Compression using Wavelet Transform

Othman O. Khalifa khalifa@iium.edu.my
Electrical and Computer Engineering Department
International Islamic University Malaysia
Gombak, P.O Box 10, 50728 Kuala Lumpur, Malaysia

Sering Habib Harding habs13@hotmail.com
International Islamic University Malaysia
Gombak, P.O Box 10, 50728 Kuala Lumpur, Malaysia

Aisha-Hassan A. Hashim aisha@iiu.edu.my
Electrical and Computer Engineering Department
International Islamic University Malaysia
Gombak, P.O Box 10, 50728 Kuala Lumpur, Malaysia

Abstract

Audio compression has become one of the basic technologies of the multimedia age. The change in the telecommunication infrastructure, in recent years, from circuit switched to packet switched systems has also reflected on the way that speech and audio signals are carried in present systems. In many applications, such as the design of multimedia workstations and high quality audio transmission and storage, the goal is to achieve transparent coding of audio and speech signals at the lowest possible data rates. In other words, bandwidth cost money, therefore, the transmission and storage of information becomes costly. However, if we can use less data, both transmission and storage become cheaper. Further reduction in bit rate is an attractive proposition in applications like remote broadcast lines, studio links, satellite transmission of high quality audio and voice over internet.

Keywords: Audio Compression, Wavelet transform.

1. INTRODUCTION

The growth of the computer industry has invariably led to the demand for quality audio data. Compared to most digital data types, the data rates associated with uncompressed digital audio are substantial. For example, if we want send high-quality uncompressed audio data over a modem, it would take each second's worth of audio about 30 seconds to transmit. This means that the data would be gradually received, stored away and the resulting file played at the correct rate to hear the sound. However, if real-time audio is to be sent over a modem link, data compression must be used.

In a digital system, the bit rate is the product of the sampling rate and the number of bits in each sample. The difference between the information rate of a signal and its bit rate is known as the redundancy. Compression systems are designed to eliminate this redundancy. These systems rely on the fact that information, by its very nature is not random but exhibits order and patterning. According to Shannon's information theory, "any signal, which is totally predictable, carries no information" [4].

Therefore, if the order and pattern can be extracted, the essence of the information can often be represented and transmitted using less data than would be required for the original signal [5].

2. AUDIO COMPRESSION TECHNIQUES

Lossless Compression

Lossless compression works by removing the redundant information present in an audio signal. This would be the ideal compression technique as there is no cost to using it other than the cost of the compression and decompression process. However, lossless compression suffers from two disadvantages. First, it offers small compression ratios, so used alone it does not meet economic needs. Also, it does not guarantee a constant output data rate as the compression ratio is very much dependent on the input data. One advantage of Lossless compression is that it can be applied to any data stream. Lossless techniques are applied in the last stages of Audio and Video coders to reduce the data rate even further. Two Lossless techniques that are in general use are: Run-Length Encoding and Entropy Encoding.

Lossy or Perceptive compression

In Lossy coding, the compressed data is not identical bit-for-bit with the original data. This method is also called Perceptive coding as it utilizes the fact that some information is truly irrelevant in that the intended recipient will not be able to perceive that it is missing. In most cases, information that is close to irrelevant is also made redundant, where the quality loss is small compared to the data savings.

The objective of Lossy compression is to get maximum benefit, i.e., compression ratio or bit rate reduction, at reduced cost, i.e., loss in quality.

To pinpoint the portions of the audio signal that is redundant involves using psychoacoustic analysis to determine a masking threshold below which the Power of the signal is not strong enough to be heard by the human ear. The figure below illustrates this point.

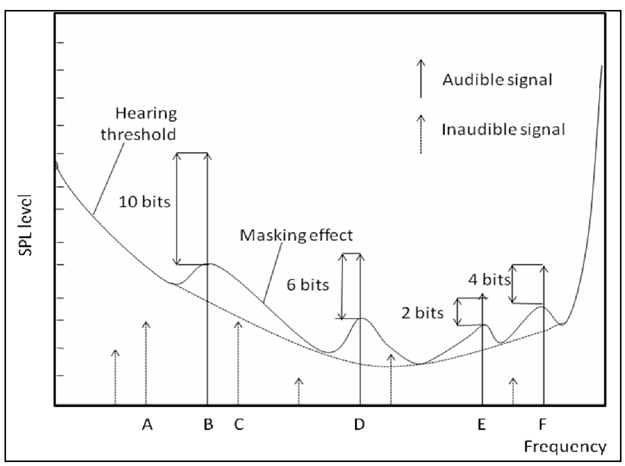

FIGURE 1: The bit allocation algorithm assigns bits according to audibility of sub band signals. Inaudible tones are not assigned bits, and are not coded [3]

3. WAVELET TRANSFORMATION

A *wavelet* is defined as a "small wave" that has its energy concentrated in time to provide a tool for the analysis of transient, non-stationary, or time-varying phenomena. It has the oscillating wave-like properties but also has the ability to allow simultaneous time and frequency analysis [7]. Wavelet Transform has emerged as a powerful mathematical tool in many areas of science and engineering, more so in the field of audio and data compression.

