SPEECH SIGNAL PROCESSING: SOME ASPECTS

THESIS

Submitted to

THE UNIVERSITY OF CALICUT

in fulfilment of the requirement for the award of the degree of

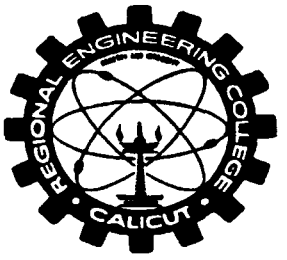
DOCTOR OF PHILOSOPHY

in

ELECTRONICS ENGINEERING

by

P. S. SATHIDEVI



Department of Electronics Engineering

Calicut Regional Engineering College

CALICUT - 673601, KERALA

AUGUST 2001

Certificate

This is to certify that the thesis entitled **"Speech Signal Processing : Some Aspects"** being submitted by **P. S. Sathidevi** to the University of Calicut, for the award of the degree of **Doctor of Philosophy,** is a record of the bonafide research work carried out by her under my supervision and guidance. The results contained in this thesis have not been submitted to any other University or Institute for the award of any Degree or Diploma.

Y. Venkataramani

Date: 11 - 08 - 2001

Dr. Y. Venkataramani
Professor (Retd.)
Department of Electronics Engineering
Calicut Regional Engineering College
Calicut – 673 601

Dept. of Electronics Engg. Calicut Regional Engineering College Calicut - 673601 Kerala

Forwarded to the University for Evaluation

17/8/2001

Dedicated

To

My Parents

ACKNOWLEDGEMENTS

I express my profound gratitude to my guide **Dr. Y. Venkataramani**, Former Professor, Department of Electronics Engineering, for his invaluable guidance, tremendous support, advice, encouragement and constructive suggestions throughout the research work. I shall always remain grateful to him.

I express my sincere thanks to **Dr. E. Gopinathan**, Professor and Head, Department of Electronics Engineering, for providing all the helps, facilities and encouragement rendered for the completion of the research work.

I wish to express my sincere thanks to **Dr. M.P. Chandrasekharan**, Principal, Calicut Regional Engineering College, for providing all the requisite facilities without which the research work would not have been completed. I express my heartfelt thanks to **Dr. B. N. Nagaraj**, Dean (Academic) for all the helps rendered during the period of this research work.

I wish to thank **Dr. P. Janardhanan,** Professor, Department of Electronics Engineering, **Dr. V. K. Govindan**, Professor and Head, Department of Computer Engineering, and **Dr. M. Chandramohan**, Professor and Head, Department of Mathematics, the members of the Doctoral Committee for reviewing the work and extending valuable suggestions regarding my research work.

I gratefully acknowledge **Dr. Rajalakshmi,** Lecturer, Audiology department, All India Institute of Speech and Hearing, Mysore for helping me in conducting clinical tests on hearing impaired subjects. I will never forget the assistance rendered by Dr. Rajalaksmi without which the hearing compensation algorithm would not have been implemented successfully. I wish to thank the Director, AIISH and all faculty members of Audiology department, AIISH for providing me all the helps and facilities for data

collection and clinical experiments. I also acknowledge the individuals who participated in the experiments.

I express my sincere thanks to **Sri. Sasidharan**, Audiologist, Baby Memorial Hospital, Calicut, Kerala and **Dr. M. Jacob**, E.N.T. Surgeon, for their valuable suggestions which helped me a lot in developing hearing compensation techniques.

I would like to thank **Dr. Saly George, Dr. Elizabeth Elias, Smt. Laila. P. Daniel, Sri. R. Suresh** and **Sri. Sanjay Kumar** for helping me in one way or other. I thank my students **Deleep. R. Nair, Sijo. N. Lukose** and **Alok. G. Singh** for all the help I received from them.

I am thankful to all my colleagues in the Department of Electronics Engineering and the Department of Electrical Engineering for their encouragement during the course of the research work.

The partial financial support received from Ministry of Human Resources and Development, Govt. of India, under the scheme of R&D projects is gratefully acknowledged.

Finally, I am very happy to appreciate my husband **C. Muraleedharan** and children **Aparna** and **Gopal** for their encouragement and wholehearted co-operation for the completion of the research work.

P. S. Sathidevi

ABSTRACT

Speech signal processing is an active research area in the field of digital signal processing. Some of the important aspects of digital speech processing are high quality coding (perceptual coding) of speech and audio, speech recognition, enhancement and modification of speech and audio, text-to-speech synthesis, speech-to-text synthesis *etc.* The first step in the processing of any signal is signal analysis. A well known signal analysis tool is Fourier Transform. This transforms a signal in the time domain to a signal in the frequency domain. The serious drawback of Fourier analysis is that while transforming a signal into the frequency domain, the time information will be lost. Hence, Fourier analysis is not suitable for analyzing non-stationary signals. Consequently various techniques have been developed for time-frequency analysis of non-stationary signals. Short Time Fourier Transform (STFT) is one such technique. Here, signal is windowed into small sections and then each section is analyzed using Fourier transform. In this transform, the window size is same for all frequencies and hence every spectral component will be resolved equally.

The Wavelet Transform is the most recent solution to overcome the shortcomings of the Fourier Transform. In wavelet analysis, window size is not same for all frequencies. Here, high frequency components are analysed with narrow windows to give good time resolution and low frequency components are analysed with wider windows to give good frequency resolution. This approach makes sense especially when the signal at hand has high frequency components for short durations and low frequency components for long durations. Fortunately, signals like speech and audio that are encountered in practical applications are of this type.

This research work is mainly focused on the application of wavelet analysis in the following major fields of digital speech processing.

1. Perceptual coding (high quality compression) of speech and audio signals for teleconferencing and multimedia applications.

2. Enhancement and modification of speech and audio for the hearing impaired people (especially for patients with sensorineural hearing loss).

Perceptual coding of Speech and Audio: A particularity of speech is that a good production model can be identified. The vocal cords produce an excitation function which can be roughly classified into voiced and unvoiced excitation. The vocal tract, mouth, and lips act as a filter on this excitation signal. Therefore, very high compression systems for speech are based on identifying the parameters of this speech production model. At the decoder, the speech is synthesized following the production model and using the parameters identified at the encoder. This approach leads to very high compression ratios. Certain applications require speech compression with greater than telephone quality (for example, audio conferencing). This is often called wideband speech since the sampling rate is raised from 8 kHz to 11–22 kHz. Due to a desire for high quality, more attention is focused on the perception process, since the goal is to attain a perceptually transparent coding. In this sense, wideband speech coding is similar to audio coding with only one difference that delay constraint is stringent for real time interactive speech compression, while being relaxed in the audio compression case, since the latter is usually performed off line.

Perceptual audio coding is a lossy compression scheme, that minimizes the number of bits required to represent audio signals, while maintaining high fidelity. This is accomplished through two processes, namely, irrelevancy reduction and redundancy removal. Irrelevancy reduction is achieved by shaping the coding distortion (quantization noise) such that it cannot be perceived by the human ear. That is, by keeping the quantization noise just below the level (masking threshold) where it would become noticeable. Perceptually relevant signal components are accurately represented using a large number of bits. Perceptually unimportant or imperceptible components, in contrast, are represented using very few bits and in some cases are altogether discarded. Moving Picture Experts Group (MPEG) audio coding algorithm is the most important international standard, currently in use, for perceptual coding of digital audio.

In this research work, following wavelet based perceptual audio coding schemes of varying computational complexity, quality and compression ratios, as alternatives to ISO/MPEG international audio coding standard are developed.

1. Discrete Wavelet Transform (DWT) based Audio Coder. **2.** Discrete Wavelet Packet (DWP) based Audio Coder. **3.** Computationally Efficient (Low Complexity and Cost Effective) Wavelet Packet based Audio Coder. **4.** Low Complexity Scalable (which supports most of the industrial audio sampling frequencies) Wavelet Packet Based Audio Coder and **5.** Fixed or Constant Bit Rate Wavelet Packet Based Audio Coder. SNR (Signal-to-Noise Ratio) scalability is also incorporated in scheme 5 using two stages of compression. Discrete Wavelet Transform and Discrete Wavelet Packets are implemented using filterbanks with high energy compaction properties.

The performance of all the above schemes are validated through subjective listening tests. The Compact Disc (CD) attached with this thesis, contains, some of the original audio signals and, the signals encoded and reconstructed using various schemes proposed in the present work. A model (psychoacoustic model) of the human auditory system (cochlea) is incorporated in these coders, to identify the perceptually relevant and irrelevant parts of the signal. Simultaneous masking properties of the human ear, in the frequency domain, is included in these schemes to place quantization noise in the least sensitive regions of the spectrum.

A serious artifact of MPEG standard (which is based on Discrete Cosine Transform) due to the use of fixed time-frequency resolution analysis windows, known as *"Pre-echo distortion",* is almost eliminated in all the above schemes since high frequency components (like transients) are analyzed with narrow windows (good time resolution) and low frequency components are analyzed with wide windows (good frequency resolution). That is, in contrast to DCT, wavelet transform provides a non-uniform time resolution. In wavelet filtering, the impulse response of high frequency filters are short, while low frequency filters have a long impulse response. Hence, pre-echo problem is eliminated by wavelet filtering. The performance of the above schemes are tested using various wavelet families listed below.

1. Haar wavelet 2. Daubechies wavelet 3. Symlet wavelet 4. Coiflet wavelet and 5. Biorthogonal Spline wavelet.

The above wavelet families are specifically chosen because their properties are different in terms of regularity, compact support, frequency selectivity, linear phase etc. One of the objectives of the present work was to find which features of a wavelet are more relevant in the case of perceptual audio compression and hence to determine the best wavelet basis. But it was found that the above properties are less relevant as far as perceptual audio coding is concerned.

A novel theoretical result, "the compression ratio not only depends on the transform (or basis function), but also depends on the statistical features of the source signal" is experimentally proved here. Three different optimization methods are developed in this thesis, to select the optimum wavelet basis from a predefined library of wavelets, according to the time varying characteristics of the signal. Hence, each frame (of size 512 samples) of the audio signal is represented by the optimum wavelet basis such that compression ratio is maximum.

In the case of stationary audio segments, as expected, sinusoidal basis gave more compact representation than wavelet basis. This is because the low order Finite Impulse Response analysis filters employed in discrete wavelet transform and discrete wavelet packet decompositions are characterised by poor frequency selectivity. For a stationary signal, masking thresholds will be quite localized in frequency and hence calls for fine frequency resolution and coarse time resolution. Therefore wavelet bases will not provide compact representations for stationary signal segments. In the case of non stationary segments, using wavelet analysis, the high frequency energy of the signal is localized to a smaller portion of time. This is useful for preserving the transient like segments in the audio signals. In cases where this does not happen, quantization noise produced in the frequency domain will be transferred to the time domain as pre-echo noise and become audible. Hence, a switching algorithm is also developed as a part of the present work, to switch between sinusoidal basis (DCT) and optimum wavelet basis (DWT/DWP) according to the characteristics of the signal frame (stationary/non-stationary).

The compression ratios obtained with the proposed audio coding schemes are comparable to ISO/MPEG audio coding standards (between 6 and 12) with almost transparent quality (Mean Opinion Score, MOS > 4.0). All these schemes are implemented using 1) Scalar quantization with dynamic bit allocation (as in MPEG), 2) A new vector quantization scheme (named as "Hit book method") in which length of the code book can be adaptively changed according to the psychoacoustic model requirement and 3) Combined scalar and vector quantization. Maximum compression is achieved with the 'Hit book' method with little degradation in quality.

One of the above coding schemes can be suitably chosen by a user, according to the available bandwidth of the channel, preferred application, the processing power available, quality to be met *etc.*

Enhancement and Modification of speech and audio for the hearing impaired: Approximately 7.5 percent of the population has some degree of hearing loss and about 1.0 percent has a loss that is moderately severe or greater. A majority of the hearing impaired population would benefit significantly from improved methods of acoustic amplification. Two major types of hearing loss are conductive deafness and sensorineural deafness. Conductive deafness is caused by degraded transmission of the acoustic energy to the sense organ (cochlea) and can be modeled by a linear distortion. Hence this can be adequately compensated by using normally available analog hearing aids. On the other hand, sensorineural deafness is caused by abnormal function of the cochlea, the auditory nerve or both. This impairment is mainly caused by prolonged or excessive exposure to noise, aging, hereditary factors etc. Recruitment (growth or increase) of loudness commonly occurs in conjunction with sensorineural hearing impairments. Loudness perception varies from person to person and with frequency. If the patient suffers from recruitment of loudness, the perceived loudness grows more rapidly with an increase in sound intensity than it does in the normal ear. Threshold of discomfort level can be the same or even lowered. Hence, in the case of such patients, some frequency bands may need amplification, some frequency bands may not need any amplification and some may need even attenuation.

In this research work, two new digital hearing aid algorithms (hearing compensation methods) - one using Discrete Wavelet Transform (DWT) and the other using Discrete Wavelet Packet (DWP) – according to the two common audiogram standards (ISO) are developed and implemented. In these schemes, first the signal is decomposed into the required number of frequency bands (matching to common audiogram standards) using wavelet analysis. Then, according to the patient's hearing loss in each frequency band and the intensity of the signal in that band, gain factor/attenuation factor is determined and the wavelet coefficients for each band are correspondingly modified. The enhanced signal for the patient is reconstructed by taking the inverse wavelet transform of the modified coefficients.

Simulated digital hearing aid using discrete wavelet transform technique is validated through clinical tests on twelve hearing impaired subjects, at *All India Institute of Speech and Hearing (AIISH), Mysore, Karnataka, INDIA.*

Background noise is a major problem with hearing aids. The noise will not only mask consonants but also its amplification is distracting and often painful. Hence, a wavelet based denoising technique is also incorporated in this algorithm which serves as a pre-processor for the frequency dependent compensation for hearing impairments. The wavelet coefficients resulting from the denoising stage are used for the gain calculations.

Advantages of the hearing compensation algorithm developed in the present work are,

(i) The calculation of the gain is customized to the common audiogram standards.

(ii) The gain for consonants and stops, which contain predominantly high frequencies will be calculated from a smaller window than vowels (which contain more low frequencies), while other multichannel compression schemes use the same window size for each frequency band.

(iii) The gain for each frequency band is calculated according to the intensity of the signal in each frequency band and the hearing loss of the patient in the respective frequency band.

Clinical Test results : All subjects rated the quality of the perceived signal with the new digital hearing aid (using DWT), well above that of analog hearing aid, in the *presence of noise* for majority of the audio signals tested. The performance of this new aid *in quiet* was rated equally with that of analog hearing aid, by most of the subjects. But two subjects rated this new aid well above the analog hearing aid, *in quiet as well as in the presence of noise*. The proposed wavelet approach to the combined problem of noise reduction and frequency dependent compensation for hearing impairments, offers high quality and flexibility since the parameters can be modified to fit the individual hearing loss and intensity of time varying signal characteristics. Clinical test results reveal that wavelet analysis is a promising tool for the design of efficient digital hearing aids.

CONTENTS

Chapter 8. CONCLUSIONS

LIST OF FIGURES

LIST OF TABLES

PRINCIPAL SYMBOLS AND ABBREVIATIONS

AGC	automatic gain control
b_k	high pass filter coefficients
BM	basilar membrane
BPN	back propagation network
c_k	low pass filter coefficients
coifN	coiflet wavelet of order N
CR	compression ratio
CWT	continuous wavelet transform
dbN	daubechies wavelet of order N
DCT	discrete cosine transform
DFT	discrete fourier transform
DWP	discrete wavelet packets
DWT	discrete wavelet transform
EY	entropy
FFT	fast fourier transform
MDCT	modified discrete cosine transform
MOS	mean opinion score
MPEG	moving pictures experts group
NMR	noise-to-mask ratio
OHC	outer hair cells
QMF	quadrature mirror filter
$R_{xx}(k)$	auto correlation
symN	symlet wavelet of order N
S	skewness
SMR	signal-to-mask ratio
SNR	signal-to-noise ratio
SPL	sound pressure level
SSNR	segmental signal-to-noise ratio
STFT	short time fourier transform

$T_g(i)$	global masking threshold
$T_q(f)$	absolute threshold of hearing in quiet
UCL	un - comfortable level
VQ	vector quantization
Z(n)	zero crossing rate
ϕ (t)	scaling function
μ	mean
σ	std.deviation
ψ (t)	wavelet function

Chapter 1

INTRODUCTION

1.1 Introduction

Audio Coding or audio compression algorithms are used to obtain compact digital representations of audio signals for the purpose of efficient transmission or storage. The central objective in audio coding [1 - 5] is to represent the signal with minimum number of bits while achieving transparent signal reproduction, *i.e.,* generating output audio that cannot be distinguished from the original input, even by a sensitive listener. Typical audio signal classes are telephone speech, wideband speech, and wideband audio. These differ in bandwidth, dynamic range, and in listener expectation of offered quality. The quality of the telephone-bandwidth speech is acceptable for telephony and for some video telephony services. Higher bandwidths (7 kHz for wideband speech) may be necessary to improve the intelligibility and naturalness of speech. Wideband (high fidelity) audio representation including multichannel audio needs a bandwidth of at least 20 kHz. The conventional digital format for these signals is Pulse Code Modulation (PCM).

Although high bit-rate channels and networks have become more easily accessible, low bit-rate coding of audio signals has retained its importance. The main motivations for low bit-rate coding are the need to minimize transmission costs or to provide cost-effective storage, the demand to transmit over channels of limited capacity such as mobile radio channels, and to support variable-rate coding in packet oriented networks.

We have seen rapid progress in bit-rate compression techniques for speech and audio signals [5-15]. Linear Prediction [7,8,9], Subband coding [10,12,13], Transform coding [10,12,13], as well as various forms of Scalar and Vector

Quantization [5,16], and Entropy coding [16,17] techniques have been used to design efficient coding algorithms that can achieve substantially more compression than was thought possible only a few years ago. Recent results in speech and audio coding [1,13] indicate that an excellent coding quality can be obtained with bit rates of 0.5 to 1 b/sample for speech and wide band speech and 1 to 2 b/sample for hi-fi audio. Speech and audio coding are similar in that, in both cases, quality is based on the properties of human auditory perception [15]. A good speech production model [8,9,18] is available. Hence, low bit rate coding of speech using these model parameters is possible. However no such model exists for audio signals. This work is focussed on *perceptual coding (high quality coding) of speech and audio*, in which a model of the human auditory system [19,20] is incorporated.

The basic task of a perceptual audio coding [1,21,22,23] system is to compress the digital audio data in a way that

- the compression is as efficient as possible, i.e., the compressed file is as small as possible and
- the reconstructed (decoded) audio sounds exactly (or as close as possible) to the original audio before compression.

Other requirements for audio compression techniques include low complexity and flexibility for different application scenarios.

Perceptual audio coding uses knowledge from *psychoacoustics* to reach the target of efficient but inaudible compression. Perceptual encoding is a lossy compression technique, i.e., the decoded file is not a bit-exact replica of the original digital data, while maintaining high fidelity. This is accomplished through two processes, namely, irrelevancy reduction and redundancy removal. Irrelevancy reduction is achieved by shaping the coding distortion (quantization noise) such that it cannot be perceived. The idea is to distribute quantization noise in both time and frequency such that it is not detectable by the human auditory system. Perceptually relevant signal components are accurately represented using a large number of bits, while perceptually unimportant or imperceptible components, in contrast, are represented using very few bits and in some cases are altogether discarded. For example, frequency components that fall below the threshold of hearing can be safely

discarded. The irrelevancy reduction/distortion control step is central to the success of perceptual coding schemes. Redundancy removal is also of vital importance to the perceptual coder. Whereas irrelevancy reduction exploits the properties of auditory perception, redundancy removal identifies and removes statistical redundancies [1,16,17]. Redundancy removal is typically applied after irrelevancy reduction.

A generic perceptual audio coder [1,21,22,23] is shown in Fig.1.1. It consists of the following building blocks.

- Filter bank - A filter bank is used to decompose the input signal into subsampled spectral components (time/frequency mapping). Together with the corresponding filter in the decoder it forms an analysis/synthesis system.
- Perceptual Model (Psychoacoustic model) - Using either time domain input signal and/or the output of the analysis filter bank, an estimate of the actual (time and frequency dependent) masking threshold (this is the level at which distortion becomes just noticeable) is computed using rules known from psychoacoustics. This is called psychoacoustic model of the perceptual encoding system.

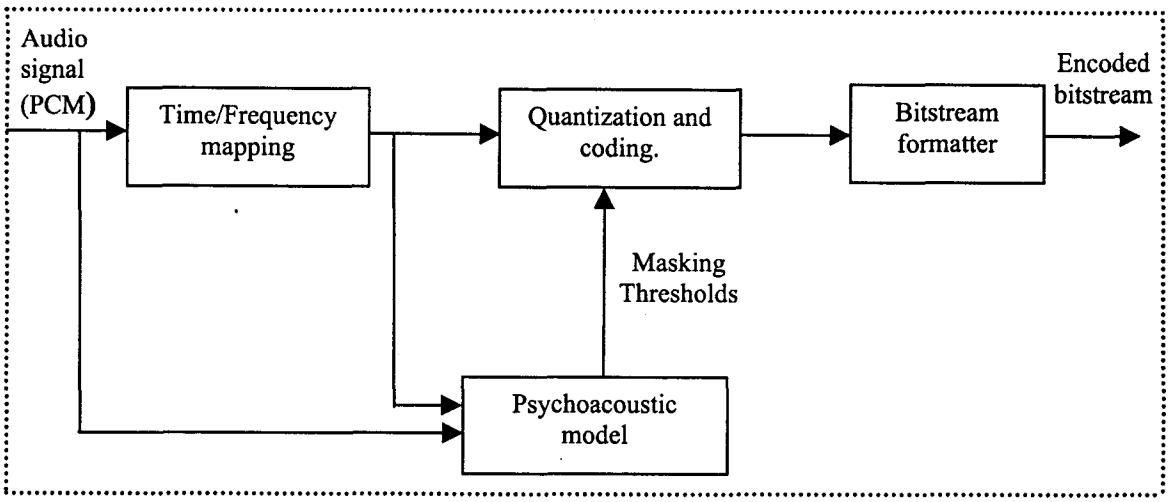


Fig. 1.1: A generic perceptual audio coder

- Quantization and Coding - The spectral components are quantized and coded with the aim of keeping the noise, which is introduced by quantization, below

the masking threshold. This quantization step can be done in different ways. Additional noiseless (entropy) coding can be done after quantization.

- Encoding of Bit stream - A bit stream formatter is used to assemble the bit stream, which typically consists of the quantized and coded spectral coefficients and some side information, e.g., bit allocation information.

Considerable research has been devoted to the development of algorithms for perceptually transparent coding of high fidelity digital audio. As a result, many algorithms have been proposed, and several have now been international and/or commercial product standards. Next section reviews algorithms for perceptual audio coding including research and standardization activities.

1.2 Literature Review

A survey of several different classes of audio coders is discussed in [1]. Audio coders in general are either transform coders or subband coders. Painter and Spanias [1] also considered coders based on sinusoidal signal models as well as based on linear prediction.

1.2.1 Transform Coders

Transform coding algorithms for high - fidelity audio make use of unitary transforms for the time/frequency analysis section in Figure 1.1. These algorithms typically achieve high resolution spectral estimates at the expense of adequate temporal resolution. Many transform coding schemes for wide band speech and high-fidelity audio have been proposed.

In the mid-1980's, Krahe applied psychoacoustic bit allocation principles to a transform coding scheme [24]. Schroeder later extended these ideas into multiple adaptive spectral coding (MSC) [25]. The MSC utilizes a 1024 -point DFT, and then groups coefficients into 26 subbands. DFT magnitude and phase components are quantized and encoded in a two-step successive refinement procedure that relies upon a perceptual bit allocation. Schroeder reported nearly transparent coding of CD quality audio at 132kb/s.

Brandenburg in 1987 proposed a 132 kb/s algorithm known as Optimum Coding in the Frequency domain (OCF) [26], which is in some respects similar to the well-known " Adaptive Transform Coder" (ATC) [13] for speech. OCF works as follows: The input signal is first buffered into 512 sample blocks and transformed to the frequency domain using DCT. Next, transform components are quantized and entropy coded. A single quantizer is used for all transform components. Adaptive quantization and entropy coding work together in an iterative procedure to achieve a fixed bit-rate.

Brandenburg in 1988 reported an enhanced OCF (OCF-2) [27], which achieved subjective quality improvements at a reduced bit rate of only 110 kb/s. The improvements were realized by replacing the DCT with modified DCT (MDCT) and adding a pre-echo detection compression scheme. OCF-2 contains the first reported application of the MDCT to audio coding. The 50% time overlap associated with MDCT increases the effective time resolution and consequently improves the reconstruction quality.

In 1988 itself, Brandenburg reported further OCF improvements (OCF-3) [28] in which better quality was realized at a lower bit-rate (64 kb/s) with reduced complexity. This was achieved through differential coding of spectral components, an enhanced psychoacoustic model, and an improved rate-distortion loop.

Johnston developed several DFT based transform coders for audio during the late 1980's. Johnston's work in perceptual entropy forms the basis for a transform coder reported in 1988 [22] that achieves transparent coding of FM quality monaural audio signals. The idea behind the perceptual transform coder (PXFM) is to estimate the amount of quantization noise that can be inaudibly injected into each transform domain subband using Perceptual Entropy (PE) estimates. The coder works as follows: The signal is first windowed into overlapping segments and transformed using a 2048-point FFT. Just Noticeable Distortion (JND) thresholds are estimated for each critical band. Then, an iterative quantization loop adapts a set of 128 subband quantizers to satisfy the JND thresholds until the fixed bit rate is achieved. Finally quantization and bit packing are performed.

Johnston and Brandenburg collaborated in 1990 to produce a hybrid coder, known as AT&T hybrid coder [29] that uses subband and transform coding algorithms. The idea behind this hybrid coder is to improve time and frequency resolution relative to OCF and PXFM by constructing a filter bank that more closely resembled the auditory filter bank. This is accomplished at the encoder by first splitting the input signal into four octave-width subbands using a QMF filter bank. The decimated output sequence from each subband is then followed by one or more transforms to achieve the desired time/frequency resolution. Both DFT and MDCT methods were investigated.

In 1989, Mahieux et al. [30] proposed a DFT based audio coding system, that introduced a novel scheme (CNET coder), to exploit DFT inter block redundancy. Nearly transparent quality was reported for 15 kHz (FM-grade) audio at 96 kb/s.

In 1990, Mahieux and Petit reported on the development of a MDCT based transform coder for which they reported transparent CD quality at 64 kb/s [31]. This algorithm introduced a novel " spectrum descriptor" scheme for representing the power spectral envelope. The coder was reported to perform well for broad band signals with many harmonics but had some problems in the case of spectrally flat signals. More recently, Mahieux and Petit enhanced their 64 kb/s algorithm by incorporating a sophisticated pre-echo distortion and post-filtering scheme [1].

The MSC, OCF, PXFM, AT&T hybrid and CNET audio transform coders were eventually clustered into a single proposal by the ISO/IEC JTC1/SC2 WG11 committee. As a result, Schroeder, Brandenburg, Johnston, Herre, and Mahieux collaborated in 1991 to propose a flexible coding algorithm, ASPEC (Adaptive Spectral Entropy Coding), which incorporated the best features of each coder in the group, for acceptance as the new MPEG audio compression standard [23]. ASPEC was claimed to produce better quality than any of the individual coders at 64 kb/s.

Paraskevas and Mourjopoulos [32] have reported on a Differential Perceptual Audio Coder (DPAC), which makes use of a novel scheme for exploiting long-term correlations. The authors have concluded that DPAC may be preferable to algorithms such as PXFM for low bit-rate non-transparent applications.

The algorithms described, thus far, rely upon scalar quantization of transform coefficients. Vector Quantization (VQ) has also been applied to transform coding of audio, although on a much more limited scale.

Gersho and Chan investigated VQ schemes for coding DCT coefficients subject to a constraint of minimum perceptual distortion. They have reported on a variable rate coder [33], which achieves high quality in the range of 55-106 kb/s for audio sequences band limited to 15 kHz (32 kHz sampling rate). After computing the DCT on 512 sample blocks, the algorithm utilizes a novel multistage tree structured VQ (MSTVQ) scheme for quantization of normalized vectors, with each vector containing four DCT coefficients.

Gersho and Chan later enhanced their algorithm by improving the quantization of transform coefficients. In this new approach [34] constrained - storage VQ (CS-VQ) techniques are combined with the MST VQ (CS-MSTVQ) techniques from the original coder. Relative to their first VQ/DCT coder, the authors reported savings of 10-20 kb/s with no reduction in quality due to CS-VQ.

Iwakami et al. developed transform domain weighted interleave vector quantization (TWIN- VQ), an MDCT based coder [35], which also involves transform coefficient VQ. This algorithm exploits LPC analysis, spectral interframe redundancy, and interleaved VQ. The authors have claimed higher subjective quality than MPEG-1 Layer II at 64 kb/s for 48 kHz CD quality audio, as well as higher quality than MPEG-1 Layer II for 32 kHz audio at 32 kb/s.

1.2.2 Subband Coders

Subband coders also exploit signal redundancy and psychoacoustic irrelevancy in the frequency domain. Instead of unitary transforms, these coders rely upon frequency domain representations of the signal obtained from banks of band pass filters. The audible frequency spectrum (20 Hz - 20 kHz) is divided into frequency subbands using a bank of band pass filters. The output of each filter is then sampled and encoded. At the receiver, the signals are demultiplexed, decoded, and then

summed to reconstruct the signal. Numerous subband coding algorithms for high fidelity audio have appeared in the literature since the late 1980's.

Theile et al. developed a subband audio coder [36], MASCAM (Masking Pattern Adapted Subband Coding), based upon a tree structured QMF bank that was designed to mimic the human auditory filterbank. The coder has 24 non-uniform subbands, with bandwidths of 125 Hz below 1 kHz, 250 Hz in the range 1-2 kHz, 500 Hz in the range 2-4 kHz, 1 kHz in the range 4-8 kHz, and 2 kHz from 8 to 16 kHz. The prototype filter has 64 taps. Subband output sequences are processed in 2- ms blocks. A normalization scale factor from each subband is quantized and transmitted for each block. Subband bit allocations are derived from a simplified psychoacoustic analysis. The original coder considered only in-band simultaneous masking. Later, as described in [37], interband masking and temporal masking were added to the bit-rate calculation. The MASCAM coder was reported to achieve high quality results for 15 kHz bandwidth input signals at bit rates between 80-100 kb/s per channel.

Veldhuis [38] developed a subband coder at Philips, based on 20 and 26 band non-uniform filter banks. Philips coder relies upon a highly simplified masking model, which considers only the upward spread of simultaneous masking. Given SNR targets due to the masking model, uniform ADPCM is applied to the normalized output of each subband. The Philips coder was claimed to deliver high quality coding of CD quality signals at 110 kb/s for the 26 band version and 180 kb/s for the 20 band version.

MUSICAM (Masking pattern adapted Universal Subband Integrated Coding And Multiplexing) algorithm [39], based upon MASCAM and Philips coders was successful in the 1990 ISO/IEC competition for a new audio coding standard. This formed the basis for MPEG-1 and MPEG-2 audio Layers I & II. MUSICAM makes several practical trade offs between complexity delay and quality. By utilizing a uniform bandwidth, 32 band polyphase filterbank, instead of a tree structured QMF bank, both complexity and delay are greatly reduced relative to MASCAM and Philips coders. These improvements are realized at the expense of using a sub optimal filter bank, however, in the sense that filter bandwidths (constant 750 Hz for 48 kHz sample rate) no longer correspond to the critical bands. Despite these excessive filter

bandwidths at low frequencies, high quality coding is still possible with MUSICAM due to its enhanced psychoacoustic analysis. High resolution spectral estimates are obtained through the use of a 1024 point FFT in parallel with the polyphase filterbank. This parallel structure allows for improved estimation of masking thresholds. The MUSICAM psychoacoustic analysis procedure is essentially the same as the MPEG-1 [1,40,41] psychoacoustic model 1, described in Chapter 2.

Sinha and Tewfik developed a variable-rate wavelet based coding scheme using globally adapted Daubechies wavelet, for which they have reported nearly transparent coding of CD quality audio at 48-64 kb/s [42]. The encoder exploits redundancy using a VQ scheme and irrelevancy using a Wavelet Packet signal decomposition combined with perceptual masking thresholds. In this scheme, the same analysis wavelet is applied to the entire decomposition. The authors reached several useful conclusions regarding the optimal compact support (K-coefficient) wavelet basis when selecting from among the Daubechies orthogonal wavelet bases. The decomposition here is structured such that its 29 frequency subbands roughly correspond to the critical bands of the auditory filterbank.

Pramila Srinivasan and Jamieson [43] have proposed a Wavelet Packet Based audio coding scheme in which a signal specific perceptual best basis is constructed by adapting the WP tree structure on each frame such that perceptual entropy and ultimately the bit rate are minimized. While the tree structure is signal adaptive, the analysis filters are time invariant and obtained from the family of Spline based biorthogonal wavelets. For informal listening tests over coded program material that included violin, flute, sitar, vocals/orchestra and sax, the coded outputs at rates in the vicinity of 45 kb/s were reported to be indistinguishable from the originals with the exceptions of the flute and sax.

Philippe et al. [44] measured the impact on perceptual coding gain, of wavelet regularity, AR (1) coding gain, and filter bank frequency selectivity. The study compared performance among orthogonal Rioul, Orthogonal Onno, and the Biorthogonal Wavelets in a Wavelet Packet coding scheme with 29 bands. Using filters of lengths varying between 4-120 taps, minimum bit-rates required for transparent

coding with the usual perceptual subband bit allocation were measured for each wavelet. For a given filter length, the results suggested that neither regularity nor frequency selectivity mattered significantly. Authors concluded that AR (1) coding gain [12] is a legitimate criterion for WP filters selection in perceptual audio coding schemes.

Hamdy et al. developed a novel hybrid coder [45] designed to exploit the efficiencies of both harmonic and wavelet signal representations. For each analysis frame, the encoder chooses a compact signal representation from combined sinusoidal and wavelet bases. The encoder works as follows: First, Thomson's analysis model is applied to extract sinusoidal parameters for each input frame. Harmonic synthesis using the McAulay and Quatieri reconstruction algorithms for phase and amplitude interpolation is next applied to obtain a residual sequence. Then, the residual is decomposed into WP subbands. This coder was reported to achieve nearly transparent coding over a wide range of CD quality source material at bit rates in the vicinity of 44 kb/s.

Princen and Johnston have developed a high quality audio coder based upon a signal adaptive filter bank [46] for which they have reported quality better than the sophisticated MPEG -1 Layer III algorithm at both 48 and 64 kb/s. The analysis filter bank for this coder consists of a two stage cascade. The first stage is a 48 band non-uniform modulated filter bank split into four uniform bandwidth sections. There are eight uniform subbands from 0 to 750 Hz, four uniform subbands from 750 to 1500 Hz, 12 uniform subbands from 1.5 to 6 kHz, and 24 uniform subbands from 6 to 24 kHz. The second stage in the cascade optionally decomposes non-uniform bank outputs with on/off switchable banks of finer resolution uniform subbands. During filter bank adaptation, a suitable overall time-frequency resolution is attained by selectively enabling or disabling the second stage filters for each of the four uniform band width sections. Uniform PCM is applied to subband samples under the constraint of perceptually masked quantization noise.

1.2.3 Sinusoidal Coders

Although sinusoidal signal models have been applied successfully in speech coding [7,8], relatively little work was reported on perceptual audio coding using sinusoidal models. The existing sinusoidal coders were developed in a speech coding context and tended to minimize Mean Square Error (MSE). Perceptual properties were introduced later [22].

Edler et al. have developed the sinusoidal ASAC (Analysis/ Synthesis Audio Codec) for robust coding of general audio signals at rates between 6 and 24 kb/s [1], at the University of Hannover and proposed for MPEG-4 standardization [47] in 1995. Initially, ASAC segments input audio analysis frames over which the signal is assumed to be nearly stationary. Sinusoidal synthesis parameters are then extracted according to perceptual criteria, quantized, encoded and transmitted to the decoder for synthesis. The algorithm distributes synthesis parameter across basic and enhanced bit streams to allow scalable output quality at bit rates of 6 and 24 kb/s.

Purnhagen et al. at the University of Hannover have developed in conjunction with Deutsche Telekom Berkom, an "Object based algorithm", known as the "Harmonic and Individual Lines plus Noise" (HILN) [48], which is architecturally very similar to ASAC, with some modifications. In the enhanced analysis-synthesis system, harmonic analysis is applied first, followed by individual spectral line analysis, followed by shaped noise modeling of the two stage residual. Results from subjective listening tests at 6 kb/s showed significant improvements for HILN over ASAC.

1.2.4 Linear Prediction Coders

Singhal at Bell Labs [49] reported that analysis- by - synthesis multi - pulse excitation of sufficient pulse density can be applied to correct for LP envelope errors introduced by bandwidth expansion and quantization. This algorithm uses a twenty fourth order LPC synthesis filter while optimizing pulse positions and amplitudes to minimize perceptually weighted reconstruction errors. The proposed MPLPC audio coder achieved about SNR's of 30-40 dB at a bit rate of 128 kb/s.

Boland and Deriche have reported output quality comparable to MPEG-1 Layer II at 128 kb/s for an LPC audio coder operating at 96 kb/s [50] in which the prediction residual was transform coded using a three level DWT based on a four band non uniform filter bank.

Excitation sequences modeled as a sum of sinusoids were investigated [52,53] in order to capitalize on the experimentally observed tendency of the prediction residuals for high - fidelity audio to be spectrally impulsive rather than flat. In coding experiments using 32 kHz sampled input audio, subjective and objective quality improvements relative to the MPLPC coders were reported for the sinusoidal excitation schemes, with high quality output audio reported at 72 kb/s.

1.2.5 Audio Coding Standards

An International Standards Organization / Moving Picture Experts Group (ISO/MPEG) audio coding standard for stereo CD quality audio was adopted in 1992 after four years of extensive collaborative research by audio coding experts world wide. An overview of MPEG audio coding standard [23] is given in Chapter 2.

ATRAC (Adaptive TRansform Acoustic Coding) coding method developed by Sony for use in its Re-writable Mini Disc system [51] makes combined use of subband and transform coding techniques to achieve CD quality 256 kb/s coding of 44.11 kHz stereo 16 bit PCM input data. ATRAC QMF filter bank is followed by adaptive MDCT analysis. MDCT components are quantized and encoded according to a psychoacoustically derived bit allocation.

Dolby Laboratories originally developed 320 kb/s AC-3 [54] perceptual audio coder for High Definition Television (HDTV). The PCM input signal is first windowed using a proprietary function and then segmented into 50% overlapping 10.66ms blocks (512 samples). The block size is reduced to 5.33 ms during transient conditions to compensate for pre-echoes. After segmentation, a modified DCT (MDCT) filter bank with 93.75 Hz frequency resolution is used to decompose the signal. Transform components are quantized using a psychoacoustically derived dynamic bit allocation scheme [1].

Like the MPEG coders, Lucent Technologies Perceptual Audio Coder (PAC) [55] is flexible in that it supports monophonic, stereophonic and multiple channel modes. Depending upon desired quality, PAC supports several bit rates. The PAC system described in [55], achieves very high quality coding of stereophonic inputs at 96 kb/s. PAC relies on MDCT alone to reduce the complexity in the filter bank section.

1.3 Thesis Aims and Objectives

Studies on various perceptual audio coding schemes and standards revealed that transform coders and subband coders can provide almost transparent audio quality. One major artifact of transform coders and subband coders (including ISO/MPEG audio coding standard) employing uniform filterbank for time/frequency analysis is **"Pre-echo distortion"**, which usually occurs when a signal with a sharp attack begins near the end of a transform block immediately following a region of low energy. Quantization noise will spread evenly in time throughout the reconstructed block. This results in unmasked distortion throughout the low energy region preceding the "signal attack" time, at the decoder.

Wavelet Packet (WP) based coders described in the literature are computationally very complex and their performance is not satisfactory for coding of stationary signals or harmonic signals like those from a piano. This is in part because the low order FIR analysis filters typically employed in WP decomposition are characterised by poor frequency selectivity and therefore wavelet bases tend not to provide compact representations for stationary signals. Most of the authors use Daubechies wavelets, even though many other wavelet bases are available. Audio coders based on Discrete Wavelet Transform (DWT) (which is computationally less complex than WP), have not been developed so far. Hence, initially the main objective of this research work was to develop DWT and DWP based audio coding schemes using a number of wavelet families or bases with different properties, to study their performance in perceptual audio coding and then to identify the best wavelet to achieve maximum compression.

Choice of an inappropriate filter bank can result in lower output quality or in a need for higher digital audio bit rates. The properties of the filter bank must be matched to the characteristics of the incoming signal. Impulse response of the filter is decided by the basis function. Hence, three different optimization methods are also developed as part of the present research work, to select the best wavelet basis (that is, the wavelet which gives maximum compression ratio) from a library of wavelets (with different properties), for each audio segment. If a signal is stationary, masking thresholds will be more localized in frequency and hence sinusoidal basis is the optimum basis for its representation. Therefore, a switching algorithm is also developed to switch between sinusoidal/optimum wavelet basis according to the time varying characteristics of the signal.

Most of the audio coding algorithms are based upon scalar quantization. Only very little work has been done using Vector Quantization (VQ). VQ schemes employed in these coders use code books of fixed size. But psychoacoustic model output necessitates the need of a code book whose length can be adaptively varied. Hence, in this research work, a new VQ scheme (named as 'Hit Book Method') in which length of the code book can be adaptively changed according to the psychoacoustic model requirement, is also proposed.

It became clear during the preliminary study that the wavelet transforms had not only a great compression potential for low bit rate coding of audio signals but also that it was amenable to implementation of digital hearing aid algorithms (enhancement and modification of speech and audio for the hearing impaired) as well. Thus the aims and objectives of the present work include:

- the design and implementation of variable bit rate and fixed bit rate DWT and DWP based perceptual audio coding schemes with varying complexity, quality, compression ratio, encoding delay *etc.*, as alternatives to ISO/MPEG audio coding standards.
- development of optimization techniques to identify the most matching wavelet basis for each audio frame, from a predefined library of wavelets with different properties such that maximum compression is achieved.

- development of a switching scheme to switch between sinusoidal/optimum wavelet basis according to the time varying characteristics of the audio signal.
- the investigation and applicability of wavelet transforms and wavelet packets in the design and implementation of digital hearing aid algorithms to enhance and modify speech and audio, especially for sensorineural hearing impaired people.

The software implementation also became an important goal as advances in computer technology had made an entirely software based codec (coder + decoder) feasible. The experimental results obtained with various wavelet bases and sinusoidal basis revealed a novel theoretical result: *the performance of a transform (or basis) in compressing a signal not only depends on the properties of the basis function, but also depends on the statistical features of the source signal itself.*

1.4 Thesis Organisation

This thesis is divided into eight chapters. This first chapter gives introduction to perceptual audio coding and literature review on the same. Next, the aims and objectives of the present investigation is followed by this section which deals with the organisation of this research and thesis.

The second chapter **"An Overview of MPEG Audio"**, describes the basic concepts behind the MPEG Audio coding international standard. Emphasis is given to psychoacoustic principles of perceptual audio coding such as: critical bands, masking and hearing threshold in quiet.

The third chapter, **"Wavelet Theory"**, introduces fundamentals of wavelets, and implementation of discrete wavelet transforms / discrete wavelet packets. This chapter also provides details and properties of wavelets used in this research work.

The use of wavelet transforms for audio compression gives rise to many questions. What kind of decomposition has to be used? Has the transform to adapt to the signal? Which wavelet filters are well suited to the signal ? These points are addressed in Chapter4. Most of the authors use Daubechies filters, while many other filters are available and may be efficient. Thus the fourth chapter, "**Discrete Wavelet**

Transform based Perceptual Audio Coder" deals with comparative performance evaluation of various wavelet filter families of different properties. Three optimization methods for selecting the best wavelet basis from a predefined library of wavelets are also developed in this chapter. A novel switching algorithm to switch between sinusoidal/optimum wavelet basis according to the time varying characteristics of the audio signal is also proposed in this chapter. A new Vector Quantization (VQ) scheme, in which length of the code book can be adaptively changed according to the psychoacoustic model requirement is also proposed here.

In the fifth chapter, **"Wavelet Packet Based Audio Coding Schemes",** three variable bit rate schemes, which give higher compression ratios than the DWT based coder described in Chapter 4, are proposed. The performance of these coding schemes are compared with DWT based audio coder developed in Chapter 4, in terms of computational complexity, encoding delay, perceptual quality and compression ratio. In the first scheme, analysis filterbank (Discrete Wavelet Packet) splits the audio signal into 27 subbands, closely mimicking the human auditory system. Psychoacoustic analysis is done with a separate high resolution FFT stage. Higher compression ratio is achieved by this scheme at the expense of increase in computational complexity. **"A Low Complexity (Computationally More Efficient) DWP Based Audio Coder"** is proposed as the second scheme. In this scheme, computational load is reduced by integrating the psychoacoustic model design into the design of analysis filterbank itself. This is implemented at the expense of a slight reduction in the perceptual quality. A **"Scalable Perceptual Audio Coder Using Wavelet Packets",** which can support most of the industrial audio sampling frequencies (namely 11.025 kHz, 22.05 kHz and 44.1 kHz) and suitable for an advanced user is also proposed in this chapter.

The perceptual coders developed in Chapters 4 and 5 are of variable bit rate type. The sixth chapter, **"Constant Bit Rate Wavelet Packet Based Audio Coder",** describes a coding scheme suitable for constant bit rate transmission. A two stage SNR (Signal to Noise Ratio) scalable audio coding scheme is also proposed in this chapter.

The seventh chapter, **"Enhancement and Modification of Speech and Audio for the Hearing Impaired",** gives a brief introduction to different types of hearing

losses, and the principles of various hearing compensation methods/algorithms. A new digital hearing aid algorithm using DWT analysis, matching to the standard audiogram specifications (octave steps), is designed and implemented here. A WP based hearing aid algorithm according to the audiogram standard using half octave steps is also designed in this chapter.

The thesis is concluded by summarising the research performed and indicating avenues for future research and investigation, in Chapter 8. Line diagram of the research work is given in Figs. 1.2 and 1. 3.

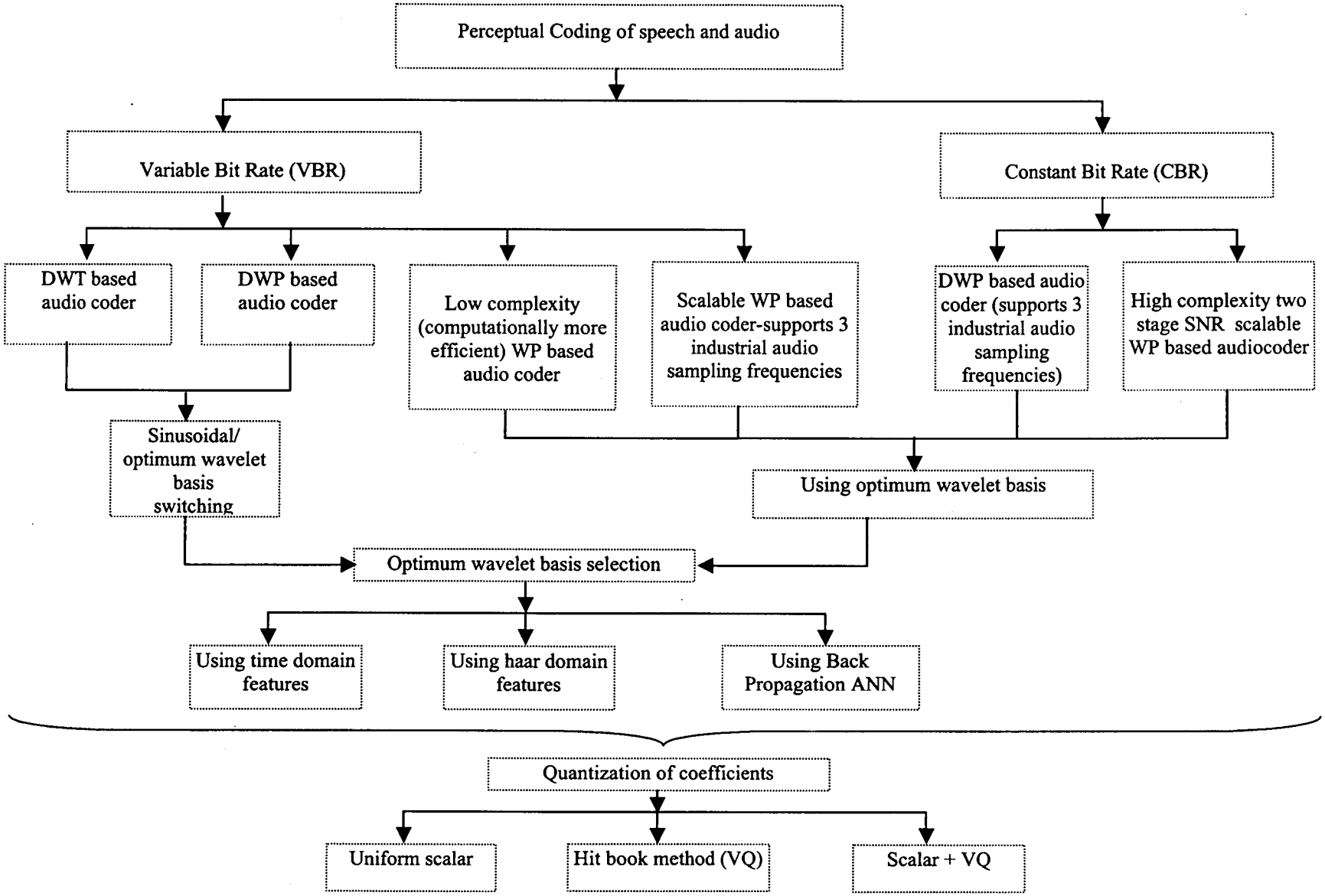


Fig.1.2: Schematic Diagram of the programme of the present work on perceptual coding of speech and audio

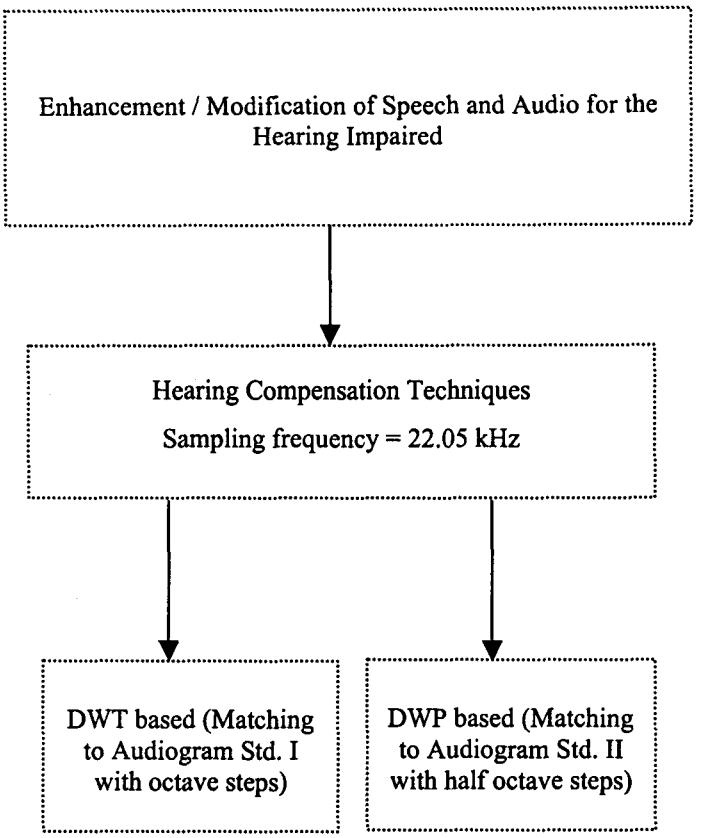


Fig.1.3: Schematic Diagram of the Programme of the Present Work on Enhancement/Modification of Speech/Audio for the Hearing Impaired

Chapter 2

AN OVERVIEW OF MPEG AUDIO

2.1 Introduction

The MPEG Audio compression algorithm is the first International Standard for the digital compression of high-fidelity audio [40,41]. This standard is the result of three years of collaborative work by an international committee of high-fidelity audio compression experts known as the Moving Picture Experts Group (MPEG). The International Organization for Standards and the International Electrotechnical Commission (ISO/IEC) adopted this standard [23] at the end of 1992.

2.2 Features and Applications of MPEG Audio

Sampling Rate : The audio sampling rate can be 48 kHz, used in professional sound equipment, 44.1 kHz used in consumer equipment like CD audio or 32 kHz used in some communications equipment.

Operating Mode: MPEG 1 audio works for both mono and stereo signals. A technique called joint stereo coding can be used to do more efficient combined editing of the left and right channels of a stereophonic audio signal. The operating modes are

- Single channel
- Dual Channel (two independent channels, for example, containing different language versions of the audio)
- Stereo (no joint stereo coding)
- Joint Stereo

Predefined Bit-Rates: The MPEG compressed bit stream can have one of several predefined fixed bit-rates ranging from 32 kb/s per channel to 448 kb/s. A raw PCM

audio stream is about 705 kb/s. Hence 32 kb/s corresponds to a compression ratio of about 22. Normal compression ratio is more like 4:1 (Layer I), 6:1 (Layer II) and 12:1 (Layer III). 96 kb/s is considered transparent for most practical purposes. This means that we will not notice any difference between the original and the compressed signal for rock'n roll or popular music. For more demanding material like piano concerts and such, we will need to go up to 128 kb/s.

Compression Layers: The MPEG committee chose to recommend three compression methods and named them Audio Layer I, II, and III. This provides increasing quality/ compression ratios with increasing complexity and demands on processing power.

Layer I is the simplest, a polyphase filter bank with a psychoacoustic model. It best suits bit-rates above 128 kb/s per channel. Philips' Digital Compact Cassette (DCC) uses Layer I at 192 kb/s per channel.

Layer II adds more advanced bit allocation techniques and greater accuracy. It has intermediate complexity and targets bit-rates around 128 kb/s per channel. Possible applications include Digital Audio Broadcasting (DAB).

Layer III adds a hybrid filter bank and non uniform quantization plus advanced features like Huffman coding, higher frequency resolution and bit reservoir technique. It is the most complex but offers the best audio quality, at bit rates around 64 kb/s per channel. This layer suits audio transmission over ISDN.

