Comparison of WT Based Speech Compression Techniques Using VC++

K.Arun Kumar[1] and M. Vinod Kumar Reddy[2]
[1]CVR College of Engineering/ECE Department, Hyderabad, India
Email: arun.katkoori@gmail.com
[2]CVR College of Engineering/ECE Department, Hyderabad, India
Email: vinodreddy.488@gmail.com

Abstract : The main purpose of this paper is to compresses the speech signal using wavelet transform. Psychoacoustics is the scientific study of sound perception. From the psychoacoustic point of view, we have selected Wavelet analysis for the digital speech compression. Also Wavelet Transform eliminates the irrelevancies and redundancies present in the speech signal. The two popoular models of Wavelet Transform for speech compression are-Filter bank model and Lifting Scheme.Filter Bank model is also called Subband Filtering model. Both models decompose the speech signal into approximate and detailed components. But the lifting scheme is fast compared to the filtering model. We have tested some of the lifting scheme WT algorithms like Haar, Daubechies series, Cohen-Daubechies and Cohen-Daubechies-Feauveau bidirectionnel wavelet and have implemented them.

Keywords- Wavelet Transform, Psycoacoustics, Speech compression, Haar wavelet, Daubechies series, Cohen-Daubechies-Feauveau wavelets.

I. INTRODUCTION

The generation of human speech is very complicated. It involves lungs, vocal tracts and vocal folds. The message is formulated in the brain and is converted into an acoustic signal by vocal tract system. The vocal tract system is similar to the electronic parts like power supply, oscillator and resonator.[2] The structure of speech generation contains lungs, ribcage and abdominal muscles. Controlled airstream is produced between the vocal folds by this combination. From the lungs, the chest cavity expands and contracts to force air. By opening and closing of glottis the resistance of the air is changed. Finally a sound is produced from of the vibration of vocal cords.[2]

The human speech generation system is shown in Figure 1.
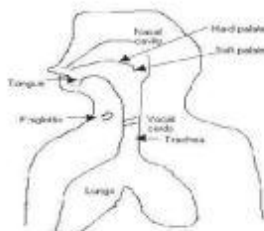

Figure 1 Generation of Human speech

The generation of human speech is analogous to the electronic circuit as shown in below figure.
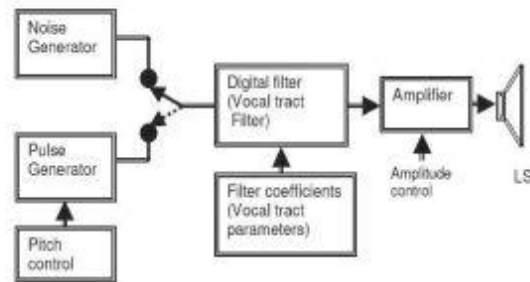

Figure 2 Electronic structure of speech generation

For efficient transmission and storage, we compress the signal. If the signal is uncompressed, it requires large memory. Consider the telephone level speech signal in the range of 300 Hz to 3400 Hz coded on at 8 bits per sample, and sampling rate of 8000 samples per second is defined as an uncompressed speech signal. [5]. This non-compressed signal requires storage capacity of 28.8Mbytes for one-hour duration and transmission rate of 64Kbps. The main idea of speech compression algorithm is to represent the non-compressed speech with less number of bits and optimum speech quality. Wavelets are used for the speech compression.

II. SPEECH COMPRESSION MODELS

With modern telecommunication, speech compression plays an important role. Speech compression is also speech coding. It is the process of indicating digital speech signal with a small number of bits with the normalized speech quality and low computational complexity. [1] Also these techniques will remove the unimportant components from the original speech.

The main requirements of the speech compression algorithms are

1) High compression ratio- The compression ratio gives memory size and transmission rate required after compression.

2) Low computational complexity- For real-time encoding and decoding, to minimize the power and delay coding, the computational complexity should be minimum.

