OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Speech Compression Using Wavelets

P. srinivasa rao[1], G.krishnaveni[2], G. Prasanna kumar [3], G.satyanandam[4], CH.parimala[5], K.ramteja[6]

[1](Department of Electronics & Communication Engineering, Asst.professor, St.Ann's college of engineering & technology, India)
[2, 3, 4, 5, 6](Department of Electronics & Communication Engineering ,Student, St.Ann's college of engineering & technology, India)

Abstract: *In the recent years, large scale information transfer by remote computing and the development of massive storage and retrieval systems have witnessed a tremendous growth. To cope up with the growth in the size of databases, additional storage devices need to be installed and the modems and multiplexers have to be continuously upgraded in order to permit large amounts 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. In this we investigated for optimum wavelet, optimum level, and optimum scaling factor.*

I. Introduction

Speech Compression is a method to convert human speech into an encoded form in such a way that it can later be decoded to get back the original signal .Compression is basically to remove redundancy between neighboring samples and between adjacent cycles. Major objective of speech compression is to represent signal with lesser number of bits. The reduction of data should be done in such a way that there is acceptable loss of quality.

II. Compression

Compression is a process of converting an input data stream into another data stream that has a smaller size. Compression is possible only because data is normally represented in the computer in a format that is longer than necessary i.e. the input data has some amount of redundancy associated with it. The main objective of compression systems is to eliminate this redundancy. When compression is used to reduce storage requirements, overall program execution time may be reduced. This is because reduction in storage will result in the reduction of disc access attempts. With respect to transmission of data, the data rate is reduced at the source by the compressor (coder) ,it is then passed through the communication channel and returned to the original rate by the expander(decoder) at the receiving end. The compression algorithms help to reduce the bandwidth requirements and also provide a level of security for the data being transmitted. A tandem pair of coder and decoder is usually referred to as codec.

2.1 Types of compression

There are mainly two types of compression techniques - Lossless Compression and Lousy Compression.

2.1.1 Lossless compression

It is a class of data compression algorithm that allows the exact original data to be reconstructed from the exact original data to be reconstructed from the compressed data. It is mainly used in cases where it is important that the original signal and the decompressed signal are almost same or identical. Examples of lossless compression are Huffman coding.

2.1.2 Lousy compression

It is a data encoding method that compresses data by removing some of them. The aim of this technique is to minimize the amount of data that has to be transmitted. They are mostly used for multimedia data compression. The rest of the paper is organized as follow; section 2 gives the Theoretical background about the speech compression schemes. The speech compression techniques are described in section 3& Section 4 evaluates the performance of the proposed technique followed by the conclusion.

III. Techniques for speech compression

Speech compression is classified into three categories,

3.1 Waveform coding

The signal that is transmitted as input is tried to be reproduced at the output which would be very similar to the original signal.

3.2 Parametric coding

In this type of coding the signals are represented in the form of small parameters which describes the signals very accurately. In parametric extraction method a preprocessor is used to extract some features that can be later used to extract the original signal.

3.3 Transform coding

This is the coding technique that we have used for our paper. In this method the signal is transformed into frequency domain and then only dominant feature of signal is maintained. In transform method we have used discrete wavelet transform technique and discrete cosine transform technique. When we use wavelet transform technique, the original signal can be represented in terms of wavelet expansion.

Similarly in case of DCT transform speech can be represented in terms of DCT coefficients. Transform techniques do not compress the signal, they provide information about the signal and using various encoding techniques compressions of signal is done. Speech compression is done by neglecting small and lesser important coefficients and data and discarding them and then using quantization and encoding techniques.
Speech compression is performed in the following steps.

1. Transform technique
2. Thresholding of transformed coefficients
3. Quantization
4. Encoding

3.3.1 Transform technique

DCT and DWT methods are used on speech signal. Using DCT, reconstruction of signal can be done very accurately; this property of DCT is used for data compression. Localization feature of wavelet along with time frequency resolution property makes DWT very suitable for speech compression. The main idea behind signal compression using wavelets is linked primarily to the relative scarceness of the wavelet domain representation of signal.