A signal or function *f (t)* can often be better analyzed, described, or processed if expressed as a linear decomposition by:

$$f(t) = \sum_{\ell} a_{\ell}\, \psi_{\ell}(t) \qquad (1)$$

where ◎ is an integer index for the sum, a_{ℓ} is the expansion coefficients and $\psi_{\ell}(t)$ is the set of real-valued functions of *t* called the expansion set. If the expansion is unique, the set is called a basis for the functions that could be represented. If the basis is orthogonal, then the coefficients can be calculated by the *inner product*

$$a_k = \langle f(t), \psi_k(t) \rangle = \int f(t)\psi_k(t)\, dt. \qquad (2)$$

A single a_k coefficient is obtained by substituting (1) into (2) and therefore for the *wavelet expansion,* a two-parameter system is constructed such that (1) becomes

$$f(t) = \sum_{k}\sum_{j} a_{j,k}\, \psi_{j,k}(t) \qquad (3)$$

Where both *j* and *k* are integer indices and $\psi_{j,k}(t)$ is the wavelet expansion that usually forms an orthogonal basis. The set of expansion coefficients $a_{j,k}$ are called the discrete wavelet transform of *f (t)* and (3) is its inverse.

All wavelet systems are generated from a single scaling function or wavelet by simple scaling and translation. This two-dimensional representation is achieved from the function $\psi(t)$, also called the mother wavelet, by

$$\psi_{j,k}(t) = 2^{j/2}\, \psi(2^j t - k) \qquad j,k \in \boldsymbol{Z} \qquad (4)$$

Wavelet systems also satisfy multi-resolution conditions. In effect, this means that a set of scaling functions can be determined in terms of integer translates of the basic scaling function by

$$\varphi_k(t) = \varphi(t-k) \quad k \in \boldsymbol{Z} \quad \varphi \in L^2 \qquad (5)$$

It can therefore be seen that if a set of signals can be represented by $\varphi(t-k)$; a larger set can represented by $\varphi(2t-k)$, giving a better approximation of any signal.

Hence, due to the spanning of the space of $\varphi(2t)$ by $\varphi(t)$, $\varphi(t)$ can be expressed in terms of the weighted sum of the shifted $\varphi(2t)$ as

$$\varphi(t) = \sum_{n} h(n)\sqrt{2}\,\varphi(2t-n), \qquad n \in \boldsymbol{Z} \qquad (6)$$

Where the coefficients $h(n)$ may be real or complex numbers called the scaling function coefficients.

However, the important features of a signal can better be described, not by $\varphi_{j,k}(t)$, but by defining a slightly different set of functions $\psi_{j,k}(t)$ that span the differences between the spaces spanned by the various scales of the scaling function. These functions are the wavelets and, they can be represented by a weighted sum of shifted scaling function $\varphi(2t)$ defined in (6) by

$$\psi(t) = \sum_{n} h_1(n)\sqrt{2}\,\varphi(2t-n), \qquad n \in \boldsymbol{Z} \qquad (7)$$

The function generated by (7) gives the prototype or mother wavelet $\psi(t)$ for a class of expansion functions of the form given by (4).

$$f(t) = \sum_{k=-\infty}^{\infty} c(k)\,\varphi_k(t) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} d(j,k)\,\psi_{j,k}(t) \qquad (8)$$

The coefficients in this wavelet expansion are called the discrete wavelet transform (DWT), of the signal $f(t)$. For a large class of signals, the wavelet expansion coefficients drop off rapidly as *j* and *k* increase. As a result, the DWT is efficient for image and audio compression.