3) Limited storage requirements- Generally the speech coding algorithms use speech buffers between intermediate stages. So the size of the buffer is low.

4) Floating-point Multiplications- The algorithms require many numbers of floating-point multiplications. We can replace floating-point operations with integer operations.

5) Immune to noise- Sometimes noise may distort the speech signal. Speech compression technique should control noise.

6) Minimum audible distortion-At a given bit rate, the speech compression algorithm will provide minimum audible distortion.

A .Waveform Coders

Waveform coders are used to represent the speech signal without underlying the speech model. They use scalar and vector quantization techniques to represent the samples of speech signal. [5] Robust than parametric coders but have the disadvantage of operated at higher data rates than vocoders.[3] Some of wave-coders are discussed below.

i) 64Kbps G.711 PCM, DPCM and DM

It uses scalar quantization model. It is simple, but costly in terms of data-rates. There are two types of PCMs-Uniform and Non-uniform. In uniform PCM, the step-size is fixed irrespective of input speech signal, whereas in non-uniform PCM, the step size is changed. If the speech is coded at 64Kbps using μ-law and A-law called 'uncompressed' standard (G7.11 standard). This format is used as a reference for different speech compression techniques.

In DPCM, the difference between original and estimated samples is quantized and transmitted. In DM, the error sample is given to the comparator. If error is positive, output of the comparator is positive pulse else if error is negative, output of the comparator is negative pulse. For data rates at and below 32Kbps, DM and DPCM have better performance than PCM.

ii) 32Kbps G.721 CCITT ADPC,

It also uses scalar quantization model. It provides better quality and less memory compared to original PCM.[7] The coding of speech is done for 4-bits per sample instead of 8-bits per sample.

iii) Vector Quantizer

In this, the sample space is converted into symbol space with quantization error and low bit rate. This method is too difficult to implement, but is used in today's fast digital signal processors.

B. Sub-band and Transform Coders

The Sub-band coder is a waveform coder's method in f-domain. [8] At Txr, The speech is passed through an analysis filter bank and by down-sampling method; the bandwidth of each sub band of the input speech signal is reduced. At the Rxr, by up-sampling and synthesis filter bank, the original speech is retrieved. [9] The structure of the sub band coder is shown in figure 3.
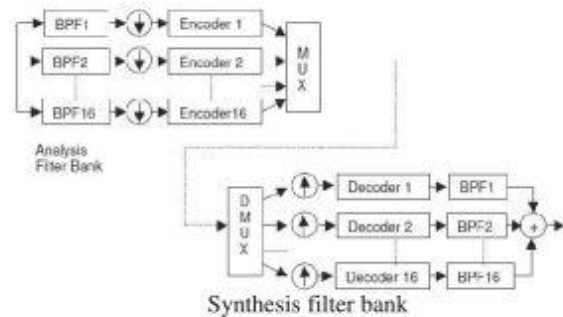

Figure 3 Sub band coder structure

Transform coder in other hand, a waveform coder in transform domain. At Txr, by using any of these transform techniques, DCT,DFT,WHT, WT, the speech signal is quantized and encoded.[4] At Rxr, it is decoded and inverse transformed. The structure of transforming coder is shown below.
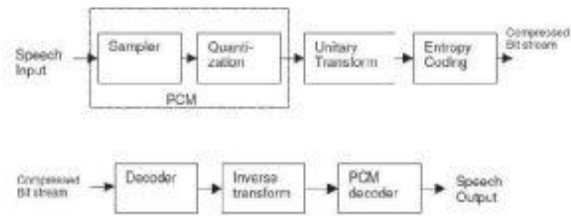