A) Continuous wavelet transforms (CWT)

This chapter provides a motivation towards the study of wavelets as a tool for signal processing. The drawbacks inherent in the Fourier methods are overcome with wavelets. This fact is demonstrated here.
It must be reiterated that the discussion in this chapter is by no means comprehensive and exhaustive. The concepts of time-frequency resolution have been avoided for the sake of simplicity. Instead, the development endeavors to compare the Wavelet methods with the Fourier methods as the reader is expected to be well conversant with the latter.

Consider the following figure which juxtaposes a sinusoid and a wavelet

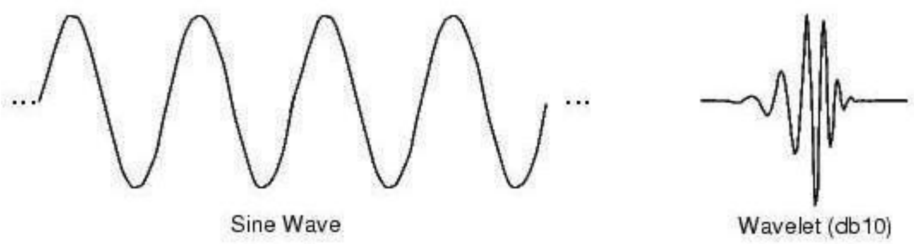

Fig 3.1: comparing sine wave and a wavelet

As has already been pointed out, wavelet is a waveform of effectively limited duration
That has an average value of zero. Compare wavelets with sine waves, which are the basis of Fourier analysis. Sinusoids do not have limited duration -- they extend from minus to plus infinity. And where sinusoids are smooth and predictable, wavelets tend to be irregular and asymmetric.

Fourier analysis consists of breaking up a signal into sine waves of various Frequencies. Similarly, wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet.

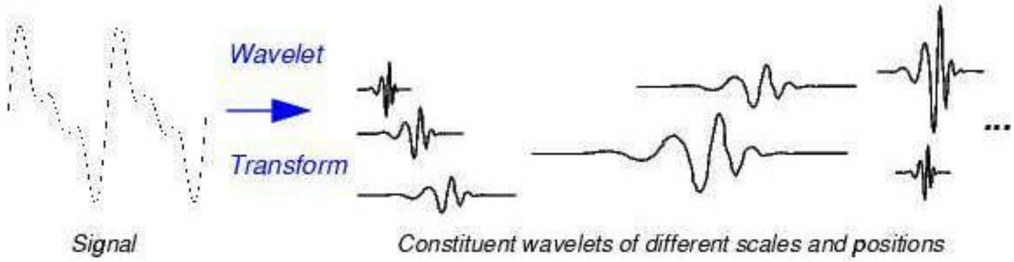

Fig3.2: constituent wavelets of different scales and positions

The above diagram suggests the existence of a synthesis equation to represent the original signal as a linear combination of wavelets which are the basis function for wavelet analysis (recollect that in Fourier analysis, the basic functions are sines and cosines). This is indeed the case. The wavelets in the synthesis equation are multiplied by scalars. To obtain these scalars, we need an analysis equation, just as in the Fourier case. We thus have two equations, the analysis and the synthesis equation. They are stated as follows:

1. Analysis equation or CWT equation:

$$C(a,b) = \int_{-\infty}^{\infty} f(t).\frac{1}{\sqrt{|a|}}\Psi*\left(\frac{t-b}{a}\right)d(t) \ldots\ldots\ldots (3.1)$$

2. Synthesis equation or ICWT:

$$f(t) = \frac{1}{K}\int_{a=-\infty}^{\infty}\int_{b=-\infty}^{\infty}\frac{1}{|a|^2}C(a,b)\frac{1}{\sqrt{|a|}}\Psi\left(\frac{\mathrm{t-b}}{\mathrm{a}}\right).d(a).d(b) \ldots\ldots\ldots\ldots (3.2)$$

B) Continuous-time Wavelet

Consider a real or complex-valued continuous-time function y(t) with the following Properties:

