]$ of length N is

$$B[m] = \left(\frac{2}{N}\right)^{\frac{1}{2}} c_m \sum_{m=0}^{N-1} b[n]\cos\left[\frac{(2n+1)m\pi}{2N}\right] \quad (6)$$

Where, $m = 0, 1…….N\text{-}1.$ (7)

Similarly, the inverse transformation IDCT is defined as

$$b[n] = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{m=0}^{N-1} c_m B[m]\cos\left[\frac{(2n+1)m\pi}{2n}\right] \quad (8)$$

In the both equations (1) and (2), c_m is defined as

$$c_m = \begin{cases} \left(\frac{1}{2}\right)^{\frac{1}{2}} & \text{for } m=0 \\ 1 & \text{for } m \neq 0 \end{cases} \quad (9)$$

The coefficient B [0], which is directly related to the average value of the time-domain block, is termed as the *DC coefficient*, and the remaining coefficients are called *AC coefficients*.

C. Descrete Wavelet Transform

The Discrete Wavelet transform (DWT) is a special case of Wavelet Transform (WT) that provide a compact representation of a signal in time and frequency. Basic principal of wavelet transform is that it decompose the given signal in too many function by using property of translation and dilation of a single function called a mother wavelet, mother wavelet is defined as $\psi(t)$

$$\psi_{a,b}(t) = |a|^{\frac{-1}{2}}\psi\left(\frac{t-b}{a}\right) \quad (10)$$

Where, a is defines the dilation factor applied to the mother wavelet, and b is a translation factor.

Wavelet transform give the multiresolution decomposition of signal [26]. DWT decompose a signal at several n levels in different frequency bands. At each step of DWT decomposition, there are two outputs: scaling coefficients $x^{j+1}(n)$ and the wavelet coefficients $y^{j+1}(n)$. These coefficients are:

$$x^{j+1}(n) = \sum_{i=1}^{2n} h(2n-i)x^j(n) \quad (11)$$

and
$$y^{j+1}(n) = \sum_{i=1}^{2n} g(2n-i)x^j(n) \quad (12)$$

Where, the original signal is represented by $x^0(n)$ and j shows the scaling number. Here $g(n)$ and $h(n)$ represent the low pass and high pass filter, respectively.

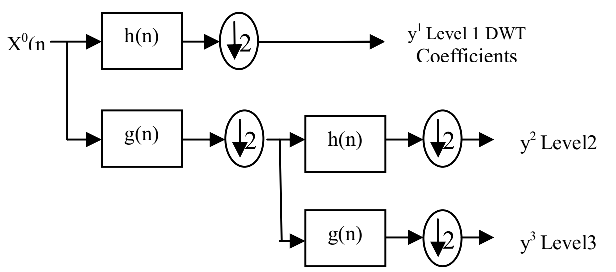

Fig.1 Filter bank representation of DWT decomposition

The output of scaling function is input of next level of decomposition, known as approximation coefficients as shown in Fig. 1.

In order to reconstruct the original signal, at each level of reconstruction, approximation components and the detailed components are up by 2 and the detailed components are up sampled by 2, and then convolved which is shown in Fig. 2

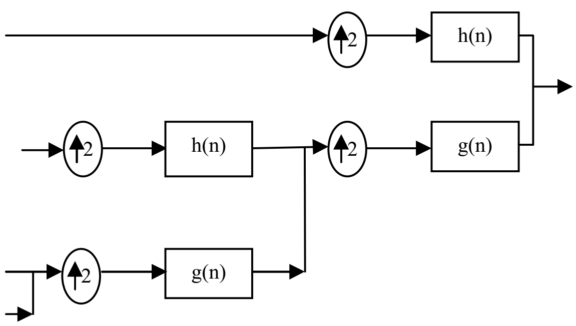

Fig. 2. IDWT filter bank representation

III. Compression Methodology

In this paper we achieved a compression signal by using Transform Method i.e. FFT, DCT and DWT on the speech signal. Compression is achieved neglecting a small coefficient as insignificant data and thus discarding them. The process of compression a speech signal using Transform Method is carried out by using number of stages [1]. The algorithm of speech compression is carried out in three stages: (i) Decomposition, (ii) Threshold & Quantization, (iii) Entropy encoding. In first stage i.e. in decomposition step input speech signal is decomposition in to different resolutions or frequency bands, after decomposition of signal; compression involves truncation of coefficients [1]. Experiments are conducted on a male spoken sentence (i.e. CONFIDENCE). Apply a thresholding (global), which suggested that a fixed percentage of wavelet coefficients should be zero Further, uniform quantizer is applied in these coefficients. The actual compression is achieved at this stage and then compression achieved based on the entropy encoding techniques (Huffman). Finally compressor system gives the compressed data value of speech signal. Fig. 3 shows the method of compression based on the transform techniques.

