A Simple Lossy Audio Coding Scheme Based on the DWT

Vikas Kumar Gupta*, Prashant Kumar Khevariya* and José Salvado**
* Department of Electrical Engineering
Indian Institute of Technology – Roorkee
Haridwar, Uttarakhand
Roorkee, India
Telf: +91 1332 285239; Fax: +91 1332 277587; e-mail: {vik4uuee, khareuee}@iit.ernet.in

** Department of Electrical Engineering
EST – IPCB
Av. do Empresário, s/n – Castelo Branco
Telf: +351 272 339 300; Fax: +351 272 339 399; e-mail: jsalvado@ieee.org

***Abstract* — This paper proposes a simple scheme for audio coding that does not use perceptual models. The audio coder is based on the discrete wavelet transform to decorrelate signals, computed through the lifting scheme, and followed by Huffman coding. The evaluation of the coding scheme is presented by using some .wav audio test files, coded for different conditions, and also includes subjective evaluation. Experimental results show the compression ratios achieved, the degradation of the signals expressed as values of signal to noise ratio and the changes in spectrum information.**

1. Introduction

Audio coding or audio compression algorithms are used to obtain compact digital representations of high-fidelity (wideband) audio signals for the purpose of efficient transmission over larger distances and storage. This becomes a necessity because of the bandwidth constraint and to reduce the transmission time between the wirelessly connected systems. The main objective in audio coding is to represent the signal with a minimum number of bits while achieving transparent signal reproduction i.e. generating output audio that cannot be distinguished from the original input signal.

Based on the signal model or in the analysis/synthesis techniques to encode audio signals one can classify audio coders into one of the following four types [1]: linear predictive, transform based, sub-band and sinusoidal. Algorithms can also be classified as lossy or lossless, according to the nature of the coding scheme. In lossy audio schemes compression is achieved by exploring information that is perceptually irrelevant. On the other hand lossless audio coding schemes the information is "merely" packed in order to obtain a different and more compact (though efficient) representation of data. Lossless audio coding techniques yield in general high quality audio without artifacts at high frequency or at high rates. Lossy audio coding techniques allow for better compression ratios (e.g. up to 25:1 and more) but are more susceptible to high rate artifacts.

With the improvement of circuit technology and telecommunication bandwidth, there have been a great amount of interest in the high-quality, in which a large amount of data need to be processed, compressed and transmitted in real time. But there exist a trade-off between the compression ratio and reconstruction quality. Higher compression ratios imply more degradation on the sound quality, when reconstructed. Any compression scheme should consider these facts and should also provide a good control mechanism to use the limited bandwidth and maintain acceptable quality, keeping in mind the computational complexity. The computational complexity directly relates to the processing speed and it should be as less as possible so as to reduce the propagation delay during real time processing.

Typical perceptual audio compression schemes, e.g. MPEG audio, represented generically in figure 1, use signal analysis, psychoacoustic models and bit allocation and coding blocks. At the ISO/MPEG layer III (MP3) coding scheme, the time to frequency mapping block includes a polyphase analysis filter banks followed by decimation of a factor of 32, feeding a modified discrete cosine transform (MDCT) and adaptive segmentation block, also connected with the psychoacoustic model. The bit allocation block includes block companding, quantization and Huffman coding [3].

This paper proposes a simple audio coding and decoding (CoDec) scheme based on sub-band analysis with polyphase filter banks, using the discrete wavelet transform (DWT). In the proposed CoDec we do not use any perceptual model in order to reduce computational complexity. To evaluated the audio CoDec we used some **.WAV** test files, and two families of wavelets: the Daubechies 2 (DB2) orthogonal wavelet and the biorthogonal (4,4) wavelet. Evaluation tests included the measurement of the compression ratio, the signal to noise ratio, spectrum modification and perceptual quality.

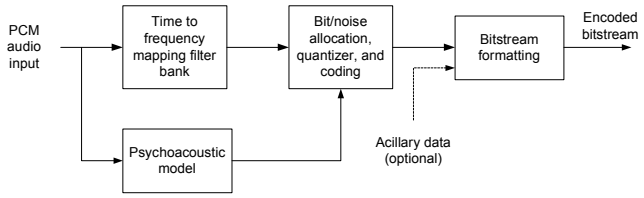

Fig .1. Diagram of a generic perceptual audio coding scheme[2].