4. COMPARISON BETWEEN WAVELET TRANSFORM AND FOURIER TRANSFORM.

The DWT is very similar to a Fourier series, but in many ways, is much more flexible and informative. It is a tool which breaks up data into different frequency components or sub bands and then studies each component with a resolution that is matched to its scale. Unlike the Fourier series, it can be used on non-stationary transient signals with excellent results.
The Fourier Transform is given by:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\,dt \qquad (9)$$

It involves the breaking up of a signal into sine waves of various frequencies. The advantages of Wavelets over Fourier methods in analyzing physical situations stem from the fact that sinusoids do not have a limited duration but instead extend from minus to plus infinity.

In Fourier transform domain, we completely lose information about the localization of the features of an audio signal. Quantization error on one coefficient can affect the quality of the entire audio file. The wavelet expansion allows a more accurate local description and separation of signal characteristics. A wavelet expansion coefficient represents a component that is itself local and is easier to interpret.

The Fourier basis functions have infinite support in that a single point in the Fourier domain contains information from everywhere in the signal. Wavelets, on the other hand, have compact or finite support and this enables different parts of a signal to be represented at different resolution.

Wavelets are adjustable and adaptable and can therefore by designed for adaptive systems that adjust themselves to suit the signal. Fourier Transform, however, is suitable only if the signal consists of a few stationary components. Also, the amplitude spectrum does not provide any idea how the frequency evolve with time.

All wavelets tend to zero at infinity, which is already better than the Fourier series function. Furthermore, wavelets can be made to tend to zero as fast as possible. It is this property that makes wavelets so effective in signal and audio compression.

5. AUDITORY MASKING

Auditory masking is a perceptual property of the human auditory system that occurs whenever the presence of a strong audio signal makes a temporal or spectral neighborhood of a weaker audio signal imperceptible. If two sounds occur simultaneously and one is masked by the other, this is referred to as simultaneous masking. A sound close in frequency to a louder sound is more easily masked than if it is far apart in frequency. For this reason, simultaneous masking is also sometimes called frequency masking.

A weak sound emitted soon after the end of a louder sound is masked by the louder sound. In fact, even a weak sound just before a louder sound can be masked by the louder sound. These two effects are called forward and backward temporal masking respectively [3].

It is of special interest for perceptual audio coding to have a precise description of all masking phenomena to compute a masking threshold that can be used to compress a digital signal. Using this, it is possible to reduce the SNR and therefore the number of bits. In the perceptual audio coding schemes, these masking models are often called psychoacoustic models. Psychoacoustics research also reveals the existence of an absolute threshold. The minimum threshold of hearing describes the minimum level at which the ear can detect a tone at a given frequency. It is normally referenced to 0dB at1kHz.

6. PSYCHOACOUSTIC MODEL

The human auditory system has a dynamic frequency range from about 20Hz - 20 kHz, and the intensity of the sound as perceived by us varies. However, we are not able to perceive sounds equally well at all frequencies. In fact, hearing a tone becomes more difficult close to the extreme frequencies (i.e. close to 20 Hz and 20 kHz). Further study exhibits the concept of critical bands which is the basis of audio perception.

A critical band is a bandwidth around a center frequency, within which sounds with different frequencies are blurred as perceived by us [8]. Critical bands are important in perceptual coding because they show that the ear discriminates between the energy in the band and the energy outside the band. It is this phenomenon that promotes masking.

In this implementation, the following were determined:

- Tone maskers
- Noise maskers
- Masking effect due to these maskers

Tone Masker

Determining whether a frequency component is a tone requires knowing whether it has been held constant for a period of time, as well as whether it is a sharp peak in the frequency spectrum, which indicates that it is above the ambient noise of the signal.

A frequency f (with FFT index k) is a tone if its power P[k] is:

1. greater than P[k-1] and P[k+1], i.e., it is a local maxima
2. 7 dB greater than the other frequencies in its neighborhood, where the neighborhood is dependent on f:
   - If 0.17 Hz < f <5.5 kHz, the neighborhood is [k-2…k+2].
   - If 5.5 kHz ≤ f <11 kHz, the neighborhood is [k-3…k+3].
   - If 11 kHz ≤ f <20 kHz, the neighborhood is [k-6…k+6].