A. Thresholding

After decomposing of the signal, threshold is applied to coefficients for each level from 1to N [1,7]. So many of the wavelet coefficients are zero or near to zero so due to thresholding near to zero coefficients are equal to zero. By applying a hard thresholding the coefficient below the level is zero so produce a many consecutive zero's which can stored in much less space and transmission speed is up, and in the case of global thresholding the value are set manually, this value are chosen from the range (0…….C_{max}) where C_{max} is maximum coefficient in the decomposition [26,28].

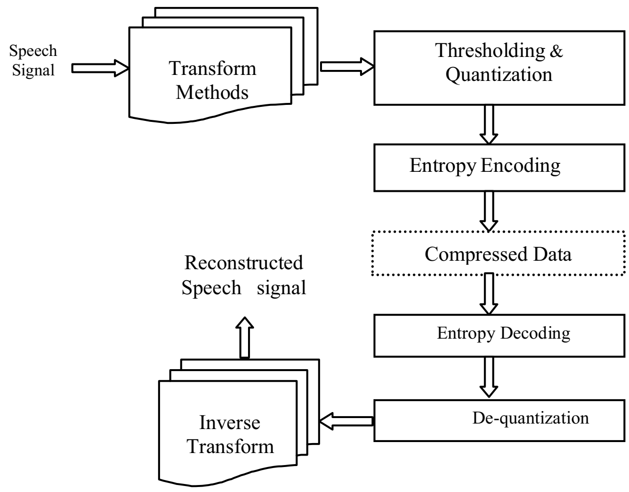

Fig. 3. Block Diagram of Compression methodology

B. Quantization

Aim of this step is to decreases the information which found in the wavelet coefficients in such a way so no error is formed. We quantize the wavelet coefficients using uniform quantization, the computation of step size depend on three parameters [7] are:

(i) Maximum value, M_{max} in the signal matrix
(ii) Minimum value, M_{min} in the signal matrix
(iii) Number of quantization level, L

Once these parameters are found, then step size,

$$\Delta = \left(M_{\max} - M_{\min} \big/ L \right) \tag{12}$$

Then, the input is divided in to L+1 level with equal interval size ranging from M_{min} to M_{max} to plot quantization table. When quantization step is done, then quantization value is fed to the next stage of compression. Three parameters defined above are stored in the file because to create the quantization table during reconstruction step for dequantization [11].

C. Huffman Encoding

In quantization process, the quantized data contains some redundant data, means repeated data, and it was wastage of space. To overcome this problem, Huffman encoding [7] is exploited. In Huffman encoding, the probabilities of occurrence of the symbols in a signal are computed. These symbols indices in the quantization table. Then these symbols are arranged according to the probabilities of occurrence in descending order and build a binary tree and codeword table.

To reconstruct the speech signal, we will reverse the three processes which perform in this paper (Wavelet Transform, Quantization, and Huffman Coding).

IV. Results and Discussion

Speech signal compression is achieved based on transformation techniques using threshold and uniform quantization with Huffman coding. Here performance measured in term of *CR*, *SNR*, *PSNR* and *NRMSE*. Here results are obtained using different transformation techniques such as DCT, FFT, DWT (db10, db6, bior6.8 and coif2, coif5). Table 1 illustrate the performance of compression using transform methods, average compression ratio of FFT are 3.72%, in case of DCT compression ratio is 5.86% and DWT based average compression is 7.47%, 7.49% and 7.45% for the Debauchees (db), Biorthogonal (bior) and Coiflet (coif) wavelets respectively.