1. The function integrates to zero

$$\int_{-\infty}^{\infty}\Psi(t).d(t) = 0 \ldots\ldots\ldots\ldots (3.3)$$

2. It is square integrable or, equivalently, has finite energy

$$\int_{-\infty}^{\infty}|\Psi(t)|^2.d(t) < \infty \ldots\ldots\ldots\ldots (3.4)$$

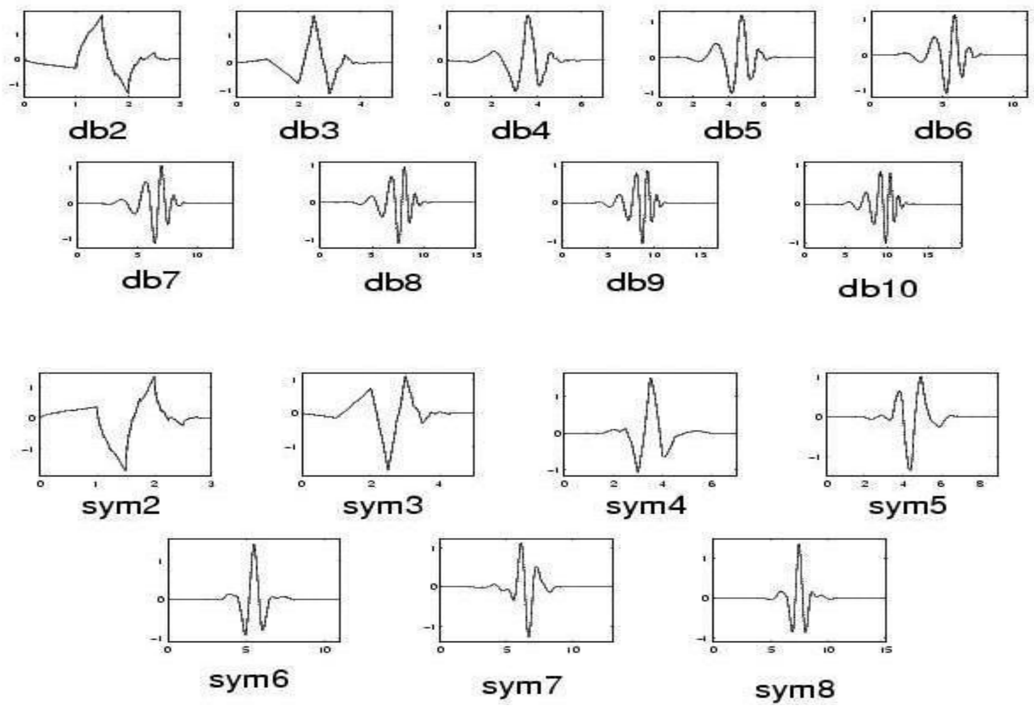

Fig3.3: some wavelet functions

C) Discrete wavelet transforms (DWT)

A discrete wavelet transform can be defined as a „small wave" that has its energy concentrated in time, and it provides a tool for the analysis of transient, non-stationary or time varying phenomenon. It has oscillating wave like property. Wavelet is a waveform of limited duration having an average value zero. They are localized in space. Wavelet transform provides a time-frequency representation of the signal. In DWT, the signal is decomposed into set of basic functions also known as „WAVELETS". Wavelets are obtained from a single MOTHER WAVELET by delay and shift in.

$$\Psi(t) = \frac{1}{\sqrt{a}} \Psi \frac{(t-b)}{a} \ldots\ldots\ldots\ldots (3.5)$$

Where "a" is the scaling parameter and „b" is the shifting parameter.DWT uses multi resolution technique to analyze different frequencies. In DWT, the prominent information in the signal appears in the lower amplitudes. Thus compression can be achieved by discarding the low amplitude signals.

D) Discrete cosine transforms (DCT)

Discrete Cosine Transform can be used for speech compression because of high correlation in adjacent coefficient. We can reconstruct a sequence very accurately from very few DCT coefficients. This property of DCT helps in effective reduction of data.