The rest of the paper is organized as follows. The discrete wavelet transforms and its implementation through the lifting scheme is explained in Section II. In Section III the structure of the proposed audio CoDec is presented and explained and compared with the most common coding schemes for audio signals. Evaluation tests and experimental results are shown and discussed in section IV and finally, in Section VI, we present the main conclusions.

The work that supports this paper was carried out by the first two authors during a two-month internship at EST-IPCB, on summer 2007 [4].

2. The DWT and the Lifting Scheme

In signal analysis, a signal $f(t)$ is represented as a weighted sum of building blocks of basis functions

$$f(t) = \sum_k c_k \psi_k(t) \qquad (1)$$

where ψ_k are basis functions and c_k are weight coefficients. Since the basis functions are fixed, the information about the signal is carried by the coefficients. Choosing sinusoids as the basis functions in (1) yields the Fourier representation. To detect spikes or well localized high frequency components in the signal, one needs a representation which contains information about both, the time and frequency behaviour of the signal. However, resolution in time (Δt) and resolution in frequency $(\Delta\omega)$ cannot be arbitrarily small at the same time as their product is lower bounded by the Heisenberg's uncertainty principle

$$\Delta t \cdot \Delta\omega \geq 1/2 \qquad (2)$$

This means that one must trade off time resolution for frequency resolution, or vice versa. However, low-frequency events are usually spread in time (non local) and high-frequency events are usually concentrated (localized) in time. Therefore, it is possible to obtain good time-frequency information of a signal by choosing the basis functions to act as cascaded octave band-pass filters which repeatedly split the bandwidth of the signal in half. On the other hand sinusoids cannot provide information about the time behaviour of a signal as they have infinite support. The solution is to use basis functions that have finite (compact) support and different widths.

A. The Discrete Wavelet Transform

For a wavelet representation [5], the set of basis functions $\{\psi_k\}$ are scaled and translated versions of the same prototype function $\psi(t)$ known as *mother wavelet*. Scaling is achieved by multiplying t by a scale factor, normally a power of two, $\psi(2^a t), a \in \mathbb{Z}$. Since the prototype function has finite support, it can be scaled and translated to cover the all signal *i.e.* $\psi(2^a t - b), b \in \mathbb{Z}$. The wavelet decomposition of a signal can be represented as

$$f(t) = \sum_a \sum_b c_{ab} \psi_{ab}(t) \qquad (3)$$

Where $\psi_{ab}(t) = 2^{a/2}\psi(2^a t - b)$ and c_{ab} are coefficients that can be computed trough the wavelet transform.
The wavelet transform of a signal $x(t)$ is defined as

$$W_x(a,b) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \qquad (4)$$

where $a \in \mathbb{R}^+$ is the scale or dilation parameter, $b \in \mathbb{R}$ is the translation parameter, $\psi^*(t)$ denotes the complex conjugate of $\psi(t)$ and $1/\sqrt{a}$ normalizes energy along scales. By taking the discrete values of the scale and translation parameters $\left(a = 2^{-j}, b = k2^{-j}\right)$ one can obtain the discrete wavelet transform (DWT) given by:

$$w_{k,j} = W_x(2^{-j}, k2^{-j}) = \frac{1}{\sqrt{2^{-j}}} \int_{\mathbb{R}} x(t) \psi^* \left(\frac{t - k2^{-j}}{2^{-j}} \right) dt \,. \qquad (5)$$

The basic idea of the wavelet transform is to exploit the high correlation structure in most real signals, and build a sparse approximation. The correlation structure is typically local in space (or time) and frequency; the samples that are close each other are more correlated than ones that are far apart. Wavelet analysis has good scale-frequency and localization properties, allows the decomposition of signals with different levels of detail. For a multi-resolution [6] decomposition scheme of a signal with p levels, one has to successively apply the analysis filters to the resulting approximation sub-sequence.