Thus a wide range of trade-offs between codec complexity and compressed audio quality is offered by the three layers. The reason for recommending three layers was partly that the testers felt that none of these coders was 100% transparent to all material and partly that the best coder (Layer III) was so computation intensive that it would seriously impact the acceptance of the standard.

The specifications say that a valid Layer III decoder shall be able to decode Layer I, II or III MPEG Audio stream. A Layer II decoder shall be able to decode Layer I and Layer II streams. This is the so called "Backward Compatibility"(BC).

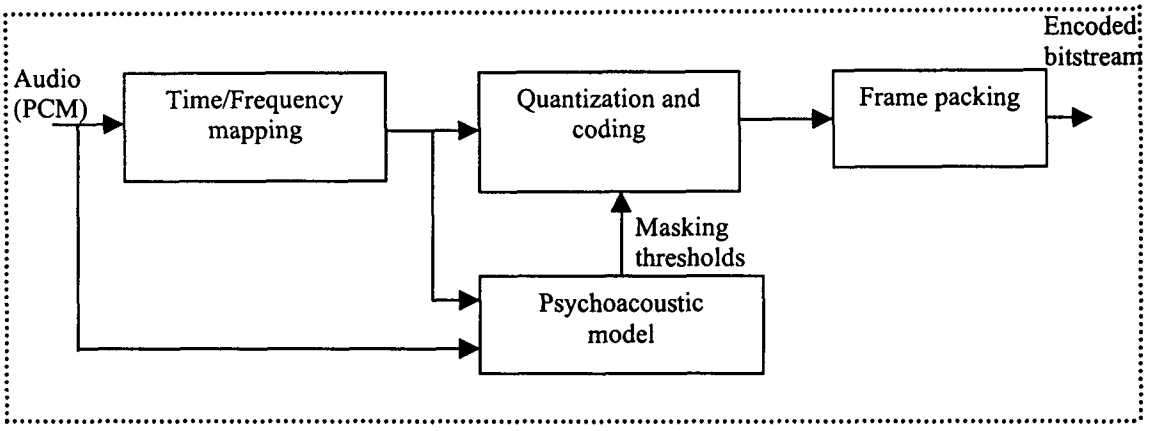


Fig. 2.1: Basic structure of the ISO/MPEG audio encoder

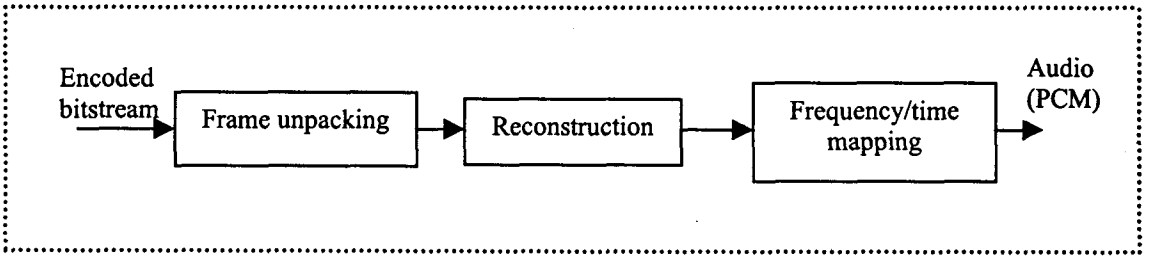


Fig. 2.2: ISO/MPEG Audio decoder

The specifications say that a valid Layer III decoder shall be able to decode Layer I, II or III MPEG Audio stream. A Layer II decoder shall be able to decode Layer I and Layer II streams. This is the so called "Backward Compatibility"(BC).

2.3 Overview

The basic structures of perceptual audio encoder and decoder are shown in Figs. 2.1 and 2.2. Encoder consists of the following four main parts:

- A time/frequency mapping (filter bank) is used to decompose the input signal into subsampled spectral components. Depending on the filter bank used, these are called subband values (low frequency resolution together with high time resolution) or transform coefficients.

- The output of this filter bank or separate calculation of frequency content, is used to calculate an estimate of the actual time dependent masking threshold using rules known from psychoacoustics.
- The subband samples or frequency lines are quantized and coded with the aim of keeping the noise, introduced by quantizing, below the masking threshold. Depending on the quantization and coding algorithm, this step is done in very different ways.
- In the last step, a frame packing is used to assemble the bit stream, which typically consists of the quantized and coded mapped samples and some side information. Entropy coding is done to remove statistical redundancies.

2.4 Filter Banks

The following list provides a short overview of the most common filter banks [58,59] used for coding of high quality audio signals:

- Discrete Fourier Transform (DFT) or Discrete Cosine Transform (DCT)

These were the first transforms used in transform coding of audio signals. They implement equally spaced filter banks with at least 128 to 512 bands at a low computational complexity. They do not provide critical sampling, i.e., the number of time/frequency components is greater than the number of time samples represented by one block length. Another disadvantage of these transforms are possible blocking artifacts.

- Polyphase Filter Banks

These are equally spaced filter banks which combine the filter design flexibility of generalized QMF banks with low computational complexity [58]. A polyphase filter bank using 32 bands is used for Layer I and Layer II of the MPEG coder. The main disadvantage of the polyphase filterbank is that, it is not of "perfect reconstruction" type.

- Modified Discrete Cosine Transform (MDCT)

MDCT using time domain aliasing cancellation is proposed in [56,60]. This transform cobines with a good frequency resolution provided by a sine window and the

computational efficiency of a fast FFT like algorithm [59]. Typically 128 to 512 equally spaced bands are used.

- Hybrid structures (eg. Polyphase +MDCT)

Using hybrid structures as proposed in [61], it is possible to combine different frequency resolution at different frequencies with moderate implementation complexity. A hybrid scheme consisting of a polyphase filter bank and a MDCT is used in Layer III. However it does not exploit the human auditory system's frequency dependent behaviour.

- Quadrature Mirror Filters, QMF tree filter banks.

Different frequency resolutions at different frequencies is possible. Typical QMF tree filter banks uses up to 32 bands. The computational complexity is also low. The advantage of QMF filterbanks is that near perfect stop band rejection is possible. Theoretically MDCT and polyphase filter banks belong to the same class of time to frequency domain mappings, called Lapped Orthogonal Transform [58].

2.4.1 Polyphase Filterbank

The polyphase filter bank is common to all three layers of MPEG Audio. This filter bank divides the audio signal into 32 equal width frequency subbands. It should be noted that the polyphase filterbank and its inverse are not lossless transformations. Even without quantization, the inverse transformation cannot perfectly recover the original signal.

The ISO / MPEG Audio standard [23] describes steps for computing the polyphase filterbank, analysis and synthesis algorithms.
The analysis algorithm is given by the following equation:

$$S_i[t] = \sum_{k=0}^{63} \sum_{j=0}^{7} M[i][k] \times (C[k + 64\ j] \times [k + 64\ j]) \qquad (2.1)$$

where i is the subband index and ranges from 0 to 31; $S_i[t]$ is the filter output sample for subband i at time t, where t is an integer multiple of 32 audio sample intervals;

C[n] is one of 512 coefficients of the analysis window defined in the ISO/MPEG audio standard: X [n] is an audio input sample read from a 512 sample buffer ; and

$$M[i][k] = \cos\left[\frac{(2i+1)\times(k-16)\times T}{64}\right] \tag{2.2}$$

are the analysis matrix coefficients.

The function within the paranthesis in Eq. (2.1) is independent of the value of i, and $M[i][k]$ is independent of j, so the 32 filter outputs need , 512+(32x64)=2,560 multiplications and (64x7)+(32x63)=2,464 additions or roughly 80 multiplications and additions per output.

However, the polyphase filterbank is one of the most computational intensive operations in MPEG coding. For example, MPEG audio decoding showed that the polyphase synthesis operation represented 40% of the overall decoding time. Hence, fast algorithms are of prime importance here, especially for applications such as real time audio encoding and decoding. Substantially further reductions in multiplications and additions are possible with a fast Discrete Cosine Transform or a Fast Fourier Transform implementation. For example, the original 2048 multiply - accumulate operations in the matrixing operation, can be reduced to 80 multiplications and 209 additions by using 32 point Lee's fast DCT algorithm [62]. Overall this reduces the original 512+(32x64)=2,560 multiplications down to 512+80=592 and the additions from (64x7)+(32x63)=2,464 down to (64x7)+209=657 or roughly 20 multiplications and additions per output. Note also that this polyphase filterbank is critically sampled. For every 32 input samples, the filterbank produces 32 output samples.

Equation 2.1 can be rewritten as

$$S_i[t] = \sum_{n=0}^{511} X[t-n]\times H_i[n] \tag{2.3}$$

where X [t] is an audio sample at time t, and

$$H_i[n] = h[n]\times\cos\left[\frac{(2i+1)\times(k-16)\times\pi}{64}\right] \tag{2.4}$$

with h [n] = - C [n] if the integer part of (n/64) is odd and h [n] = C [n] otherwise, for n = 0 to 511. In this notation, each subband of the filterbank has its own band pass

filter response, $H_i[n]$. The coeffiicients, h[n], correspond to the prototype low-pass filter response for the polyphase filterbank. Eq.(2.4) clearly shows that each is a modulation of the prototype response with a cosine term to shift the low pass response to the appropriate frequency band. Hence these are called polyphase filters. The polyphase analysis and synthesis algorithms in pseudo code are following:

ANALYSIS ALGORITHM	SYNTHESIS ALGORITHM

• Begin
• for i=511 down to 32 do
X [i]=X [i - 32]
• for i = 31 down to 0 do
X [i] = next_input_audio_sample
• Window by 512 coefficients, produce vector Z for i = 0 to 511 do $Z_i = C_i \times X_i$
• Partial calculation
for k = 0 to 63 do
$Y_k = \sum_{j=0}^{7} Z_{k+64\,j}$
• Calculate 32 samples by matrixing
for i = 0 to 31 do
$S_i = \sum_{k=0}^{63} M_{ik} * Y_k$ where
$M_{ik} = \cos\left[\frac{(2i+1)*(k-16)*\pi}{64}\right]$
• Output 32 subband samples
	• Begin
	• Input 32 new subband samples S_i
	i = 0 31
	• Shifting for i = 1023 down to 64 do
	V [i] = V [i - 64]
	• Matrixing for i = 0 to 63 do
	$V_i = \sum_{k=0}^{31} N_{ik} * S_k$ where
	$N_{ik} = \cos\left[\frac{(16+i)*(2k+1)*\pi}{64}\right]$
	• Build a 512 values vector U
	for i = 0 to 7 do; for j = 0 to 31 do
	U [i * 64 + j] = V [i * 128 + j]
	U [i *64 + 32 +j] = V [i * 128 + 96 +j]
	• Window by 512 coefficients, produce vector W.
	for i = 0 to 511 do
	$W_i = U_i * D_i$
	• Calculate 32 samples for j = 0 to 31 do
	$S_j = \sum_{i=0}^{15} W_{j+32\,i}$
	• Output 32 reconstructed PCM samples

2.5 Psychoacoustic Principles

The number of bits needed to represent an audio signal can be reduced without affecting the perceptual quality by examining the perception of sound by a human listener, identifying the components that will not be audible and throwing these components.

2.5.1 A walk through the Human auditory system

The main components of the human auditory system are shown in Fig.2.3 [19].

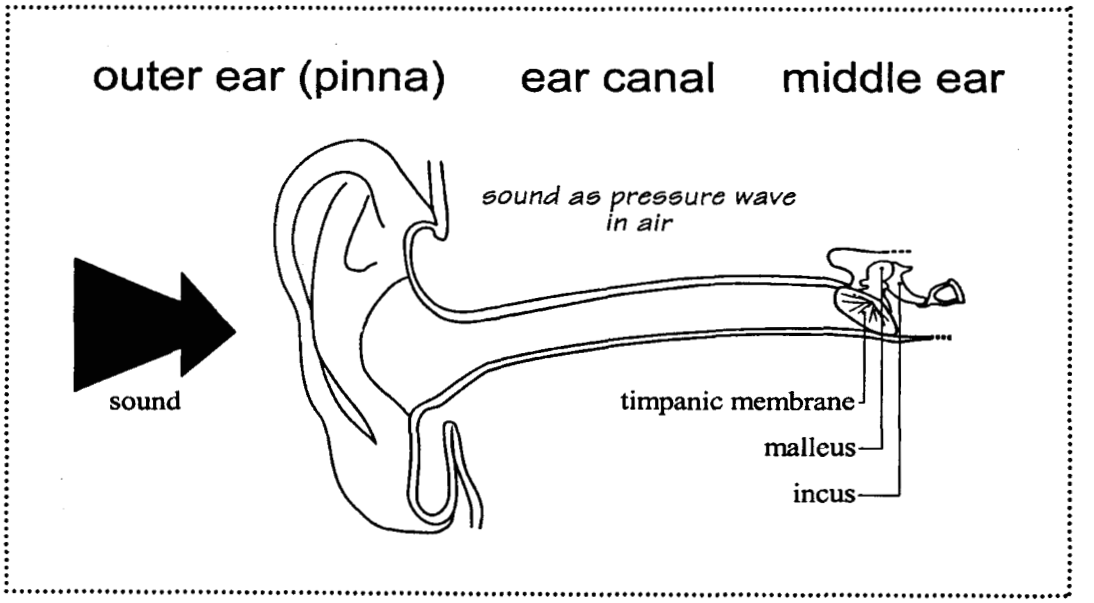


Fig.2.3: Human auditory system

Sound waves incident from different angular positions are spectrally shaped by the pinna in a direction dependent manner. The ear canal further filters the waveform, before it passes through two small bones, and on to the cochlea. The ear canal is the resonant cavity between the outer and middle ear, which has a resonance at around 3-5 kHz. Hence it attenuates higher and lower frequencies. Cochlea is a fluid filled coil within the ear, and is partially partitioned by the Basilar Membrane (BM) (see Fig.2.4). Sensory cells (outer hair cells and inner hair cells) are distributed along the basilar membrane (see Fig. 2.5).

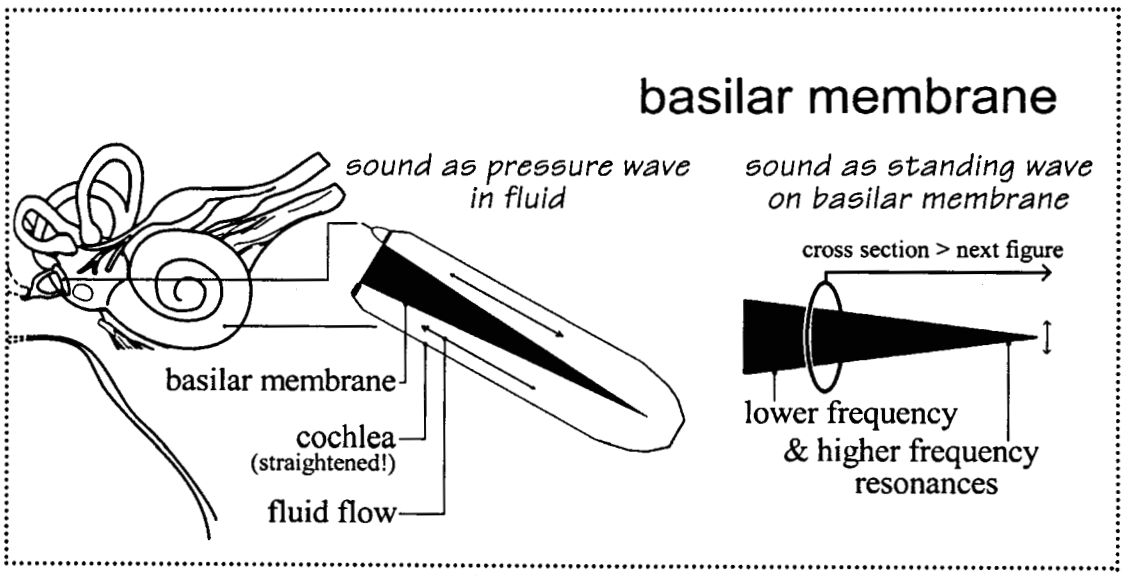


Fig.2.4: Cochlea

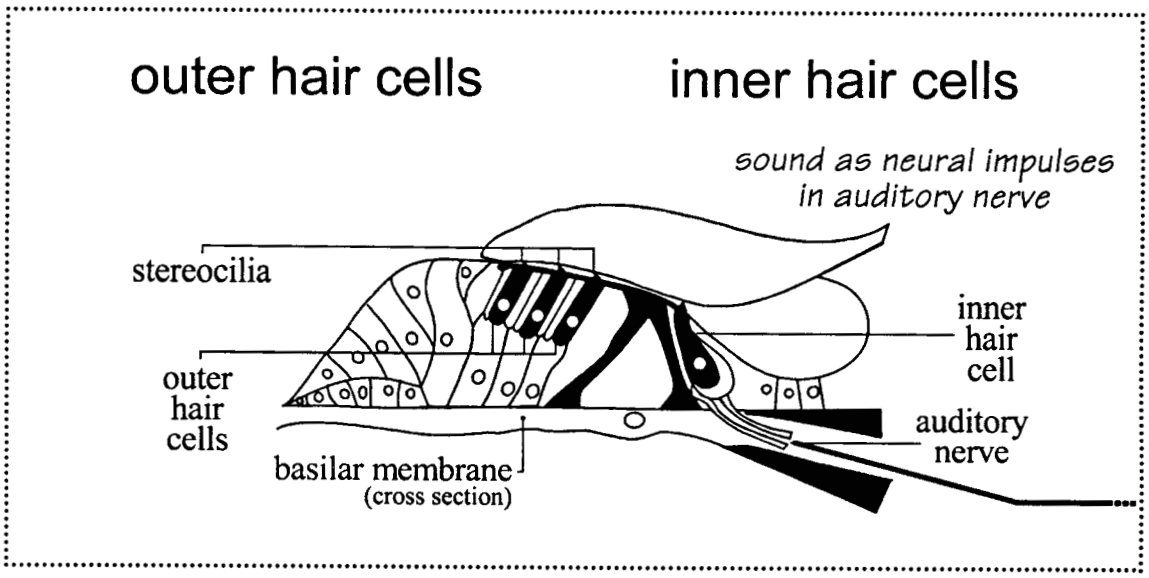


Fig.2.5: Cross section of basilar membrane

The different points of the basilar membrane resonate at different frequencies. Thus the BM acts as a spectrum analyser. The spacing of frequency resonances along the BM is not linear with frequency. The resonant frequencies of various points along the BM are shown in Fig 2.6. The scale that relates the resonant frequency to position on BM is called the Bark scale or Critical Band sale [20,63,64]. It approximates to a log scale.

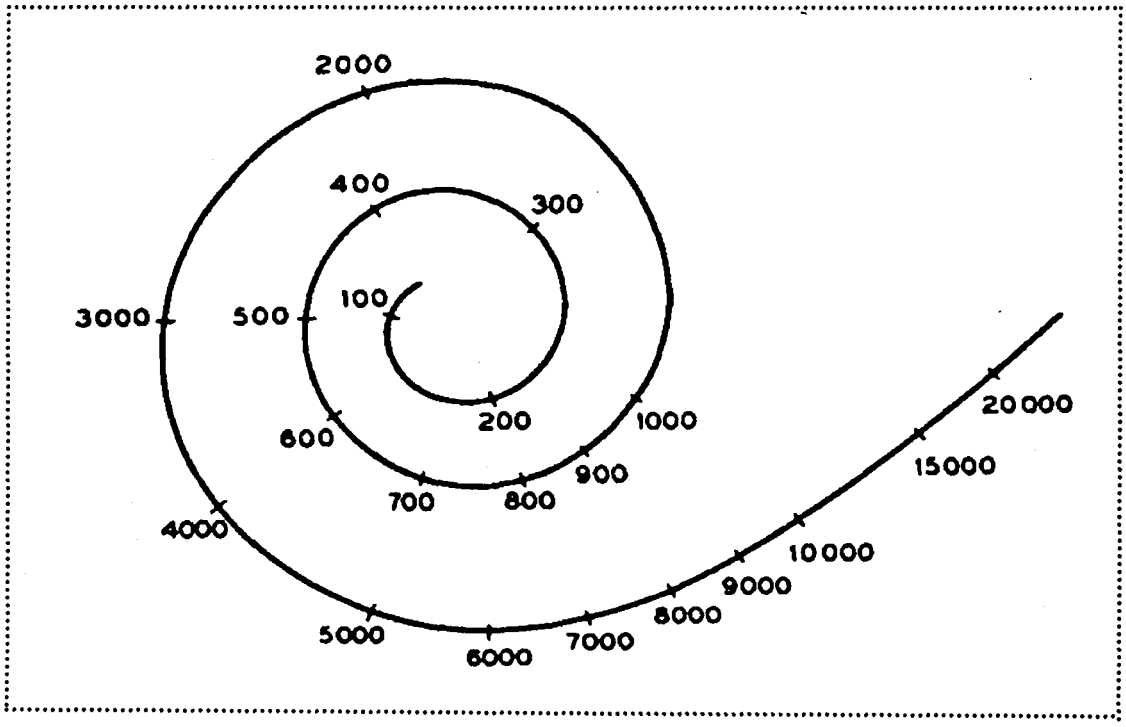


Fig.2.6: Resonant frequencies of various points along the basilar membrane

. Sound waves enter the cochlea and set the fluid within it in motion. The movement of the fluid stimulates the hair cells of BM. Auditory nerve endings carry these stimuli to the auditory centre of the brain. Interpretations of these impulses by the brain results in hearing.

2.5.2 Absolute Threshold of Hearing

The absolute threshold of hearing [20,63,64] characterizes the amount of energy needed in a pure tone such that a listener in a noiseless environment can detect it. The absolute threshold is typically expressed in terms of dB SPL (Sound Pressure Level). The SPL gives the level (intensity) of sound pressure in decibels (dB) relative

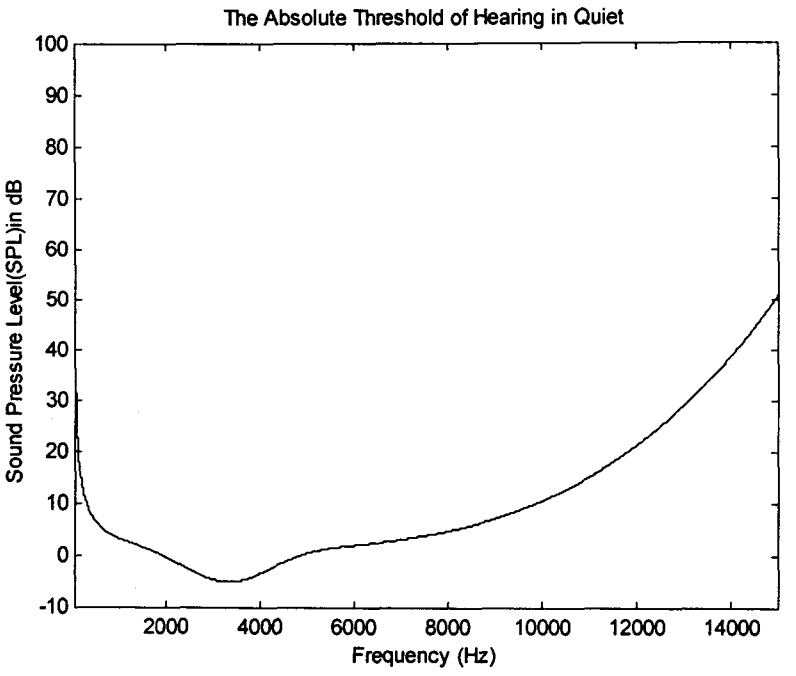


Fig 2.7: Hearing threshold in quiet

to an internationally defined reference level, i.e. L_{SPL} = 20 log 10 (P/Po) dB, where L_{SPL} is the SPL of stimulus, 'P' is the sound pressure of stimulus in Pascal's, and 'Po" is the standard reference level of 2 x 10^{-5} N/m^2. The quiet threshold is well approximated by the non-linear function

$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5\exp(-0.6(f/1000-3.3)^2) + 10^{-3}(f/1000)^4 \text{ (dB SPL)} \qquad (2.5)$$

which is representative of a young listener with acute hearing. When applied to signal compression, Tq (f) could be interpreted as a maximum allowable energy level for coding distortions introduced in the frequency domain. Variation of threshold in quiet with frequency is given in Fig.2.7. It is the outer ear canal that is responsible for the high sensitivity of hearing at frequencies near 4 kHz, indicated by the dip of threshold in quiet around 4 kHz.

2.5.3 Critical Bands

Using the absolute threshold to shape the coding distortion spectrum represents the first step toward perceptual coding. The detection threshold for spectrally complex quantization noise is a modified version of the absolute threshold, with its shape determined by the stimuli present at any given time. Since stimuli are in general time

varying, the detection threshold is also a time varying function of the input signal. Ear performs spectral analysis as follows. A frequency-to-place transformation takes place in the cochlea (inner ear), along the basilar membrane. A sound wave generated by an acoustic stimulus moves the eardrum and the attached ossicular bones, which in turn transfer the mechanical vibrations to the cochlea. Once excited by mechanical vibrations at its input, the cochlear structure induces travelling waves along the length of the basilar membrane. Neural receptors are connected along the length of the basilar membrane. The travelling wave generate peak responses at frequency- specific membrane positions, and therefore different neural receptors are effectively tuned to different frequency bands according to their locations. For sinusoidal stimuli, the travelling wave on the basilar membrane propagates from the oval window, until it nears the region with a resonant frequency. The wave then slows, and the magnitude increases to a peak. The location of the peak is referred to as the best place or characteristic place for the stimulus frequency, and the frequency that best excites a particular place is called the 'best frequency' or 'characteristic frequency'. Thus a frequency-to-place transformation occurs. An example is given in Fig.2.8.

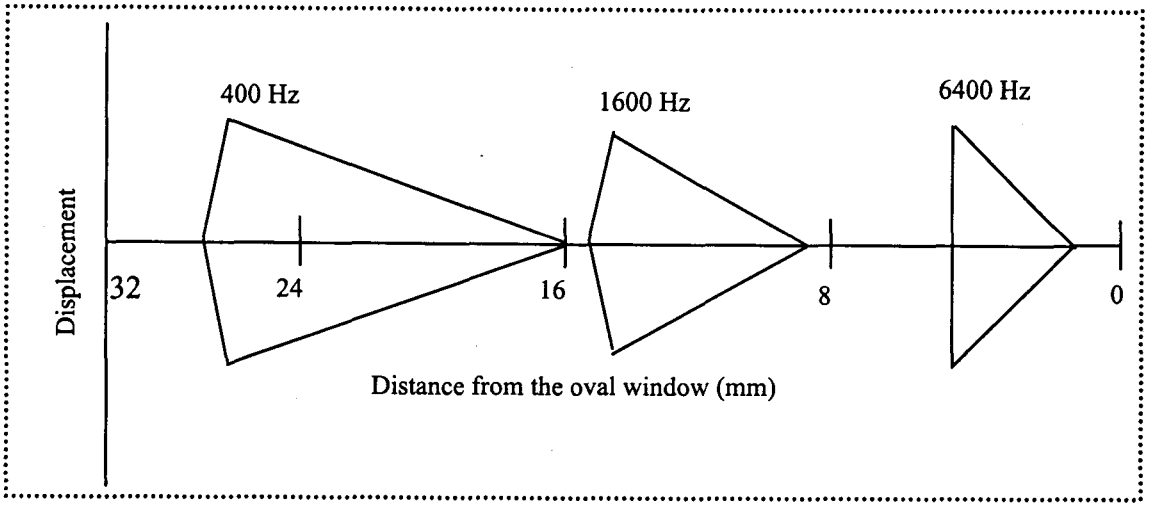


Fig.2.8: The frequency-to-place transformation

The above figure gives a schematic representation of the travelling wave envelopes that occur in response to an acoustic tone complex containing sinusoids of 400,1600, and 6400 Hz. Peak responses for each sinusoid are localized along the

membrane surface, with each peak occurring at a particular distance from the oval window (cochlear window). As a result of the frequency-to-place transformation, the cochlea can be viewed from a signal-processing perspective as a bank of highly overlapping band pass filters. The cochlear filter pass bands are of non-uniform bandwidth, and the bandwidth increases with increasing frequency.

Table 2.1. Idealized critical band filter bank

Band No.	Lower Edge (Hz)	Upper Edge (Hz)	BW (Hz)
1	0	100	100
2	100	200	100
3	200	300	100
4	300	400	100
5	400	510	110
6	510	630	120
7	630	770	140
8	770	920	150
9	920	1080	160
10	1080	1270	190
11	1270	1480	210
12	1480	1720	240
13	1720	2000	280
14	2000	2320	320
15	2320	2700	380
16	2700	3150	450
17	3150	3700	550
18	3700	4400	700
19	4400	5300	900
20	5300	6400	1100
21	6400	7700	1300
22	7700	9500	1800
23	9500	12000	2500
24	12000	15500	3500
25	15500	22000	6500

The critical bandwidth is a function of frequency that quantifies the cochlear filter pass bands. Approximate critical bands of auditory system are shown in Table 2.1. The critical band can be loosely defined as the bandwidth at which subjective responses change abruptly. For example, the perceived loudness of a narrowband noise source at constant sound pressure level remains constant even as the bandwidth is increased up to the critical bandwidth. The loudness then begins to increase. For an average listener, the critical bandwidth is approximated by

$$BW_c(f) = 25 + 75\,[1+1.4(f/1000)^2]^{0.69}\ \text{Hz} \tag{2.6}$$

A distance of one critical band is referred as one bark in literature. The function

$$z(f) = 13\arctan(0.00076f) + 35\arctan[(f/7500)^2] \tag{2.7}$$

is used to convert from frequency scale to bark scale.

2.5.4 Masking

Masking [1,20,21,63,64] refers to a process where one weak sound is rendered inaudible because of the presence of another strong sound. Simultaneous masking is a frequency domain phenomenon within critical bands when two or more stimuli are simultaneously present.

The mechanism underlying simultaneous masking phenomena is that the presence of a masker creates an excitation of sufficient strength on the basilar membrane at the critical band location to effectively block the transmission of a weaker signal. There are two types of masking, namely Noise Masking Tone (NMT) and Tone Masking Noise (TMN).

a. Noise - Masking - Tone (NMT)

In the NMT scenario, a narrow band noise masks a tone within the same critical band, provided that the intensity of the masked tone is below a predictable threshold directly related to the intensity of the masking noise. At the threshold of detection for the masked tone, the minimum signal-to-mask ratio (SMR), i.e. the smallest difference between the intensity (SPL) of the masking noise and the intensity of the masked tone occurs when the frequency of the masked tone is close to the masker's centre

frequency. In most studies, the minimum SMR tends to lie between -5 and +5 dB. Fig. 2.9 (a) shows the NMT scenario. In this figure, a critical band noise masker centered at 410 Hz with an intensity of 80-dBSPL masks a 410 Hz tone, and the resulting SMR at the threshold of detection is 4 dB. Masking power decreases for probe tones above and below the frequency of the SMR tone, in accordance with a level-and frequency-dependent spreading function (discussed in Section 2.5.6).

b. Tone- Masking- Noise (TMN)

In the case of TMN, a pure tone occurring at the center of a critical band masks noise of any sub critical bandwidth or shape, provided the noise spectrum is below a predictable threshold related to the strength of the masking tone. At the threshold of detection for a noise band masked by a pure tone, it was found that the minimum SMR, i.e. the smallest difference between the intensity of the masking tone and the intensity of the masked noise, occurs when the masker frequency is close to the center frequency of the probe noise. Minimum SMR for TMN tends to lie between 21-28 dB. This is shown in Fig.2.9 (b). In the figure, a narrow band noise of one Bark band width centered at 1 kHz is masked by a 1 kHz tone of intensity 80 dB SPL.

2.5.5 Asymmetry of Masking

The NMT and TMN examples in Fig. 2.9 clearly show an asymmetry in masking power between the noise masker and the tone masker. In spite of the fact that both maskers are presented at a level of 80 dB SPL, the associated threshold SMR's differ by about 20 dB. For each temporal analysis interval, a codec's perceptual model should identify across the frequency spectrum noise- like and tone-like components. The model should apply the appropriate masking relationships in a frequency specific manner. In conjunction with the spread of masking, NMT and TMN properties can be used to construct a global masking threshold.

2.5.6 The Spread of Masking

Simultaneous masking effects are not band limited to within the boundaries of a single critical band. Interband masking also occurs, i.e., a masker centred within one

critical band has some predictable effect on detection thresholds in other critical bands. This effect, also known as spread of masking, is often modeled in coding applications by an approximately triangular spreading function.

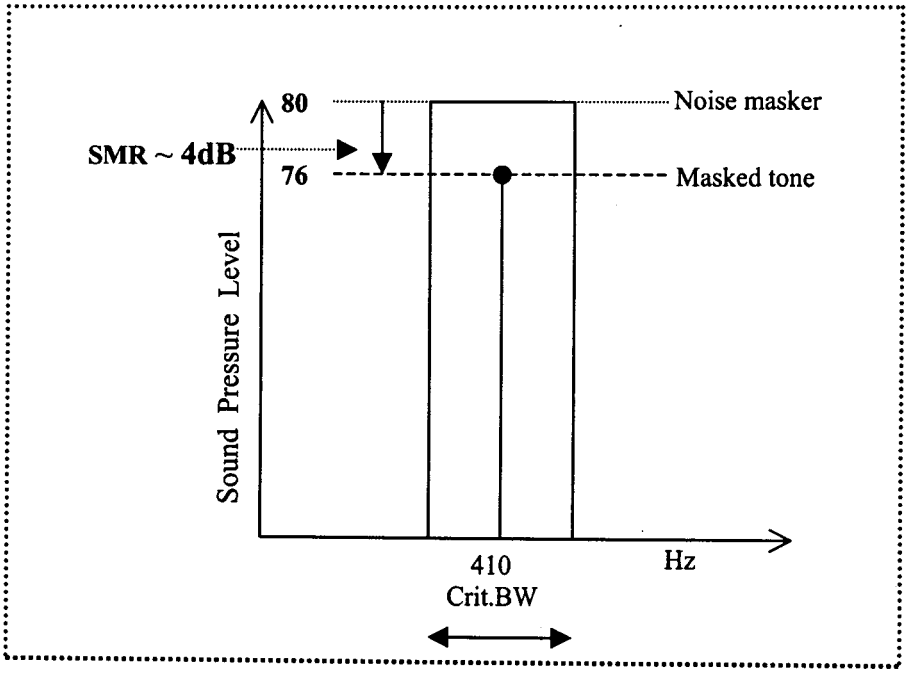


Fig. 2.9 (a): Noise masking tone.

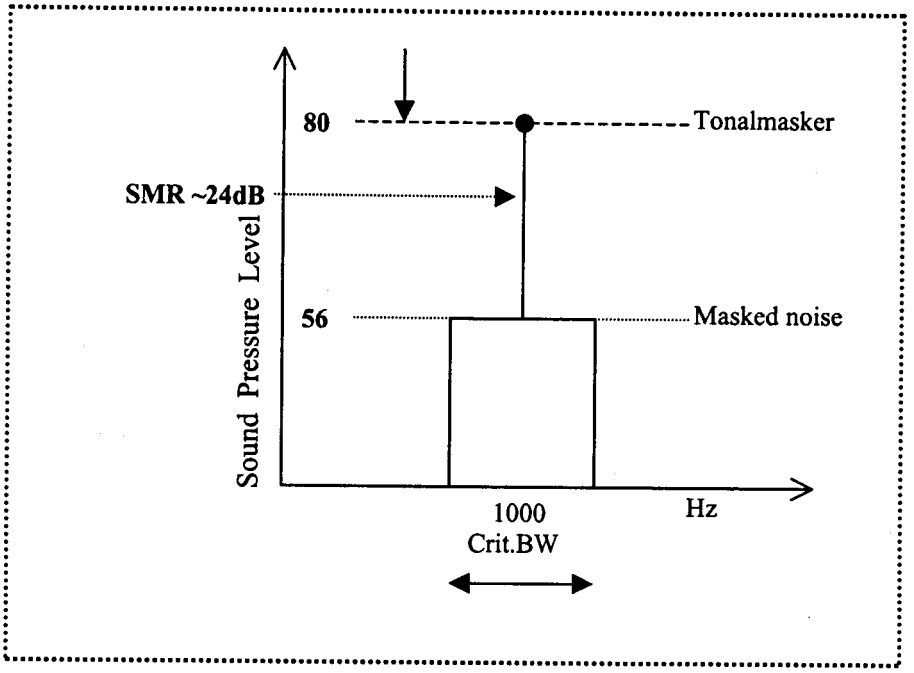


Fig.2.9 (b): Tone masking noise

An analytical expression for the spreading function can be given as:

$$SF\ dB\ (x) = 15.81 + 7.5\ (x + 0.474) - 17.5(1+(x+0.471)^2)^{1/2}\ dB. \quad (2.8)$$

where 'x' has unit of Barks.

After critical band analysis is done and spread of masking has been accounted for, masking thresholds in perceptual coders are established by the relations.

$$T_{HN} = E_T - 14.5 - B \quad (2.9)$$

$$T_{HT} = E_N - K \quad (2.10)$$

T_{HN} and T_{HT} are noise and tone masking thresholds, respectively, due to TMN and NMT. E_N and E_T are critical band noise and tone masker energy levels respectively. B- critical band number; K is typically set to 5 dB.

The above equations capture only the contributions of individual tone like or noise-like maskers. In the actual coding scenario each frame typically contains a collection of both masker types. After they have been identified, these individual masking thresholds are combined to form a global masking threshold. The global masking threshold comprises an estimate of the level at which quantization noise becomes just noticeable.

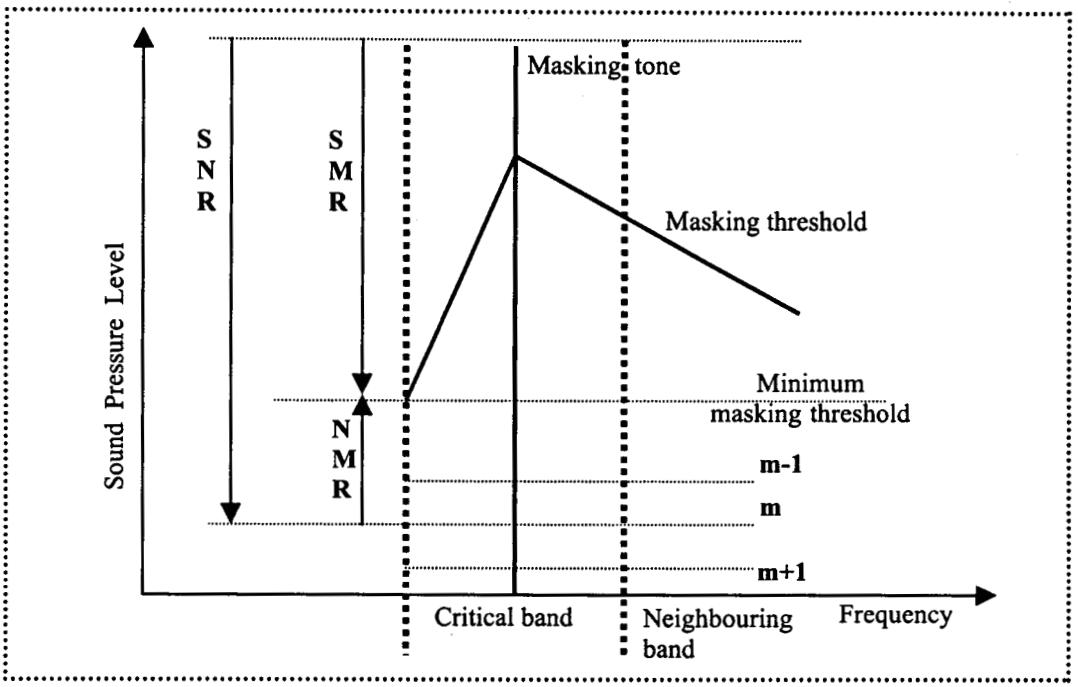


Fig.2.10: Schematic representation of simultaneous masking

Notions of critical bandwidth and simultaneous masking in the audio coding context give rise to some convenient terminology illustrated in Fig.2.10. Consider the case of a single masking tone occurring at the center of a critical band. This generates an excitation along the basilar membrane that is modeled by a spreading function and a corresponding masking threshold. For the band under consideration, the minimum masking threshold denotes the spreading function in-band minimum. Assuming the masker is quantized using an m-bit uniform scalar quantizer, noise might be introduced at level m. SMR and noise-to-mask ratio (NMR) denote the log distances from the minimum masking threshold to the masker and noise levels respectively.

2.6 Psychoacoustic Model Implementation (Layer I)

In this section emphasis is given on the implementation details of the Psychoacoustic model I, as has been used in common by both MPEG and the wavelet based codecs proposed in this thesis.

Step 1 Spectral analysis and SPL Normalisation

The goal of this step is to obtain a high resolution spectral estimate of the input signal. A 512 point FFT (Fast Fourier Transform) is used for the purpose.

First the input samples, s (n), are normalised according to the FFT length N and the number of bits per sample b.

$$X(n) = \frac{s(n)}{N(2^{b-1})} \tag{2.11}$$

Next, a *power spectral density* (PSD) estimate is obtained using 512 point FFT.

$$P(k) = P_N + 10\log_{10}\left|\sum_{n=0}^{N-1} w(n)x(n)e^{-j\frac{2\pi kn}{N}}\right|^2 \qquad 0 < k < \frac{N}{2} \tag{2.12}$$

where the power normalization term P_N is fixed at 90 dB and the Hann window, w(n) is defined as

$$w(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N}\right)\right] \tag{2.13}$$

Since play back levels are unknown during psychoacoustic analysis, the normalisation term P_N is used to estimate SPL (Sound Pressure Level) conservatively

from input power. For example, a full scale sinusoid which is precisely resolved by the 512 -point FFT in bin k_0 will yield a spectral line, P (k_0), having 84 dB SPL. With 16 -bit sample resolution, SPL estimates for very low amplitude input signals will be at or below the absolute threshold.

Step 2 : Tonal and Noise Masker Identification

Local Maxima in the PSD which exceed neighbouring components within a certain bark distance by at least 7 dB are taken as tonal components. The tonal set S_T is defined as

$$S_T = \left\{ P(k) \middle| \begin{matrix} P(k) > P(k \pm 1) \\ P(k) > P(k \pm \Delta_k) + 7\,\text{dB} \end{matrix} \right\} \tag{2.14}$$

where

$$\Delta_k \in \begin{cases} 2 & 2 < k < 63 & 0.17 - 5.5\,\text{kHz} \\ [2,3] & 63 \le k \le 127 & 5.5 - 11\ \text{kHz} \\ [2,6] & 127 \le k \le 256 & 11 - 20\ \text{kHz} \end{cases}$$

The tonal maskers P_{TM} (k) are computed for the peaks obtained from the above step and listed in S_T as

$$P_{TM}(k) = 10 \log_{10} \sum_{j=-1}^{1} 10^{0.1P(k+j)} \tag{2.15}$$

The remaining spectral components in each critical band not within a certain bark distance (as explained earlier) of tonal components are added up into a single noise masker,

$$P_{NM}(\bar{k}) = 10 \log_{10} \sum_{j} 10^{0.1P(j)} \quad \text{dB} \tag{2.16}$$

$$\forall P(j) \notin \{P_{TM}(k, k \pm 1, k \pm \Delta_k)\}$$

where $\bar{k}$ is defined as the geometric mean spectral line of the critical band, *i.e.*,

$$\bar{k} = \left(\prod_{j=l}^{u} j \right)^{1/(l-u+1)}$$ where l and u are the lower and upper specral line boundaries of the critical band, respectively.

Step 3 Decimation of Maskers

In this step the number of maskers is reduced using two criteria. First, any tonal or noise masker below absolute threshold are discarded. That is, only maskers that satisfy the inequality given in Eq. 2.17 are retained.

$$P_{TM,NM}(k) \geq T_q(k) \tag{2.17}$$

Next, a sliding 0.5- Bark-wide window is used to replace any pair of maskers occurring within a distance of 0.5 Bark by the stronger of the two.

Step 4 Calculation of Individual Masking Threshold

Each individual masking threshold represents the masking contribution at a particular frequency bin, say i,(due to a tone or noise masker located at frequency bin, say j). Total masking thresholds are given by,

$$T_{TM}(i,j) = P_{TM}(j) - 0.275z(j) + SF(i,j) - 6.025 \quad \text{dB SPL} \tag{2.18}$$

where $P_{TM}(j)$ is the SPL of the tonal masker in frequency bin j, z(j) is the bark frequency of bin j, and SF(i, j) is the spread of masking from masker bin j to maskee bin i, and is given by the expression,

$$SF(i,j) = \begin{cases} 17\Delta_z - 0.4P_{TM}(j) + 11 & -3 \leq \Delta_z < -1 \\ (0.4P_{TM}(j) + 6)\Delta_z & -1 \leq \Delta_z < 0 \\ -17\Delta_z & 0 \leq \Delta_z < 1 \\ (0.15P_{TM}(j) - 17)\Delta_z - 0.15P_{TM}(j) & 1 \leq \Delta_z < 8 \end{cases}$$

Individual noise masker thresholds are given by,

$$T_{NM}(i,j) = P_{NM}(j) - 0.175z(j) + SF(i,j) - 2.025 \quad \text{dB SPL} \tag{2.19}$$

Step 5 Calculation of Global and Minimum Masking Thresholds

The individual masking thresholds are combined to estimate a *global masking threshold* for each frequency. Global masking threshold is given by the sum,

$$T_g(i) = 10\log_{10}\left(10^{0.1T_q(i)} + \sum_{l=1}^{L} 10^{0.1T_{TM}(i,l)} + \sum_{m=1}^{M} 10^{0.1T_{NM}(i,m)}\right) \quad \text{dB SPL} \tag{2.20}$$

where,

$T_q(i)$	:absolute threshold for frequency bin i;
$T_{TM}(i,l)$ and $T_{NM}(i,m)$	:individual masking thresholds from step 4;
L and M	:numbers of tonal and noise maskers, respectively, identified during step 3

The minimum value of global masking threshold in each critical band is taken as the minimum masking threshold of that particular critical band. From this the SMR (Signal-to-Mask Ratio) in each critical band is calculated. The bit allocation is then done on the basis of SMR s calculated in various subbands.

2.7 Summary

Basic concepts of ISO/MPEG audio coding standard are presented in this chapter. Human auditory system and psychoacoustic properties like absolute threshold of hearing, critical bands, masking *etc.* are discussed briefly. Implementation details of MPEG Layer I psychoacoustic model is also described here. Draw backs of MPEG standard by using uniform filterbank for time/frequency analysis are:

- Analysis does not match with the properties of speech and audio signals.

  Speech and audio signals are non-stationary and hence fixed time-frequency resolution windows are not suitable for their analysis. These signals call for narrow windows in the analysis of high frequency components and wide windows in the analysis of low frequency components.

- Analysis does not match with the properties of human auditory system.

  Human ear analyses various frequency components with different resolutions. Critical bands are non-uniform and the bandwidth of the critical bands increases as the frequency increases. That is, our ear analyses high frequency components with good time resolution and low frequency components with good frequency resolution.

- Filterbank and its inverse do not yield perfect reconstruction. This introduces errors even in the absence of quantization error.

Hence, for the efficient exploitation of perceptual irrelevancies in audio coding, analysis filterbank should match to the properties of human auditory system. MPEG standard faces with a serious artifact known as **pre-echo distortion,** because of employing uniform filterbank for analysis purpose. Pre-echo is noise, spread out over some time, even before the music event, causing the noise. To avoid pre-echo distortion, high frequency components should be analysed with narrow windows (good time resolution) and low frequency components should be analysed with wider windows (good frequency resolution).

Since, critical bands are of almost constant Q type, in order to fully exploit the masking thresholds in various frequency bands and to place the quantization noise in the least sensitive regions of the spectrum, the analysis filterbank should be of either constant Q type or whose subbands mimic the various critical bands of the human auditory system.

Major attraction of *wavelet analysis* is that it uses basis functions which are well localized in time and frequency. Hence, wavelet transform will concentrate the energy of a signal in very few transform coefficients. Non-uniform filterbank is used for the implementation of wavelet transform. A filterbank emulating the human hearing process can be constructed. Unlike Fourier analysis in which basis functions are only sines and cosines, a number of wavelet basis functions are available in the literature. Hence, wavelet analysis is more flexible in the sense that each audio frame can be represented with the most matching wavelet basis. DWT provides a good approximation to the Karhunen- Loeve transformation (KLT) of a wide class of stationary and non-stationary process. In this transform, high frequency components of the signal are analysed using narrow windows and low frequency components are analysed using wide windows. Hence, wavelet analysis is readily applicable to the task of perceptual audio coding. Brief theory of wavelets and its implementation details are discussed in the next chapter. Wavelet based perceptual audio coding schemes with various features are proposed in Chapters 4-6.

Chapter 3

WAVELET THEORY

3.1 Introduction

A wave is usually defined as an oscillating function of time or space, such as a sinusoid. Fourier analysis is wave analysis. It uses global (non-local) sine and cosine functions (extending from - ∞ to + ∞ and hence having infinite energy) as bases and hence is suitable for periodic, time-invariant or stationary phenomena. A *wavelet* is a small 'wave' which has its energy concentrated in time to give a tool for the analysis of transient, non-stationary or time varying phenomena. Hence, wavelet analysis [66-84] use bases that are localized in time and frequency to represent nonstationary signals (most of natural and human made signals are transient in nature) more effectively. As a result, a wavelet representation is much more compact and easier to implement. Using the powerful *multiresolution analysis*, we can represent a signal by a finite sum of components at different resolutions, so that each component can be processed adaptively based on the objectives of the application. This capability to represent signals compactly and in several levels of resolution is the major strength of *wavelet analysis*.

3.2 Continuous Wavelet Transform (CWT)

Consider a real or complex value continuous time function $\Psi(t)$ with the following properties.

1. The function integrates to zero.

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{3.1}$$

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

$$\int_{-\infty}^{\infty} |\psi(t)|^2 \, dt < \infty \tag{3.2}$$

3. Admissibility condition

$$C \equiv \int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{|\omega|} d\omega \qquad 0 < C < \infty \qquad (3.3)$$

where $\psi(\omega)$ is the Fourier Transform of $\psi(t)$.

The function $\psi(t)$ is mother wavelet or wavelet, if it satisfies the above three conditions. While the admissibility condition is useful in formulating a simple inverse wavelet transform, properties 1 and 2 suffice to define the CWT. Property 2 implies that most of the energy in $\psi(t)$ is confined to a finite duration. Property 1 is suggestive of a function that is oscillatory. Thus, in contrast to a sinusoidal function, it is a "small wave" or a wavelet.

Let f (t) be any square integerable function. The CWT of f (t) with respect to a wavelet $\psi(t)$ is defined as

$$W_{(a,b)} \equiv \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{|a|}} \psi^* \left(\frac{t-b}{a} \right) dt \qquad (3.4)$$

where 'a' and 'b' are real and * denotes complex conjugation. Thus, the wavelet transform is a function of two variables. f (t) and ψ (t) belong to $L^2_{(R)}$, the set of square integerable functions, also called the set of energy signals. The wavelet basis function and wavelet transform can be written as

$$\psi_{(a,b)}(t) = \frac{1}{\sqrt{|a|}} \psi \left(\frac{t-b}{a} \right) \qquad (3.5)$$

$$W_{(a,b)} = \int_{-\infty}^{\infty} f(t) \psi^*_{(a,b)}(t) dt \qquad (3.6)$$

The normalizing factor of $1/\sqrt{(|a|)}$ in Eq.3.5 ensures that the energy stays the same for all values of 'a' and 'b' i.e.,

$$\int_{-\infty}^{\infty} \left|\psi_{(a,b)}(t)\right|^2 dt = \int_{-\infty}^{\infty} \left|\psi(t)\right|^2 dt \qquad (3.7)$$

For any given value of a, the function $\psi_{a,b}(t)$ is a shift of $\psi_{a,0}(t)$ by an amount b along the time axis. Thus, the variable b represents time shifts or translation.

Time scaled and amplitude scaled version of $\psi(t)$ can be written as

$$\psi_{(a,0)}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t}{a}\right) \qquad (3.8)$$

Since 'a' determines the amount of time scaling or dilation, it is referred to as the scale or dilation variable. If $a>1$, there is a stretching of $\psi(t)$ along the time axis, and if $0 < a < 1$ there is a contraction of $\psi(t)$.

Time and frequency resolutions of wavelet transform is illustrated in Fig.3.1. Every box in the figure corresponds to a value of wavelet transform in the time frequency plane. Note that boxes have a certain non zero area, which implies that the value of a particular point in the time- frequency plane, cannot be known. All the points in the time-frequency plane that falls into a box are represented by one value of WT. Although the widths and heights of the boxes change, the area is constant. That is each box represents an equal portion of the time-frequency plane, but giving different proportions to time and frequency. Note that at low frequencies, the heights of the boxes are shorter, which corresponds to better frequency resolutions, but widths are longer, which corresponds to poorer time resolution. At higher frequencies the width of the boxes decreases, and height of boxes increases. The areas of boxes are same and determined by Heisenberg's inequality. The area of each mother wavelet is fixed, where as different mother wavelets can result in different areas. However all areas are lower bounded by $1/4\pi$. That is, we cannot reduce the areas of the boxes as much as

we want due to uncertainty principle. On the other hand, for a given mother wavelet the dimensions of the boxes can be changed, while keeping the 'area same'.

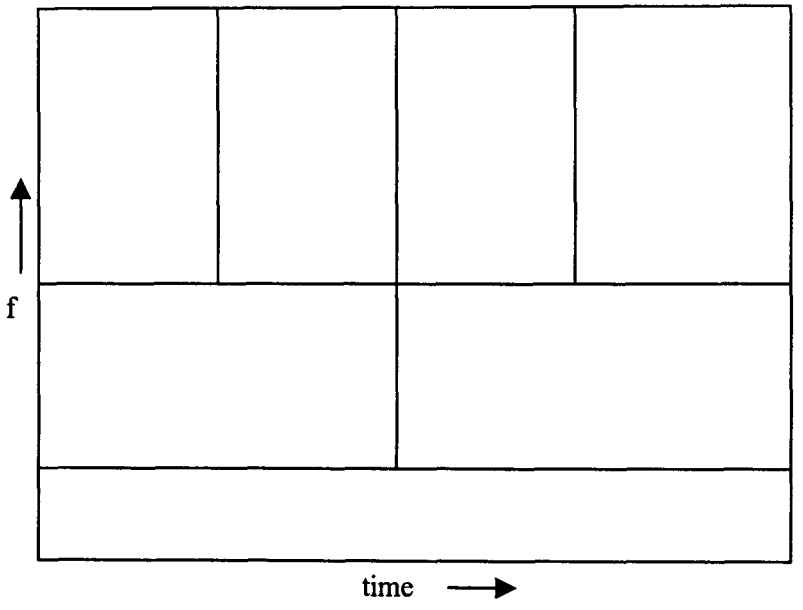


Fig. 3.1: Time and frequency resolution

The wavelet expansion set is not unique. There are many different wavelet systems that can be used very effectively, but all seem to have the following general characteristics [67].