It's shown that the higher compression is achieved using DWT decomposition at low cost of fidelity (*SNR*, *PSNR* and *NRMSE*) degradation as compare to the FFT and DCT. Further, as comparison biorthogonal wavelet is give the higher signal compression. These results are shown the Comparative analyses of different transform method in the fig.4 at the different threshold values.

V. Conclusion

Speech Compression is way to representing a Speech signal with minimum data values and favorable in case of storage and transmission. Simulation results included in this paper clearly show the key advantageous features of the wavelet filters over others transforms in the field of speech signal processing. It is found that the wavelet filters significantly improves the reconstruction or fidelity assessments of the compressed speech signal and also yields higher compression ratio as compared to FFT and DCT. Therefore, it is concluded that it can be very effectively used for the speech signal compression.

TABLE I
COMPARISON OF PERFORMANCE BASED ON DIFFERENT TRANSFORM METHOD

Transform Techniques	Signal	CR	SNR	PSNR	NRMSE
FFT	Hello	3.18	10.6668	36.0315	0.1614
DCT	Hello	6.38	11.3312	36.3637	0.1335
Db6	Hello	7.30	21.6832	41.5397	0.0680
Db10	Hello	7.30	22.3895	41.8928	0.0549
Bior6.8	Hello	7.34	18.9128	40.1545	0.0989
Coif2	Hello	7.31	21.3130	41.3546	0.0715
Coif5	Hello	7.31	19.5119	40.4540	0.0855
FFT	Apple	2.67	11.1080	36.5001	0.1597
DCT	Apple	5.60	11.7651	36.8287	0.1308
Db6	Apple	7.83	14.9327	38.4125	0.1609
Db10	Apple	7.88	14.9455	38.4189	0.1535
Bior6.8	Apple	7.89	14.4994	38.1959	0.1776
Coif2	Apple	7.84	14.4293	38.1608	0.1826
Coif5	Apple	7.88	15.7235	38.8079	0.1394
FFT	Confidence	5.31	17.06	42.57	0.0774
DCT	Confidence	5.62	24.07	46.08	0.0180
Db6	Confidence	7.23	22.51	45.31	0.0841
Db10	Confidence	7.30	22.63	45.36	0.0834
Bior6.8	Confidence	7.24	22.34	45.21	0.0895
Coif2	Confidence	7.18	22.49	45.29	0.0874
Coif5	Confidence	7.22	22.37	45.23	0.0888