B. The Lifting Scheme

Daubechies and Sweldens proposed and described how any discrete wavelet transform (or a two sub-band filtering scheme) with finite filters can be decomposed into a finite sequence of simple filtering steps and this ladder structure is called *Lifting Steps* [7]. Mathematically, this decomposition corresponds to a polyphase matrix representation of the wavelet filters or DWT, as shown in figure 2. Consider that all filters are of FIR type (for more extensive sub-band transform). At the analysis side one has a pair of analysis filters $\tilde{h}$ (low-pass) and $\tilde{g}$ (high-pass) followed by sub-sampling by a factor of 2.

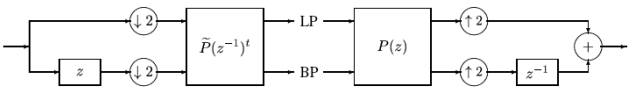

Fig .2. Polyphase representation of the DWT

At the synthesis side coefficients are first up-sampled and then passed trough a pair of synthesis filters *h* and *g* (low and high pass, respectively).
Then one has to factor the polyphase matrices into elementary matrices, using the Euclidean algorithm.
In [7] Daubechies and Sweldens showed how to obtain the algorithms by factoring the wavelet transform into lifting steps for the orthogonal and bi-orthogonal wavelet families. They also showed that with the lifting scheme there is significant speed-up on DWT calculation over the standard implementation, ranging from around 54% for the Daubechies D4, or 64% for the 9/7 filter wavelet pair, up to 100% for the $\left(N, \widetilde{N}\right)$ interpolating.

3. Structure of the Proposed Audio CoDec

In audio compression schemes, besides the transform representation of signals to explore local redundancies, entropy coding techniques are also employed in conjunction with the quantization and bit-allocation modules in order to obtain improved coding efficiencies. In entropy coding, the information symbols are mapped into codes based on the frequency of each symbol. Higher occurring magnitudes are encoded with shorter code words and vice versa. Several coding schemes have been proposed namely Rice coding, Golomb coding and arithmetic coding and Huffman coding, which is probably the most popular.
We developed a prototype of a simple sub-band audio coding scheme, in C/C++ language and also in MATLAB. The CoDec is based on discrete wavelet transform (DWT), computed through the lifting scheme, a quantization block on the zero-tree structure of the wavelet coefficients and an adaptive Huffman coding block. The block diagrams of the coding and decoding processes is presented in figure 3.
In sub-band coding, digital signals are subdivided into multiple sub-bands or frequency ranges by multirate filter banks, and then quantized according to the energy of each sub-band. The algorithm takes the advantage of wavelet transform to achieve better controllability and higher fidelity of the audio signal.
On the other hand, DWT, besides the good time-frequency analysis and synthesis capabilities, also allows for hierarchical decomposition schemes which are also scalable.

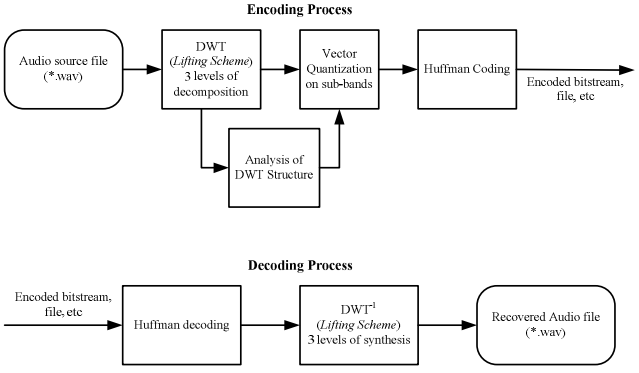

Fig .3. Simplified structures of the coding and decoding blocks

The DWT also has strong relations with sub-band analysis and filter banks and can be computed by the lifting scheme, which is based on polyphase matrix decomposition. On the other hand the wavelet filters are FIR filters that can have linear phase characteristics (in the case of biorthogonal wavelets).
The DWT is applied to the audio input signal with three levels of decomposition, thus concentrating energy in lower frequency bands (higher levels). Then a vector quantization process is applied on the DWT domain and higher frequency components are discarded. Finally an adaptive Huffman coding process explores the redundancy of the code, in order to improve compression ratio and the coded bit stream is stored in a binary file.
We implemented the algorithms of the lifted DWT for the Daubechies 4 (orthogonal wavelet) and for the 9/7 filter pair (or Daubechies 4,4 biorthogonal wavelet) in C/C++ language, with floating-point data representation, on a PC equipped with an Intel Pentium III processor at 1700 MHz with Windows XP.
We also adopt a scheme of symmetric extension of the sequences as well as used zero padding during the computation of the DWT, in order to provide enough support to the algorithm and also to avoid degradation of signal due to the limit conditions (discontinuities).