1. A wavelet system is a set of basic building blocks to construct or represent a signal or function.

2. The wavelet expansion gives a time frequency localization of the signal. This means that most of the energy of the signal is well represented by a few expansion coefficients.

3. The calculation of coefficients from the signal can be done efficiently. It turns out that many wavelet transforms can be calculated with O(N) operation i.e. the number of floating point multiplications and additions increase linearly with the length of the signal.

For the Fourier series and transform and for most signal expansion systems, the expansion functions (bases) [85,86] are chosen, then the properties of the resulting transform are derived and analyzed. For the wavelet system these properties are mathematically required then the resulting basis functions are derived. Once we decide on Fourier series, the sinusoidal basis functions are the complete set. That is not true for the wavelet. There are an infinite number of wavelets that all satisfy the above results [66-70].

The wavelet expansions and wavelet transforms have proven to be very effective and efficient in analyzing a very wide class of signals and phenomena. The properties that give this effectiveness are

1. The size of the wavelet expansion coefficients drops off rapidly with a and b for a large class of signals. This property is called being an *unconditional basis* and that is why wavelets are so effective in signal compression.

2. The wavelet expansion allows a more accurate local description and separation of signal characteristics. A Fourier coefficient represents a component that lasts for all time and therefore, temporary events must be described by a phase characteristic that allows cancellation or reinforcement over large time periods. A wavelet expansion coefficient represents a component that itself is local and easier to interpret. The wavelet expansion may allow a separation of components of a signal that overlap in both time and frequency.

3. Wavelets are adjustable and adaptable. They can be designed to fit to individual applications.

4. Generation of wavelets and the calculation of DWT [66,67,69,71] are well suited for implementation on a digital computer. This is because the defining equation of wavelet contains no derivatives or integrals, just multiplications and additions.

3.3 QMF, Wavelets and Time – Scale Algorithms

Galand in 1977 [87], faced with the impossibility of realizing subband coding using $m > 2$ bands covering the frequency space regularly and having finite length filters, (which is essential for applications), limited himself to the case of just two ($m = 2$) frequency channels. The complete analysis and synthesis scheme is given in Fig. 3.2. The low Pass Filter (LPF), say $\mathbf{F_0}$, and the High Pass Filter (HPF), say $\mathbf{F_1}$, are called Quadrature Mirror Filters (QMF) if, for all signals X of finite energy, one has:

$$\|Y_0\|^2 + \|Y_1\|^2 = \|X\|^2 \tag{3.9}$$

This condition is called the perfect reconstruction property, and in fact is nothing other than requirement of preserving energy.

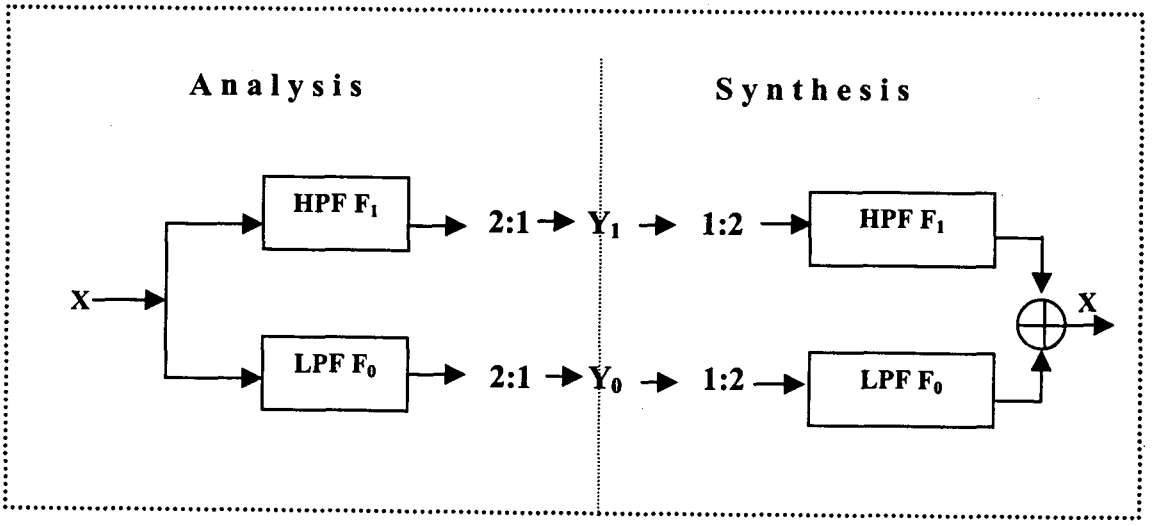


Fig. 3.2: Subband Coding Scheme

Mallat [59,71] used these QMF to construct, using a hierarchical subband coding scheme, time – scale algorithms. He considers an increasing sequence of nested grids or scales, that go from the "fine grid" to the "coarse grid". The signal to be analysed has been sampled on the fine grid, and our starting point is thus a sequence $\mathbf{f_0}$.

We process the signal $\mathbf{f_0}$ by decomposing it into its trend (approximation) and fluctuations (details). The trend then, is sampled on the next grid, it now represents a

new signal that is decomposed again into trend and fluctuation. The fluctuations (or details) are never analysed in the hierarchical scheme and the algorithm follows a "herringbone" pattern. This is shown in Fig. 3.3.

The input signal f_0 is finally represented by the sequence $r_1, r_2, \ldots\ldots, r_N$ of fluctuations and by the last trend f_N. The transformation that maps f_0 onto the sequence $(r_1, r_2, \ldots\ldots, r_N, f_N)$ is clearly orthogonal (preserving energy), and the inverse is immediately calculated, based on the perfect reconstruction property of the QMF.

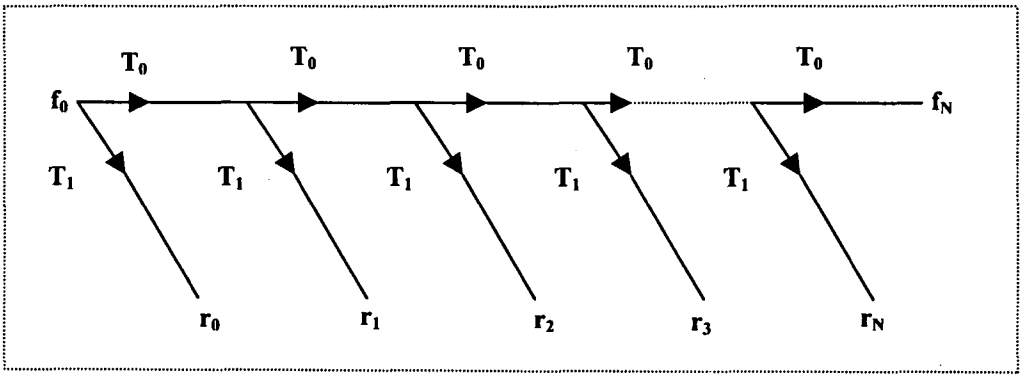


Fig.3.3: Flow diagram of time-scale analysis algorithm based on hierarchical subband coding scheme.

In the above figure, branches T_0 and T_1 stand for the LPF F_0 ,and the HPF F_1 respectively, plus the decimation process. The significance of Mallat's algorithm stems from the following two observations :

- for appropriate choice of the filters F_0 and F_1 , there are numerous cases where the fluctuations (detail coefficients) $r_1, r_2, \ldots\ldots, r_N$ are, at different stages, extremely small. Coding the signal thus comes down to coding the last trend f_N (approximation coefficients) as well as those coefficients of the fluctuations that are above the threshold fixed by the quantization, and
- the above algorithm converges to the analysis of f_0 in an orthonormal wavelet basis under certain conditions on F_0 and F_1.

That is, given two QMF $\mathbf{F_0}$ and $\mathbf{F_1}$, it is possible to associate with them two functions $\phi(t)$ and $\psi(t)$, called respectively, "the father" and "the mother" wavelet. More precisely we have:

$$\phi_{i,k}(t) = 2^{\frac{i}{2}}\phi(2^i t - k)$$
$$\psi_{j,k}(t) = 2^{\frac{i}{2}}\psi(2^i t - k) \qquad i, k \in Z \qquad (3.10)$$

Using "Mallat's program", Ingrid Daubechies discovered orthonormal wavelet bases having preselected regularity and compact support. The only previously known case was the Haar system (1909), which is not regular. Thus 80 years separated Alfred Haar's work and its natural extension by Daubechies.

The first stage of filter bank divides the spectrum of the signal into a low pass and a high pass band. The second stage then divides that low pass band into another low pass band and a high pass and so on. This results in logarithmic set of bandwidths as shown in the Fig. 3.4 below.

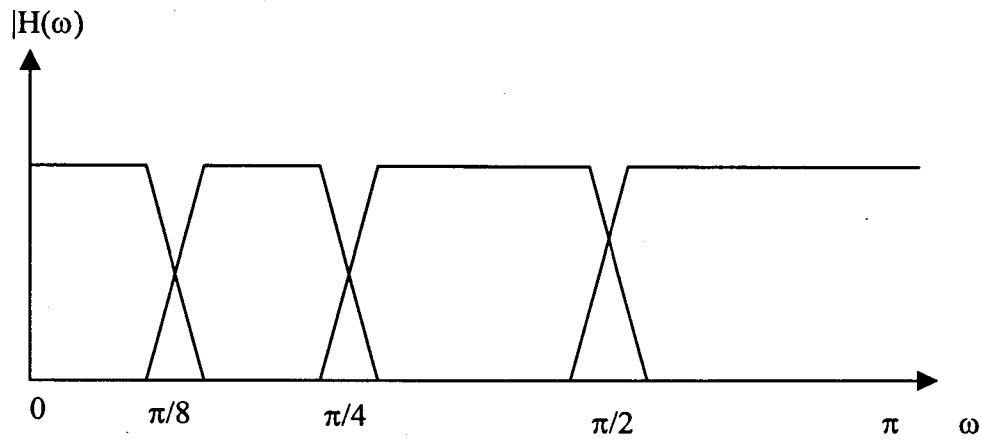


Fig.3.4: Logarithmic set of bandwidths (Constant Q behaviour of the wavelet transform)

For any practical signal that is band limited, there will be an upper level (stage) j = J, above which the wavelet coefficients are negligibly small.

3.4 Wavelet Filter Coefficients (WFC) – Fourier Domain

Let $\mathbf{F_0}(\theta)$ and $\mathbf{F_1}(\theta)$, ($\theta \in [0,2\pi]$), denote the Fourier transform of the QMF $\mathbf{F_0}$ and $\mathbf{F_1}$, respectively. Then we define:

$$m_0(\theta) = \frac{1}{\sqrt{2}} \sum_{k=0}^{k=2N-1} c_k e^{-jk\theta} \quad m_1(\theta) = \frac{1}{\sqrt{2}} \sum_{k=0}^{k=2N-1} b_k e^{-jk\theta} \tag{3.11}$$

such that

$$F_0(\theta) = \sqrt{2}\ \overline{m_0(\theta)}$$
$$F_1(\theta) = \sqrt{2}\ \overline{m_1(\theta)} \tag{3.12}$$

where bar denotes complex conjugation and j stands for the complex operator.
The construction then of an orthonormal wavelet basis, that forms a multiresolution analysis, is based on a solution of a two scale difference equation (dilation equation):

$$\phi(t) = \sqrt{2} \sum_{k=0}^{2N-1} c_k \phi(2t-k) \tag{3.13}$$

where $\phi(t)$ is known as a scaling function or father wavelet. The mother wavelet $\Psi(t)$ is then obtained from $\phi(t)$ as

$$\psi(t) = \sqrt{2} \sum_{k=0}^{2N-1} b_k \phi(2t-k) \tag{3.14}$$

Let $\phi(\theta)$ and $\Psi(\theta)$, denote the Fourier transform of $\phi(t)$ and $\Psi(t)$, respectively. Then the two functional equations (3.13) and (3.14) lead, in the Fourier domain, to the solution:

$$\phi(\theta) = m_0\left(\frac{\theta}{2}\right) m_0\left(\frac{\theta}{4}\right) \ldots\ldots\ldots\ldots m_0\left(\frac{\theta}{2^i}\right) \ldots\ldots$$
$$\psi(\theta) = m_1\left(\frac{\theta}{2}\right) \phi\left(\frac{\theta}{2}\right) \tag{3.15}$$

Where $N \geq 1$ defines the support of the above functions $\phi(t)$ and $\psi(t)$, namely [0, 2N-1], as well as the regularity of them. The wavelet filter coefficients $\mathbf{c_k}$ and $\mathbf{b_k}$ constitute the impulse response of the filters F_0 and F_1, respectively, and are real.

In order for the above solution Eq. 3.15 to be valid:

- the $\mathbf{c_k}$'s are chosen so as to make $\mathbf{F_0}$ yield, as many zeros as possible to an output data vector. This is done by requiring these wavelet filter coefficients to have a certain number of vanishing moments. When this is the case of N moments, the set of wavelet filter coefficients is said to satisfy an "approximation condition of order N" or equivalently that $F_0(\theta)$ has a N^{th} order zero at $\theta = \pi$, namely

$$F_0^{(m)}(\pi) = 0 \qquad m=0,1,\ldots\ldots\ldots, N-1, \text{ and } N \geq 1 \tag{3.16}$$

- the orthogonality or perfect reconstruction condition is required in order for the filters F_0 and F_1 to be QMF. This is expressed in Fourier domain by the following relations:

$$\frac{1}{2}\left|F_0(0)\right|^2 = 1 \qquad \text{and} \qquad \frac{1}{2}\left[\left|F_0(\theta)\right|^2 + \left|F_0(\theta+\pi)\right|^2\right] = 1 \tag{3.17}$$

It is thus relatively straightforward to obtain wavelet filter coefficient sets in Fourier domain by simply inventing a function $\mathbf{F_0}(\theta)$ satisfying the above conditions, Eqs. (3.16) and (3.17). To find the actual $\mathbf{c_k}$'s wavelet filter coefficients, applicable to an input data vector length L, we need just to apply the inverse discrete Fourier transform on $\mathbf{F_0}$

$$c_k = \frac{1}{L}\sum_{l=0}^{L-1} F_0\left(\frac{2\pi l}{L}\right) e^{-2\pi jkl/L} \tag{3.18}$$

The QMF F_1 incidentally, has the Fourier representation

$$F_1(\theta) = e^{-j\theta}\,\overline{F_0(\theta+\pi)} \tag{3.19}$$

In general the above procedure will not produce wavelet filters with compact support. In other words, all of the $\mathbf{c_k}$'s, k=0,....., 2N-1 will in general be nonzero (though they may be rapidly decreasing in magnitude). The Daubechies wavelets, or other wavelets with compact support, are specially chosen so that $\mathbf{F_0}$ is a trigonometric polynomial with only a small number of Fourier components, guaranteeing that there will be only a small number of nonzero $\mathbf{c_k}$'s.

3.5 Wavelet Filter Coefficients (WFC) – Time Domain

The Wavelet Filter Coeficients $\{c_k\}_{k=0}^{k=2N-1}$ satisfy certain constraints, or in other words Eqs. (3.16) and (3.17) in the time-domain are equivalent to:

- the "approximation of order N" or vanishing moments condition

$$\sum_{k=0}^{k=2N-1} (-1)^k k^m c_k = 0 \tag{3.20}$$

where m=0,1,....., N-1, and $N \geq 1$

- the orthogonality condition

$$\sum_{k=0}^{k=2N-1} c_k c_{k+2m} = 2\delta_{0,m} \tag{3.21}$$

where m=0,1,.....,N-1, and $N \geq 1$

- and the normalisation condition

$$\sum_{k=0}^{k=2N-1} c_k = \sqrt{2} \tag{3.22}$$

Under these constraints, the values $\mathbf{c_k}$ are the the coefficients of the Low Pass Filter $\mathbf{F_0}$ and $\mathbf{b_k}$ are the coefficients of the High Pass Filter F_1, related by:

$$b_k = (-1)^k c_{2N-1-k} \tag{3.23}$$

Thus the High Pass Filter coefficients $\mathbf{b_k}$**'s** are obtained from the Low Pass Filter coefficients $\mathbf{c_k}$'s with reversed order and alternating signs.

3.6 Wavelet Filter Coefficients (WFC) – z Domain

Another technique to generate a magnitude square function that satisfies the perfect – reconstruction (PR) QMF conditions is based on Bernstein polynomials [46]. Several well known orthonormal wavelet filters including Daubechies and Coiflet families emerge as special cases of this technique.

Let **c(k)** be a length 2N low-pass filter with the system function

$$F_0(z) = \sum_{k=0}^{2N-1} c(k) z^{-k} \tag{3.24}$$

The PR-QMF must satisfy

$$F_0(z)F_0(z^{-1}) + F_0(-z)F_0(-z^{-1}) = 1 \tag{3.25}$$

The corresponding magnitude square function $\mathbf{R_0(z)}$ then in **z** domain is given by [46]

$$F_0(z)F_o(z^{-1}) = \left|F_0(e^{j\omega})\right|^2 = R_0(z) \tag{3.26}$$

3.7 Biorthogonal Wavelets and Filterbanks

It is well known in the subband filtering community that symmetry and exact reconstruction are incompatible, if the same FIR filters are used for reconstruction and decomposition. As soon as this last requirement is given up, symmetry is possible. This added flexibility in the biorthogonal case permits use of linear phase filters and unequal length filters. This is shown in Figs 3.5 and 3.6. Detailed design procedures for linear phase wavelets and filter banks are described in [69]. In a biorthogonal filterbank with analysis/synthesis filters $F_0(z)$, $F_1(z)$, $\tilde{F}_0(z)$ and $\tilde{F}_1(z)$, perfect reconstruction with FIR filters means that

$$\tilde{F}_0(z)\ F_0(z)\ +\ \ \tilde{F}_0(-z)\ F_0(-z) = 2 \tag{3.27}$$

$$F_1(z) = -z^{2k+1}\tilde{F}_0(-z) \tag{3.28}$$

$$\tilde{F}_1(z) = z^{-2k-1} F_0(-z) \tag{3.29}$$

$$F_0(z) = F(z) \tag{3.30}$$

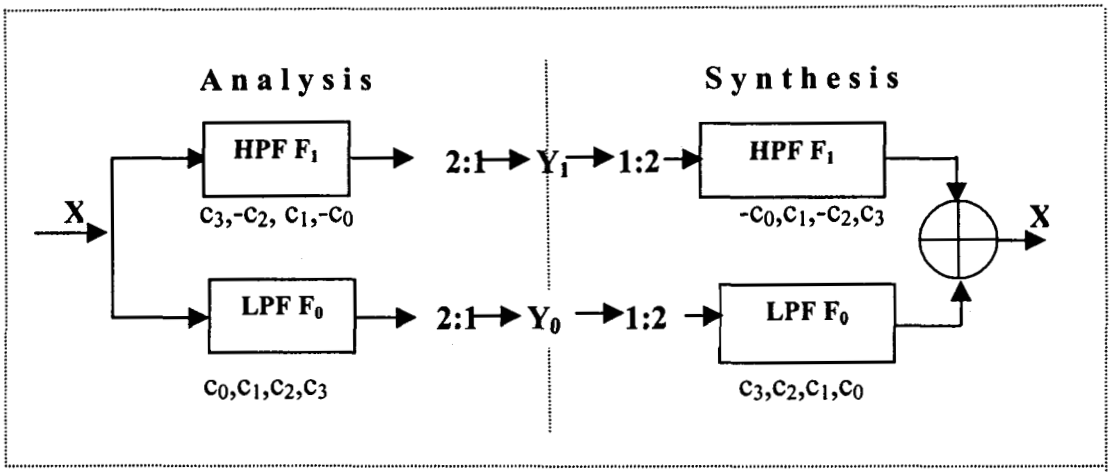


Fig. 3.5: Orthogonal subband coding scheme

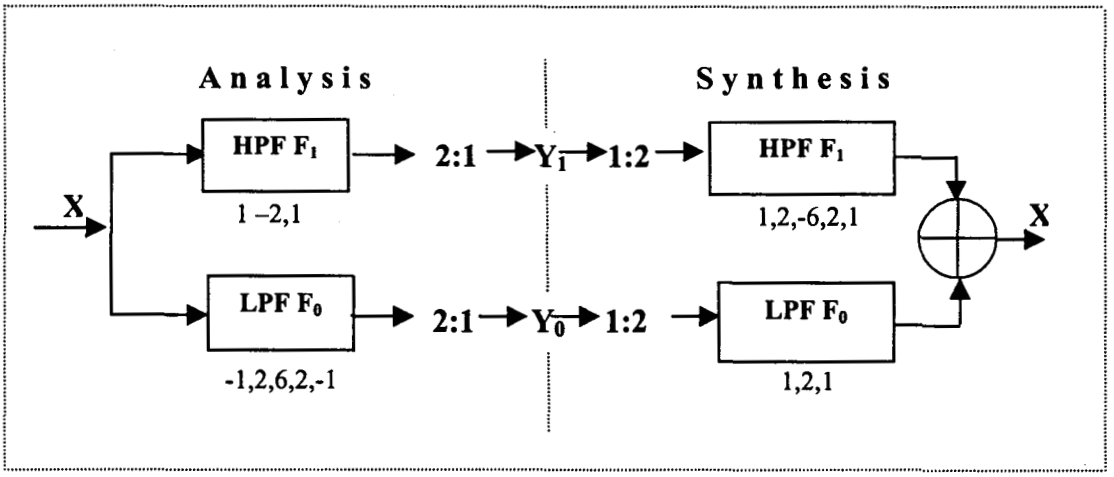


Fig. 3.6: Biorthogonal subband coding scheme

However, there is a drawback to be noted in this structure. The biorthogonal nature of the filterbank (Fig. 3.6) allows different filter lengths in the analysis section and consequently an unequal split of the signal spectrum into low and high-band segments. Since each of these bands is followed by a down-sampler of rate 2, there is an inherent mismatch between the analysis filters and the decimation factor. In turn, the synthesis or interpolation stage has the same drawback. Quantizing and encoding

these decimated subband signals can cause more degradation than would be the case for an almost linear phase solution for example, using the "less asymmetric" wavelet filter family.

3.8 Wavelet Packet Analysis

The wavelet packet (WP) method is a generalization of wavelet decomposition that offers richer range of possibilities for signal analysis. In wavelet analysis, a signal is split into an approximation (trend) and a detail (fluctuation). This approximation is then itself split into a second level approximation and detail and the process is repeated. This results in a logarithmic frequency resolution. This is called "constant Q" filtering and is appropriate for some applications, but not all. We have seen that for an n-level decomposition, there are n+1 possible ways to decompose or encode the signal as shown in Fig 3.3.

In wavelet packet system the details as well as approximations can be split. This yields 2^n different ways to encode the signal. This gives a richer structure that allows adaptation to particular signals or signal classes. The cost of this richer structure is a computational complexity of O (N log (N)), in contrast to DWT which is O (N).

The structure of a wavelet packet decomposition tree is shown in Fig.3.7. One major advantage of wavelet packet analysis is that we can have an irregular subband tree structure also, to meet a specific application (eg. A structure approximating critical bands of the human ear). From Fig. 3.7, it is clear that the signal X can be represented in various ways as described below:

$$X = A_1 + D_1$$

$$X = AA_2 + DA_2 + D_1$$

$$X = AD_2 + DA_2 + A_1$$

$$X = AA_2 + DA_2 + AD_2 + DA_2$$

But this type of representation is not possible with ordinary wavelet analysis. Hence WP analysis gives a flexible structure which can be tailored to meet the required specifications.

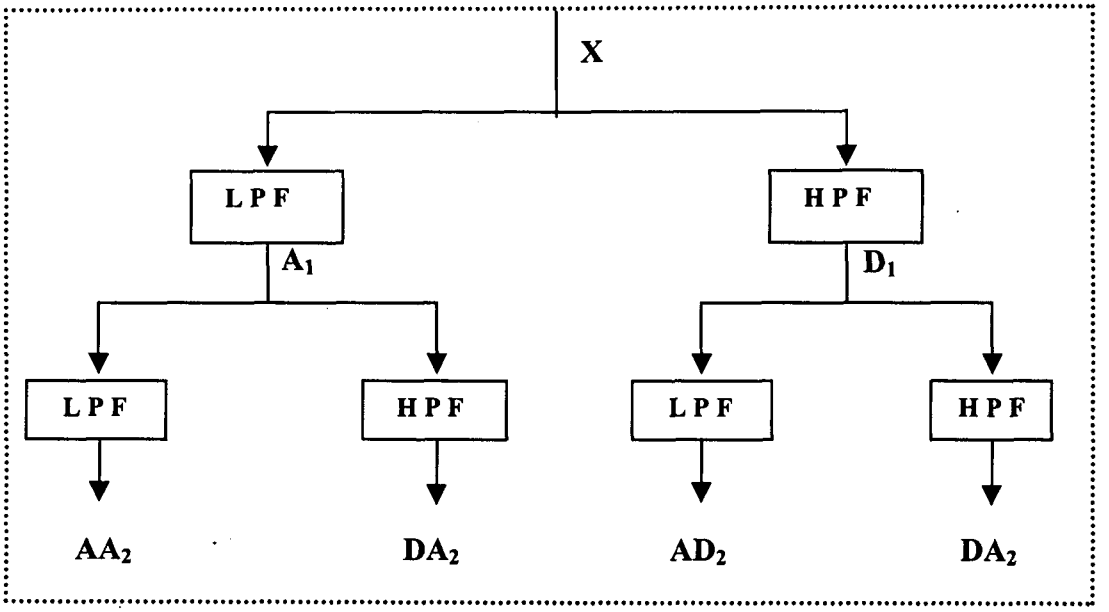


Fig. 3.7: Wavelet Packet Implementation

3.9 Properties of Wavelet Filters

Signal compression was claimed to be a major potential application of wavelets. In fact the wavelet transforms was soon recognized to be equivalent to filterbanks allowing perfect reconstruction, which was successfully applied for some time in subband coding of speech and images. The main novelty of wavelets compared to traditional subband coding is the additional requirement of regularity on the filters. Here, regularity is a smoothness requirement on a continuous-time function generated by $F_0(z)$, and can be mathematically defined as continuity of this function and its derivatives. To be more specific, if a wavelet is **m** times differentiable and its m^{th} derivative is Hölder continuous of order α, then its regularity will be $m+\alpha$ [88]. However, the actual impact of regularity is not clear at this time. Therefore, it is important to understand its role in coding systems, in competition with other filter properties and features such as vanishing moments or frequency selectivity.

For example, for fixed support width of ϕ,ψ ([0,2N-1]), or equivalently, for fixed length of filters ([2N-1, starting with the zero]), in the associated subband coding scheme, the choice of the wavelet filter coefficients $\mathbf{c_k}$'s that lead to maximum regularity is different from the choice with maximum number N of vanishing moments for ψ. The question then arises: what is more important, vanishing moments or regularity? The answer depends on the application, and is not always clear.

Another property related to wavelet filters is symmetry. Symmetric filters are often called linear phase filter. If a filter is not symmetric, then its deviation from symmetry is judged by how much its phase deviates from a linear function. More precisely, a filter with filter coefficients $\mathbf{c_k}$ is called linear phase if the phase of the function

$$f(\theta) = \sum_k c_k e^{-jk\theta} \tag{3.31}$$

is a linear function of θ, i.e. if, for some $l \in Z$, $f(\theta) = e^{-jl\theta} \mid f(\theta) \mid$. This means that the $\mathbf{c_k}$ are symmetric around l, $\mathbf{c_k} = \mathbf{c_{2l-k}}$.

All real orthonormal wavelet bases with compact support are asymmetric, except for the Haar basis (N=1). The non-existence of symmetric or antisymmetric real compactly supported wavelets should be no surprise to anybody familiar with subband coding: It had already been noted by Smith and Barnwell [89] that symmetry is not compatible with the exact reconstruction property in subband filtering.

3.9.1 Regularity and Vanishing moments

Beylkin, Coifman, and Rokhlin use compactly supported orthonormal wavelets to compress large matrices, i.e., to reduce them to a sparse form [83]. One of the things that makes their method work is the number of vanishing moments. Suppose we want to decompose a function f into wavelets. We compute all the wavelet coefficients $< f, \psi_{j,k} >$, and to compress all that information, we discard all the coefficients smaller than some threshold ε. (This is a rather primitive procedure. In practice, one chooses to allocate more precision to some coefficients than to others, by means of a

quantization rule). After thresholding, we will only retain fine-scale wavelet coefficients near singularities of f or its derivatives. The effect will be all the more pronounced if the number N of vanishing moments of ψ is large. Note that the regularity of ψ does not play a role at all in this argument, it seems that for compression of matrices-type applications the number of vanishing moments is far more important than the regularity of ψ.

For other applications, regularity may be more relevant. Suppose we want to compress the information in an image. At least some regularity might be required. Some initial experiments reported in Antonini et.al.[90] seem to confirm this, but more experiments are required for a convincing answer.

The idea of generating regular functions from a repeated interpolation schemes is not new. It appears in computer – aided geometric design, and the dependence of $\phi(t)$ on F_0 (z) was also observed by Burt and Adelson in the context of pyramid transforms [91]. These works were performed independently of wavelets where regularity does not depend on the perfect reconstruction property of filter banks. But when introducing compactly supported wavelets, Daubechies stated new problems:under which (necessary and sufficient) conditions on $F_0(z)$ do we have convergence of c_k^i to $\phi(t)$ and regularity of the limit function?

These questions are motivated by the mathematics behind regularity, where the reasoning comes from wavelet analysis for continuous-signals. But this mathematical approach seems inadequate for digital signal processing applications, because in practical systems the filterbank is iterated a finite number of times. In wavelet based image or speech compression schemes, **i** seldom exceeds 10, whereas regularity is mathematically defined when $\mathbf{i} \rightarrow \infty$.

However, we know from Mallat's work that limit functions still underly a discrete filterbank, even when it is not iterated indefinitely. Therefore, we would like to understand regularity in terms of the discrete filterbank impulse responses c_k^i and b_k^i.

Regularity does impose some "smoothness" on c_k^i and b_k^i. These are intuitive arguments in favour of this for coding applications.

- During analysis: Suppose that a smooth portion of the input is analysed by "nonregular" filters, whose impulse responses rapidly cause discontinuities as **i** increases. Then, these artificial discontinuities, not due the signal itself, appear in the wavelet transform coefficients. In other words, regularity would lead to a "better" representation of the signal by these coefficients.

- During synthesis: Suppose that an error, eg.,(a quantization error) is made in one coefficient corresponding to some decomposition level **i.** In the reconstructed signal, this results in a perturbation that is proportional to the equivalent impulse response corresponding to this level. In applications such as image compression, a perturbation presenting discontinuities is likely to "strike the eye" more than a smooth one. While in applications such as audio compression, it is likely to "irritate the ear" by producing glitches. Also, its amplitude increases for high compression rates, when transform coefficients are coarsely quantized. Therefore, it is natural to require that this perturbation be smooth.

One of the goals of this thesis is to investigate whether such arguments are relevant or not in practical perceptual audio compression systems. However, some results which have been reported by other researchers are discussed below.

Rioul [88] provides an example of image coding application, in which the effect of regularity on compression performance is measured, using orthogonal wavelets. From this performance comparison for different filter lengths it seems that the coding performance globally increases with filter length, which also increases regularity. However an asymptote is quickly attained. Above L=10 or 12, for which the regularity does not exceed 2, performance does not improve much.

Results obtained for a simple compression scheme using various coding criteria, optimised rate/distortion, and a number of paraunitary filters with balanced regularity and frequency selectivity, show that regularity may be relevant for still

image compression, at least for short filters (L 10), for which the regularity order is relatively small. Using more regular filters is probably useless, as the compression performance is not improving for longer filters.

Sinha and Tewfik [42], suggest that regularity can play a role in the coding of audio signals. In fact they claim that longer sequence yield better results: This conclusion is not surprising since longer sequences (thus more regular wavelet filters) correspond to wavelet filter banks with sharper transition bandwidths, i.e., to a better separation of frequency information.

However, Barnwell III and Richardson have a different opinion: Analysis-synthesis systems based on filterbanks, iteratively satisfy a set of constraints on the filters and the reconstruction properties of the analysis-synthesis system. To make the system into a DTWT, in general an extra set of constraints (mainly associated with regularity) must be satisfied. The primary function of the regularity constraint in wavelets is to guarantee that the total octave band decomposition will converge, in the limit, to a smooth wavelet function. In practice, however, DTWTs must be realized with a finite number of bands, so the regularity constraint generally results in subband systems whose performance is measurably worse in terms of filter performance (passband, stopband, transition band), reconstruction error, delay, compactness (length), orthogonality, computational complexity, etc. as compared to systems which do not have such constraints.

From Philippe et.al. [44], it can be noted that the most significant criterion for the design of wavelet filters for audio compression is the so called "coding gain"[58], while frequency selectivity, regularity, phase, and orthogonality seem less relevant.

3.9.2 Symmetry

If the restriction that ϕ be real is lifted, then symmetry is possible, even if ϕ is compactly supported. But do we need symmetry? Why should we care? For some applications it does not really matter at all. The numerical analysis in Beylkin, Coifman, and Rokhlin for instance, works very well with asymmetric wavelets. For

other applications, the asymmetry can be a nuisance. In image coding, for example, quantization errors will be often most prominent around edges in the images. It is a property of our visual system that we are more tolerant of symmetric errors than asymmetric ones. In other words, less asymmetry would result in greater compressibility for the same perceptual error. (Note also that "perceptually" small or large errors are difficult to quantify mathematically). The norm most often used to measure "distance" is the l^2- norm, but it is more because this is the easiest norm to handle than any other reason. All experts agree, that the l^2-norm is not a good candidate for a "perceptual" norm, but there does not seem to be agreement on a better candidate. Moreover, symmetric filters make it easier to deal with the boundaries of the image, another reason why the subband coding engineering literature often sticks to symmetry. We can recover symmetry if we give up orthogonality.

Finally a large number of vanishing moments of ψ leads to much more "compression potential" in the regions where **f** is reasonably smooth. It is not clear however, whether the high number of vanishing moments of ψ or the regularity of ψ is the most important factor ; it is possible that they are both important. However, symmetry can be achieved by using *coiflets* or even better, biorthogonal wavelet bases. Note also, that linear phase is desirable for computational purposes, because of the symmetry of the filters.

One of the purposes of this thesis is to identify the "best" wavelet from the perceptual audio compression point of view. So a number of different filter families described below, have been evaluated for their performance in low bit rate audio coding. These filter families have been chosen because of their different properties.

3.10 Wavelet Filter Families used in this work

Several families of wavelets that have different properties and proven to be especially useful are included in the present work. Wavelet filter coefficients are given in Appendix A. Scaling functions and wavelet functions are discussed in the next section.

Haar : Any discussion of wavelets begins with Haar, the first and simplest. Haar is discontinuous (Regularity = 0) and resembles a step function. The scaling and wavelet functions are shown in Fig.3.8. It represents the same wavelet as Daubechies (db1). Main properties of this wavelet are:

$\psi(x) = 1$ if $0 \leq x < \frac{1}{2}$ $\psi(x) = -1$ if $1/2 \leq x < 1$

$\phi(x) = 1$ if $0 \leq x < 1$ $\phi(x) = 0; \psi(x) = 0$ if $x \notin [0,1]$

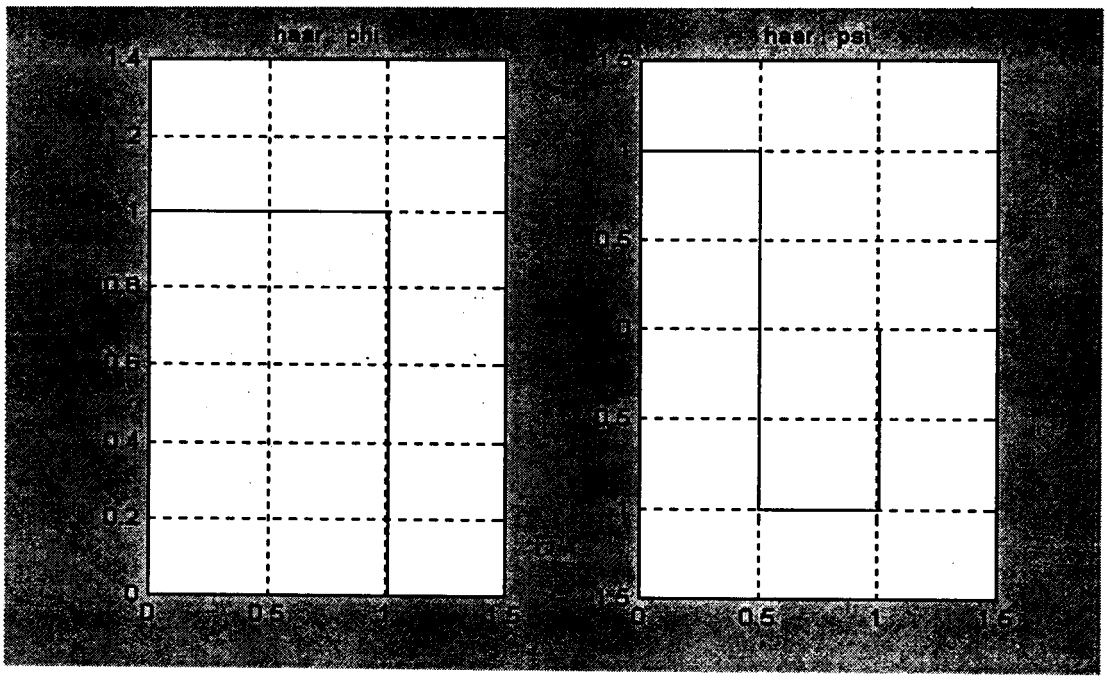


Fig. 3.8: Haar - Scaling and wavelet functions

Daubechies: Ingrid Daubechies, one of the brightest stars in the world of wavelet research, invented what are called compactly supported orthonormal wavelets - thus making discrete wavelet analysis practicable.

The names of the Daubechies family wavelets are written as dbN, where N is the order and db the "surname" of the wavelet. The db1 wavelet as mentioned above, is the same as 'Haar'. 'db4' scaling and wavelet functions are shown in Fig.3.9. These wavelets dbN have no explicit expression except for db1(Haar) wavelet.

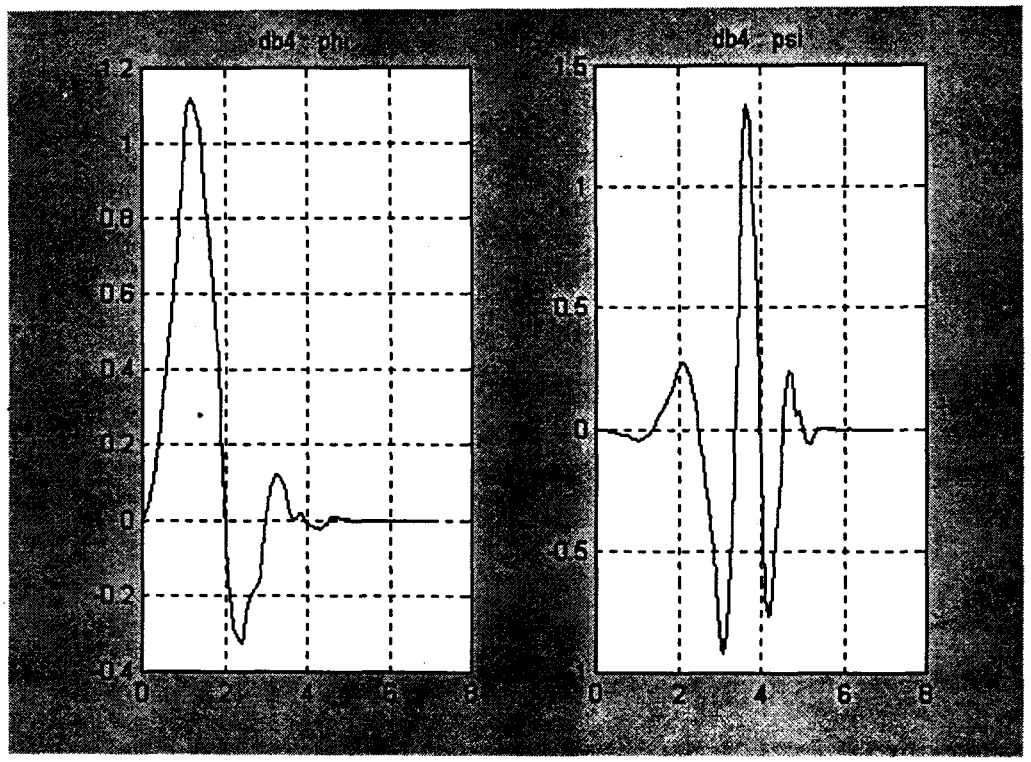


Fig. 3.9: Daubechies - Scaling and wavelet Functions

The support length of Ψ and ϕ is 2N-1. The number of vanishing moments of Ψ is N.

- dbN are not symmetrical.
- The regularity increases with order.
- The analysis is orthogonal.

Symlet Wavelets : symN The symlets are nearly symmetrical wavelets proposed by Daubechies as modifications to the db family. The properties of two wavelet families are similar. In symN (see Fig.3.10), N is the order. Since they are only near symmetric, some authors do not call them symlets. Daubechies proposed modifications for 'db' wavelets such that their symmetry can be increased while retaining great simplicity.

Coiflet Wavelets : coifN is built by I. Daubechies at the request of R.Coifman. The wavelet function ψ has 2N moments equal to 0 and the scaling function ϕ has 2N - 1 moments equal to 0. These functions are shown in Fig.3.11. In coifN, N is the order of the wavelet. The coifN ψ and ϕ are much more symmetrical than the dbNs.

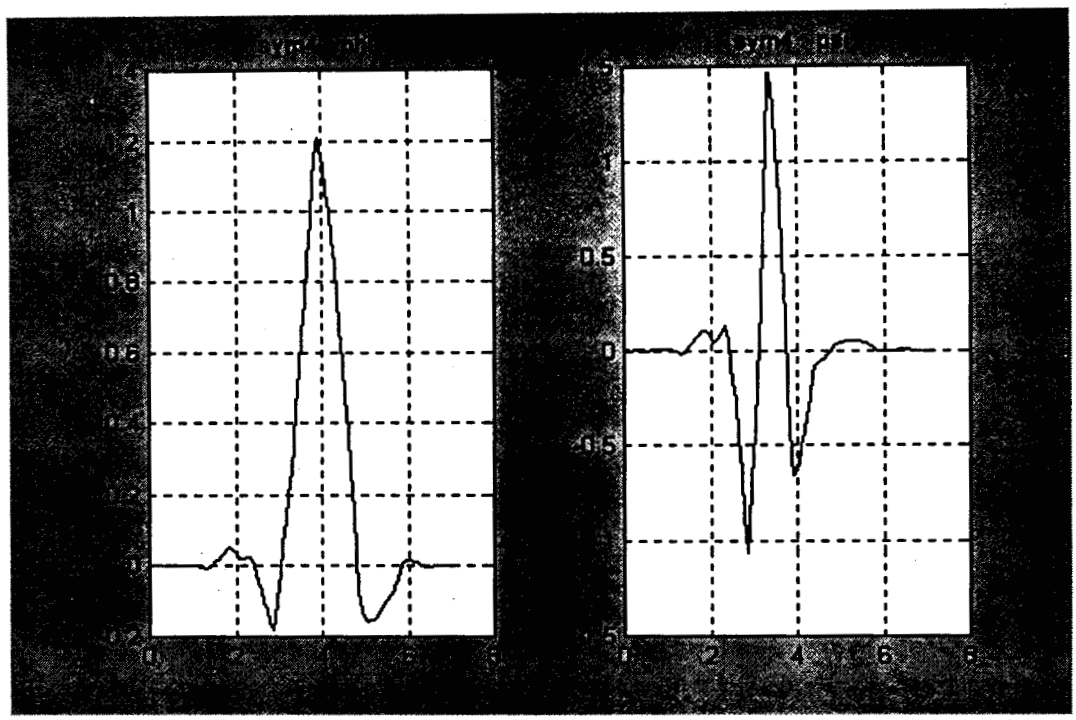


Fig. 3.10: Symlet - Scaling and wavelet functions

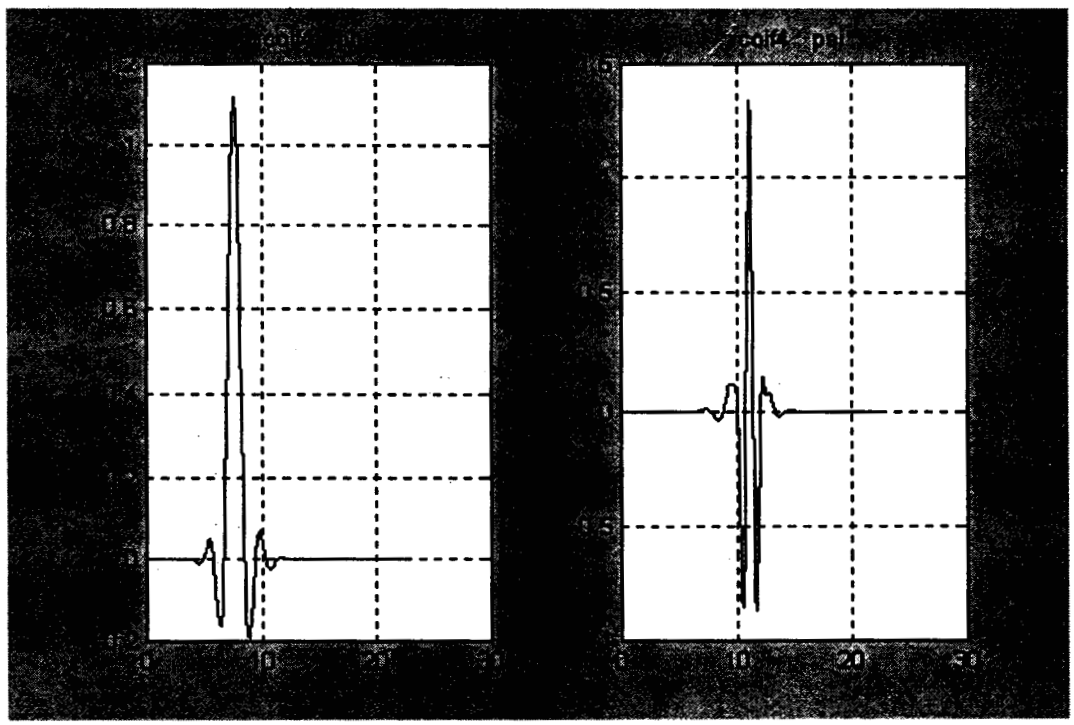


Fig. 3.11: Coiflet - Scaling and Wavelet functions

Biorthogonal Wavelet pairs : bior Nr.Nd (see Fig.3.12) This family of wavelets exhibits the property of linear phase which is mainly needed for image reconstruction. By using two wavelets, one for decomposition and the other for reconstruction, interesting properties are derived.

It is well known in the subband filtering community that symmetry and exact reconstruction are incompatible, if the same FIR filters are used for reconstruction and decomposition. Two wavelets, instead of just one, are introduced.

- One, $\tilde{\Psi}$, is used in the analysis and the coefficients of a signal S are

$$\tilde{c}_{j,k} = \int S(x)\tilde{\Psi}_{j,k}(x)dx$$

- The other, Ψ, is used in the synthesis

$$S = \sum_{j,k} \tilde{c}_{j,k}\Psi_{j,k}$$

In addition, the wavelets are related by duality in the following sense,

$$\int \tilde{\Psi}_{j,k}(x)\Psi_{j',k'}(x)dx = 0 \quad \text{as soon as } j \neq j' \text{ or } k = k' \text{ and even}$$

$$\int \tilde{\phi}_{0,k}(x)\phi_{(0,k')}(x)dx = 0 \quad \text{as soon as } k \neq k'.$$

The useful properties for analysis (eg : oscillations, zero moments) can be concentrated on the $\tilde{\Psi}$ function, whereas the interesting properties for synthesis (regularity) are assigned to the Ψ function. The separation of these two tasks proves very useful.

- Ψ, $\tilde{\Psi}$ can have very different regularity properties, Ψ being more regular than $\tilde{\Psi}$.
- The $\tilde{\Psi}$, Ψ, $\tilde{\phi}$ and ϕ functions are zero outside of a segment.
- The filters are symmetrical.

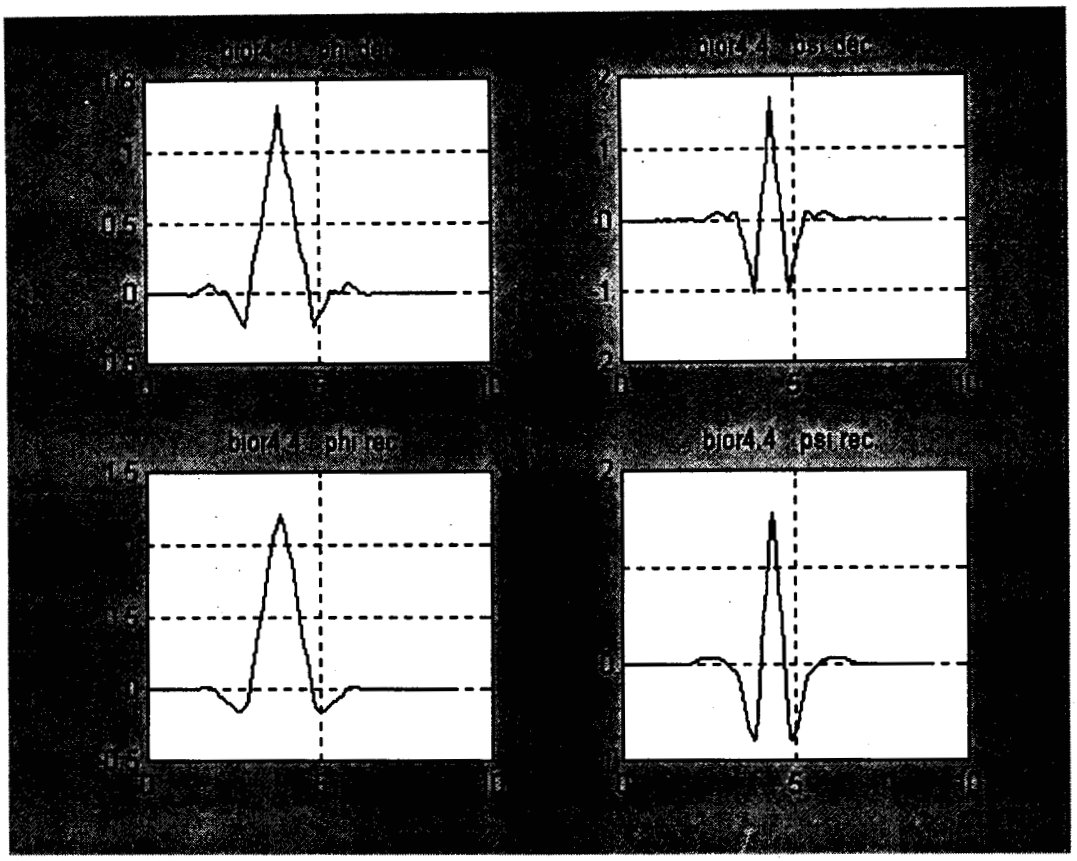


Fig. 3.12: Biorthogonal spline - Scaling and wavelet functions

The summary of wavelet families used in the present research work and associated properties are given in Table 3.1.

Table 3.1 Summary of wavelet families used and associated properties

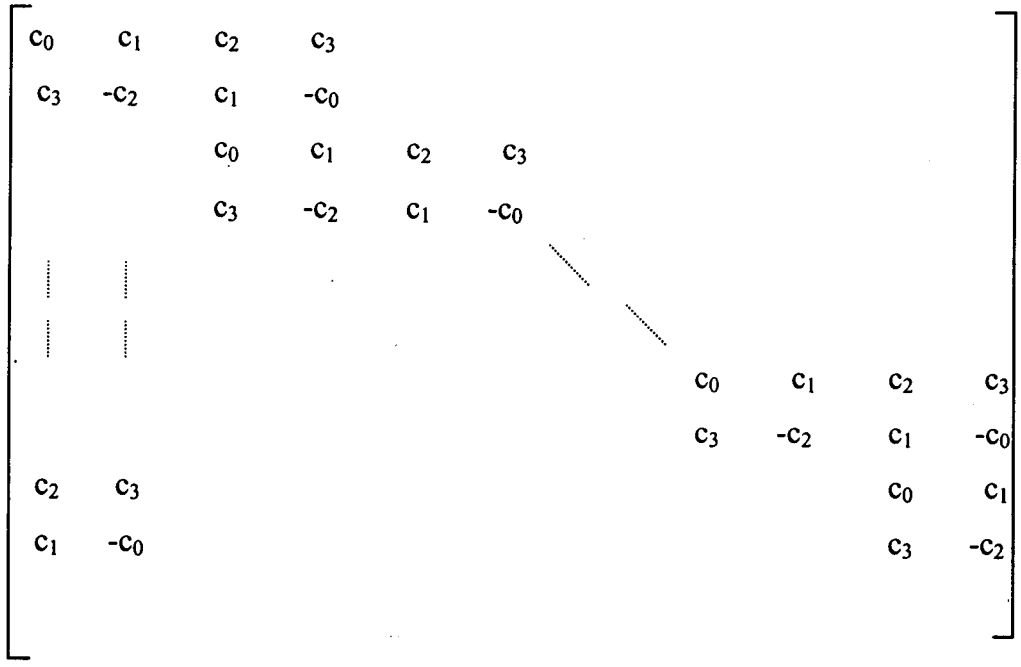
Property	haar	dbN	symN	coifN	bior Nr.Nd
Compactly supported orthogonal	•	•	•	•	
Compactly supported biorthogonal					•
Symmetry	•				•
Asymmetry		•			
Near symmetry			•	•	
Arbitrary number of vanishing moments		•	•	•	•
Arbitrary regularity		•	•	•	•
Orthogonal Analysis	•	•	•	•	
Exact Reconstruction	•	•	•	•	•
FIR filter implementation	•	•	•	•	•
Fast Algorithm	•	•	•	•	•
Explicit expression	•				for splines

3.11 Discrete Wavelet Transform (DWT) Algorithm

We have seen that the c_k's, k=0,.....,2N-1, constitute the impulse response of the QMF F_0 and that the impulse response of the QMF F_1 is related by $b_k = (-1)^k c_{2N-1-k}$.