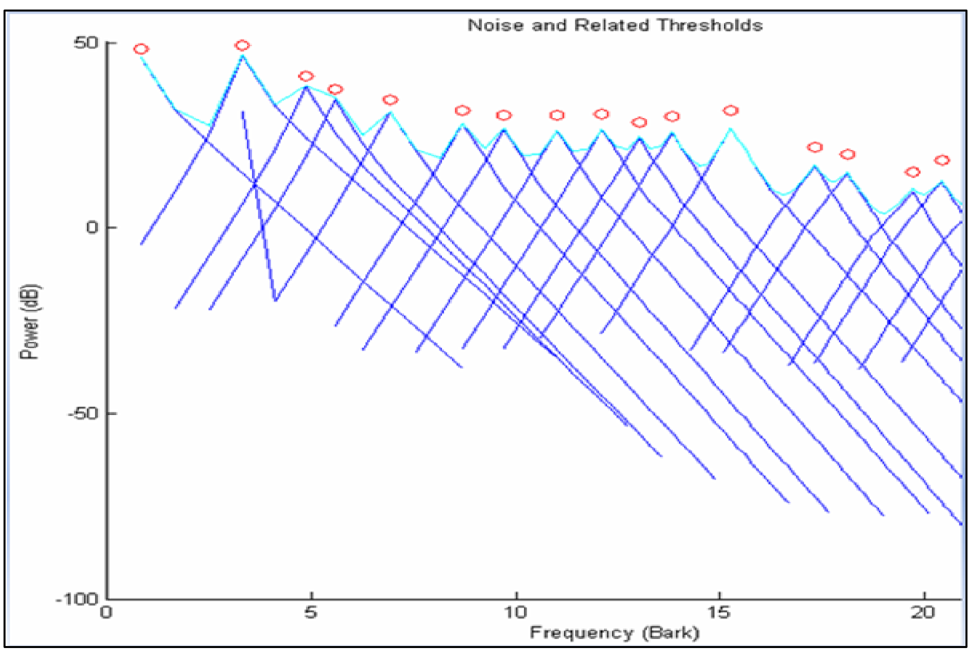

Figure 2: Thresholds resulting from each Tone masker

Noise Masker
If a signal is not a tone, it must be noise. Thus, one can take all frequency components that are not part of a tone's neighborhood and treat them like noise. Since humans have difficulty discerning signals within a critical band, the noise found within each of the bands can be combined to form one mask. Therefore, the idea is to take all frequency components within a critical band that do not fit within tone neighborhoods, add them together, and place them at the geometric mean location within the critical band.

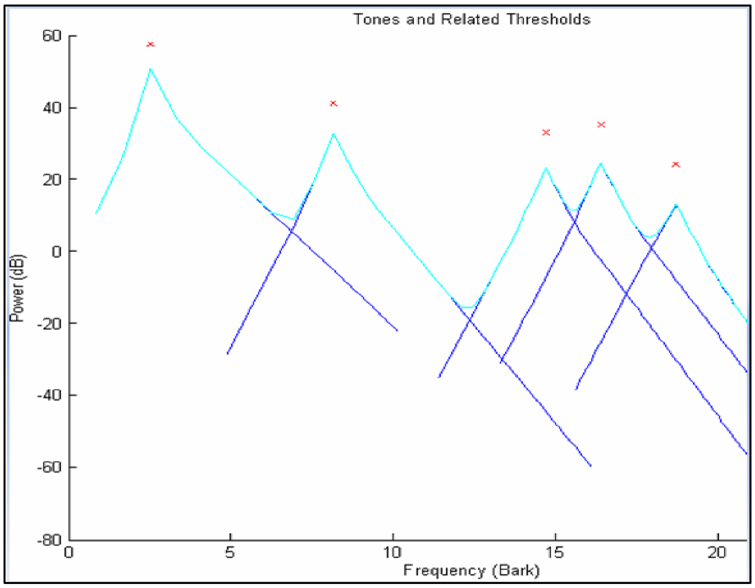

FIGURE 3: Threshold resulting from Noise Maskers

Masking Effect
The maskers which have been determined affect not only the frequencies within a critical band, but also in surrounding bands.