4. Evaluation and Experimental Results

To evaluate the overall performance of the proposed audio CoDec several objective and subjective tests were made. We measure the compression ratio and the signal to noise ratio, and present some examples of spectrum modifications. We also carried subjective quality tests, simply by collecting the opinions of several individuals on the quality of the signals, after they had listen the recovered audio files for the different test conditions. To perform the tests, some **.WAV** files with different characteristics were used, as follows:

TABEL I
LIST OF ORIGINAL 'WAVE' AUDIO TEST FILES USED

Filename	f_s [Hz]	bps[1]	# samples / ~time [s]	~file size [kB]
'**voice.wav**' [2]	22050	16	110033 / 5	215
'**flute.wav**' [3]	11025	8	96287 / 9	95
'**dontfool.wav**' [4]	11025	8	120000 / 11	118
'**heaven.wav**'	11025	8	120000 / 11	118
'**toomuch.wav**'	11025	8	100000 / 9	98

The original .WAV files were coded considering different scenarios: *i*) using two families of wavelets for the DWT, one with the Daubechies 2 (DB2) orthogonal wavelet and the other the biorthogonal (4,4) wavelet; *ii*) using adaptive quantization along the hierarchical structure on DWT

[1] bps – number of bits per sample (quantization)
[2] Speaking voice of one male individual;
[3] Sound of a musical instrument (flute) at different musical notes and scales (frequencies);
[4] Sound of a musical band (Queen) playing several musical instruments, with different dynamics, rhythms and frequency contents. This also applies to the other two remaining .WAV files.

domain, considering the degree of importance of the wavelet coefficients in each sub-band. For example, if using Q at the third (top) level, then it is lowered to $Q/2$ and $Q/4$, on second and first level, respectively.

To evaluate the CoDec we measure the compression ratio, expressed as the ratio of the original file and the output stream (or file) and the signal to noise ratio, given by

$$SNR_{[dB]} = 20\log\left(\frac{A_{signal}}{A_{noise}}\right) \qquad (6)$$

where A is the root-mean square value.

At this stage one must be aware that the dictionary used in Huffman coding block is embedded on the output stream.

Experimental results of the compression ratio are presented in tables II and III, for DWT using the 9/7 filter pair and for the DB2 wavelet, respectively, both for different values of the quantization step Q, on the DWT domain. The compression ratio varies from 1.13 up to 4.60, increasing almost linearly for values of Q from 16 to 64, being no longer linear outside that range. Compression ratio is also better for voice audio signal rather than for music, because there is more correlation in voice signals. From tables II and III one can also notice that, in general, slightly better compression ratios are achieved for the 9/7 DWT with reference to the same conditions with the DB2 DWT.

TABEL II

COMPRESSION RATIO RELATIVE TO Q FOR DWT WITH 9/7 FILTER PAIR

	Q	128	96	64	32	16
audio file						
'voice.wav'		4.60	3.95	3.35	3.00	2.82
'flute.wav'		2.13	1.82	1.44	1.30	1.23
'dontfool.wav'		2.36	1.97	1.56	1.42	1.30
'heaven.wav'		2.42	1.98	1.57	1.39	1.28
'toomuch.wav'		1.99	1.64	1.32	1.22	1.13

TABEL III

COMPRESSION RATIO RELATIVE TO Q FOR DWT WITH DB2 WAVELET

	Q	128	96	64	32	16
audio file						
'voice.wav'		4.56	3.83	3.15	2.92	2.68
'flute.wav'		2.06	1.69	1.38	1.28	1.21
'dontfool.wav'		2.34	1.93	1.53	1.37	1.30
'heaven.wav'		2.39	1.94	1.53	1.37	1.26
'toomuch.wav'		2.01	1.65	1.35	1.27	1.12