For the moment, for ease of notation we will restrict ourselves to the simplest, although the most localised, member of the class of wavelets discovered by Daubechies, with four coefficients c_0, c_1,c_2,c_3. In other words we require to satisfy the approximation condition of order N = 2. The number of coefficients increases by two

each time N is increased by one. Consider now the transformation matrix shown below, acting on a column vector of data to its right:



Blank entries to the above matrix signify zeros. Note the structure of the matrix. The odd rows generate one component of the data convolved with the filter coefficients c_0, c_1, c_2, c_3. However, the even rows perform a different convolution with coefficients c_3, $-c_2$, c_1, $-c_0$. Note how the last rows wrap-around like convolutions with periodic boundary conditions. The action of the matrix, over all, is thus to perform two related convolutions, then to decimate each of them by half (throw away half the values), and interleave the remaining halves.

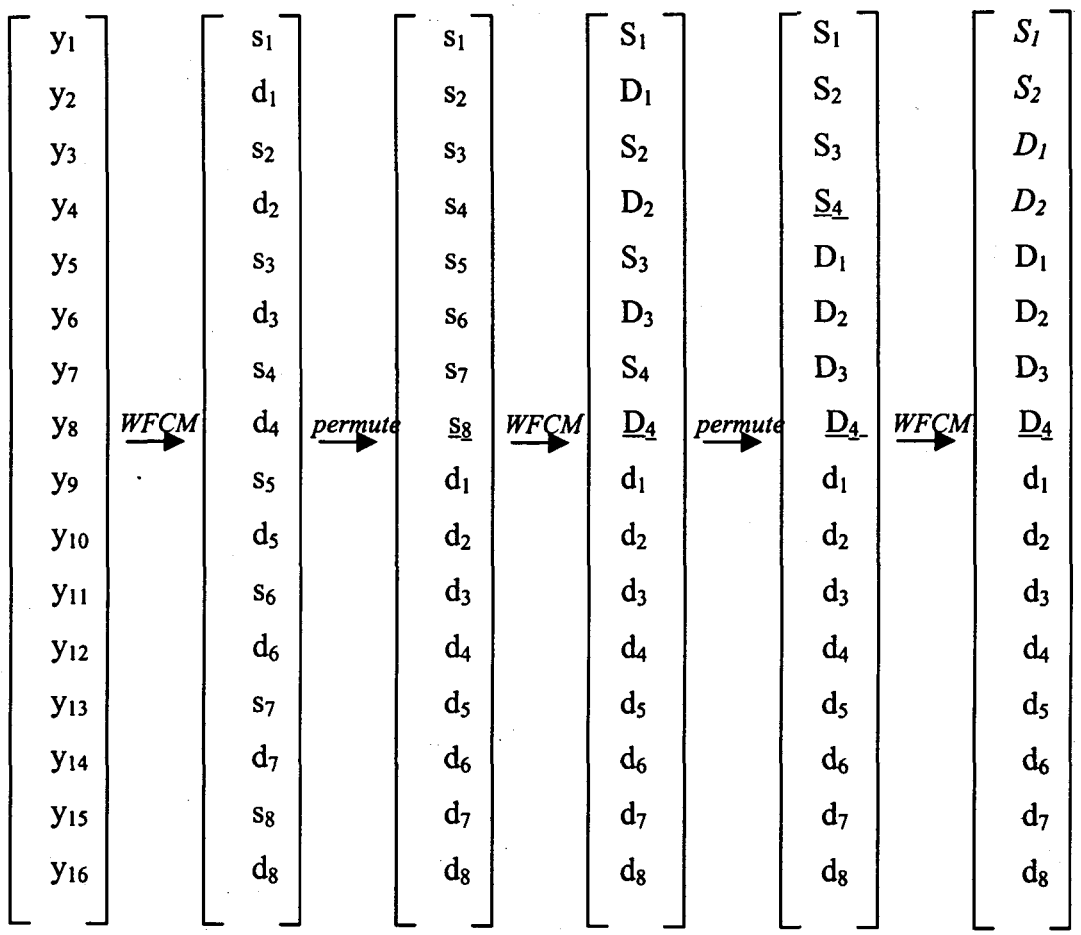
This results in the output of F_0 , decimated by half, accurately representing the data's "approximation" (trend,smooth) information. The output of F_1 ,also decimated, is referred to as the data's "detail" (fluctuation,difference) information. It must be possible to reconstruct the original data vector of length L from its L/2 approximation or **a** - components and its L/2 detail or **d** - components. That is effected by requiring the above matrix to be orthogonal, so that its inverse is just the transposed matrix. (In

fact, we require the wavelet filter coefficients of the matrix to satisfy all the conditions that we have seen before).

The DWT then consists of applying the above wavelet coefficient matrix hierarchically, first to the full data vector of length L, then to the "approximation" vector of length L/2, then to the "approximation-approximation" vector of length L/4, and so on until only a trivial number of "approximation-.....-approximation" components (usually 2) remain. The procedure is sometimes called a Pyramid Algorithm, for obvious reasons. The output of the DWT consists of these remaining components and all the "detail" components that were accumulated along the way. A diagram should make the procedure more clear.

If the length of the data vector were a higher power of two, there would be more stages of applying the wavelet coefficients matrix and permuting. The endpoint will always be a vector with two *S's* and a hierarchy of *D's* , D's and d's etc. Notice that once d's are generated, they simply propagate through to all subsequent stages.

A value d_j, of any level is termed a "wavelet transform detail coefficient" of the original data vector. The final values S_1, S_2 of any level should strictly be called "wavelet transform mother –approximation coefficients", although the term " wavelet coefficients " is often used loosely for both d's and final *S*'s. Since the full procedure is a composition of orthogonal linear operations, the whole DWT is itself an orthogonal linear operator.



To invert the DWT, one simply reverses the procedure, starting with the smallest level of the hierarchy and working the above procedure from right to left. The following inverse matrix is used instead of the first matrix.

$$\begin{bmatrix} c_0 & c_3 & & & & & & & & & c_2 & c_1 \\ c_1 & -c_2 & & & & & & & & & c_3 & -c_0 \\ c_2 & c_1 & c_0 & c_3 & & & & & & & & \\ c_3 & -c_0 & c_1 & -c_2 & & & & & & & & \\ & & & & \ddots & & & & & & & \\ & & & & & c_2 & c_1 & c_0 & c_3 & & & \\ & & & & & c_3 & -c_0 & c_1 & -c_2 & & & \\ & & & & & & & c_2 & c_1 & c_0 & c_3 \\ & & & & & & & c_3 & -c_0 & c_1 & -c_2 \end{bmatrix}$$

3.12 Computational Complexity Analysis

The implementation of DWT is based on the iteration of an elementary block (two-channel filterbank) such as filtering and down sampling.

If the complexity of the first block is C operations/input sample, (in fact, C is typically of the order of the filter length) then the upper bound on the total complexity, irrespective of the number of iterations (stages), is 2C.

The proof is immediate, since the second block has complexity C but runs at half the sampling rate and similarly, the i^{th}, block runs at 2^{i-1} times slower than the first one. Thus, the total complexity for K blocks becomes

$$C_{tot} = C + \frac{C}{2} + \frac{C}{4} + ... + \frac{C}{2^{K-1}} = 2C\left(1 - \frac{1}{2^K}\right) < 2C$$

However, while the complexity remains bounded, the delay does not. If the first block contributes a delay D, the second will produce a delay 2D and the i^{th} block a delay $2^{i-1}D$. That is the total delay becomes

$$D_{tot} = D + 2D + 4D + ... + 2^{K-1}D = (2^K - 1)D$$

Unlike the dyadic subband tree (DWT), in the implementation of the regular binary subband tree (DWP), not only the lower half-band component of the signal is splitted into two equal bands, but the higher half-band component of the signal is splitted as well, at any level of the tree. Therefore the complexity of the regular binary subband tree is dependent on the number of decomposition stages K and becomes

$$C_{tot} = C + C + C + ... + C = \log_2 (L)C = KC$$

where L denotes the transform length.

The delay is the same for both DWT and DWP, assuming parallel decomposition of the low and high band components of the signal in DWP implementation.

3.13 Summary

In this chapter, theory of Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), and Wavelet Packets has been explained. Mallat's algorithm for the fast computation of discrete wavelet transforms is described. Important features (properties) of wavelets were also discussed briefly. Various wavelet families used in this research work and their properties are presented in this chapter. Computational complexity analysis of discrete wavelet transform and discrete wavelet packets has also been done. Perceptual audio coders using discrete wavelet transform and discrete wavelet packets are proposed in the coming chapters.

Chapter 4

DISCRETE WAVELET TRANSFORM BASED PERCEPTUAL AUDIO CODER

4.1 Introduction

A discrete wavelet transform (DWT) based perceptual audio coding scheme is developed and implemented here as an alternative to ISO/MPEG International Audio Coding Standard. As already described in Chapter 2, in the MPEG standard, analysis filter bank splits the signal into 32 equally wide subbands and the lowest one corresponds to the information in the frequency band [0-690 Hz] for 44.1 kHz sampling rate. For these frequency mappings, the time resolution is the same for all frequency sub bands. But we know that the human ear analyses various frequency components with different time resolutions. We have already seen that, the ear can be modelled as a 25-channel filter bank named as critical bands and the bandwidth of these channels grows approximately as *log f*. That is, human ear uses narrow time windows for high frequency components and wider time windows for low frequency components. Therefore a filter bank with very good time resolution for high frequency components and very good frequency resolution for low frequency components, could result in a useful representation for audio coding.

A major drawback of MPEG audio standard known as " **Pre-echo**" distortion, is due to the fixed time resolution of all sub bands. Discrete wavelet transform based coding scheme proposed in this chapter analyses high frequency components using narrow windows and low frequency components using wider windows (similar to the way in which human ear analyses signals). **"Pre-echo"** distortion is almost **eliminated** in this new scheme. A new wavelet filter bank structure to split the audio signal into various sub bands according to the signal characteristics as well as matching to the properties of the human auditory system, is proposed in the present work. By integrating a FFT based psychoacoustic model to determine the kind of distortion that

the human ear will not hear, a compression ratio greater than that of MPEG Layer I and comparable to that of MPEG Layer II with almost transparent quality is achieved by this scheme.

4.2 Design of the Encoder

Block diagram of the proposed audio encoder is shown in Fig. 4.1. Sampling frequency of the audio signal is 44.1 kHz (*i.e.*, CD sampling frequency) and each audio frame is of size 512 samples (*i.e.*, 12 ms duration).

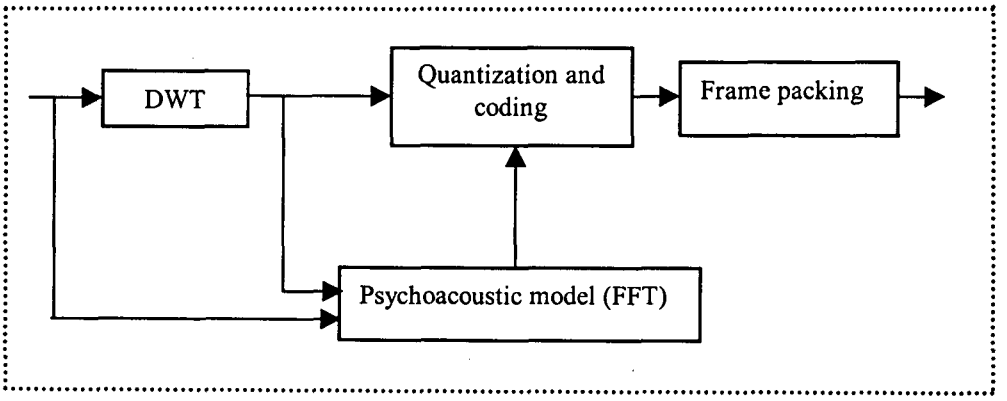


Figure 4.1: Block diagram of the proposed audio coder

- Analysis Filter bank : The signal is decomposed into nine non - uniform sub bands (octave bands) as shown in Fig.4.2, using 8 level DWT. Various subband samples are obtained using Mallat algorithm [69]. The lower and higher cut- off frequencies of each band are also indicated in Fig.4.2.

The tree structure shown in Fig.4.2 is designed in such a way that as the frequency increases, bandwidth of the subbands also increases. This matches with the variation of critical bands of the human ear. Hence, high frequency components are analysed with good time resolution and low frequency components are analysed with good frequency resolution. The mapping between subbands used in the proposed coder and critical bands of the human ear is shown Table 4.1.

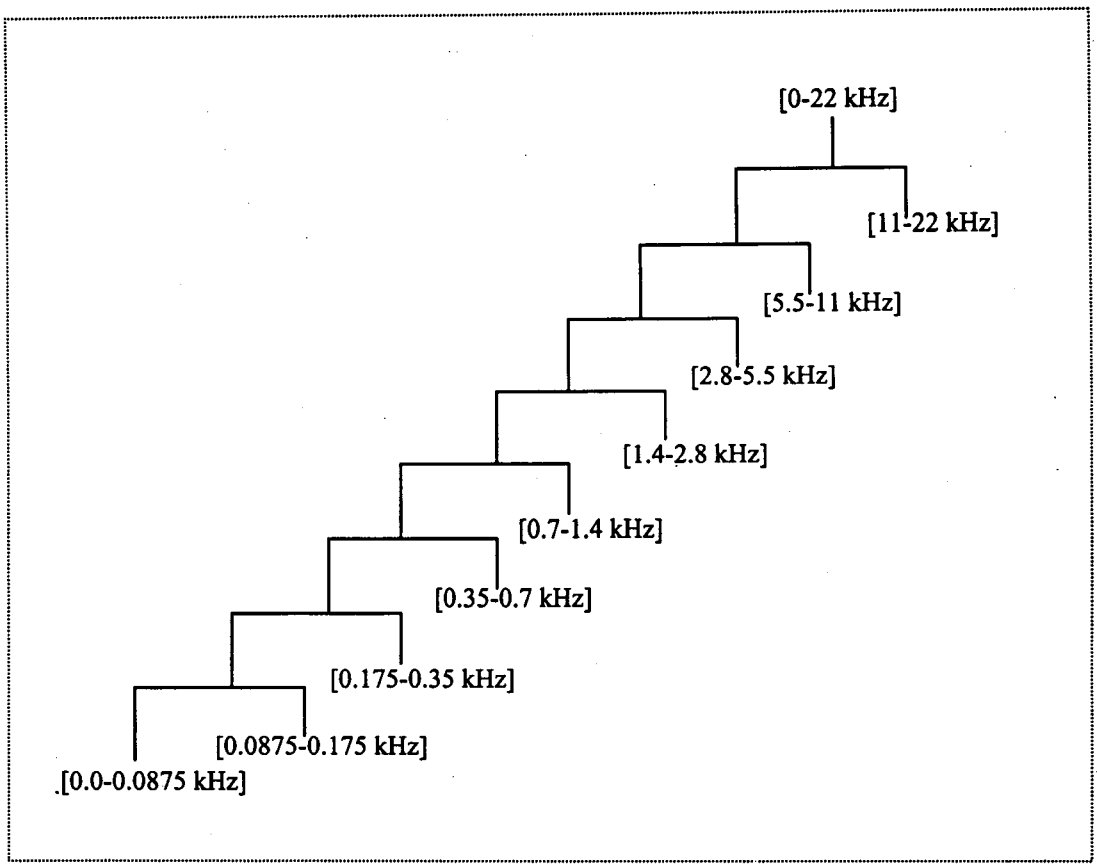


Fig. 4.2: DWT tree structure of the proposed audio coder

Table 4.1 Mapping between subbands used in the proposed coder and critical bands of the human ear.

Subband number	Frequency (Hz)	Critical band number
1	0 - 87.5	1
2	87.5 - 175	2
3	175 - 350	3 & 4
4	350 - 700	5,6 & 7
5	700 - 1400	8,9,10 &11
6	1400 - 2750	12,13,14 &15
7	2750 - 5500	16,17,18 & 19
8	5500 - 11000	21,22,23
9	11000 - 22000	24 & 25

It can be seen from the above table that some of the subbands used in this scheme cover more than one critical band. Therefore, masking thresholds for the various critical bands are calculated using psychoacoustic model and masking threshold for a particular subband is taken as the minimum of the masking thresholds of all critical bands within that subband. The samples in each subband are quantized and encoded such that quantization noise lies below the minimum masking threshold in each subband. These encoded samples are sent to the decoder. The block diagram of the decoder is shown in Fig.4.3. Inverse quantization is done at the decoder to reconstruct the subband samples(ie., DWT coefficients). Inverse DWT of these coefficients are then taken to reconstruct the PCM audio signal.

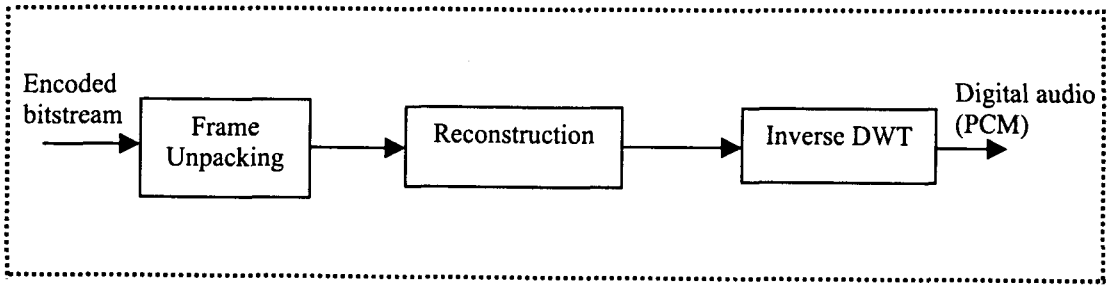


Fig. 4.3: DWT based audio decoder

4.3 Psychoacoustic Model Implementation

Fast Fourier Transform (FFT) technique is employed in this scheme to implement the psychoacoustic model (as in MPEG). For each audio frame, the signal spectrum is computed through 512-point FFT. The tonal (sinusoidal like) and non tonal (noise like) components of the signal are extracted from each audio frame and for each of them the masking effect is computed. The overall masking curve is calculated by adding along the critical bands, all the masking effects and taking the absolute curve (see Fig.2.7) of the ear into account. The transparency condition is satisfied if the reconstruction error spectrum lies under the frequency masking curve.

4.3.1 Results

Results of psychoacoustic model implementation of one frame of the audio signal 'castanets.wav' are shown in Figs.4.4 – 4.11. Local maxima obtained using

spectral analysis is shown in Fig.4.4. Tonal maskers are identified using Eq. 2.15 and are shown in Fig.4.5. Non-tonal components are calculated using Eq. 2.16 and are shown in Fig.4.6 along with the tonal maskers. Tonal and non tonal components which are above the absolute threshold of hearing are retained and other components are discarded. This is illustrated in Figs. 4.7 and 4.8. After this, tonal components too closed to each other are eliminated (see Fig.4.9). Masking thresholds for various bands are calculated using Eq.2.20 and minimum masking threshold for each critical band is obtained. These are shown in Figs.4.10 and 4.11. Quantization of DWT coefficients in various subbands are done such that quantization noise is below this minimum masking threshold.

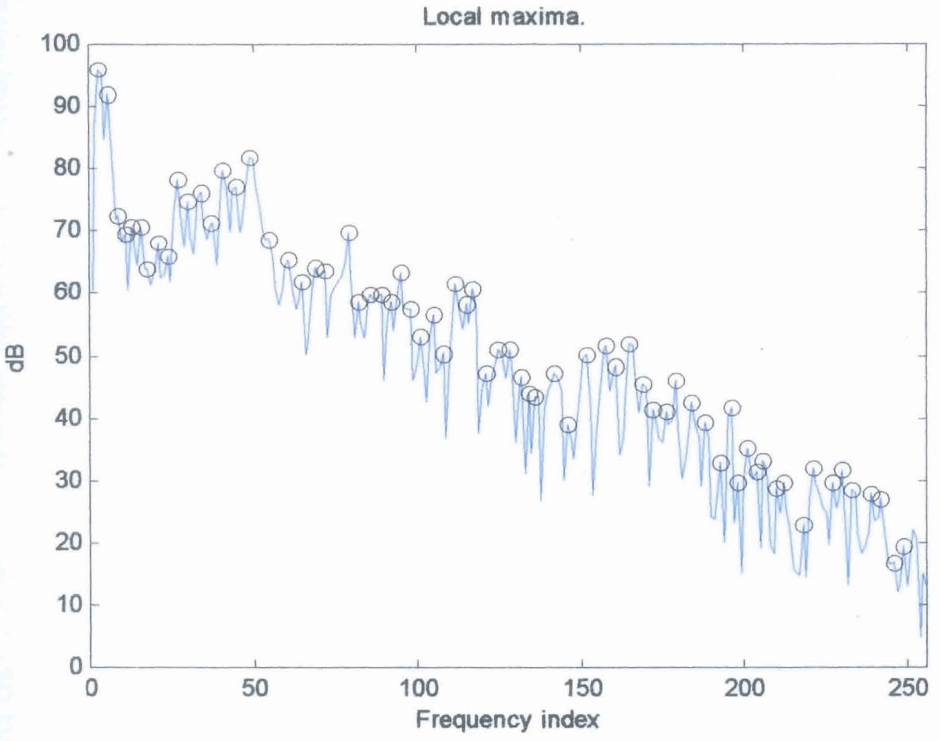


Fig. 4.4: Local maxima

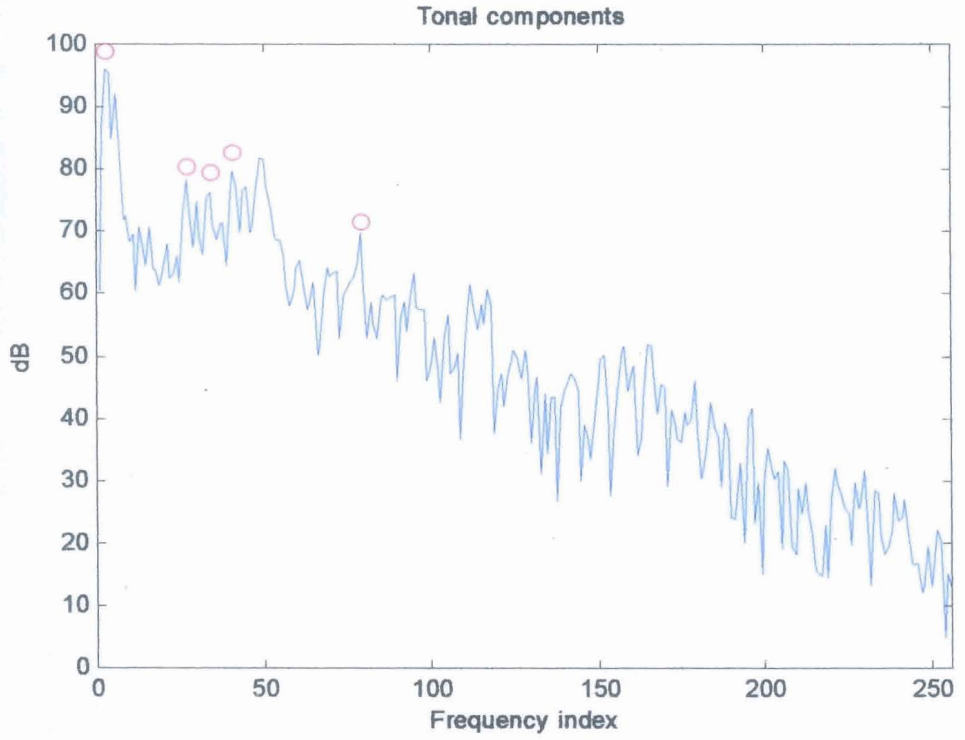


Fig. 4.5: Tonal maskers

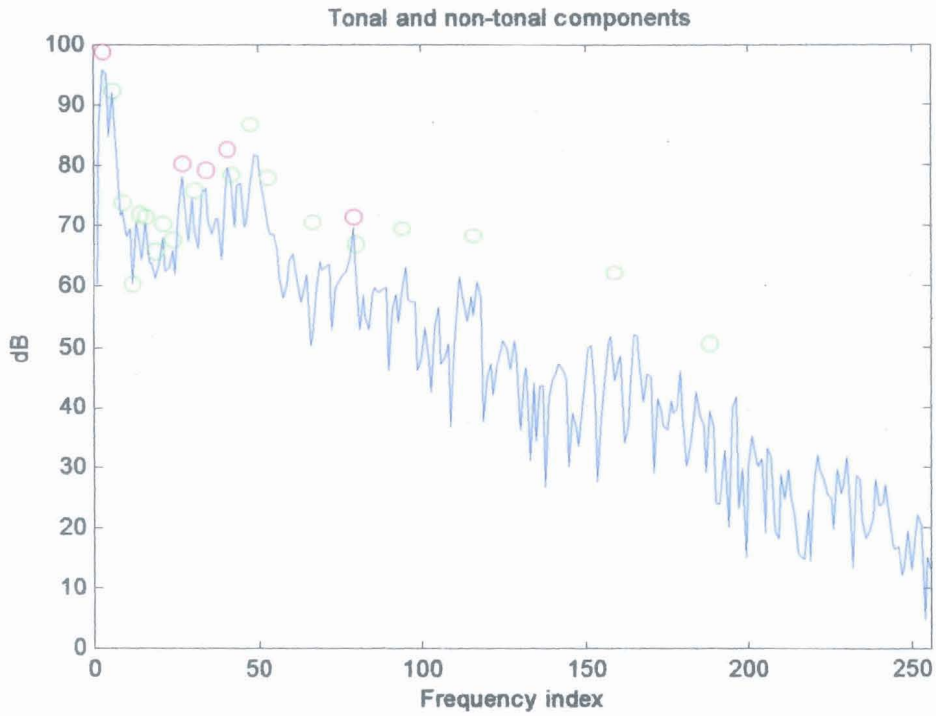


Fig 4.6: Tonal and non-tonal maskers

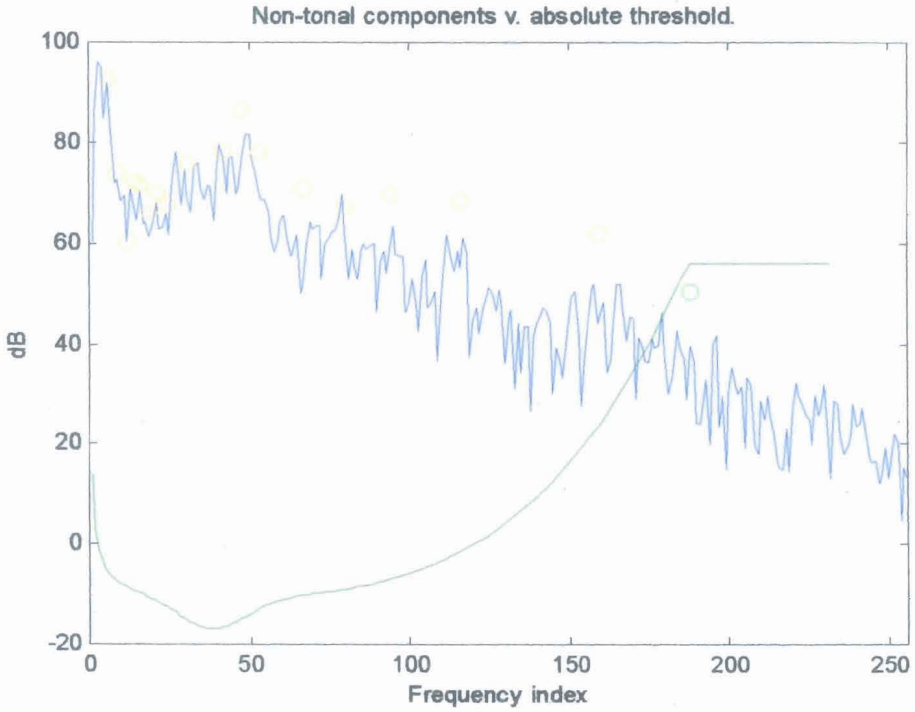


Fig 4.7: Non-tonal components vs. absolute threshold

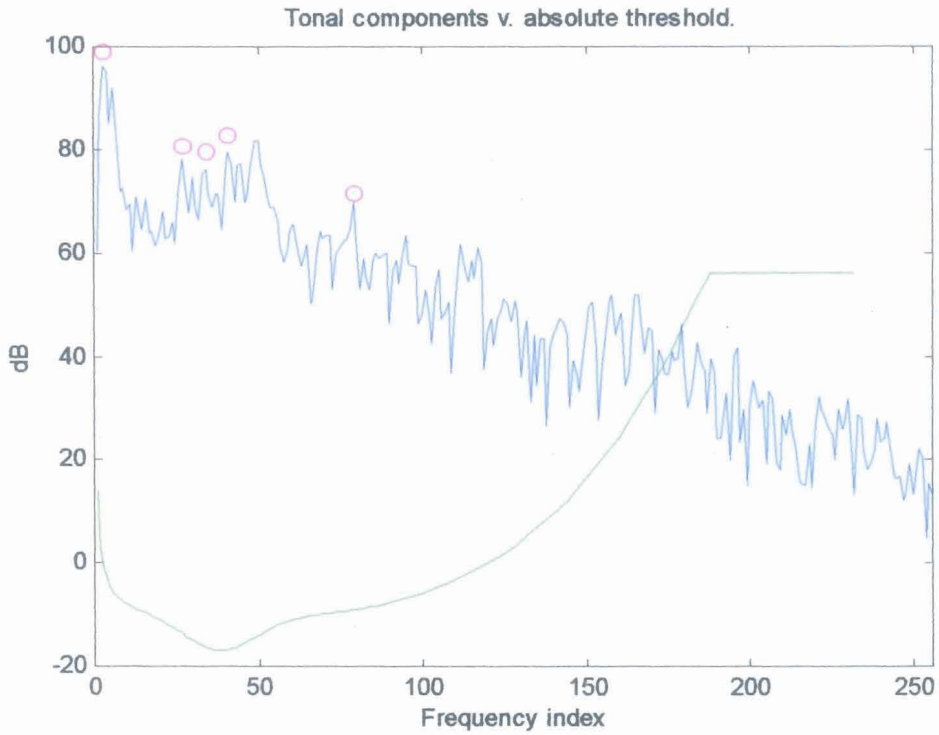


Fig. 4.8: Tonal components vs. absolute threshold

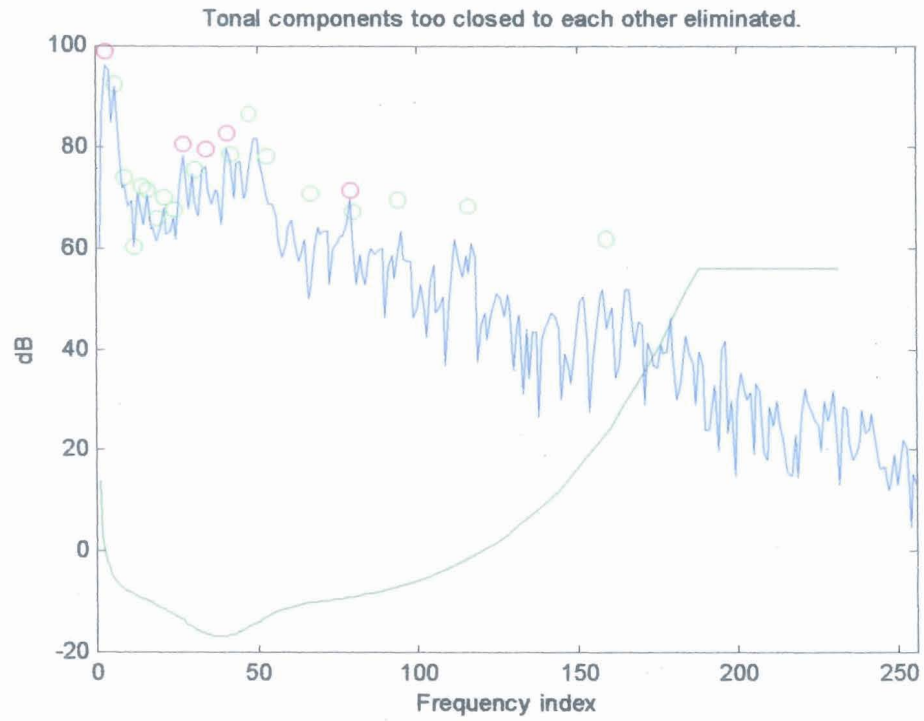


Fig 4.9: Elimination of tonal components too closed to each other

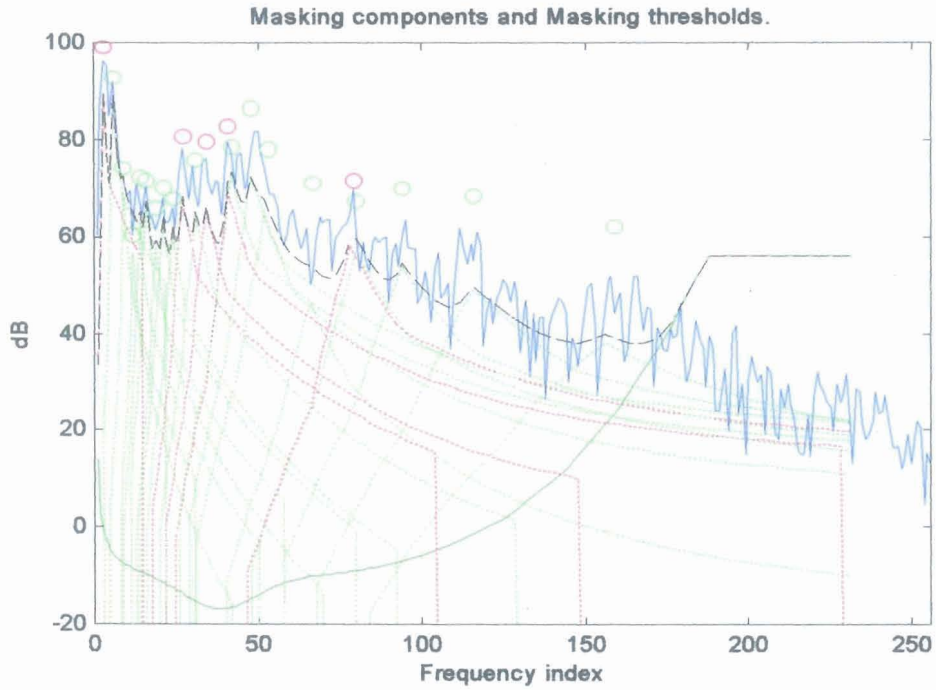


Fig 4.10: Masking thresholds

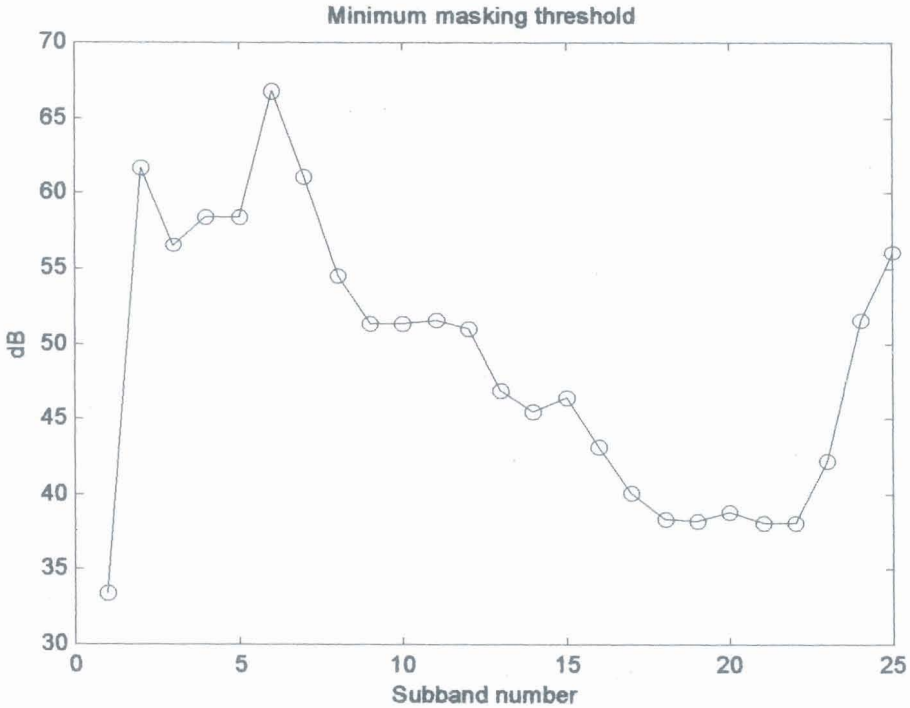


Fig.4.11: Global masking threshold

4.4 Quantization and Coding

Discrete wavelet transform coefficients which are calculated in nine subbands are quantized in such a way that quantization noise in each frequency band is below the masking threshold (Fig. 4.11) for each audio frame. In this proposed scheme, performance of the following quantization schemes are studied and compared.

1. Uniform Scalar Quantization
2. A new Vector Quantization Scheme (Here after referred as "Hit Book Method").
3. Combined Scalar and Vector Quantization.

4.4.1 Uniform Scalar Quantization

DWT coefficients containing nine frequency bands, covering frequency from 0 to 22 kHz, are subjected to scalar quantization. DWT tree structure designed for the time/frequency analysis is shown in Fig.4.3. .Uniform quantization is used here and the number of quantizer levels for each spectral component is obtained from a dynamic bit-allocation rule that is controlled by the psychoacoustics model. .In the psychoacoustic analysis model, there are 25 frequency bands covering 0 to 22 kHz. Hence, these 25 bands are required to combine to conform with the 9 subbands and for each case maximum SMR (Signal -to- Mask Ratio) is found. . The bit-allocation algorithm selects one uniform mid tread quantizer out of a set of available quantizers such that the masking requirements are met. .The bit-allocation process determines the number of code-bits to be allocated to each sub band, based on the equation ,

$$NMR\ (m) = SMR - SNR\ (m)\ (\text{in dB}) \tag{4.1}$$

NMR (m) describes the difference in dB between the SMR and the SNR (Signal to Noise Ratio) to be expected from an m-bit quantization. . The NMR (Noise-to-Mask Ratio) value is also the difference (in dB) between the level of quantization noise and the level where a distortion may just become audible in a given subband. Within a critical band, coding noise will not be audible as long as NMR (m) is negative. In the proposed scheme, masking threshold for each subband is taken as the minimum of the masking thresholds of various critical bands within that subband. More number of

quantization levels are allocated as long as SNR (m) is greater than the SMR. Compression is obtained in terms of the number of bits required for specifying the quantizer levels. For each 12ms frame, according to the SMR of nine subbands, the above procedure is done.

4.4.2 Vector Quantization

Codebook technique is used for performing vector quantization. A new vector quantization technique (named as 'Hit Book method') in which length of the code book can be adaptively changed according to the psychoacoustic model requirement, is developed and implemented here. The method of obtaining codebook is as follows. The first step is to make a raw codebook. The purpose of this codebook is to satisfy the error criteria. In the present scheme vector size is chosen as '8'. Initially a long input training sequence is taken, which is expected to contain almost all DWT vectors. The first DWT vector is taken as the first code book element. Then the second DWT vector is taken, compared with the first and if it is at a distance greater than a specified distance L, it is appended to the codebook, else it is discarded. This procedure is repeated, for each incoming DWT vector by comparing with all the existing codebook elements . The flow chart is given in Fig. 4.12. Here, the error measure is fixed in the transform domain. The values of the input samples lie in a specified range. The values of each of their DWT coefficients will also lie in specified ranges. Therefore a hyper volume can always be specified, inside which all the DWT vectors will lie.

It follows that, by using the above procedure, the length of the raw code book will eventually saturate, say to length M. The entries of the raw codebook may be visualized as centers of hyper spheres with radius (L) as the square root of the distortion measure. The raw code book has its elements interspaced at a minimum distance of L in all probable areas in the vector space. But some of the entries in the codebook are rarely accessed. These entries can be called as zero probability cells. This leads to wastage of memory space and increased processing time.

In order to find out the probability of occurrence of the each entry in the code book, a concept of 'Hit' is introduced here. A 'Hit' is said to have occurred if a match is found between the input block and entry of the code book. For example if the 'Hit'

is found to be 3, it means that, corresponding entry of the codebook was accessed three times. In order to obtain the final code book a number of input sequences are given, 'Hits' are calculated for each element of the code book and the code book elements are arranged in the descending order of hits. In this way the final codebook is formed. The flow chart for the optimization is given in Fig. 4.13. Quantization is done as follows. The DWT of the input audio frames are calculated. The coefficients are processed in blocks of eight values. These blocks are compared with the entries of the code book according to an error measure and the address corresponding to the matched entry is sent to the receiver. Since the codebook elements are at distance 'L' from each entries, the error measured is fixed at L/2. That means, if the incoming block is less than a distance L/2 with any of the codebook entries, the address corresponding to that entry is sent to the receiver.

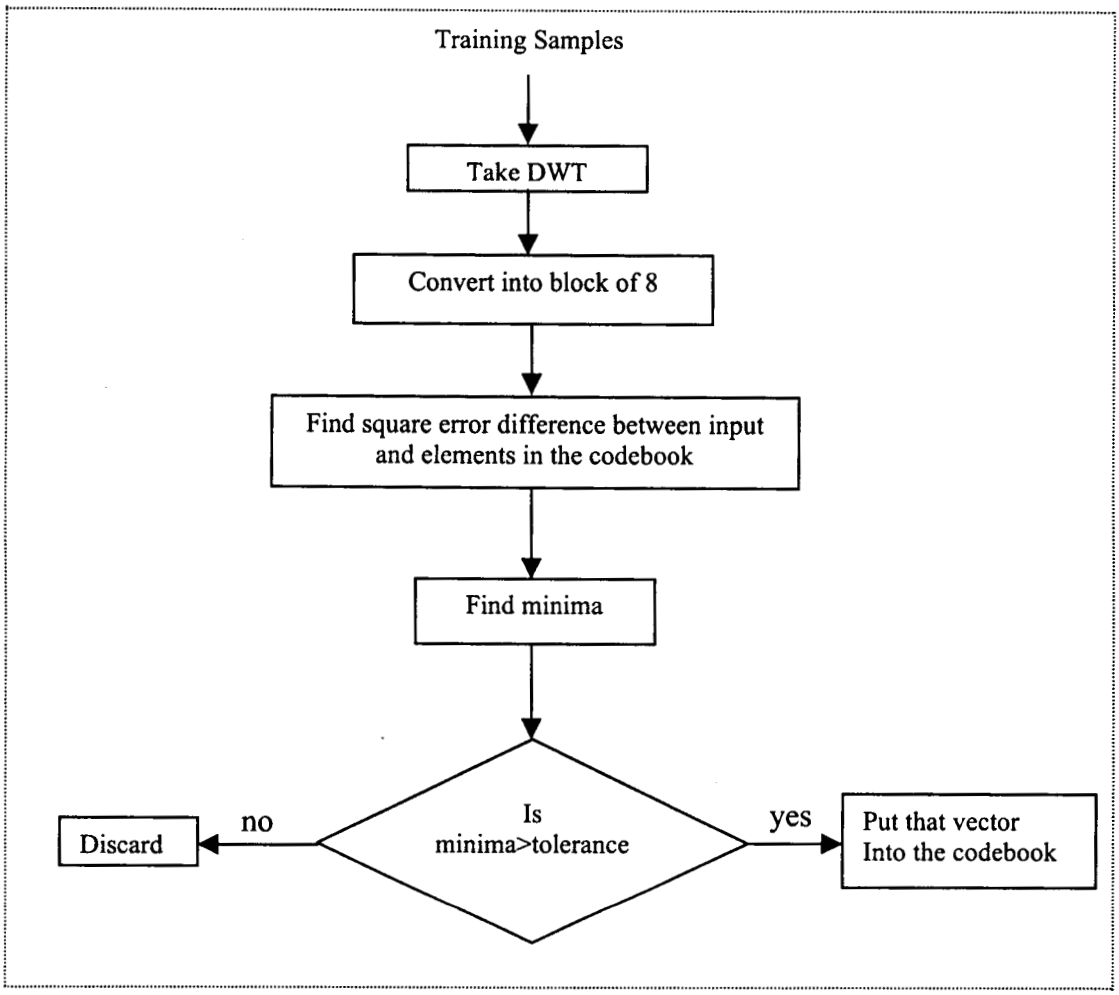


Fig.4.12: Flow chart for making the raw codebook

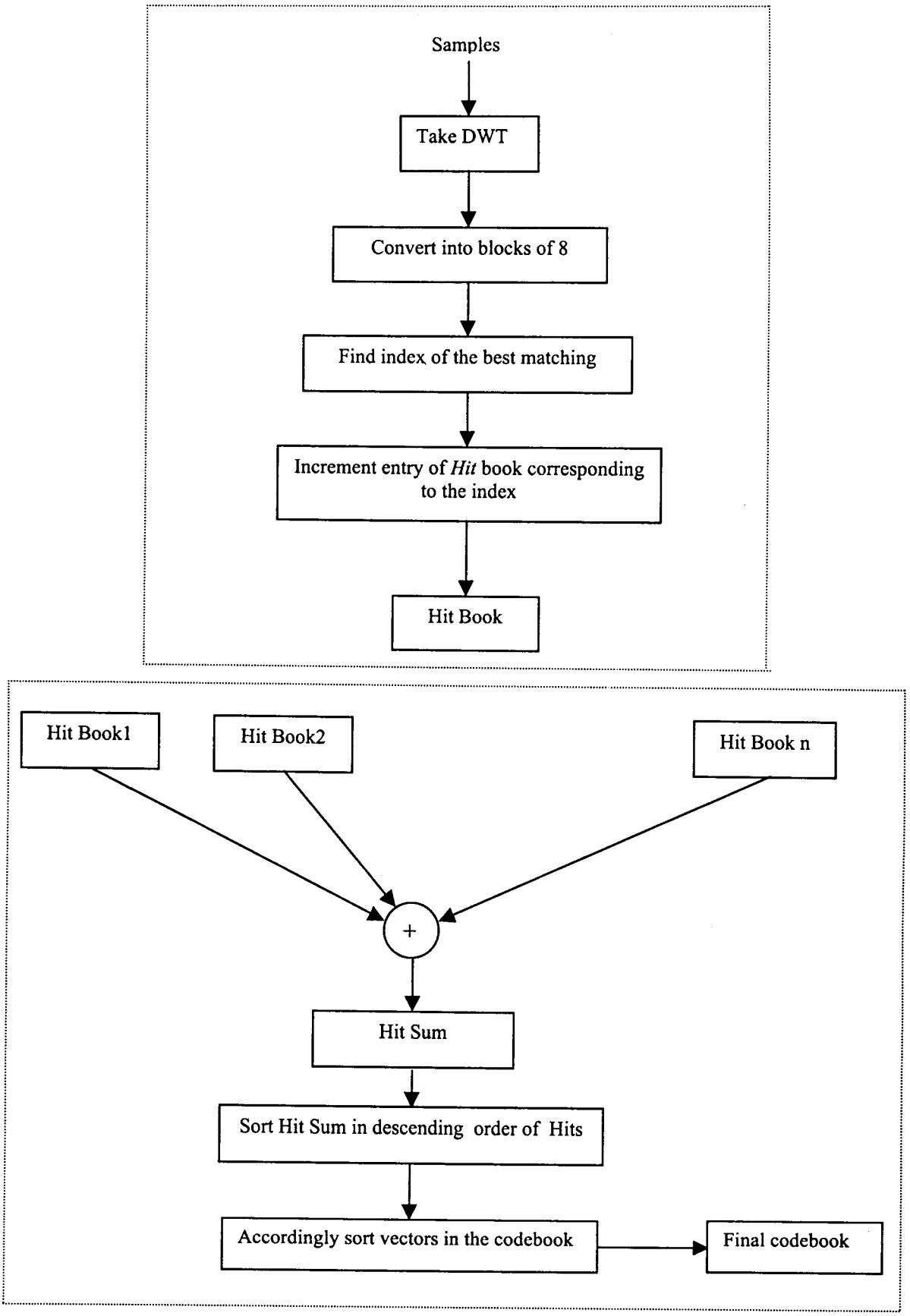


Fig.4.13: Flow Chart for optimising the codebook

At the receiver a similar codebook is placed and once the address is obtained, the corresponding entry from the codebook is obtained. The blocks are appended and then IDWT is calculated to obtain the reconstructed signal. The compression is obtained in terms of the indices of the code vectors. The flow chart of the encoder and decoder algorithms are given Fig.4.14. The advantages of this new algorithm over the most popular K-means algorithm for vector quantization are:

i) In the K-means algorithm, the length of the codebook is fixed. This means that the compression is also fixed. There is no guarantee on the quality of the reproduced signal. In this new algorithm the length of the codebook is not fixed and hence control over the compression ratio is possible.

ii) The entries of the codebook in the K-means case depend very much on the initial entries. There is always a chance of occurrence of zero probability cells. But in this algorithm the number of hits of all the entries of the code book is checked and arranged them in the descending order of hits.

iii) In K-means algorithm the codebook should be trained a number of times so as to bring down mean square error below a certain predetermined value. This is most time consuming step. But in this algorithm training of the code book is not required and so it takes lesser time.

In order to apply the psychoacoustics results, following method is used here. For each input audio frame of 12ms duration, a global masking threshold minimum is obtained for each subband from the psychoacoustics analysis. This will represent the allowable quantization noise that can be introduced in each band, without reduction in quality. Code book size is adaptively changed such that this distortion (i.e., error power spectrum) is below the minimum masking threshold. The encoder and decoder algorithms are shown in Fig.4.14.

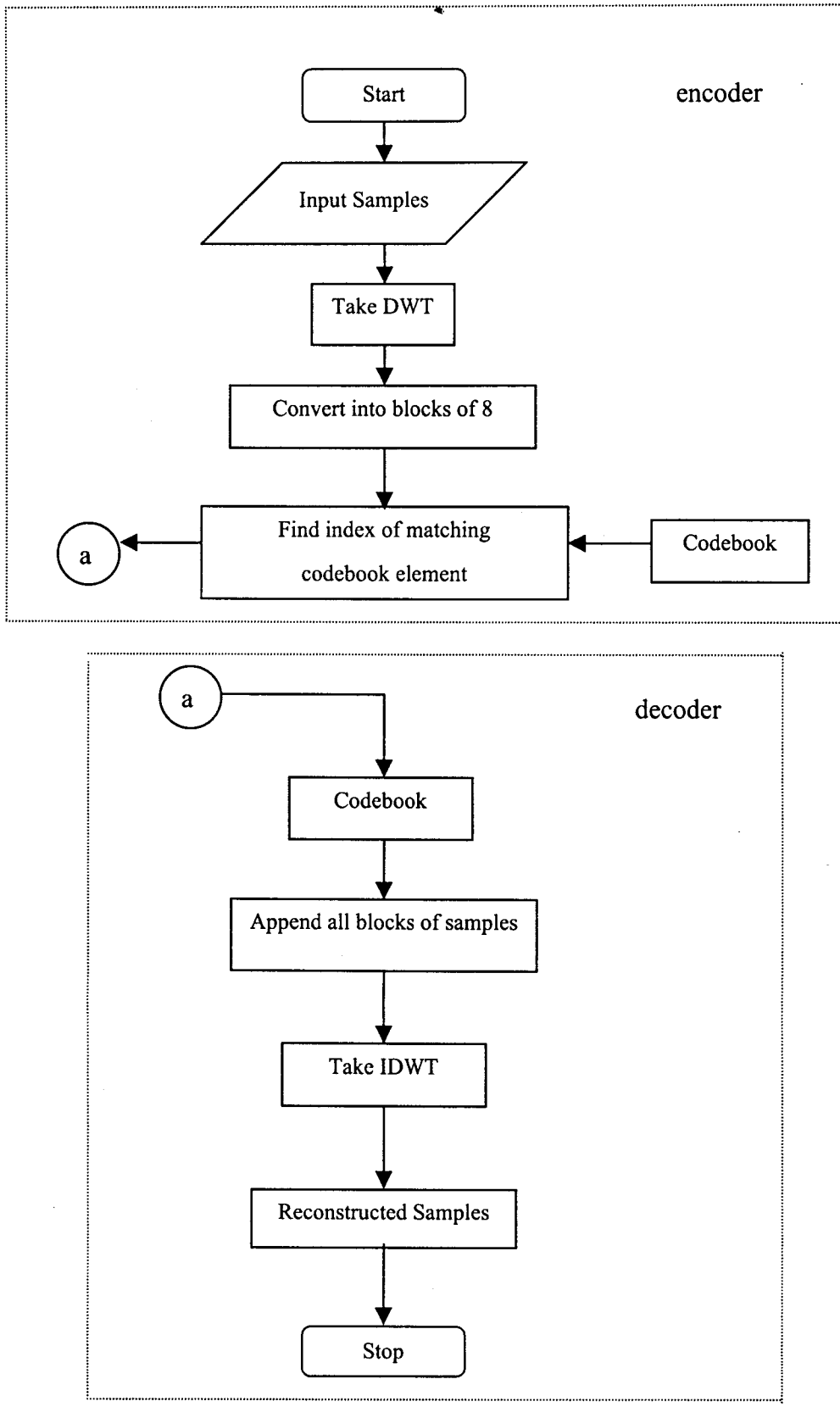


Fig. 4.14: Encoder and decoder algorithms

4.4.3 Combined Scalar and Vector Quantization

Since human ear is more sensitive to frequencies in the range 1 to 5 kHz, following quantization method is also used to study its performance in the perceptual quality and compression ratio. In this scheme, DWT coefficients in the frequency bands up to 5.5 kHz *(i.e.,* seven subbands), are quantized using uniform scalar quantization and for all other frequency bands new vector quantization ("Hit Book method") is used. Indices corresponding to the number of the quantization levels and code book vectors are finally coded using noiseless Huffman coding.

4.5 Performance of the proposed scheme using various wavelets

The studies on research activities in the field of wavelet applications revealed that most of the authors use Daubechies wavelets for signal representation, even though a number of other wavelets are available in the literature. Hence an attempt is made here to study the performance of a number of wavelets with different properties, in the perceptual coding of audio signals. Wavelets employed in the present scheme and their properties have already been described in Section 3.10.