Figure 4 Transform coder structure

In this coder, the input is processed frame by frame and each frame is processed by 'transform' technique. If s is speech signal and N is frame length then S is speech after compression, represented in matrix form as

$$S = T.s \qquad (1)$$

$$
\begin{bmatrix}
S(0) \\
S(1) \\
S(2) \\
S(3) \\
\cdot \\
\cdot \\
S(N-1)
\end{bmatrix}
=
\begin{bmatrix}
t_{1,1} & t_{1,2} & \cdots & t_{1,N} \\
t_{2,1} & t_{2,2} & \cdots & t_{2,N} \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
t_{N,1} & t_{N,2} & \cdots & t_{N,N}
\end{bmatrix}
\begin{bmatrix}
s(0) \\
s(1) \\
s(2) \\
\cdot \\
\cdot \\
\cdot \\
s(N-1)
\end{bmatrix}
$$

(2) Equation 2 is called analysis equation. And s= T⁻¹S is called synthesis equation.

III. WAVELET TRANSFORM

Wavelet is a mathematical function, used to represent data. Another definition is – wavelet is 'small wave' having its energy concentrated in the form of time. Any function can be represented in terms of wavelets similar to Fourier series. [1] It also provides time and frequency analysis simultaneously.

Let f(t) is real-time signal and Continuous Wavelet Transform of f(t) is given by

$$W(a,b) = \int_{-\infty}^{\infty} f(t)\psi *_{a,b}(t)dt \qquad (3)$$

Where 'a' & 'b' are variables, $\varphi_{a,b}(t)$ is wavelet function and is given by

$$\psi_{a,b} = \frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a}) \qquad (4)$$

The variable 'a' represents the amount of scaling or dilation of function and is also called dilation variable. [6] The variable 'b' represents a time shift or translation of the function and is also called translational variable.

The inverse wavelet transform is given by

$$f(t) = \frac{1}{|a|^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(a,b)\psi(a,b)dadb \qquad (5)$$

The block diagram of speech compression using wavelet is shown below. Firstly, the input is decompressed using wavelet transform and then small wavelets are removed by thresholding operator. [4] (Sometimes the small wavelets represent noise) The coefficients obtained after the thresholding are encoded. During the reconstruction, the compressed speech is decompressed using inverse wavelet transform.
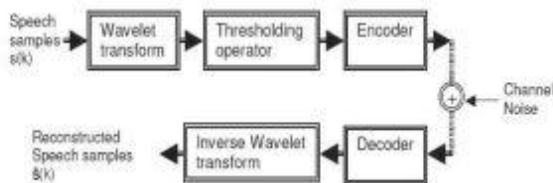

Figure 5 Wavelet Speech Compression Block Diagram

A. Lifting based Algorithms

We used some of the lifting based algorithms. Those are given by

1. Haar Wavelet algorithm

In this algorithm, initially the input samples are classified into two frames: Even and Odd frames. To get the A-coefficients, take average of even and odd samples and then to get the D-coefficients, subtract odd samples from even. This is explained in figure 6.
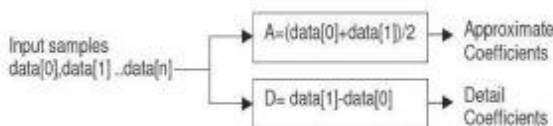

Figure 6 Haar Wavelet algorithm

Inverse transform is obtained by

data[0] = data[0]-data[1]/2
data[1] = data[1] + data[0] \qquad (6)

2. CDF Wavelet algorithm

It can be done with two methods- using lifting technique and using a filter bank technique. But lifting technique is faster than filter bank method. CDF wavelet is represented by CDF(n,ñ).where n represents primary wavelet's vanishing moments and n˜ represents dual wavelet's vanishing moments.