Tables IV and V show the values of SNR, in dB, for different values of Q and for the DWT computed using the 9/7 filter pair and the DB2 wavelet and figures 4 e 5 show the variation on these values. For values of Q from 16 to 64, SNR increases roughly 6 dB as Q changes to the half. The change on SNR for $128 \le Q < 64$ is around 10 dB as Q is scaled by a factor of 2. The SNR values for the case of the 9/7 DWT, are in general 2 dB better than those obtained for the DB2 DWT case. Observing the curves of figures 4 and 6 one can also notice that SNR varies almost linearly for $64 \le Q \le 16$, which is coherent with the values obtained for the compression ratio.

TABEL IV

VALUES OF SNR [dB] RELATIVE TO Q FOR DWT WITH 9/7 FILTER PAIR

	Q	128	96	64	32	16
audio file						
'voice.wav'		25.44	34.98	45.89	52.05	56.35
'flute.wav'		13.22	20.52	31.42	39.01	44.54
'dontfool.wav'		27.27	36.16	46.33	53.48	59.56
'heaven.wav'		21.55	31.25	42.37	50.71	56.55
'toomuch.wav'		23.16	32.88	44.93	52.91	59.03

TABEL V

VALUES OF SNR [dB] RELATIVE TO Q FOR DWT WITH DB2 WAVELET

	Q	128	96	64	32	16
audio file						
'voice.wav'		24.57	34.01	44.41	50.29	54.57
'flute.wav'		11.81	21.38	32.98	40.61	46.64
'dontfool.wav'		26.38	34.99	45.55	52.09	57.74
'heaven.wav'		20.81	30.14	41.68	49.66	54.92
'toomuch.wav'		20.23	29.78	41.79	49.56	55.85

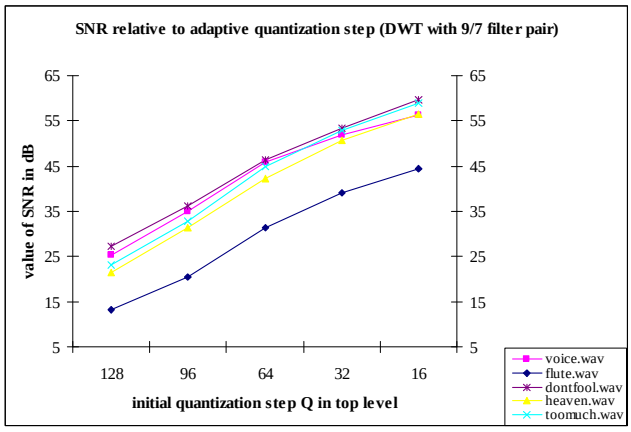

Fig .4. Variation of SNR relative to Q for 9/7 DWT

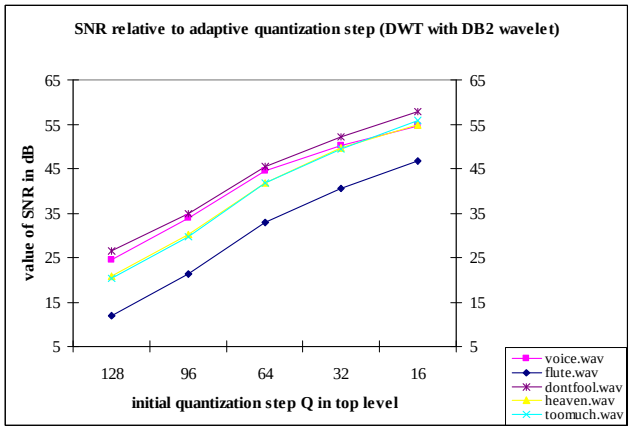

Fig .5. Variation of SNR relative to Q for DB2 DWT

The degradation on signals due to the coding process can also be noticed by observing the changes on the amplitude spectrum and power spectrum of the coded signals with respect to the original one. This can be observed in figures 6 to 8, for the case of "voice.wav" and figures 9 to 11 for the case of "heaven.wav". In both cases the original spectrum, and the spectrum of the recovered signals obtained by coding the original one with Q=16 and Q=128 (limit values), using the 9/7 DWT.