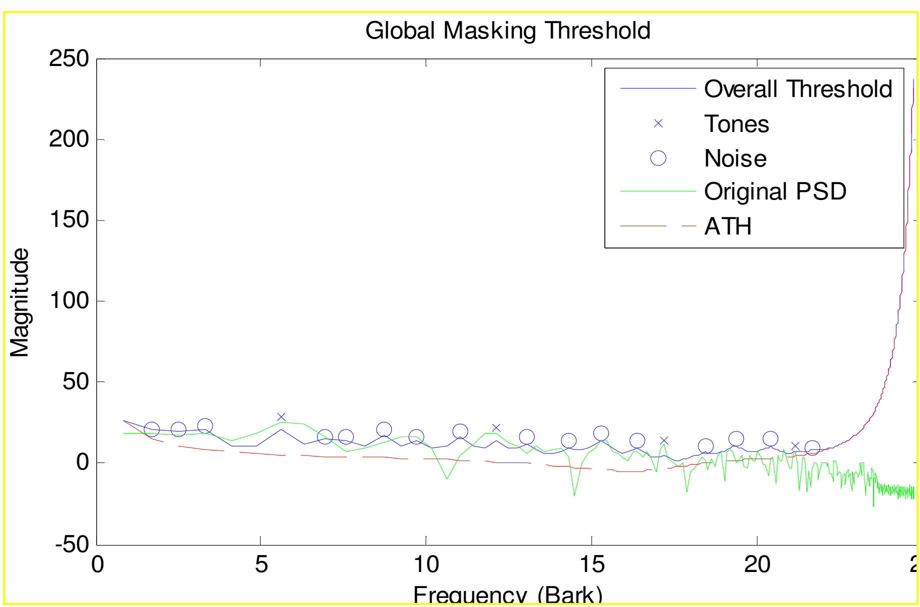

FIGURE 4: Sum of all the thresholds

After determining all these maskers, it was assumed that masking is additive and therefore the effects of the tones and noise maskers as well as their masking effect were added together to form a global threshold as shown in Figure 4 above.

7. IMPLEMENTATION

The following approach was used to compress an audio data. First, the data is divided into frames. For each frame, a wavelet representation is used to minimize the number of bits required to represent the frame while keeping any distortion inaudible. This scheme is highly successful because it reduces the number of non-zero wavelet coefficients. In addition, these coefficients may be encoded using a small number of bits. The capabilities of MATLAB's Wavelet Toolbox were utilized. The Wavelet Toolbox incorporates many different wavelet families and their coefficients. From the analysis, it was decided to use the Daubechies family of wavelets for coding audio signals.

The Wavelet Toolbox's built-in functions *dwt*, *wavedec*, *waverec* and *idwt*, were used to compute the forward and inverse wavelet transforms. *Wavedec* computes the multi-level decomposition of a signal and *waverec* reconstructs the signal from their coefficients.

8. SIMULATION

In this section, we are trying to simulate an audio codec that utilizes the wavelet transformation to perform compression of high quality audio whilst maintaining transparent quality at low bit rates.

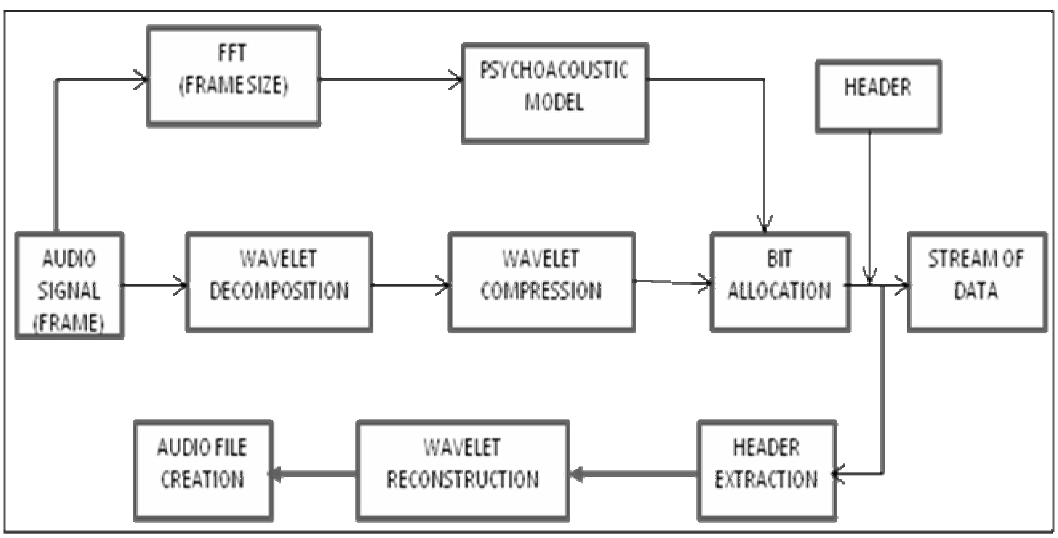

FIGURE 5: Block diagram of the MATLAB Implementation

The diagram below illustrates the MATLAB implementation used. It consists of the following features:

(a) Signal division and processing using small frames
(b) Discrete wavelet decomposition of each frame
(c) Compression in the wavelet domain
(d) A psychoacoustic model
(e) Non linear quantization over the wavelet coefficient using the psychoacoustic model
(f) Signal reconstruction
(g) Main output: Audio files.