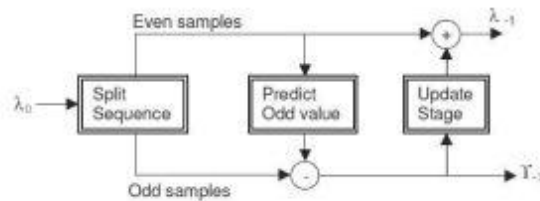CDF with lifting technique is explained in following figure 8.
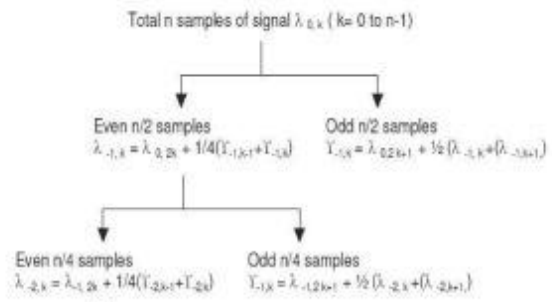

Figure 7 Steps in Lifting Technique

Figure 8 CDF Lifting Technique

Compared to Haar wavelet, CDF gives more compact representation of speech.

3. D-Wavelet algorithm

D stands for Daubechies. This algorithm provides a good smoothness and good support of compactness compared to the other algorithms. There are 4 stages of Lifting Daubechies-4.

a) Split stage:

$\lambda_{-1,k} \leftarrow \lambda_{0,2k}$ (Even samples) and $\Upsilon_{-1,k} \leftarrow \lambda_{0,2k+1}$ (Odd samples)

b) Update stage:

$$\lambda_{-1,k} \leftarrow \lambda_{-1,k} + \sqrt{3}\Upsilon_{-1,k} \qquad (7)$$

c) Predict stage:

$$\Upsilon_{-1,k} \leftarrow \Upsilon_{-1,k} - \frac{\sqrt{3}}{4}\lambda_{-1,k} - \frac{\sqrt{3}-2}{4}\lambda_{-1,k-1} \qquad (8)$$

d) Normalization stage:

$$\lambda_{-1,k} \leftarrow \frac{\sqrt{3}-1}{\sqrt{2}}\,\lambda_{-1,k} \quad \text{and} \quad \Upsilon_{-1,k} \leftarrow \frac{\sqrt{3}+1}{\sqrt{2}}\,\Upsilon_{-1,k}$$

$$(9)$$

IV. TEST SETUP FOR VC++

Microsoft Visual C++ is abbreviated as VC++ or MSVC is an Integrated Development Environment from Microsoft. It is proprietary software and used for C, C++ programming languages. It is used in speech processing, image and video processing.

Test set up is prepared in Visual C++ for speech compression using WT transform. This test setup provides following facilities for the analysis compared to Matlab.

- Selection of wavelet.
- Playing speech file.
- Select the speech file.
- Selection of threshold value.
- Selection of block length.
- Variation of threshold for lossy compression.
- Displaying speech waveform.
- Displaying original file size, compressed file size and compression ratio.
- Also displaying the parameters like signal to noise ratio, mean square error, and root mean square error for comparison between original and reconstructed speech file.
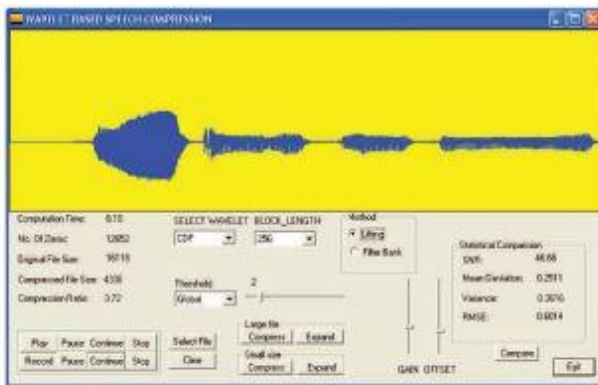

Figure 9 Test set up

V. RESULTS

We have implemented the wavelet compression techniques using vision C++ software. Here, we compare the compression ratio and quality of speech for various types of wavelet transform algorithms like Haar, CDF, cubic lifting method, Daub-4, Daub-8, Daub-16. We have also measured the statistical parameters of speech quality. They are-MSE, RMSE, SNR, and mean deviation.