Observing spectrum of figure 8 with respect to that of figure 6, for signal “voice.wav”, one can notice the significant attenuation at higher frequency components, in particular for frequencies above 6 kHz. The same applies for signal “heaven.wav” as can be seen in figures 9 to 11. The degradation on signals for values *Q* greater than 64 is significant so their quality if very poor.

This is also the results of the subjective tests carried, that also confirm a better audible quality for the cases of files coded using the 9/7 DWT.

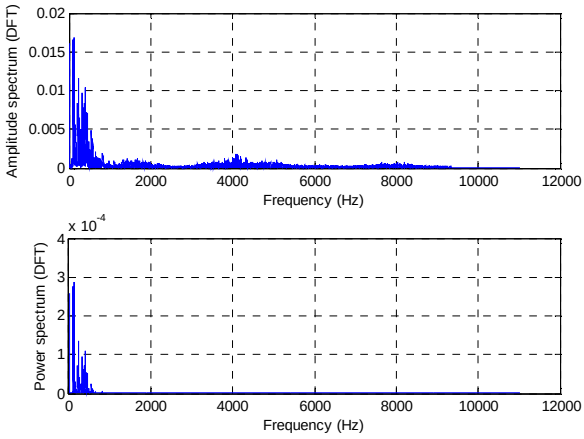

Fig .6. Amplitude and power spectrum of original ‘voice.wav’

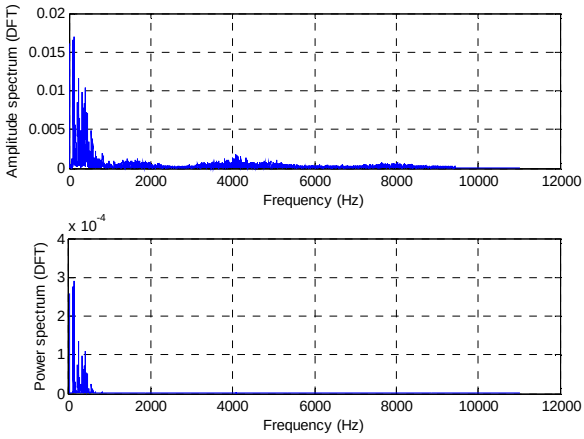

Fig .7. Amplitude and power spectrum of recovered file relative for ‘voice.wav’ coded at *Q*=16

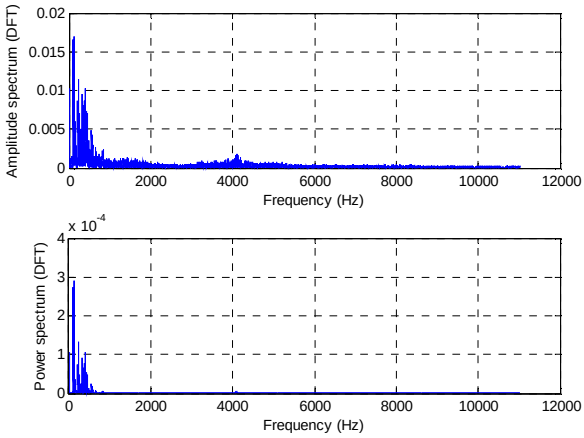

Fig .8. Amplitude and power spectrum of recovered file relative for ‘voice.wav’ coded at *Q*=128

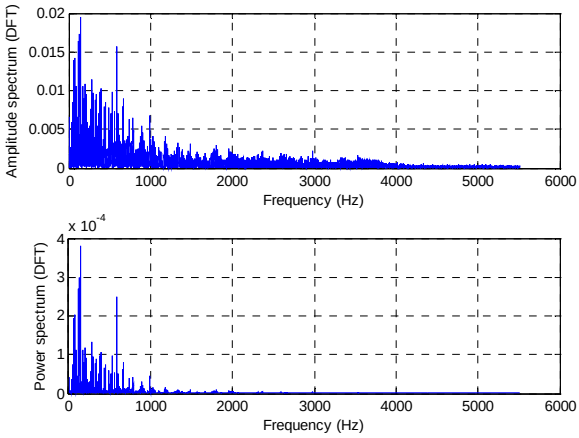

Fig .9. Amplitude and power spectrum of original ‘heaven.wav’