Performance of various wavelets in terms of the order of the wavelet, compression ratio and encoding delay are shown Figs. 4.15 – 4.18

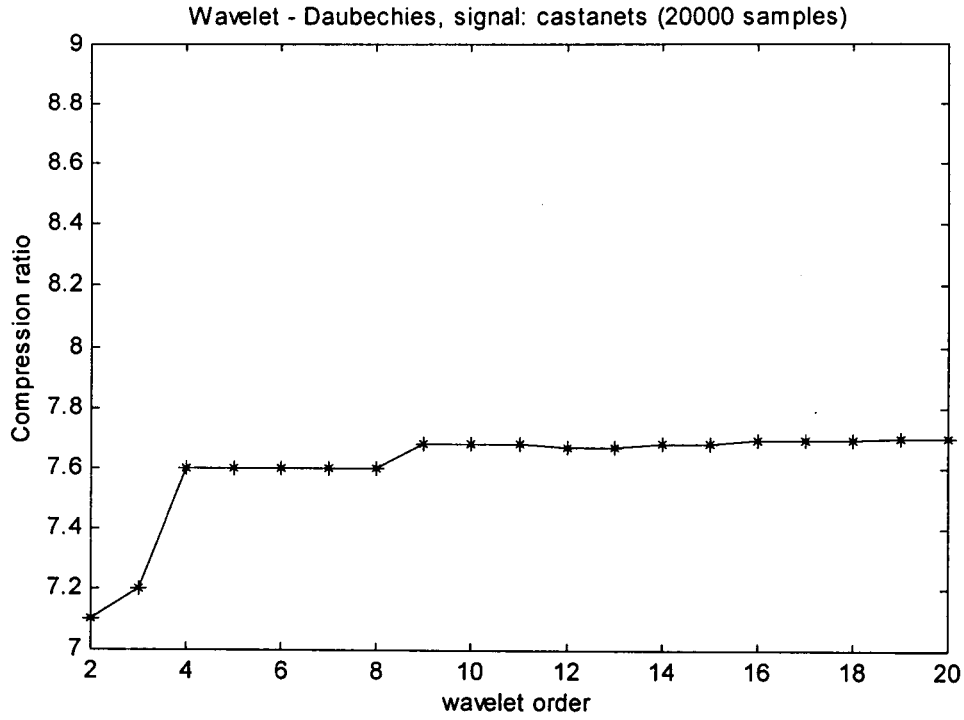


Fig. 4.15 : Variation of compression ratio with order of the wavelet (Daubechies family)

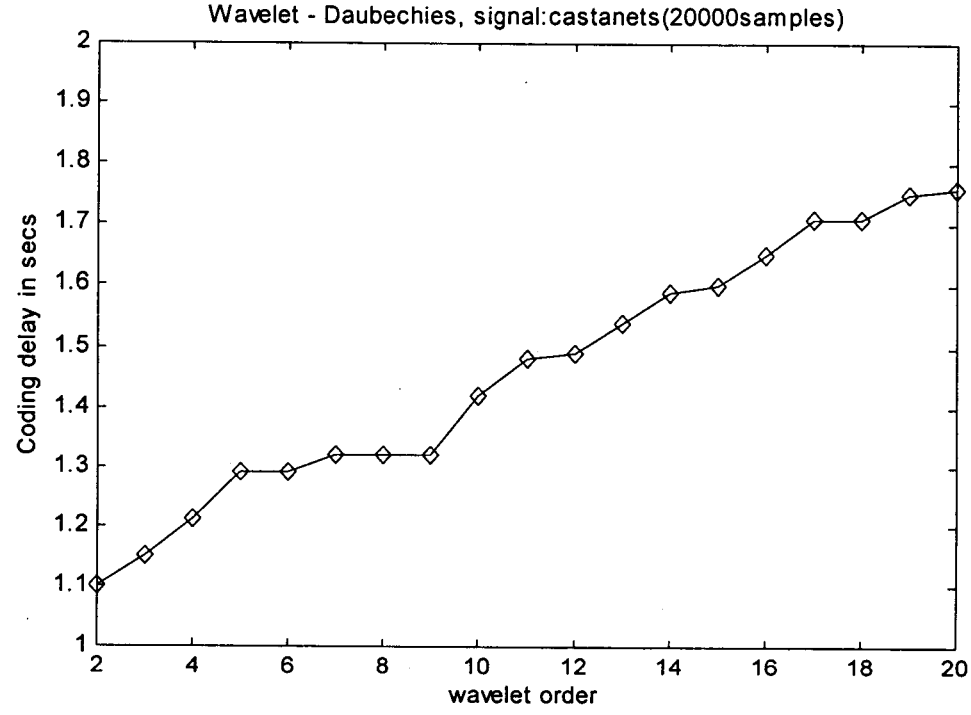


Fig. 4.16: Variation of coding delay with order of the wavelet (Daubechies family)

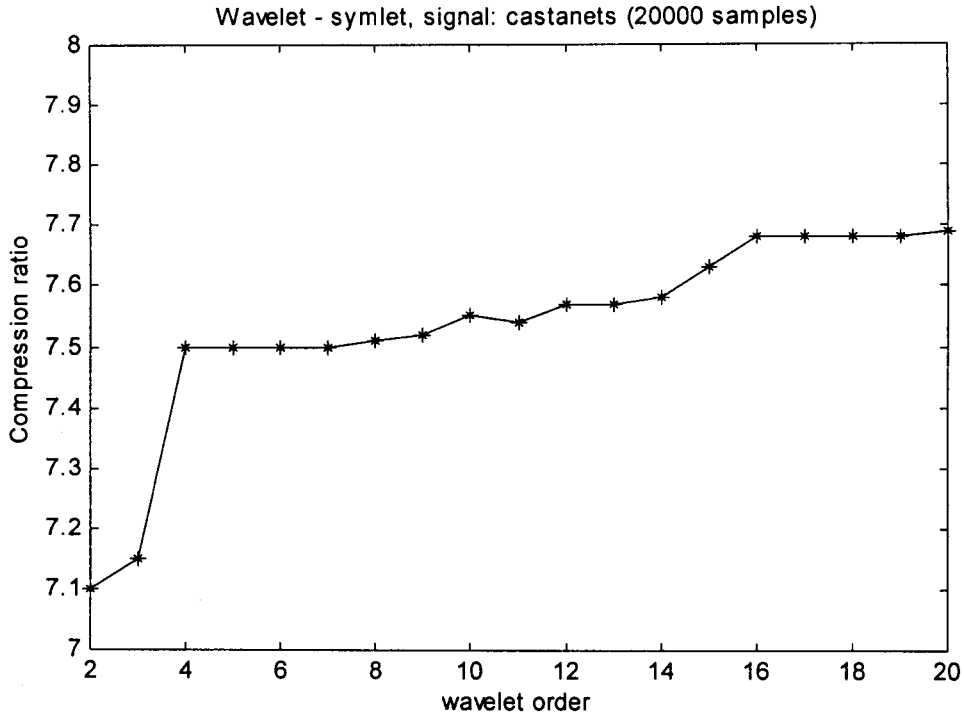


Fig.4.17: Variation of compression ratio with order of the wavelet (Symlet family)

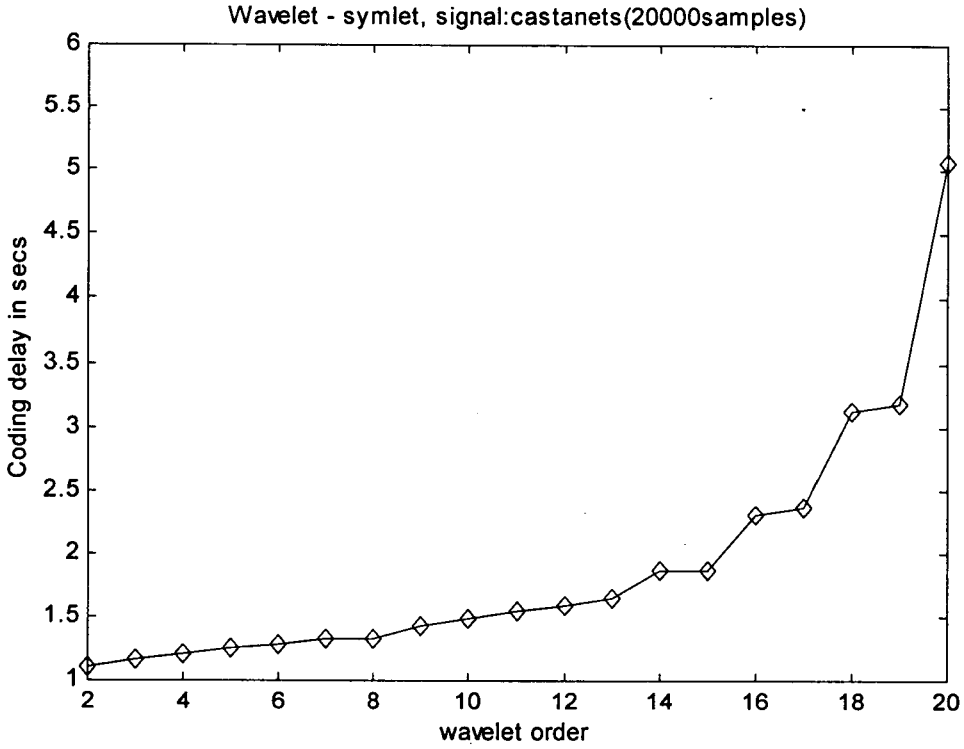


Fig. 4.18: Variation of coding delay with order of the wavelet (Symlet family)

Figs 4.15 – 4.18 reveal that the compression ratio increases when using longer and therefore more regular wavelets. This result is not surprising since longer sequences correspond to wavelet filter banks with sharper bandwidths, i.e., to a better separation of frequency information. These results have been verified for other wavelets in the library also. Hence, the following conclusions are arrived at:

- As the order of the wavelet increases, both compression ratio and computational effort or encoding delay also increase.

- But when the order of the wavelet is increased beyond 4 to 6, there is no significant improvement in the compression ratio as far as audio compression is concerned. This is in agreement with results presented in [44].

- Since todays bussword is multimedia, coding delay is a major factor which is to be kept as minimum as possible. So to achieve compression - speed trade off, the performance of short wavelets (of order 4) of various families are compared here to identify the best wavelet basis for perceptual audio coding.

Performance of fourth order wavelets for various audio signals are shown in Tables 4.2 – 4.3 and Figs. 4.19 – 4.20. These results are discussed in the next section.

Table 4.2 Compression ratios for various frames of the signal kadalinakkare.wav (malayalam music)

Wavelet name	Frame numbers									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
db4	12.8	16.5	13.8	13.3	13.29	7.16	7.47	6.48	6.36	6.6
coif4	14.4	11.5	11.13	11.5	13.26	7.12	8	6.87	5.82	6.6
sym4	11.9	15.05	11.5	12.34	12.33	6.87	9.06	7.47	6.24	6.74
haar	8.82	11.5	13.84	10.8	9.3	3.23	3.8	3.23	3.06	2.76
bior4.4	12.33	15.75	12.34	12.34	12.8	6.48	7.8	7.64	6.48	6.36

Table 4.3 Compression Ratios for various frames of the signal castanets. wav (instrumental music)

Wavelet name	Frame numbers									
	11	**1 2**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
db4	7.64	3.3	4.6	5.12	5.8	9.3	13.83	11.13	6.02	5.27
coif4	8	3.98	4.18	5.36	5.72	9.84	13.8	11.9	6.48	5.45
sym4	8.1	3.35	4.23	5.12	5.6	8.82	12.8	10.45	6.6	5.12
haar	5.28	2.8	3.17	4.28	4	6.6	9.57	8.4	4.7	4.4
bior4.4	7.64	3.2	4.63	5.2	6.02	8.83	13.81	11.9	7.2	4.83

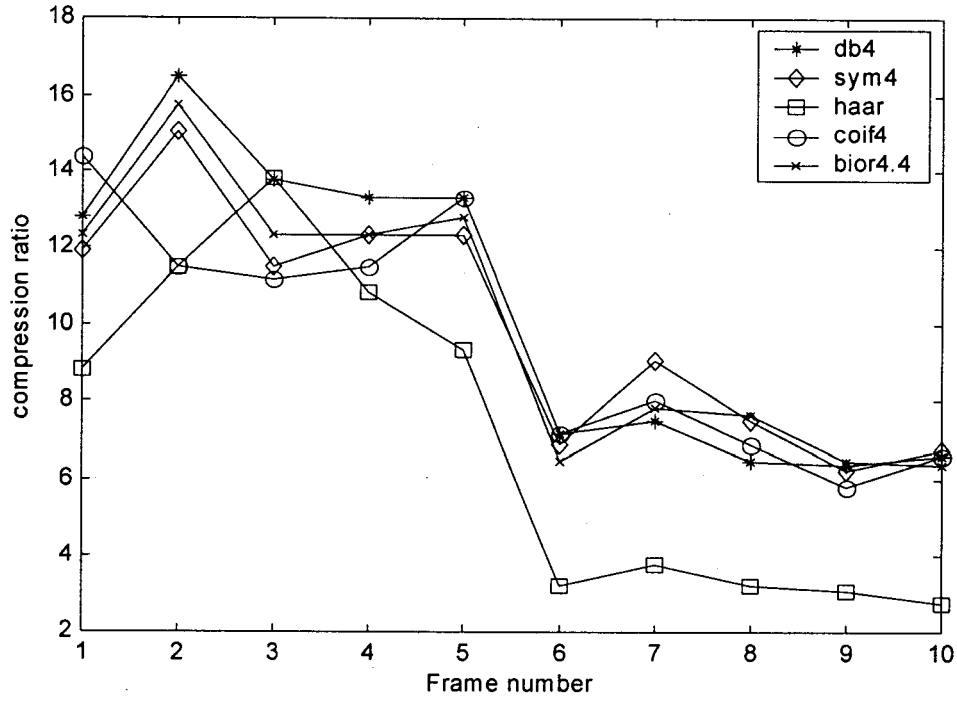


Fig. 4.19: Performance of various wavelets on different frames of audio signal ('kadalinakkare.wav')

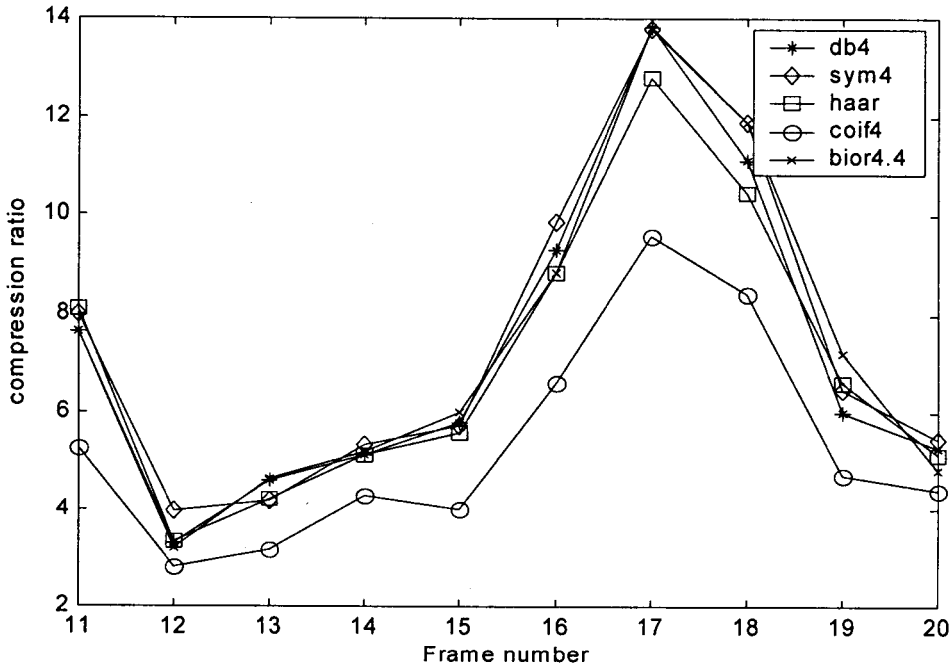


Fig 4.20: Performance of various wavelets on different frames of audio signal ('castanets.wav')

4.5.1 Performance of fourth order wavelets

The results presented in Tables.4.2 – 4.3 and Figs.4.19 - 4.20 show that

- the properties of wavelets like regularity, frequency selectivity, orthogonality and linear phase are less relevant as far as perceptual audio coding is concerned.
- The local properties/ characteristics of the source signal itself, significantly affect the compression ratio. That is, a novel theoretical result, "compression ratio not only depends on the transform or basis function, but also depends on statistical features of the signal itself" is experimentally proved here.
- Hence, best wavelet basis which gives maximum compression ratio for each frame of the audio source is to be identified from the library of wavelets.
- One direct method (brute force method) used here to identify the best wavelet basis is as follows:
    - Each audio frame is analysed using all wavelets in the library.
    - Each frame is compressed using the proposed scheme.
    - Compression ratio is calculated in each case.
    - A cost function (1/compression ratio) is assigned with each wavelet basis.
    - The wavelet whose cost function is minimum is selected as the optimum wavelet basis.
    - Quantization and encoding of the DWT coefficients corresponding to the optimum wavelet basis is done at the encoder.

Eventhough the above method (brute force method) is very accurate, it is very time consuming and hence not preferred for real time applications. Coding delay increases with the number of wavelets in the library. Hence, three different methods are also developed in the present work, to identify the best wavelet for each audio frame to achieve maximum compression with transparent quality.

4.6 Optimum Wavelet Basis Selection

Three different methods are developed and implemented here to select the optimum wavelet basis for each audio frame (of size 512 samples) to be encoded.

4.6.1 Method 1

Using the proposed perceptual audio coder, compression ratios are calculated for a number of audio frames using all wavelets in the library. About 200 audio sources are tested, covering a wide range of sounds viz., music, sounds of animals, birds, machines, musical instruments, speech (male & female), nature sounds etc. For each audio frame, the wavelet, which gives maximum compression ratio is identified. Then the above audio frames are grouped into five sets corresponding to each wavelet as shown in the Table 4.4. Out of a large number of audio frames under each group, only few are listed in this table.

Table 4.4 Grouping of audio segments

db4	haar	coif4	sym4	bior4.4
bye29	bye23	bye22	bye21	bye24
bye31	bye30	bye28	bye25	bye26
bye32	bye33	bye34	bye27	castanets13
castanets4	bye39	castanets12	bye18	castanets15
castanets5	castanets7	castanets14	bye22	castanets19
castanets6	castanets8	castanets16	castanets11	clap5
castanets17	castanets9	clap1	error5	crow1
castanets30	castanets10	clap3	error14	evil10
castanets31	castanets25	crow2	error15	evil13

- mpegtest4 means 4th frame of the audio signal mpegtest.

(contd..)

db4	haar	coif4	sym4	bior4.4
castanets33	castanets28	crow6	evil11	frog1
castanets36	else31	crow7	evil3	frog2
error1	else30	error2	kadal7	kadal8
error3	else32	error4	kadal10	kadal9
evil1	else33	error6	kadal11	lamb2
evil2	else34	error7	kadal12	lamb8
evil4	else35	evil5	kadal14	piano1
evil11	else36	evil6	kadal16	piano2
kadal2	else38	evil7	kadal17	piano5
kadal4	else39	frog3	kadal18	piano6
kadal5	kadal3	frog4	lamb12	piano9
kadal6	piano12	frog5	lamb5	piano14
kadal13	piano13	kadal1	lamb7	piano20
kadal15	piano15	lamb4	piano18	track14
kadal19	piano17	lamb6	piano7	track17
mpegtest1	piano19	piano10	track1	track18
mpegtest2	piano3	piano4	track10	track19
mpegtest3	track2	piano8	track11	track23
mpegtest4	violin14	violin1	track15	track24
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

The first method developed here, for selecting the optimum wavelet basis is based on statistical features of the signal in the time domain itself. Following seven statistical features are extracted from each audio frame.

1. Mean $$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (4.1)$$

2. Variance $$V = \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{i=n} (x_i - \mu)^2 \quad (4.2)$$

3. Skewness $$S = \frac{E(x-\mu)^3}{\sigma^3}$$ where E(x) is the expected value. (4.3)

4. Kurtosis $$k = \frac{E(x-\mu)^4}{\sigma^4} \quad (4.4)$$

5. Entropy $$EY = -\sum_i x_i^2 \log(x_i^2)$$ (Shannon entropy) (4.5)

6. Zero Crossing Rate = Number of zero crossings (number of times the sequence changes the sign) per second and is calculated by

$$Z(n) = \frac{1}{2N}\sum_{m=-\infty}^{m=+\infty} \left|\text{sign}(x(m)) - \text{sign}(x(m-1))\right| \quad (4.6)$$

where $$\text{sign}(x(m)) = \begin{cases} 1, & \text{if } x(m) \geq 0 \\ -1, & \text{otherwise} \end{cases}$$

7. Autocorrelation $$R_{xx}(k) = E[x_{n+k}\ x_n]$$ where k is *the lag.* (4.7)

These seven features are represented by a vector, 'a'. Mean of 'a' vectors of all audio frames in each group is calculated. These mean vectors are represented by FSdb4, FShaar, FScoif4, FSsym4, and FSbior4. These five mean feature vectors corresponding to five groups are stored in a dictionary. When a new audio source is to be encoded, 'a' vector for that audio source is calculated first. Then Euclidean distance D between 'a' vector and feature vectors in the dictionary is calculated. If the smallest value of D corresponds to vector FSdb4, optimum wavelet for that audio frame is selected as db4. An accuracy of 71 % is obtained with this selection method. The best wavelet obtained by brute force method and by method 1, for various audio frames are presented in the Table 4.5 .

Table 4.5 Performance of optimisation method 1

Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method1	Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method1
castanets1	haar	haar	mpegtest1	db4	db4
castanets2	haar	haar	mpegtest2	db4	db4
castanets3	haar	haar	mpegtest3	db4	db4
castanets4	db4	haar	mpegtest4	db4	haar
castanets5	db4	db4	mpegtest5	haar	haar
castanets6	db4	db4	mpegtest6	sym4	sym4
castanets7	haar	db4	mpegtest7	sym4	sym4
castanets8	haar	haar	mpegtest8	coif4	db4
castanets9	haar	haar	mpegtest9	coif4	coif4
castanets10	haar	coif4	mpegtest10	coif4	coif4
castanets11	sym4	coif4	mpegtest11	coif4	sym4
castanets12	coif4	coif4	mpegtest12	sym4	sym4
castanets13	bior4.4	bior4.4	mpegtest13	sym4	sym4
castanets14	coif4	coif4	mpegtest14	sym4	sym4
castanets15	bior4.4	bior4.4	mpegtest15	bior4.4	bior4.4
castanets16	coif4	bior4.4	mpegtest16	bior4.4	bior4.4
castanets17	db4	db4	mpegtest17	db4	bior4.4
castanets18	bior4.4	bior4.4	mpegtest18	sym4	db4
castanets19	bior4.4	bior4.4	mpegtest19	db4	db4
castanets20	coif4	coif4	mpegtest20	db4	db4
castanets21	db4	coif4	mpegtest21	bior4.4	db4
castanets22	db4	coif4	mpegtest22	bior4.4	bior4.4
castanets23	db4	haar	mpegtest23	bior4.4	bior4.4
castanets24	haar	haar	mpegtest24	db4	db4
castanets25	haar	coif4	mpegtest25	db4	db4
kadal1	coif4	coif4	piano1	bior4.4	db4
kadal2	db4	coif4	piano2	bior4.4	bior4.4

Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method1	Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method1
kadal3	haar	haar	piano3	haar	coif4
kadal4	db4	haar	piano4	coif4	coif4
kadal5	db4	db4	piano5	bior4.4	bior4.4
kadal6	db4	db4	piano6	bior4.4	bior4.4
kadal7	sym4	db4	piano7	sym4	sym4
kadal8	bior4.4	bior4.4	piano8	coif4	coif4
kadal9	bior4.4	bior4.4	piano9	bior4.4	bior4.4
kadal10	sym4	sym4	piano10	coif4	bior4.4
kadal11	sym4	sym4	piano11	db4	db4
kadal12	sym4	db4	piano12	haar	haar
kadal13	db4	db4	piano13	haar	haar
kadal14	sym4	sym4	piano14	bior4.4	bior4.4
kadal15	db4	sym4	piano15	haar	haar
kadal16	sym4	sym4	piano16	db4	db4
kadal17	sym4	sym4	piano17	haar	db4
kadal18	sym4	sym4	piano18	sym4	haar
kadal19	db4	db4	piano19	haar	haar
kadal20	db4	db4	piano20	bior4.4	haar
kadal21	bior4.4	db4	piano21	bior4.4	bior4.4
kadal22	bior4.4	bior4.4	piano22	bior4.4	bior4.4
kadal23	db4	db4	piano23	haar	haar
kadal24	sym4	db4	piano24	haar	db4
kadal25	sym4	db4	piano25	db4	db4

4.6.2 Method 2

The second method proposed here for optimum wavelet basis selection is based on statistical features of the audio signal in the wavelet domain. The various steps of the proposed method are given below.

1. Take an audio segment from one set (say, group of audio segments for which 'db4' gives maximum compression).
2. Find 4 level DWT of this segment using haar wavelet (other wavelets can also be used). Five sets of wavelet coefficients namely

   ca4 : approximation coefficients,

   cd4 : detail coefficients at level 4,

   cd3: : detail coefficients at level 3,

   cd2: : detail coefficients at level 2,

   cd1 : detail coefficients at level 1,

   which represent different frequency ranges of the audio signal are obtained.

3. Compute the following three statistical features for each section.
   a) Zero Crossing Rate {czr1, czr2, czr3, czr4, czr5}
   b) Mean value {m1, m2, m3, m4, m5}
   c) Standard deviation {std1, std2, std3, std4, std5}

4. Form a feature vector 'x' which consists of 15 components :
   {czr1, czr2, czr3, czr4, czr5, m1, m2, m3, m4, m5, std1, std2, std3, std4, std5}

5. Repeat above steps for all audio sources in this group.

6. Calculate the mean feature vector of all 'x' vectors. Let it be FSdb4.

7. Repeat the above steps to calculate similar feature vectors FScoif4, FSsym4, FShaar, and FSbior4.

8. Store the above feature vectors in a dictionary.

When a new audio frame is to be encoded, 4 level DWT of this frame using haar wavelet is obtained and feature vector 'x' is calculated. Then Euclidean distance D between 'x' vector and feature vectors in the dictionary is calculated. If the smallest value of D corresponds to FScoif4, optimum wavelet for that audio frame is selected as coif4. An accuracy of 91 % is obtained with this method. Results are shown in the Table 4.6 along with the results of brute force method.

Table 4.6 Performance of optimisation method 2

Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method2	Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method2
castanets1	haar	haar	mpegtest1	db4	db4
castanets2	haar	haar	mpegtest2	db4	db4
castanets3	haar	haar	mpegtest3	db4	db4
castanets4	db4	db4	mpegtest4	db4	haar
castanets5	db4	db4	mpegtest5	haar	haar
castanets6	db4	db4	mpegtest6	sym4	sym4
castanets7	haar	db4	mpegtest7	sym4	sym4
castanets8	haar	haar	mpegtest8	coif4	sym4
castanets9	haar	haar	mpegtest9	coif4	coif4
castanets10	haar	haar	mpegtest10	coif4	coif4
castanets11	sym4	sym4	mpegtest11	coif4	coif4
castanets12	coif4	coif4	mpegtest12	sym4	sym4
castanets13	bior4.4	bior4.4	mpegtest13	sym4	sym4
castanets14	coif4	coif4	mpegtest14	sym4	sym4
castanets15	bior4.4	bior4.4	mpegtest15	bior4.4	bior4.4
castanets16	coif4	coif4	mpegtest16	bior4.4	bior4.4
castanets17	db4	db4	mpegtest17	db4	db4
castanets18	bior4.4	bior4.4	mpegtest18	sym4	db4
castanets19	bior4.4	bior4.4	mpegtest19	db4	db4
castanets20	coif4	coif4	mpegtest20	db4	db4
castanets21	db4	db4	mpegtest21	bior4.4	bior4.4
castanets22	db4	db4	mpegtest22	bior4.4	bior4.4
castanets23	db4	db4	mpegtest23	bior4.4	bior4.4
castanets24	haar	haar	mpegtest24	db4	db4
castanets25	haar	haar	mpegtest25	db4	db4
kadal1	coif4	haar	piano1	bior4.4	bior4.4
kadal2	db4	db4	piano2	bior4.4	bior4.4

Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method2	Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method2
kadal3	haar	haar	piano3	haar	haar
kadal4	db4	db4	piano4	coif4	haar
kadal5	db4	db4	piano5	bior4.4	bior4.4
kadal6	db4	db4	piano6	bior4.4	bior4.4
kadal7	sym4	sym4	piano7	sym4	sym4
kadal8	bior4.4	bior4.4	piano8	coif4	coif4
kadal9	bior4.4	bior4.4	piano9	bior4.4	bior4.4
kadal10	sym4	sym4	piano10	coif4	coif4
kadal11	sym4	sym4	piano11	db4	db4
kadal12	sym4	sym4	piano12	haar	haar
kadal13	db4	db4	piano13	haar	haar
kadal14	sym4	db4	piano14	bior4.4	haar
kadal15	db4	db4	piano15	haar	haar
kadal16	sym4	sym4	piano16	db4	db4
kadal17	sym4	sym4	piano17	haar	haar
kadal18	sym4	sym4	piano18	sym4	sym4
kadal19	db4	db4	piano19	haar	haar
kadal20	db4	db4	piano20	bior4.4	bior4.4
kadal21	bior4.4	bior4.4	piano21	bior4.4	bior4.4
kadal22	bior4.4	bior4.4	piano22	bior4.4	bior4.4
kadal23	db4	db4	piano23	haar	haar
kadal24	sym4	db4	piano24	haar	haar
kadal25	sym4	sym4	piano25	db4	db4

The above method is tested using other wavelets also and the same results are obtained with all the wavelets in the library. There was no significant improvement in accuracy when additional statistical features like kurtosis, skewness, entropy *etc.* of DWT coefficients are included. Hence, to reduce the computational complexity,

simple haar wavelet and three statistical features of DWT coefficients in all the frequency bands are used here to select the optimum wavelet basis. Even though this method is more accurate than the previous method, coding delay is more here, because for each audio frame, haar DWT is to be computed to determine the optimum wavelet basis. Hence, if coding delay is the major constraint, first method is preferred over the second method. In the next section, a faster method with accuracy slightly less than that of second method is described.

4.6.3 Method 3

In this method, the problem of selecting optimum wavelet basis is changed into a signal classification problem. Hence, an artificial neural network (ANN) with two layers, implemented using Back Propagation approach is used for signal classification. Here, the network is trained using all audio sources listed in the Table.4.3 Therefore, when a new audio signal comes, neural network outputs the optimum wavelet basis for that segment. Using this optimum wavelet basis, discrete wavelet transform of the signal is computed and the coefficients in various subbands are quantized to meet the masking criteria.

Back Propagation Network (BPN)

This network learns a predefined set of input-output example pairs by using a two-phase propagation -adapt cycle. After an input vector has been applied as a stimulus to the first layer of network units, it is propagated through each upper layer until an output is generated. This output vector is then compared to the desired output, and an error signal is computed for each output unit.

The error signals are then transmitted backward from the output layer to each node in the intermediate layer that contributes directly to the output. However each unit in the intermediate layer receives only a portion of the total error signal, based roughly on the relative contribution the unit made to the original output. This process repeats, layer by layer, until each node in the network has received an error signal that describes its relative contribution to the total error. Based on the error signal received, connection weights are then updated by each unit to cause the network to converge

toward a state that allows all the training patterns to be encoded. After training, when presented with an arbitrary input signal, the units in the hidden layers of the network will respond with an active output if the new input contains a pattern that resembles the feature the individual units learned to recognize during training. Hence the BPN will classify previously unseen inputs according to the features they share with the training samples. The various steps are given below.

Steps:

1. Apply the input vector, $x_p = (x_p1, x_p2, \ldots\ldots, x_pN)^T$ to the input units.
2. Calculate the net input values to the hidden layer units. The net input to the j^{th} hidden unit is,

$$net^{h}_{pj} = \sum_{i=1}^{N} w^{h}_{ji} x_{pi} + \theta^{h}_{j}$$

where w^{h}_{ji} is the weight on the connection from the i^{th} input unit, and θ^{h}_{j} is the bias term. The 'h' superscript refers to quantities on the hidden layer.

3 Calculate the outputs from the hidden layer.

$$i_{pj} = f^{h}_{j}(net^{h}_{pj})$$

where f^{h}_{j} is the transfer function of the j^{th} hidden layer. Back propagation networks often use log sigmoid function and occasionally use linear transfer function.

4. Move to the output layer. Calculate the net input values to each unit.

$$net^{o}_{pk} = \sum_{j=1}^{L} w^{o}_{kj} i_{pj} + \theta^{o}_{k}$$

5. Calculate the outputs.

$$O_{pk} = f^{o}_{k}(net^{o}_{pk})$$

6. Calculate the error terms for the output units

$$\delta^{o}_{pk} = (y_{pk} - O_{pk}) f^{o'}_{k}(net^{o}_{pk})$$

where y_{pk} is the desired output value, and O_{pk} is the actual output from the k^{th} unit, $f^{o'}_{k}$ is the derivative of the transfer function.

7. Calculate the error terms for the hidden units:

$$\delta^{h}_{pj} = f^{h}_{j}{}'(net^{h}_{pj}) \sum_{k} \delta^{o}_{pk} w^{o}_{kj}$$

Error terms on the hidden units calculated before the connection weights to the output layer units have been updated.

8. Update weights on the output layer

$$w^{o}_{kj}(t+1) = w^{o}_{kj}(t) + \eta_{j} \delta^{o}_{pk} i_{pj}$$

where η is the learning- rate parameter.

9. Update weights on the hidden layer :

$$w^{h}_{ji}(t+1) = w^{h}_{ji}(t) + \eta \delta^{h}_{pj} x_{i}$$

10. Calculate error term $E_p = \frac{1}{2} \sum_{k=1}^{M} \delta^{2}_{pk}$

When this error is acceptably small for each of the training vector pairs, training is discontinued. An accuracy of 90% is obtained with this method. Results are shown in Table 4.7 along with the results of brute force technique.

Table 4.7 Performance of Optimisation Method 3

Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method3	Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method3
castanets1	haar	haar	mpegtest1	db4	db4
castanets2	haar	haar	mpegtest2	db4	db4
castanets3	haar	haar	mpegtest3	db4	db4
castanets4	db4	haar	mpegtest4	db4	haar
castanets5	db4	db4	mpegtest5	haar	haar
castanets6	db4	db4	mpegtest6	sym4	sym4
castanets7	haar	haar	mpegtest7	sym4	sym4
castanets8	haar	haar	mpegtest8	coif4	coif4
castanets9	haar	haar	mpegtest9	coif4	coif4
castanets10	haar	haar	mpegtest10	coif4	coif4
castanets11	sym4	sym4	mpegtest11	coif4	coif4
castanets12	coif4	coif4	mpegtest12	sym4	sym4
castanets13	bior4.4	bior4.4	mpegtest13	sym4	sym4

Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method3	Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method3
castanets14	coif4	coif4	mpegtest14	sym4	sym4
castanets15	bior4.4	bior4.4	mpegtest15	bior4.4	bior4.4
castanets16	coif4	db4	mpegtest16	bior4.4	bior4.4
castanets17	db4	db4	mpegtest17	db4	db4
castanets18	bior4.4	bior4.4	mpegtest18	sym4	db4
castanets19	bior4.4	bior4.4	mpegtest19	db4	db4
castanets20	coif4	coif4	mpegtest20	db4	db4
castanets21	db4	db4	mpegtest21	bior4.4	bior4.4
castanets22	db4	db4	mpegtest22	bior4.4	bior4.4
castanets23	db4	db4	mpegtest23	bior4.4	bior4.4
castanets24	haar	haar	mpegtest24	db4	db4
castanets25	haar	haar	mpegtest25	db4	db4
kadal1	coif4	coif4	piano1	bior4.4	bior4.4
kadal2	db4	db4	piano2	bior4.4	bior4.4
kadal3	haar	haar	piano3	haar	haar
kadal4	db4	haar	piano4	coif4	coif4
kadal5	db4	db4	piano5	bior4.4	bior4.4
kadal6	db4	db4	piano6	bior4.4	bior4.4
kadal7	sym4	sym4	piano7	sym4	sym4
kadal8	bior4.4	bior4.4	piano8	coif4	coif4
kadal9	bior4.4	bior4.4	piano9	bior4.4	bior4.4
kadal10	sym4	sym4	piano10	coif4	coif4
kadal11	sym4	sym4	piano11	db4	coif4
kadal12	sym4	sym4	piano12	haar	haar
kadal13	db4	db4	piano13	haar	haar
kadal14	sym4	db4	piano14	bior4.4	bior4.4
kadal15	db4	db4	piano15	haar	haar
kadal16	sym4	sym4	piano16	db4	db4

Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method3	Audio signal with frame number	Best wavelet by brute force method	Best wavelet by method3
kadal17	sym4	sym4	piano17	haar	db4
kadal18	sym4	sym4	piano18	sym4	sym4
kadal19	db4	db4	piano19	haar	haar
kadal20	db4	db4	piano20	bior4.4	haar
kadal21	bior4.4	bior4.4	piano21	bior4.4	bior4.4
kadal22	bior4.4	bior4.4	piano22	bior4.4	bior4.4
kadal23	db4	sym4	piano23	haar	haar
kadal24	sym4	sym4	piano24	haar	haar
kadal25	sym4	sym4	piano25	db4	db4

The comparison of the three methods developed in the present work to select the optimum wavelet basis for perceptual audio coding, in terms of coding delay, accuracy and the memory requirement is shown in Table 4.8. One of the optimisation methods can be suitably selected by a user, depending upon the processing power available, memory available and the delay that can be tolerated for a typical application.

Table 4.8 Comparison between the three optimisation methods

Method	Coding delay	Accuracy	Memory requirement
I	Medium	71%	Minimum
II	Maximum	91%	Medium
III	Minimum	90%	Maximum

4.7 Performance of Discrete Cosine Transform (DCT)

ISO/MPEG audio coding standard is based on DCT and is implemented with uniform filter banks as described in Chapter 2. Hence DWT block in the proposed audio coder is replaced by DCT and its performance on perceptual audio coding is also evaluated. The results are shown in Tables 4.9 and 4.10 for various frames of 'male speech' (which is quasi-stationary) and for various frames of 'castanets' (an instrumental music which contains sharp attacks or transients). From these tables it can be seen that,

- If the audio frame is stationary, DCT gives more compact representation than DWT, as expected. The reason is that low order FIR analysis filters typically employed in wavelet decompositions are characterised by poor frequency selectivity and therefore wavelet bases tend not to provide compact representations for stationary signals. Masking thresholds calculated for various subbands of a stationary signal will highly be localized in frequency and hence such signal calls for a high frequency resolution analysis filterbank.

Table 4.9 Performance of DCT on 'Male Speech' Segments

Audio signal segment	Compression ratio with optimum wavelet basis	Compression ratio with sinusoidal basis (DCT)
1	6.3	6.8
2	5.8	6.4
3	4.1	2.6
4	6.8	8.8
5	6.3	7.9
6	5.0	6.2
7	6.8	7.2
8	3.6	2.0
9	5.3	7.0
10	5.7	6.3

Table 4.10 Performance of DCT on 'castanets' Segments

Audio signal segment	Compression ratio with optimum wavelet basis	Compression ratio with sinusoidal basis (DCT)
1	7.2	3.1
2	4.4	3.6
3	8.3	4.2
4	8.5	4.4
5	7.1	4.3
6	7.8	4.7
7	3.0	2.1
8	2.6	1.9
9	8.5	5.2
10	8.7	4.6

In the case of 'castanets', which is highly non-stationary, DWT gives more compact representation than DCT. Hence, a switching algorithm is also developed in the present work, to switch between DCT and DWT depending upon the time varying characteristics of the audio signal, so that each frame is compactly represented using optimum basis (i.e., either using sinusoidal basis or using optimum wavelet basis).

4.7.1 Switching Algorithm

If a segment is stationary, the analysis filter bank should be DCT, but in the event of non-stationarity, it should switch to the wavelet filter bank (DWT). The following algorithm is developed for this switching.

Steps

1. Divide the audio frame into 8 blocks (*ie.*, 64 samples/block).
2. Calculate Mean and Autocorrelation of various blocks.
3. If the difference in mean between various blocks and difference in autocorrelation between various blocks is less than a threshold value, 'thr', that frame is stationary. Value of the 'thr' is found experimentally. 'thr' for mean is obtained as 0.3 and 'thr' for autocorrelation is obtained as 0.7.

Hence, discrete wavelet transform based audio coder in Fig.4.1 is enhanced into the coding scheme as shown in Fig.4.21, by incorporating the above switching algorithm also.

4.7.2 Enhanced Audio Coder

Fig.4.21 shows the block diagram of the enhanced wavelet based audio coder. With this scheme, each audio frame is represented using either sinusoidal or optimum wavelet basis according to the time varying characteristics of the audio source.

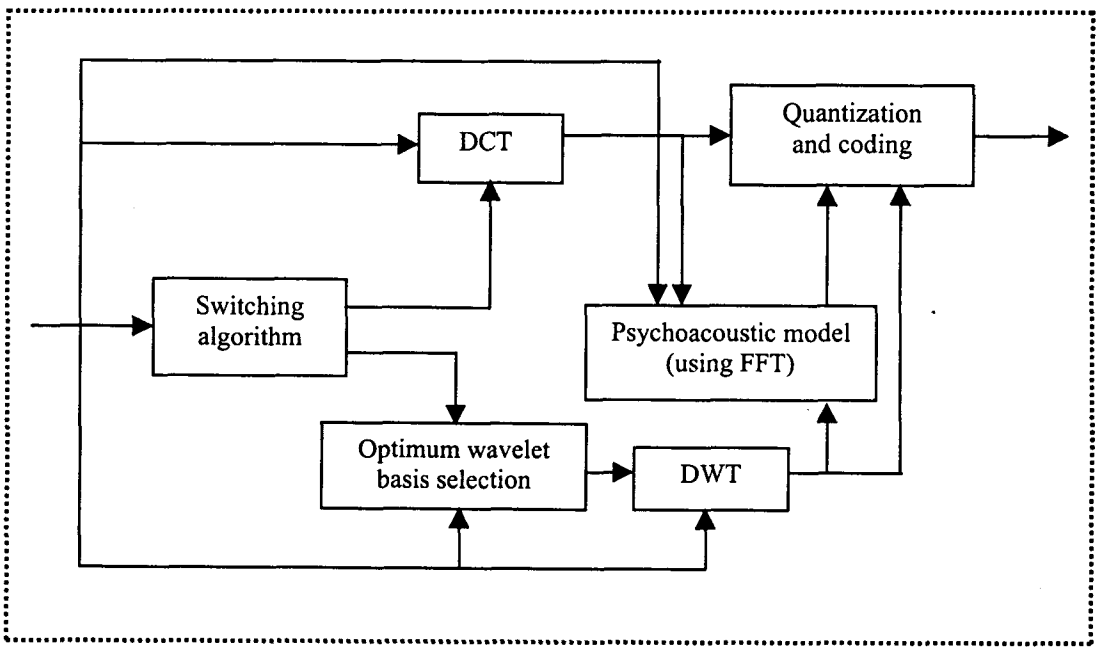


Fig. 4.21: Enhanced DWT based audio coder

4.8 Advantages of the Switched Filterbank

Switched filterbank does not increase the computational complexity. Both the DCT and the wavelet filtebank are particularly suitable for encoding a different class of signals. A high resolution DCT leads to a very compact representation for stationary signals. However, signals that contain transients or sharp attacks cannot be represented compactly in the DCT filterbank. These signals require a higher time resolution at high

frequencies both for compact representation and optimal exploitation of perceptual irrelevancies. Wavelet filterbanks are quite attractive for encoding of such signals. Besides the fact that, wavelet representation of such signals is more compact than the representation derived from a high resolution DCT, wavelet filters have desirable temporal characteristics. In a wavelet filterbank, the high frequency filters typically have a compact impulse response. This prevents excessive time spreading of quantization errors during synthesis (perceptible as the so called "pre-echo"). Pre-echo distortion in MPEG coder based on DCT and pre-echo elimination in wavelet based coder are shown Figs.4.22 and 4.23. The signal shown in the figure corresponds to few samples of "castanets"(an instrumental music).

4.9 Audio Quality Evaluation and Measurements

Results in the field of low bit rate audio compression are highly subjective. When comparing two compression schemes, there is no universally accepted standard for audio quality. A true comparison can only be ever made by expensive and time-consuming listening tests. To counter this problem, audio quality is usually expressed in terms of segmental signal to noise ratio (SSNR) which can be calculated from original and reconstructed data. However the human auditory system does not hear distortion in terms of segmental signal to noise ratios, SSNR is only used to indicate the statistical error introduced by compression. Therefore, power spectra comparative results are also taken into account in this thesis and of course subjective listening tests have been done using the Mean Opinion Score (MOS) scale. The aim of this section is to define all of these objective and subjective measurements which have been used to present the results throughout this thesis.

4.9.1 Objective Evaluation

4.9.1.1 Signal-to-Noise Ratio (SNR) measurements: An objective measure that is applicable to waveform coding systems is the signal to noise ratio (SNR) measurements. Signal-to-Noise Ratio measures are only appropriate if both the original and distorted signals are time aligned.

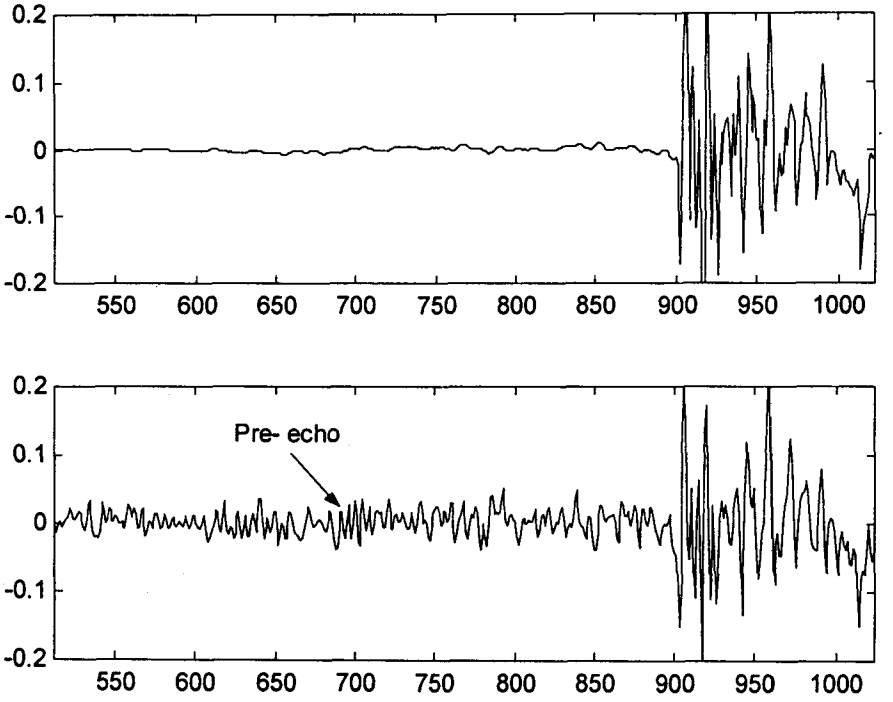


Fig. 4.22: Pre -echo distortion in MPEG standard

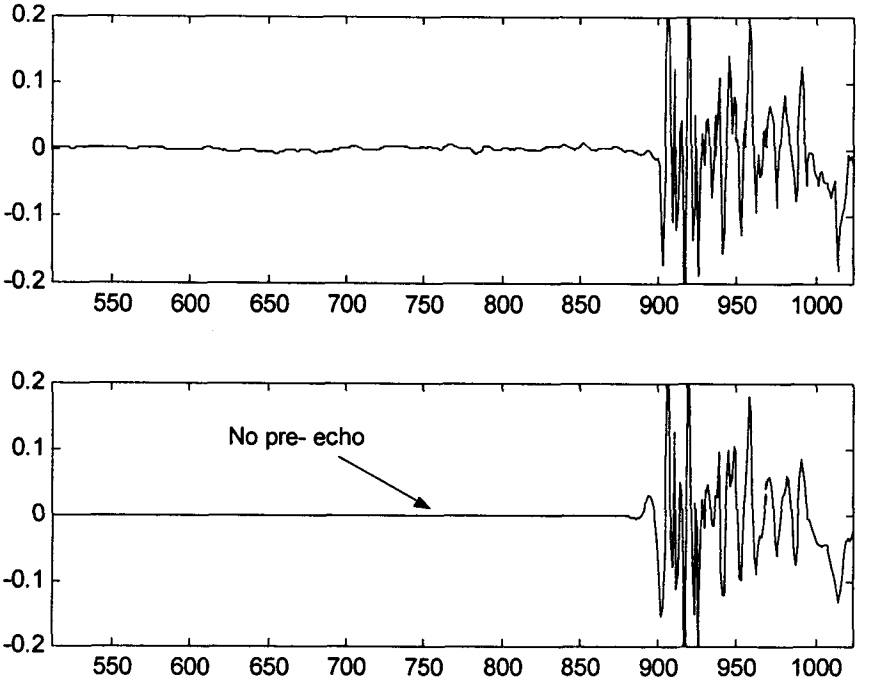


Fig.4.23: Pre- echo distortion elimination in the proposed coder

The classical signal-to-noise ratio expressed in dB is

$$\text{SNR} = 10\log_{10}\left[\frac{\sum_n s(n)^2}{\sum_n [\hat{s}(n) - s(n)]^2}\right]$$ where n is the sample index, s(n) is the original PCM audio signal and $\hat{s}(n)$ is the decoded uncompressed audio signal..

One of the characteristics of audio signals, like speech signals, is its non-stationary nature (i.e., time – varying) which results in some segments with high energy and other segments with low energy. If the error energy is more or less constant, the resulting SNR is deceptively high because the perceptual effects of the noise in the regions of lower error energy level are more severe.

4.9.1.2 Segmental Signal-to-Noise Ratio Measurements

Segmental SNR measurements are calculated since classical SNR measurements give poor results for a broad range of audio distortions across the audio bandwidth and they are not adequate for adaptive coding systems that exploit the local stationarity of waveforms. Segmental SNR can be computed by the same classical SNR measurement over short (12ms) segments of the audio signal and summed over all segments in that waveform. This results in a better estimator of audio sound quality since it penalizes coders whose performance is more time variable. The expression for segmental SNR expressed in dB is given below

$$\text{Segmental SNR} = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10}\left[\frac{\sum_{n=N*m}^{N*m+N-1} s(n)^2}{\sum_{n=N*m}^{N*m+N-1} [\hat{s}(n) - s(n)]^2}\right]$$ where N is the segment length of 12 ms or 512 samples data, at 44.1 kHz sampling rate, M is the total number of segments in the audio signal waveform, s(n) is the original PCM audio signal and $\hat{s}(n)$ is the decoded uncompressed audio signal.

Such measurements show certain physical differences between the original signals and compressed signal waveforms, but do not necessarily give a true representation of the way it would be perceived. In addition, the perceptual quality of a coder does not correlate well with the SSNR obtained from different coders for diverse

inputs such as audio. Therefore, the SSNR values for the various audio signals should not be used as an ultimate indication of coder performances. Nevertheless, it would not be unreasonable to predict subjective quality from the objective measurements of various audio signals applied to each coder.

4.9.2 Power Spectra Measurements

The Discrete Fourier Transform (DFT) of a single frame from a continuous time process is often not a good indication of the true spectrum of the signal. The solution to this dilemma is to take multiple DFTs from successive short sequences from the same signal source and take the time average of the power spectrum. Therefore, 512 - point, Hanning- windowed, non-overlapping, short-time DFT-spectra blocks are usually calculated and normalised for 96dB dynamic range. These power spectra measurements are also known as averaged periodograms.

4.9.3 Subjective Evaluation

Musical signals that code according to the auditory-masking models do not reflect the way they will be perceived by the human ear when measured using conventional objective methods. It is proposed that critical listening can reveal artifacts that existing measurement techniques fail to detect and vice-versa. Again, it must be emphasised that the ear is the final judge of accuracy in any sound restoration process. No matter how good the results obtained from an objective measurement, it is still not sufficient to conclusively evaluate a processing technique. Subjective test-measurements should therefore take part in evaluating the perceived quality of any perceptual - based codec.

Perceived signal quality is often measured on a five-point scale that is well known as the Mean-Opinion-Score (MOS) scale. It consists of an average over a large number of audio signals and listeners evaluating signal quality. The five points in the MOS impairment scale [1] are defined as :

5 imperceptible

4 perceptible but not annoying

3 **slightly annoying**

2 **annoying**

1 **very annoying**

During this thesis, informal MOS test measurements have been done to evaluate the quality of compressed signals with ten subjects including the author. Therefore, a Compact Disc (CD) is provided for the assessment of the some of the audio signals. However, it should be noted that some of these signals are presented to the audience of international conferences and seminars as well, with excellent comments.

Power Spectra Measurements

Power spectra measurements have already been explained in Section.4.9.2. Fig.4.24.shows the power spectra of the error signal (i.e., the difference between the original and reconstructed signals) for one frame of an audio signal 'castanets', along with the power spectrum of the original signal, for the sake of comparison.