For the Haar transform, the compression ratio and speech quality is shown in the following figure. Simple algorithm and compression ratio of 3.31 is achieved.
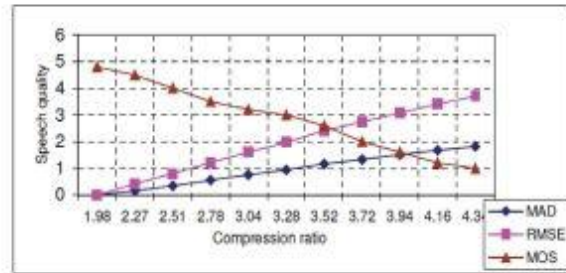

Figure 10 Compression ratio vs speech quality for Haar

For the CDF transform, the compression ratio and speech quality is shown in figure below. Compression ratio of 4.5 is achieved. [10] If the threshold value is increased, we can't reconstruct the original speech signal.
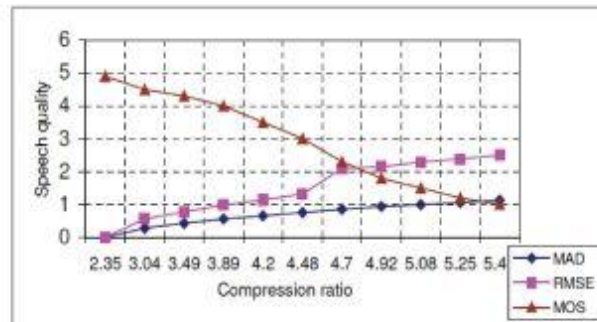

Figure 11 Compression ratio vs Speech quality for CDF

For cubic prediction lifting method, the compression ratio and speech quality is shown below. Here we achieve the compression ratio of 5.34.Compared to the above two algorithms, this algorithm got the loss-less compression.
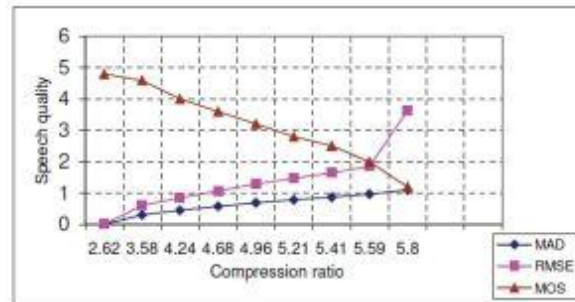

Figure 12 Compression ratio vs Speech quality for Cubic prediction

For Daub-4 wavelet algorithm, compression ratio and speech quality is given below. The compression ratio of 3.9 is achieved.
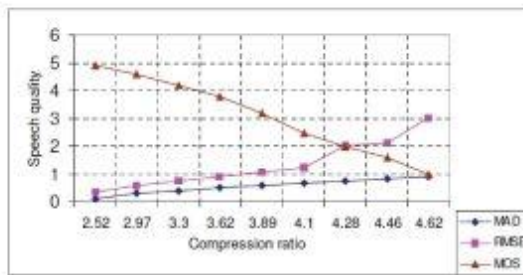
Figure 13 Compression ratio vs Speech quality for Daub-4

For the Daub-8 wavelet algorithm, compression ratio and speech quality is given below. The compression ratio of 4.2 is achieved.
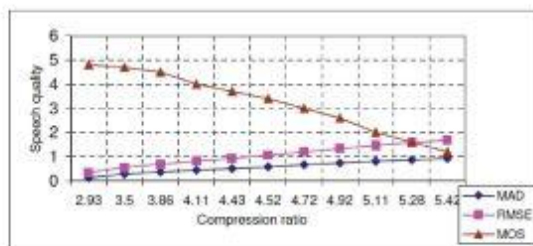