DCT of 1-D sequence x (n) of length N is given by

$$X(m) = [\ \frac{2}{N}\ ]^{1/2} C_m \sum_{m=0}^{N-1} X(n) \cos[\frac{(2n+1)m\pi}{2N}] \ldots\ldots\ldots\ldots\ldots\ldots..(3.6)$$

Where m=0, 1, - - - - - -, N-1

The inverse discrete cosine transform is

$$X(n) = [\frac{2}{N}]^{1/2} \sum_{m=0}^{N-1} Cm\, X(m) \cos[\frac{(2n+1)m\pi}{2N}] \ldots\ldots\ldots\ldots\ldots\ldots..(3.7)$$

In both equations Cm can be defined as

Cm= (1/2)1/2 for m=0.

=1form≠0

3.3.2 Thresholding

After the coefficients are received from different transforms, thresholding is done. Very few DCT coefficients represent 99% of signal energy; hence Thresholding is calculated and applied to the coefficients. Coefficients having values less than threshold values are removed.

3.3.3 Quantization

It is a process of mapping a set of continuous valued data to a set of discrete valued data. The aim of quantization is to reduce the information found in threshold coefficients. This process makes sure that it produces minimum errors. We basically perform uniform quantization process.

3.3.4 Encoding

We use different encoding techniques like Run Length Encoding and Huffman Encoding. Encoding method is used to remove data that are repetitively occurring. In encoding we can also reduce the number of coefficients by removing the redundant data. Encoding can use any of the two compression techniques, lossless or lossy. This helps in reducing the bandwidth of the signal hence compression can be achieved. The compressed speech signal can be reconstructed to form the original signal by decoding followed by dequantization and then performing the inverse-transform methods. This would reproduce the original signal.

IV. Weaknesses of Fourier analysis

This chapter develops the need and motivation for studying the wavelet transform. Historically, Fourier Transform has been the most widely used tool for signal processing. As signal processing began spreading its tentacles and encompassing newer signals, Fourier Transform was found to be unable to satisfy the growing need for processing a bulk of signals. Hence, this chapter begins with a review of Fourier Methods Detailed explanation is avoided to rid the discussion of insignificant details. A simple case is presented, where the shortcomings of Fourier methods is expounded. The next chapter concerns wavelet transforms, and shows how the drawback of FT is eliminated.

4.1 Review of Fourier Methods

For a continuous –time signal x(t) , the Fourier Transform (FT) equations are

$$X(f) = \int_{-\infty}^{\infty} x(t).\, e^{-2j\pi ft}\, dt \ldots\ldots\ldots..(4.1)$$

$$x(t) = \int_{-\infty}^{\infty} X(f).\, e^{2j\pi ft}\, df \ldots\ldots\ldots\ldots(4.2)$$

Equation (2.1) is the analysis equation and equation (2.2) is the synthesis equation.

The synthesis equation suggests that the FT expresses the signal in terms of linear combination of complex exponential signal. For a real signal, it can be shown that the FT synthesis equation expresses the signal in terms of linear combination of sine and cosine terms.

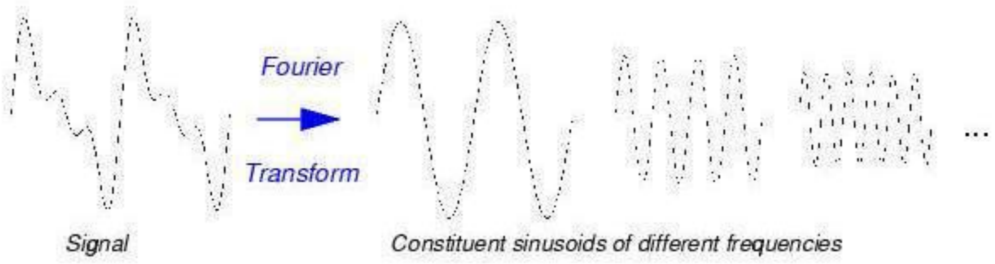

Fig 4.1: constituent sinusoids of different frequencies

The analysis equation represents the given signal in a different form; as a function of frequency. The original signal is a function of time, whereas the after the transformation, the same signal is represented as a function of frequency. It gives the frequency components

Fig4.2: Fourier transform

Thus the FT is a very useful tool as it gives the frequency content of the input signal. It however suffers from a serious drawback. It is explained through an example in the sequel.