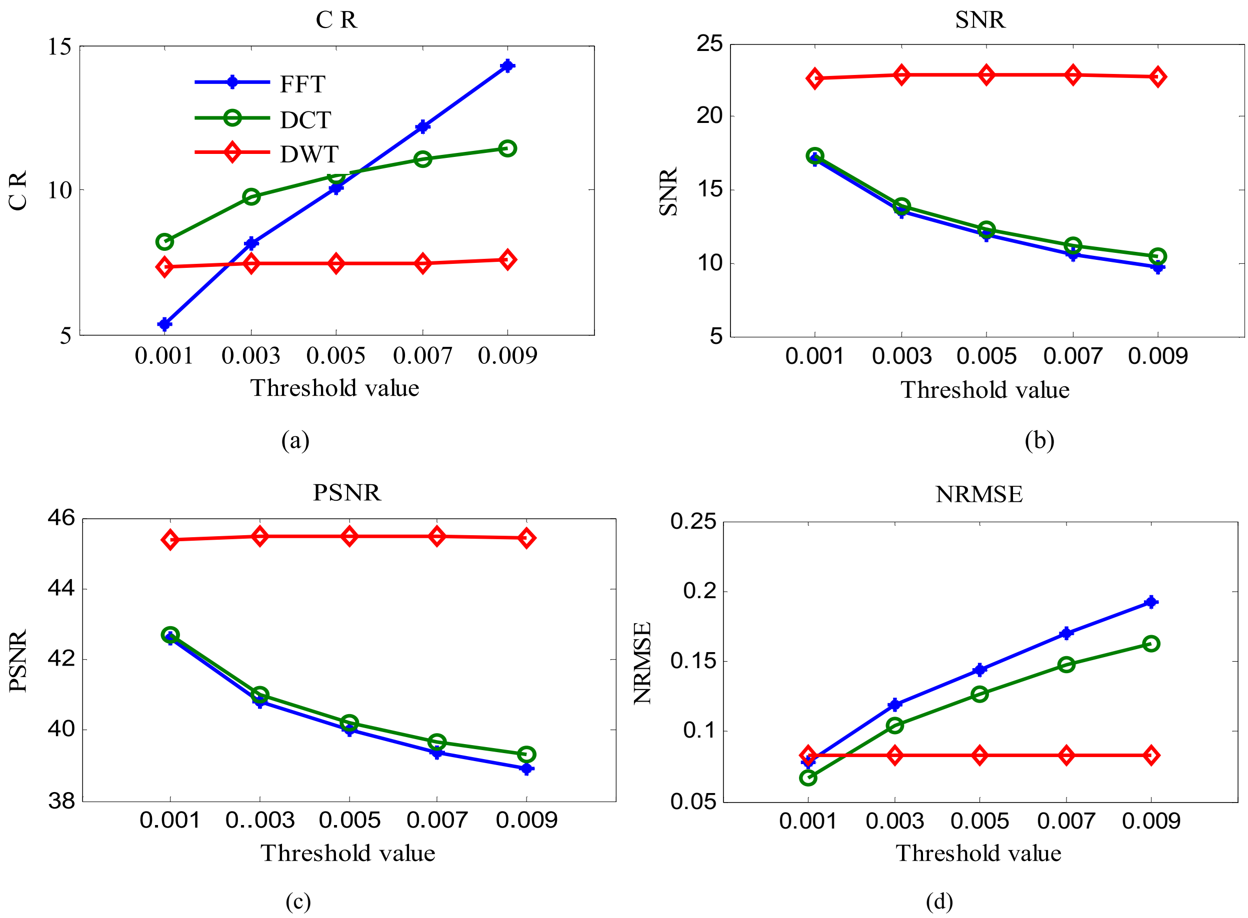

Fig.4. A comparative analysis of the performance in different transforms (DCT, FFT and DWT)
(a) *C R* (b) *SNR* (C) *PSNR* (d) *NRSME*

REFERENCES

[1] H. Elaydi, "Speech Compression Using Wavelet", "site.iugaza.edu.ps/helaydi/files/2010/02/Elaydi.pdf."

[2] W. Chong and J. Kim, "Speech and Image Compression by DCT, Wavelet, and Wavelet Packet", Proceeding in Information, Communication and Signal processing IEEE, Vol.3, pp.1353-1357, Aug 2002.

[3] W. Kinsner and A. Langi, "Speech and Image Signal Compression with Wavelet", IEEE Wescanex conference Proceedings, New York, N Y, pp. 368-37, 1993.

[4] N. Junejo, N. Ahmed, M. Ali Unar, and A. Q. K Rajeput, "Speech and Image Compression Using Discrete Wavelet Transform", Advances in Wired and Wireless Communication, IEEE/Sarnoff Symposium, pp.45-48, 09 May 2005.

[5] F. W. Zaki , H. Hashish and S. H. behir, "Speech Compression Using wavelet Transform", Proceedings of the Eighteenth National Radio Science Conference, 2001, Vol. 2, pp 467-474, March 27-29, 2001.

[6] N. M. Hosny S. H. El-Ramly and M. H. El- Said, "Noval Techniques For Speech Compression Using Wavelet Transform", Eleventh Int. Conf. on Microelectronics (ICM99), pp.225-229, 2002.

[7] Abdul Mawla M. A. najih, A. R. bin Ramli, V. Prakash, and Syed A. R., "Speech Compression Using Discreet Wavelet Transform", Proceedings of 4th National Conf. on Telecommunication Technology IEEE, pp.1-4, March 2003.

[8] R. S. H. Istepanian, A. Sungoor, and J. C. Nebel ," Linear Predictive Coding and Wavelet Decomposition for Robust Microarray Data Clustering" , pp.4629-4632, Oct. 2007, IEEE.

[9] M.R. Madane, Z. Shah, R. Shah, and S. Thakur, "Speech Compression Using Linear predictive Coding", Proceeding of the international workshop on machine Intelligence Research (MIR Day, GHRCE- Nagpur) © 2009 MIR Labs.

[10] P. Venkateswaran, A. sanyal, S. Das, R. Nandi, S. K. Sanyal, " An Efficient Time Domain Speech Compression Algorithm Based On LPC and Sub-Band Coding Techniques", Journal of Communication, vol.4, No.6, july2009.