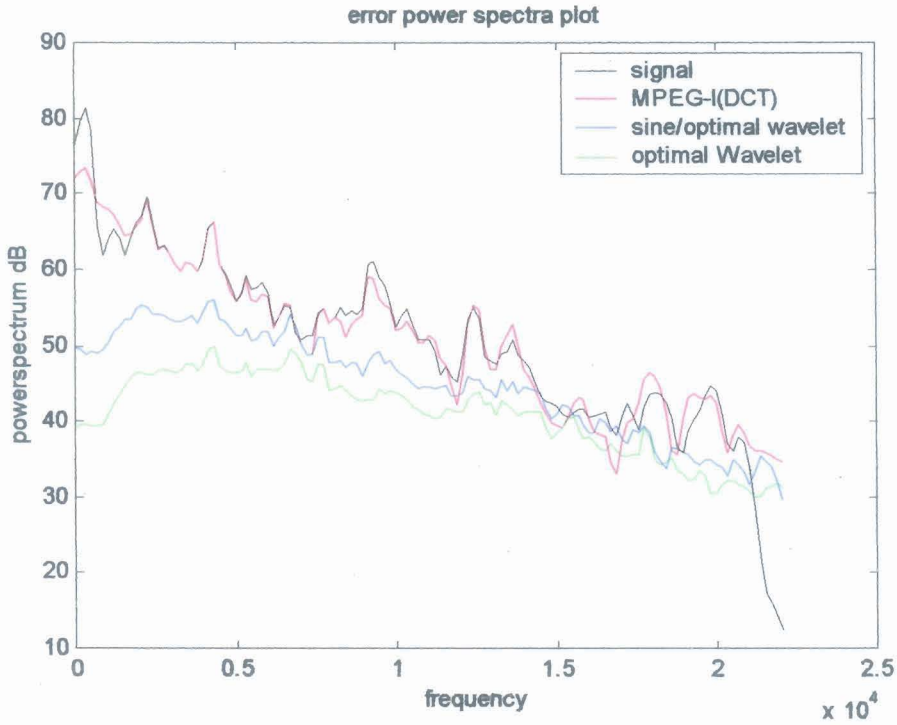


Fig.4.24: Error power spectra plot (one frame of 'castanets')

It can be observed from the above figure that with optimal wavelet and sine/optimal wavelet, the requantization noise is below the signal power spectrum upto around 11 kHz. However, with DCT, error power spectrum crosses the signal spectrum at some frequencies. In comparison to the power spectrum of the original signal, let us distinguish between three frequency regions.

At low frequencies, up to approximately 1 kHz, requantization noise introduced by the wavelet based coder is lower than the energy of the original signal and is therefore masked. But with DCT, quantization noise is not fully masked.

At medium frequencies, between 1- 4 kHz, where the human hearing is most sensitive, the error energy associated with the optimal wavelet basis and sinusoidal/optimum wavelet basis is below the energy of the original signal and therefore masked. But the quantization noise due to DCT is not masked.

At high frequencies, above 4 kHz, it seems that quantization noise due to DCT is above the energy of the original signal. At frequencies above 11 kHz, quantization noise due to optimum wavelet and sinusoidal/optimum wavelet is above the energy of the original signal. However, here the importance of the hearing threshold in quiet has to be taken into consideration. This means that at high frequencies the error energy is allowed to be higher than the energy of the original signal since it is subjectively imperceptible. Error energy associated with sine/optimum wavelet basis is above that of with optimum wavelet basis. This directly implies that the compression ratio with sine/optimum wavelet basis is more than with optimum wavelet basis.

It should be noted that the results obtained with subjective listening tests (i.e., MOS), Compression Ratio and SSNR values calculated are in close agreement with the error power spectrum plots.

Performance of the enhanced DWT audio coder on some typical audio signals are presented in Tables 4.11- 4.25 and Figs.4.25 - 4.28. The performance of DCT on audio coding in terms of SSNR, compression ratio and MOS values is shown in Table 4.26. The corresponding audio files in *.wav* format are provided in the attached CD.

The results presented in Tables 4.11– 4.26 show that compression ratio obtained for all audio signals is maximum with the new vector quantization scheme. But Mean Opinion Score(MOS) and SSNR are minimum in this case. The main reason for this is that the maximum length of the code book was restricted to 2^{12}, according to the distance criterion selected for the code book design. But, in the case of combined scalar and vector quantization, MOS and compression ratio are higher than that of with vector quantization (VQ) alone because VQ is used only in high frequency region where human ears are less sensitive. As a result, more amount of quantization noise can be hidden under the masking curve in these frequency regions and hence number of bits required for representing the signal components in these regions are also less.

cast111.wav, cast112.wav, cast113.wav, cast121.wav *etc.* are the reconstructed versions of the same audio signal 'castanets.wav' using different optimisation methods and quantization schemes. The optimisation method and quantization scheme used in each case are also mentioned in the tables. Same names are given for all these signals stored in the attached CD also.

Power spectra plots of original signal and reconstructed signals are presented in Figs. 4.25 – 4.28 and these are in close agreement with the results shown in Tables 4.11- 4.26. In Fig. 4.25 (DCT and Scalar Quantization), it can be seen that error power spectrum crosses the signal power spectrum at some frequencies and hence requantization noise will not be fully masked. But in the case of DWT, error power spectrum lies below the signal power spectrum upto 20 kHz in the case of scalar quantization and combined scalar plus vector quantization. Hence the quantization noise will be masked in these cases. But, in the case of vector quantization scheme, the error power spectrum is very closer to the signal power spectrum and they cross each other in the low frequency region. Hence, in this region, quantization noise is not fully masked and this resulted in the decrease of Mean Opinion Score values in the case of DWT and vector quantization.

Table 4.11 Results of Compression with Optimisation Method 1 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast111.wav	DCT/DWT	Scalar Quantization	8.46	13.6	4.6
cast112.wav	DCT/DWT	Vector Quantization	12.57	10	3.8
cast113.wav	DCT/DWT	Scalar + Vector Quatization	9.4	11.88	4.5

Table 4.12 Results of Compression with Optimisation Method 1 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal111.wav	DCT/DWT	Scalar Quantization	6.986	22.15	4.2
kadal112.wav	DCT/DWT	Vector Quantization	14.12	13.37	3.7
kadal113.wav	DCT/DWT	Scalar + Vector Quantization	9.22	18	4.2

Table 4.13 Results of Compression with Optimisation Method 1 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest111.wav	DCT/DWT	Scalar Quantization	9.44	22	4.3
mpegtestt112.wav	DCT/DWT	Vector Quantization	15.78	16	3.8
mpegtest113.wav	DCT/DWT	Scalar + Vector Quantization	12.2	18	4.3

Table 4.14 Results of Compression with Optimisation Method 1 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3111.wav	DCT/DWT	Scalar Quantization	7.9	17.7	4.2
else3112.wav	DCT/DWT	Vector Quantization	12.07	14.4	3.8
else3113.wav	DCT/DWT	Scalar + Vector Quantization	10.45	15.7	4.2

Table 4.15 Results of Compression with Optimisation Method 1 'sitar.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar111.wav	DCT/DWT	Scalar Quantization	4.9	22.4	4.6
sitar112.wav	DCT/DWT	Vector Quantization	10.6	13.2	3.9
sitar113.wav	DCT/DWT	Scalar + Vector Quantization	6.74	18.2	4.6

Table 4.16 Results of Compression with Optimisation Method 2 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast121.wav	DCT/DWT	Scalar Quantization	8.64	12.7	4.6
cast122.wav	DCT/DWT	Vector Quantization	12.7	9.9	3.8
cast123.wav	DCT/DWT	Scalar + Vector Quatization	9.6	11.8	4.5

Table 4.17 Results of Compression with Optimisation Method 2 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal121.wav	DCT/DWT	Scalar Quantization	8.37	19.5	4.2
kadal122.wav	DCT/DWT	Vector Quantization	15.02	12.7	3.7
kadal123.wav	DCT/DWT	Scalar + Vector Quantization	10.1	17.1	4.2

Table 4.18 Results of Compression with Optimisation Method 2 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest121.wav	DCT/DWT	Scalar Quantization	11.78	19.3	4.3
mpegtestt122.wav	DCT/DWT	Vector Quantization	17.02	15.2	3.8
mpegtest123.wav	DCT/DWT	Scalar + Vector Quantization	14.58	16.9	4.3

Table 4.19 Results of Compression with Optimisation Method 2 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3121.wav	DCT/DWT	Scalar Quantization	9.3	16.6	4.2
else3122.wav	DCT/DWT	Vector Quantization	13.65	13.7	3.8
else3123.wav	DCT/DWT	Scalar + Vector Quantization	11.46	15.1	4.2

Table 4.20 Results of Compression with Optimisation Method 2 'sitar.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar121.wav	DCT/DWT	Scalar Quantization	6.2	19.3	4.6
sitar122.wav	DCT/DWT	Vector Quantization	10.8	12.9	3.9
sitar123.wav	DCT/DWT	Scalar + Vector Quantization	7.6	16.7	4.6

Table 4.21 Results of Compression with Optimisation Method 3 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast131.wav	DCT/DWT	Scalar Quantization	8.6	12.9	4.6
cast132.wav	DCT/DWT	Vector Quantization	12.7	9.9	3.8
cast133.wav	DCT/DWT	Scalar + Vector Quatization	9.6	11.8	4.5

Table 4.22 Results of Compression with Optimisation Method 3 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal131.wav	DCT/DWT	Scalar Quantization	8.37	19.5	4.2
kadal132.wav	DCT/DWT	Vector Quantization	15.02	12.7	3.7
kadal133.wav	DCT/DWT	Scalar + Vector Quantization	10.1	17.1	4.2

Table 4.23 Results of Compression with Optimisation Method 3 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest131.wav	DCT/DWT	Scalar Quantization	11.78	19.3	4.3
mpegtestt132.wav	DCT/DWT	Vector Quantization	17	15.2	3.8
mpegtest133.wav	DCT/DWT	Scalar + Vector Quantization	14.58	16.8	4.3

Table 4.24 Results of Compression with Optimisation Method 3 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3131.wav	DCT/DWT	Scalar Quantization	9.3	16.6	4.2
else3132.wav	DCT/DWT	Vector Quantization	13.6	13.7	3.8
else3133.wav	DCT/DWT	Scalar + Vector Quantization	10.9	15.4	4.2

Table 4.25 Results of Compression with Optimisation Method 3 'sitar.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar131.wav	DCT/DWT	Scalar Quantization	6.2	19.3	4.6
sitar132.wav	DCT/DWT	Vector Quantization	10.8	12.4	3.9
sitar133.wav	DCT/DWT	Scalar + Vector Quantization	7.5	16.9	4.6

Table 4.26 Results of Compression with DCT and Scalar Quantization

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
castdct11.wav	DCT	Scalar Quantization	7.7	10	3.5
kadaldct11.wav	DCT	Scalar Quantization	8.1	15	4.1
megtestdct11.wav	DCT	Scalar Quantization	8	23	4.2
else3dct11.wav	DCT	Scalar Quantization	9.2	16	3.9
sitardct11.wav	DCT	Scalar Quantization	6.78	14.5	4.4

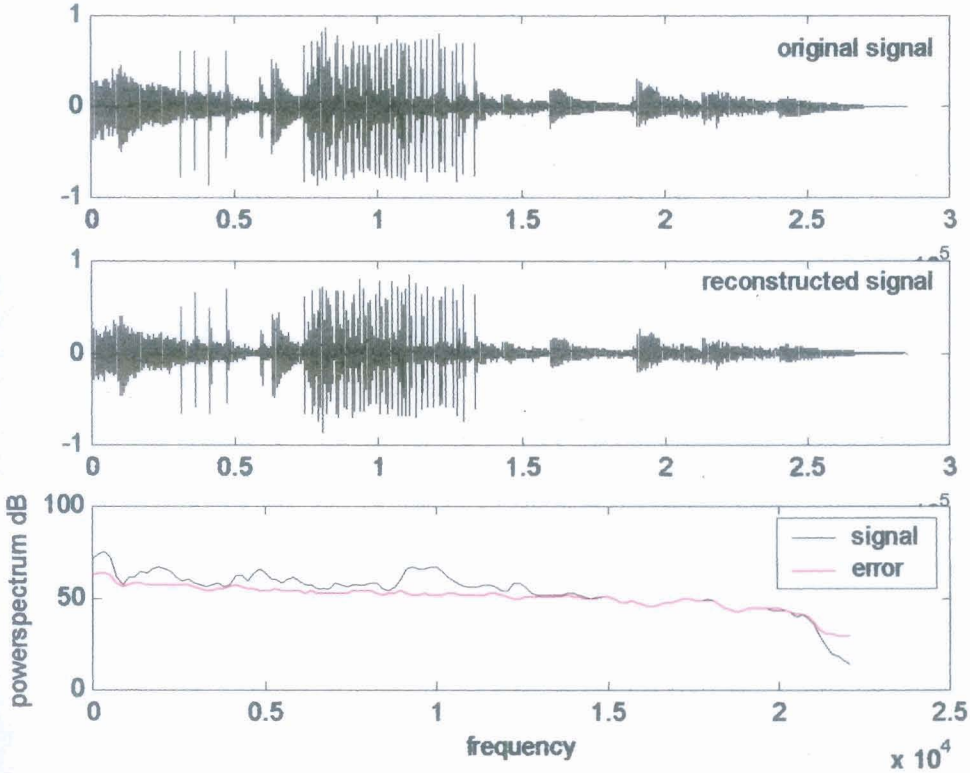


Fig.4.25: Power Spectra Plot -castanets: DCT and Scalar Quantization -castdct1.wav

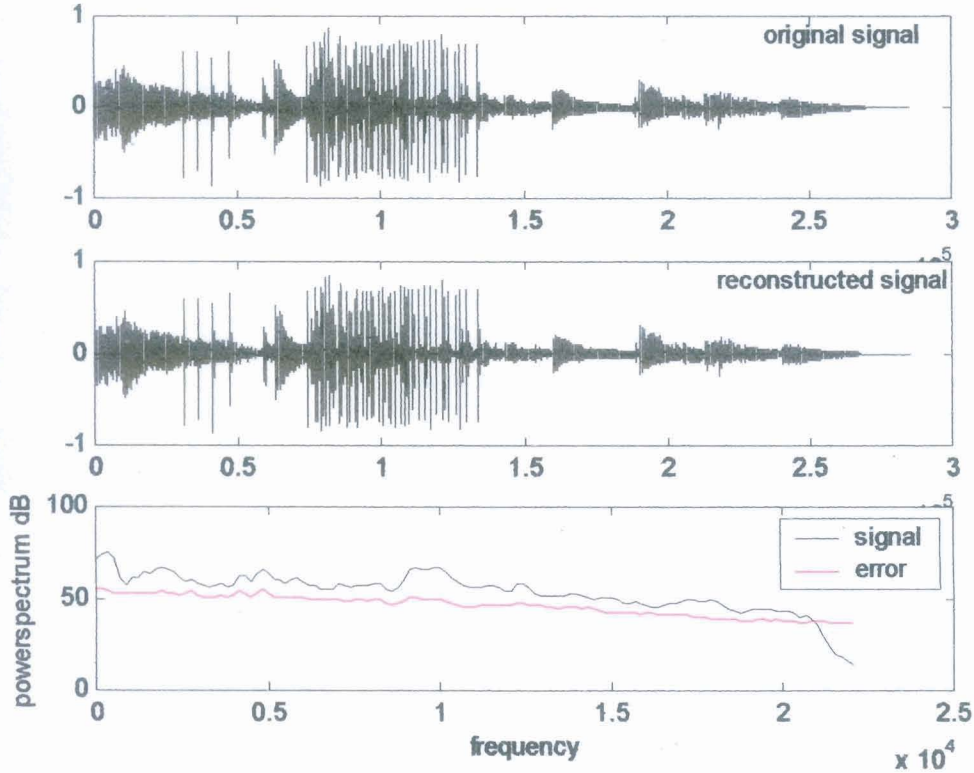


Fig.4.26: Power Spectra Plot - castanets: DWT and Scalar Quantization –cast111.wav

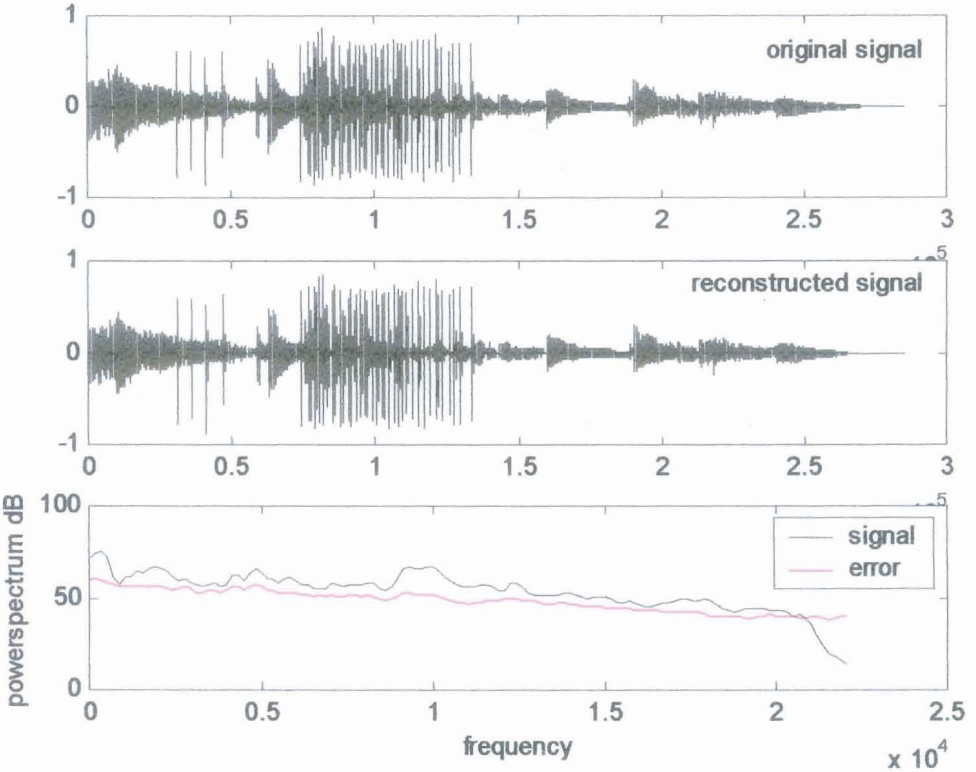


Fig.4.27: Power Spectra Plot -castanets: DWT and Vector Quantization –cast112.wav

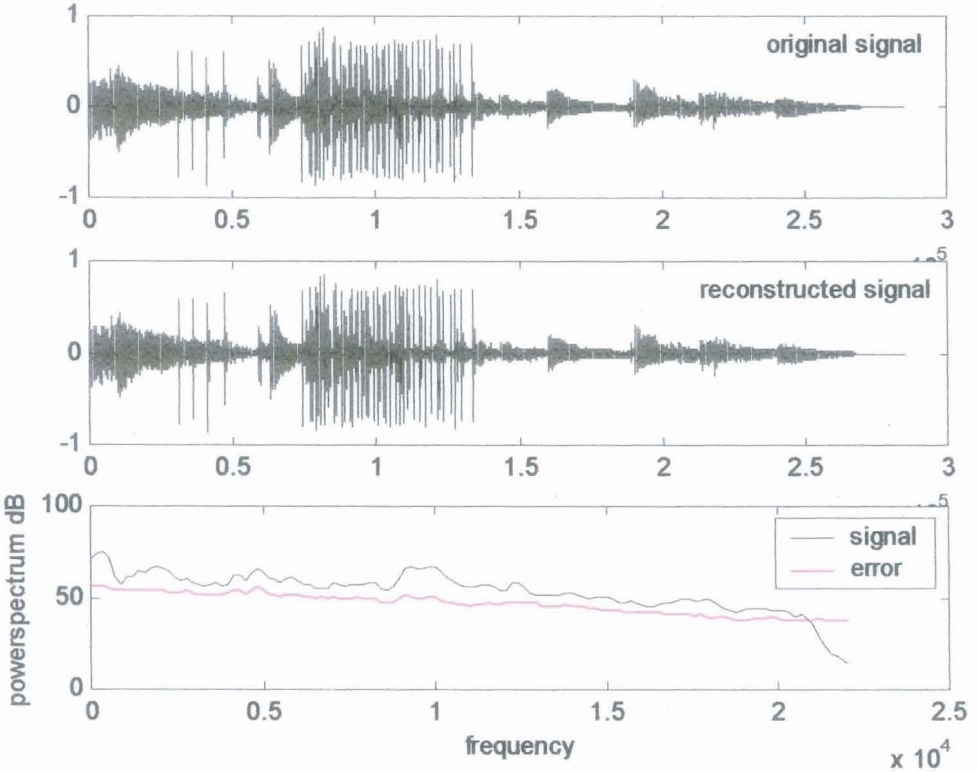


Fig.4.28: Power Spectra Plot -castanets: DWT and Scalar + VQ–cast113.wav

4.10 Summary

In this chapter, a Discrete Wavelet Transform based perceptual audio coding scheme has been developed and implemented. Performance of various wavelet families are studied and compared. A novel theoretical result : "Compression ratio not only depends on the transform (basis) used, but also depends on the statistical features of the source signal itself" is experimentally proved here. Hence, three different optimisation methods have been developed to identify the best wavelet basis for each frame of the audio source according to the local characteristics/statistical features of the signal. Mean Opinion Score and Compression ratio are comparable to MPEG standard. A major artifact of MPEG scheme - **"Pre-echo"** distortion is almost eliminated here. A new vector quantization scheme named as Hit Book method has also been developed in this chapter. The implementation results show that this Hit Book method gives maximum compression ratio with little degradation in quality. Performance of DCT on which MPEG audio standard is based, is also tested. The experimental results revealed that DCT gives more compression for stationary audio segments than DWT, as expected, and hence a switching algorithm is also developed to switch between DCT and DWT according to the time varying characteristics of the audio signal. The enhanced wavelet based coder represents each frame of the audio signal either using sinusoidal basis or using optimal wavelet basis according to the statistical features of the signal.

Eventhough the scheme developed here is less complex, and performance is comparable to MPEG standard, one drawback is that some of the subbands cover more number of critical bands as shown in the Table 4.1. Masking threshold in each critical band is calculated and masking threshold for a subband is taken as the minimum of masking thresholds calculated for all critical bands in that subband. Hence, efficient exploitation of perceptual irrelevancies could not be met here to achieve maximum compression as possible. Wavelet packet based audio coding schemes developed in the next chapter, provides more compression ratio than this scheme, at the expense of increased computational complexity at the analysis stage.

Chapter 5

WAVELET PACKET BASED AUDIO CODING SCHEMES

5.1 Introduction

In the previous chapter, a discrete wavelet transform based perceptual audio coder with nine subbands and using optimum wavelet basis for each audio frame representation, is developed and implemented. Enhancement of the above coder by incorporating a novel switching scheme to switch between DCT and DWT is also described. Since, in this scheme some of the subbands cover more than one critical band of the human auditory system, maximum amount of quantization noise which will be perceptually inaudible, could not be introduced in all the subbands to achieve maximum compression. Hence, wavelet packet based audio coding schemes, with higher compression ratios have been developed and implemented here. In the first scheme, each audio frame is decomposed into 27 subbands closely mimicking the human auditory system. Hence, quantization noise can be more effectively distributed into various subbands to achieve more compression. Performance of the wavelet packet based audio coder is validated through informal subjective listening tests using optimum wavelet basis and with three quantization schemes described in Chapter 4. Enhanced version of this scheme (using switching algorithm) is also implemented and validated through subjective listening tests.

Second scheme is a low complexity (computationally more efficient) wavelet packet based coding scheme. MPEG, DWT based audio coder, and the first scheme described above, use separate high resolution FFT stage for psychoacoustic model implementation. Hence, a new low complexity optimum wavelet packet based audio coder, in which psychoacoustic model design is integrated into the design of the analysis filter bank, is developed and implemented. The compression ratio obtained is same as that of scheme 1, but Mean Opinion score (MOS) is slightly less in some cases.

One reason is due to the use of wavelet packet (whose frequency resolution is less than that of FFT) for the implementation of psychoacoustic model. Second reason is the absence of switching between sinusoidal basis and wavelet basis.

Third scheme is a scalable perceptual audio coder using wavelet packets, which supports most of the industrial audio sampling frequencies for transmission of speech/audio over internet, teleconferencing and other multimedia applications. Sampling frequencies supported in the proposed scheme are:

1. 11.025 kHz – Wide band speech for transmission over Internet, teleconferencing and some multimedia applications.
2. 22.05 kHz – Wideband speech and some audio signals for multimedia applications.
3. 44.1 kHz – High fidelity CD quality music for storage, transmission over Internet and multimedia applications.

5.2 Implementation of the first scheme

The block diagram of the proposed wavelet packet based audio coder is shown in Fig.5.1

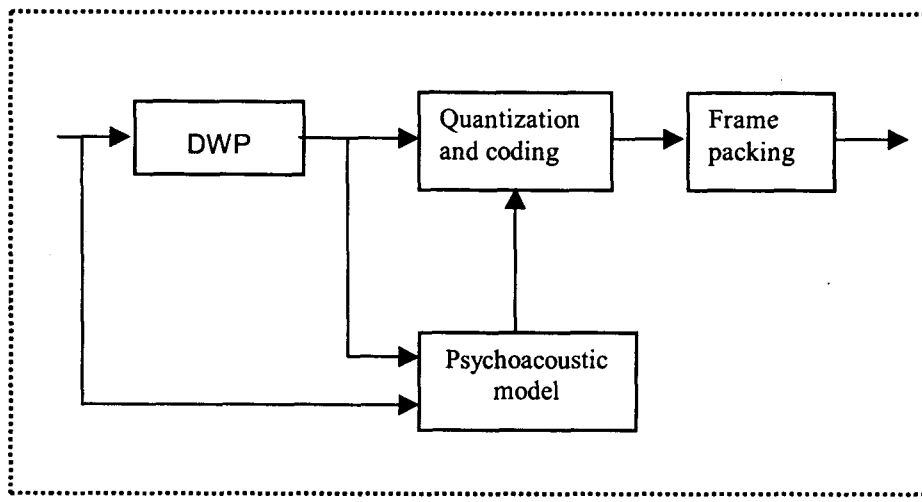


Fig.5.1: Block diagram of the wavelet packet based audio coder (scheme 1)

- **Analysis filter bank (DWP):** The signal is decomposed into 27 subbands closely mimicking the human auditory system. The wavelet packet filter bank tree structure shown in Fig.5.2 is designed for the decomposition of the audio signal. The lower and upper ends of each frequency band are also shown in the figure. The mapping between sub bands used in the proposed coder and the critical bands of the human auditory system are shown in Table 5.1.
- Sampling frequency of the audio signal = 44.1 kHz (CD sampling frequency)
- Frame size = 512 samples (12 ms duration)

5.3 Psychoacoustic Model Implementation for scheme 1

Fast Fourier Transform (FFT) technique is employed here to implement the psychoacoustic model. For each audio frame, the signal spectrum is computed through FFT. The tonal (sinusoidal like) and non tonal (noise like) components of the signal are extracted from each audio frame and for each of them the masking effect is computed. The overall masking curve is calculated by adding along the critical bands, all the masking effects and taking the absolute curve of the ear (see Fig.2.7) into account. The transparency condition is satisfied if the reconstruction error spectrum lies under the frequency masking curve. Thus, implementation of psychoacoustic model is similar to that described in Section 4.3.

5.4 Quantization and Coding

Wavelet packet coefficients which are calculated in 27 subbands are quantized in such a way that quantization noise in each frequency band is below the masking threshold. In this proposed scheme also, performance of the following quantization schemes are studied and compared.

1. Uniform scalar quantization
2. Hit book method (Vector quantization).
3. Combined scalar and vector quantization.

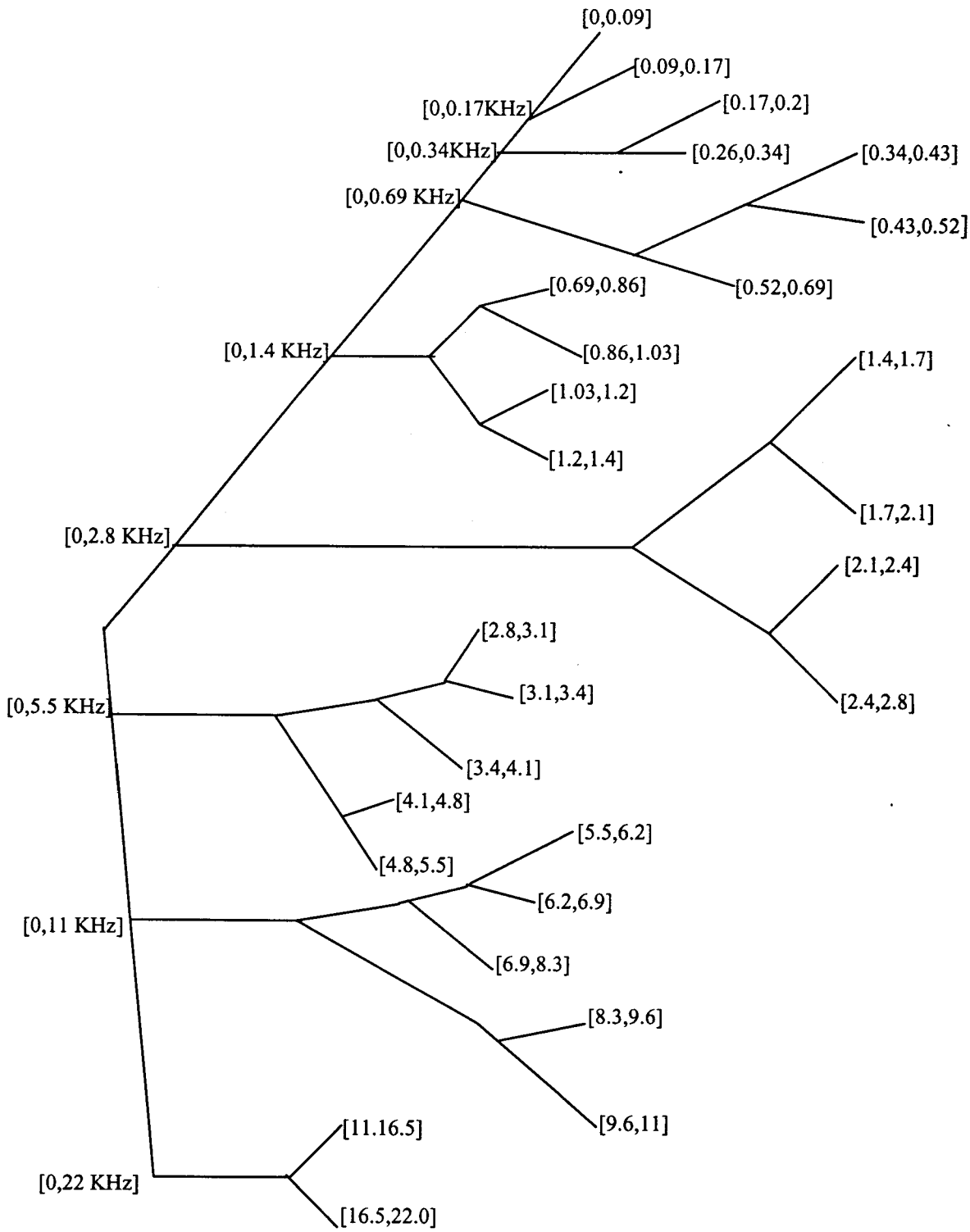


Fig.5.2: Wavelet packet decomposition

Table 5.1 Mapping between subbands used in the proposed scheme and critical bands of the human ear

Frequency band (kHz)	Critical band number
0 – 0.09	1
0.09 – 0.17	2
0.17 – 0.26	3
0.26 – 0.34	4
0.34 – 0.43	4
0.43 – 0.52	5
0.52 – 0.69	6
0.69 – 0.86	7
0.86 – 1.03	8
1.03 – 1.2	9&10
1.2 – 1.4	11
1.4 – 1.7	12
1.7 – 2.1	13
2.1 – 2.4	14
2.4 – 2.8	15
2.8 – 3.1	16
3.1 – 3.4	17
3.4 – 4.1	18
4.1 – 4.8	19
4.8 – 5.5	19
5.5 – 6.2	20
6.2 – 6.9	21
6.9 – 8.3	22
8.3 – 9.6	22
9.6 – 11	23
11 – 16.5	24
16.5 – 22	25

5.5 Performance of the proposed scheme (Scheme 1)

Performance of the proposed wavelet packet based audio coder is tested with the optimum wavelet basis and the switching scheme developed in the last chapter. The block diagram of the enhanced wavelet packet based audio coder is shown in Fig.5.3. The implementation results are presented in Tables 5.2 - 5.16 and Figs.5.4 - 5.9. The reconstructed versions of the same audio signal with same optimisation method and different quantization schemes are listed in each table. The values of compression ratio, SSNR and MOS obtained in each case are given in these tables. Higher compression ratios are obtained in most of the cases with MOS values almost same as that of DWT based audio coder. Error power spectra plots shown in Figs. 5.4 –5.9 are in close agreement with the results shown in Tables 5.2-5.16.

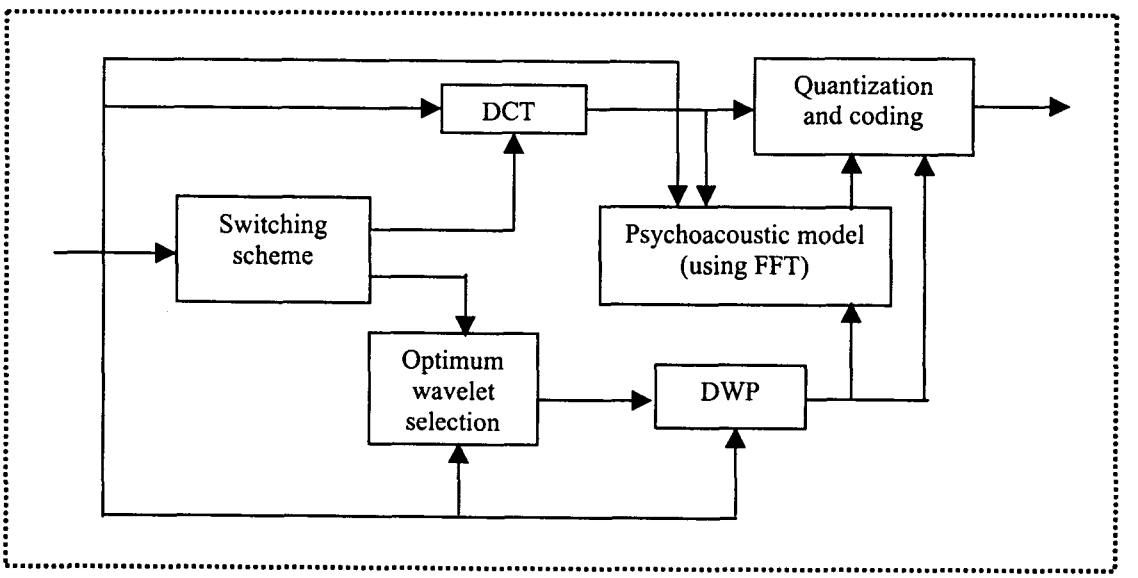


Fig.5.3: Block diagram of the enhanced wavelet packet based audio encoder

Table 5.2 Results of Compression with Optimisation Method 1 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast211.wav	DCT/DWP	Scalar Quantization	8.8	13.1	4.5
cast212.wav	DCT/DWP	Vector Quantization	13.57	9.7	3.8
cast213.wav	DCT/DWP	Scalar + Vector Quatization	9.9	11.68	4.4

Table 5.3 Results of Compression with Optimisation Method 1 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal211.wav	DCT/DWP	Scalar Quantization	7.58	21.5 dB	4.2
kadal212.wav	DCT/DWP	Vector Quantization	14.82	13.1 dB	3.7
kadal213.wav	DCT/DWP	Scalar + Vector Quantization	10.12	17.5 dB	4.2

Table 5.4 Results of Compression with Optimisation Method 1 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest211.wav	DCT/DWP	Scalar Quantization	10.9	21	4.3
mpegtest212.wav	DCT/DWP	Vector Quantization	16.18	15.1	3.7
mpegtest213.wav	DCT/DWP	Scalar + Vector Quantization	12.9	17.6	4.3

Table 5.5 Results of Compression with Optimisation Method 1 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3211.wav	DCT/DWP	Scalar Quantization	8.6	17.1 dB	4.2
else3212.wav	DCT/DWP	Vector Quantization	13.01	13.9 dB	3.8
else3213.wav	DCT/DWP	Scalar + Vector Quantization	11.5	14.9 dB	4.2

Table 5.6 Results of Compression with Optimisation Method 1 'sitar111.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar211.wav	DCT/DWP	Scalar Quantization	5.4	22.1	4.5
sitar212.wav	DCT/DWP	Vector Quantization	11.3	12.8	3.9
sitar213.wav	DCT/DWP	Scalar + Vector Quantization	7.3	17.8	4.5

Table 5.7 Results of Compression with Optimisation Method 2 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast221.wav	DCT/DWP	Scalar Quantization	8.92	13.01	4.5
cast222.wav	DCT/DWP	Vector Quantization	13.77	9.1	3.8
cast223.wav	DCT/DWP	Scalar + Vector Quatization	10.2	11.2	4.5

Table 5.8 Results of Compression with Optimisation Method 2 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal221.wav	DCT/DWP	Scalar Quantization	7.8	21.1	4.2
kadal222.wav	DCT/DWP	Vector Quantization	15.02	12.8	3.7
kadal1223.wav	DCT/DWP	Scalar + Vector Quantization	10.6	16.9	4.2

Table 5.9 Results of Compression with Optimisation Method 2 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest221.wav	DCT/DWP	Scalar Quantization	11.4	20.2	4.2
mpegtestt222.wav	DCT/DWP	Vector Quantization	17.08	14.2	3.8
mpegtest223.wav	DCT/DWP	Scalar + Vector Quantization	13.4	17.1	4.3

Table 5.10 Results of Compression with Optimisation Method 2 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3221.wav	DCT/DWP	Scalar Quantization	8.9	16.8	4.2
else3222.wav	DCT/DWP	Vector Quantization	13.9	13.1	3.7
else3223.wav	DCT/DWP	Scalar + Vector Quantization	12.1	14.2	4.2

Table 5.11 Results of Compression with Optimisation Method 2 'sitar111.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar221.wav	DCT/DWP	Scalar Quantization	5.8	21.8	4.5
sitar222.wav	DCT/DWP	Vector Quantization	11.9	12.2	3.9
sitar223.wav	DCT/DWP	Scalar + Vector Quantization	7.7	17.1	4.5

Table 5.12 Results of Compression with Optimisation Method 3 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast231.wav	DCT/DWP	Scalar Quantization	8.92	13.01	4.6
cast232.wav	DCT/DWP	Vector Quantization	13.77	9.1	3.8
cast233.wav	DCT/DWP	Scalar + Vector Quatization	10.2	11.2	4.5

Table 5.13 Results of Compression with Optimisation Method 3 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal231.wav	DCT/DWP	Scalar Quantization	7.7	21.2	4.2
kadal232.wav	DCT/DWP	Vector Quantization	15.02	12.8	3.7
kadal233.wav	DCT/DWP	Scalar + Vector Quantization	10.4	17.1	4.2

Table 5.14 Results of Compression with Optimisation Method 3 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest231.wav	DCT/DWP	Scalar Quantization	11.8	19.2	4.3
mpegtestt232.wav	DCT/DWP	Vector Quantization	17.08	14.2	3.7
mpegtest233.wav	DCT/DWP	Scalar + Vector Quantization	14.6	16.1	4.3

Table 5.15 Results of Compression with Optimisation Method 3 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3231.wav	DCT/DWP	Scalar Quantization	9.9	15.1	4.2
else3232.wav	DCT/DWP	Vector Quantization	13.9	13.1	3.8
else3233.wav	DCT/DWP	Scalar + Vector Quantization	12.1	14.2	4.2

Table 5.16 Results of Compression with Optimisation Method 3 'sitar111.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar231.wav	DCT/DWP	Scalar Quantization	6.6	18.9	4.5
sitar232.wav	DCT/DWP	Vector Quantization	11.8	12.3	3.9
sitar233.wav	DCT/DWP	Scalar + Vector Quantization	7.51	16.4	4.5

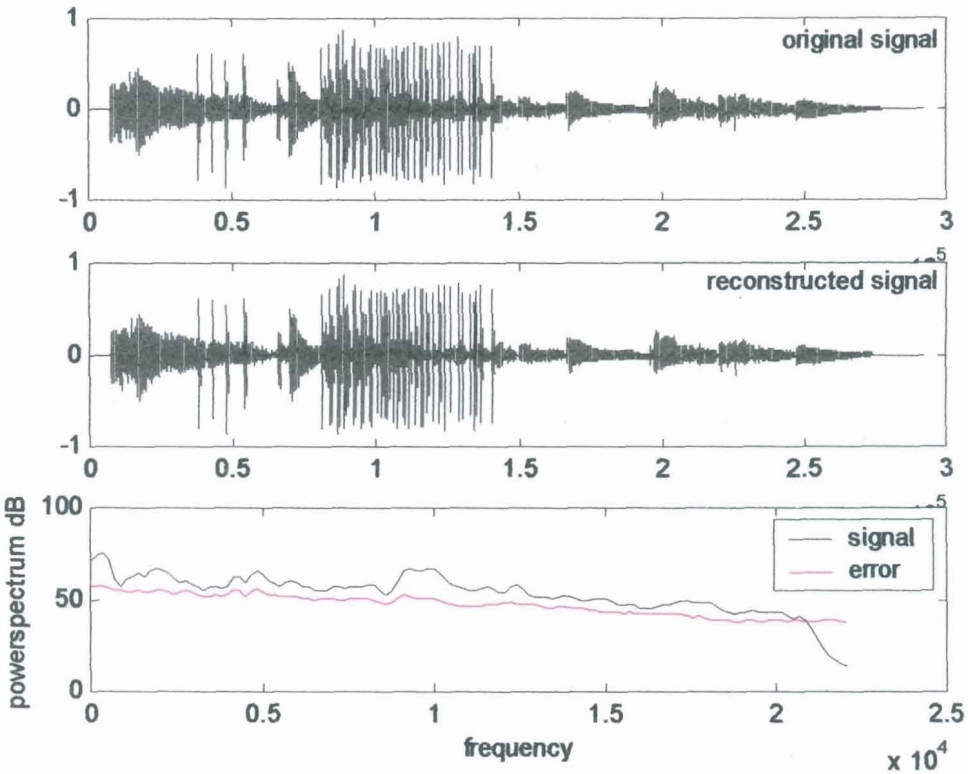


Fig. 5.4: Power spectra plot -castanets: DWP and scalar quantization –cast211.wav

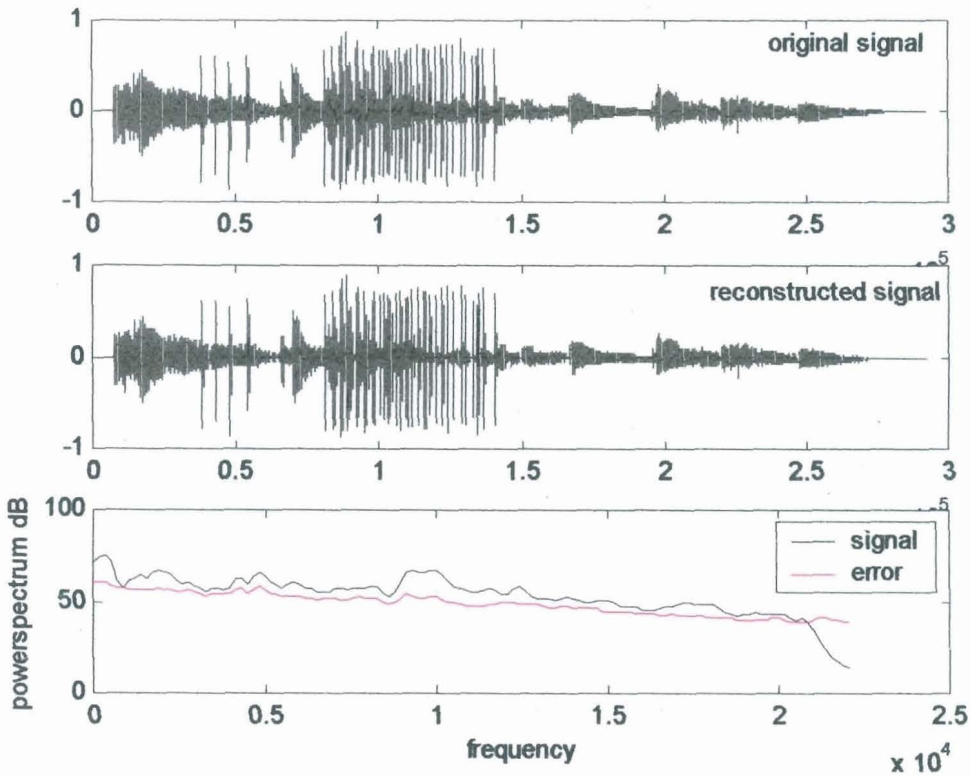


Fig.5.5: Power spectra plot –castanets: DWP and vector quantization –cast212.wav

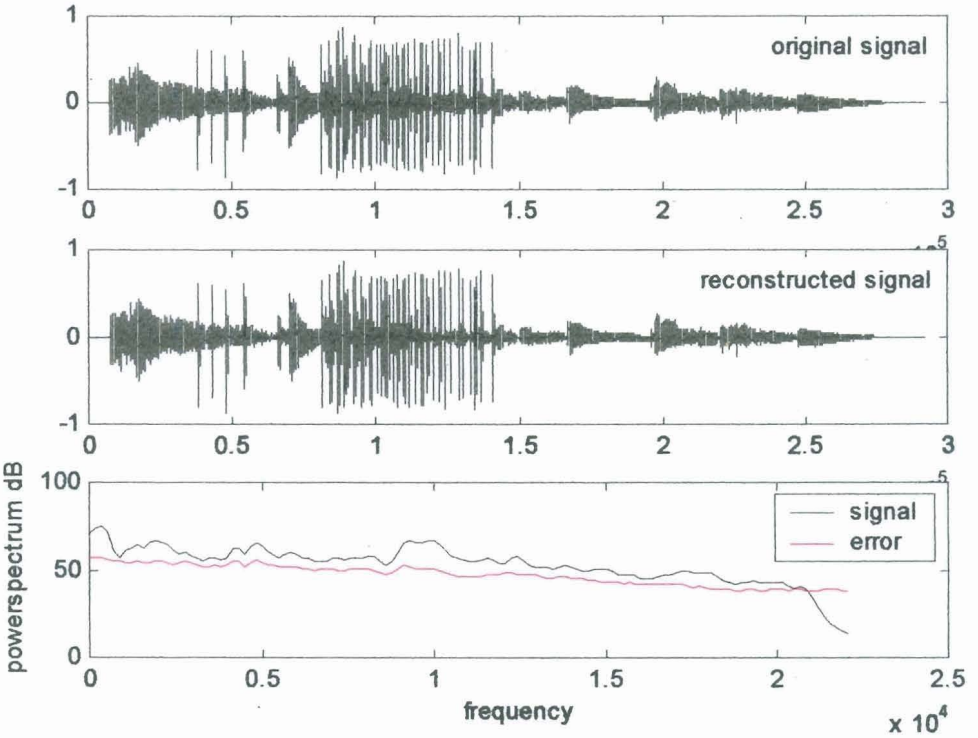


Fig.5.6: Power spectra plot -castanets: DWP and scalar + vq- cast213.wav

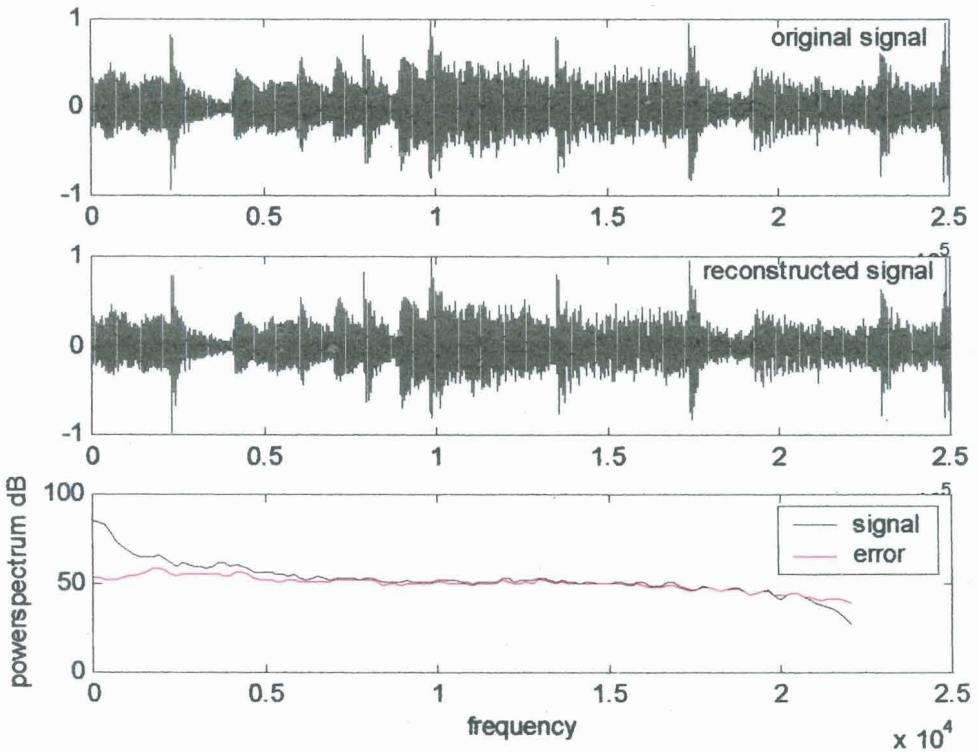


Fig.5.7: Power spectra plot –else3: DWP and scalar quantization - elsemono211.wav

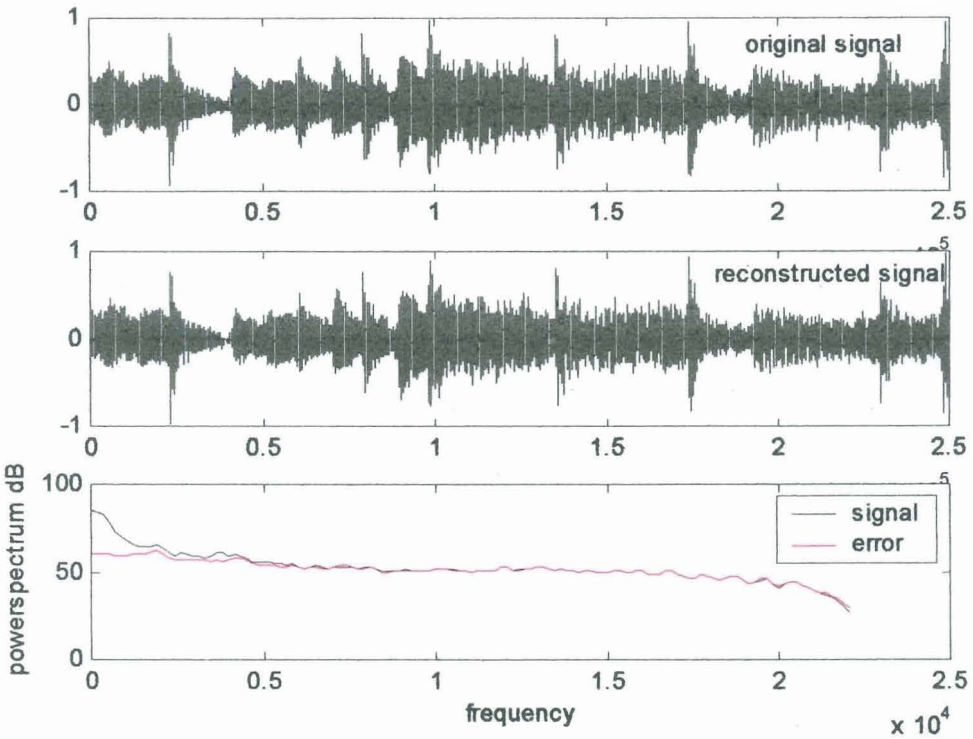


Fig.5.8: Power spectra plot –else3: DWP and vector quantization -elsemono212.wav

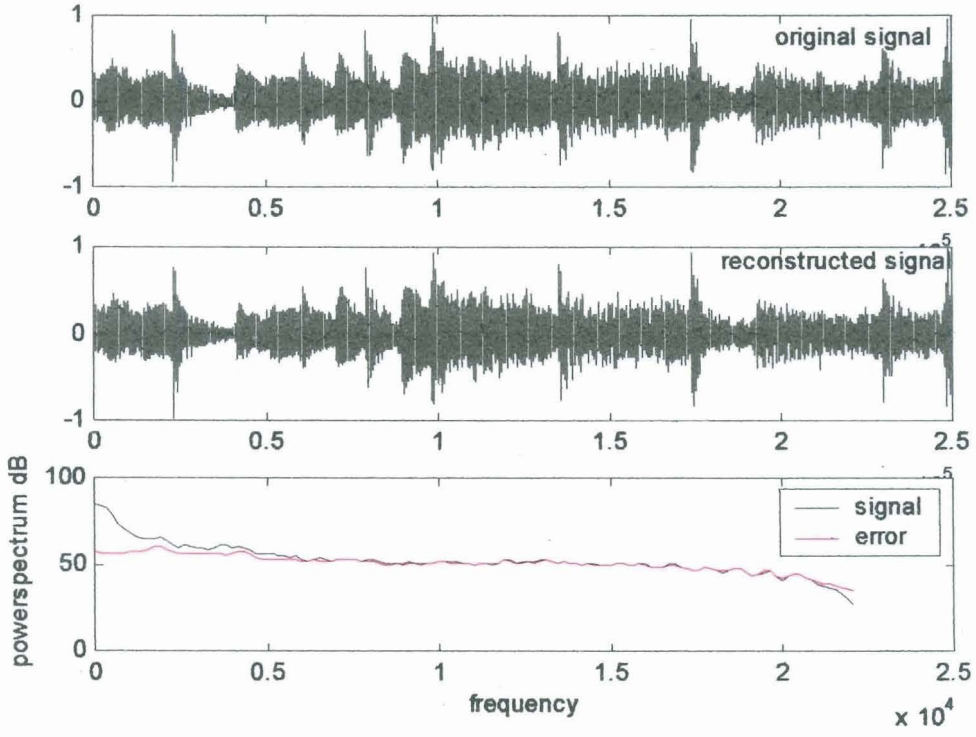


Fig. 5.9: Power spectra plot –else3: DWP and scalar + vq - elsemono213.wav

5.6 Implementation of the second scheme (Low complexity scheme)

The block diagram of the proposed low complexity wavelet packet based audio coding scheme which uses same wavelet packet tree structure for analysis filterbank as well as for the implementation of psychoacoustic model is given in Fig.5.10.