4.2 Shortcomings of FT

Ex: 2.1- Consider the following 2 signals

x1(t) = sin(2*p*100*t) 0 <= t < 0.1 sec
= sin(2*p*500*t) 0.1 <= t < 0.2 sec
x2(t) = sin(2*p*500*t) 0 <= t < 0.1 sec
= sin(2*p*100*t) 0.1 <= t < 0.2 sec

A plot of these signals is shown below.

(Note: A time interval of 0 to 0.2 seconds was divided into 10,000 points. The sine of each point was computed and plotted. Since the signal is of 10,000 points, 16,384 point FFT was computed which represents the frequency domain of the signal.)

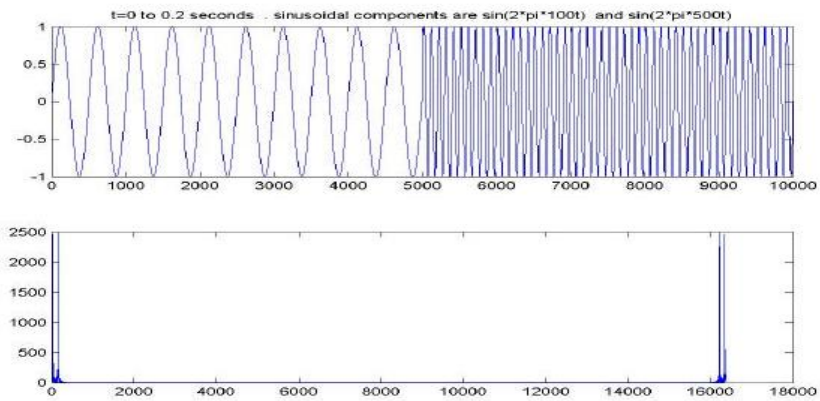

Fig4.3:signalX1 (t) and its FFT

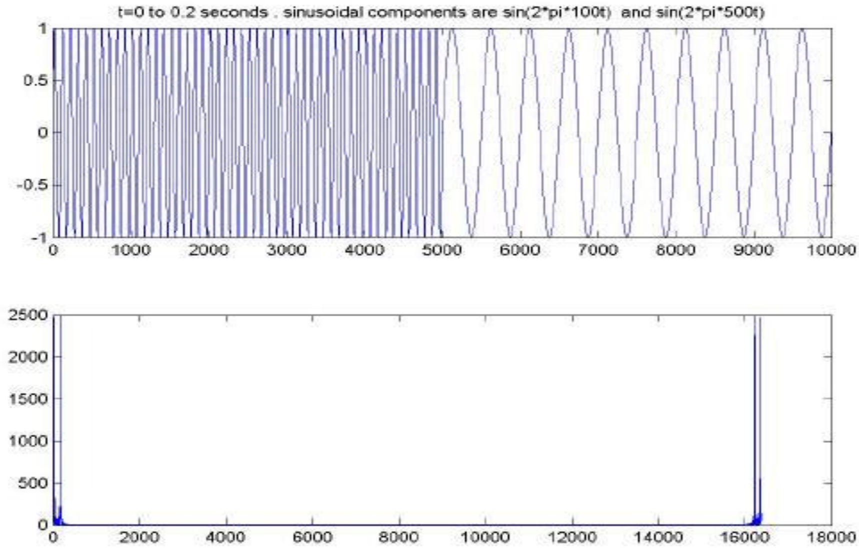

Fig4.4:signalX2 (t) and its FFT

The above example demonstrates the drawback inherent in the Fourier analysis of signals. It shows that the FT is unable to distinguish between two different signals. The two signals have same of giving time information of signals.

In general, FT is not suitable for the analysis of a class of signals called "Non stationary signals". This led to the search of new tools for analysis of signals. One such tool that was proposed was the "Short time Fourier transforms" (STFT). This STFT too suffered from a drawback1 and was supplanted by "Wavelet transform".