Figure14 Compression ratio vs Speech quality for Daub-8

For Daub-16, compression ratio of 4.61 is achieved. Compression ratio and speech quality of this wavelet is given by
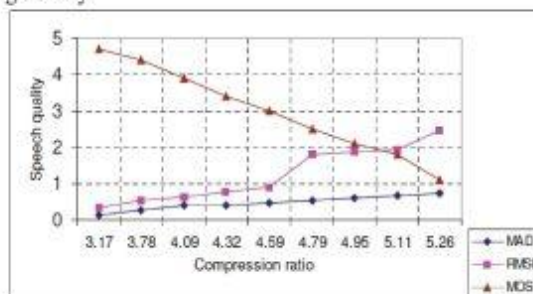

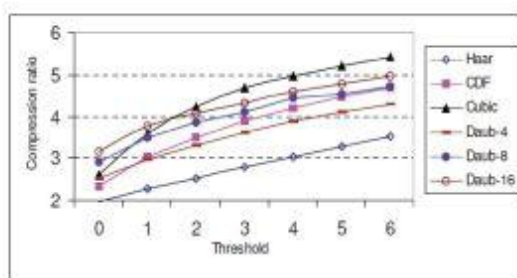Figure 15 Compression ratio vs Speech quality for Daub-16

Figure 16 Compression ratio of various wavelet algorithms

Compression ratios of different wavelet algorithms like Haar, CDF, Cubic, Daub-4, 8, and 16 are compared in the figure above.

VI. CONCLUSIONS

It is concluded that wavelet based algorithms are the best for speech compression. We tested and implemented some of the lifting scheme algorithms like Haar, CDF and Daub algorithms. They gave best compression ratio and low computational complexity. The frame size of each wavelet is considered as 256 bytes. VC++ software is used. It selects the wavelet, plays the speech file, then applies threshold and finally displays the output speech waveform with parameters like SNR, MAD, MSE, and RMSE. The lifting scheme of wavelet transform is very good from the point of view of computational complexity and loss-less compression. Some algorithms gave lossy-compression and some gave loss-less compression.

REFERENCES

[1] A. Verbuch, B. Gutman "Speech compression using wavelet packet transform & vector quantisation", SPIE Proceedings Vol. 2569: 2014.

[2] Pramila shrinivasan & Leah H. Jamieson "high quality audio compression using adaptive wavelet packet decomposition and psycho-acoustic modelling" ,IEEE transactions on signal processing, vol. 46, no. 4, april 2010

[3] P.S. Sathidevi & Y. Venkataramani "Applying wavelet analysis for coding of speech & audio signals for multimedia applications" TEXTBOOK.

[4] Amara Graps, IEEE Computational Science and Engineering, 2011 vol. 2 No.2 "An introduction to wavelets".

[5] Howard L. Resinkoff & Raymond o. Wells "Wavelet analysis : the scalable structure of information"

[6] Andreas S. Spanias "Speech coding: A tutorial review", Proceedings of the IEEE, Vol. 82, No.10, October 2014.

[7] N. Benevuto et al, "The 32Kb/s coding standard," AT&T Technical Journal, Vol. 65(5), pp. 12-22, Sept.-Oct. 2006

[8] R. Crochiere, S. Webber, and J. Flanagan, "Digital Coding of Speech in Sub-bands" The Bell Tech. J., Vol. 55(8), p. 1069, Oct. 2006.

[9] Langlias, Masson, Montagna "Real-time implementation of 16 kbps subband coder with vector quantization" Proceedings of EUSIPCO-86, Signal processing-III, part-1, page 419-422.
     http://www.wavelet.org

[11] P.P. Vaidyanathan, Proceedings of the IEEE, Vol.41, No. 2, 1993 "Multi-rate digital filters, filter banks, polyphase networks and applications: A tutorial review".

[12] Raghuveer Rao & Bopardikar. "Wavelet transform Introduction to theory & applications", Addison-Wesley, Pearson Education.