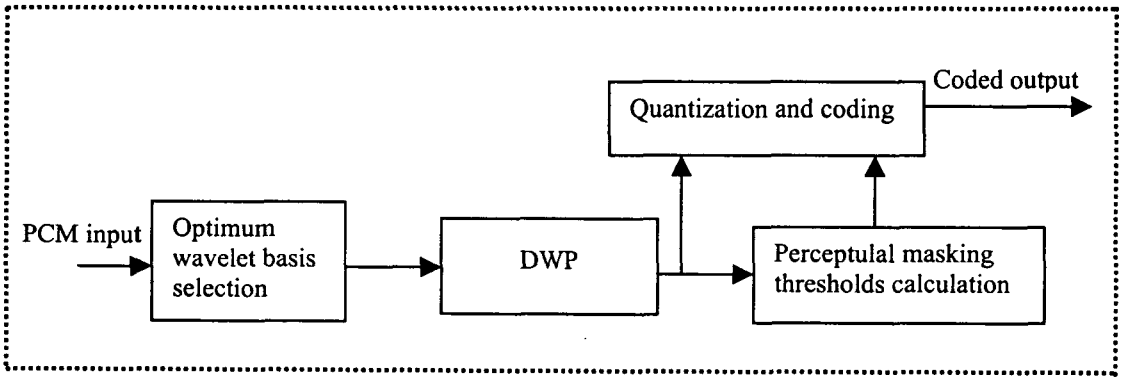
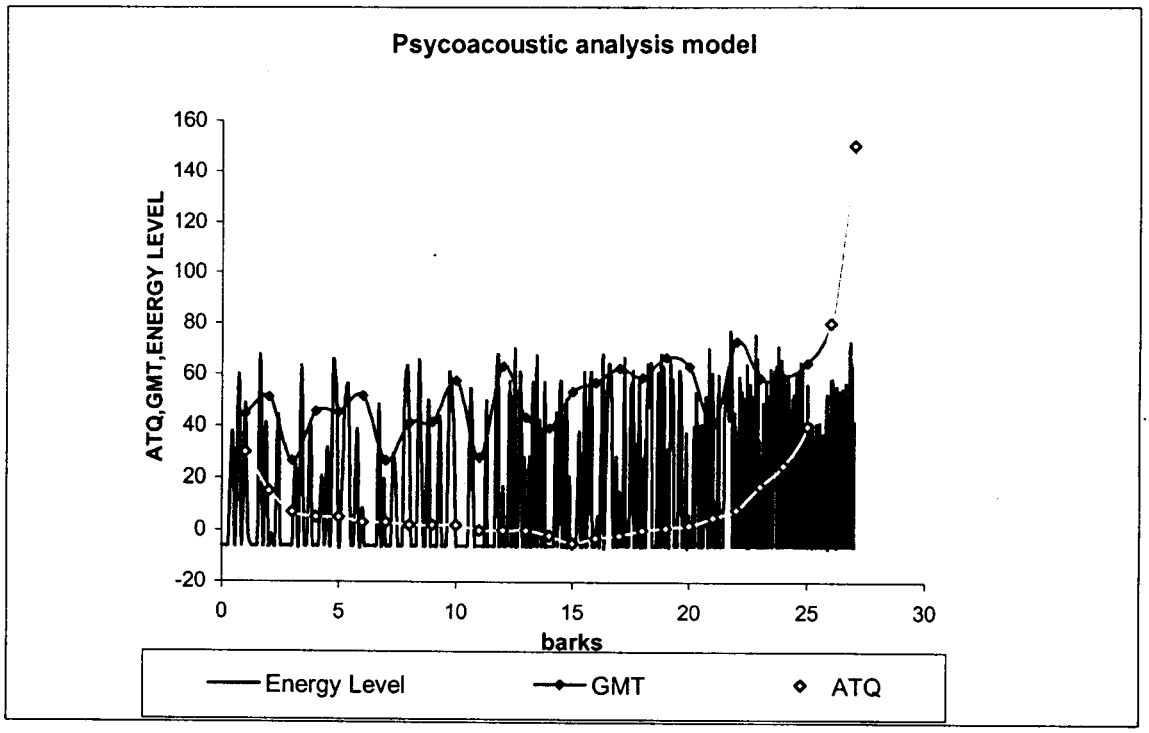


Fig.5.10: Block diagram of the proposed low complexity wavelet packet based audio coder

Each frame of the audio signal is decomposed into 27 subbands (mimicking the human auditory system) using optimum wavelet basis. Wavelet packet tree structure used for this decomposition is same as the structure shown in Fig.5.2. Instead of determining the masking thresholds in the Fourier domain, the wavelet packet coefficients from the analysis filterbank are used to calculate signal energy and perceptual masking thresholds for 27 subbands (see Fig.5.11). Quantization and coding of each subband samples are done such that requantization noise in each subband is below the masking threshold in the respective band (see Fig.5.12). Mapping between the various subbands used in this scheme and critical bands of the human ear are same as before and as shown in Table 5.1. The performance of this scheme is tested with three quantization schemes listed earlier and three optimisation methods developed in Chapter 4. Implementation results are presented in Tables 5.17-5.31 and Figs.5.13 – 5.18 for some typical audio signals. Error power spectra plots shown in Figs. 5.13-5.18 agree with the results given in Tables 5.17 – 5.31. Original and reconstructed versions of some typical audio signals are provided in the attached CD.



ATQ---Absolute Threshold in Quiet. GMT---Global Masking Threshold.

Fig. 5.11: . Global Masking Threshold for frame 10 (12ms) of castanets.wav

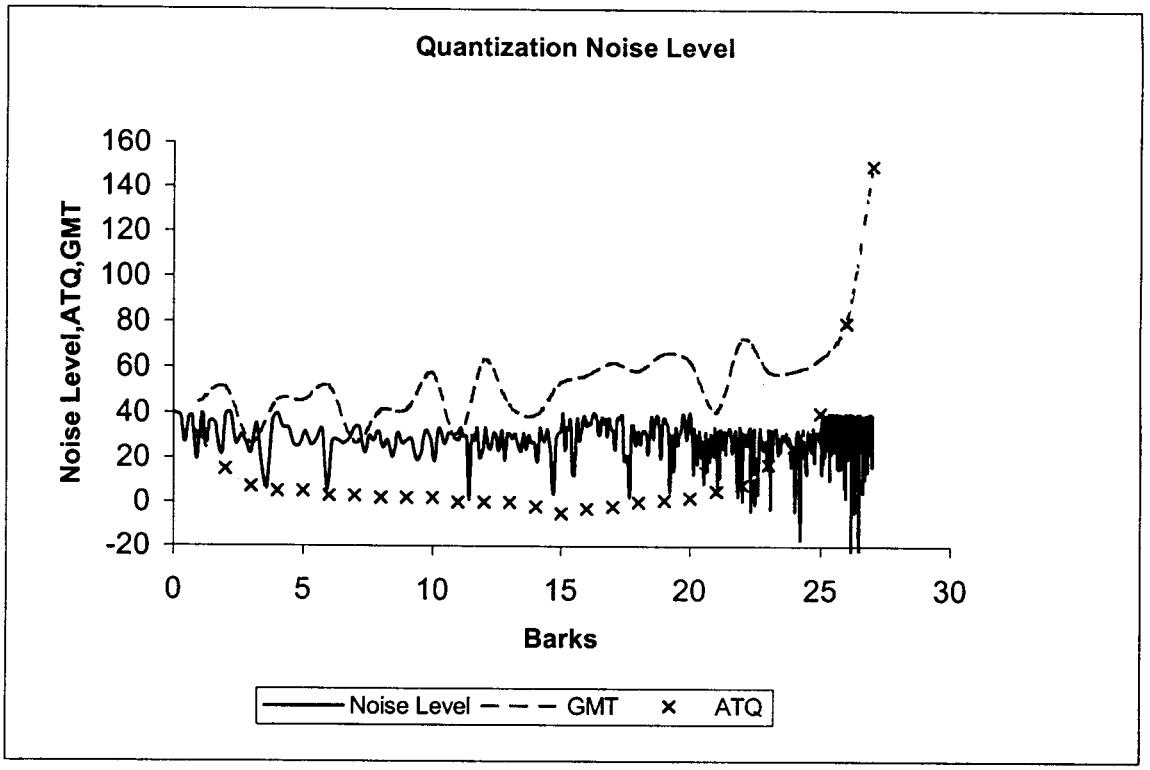


Fig.5.12: .Quantization noise level

Table 5.17 Results of Compression with Optimisation Method 1 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast311.wav	DWP	Scalar Quantization	8.8	12.7	4.4
cast312.wav	DWP	Vector Quantization	13.57	9.3	3.8
cast313.wav	DWP	Scalar + Vector Quatization	9.9	11.4	4.4

Table 5.18 Results of Compression with Optimisation Method 1 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal311.wav	DWP	Scalar Quantization	7.5	21.6	4.1
kadal312.wav	DWP	Vector Quantization	14.8	13.4	3.7
kadal313.wav	DWP	Scalar + Vector Quantization	10.1	17.7	4.1

Table 5.19 Results of Compression with Optimisation Method 1 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest311.wav	DWP	Scalar Quantization	10.9	21	4.0
mpegtestt312.wav	DWP	Vector Quantization	16.18	15.1	3.6
mpegtest313.wav	DWP	Scalar + Vector Quantization	12.9	17.6	4.0

Table 5.20 Results of Compression with Optimisation Method 1 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3311.wav	DWP	Scalar Quantization	8.6	17.1	4.1
else3312.wav	DWP	Vector Quantization	13.01	13.9	3.7
else3313.wav	DWP	Scalar + Vector Quantization	11.5	14.9	3.9

Table 5.21 Results of Compression with Optimisation Method 1 'sitar.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar311.wav	DWP	Scalar Quantization	5.4	22.1	4.5
sitar312.wav	DWP	Vector Quantization	10.9	13.1	3.9
sitar313.wav	DWP	Scalar + Vector Quantization	7.3	17.8	4.5

Table 5.22 Results of Compression with Optimisation Method 2 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast321.wav	DWP	Scalar Quantization	8.92	13.01	4.3
cast322.wav	DWP	Vector Quantization	13.77	9.1	3.8
cast323.wav	DWP	Scalar + Vector Quatization	10.2	11.2	4.2

Table 5.23 Results of Compression with Optimisation Method 2 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal321.wav	DWP	Scalar Quantization	8.8	19.4	4.1
kadal322.wav	DWP	Vector Quantization	15.0	12.8	3.7
kadal1323.wav	DWP	Scalar + Vector Quantization	10.6	16.9	4.0

Table 5.24 Results of Compression with Optimisation Method 2 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest321.wav	DWP	Scalar Quantization	11.4	20.2	4.0
mpegtest322.wav	DWP	Vector Quantization	17.08	14.2	3.5
mpegtest323.wav	DWP	Scalar + Vector Quantization	13.4	17.1	3.8

Table 5.25 Results of Compression with Optimisation Method 2 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3321.wav	DWP	Scalar Quantization	8.9	16.8	4.1
else3322.wav	DWP	Vector Quantization	13.9	13.1	3.7
else3323.wav	DWP	Scalar + Vector Quantization	12.1	14.2	4.1

Table 5.26 Results of Compression with Optimisation Method 2 'sitar.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar321.wav	DWP	Scalar Quantization	5.8	21.8	4.3
sitar322.wav	DWP	Vector Quantization	11.9	12.2	3.7
sitar323.wav	DWP	Scalar + Vector Quantization	7.7	17.1	4.2

Table 5.27 Results of Compression with Optimisation Method 3 'castanets.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
cast331.wav	DWP	Scalar Quantization	8.92	13.01	4.4
cast332.wav	DWP	Vector Quantization	13.77	9.1	3.8
cast333.wav	DWP	Scalar + Vector Quatization	10.1	11.2	4.4

Table 5.28 Results of Compression with Optimisation Method 3 'kadalinakkare.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
kadal331.wav	DWP	Scalar Quantization	8.7	19.2	4.2
kadal332.wav	DWP	Vector Quantization	15.02	12.8	3.7
kadal333.wav	DWP	Scalar + Vector Quantization	10.4	17.1	4.2

Table 5.29 Results of Compression with Optimisation Method 3 'mpegtest.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
mpegtest331.wav	DWP	Scalar Quantization	11.8	19.2	4.0
mpegtestt332.wav	DWP	Vector Quantization	17.08	14.2	3.4
mpegtest333.wav	DWP	Scalar + Vector Quantization	14.6	16.1	3.9

Table 5.30 Results of Compression with Optimisation Method 3 'else3.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
else3331.wav	DWP	Scalar Quantization	9.9	15.0	4.1
else3332.wav	DWP	Vector Quantization	13.9	13.1	3.8
else3333.wav	DWP	Scalar + Vector Quantization	12.1	14.2	3.9

Table 5.31 Results of Compression with Optimisation Method 3 'sitar111.wav'

Reconstructed Signal	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
sitar331.wav	DWP	Scalar Quantization	6.6	18.9	4.3
sitar332.wav	DWP	Vector Quantization	11.6	12.3	3.9
sitar333.wav	DWP	Scalar + Vector Quantization	7.5	16.4	4.2

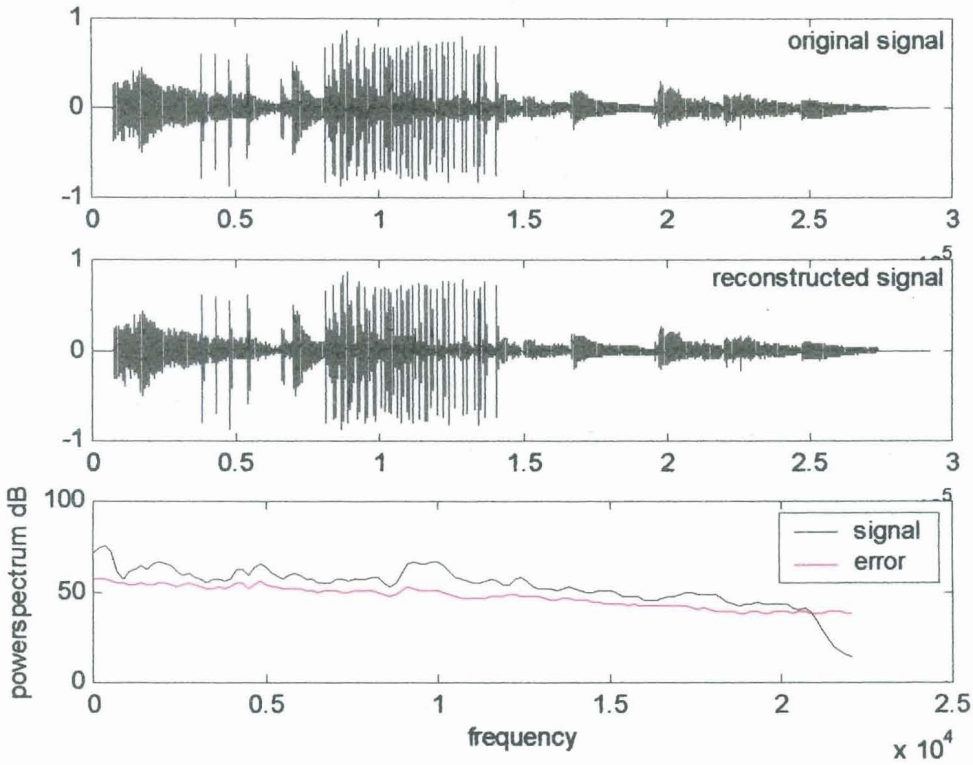


Fig.5.13: Power spectra plot: castanets - Low complexity scheme with scalar quantization -cast311.wav

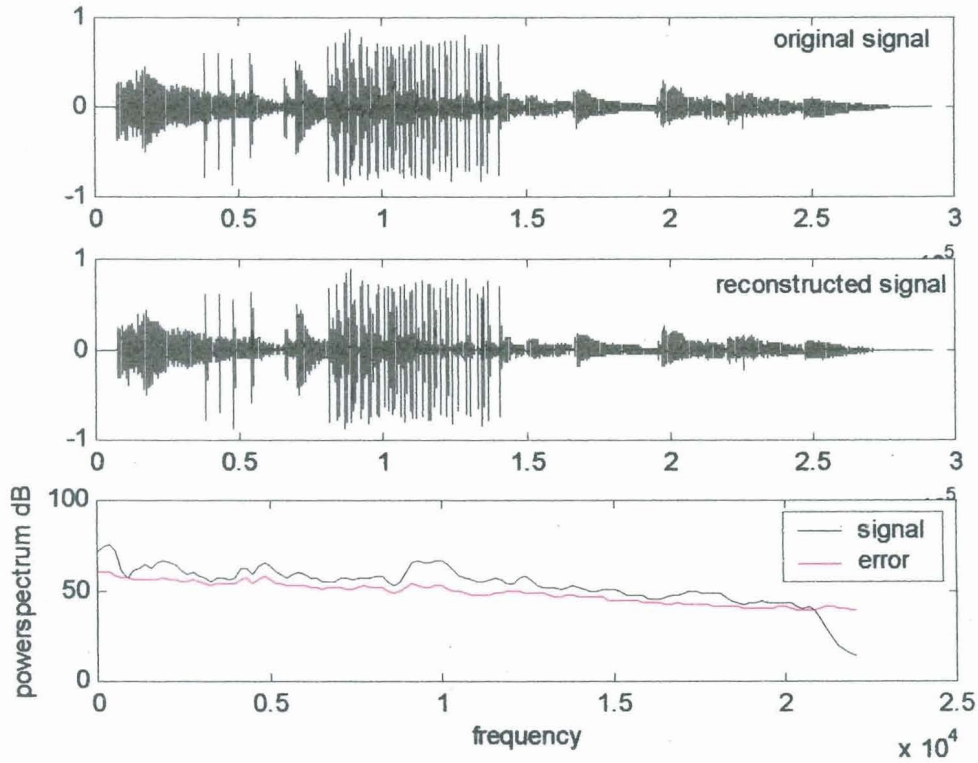


Fig. 5.14: Power spectra plot: castanets - Low complexity scheme with vector quantization - cast 312.wav

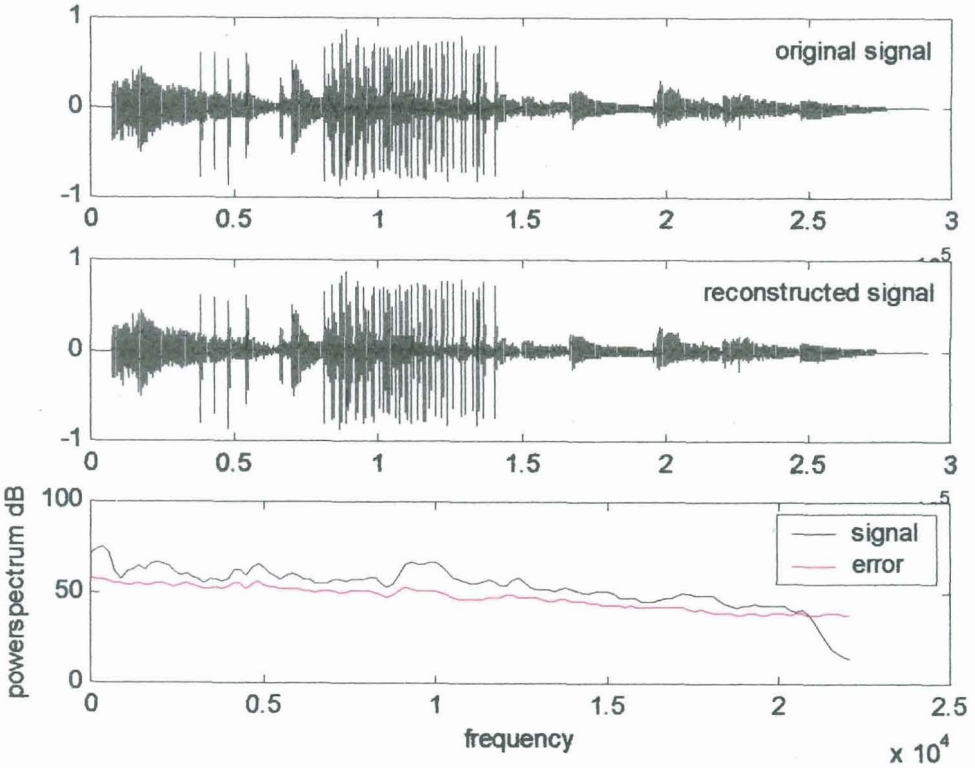


Fig. 5.15: Power spectra plot: castanets - Low complexity scheme with Scalar + VQ - cast313.wav

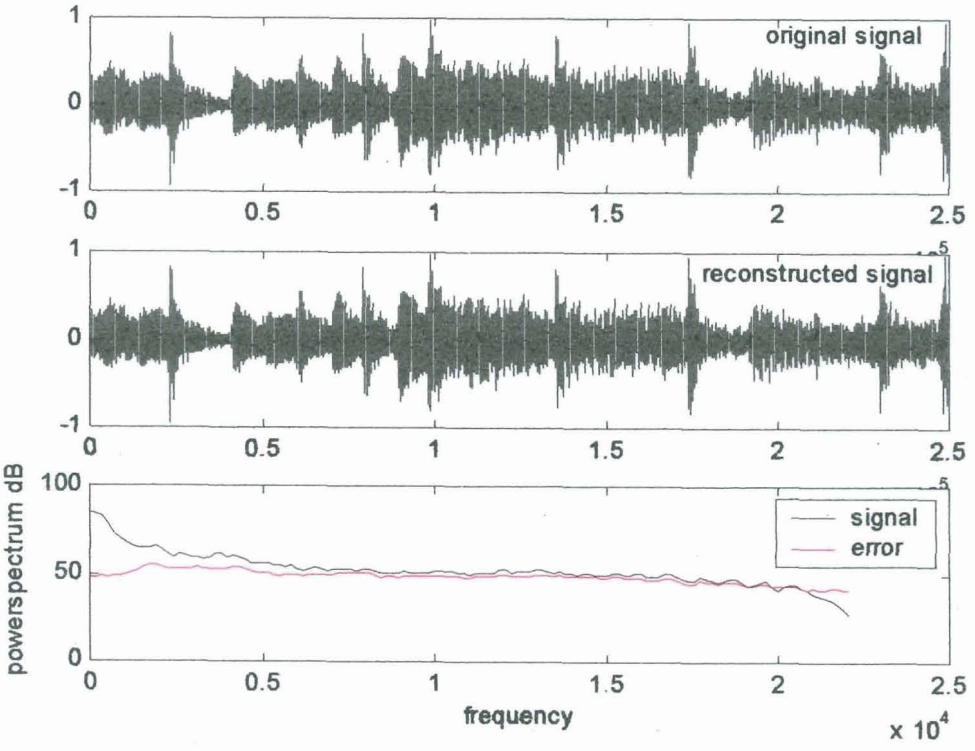


Fig.5.16: Power spectra plot: else3 - Low complexity scheme with Scalar quantization -else3311.wav

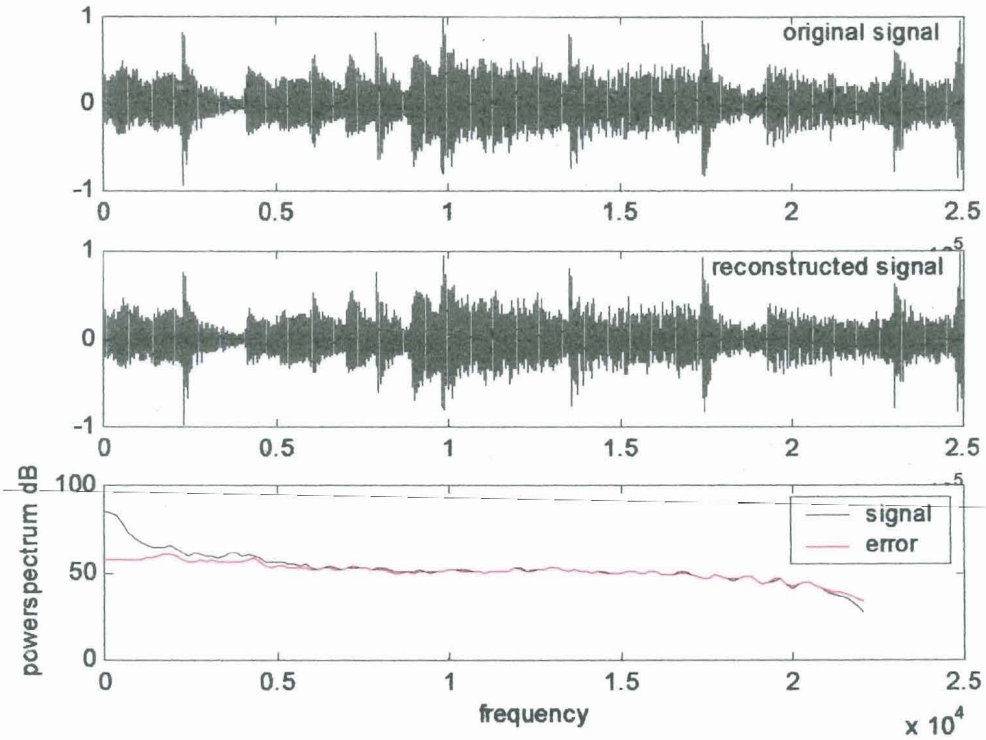


Fig. 5.17: Power spectra plot: else3 - Low complexity scheme with Vector quantization - else3312.wav

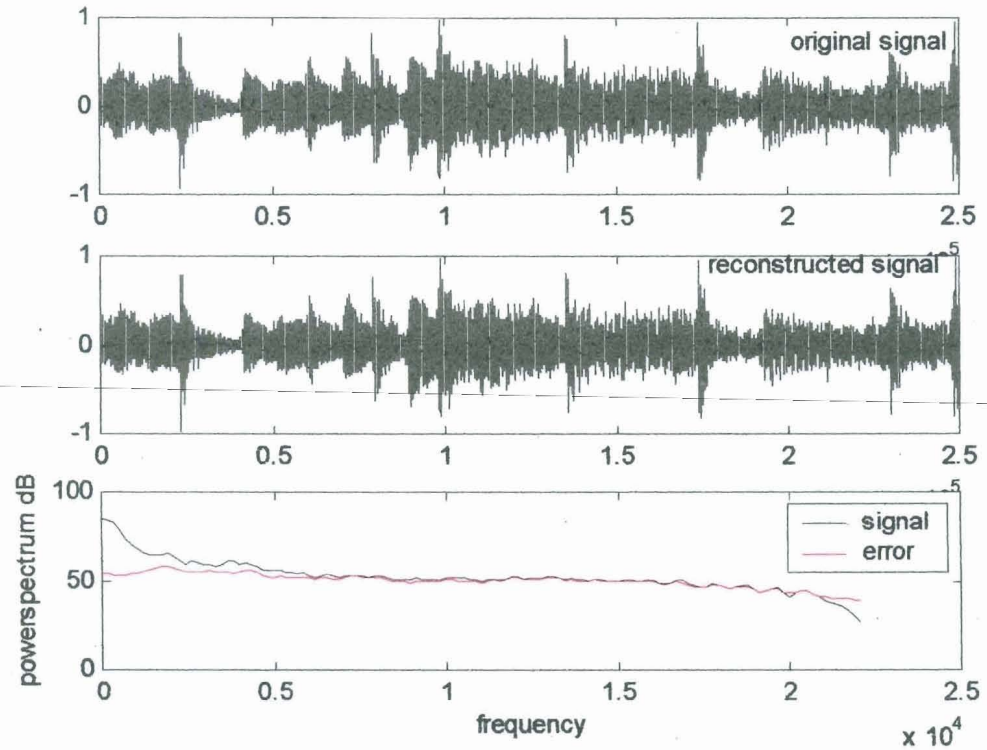


Fig. 5.18: Power spectra plot: else3 - Low complexity scheme with Scalar +VQ- else3313.wav

5.7 Implementation of the scalable perceptual audio coder (scheme 3)

Analysis filter bank: Block diagram of the proposed scheme is shown in Fig.5.19. In this scheme, a flexible wavelet packet tree structure as shown in Fig. 5.20 is designed to support three sampling frequencies as indicated. The lower and higher cut-off frequencies of each band are also illustrated. This tree structure approximately achieves the frequency resolution of the critical bands. Number of subbands are different for different sampling frequencies. The 27 subband tree structure proposed earlier is used in this coder for 44.1 kHz sampled audio. The tree structure for audio sources of other bandwidth are obtained by modifying the 27 band tree structure. For example, for 22.05 kHz sampled audio, the higher two bands representing frequency range between 11 kHz and 22 kHz are truncated, while the lower 25 bands are preserved. Similarly, for 11 kHz sampled audio, frequency bands above 5.5 kHz are dropped. This results in a tree structure with 20 subbands for 11 kHz sampled audio.

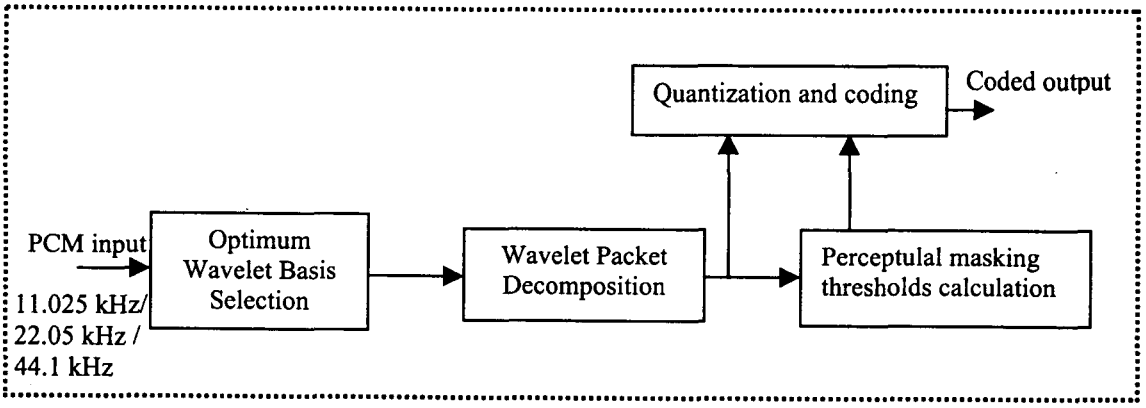


Fig. 5.19: Block diagram of scalable perceptual audio coding scheme

5.7.1 Psychoacoustic model implementation

Psychoacoustic model is driven by analysis filter bank coefficients (i.e., wavelet packet coefficients) as described earlier in section 5.6 . Hence, computational complexity is less in this scheme also. In the case of 11.025 kHz sampled signals, we need have to calculate masking thresholds for only 20 subbands and in the case of 22.05 kHz sampled signals we need have to calculate those for only 25 subbands.

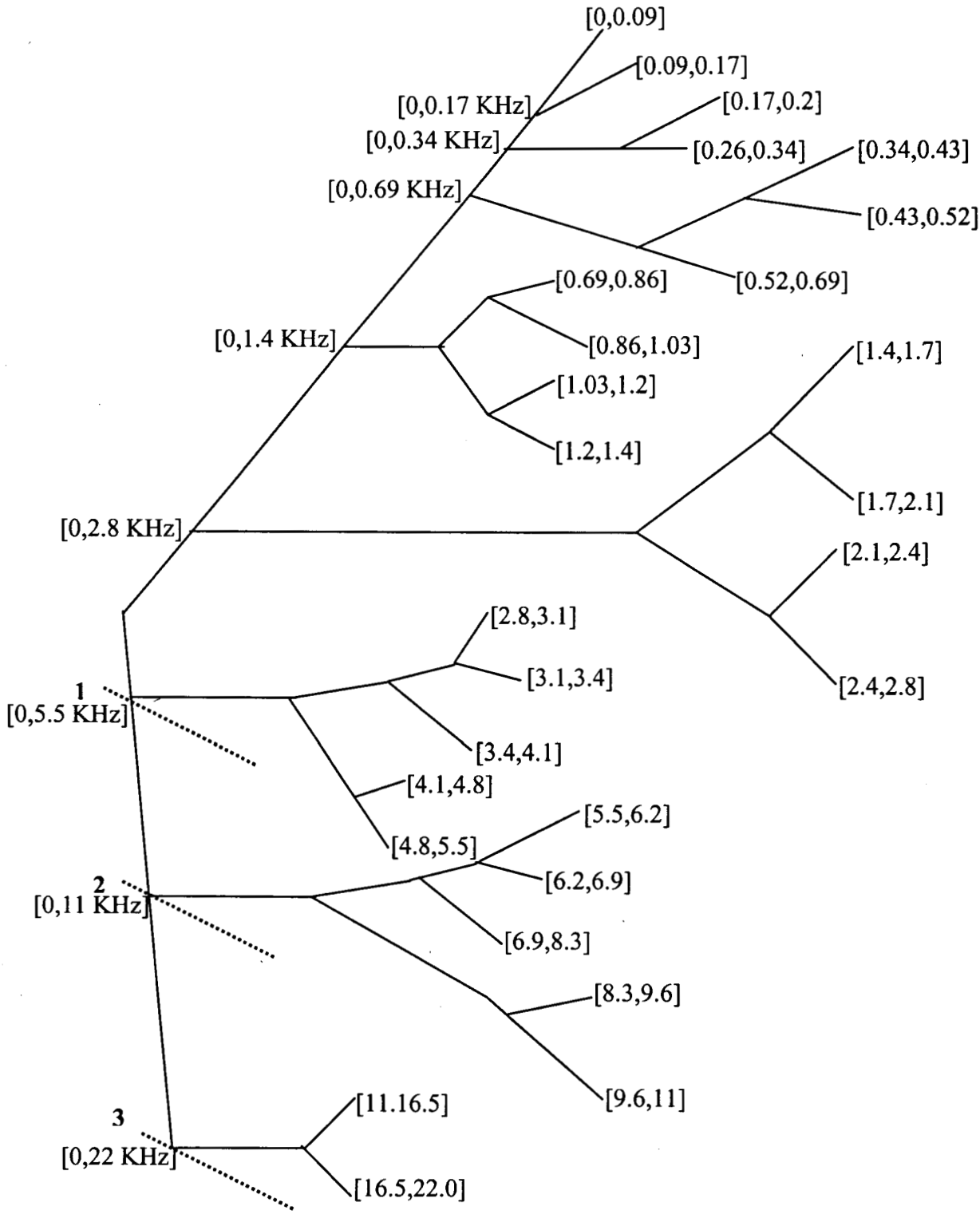


Fig.5.20: Flexible Wavelet Packet Tree structure designed to support three sampling frequencies (11.025 kHz, 22.05 kHz and 44.1 kHz).

5.8 Results

Results of implementation of the proposed scalable audio coding scheme, supporting three industrial audio sampling frequencies, are shown in Tables 5.32 -5.34 and Figs.5.21 – 5.36 for some typical audio signals with different sampling frequencies. This scheme is tested with all the quantisation schemes mentioned earlier and the three optimisation methods developed in Chapter 4. Performance of the codec is validated through subjective listening tests. Compression ratio, SSNR and MOS values are calculated for audio signals of different sampling frequencies. Some of the reconstructed signals are stored in the attached CD along with the original signals for the sake of comparison. Error power spectrum plots shown in Figs.5.21-5.36 are in agreement with the results presented in Tables 5.32-5.34. The use of the proposed flexible wavelet packet tree structure in the audio coder allows users to encode various audio signals sampled at different sampling frequencies according to their need/requirement.

Table 5.32 Performance of the scheme with Optimisation Method 2 (Sampling Frequency =11.025 kHz)

Audio Signal (.wav)	Sampling Frequency	Analysis	Quantization Scheme	Compression Ratio	SSNR (dB)	MOS
male421	11.025kHz	DWP	Scalar Quantization	6.13	17.3	4.0
female421				4.8	17.9	4.2
clap421				6.5	15.8	4.8
ring421				1.1	20.3	4.0
clarinet421				4.01	18	4.1
male422	11.025kHz	DWP	Vector Quantization	9.92	13.1	3.9
female422				6.94	14.5	4.1
clap422				7.6	15.1	4.8
ring422				1.3	15.2	3.9
clarinet422				4.7	15.6	4.1

Note: Sampling Frequency of the signals which are listed above is 11.025 kHz. Hence, maximum frequency of the signals is 5.5 kHz. Therefore all subbands are scalar quantized in the first set and vector quantized in the second set. Combined scalar and vector quantization was not done in the case of these signals sampled at 11.025 kHz since maximum frequency of the signal lies in the sensitive frequency region of the human ear.

Table 5.33 Performance of the scheme with Optimisation Method 2 (Sampling Frequency = 22.05 kHz)

Audio Signal (.wav)	Sampling Frequency	Analysis	Quantization Scheme	Compre-ssion Ratio	SSNR (dB)	MOS
pup421	22.05 kHz	DWP	Scalar Quantization	4.3	18	4.3
whistle421				1.6	14.1	4.2
drums421				11.1	22.7	4.4
female2421				7.3	22.7	4.3
crow421				5.9	17.8	4.8
pup422	22.05 kHz	DWP	Vector Quantization	6.1	13.4	4.2
whistle422				2.73	9.9	4.0
drums422				18.4	13.1	4.0
female2422				18.7	13.5	3.9
crow422				9.86	13.4	4.7
pup423	22.05 kHz	DWP	Scalar + Vector Quatization	5.05	15.8	4.3
whistle423				2.04	11.8	4.1
drums423				15.7	15.6	4.1
female2423				14.9	15.6	4.1
crow423				7.3	15.3	4.8

Table 5.34 Performance of the scheme with Optimisation Method 2 (Sampling Frequency = 44.1 kHz)

Audio Signal (.wav)	Sampling Frequency	Analysis	Quantization Scheme	Compre-ssion Ratio	SSNR (dB)	MOS
castanets421	44.1 kHz	DWP	Scalar Quantization	8.92	13.1	4.3
mpegtest421				11.4	20.2	4.0
kadal421				8.8	19.4	4.1
else3421				8.9	16.8	4.1
sitar421				5.8	21.8	4.3
castanets422	44.1 kHz	DWP	Vector Quantization	13.77	9.1	3.8
mpegtest422				17.08	14.2	3.5
kadal422				15.0	12.8	3.7
else3422				13.9	13.1	3.7
sitar422				11.9	12.2	3.7
castanets423	44.1 kHz	DWP	Scalar + Vector Quatization	10.2	11.2	4.2
mpegtest423				13.4	17.1	3.8
kadal423				10.6	16.9	4.0
else3423				12.1	14.2	4.1
sitar423				7.7	17.1	4.2

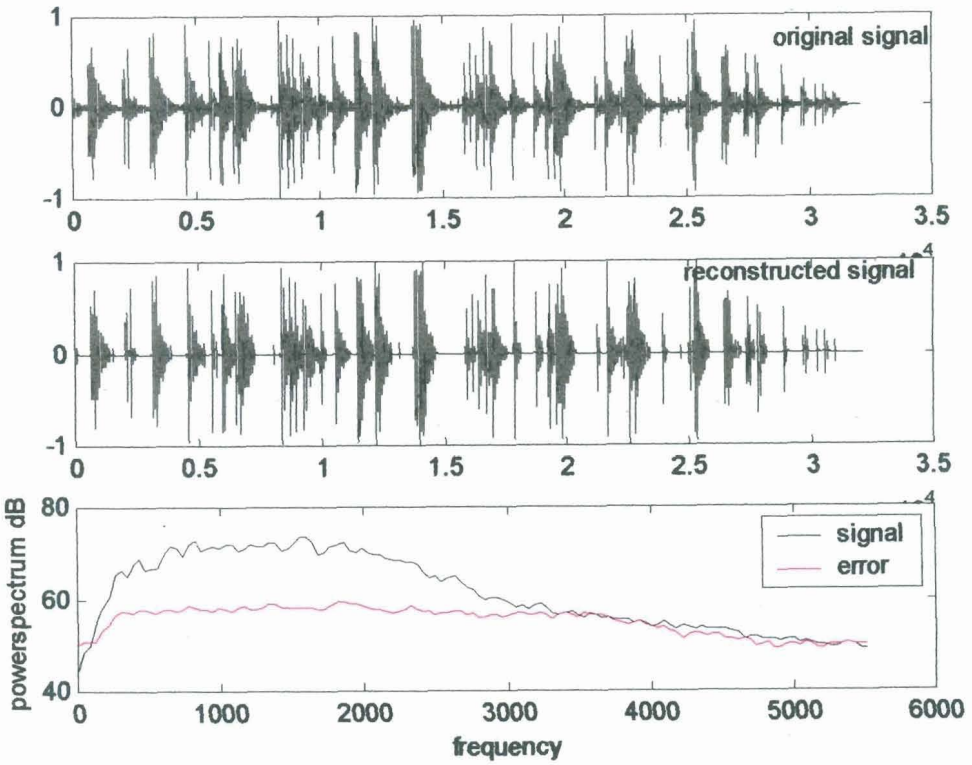


Fig.5.21: Power spectra plot –clap : Flexible DWP and scalar quantization -clap421.wav

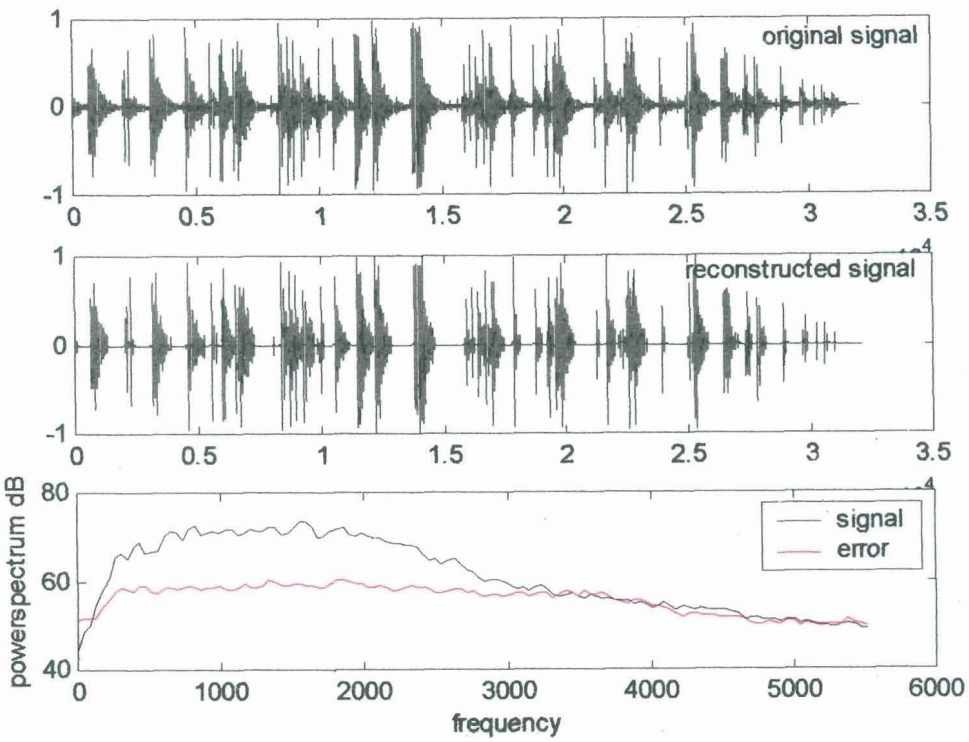


Fig. 5.22: Power spectra plot –clap : Flexible DWP and vector quantization clap422.wav

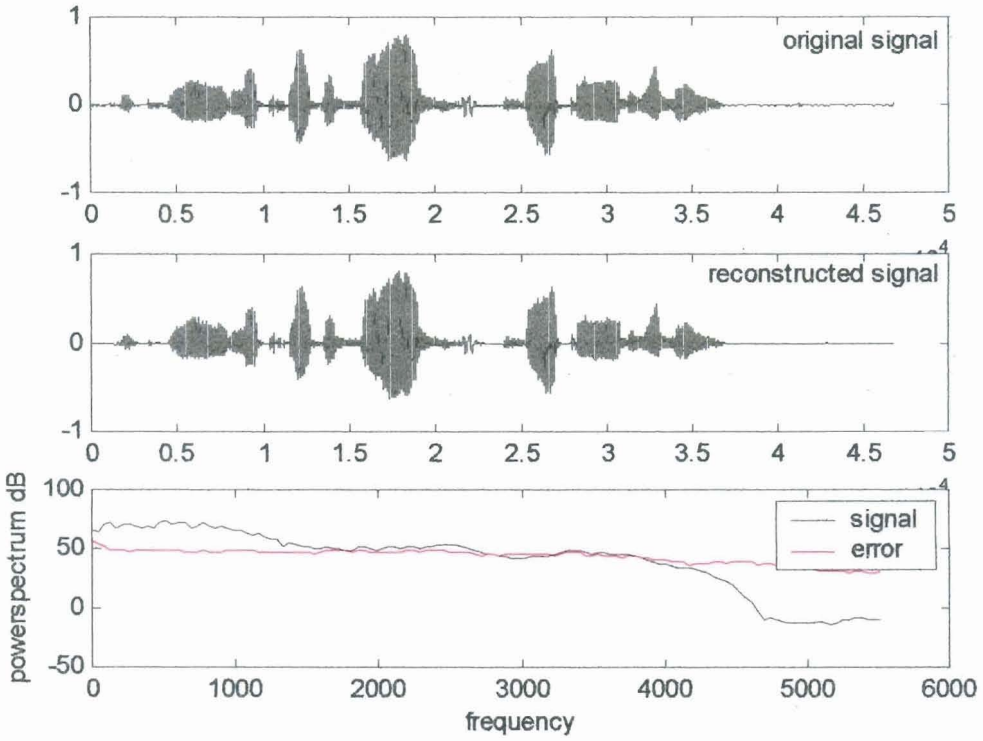


Fig.5.23: Power spectra plot –male: Flexible DWP and scalar quantization -male421.wav

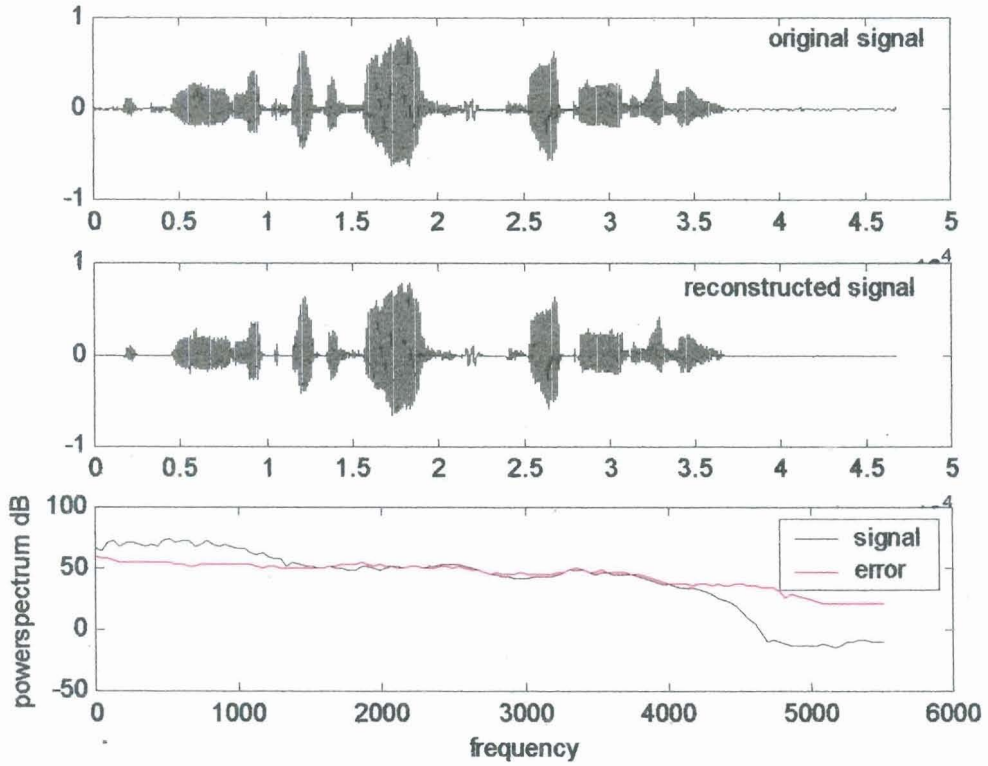


Fig.5.24: Power spectra plot-male: Flexible DWP and vector quantization-male422.wav

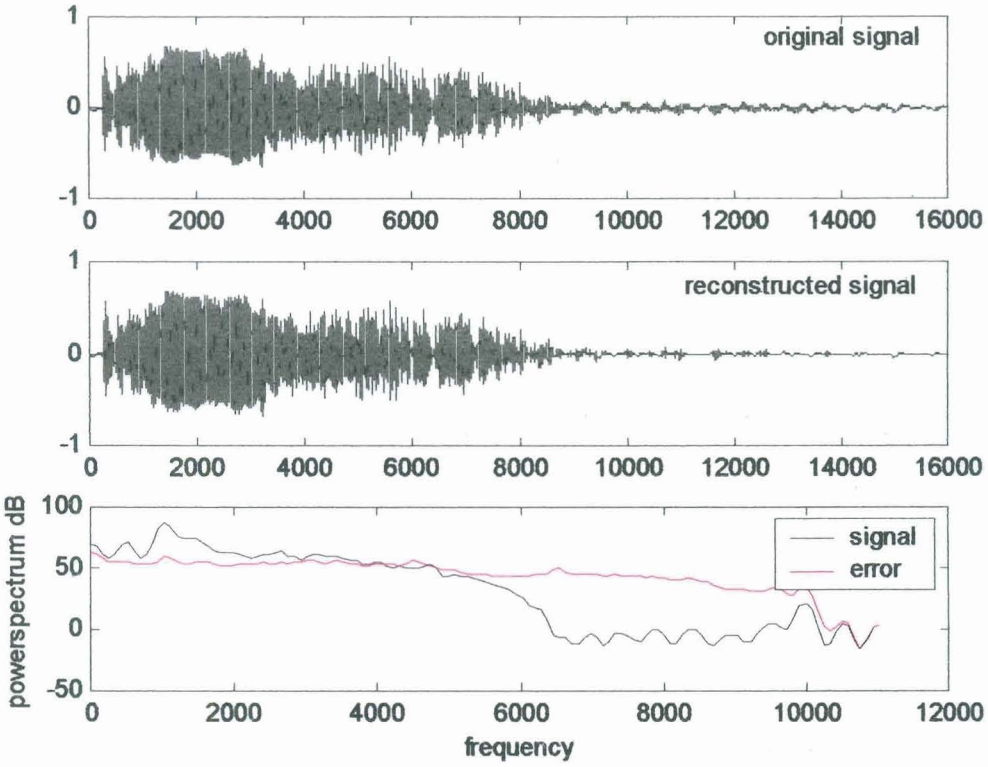


Fig. 5.25: Power spectra plot –crow: Flexible DWP and scalar quantization -crow421.wav

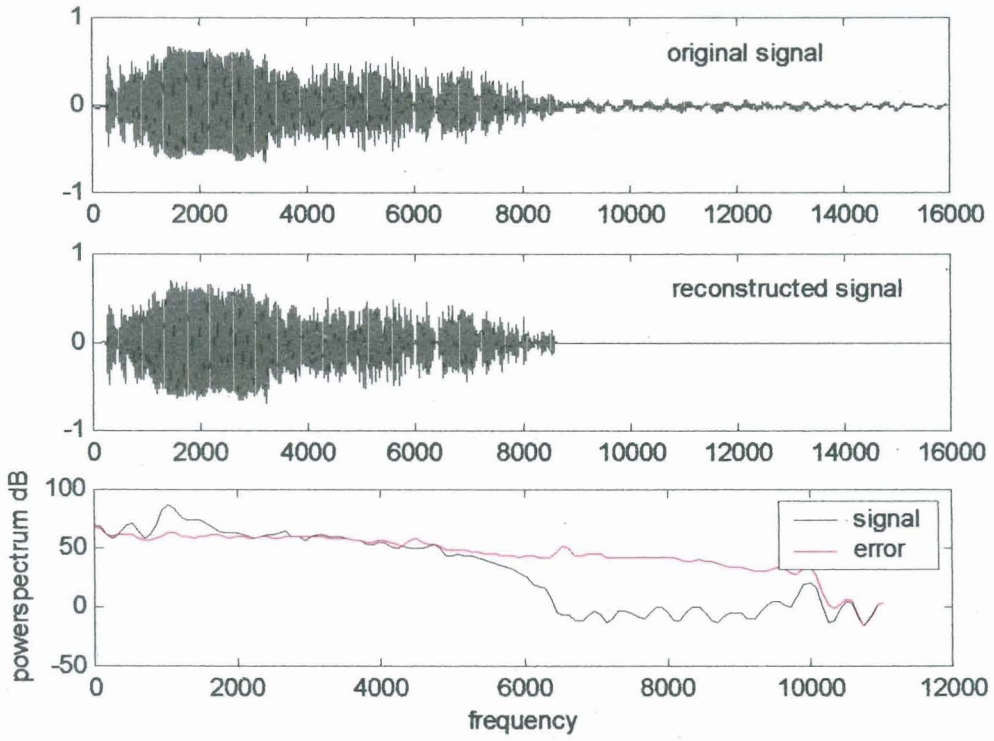


Fig.5.26: Power spectra plot –clap: Flexible DWP and vector quantization -crow422.wav