V. Procedure

5.1 Wavelet based compression techniques

Wavelets concentrate speech signals into a few neighboring coefficients. By taking the wavelet transform of a signal, many of its‟ coefficients will either be zero or have negligible magnitudes. Data compression can then be done by treating the small valued coefficients as insignificant data and discarding them. Compressing a speech signal using wavelets involves the following stages.

5.2Choice of wavelets

Choosing mother-wavelet function which is used in designing high quality speech coders is of prime importance. Choosing a wavelet having a compact support in time and frequency in addition to a significant number of vanishing moments is important for wavelet speech compressor. Different criteria can be used in selecting an optimal wavelet function. The objective is to minimize the error variance and maximize signal to noise ratio. They can be selected based on the energy conservation properties. Better reconstruction quality is provided by wavelets with more vanishing moments, as they introduce lesser distortion and concentrate more signal energy in neighboring coefficients.

However the computational complexity of DWT increases with the number of vanishing moments. Hence it is not practical to use wavelets with higher number of vanishing moments. Number of vanishing moments of a wavelet indicates the smoothness of a wavelet function and also the flatness of the frequency response of the wavelet filters. Higher the number of vanishing moments, faster is the decay rate of wavelet coefficients. It leads to a more compact signal representation and hence useful in coding applications. However, length of the filters increases with the number of vanishing moments and the hence complexity of computing the DWT coefficients increases.

5.3 Decomposition of wavelets

Wavelets decompose a signal into different resolutions or frequency bands. Signal compression is based on the concept that selecting small number of approximation coefficients and some of the detail coefficients can represent the signal components accurately. Choosing a decomposition level for the DWT depends on the type of signal being used or parameters like entropy.

5.4 Truncation of coefficients

Compression involves truncating wavelet coefficients below threshold. Most of the speech energy is high-valued coefficient. Thus the small valued coefficients can be truncated or zeroed and can then be used for reconstruction of the signal. This compression technique provided lesser signal-to-noise ratio.

.

5.5 Encoding coefficients

Signal compression is achieved by first truncating small-valued coefficients and then encoding these coefficients. High-magnitude coefficients can be represented by storing the coefficients along with their respective positions in the wavelet transform vector. Another method for compression is to encode consecutive zero valued coefficient with two bytes. One byte indicates the sequence of zeros in the wavelet transforms vector and the second byte represents the number of consecutive zeros. For further data compression a suitable bit-encoding format can be used. Low bit rate representation of signal can be achieved by using an entropy coder like Huffman coding.

5.6 Calculating threshold

Two different thresholding techniques are used for the truncation of coefficients i.e. global thresholding and level thresholding.

- **Global Thresholding-** It takes the wavelet expansion of the signal and keeps the largest absolute value coefficient. In this we manually set a global threshold. Hence only a single parameter needs to be selected in this case.
- **Level Thresholding-** It applies visually determined level dependent thresholds to each of the decomposition level in the wavelet transform.

5.7 Encoding zero value functions

In this method, consecutive zero valued coefficients are encoded with two bytes. One byte specifies the starting string of zeros and the second byte keeps record of the number of successive zeros. This encoding method provides a higher compression ratio.

VI. DCT based compression technique

The given sound file is read. The vector is divided into smaller frames and arranged into matrix form. DCT operation is performed on the matrix. DCT operation is performed and the elements are sorted in their matrix form to find components and their indices.

The elements are arranged in descending order. After the arrangement has been done, a Threshold value is decided. The coefficients below the threshold values are discarded. Hence reducing the size of the signal which results in compression. The data is then converted back into the original form by using reconstruction process. For this we perform IDCT operation on the signal. Now convert the signal back to its vector form. Thus the signal is reconstructed.

VII. Applications of compression

1. The use of compression in recording applications is extremely powerful. The playing time of the medium is extended in proportion to the compression factor.
2. In the case of tapes, the access time is improved because the length of the tape needed for a given recording is reduced and so it can be rewound more quickly.
3. In digital audio broadcasting and in digital television transmission, compression is
   Used to reduce the bandwidth needed.
4. The time required for a web page to be displayed and the downloading time in case of files is greatly reduced due to compression.