[11] H. M. Magboun, N. Ali, M. A. Osman, S. A. Alfandi, "Multimedia Speech Compression Techniques", 3rd IEEE International Conference on Computer Science and Information (ICCSIT), pp. 498-502, 2002.

[12] I. S. Burnett, and G. J. Bradley, "New Techniques for Multi-Prototype Waveform Coding at 2.84 kb/s." International Conference on Acoustics, Speech, and Signal Processing, 1995, ICASSP-95, Vol. 1, pp.261-264.

[13] E. Shlomot, V. Cuperman and A. Gersho, "Combined Harmonic and Waveform Coding of Speech At low Bit Rates", International Conference on Acoustics, Speech, and Signal Processing, ICASSP-98, Vol 2, pp. 585-588, Aug 2002.

[14] W. Kleijn and K. Paliwal, eds., "Speech Coding and Synthesis", Amsterdam: Elsevier Science Publishers, 1995.

[15] H. S. Malvar, "Lapped transform for Efficient Transform/Subband Coding" IEEE Transaction on acoustics, speech and audio processing", Vol. 38, No. 6, June 1990.

[16] J. Kovacevic, "Subband Coding System Incorporating Quantizer Models", IEEE Transactions on image processing, vol. 4, No. 5, may 1995.

[17] K. Ramchandran, M. Vetterli, and C. Herley, "Wavelet, Subband Coding, and Best Bases", Proceeding of the IEEE, Vol. 84, No. 4, April 1996.

[18] J. McAuley, Ji Ming, D. Stewart, and P. Hanna, "Subband Correlation and Robust Speech Recognition" ,IEEE Transactions on speech and audio procession, Vol. 13, No. 5, September 2005

[19] S. Mallat, "A Wavelet Tour of Signal Processing ", Academic Press, San Diego, 1998.

[20] G. R. Redinbo, "Fault-Tolerant FFT Data Compression", Proceeding on Pacific Rim International Symposium on IEEE, Vol.29, pp. 2095-2105, Dec 2001

[21] N Al-Hinai, K. Neville, A. Z. Sadik and Z. M. Hussain, "Compressed Image Transmission over FFT-OFDM: A Comparative Study", Proceeding on Telecommunication Networks and Application Conference (ATNAC 2007) IEEE.pp.465-469, Nov. 2008.

[22] M. S. Alam and N. M. S. Rahim, "Compression of ECG Signal Based on Its Deviation from a Reference Signal Using Discrete Cosine Transform", proceeding on Electrical and Computer Engineering (ICECE 2008) IEEE, pp.53-58, January 2009.

[23] G. R. Redinbo and R. Manomohan, "Failt Tolerance in Computing, Compressing, and Transmitting FFT Data", Transaction on Communication in IEEE, pp. 2095-2105, Aug. 2002.

[24] Y. Yusong, W. Chunmei, Su Guangda, S. Qingyun "Invertible Integer DCT Applied on Progressive until Lossless Image Compression", Proceeding of the 3rd International Symposium on Image and Signal Processing and Analysis (ISPA 2003), IEEE, vol. 2, pp. 1018-1023, May 2004.

[25] C. Souano, M. Atri, M. Abid, K. Torki and R. Tourki, "Design of New Optimized Architecture Processor for DWT", Vol. 6, pp. 297-312, Aug. 2000.

[26] J. Karam, "End Point Detection For wavelet Based Speech Compression", world Academy of science, Engineering technology 37 2008.

[27] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.11, No.7, July 1989.

[28] Karam, "Various Speech Processing Techniques for Speech Compression and Reconignition", world Academy of science, Engineering ans technology 37 2008.

[29] J. Pang, S. Chauhan, and J. M. Bhlodia, "Speech Compression FPGA Design by Different Discrete Wavelet Transform Schemes", Proceeding in Advances in Electrical and Electronics Engineering (WCECS '08) IEEE, pp.21-29, Sept. 2009.

[30] E. B.Fgee, W. J. Phillips and W. Robertson, "Comparing Audio Compression Using Wavelet with Other Audio Compression Schemes", IEEE Transaction Electrical and Computer Engineering IEEE, Vol.2, pp.698-701, Aug 2002.