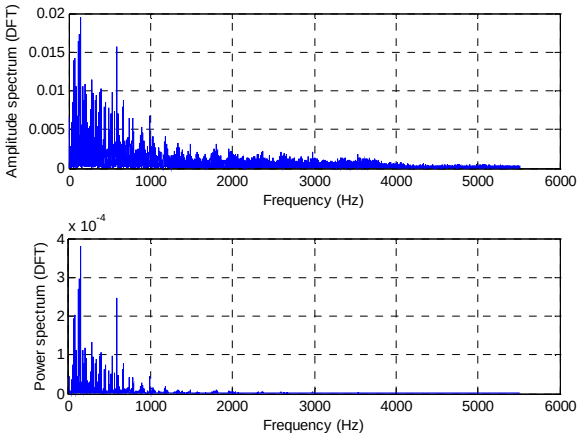

Fig .10. Amplitude and power spectrum of recovered file relative for ‘heaven.wav’ coded at *Q*=16

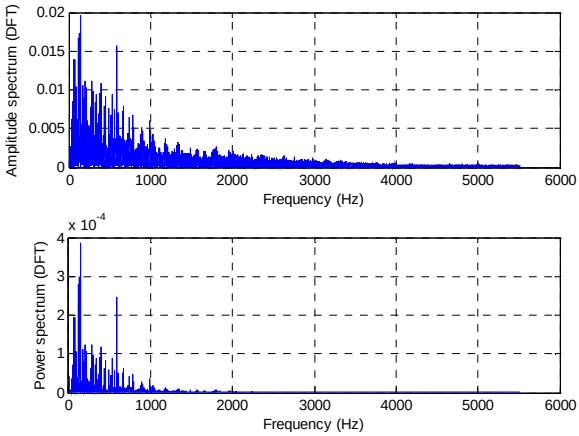

Fig .11. Amplitude and power spectrum of recovered file relative for ‘heaven.wav’ coded at *Q*=128

5. Conclusions

A simple audio coding and decoding scheme based on the discrete wavelet transform is presented in this paper. The CoDec was developed in C/C++ language, using floating point data representation, and implemented in a PC equipped with an Intel Pentium III processor at 1700 MHz, with Windows XP operating system. The CoDec was evaluated for DWT using the 9/7 filter pair and for DB2

wavelet, and for different conditions of the adaptive quantization. Experimental results demonstrate that in general there is higher compression ratios and higher quality when using the 9/7 DWT.
Spectrum analysis and subjective tests also tend to confirm the objective evaluation but also demonstrate that quality is very poor for values of Q greater than 64.

Acknowledgments

This work was carried out during a short term internship at EST-IPCB, Portugal, from May to July 2007, whose report was later submitted to the Indian Institute of Technology, Roorkee, India.

References

[1] A. Spanias, T. Painter, V. Atti, *Audio Signal Processing and Coding*, Wiley–Interscience, 2007, ISBN 978-0-471-79147-8.

[2] Michael Blackstock, "Summary of a Tutorial on MPEG Audio Compression" available on line at http://www.cs.ubc.ca/~michael/MPEG_audio/MPEG_audio_summary.htm

[3] D. Huffman, "A Method for the Construction of Minimum Redundancy Codes" *Proceedings of the I.R.E.* pp. 1098-1101, September 1952.

[4] Vikas K. Gupta, Prashant K. Khevariya, *A Sub-band Audio CODEC using the DWT*, internal project report, IIT–Roorkee, EST-IPCB, 2007.

[5] I. Daubechies, Ten Lectures on Wavelets, SIAM – Society for Industrial and Applied Mathematics, 1992.

[6] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation" *IEEE Trans. On Pattern Analysis and Machine Intelligence.* vol. 7, pp. 674-693, 1989.

[7] I. Daubechies, W. Sweldens, "Factoring Wavelet Transforms into Lifting Steps" *Jean Fourier Analysis and Applications*, vol. 4, Nr3, pp. 247-269, 1998.