VIII. Compression terminology

- Compression ratio:- The compression ratio is defined as
  Compression ratio = size of the output stream/size of the input stream. A value of 0.6 means that the data occupies 60% of its original size after compression. Values greater than 1 mean an output stream bigger than the input stream. The compression ratio can also be called bpb(bit per bit),since it equals the no. of bits in the compressed stream needed, on an average, to compress one bit in the input stream.
- Compression factor:- It is the inverse of compression ratio. Values greater than 1 indicate compression and less than 1 indicates expansion

8.1 Aim, scope and limitations of this thesis

The primary objective of this thesis is to present the wavelet based method for the compression of speech. The algorithm presented here was implemented in MATLAB the said software is provided in the accompanying CD. Readers may find it useful to verify the result by running the program

Since this thesis is an application of wavelets, it was natural to study the basics of wavelets in detail. The same procedure was adopted in writing this thesis, as it was felt 5 that without minimal background in wavelets, it would be fruitless, and also inconvenient to explain the algorithm.

However, the wavelet itself is an engrossing field, and a comprehensive study was beyond the scope of our undergraduate 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.

At this stage, it is worthwhile mentioning two quotes by famous scientists

'So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality.' --Albert Einstein 'As complexity rises, precise statements lose meaning and meaningful statements lose precision.' --Lotfi Zadeh 1

The inclusion of the above quotes is to highlight the fact that simplicity and clarity are often the casualties of precision and accuracy, and vice-versa.

In this thesis, we have compromised on the mathematical precision and accuracy to make matters simple and clear. An amateur in the field of wavelets might find this work useful as it is relieved of most of the intimidating vector analysis and equations, which have been supplanted by simple diagrams. However, for our own understanding, we did found it necessary, interesting and exciting to go through some literature which deal with the intricate details of wavelet analysis, and sufficient references have been provided wherever necessary, for the sake of a fairly advanced reader. Some of the literature that we perused has been included in the CD.

The analysis that we undertook for wavelets includes only the orthogonal wavelets. This decision was based on the extensive literature we read on the topic, wherein the suitability of these wavelets for speech signals was stated. Another topic that has been deliberately excluded in this work is the concept of MRA, which bridges the gap between the wavelets and the filter banks and is indispensable for a good understanding of Mallet's Fast Wavelet Transform Algorithm. Instead, we have assumed certain results and provided references for further reading.

Secondly, the sound files that we tested were of limited duration, around 5 seconds. Albeit the programs will run for larger files (of course, the computation time will be longer in this case), a better approach towards such large files is to use frames of finite length. This procedure is more used in real-time compression of sound files, and is not presented here.

Encoding is performed using only the Run Length Encoding. The effect of other encoding schemes on the compression factor has not been studied.

This thesis considers only wavelets analysis, wherein only approximation coefficients are split. There exists another analysis, called wavelet packet analysis, which splits detail coefficients. This is not explored in this thesis.

IX. Conclusion and future scope

In this project compress the data by optimization of wavelet, scale, and level. This technology is needed in the field of speech to satisfy transfer requirements of huge speech signals via communication companies and decreasing storage equipment is another need.

The main objective was to develop an appreciation for wavelet transforms, discuss their application in compression of human speech signals and study the effect of a few parameters on the quality of compression.

The parameters studied are: Sampling frequency, type of wavelet, threshold, file. Here using only hear, daubechies wavelets etc, if apply the advanced wavelets like biorthogonal wavelets achieve better performance.

Encoding is performed using only the Run Length Encoding. Higher compression ratios are expected with coding techniques like Huffman coding

REFERENCES

[1.] Robi Polikar,"The Wavelet Tutorial Part I, Part II and Part III".
[2.] J. Pang, S.Chauhan,"FPGA Design of Speech
[3.] Compression Using DWT" Proceeding of World
[4.] Congress on Engineering and Computer science October 22-24, 2008, San Francisco, USA.
[5.] Fundamentals of speech recognition by Lawrence Rabiner.