9. RESULTS

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

- Signal to Noise Ratio (SNR)
- Peak Signal to Noise Ratio (PSNR)
- Normalized Root Mean Square Error (NRMSE)
- Compression Ratios

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

1. Signal to Noise Ratio

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

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

2. Peak Signal to Noise Ratio

$$\mathrm{PSNR} = 10\log_{10}\frac{NX^2}{\|x-r\|^2} \qquad (11)$$

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

3. Normalized Root Mean Square Error

$$\mathrm{NRMSE} = \sqrt{\frac{\overline{(x(n)-r(n))^2}}{\overline{(x(n)-\mu_x(n))^2}}} \qquad (12)$$

x(n) is the speech signal, *r(n)* is the reconstructed signal, and µ*x(n)* is the mean of the speech signal.

4. Compression Ratio

$$C = \frac{\mathrm{Length}(x(n))}{\mathrm{Length}(cWC)} \qquad (13)$$

cWC is the length of the compressed wavelet transform vector.

Wavelets	Zeros (%)	Retained Energy (%)	SNR	PSNR	NRMSE
Haar	44.9	99.62	30.4	41.55	0.0018
Db4	47.2	99.79	31.7	43.49	0.0010
Db6	50.1	99.86	32.2	42.50	0.0015
Db8	50.7	99.92	34.1	43.19	0.0012
Db10	53.6	99.98	34.5	45.20	0.012

TABLE 1: Performance of test signal ‘testsig.wav’ over different wavelets

Wavelet	Compression score
Haar	0.48
Db4	0.56
Db6	0.88
Db8	1.32
Db10	1.88

TABLE 2: Compression score of 'testsig.wav' over different wavelets

10. DISCUSSION AND CONCLUSION

The demand for compression technology increases every year in parallel with the increase in aggregate bandwidth for the transmission of audio and video signals. As a result, the Wavelet-based approach plays an important role in the scheme of things as Perceptual coding of audio signals found its way to a growing number of consumer applications.

The foremost criterion for audio compression technology is to achieve a certain signal quality at a given bit-rate as this directly translates to cost savings by getting a higher compression ratio at the same quality of service. Wavelet-based compression is claimed to be more efficient at low bit rates but are actually less successful than discrete cosine transform (DCT) -based systems in achieving good efficiency at near-transparent compression ratios.

Computational complexity also limits the algorithmic implementation of a codec. As a result, algorithmic delay becomes an important constraint especially for two-way communications applications. In that respect, it is notable that Wavelet compression does require more computational power than DCT-based compression.

The wavelet based compression software designed reaches a signal to noise ratio of 34.5 db at a compression ratio of 1.88 using the Daubechies 10 wavelet. The performance of the wavelet scheme in terms of compression scores and signal quality is incomparable with other good techniques such as MP3 codecs; however the implemented scheme performs reasonably well with an average fidelity and with much less computational burden. In addition, using wavelets, the compression ratio can be easily varied, while most other compression techniques have fixed compression ratios.

REFERENCES

11. D. Sinha and A. Tewfik. "Low Bit Rate Transparent Audio Compression using Adapted Wavelets", IEEE Trans. ASSP, Vol. 41, No. 12, December 1993.
12. P. Srinivasan and L. H. Jamieson. "High Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling", IEEE Transactions on Signal Processing, Vol. 46, No. 4, April 1998.
13. Pohlmann, K.C., 2000. *Principles of Digital Audio*. New York: McGraw-Hill,
14. Proakis, J.G., Manolakis, D.G., 1996. *Digital Signal Processing: Principles, Algorithms and Applications*. New Jersey: Prentice-Hall
15. Symes, P., 2001. *Digital Video Compression*. New York: McGraw-Hill.
16. Watkinson, J., 1995. *Compression in Video and Audio*. Oxford: Focal Press.
17. Burrus, C.S., Gopinath, R.A., Guo, H., 1998. *Introduction to Wavelets and Wavelet Transforms*. New Jersey: Prentice-Hall.

18. Khars, M., & Brandenburg, K. (edit.). 1998. *Applications of Digital Signal Processing to Audio and Acoustics.* Massachusetts: Kluwer Academic Publishers.
19. Rowden, C. (edit.). 1992. *Speech Processing*, Berkshire,UK: McGraw-Hill.
20. Khalifa O. O., Review of Wavelet theory and its application to Image Data Compression, International Islamic University Malaysia Engineering Journal, Vol.4, No.1, p25-43, 2003.