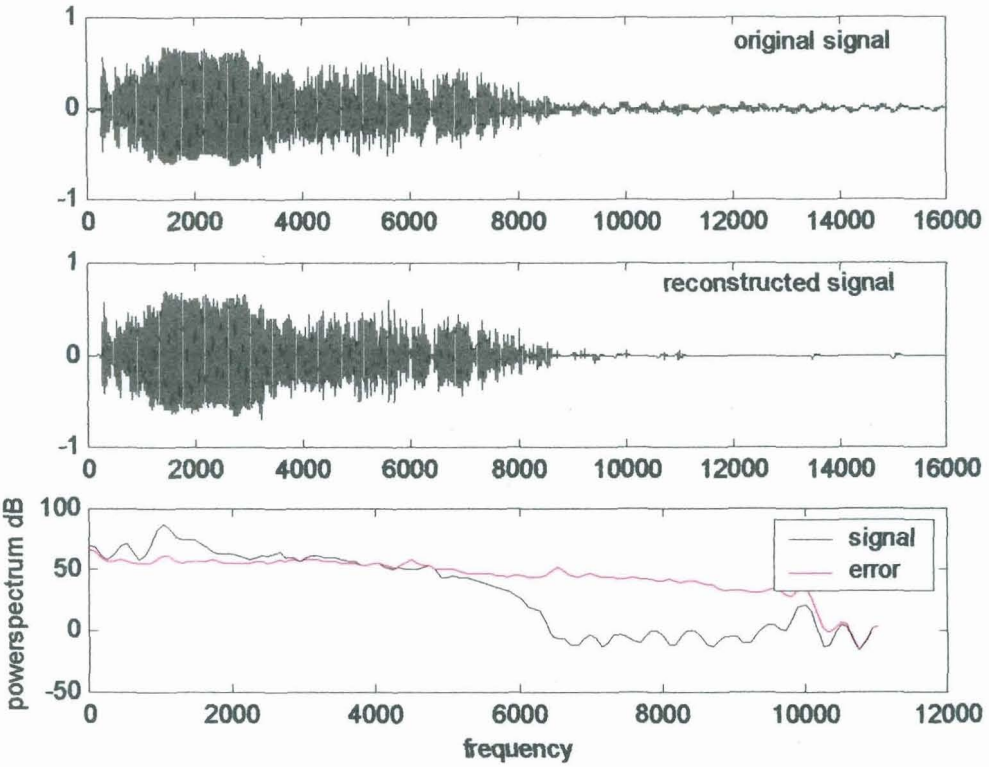


Fig.5.27: Power spectra plot –crow: Flexible DWP and Scalar +VQ -crow423.wav

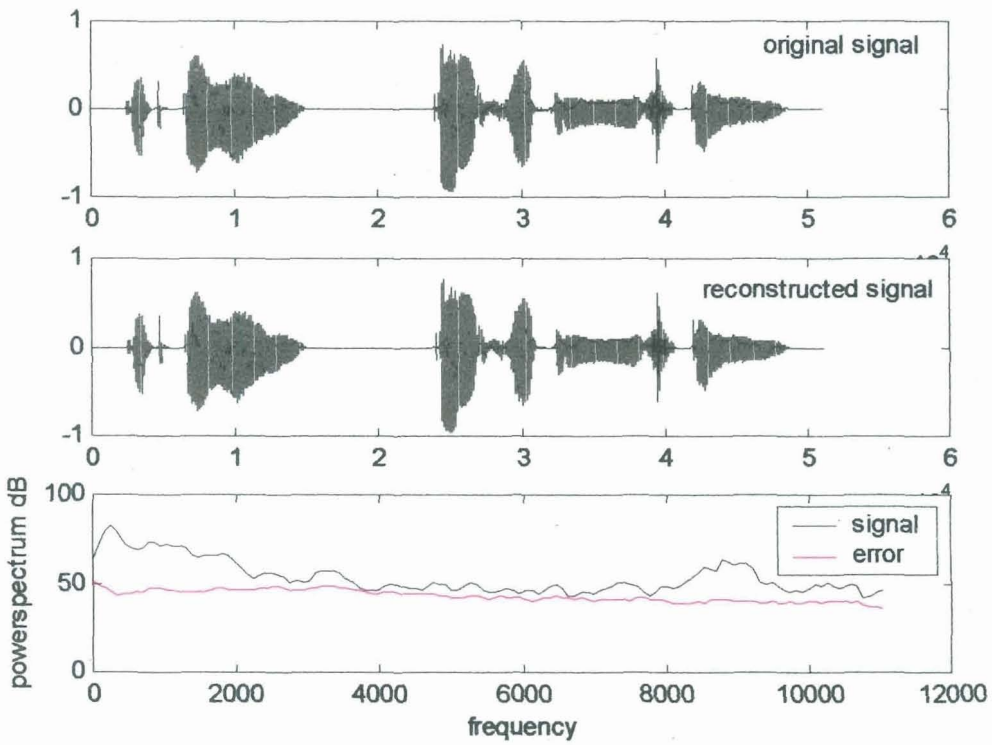


Fig. 5.28: Power spectra plot –female2:Flexible DWP and Scalar Quantization -female2421.wav

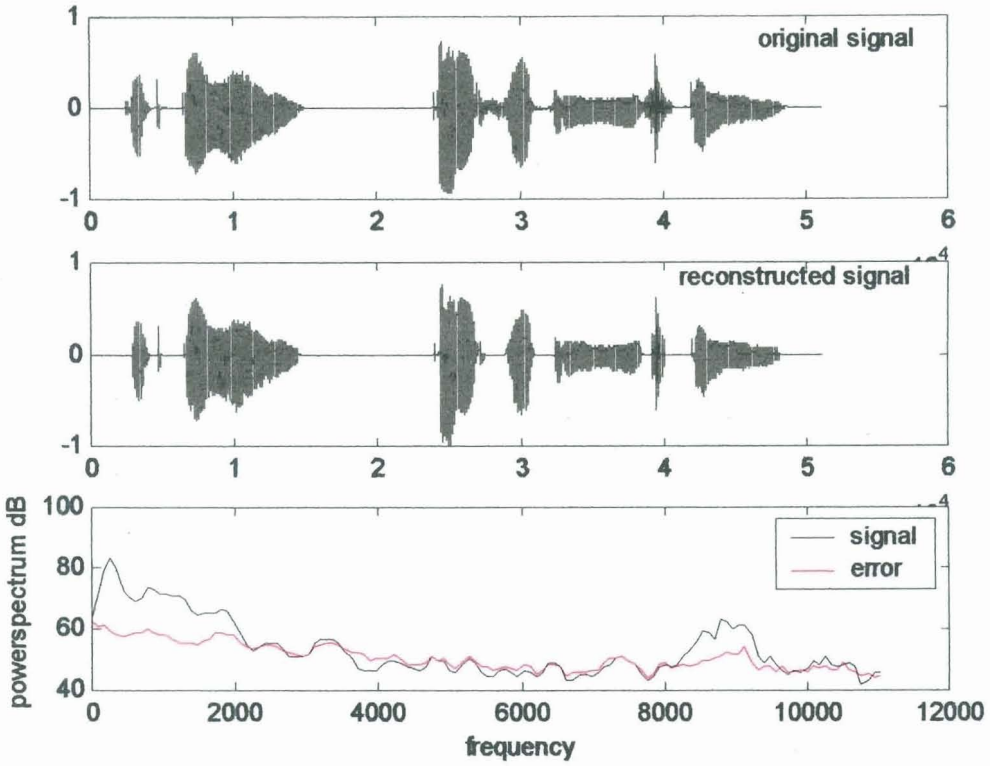


Fig.5.29: Power spectra plot –female2: Flexible DWP and vector quantization- female2422.wav

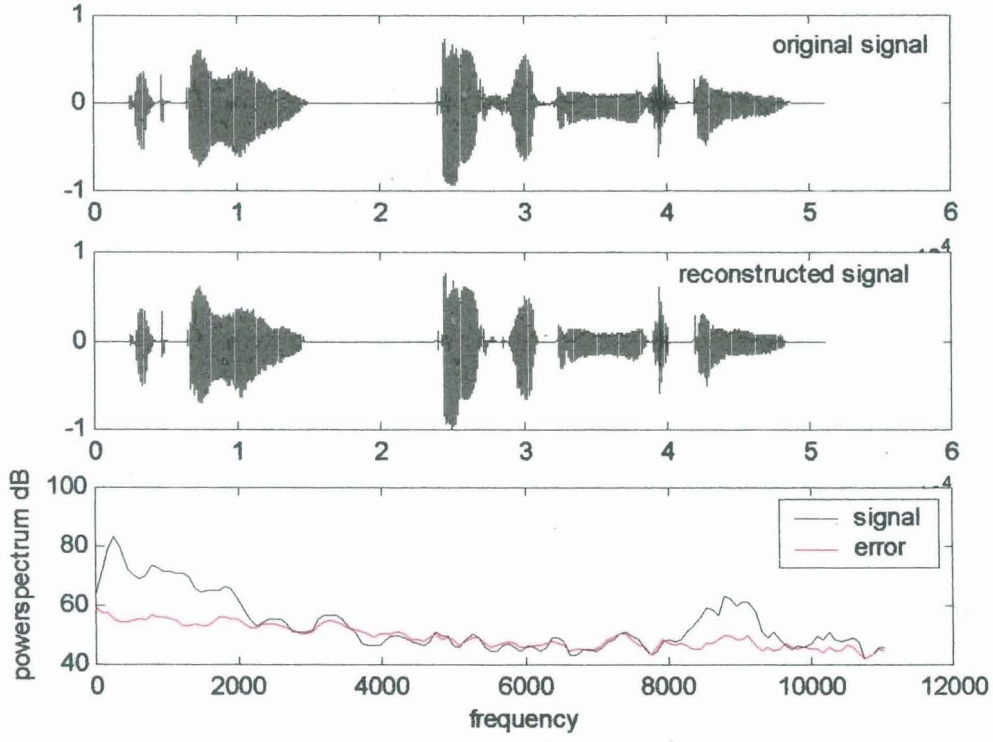


Fig.5.30: Power spectra plot –female2: Flexible DWP and Scalar +VQ -female2423.wav

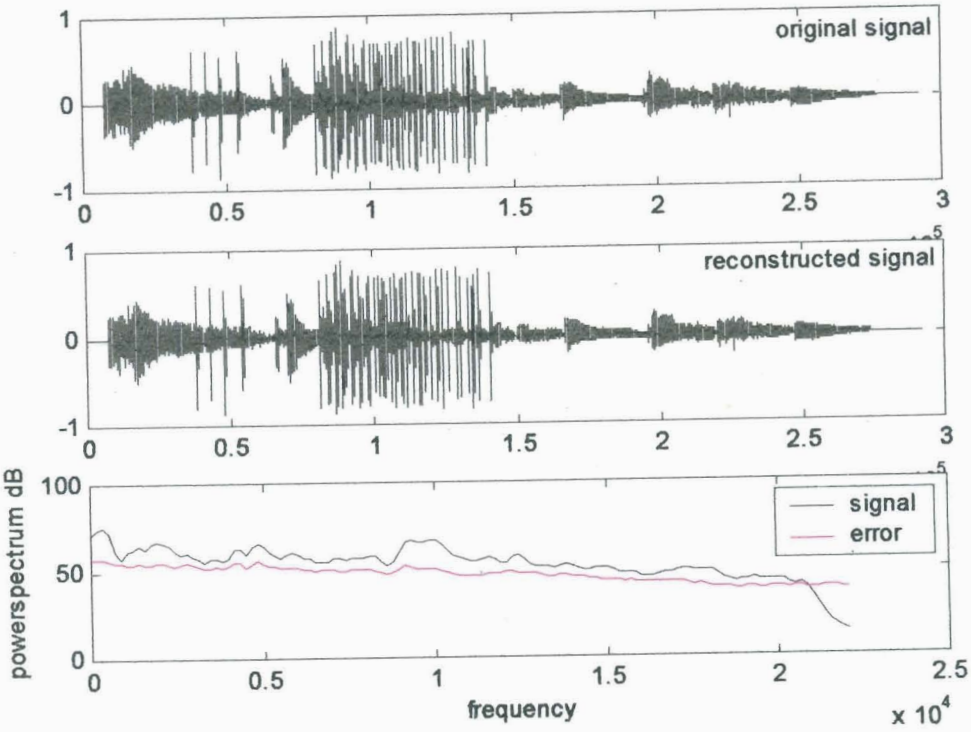


Fig.5.31: Power spectra plot –castanets: Flexible DWP and scalar quantization –cast421.wav

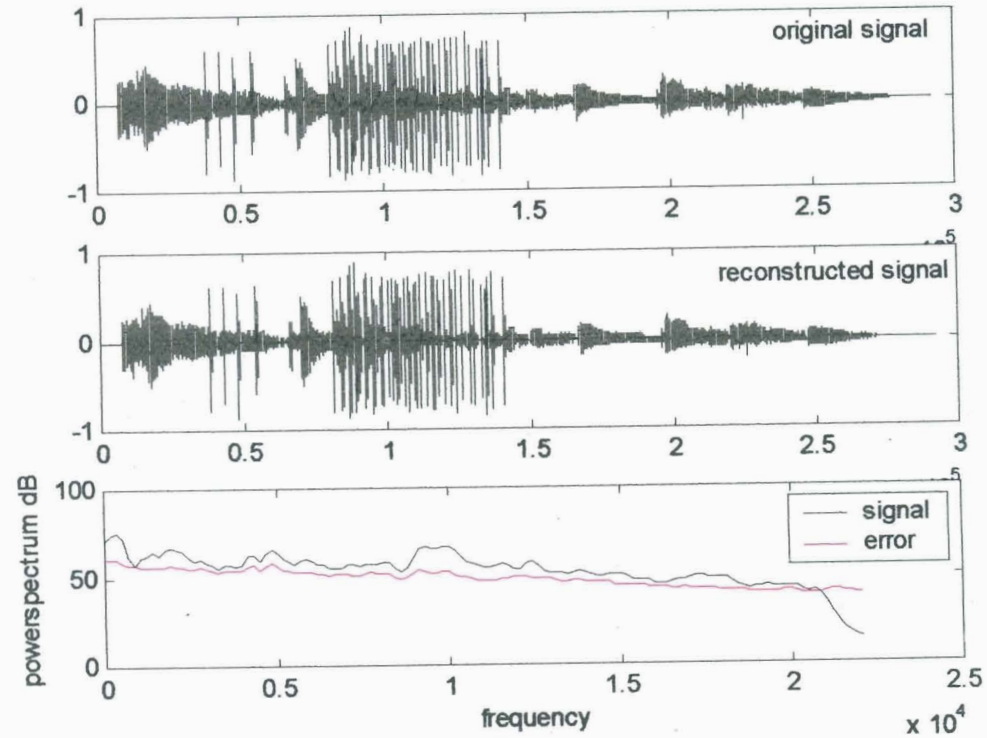


Fig.5.32: Power spectra plot-castanets:Flexible DWP and vector quantization –cast 422.wav

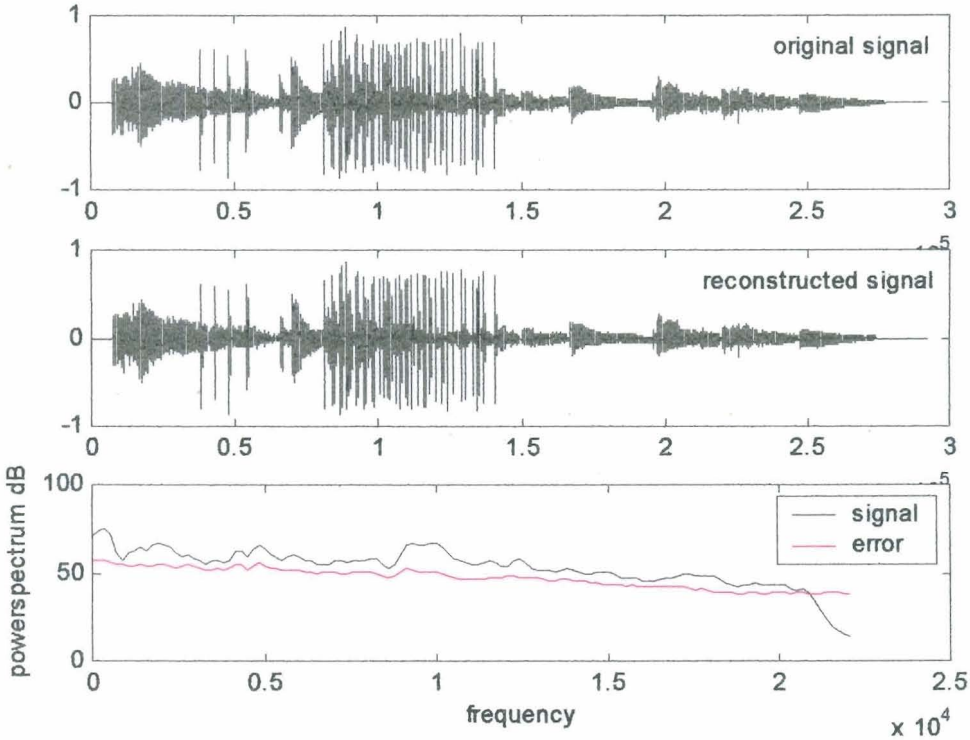


Fig.5.33: Power spectra plot –castanets : Flexible DWP and Scalar+VQ -cast423.wav

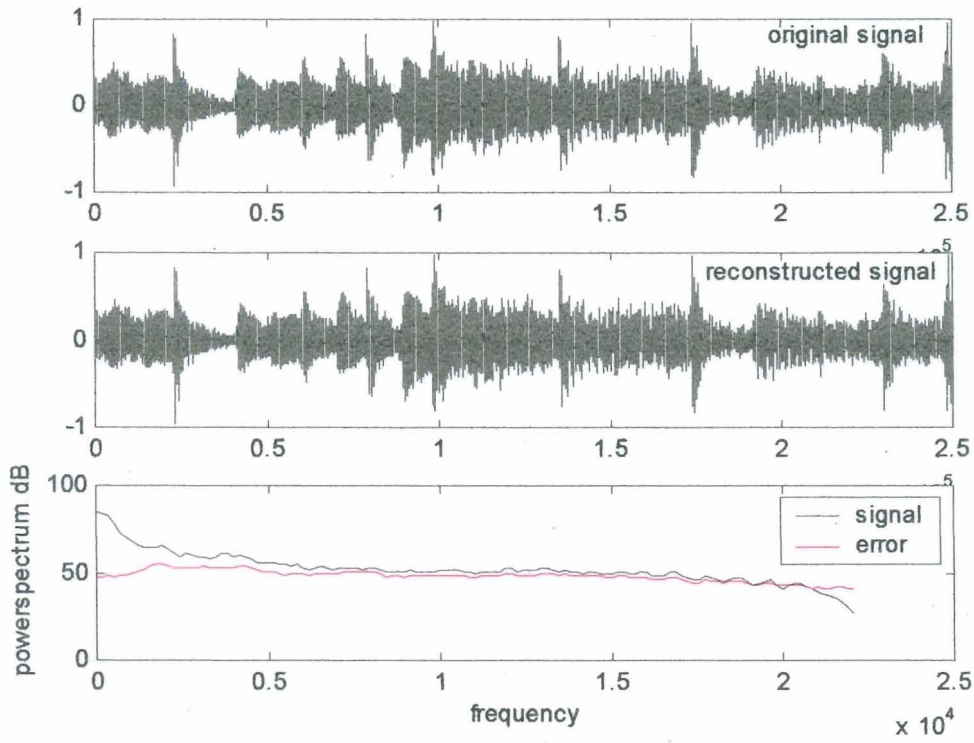


Fig.5.34: Power spectra plot –else3 : Flexible DWP and scalar quantization –else3421.wav

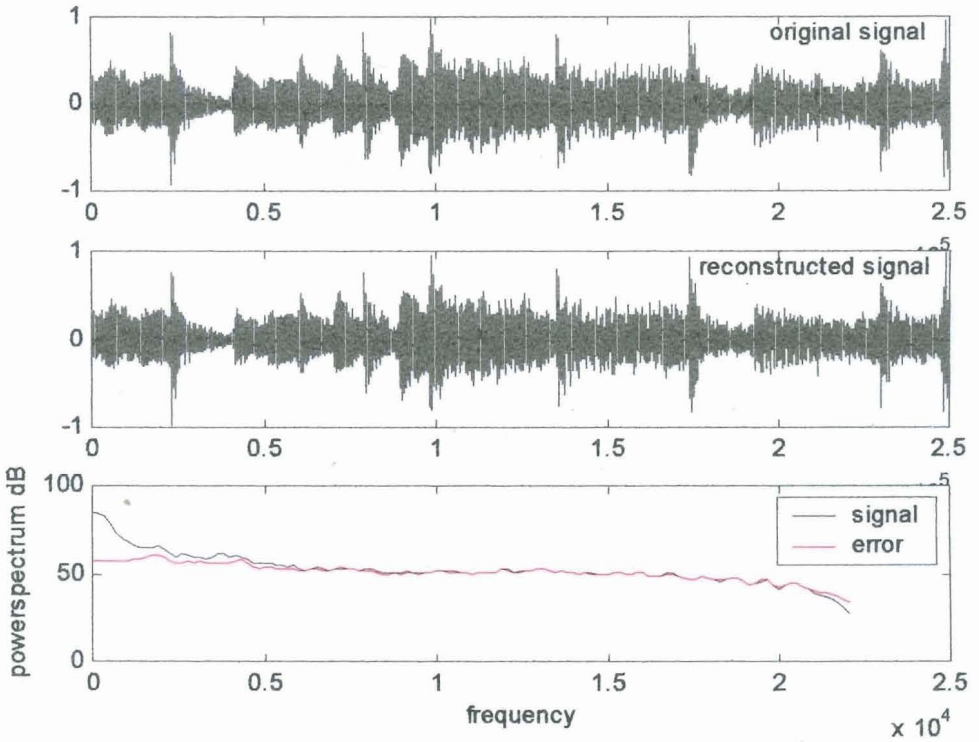


Fig.5.35: Power spectra plot –else3 : Flexible DWP and vector quantization – else3422.wav

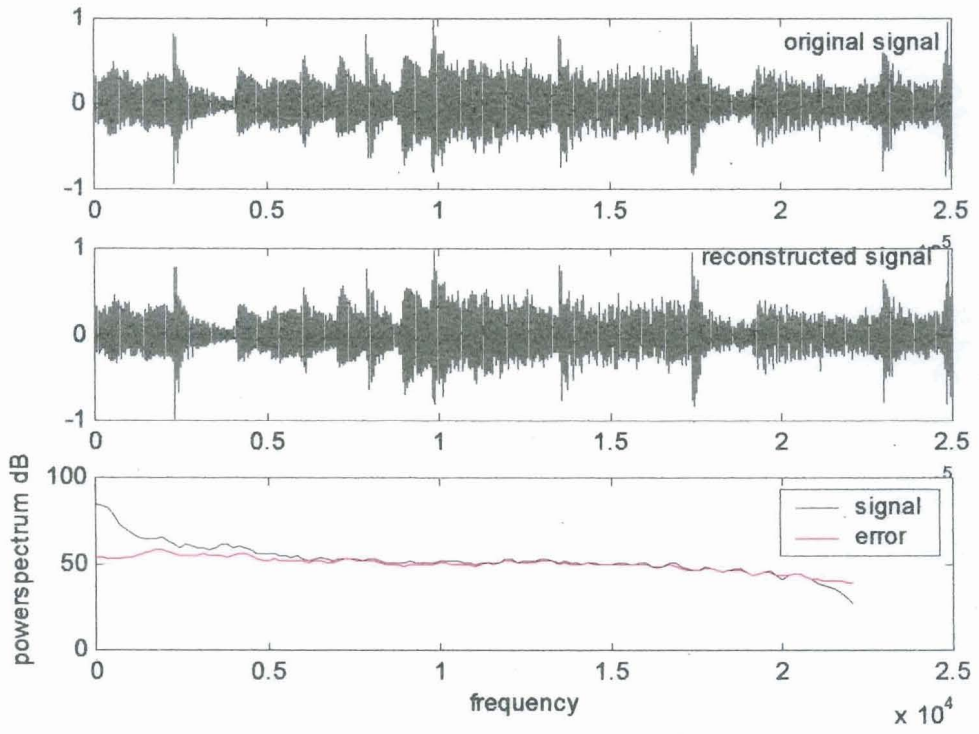


Fig.5.36: Power spectra plot –else3: Flexible DWP and Scalar + VQ– else3423.wav

5.9 Summary

Compression ratios obtained with the wavelet packet based audio coding schemes are higher than that of wavelet transform based audio coder developed in Chapter 4, at the expense of increase in computational complexity at the analysis stage. Higher compression ratios have been achieved by decomposing the signal into 27 subbands closely mimicking the human auditory system such that maximum quantization noise could be effectively introduced into the various subbands without affecting the perceptual quality. Perceptual quality (measured by MOS values) obtained with the first scheme is same as that of DWT based coder in the case of most of the audio signals tested. Main drawback of MPEG international audio coding standard, DWT based scheme and the first scheme using DWP proposed in this chapter, is the use of a separate high resolution FFT stage for the psychoacoustic model implementation. To implement these codecs in real time, two digital signal processors will be required which will increase the overall cost of the codec. Hence, a low complexity (computationally more efficient) wavelet packet based audio coder (with a slight reduction in MOS values in the case of some audio signals) has been developed and implemented as scheme 2.

In the second scheme, computational complexity is reduced by using the same wavelet packet tree structure for analysis filter bank as well as for psychoacoustic model implementation. That is, wavelet packet coefficients from the analysis filterbank directly drives the psychoacoustic model. Compression ratio obtained is same as that of first scheme using wavelet packets, in the case of most of the audio signals tested. MOS values obtained with scheme 2 are slightly less than that with scheme 1, in the case of some audio signals. This is due to the use of wavelet packet transform (whose frequency resolution is less than that of FFT) for the implementation of psychoacoustic model and the absence of the switching scheme to switch between sine/wavelet basis. It may be noted that switching scheme cannot be incorporated in this coder, since the analysis filterbank directly drives the psychoacoustic model.

DWT based audio coder and DWP based audio coders (schemes 1 and 2) use analysis filterbank suitable for CD sampling frequency (44.1 kHz) and these codecs are mainly suitable for encoding hi-fi music signals. Hence, a scalable perceptual audio coding scheme (scheme3) suitable for most of the industrial audio sampling frequencies, using a flexible wavelet packet tree structure has also been developed in this chapter. Third scheme supports three different sampling frequencies and hence, is suitable for an advanced user.

All the schemes developed in this thesis so far are variable bit rate coders and bit rate will meet the psychoacoustic model requirement. If the available channel bandwidth is less than the bandwidth required by the psychoacoustic model, then bit rate is to be again reduced (or compression is to be increased). If the channel bandwidth available is more than the bandwidth required by psychoacoustic model, more bits can be allocated to various subbands or ancillary data can be sent along with the audio signal, if constant bit rate transmission is preferred. Hence, a constant bit rate audio coder using discrete wavelet packets is developed and implemented in the next chapter.

Chapter 6

CONSTANT BIT RATE WAVELET PACKET BASED AUDIO CODER

6.1 Introduction

Perceptual audio coding schemes using discrete wavelet transforms and discrete wavelet packets were discussed in Chapters 4 and 5. All these schemes were variable bit rate audio coders. A constant bit rate discrete wavelet packet based audio coder is designed and implemented in this chapter. Signal-to-Noise Ratio (SNR) scalabilty feature is also added to this constant bit rate encoding scheme by employing two stages of compression at the encoder.

6.2 Implementation

Each frame of the audio signal is decomposed into 27 subbands using wavelet packet analysis. Depending upon the sampling frequency of the audio signal, number of subbands are selected. Perceptual masking thresholds are calculated for the various subbands. Quantization and coding of subband samples are done such that quantization noise is less than the perceptual masking threshold in each subband. The bit rate or compression ratio is then calculated. If the total number of bits required to meet the psychoacoustic model requirement exceeds the budget, number of bits are reduced by one, in each subband, starting from the band with the highest frequency. This is repeated until the bit rate becomes less than or equal to the budget. On the other hand, if the total number of bits required is less than the budget, the number of bits allocated to each band are increased by one (instead of this, ancillary data can be sent), starting from the band with the lowest frequency. This is repeated until the budget is reached. This method of allocating bits to various subbands is justifiable because our ears are less sensitive to high frequency components.

6.3 Results

Performance of the proposed audio coding scheme is evaluated for various audio signals with different sampling frequencies using three optimisation methods and various quantization schemes. The results of implementation are shown in Tables 6.1-6.4 and Figs. 6.1- 6.15 for four different compression ratios namely, 6, 10 ,15 and 25. From these tables it can be seen that for any signal, as the compression ratio increases, SSNR decreases and MOS value either decreases or remains same. The information on masked quantization noise can be observed from the error power spectra plots (Figs. 6.1-6.15) for various compression ratios. At higher compression ratios , MOS values are less than 3.0 in the case of some audio signals.

The results presented in the Tables 6.1-6.4 reveal that perceptual quality (measured by MOS values) vary from signal to signal eventhough the compression ratio remains same. Hence, in the case of constant bit rate audio coding scheme, perceptual quality mainly depends on the characteristics of the encoded signal. For example, when the compression ratio=25, MOS value obtained for the signal 'clap' is 3.8 while its values are 1.0 and 2.0 for the signals 'ring' and 'clarinet', respectively. Original signal, reconstructed signals with various compression ratios and error power spectra plots for some typical audio signals are shown in Figs.6.1-6.15. These error power spectra plots obtained are in close agreement with the results shown in Tables 6.1-6.4. Some of the reconstructed audio signals are provided in the attached CD along with the original signals for the purpose of comparison. Signal to noise ratio scalability feature is also incorporated into the above scheme and its implementation details are discussed in the Section 6.4.

Table 6.1 Performance of the Constant Bit rate DWP coder
(Compression Ratio = 6, Optimisation Method 3, Scalar Quantization)

Audio Signal (.wav)	Sampling Frequency (kHz)	SSNR (dB)	MOS
male5316	11.025	18.3	4.1
female5316	11.025	15	4.2
clap5316	11.025	16	4.8
ring5316	11.025	3.7	1.0
clarinet5316	11.025	11.3	3.5
pup5316	22.05	13.2	4.1
whistle5316	22.05	4.97	2.8
drums5316	22.05	41.2	4.5
female25316	22.05	24.1	4.3
crow5316	22.05	16.5	4.7
cast5316	44.1	14.7	4.6
mpegtest5316	44.1	24.7	4.6
kadal5316	44.1	22.5	4.2
else5316	44.1	18.4	4.4
sitar5316	44.1	18.5	4.2

Table 6.2 Performance of the Constant Bit rate DWP coder
(Compression Ratio = 10, Optimisation Method 3, Scalar Quantization)

Audio signal (.wav)	Sampling Frequency (kHz)	SSNR (dB)	MOS
male53110	11.025	11.5	4.0
female53110	11.025	13	4.2
clap53110	11.025	14.3	4.8
ring53110	11.025	3.2	1.0
clarinet53110	11.025	8.3	3.2
pup53110	22.05	9.8	4.1
whistle53110	22.05	3.8	2.7
drums53110	22.05	30.1	4.3
female253110	22.05	20.5	4.1
crow53110	22.05	14.7	4.3
cast53110	44.1	12.73	4.2
mpegtest53110	44.1	23.2	4.6
kadal53110	44.1	17.2	4.0
else53110	44.1	15.1	4.4
sitar53110	44.1	13.5	4.2

Table 6.3 Performance of the Constant Bit rate DWP coder
(Compression Ratio = 15, Optimisation Method 3, Scalar Quantization)

Audio signal (.wav)	Sampling Frequency (kHz)	SSNR (dB)	MOS
male53115	11.025	8.9	3.2
female53115	11.025	9.5	3.6
clap53115	11.025	10.1	4.4
ring53115	11.025	2.4	1.0
clarinet53115	11.025	6.8	2.9
pup53115	22.05	6.2	3.9
whistle53115	22.05	2.5	1.9
drums53115	22.05	22.5	4.1
female253115	22.05	16.5	3.8
crow53115	22.05	11	4.1
cast53115	44.1	9.4	3.9
mpegtest53115	44.1	13.4	3.5
kadal53115	44.1	12.8	3.8
else53115	44.1	13.7	4.0
sitar53115	44.1	11	4.1

Table 6.4 Performance of the Constant Bit rate DWP coder (Compression Ratio = 25, Optimisation Method 3, Scalar Quantization)

Audio signal (.wav)	Sampling frequency (kHz)	SSNR (dB)	MOS
male53125	11.025	6.6	2.8
female53125	11.025	8.1	3.0
clap53125	11.025	7.7	3.6
ring53125	11.025	1.22	1.0
clarinet53125	11.025	2.5	2.0
pup53125	22.05	3.0	2.8
whistle53125	22.05	1.0	1.3
drums53125	22.05	11.2	3.7
female253125	22.05	9.06	3.0
crow53125	22.05	6.1	3.7
cast53125	44.1	5.4	3.1
mpegtest53125	44.1	9.3	2.9
kadal53125	44.1	6.9	3.0
else53125	44.1	6.7	3.1
sitar53125	44.1	5.4	3.2

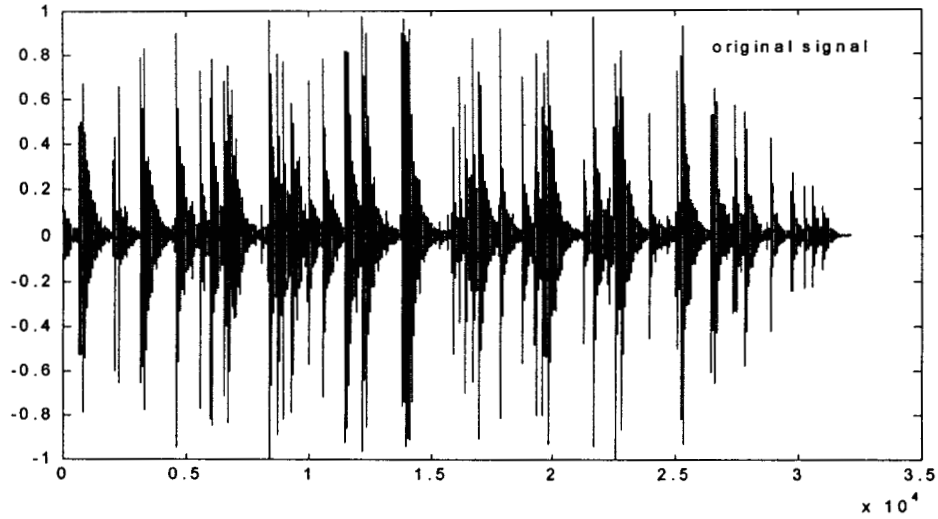


Fig. 6.1: (a) Original signal ('clap')

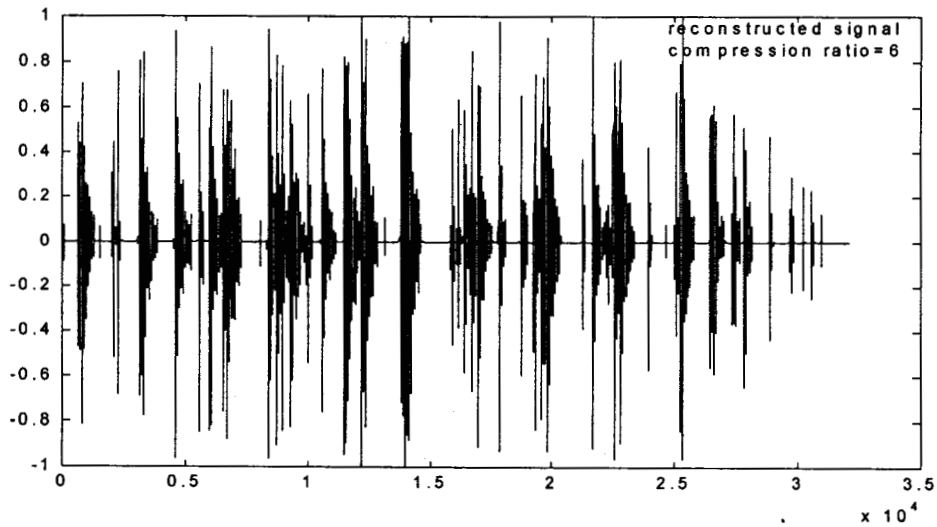


Fig. 6.1: (b) Reconstructed signal (CR=6)

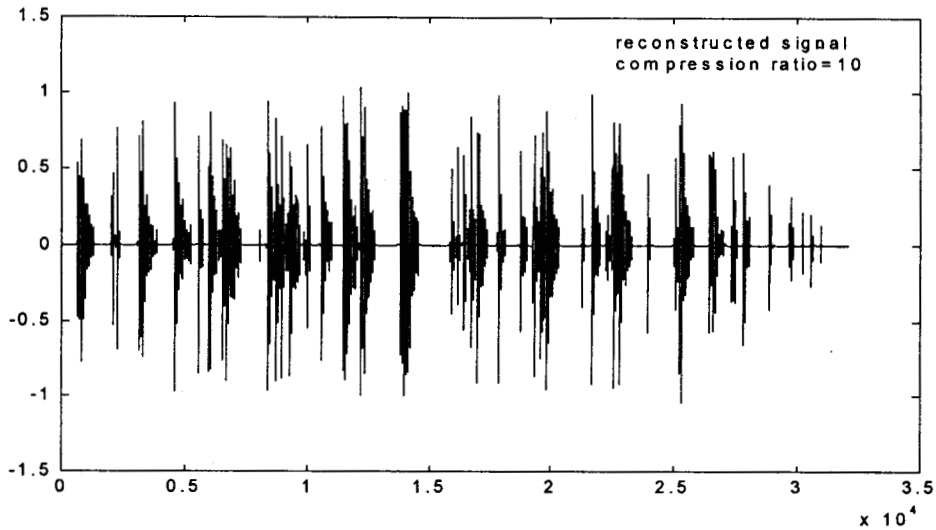


Fig. 6.1: (c) Reconstructed signal (CR=10)

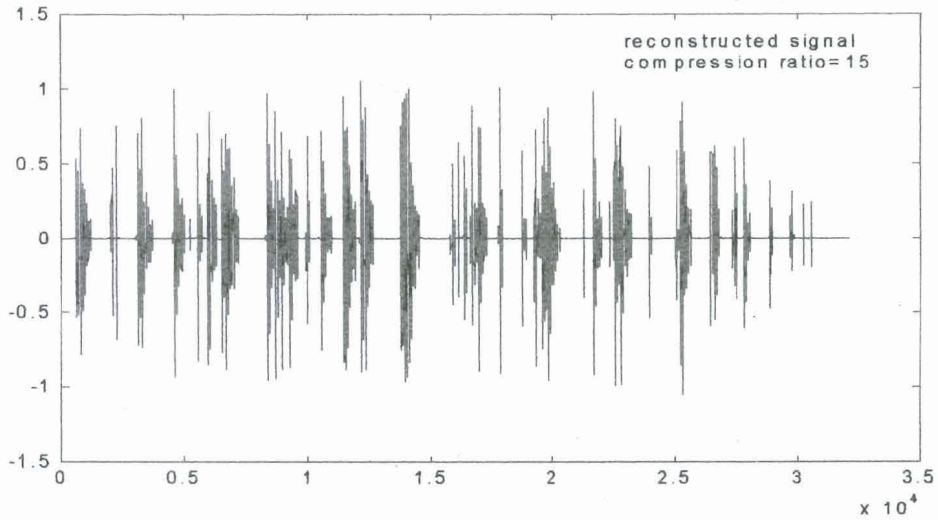


Fig. 6.1: (d) Reconstructed signal (CR=15)

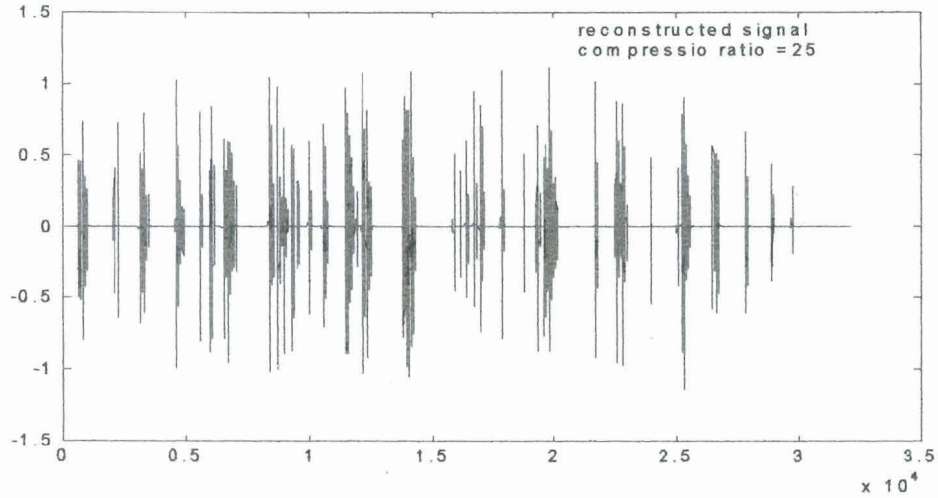


Fig. 6.1: (e) Reconstructed signal (CR=25)

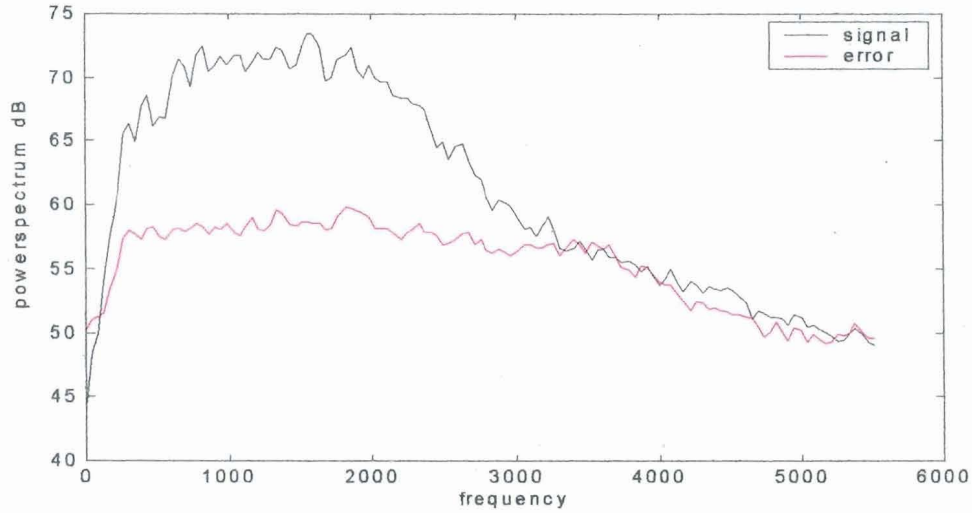


Fig. 6.2: Power spectra plot (clap, CR=6)

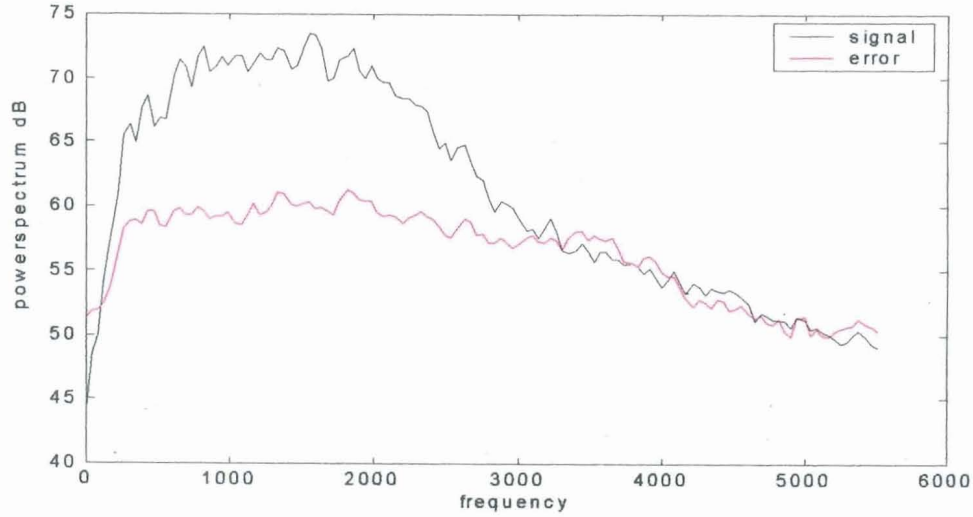


Fig. 6.3: Power spectra plot (clap, CR=10)

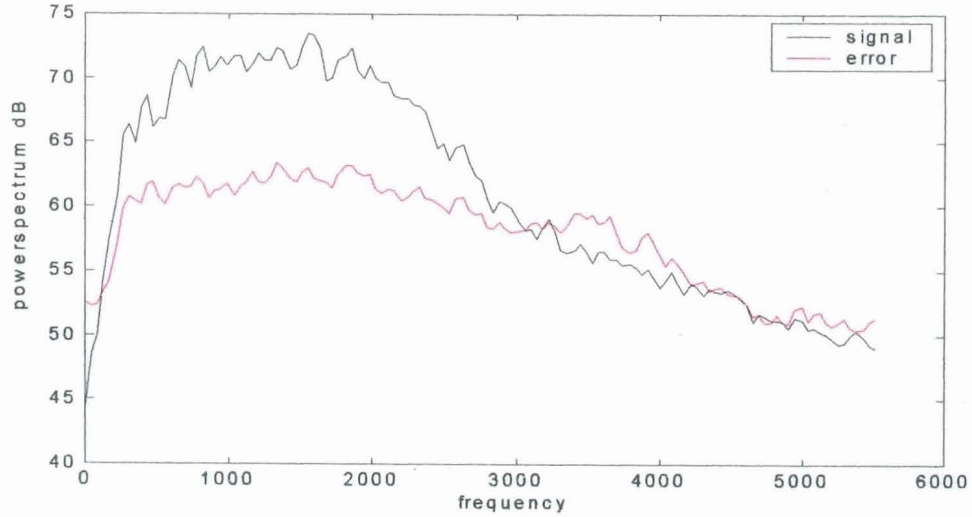


Fig. 6.4: Power spectra plot (clap, CR=15)

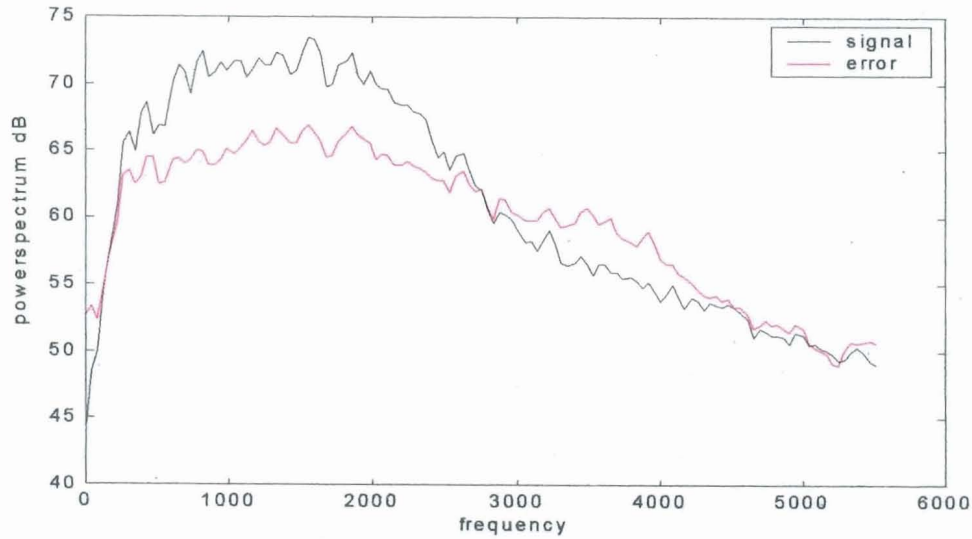


Fig. 6.5: Power spectra plot (clap, CR=25)

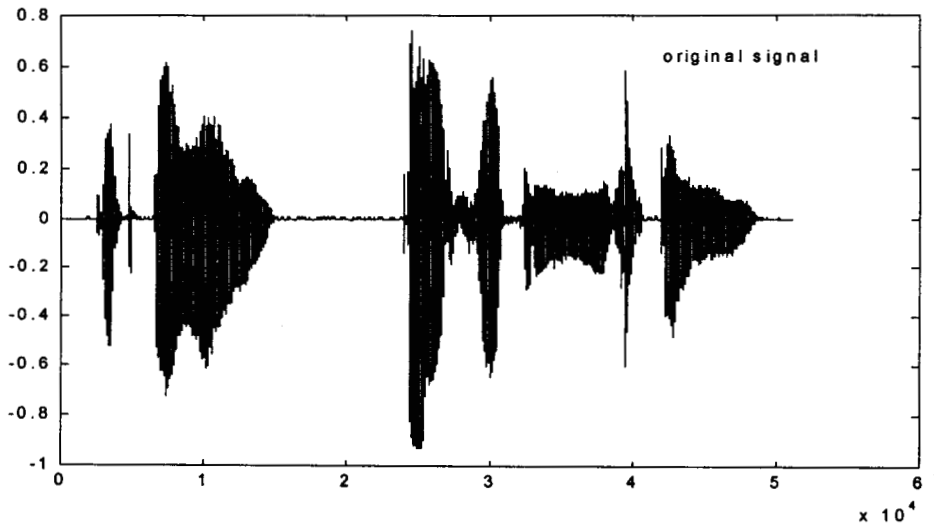


Fig.6.6: (a) Original signal ('female2')

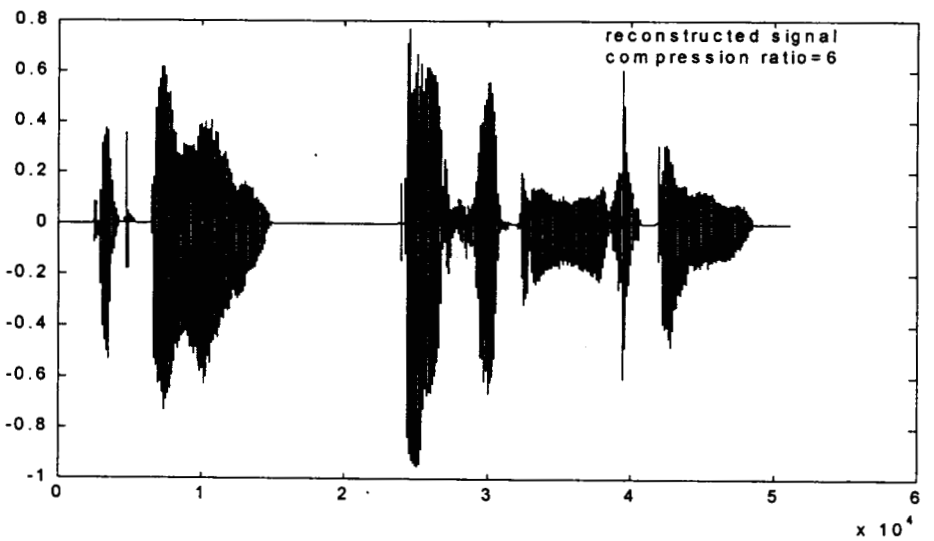


Fig.6.6: (b) Reconstructed signal (CR=6)

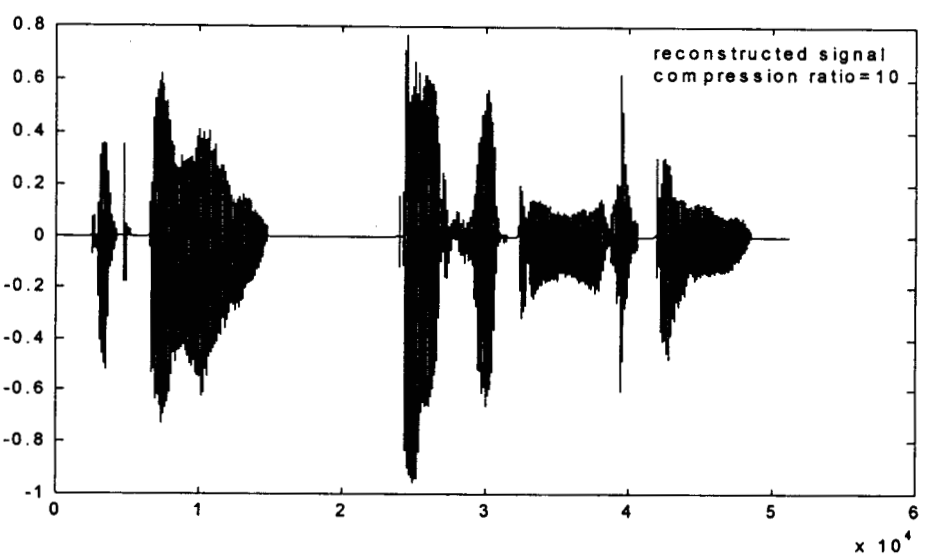


Fig.6.6: (c) Reconstructed signal (CR=10)

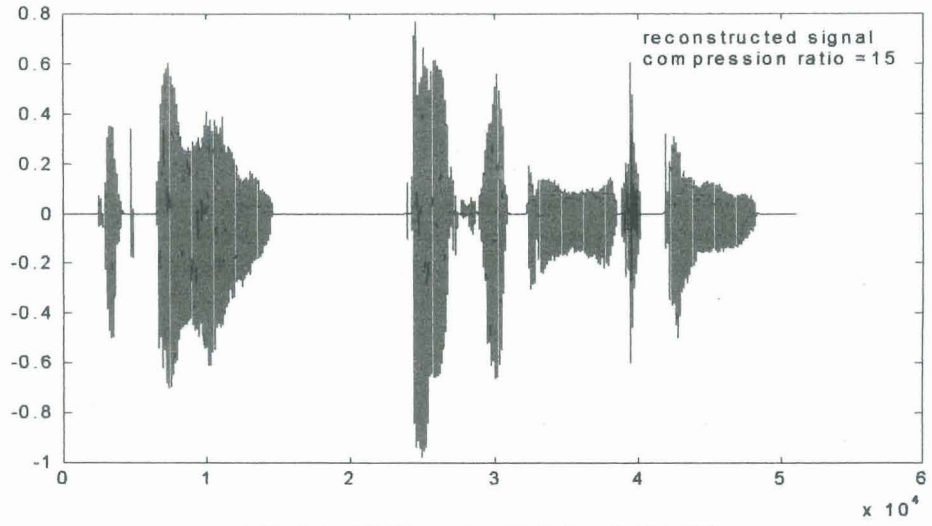


Fig.6.6: (d) Reconstructed signal (CR=15)

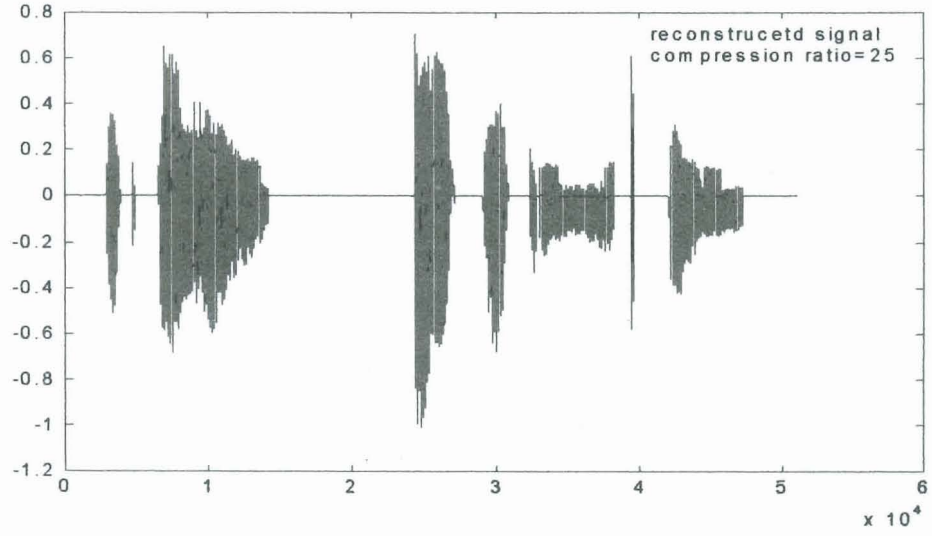


Fig.6.6: (e) Reconstructed signal (CR=6)

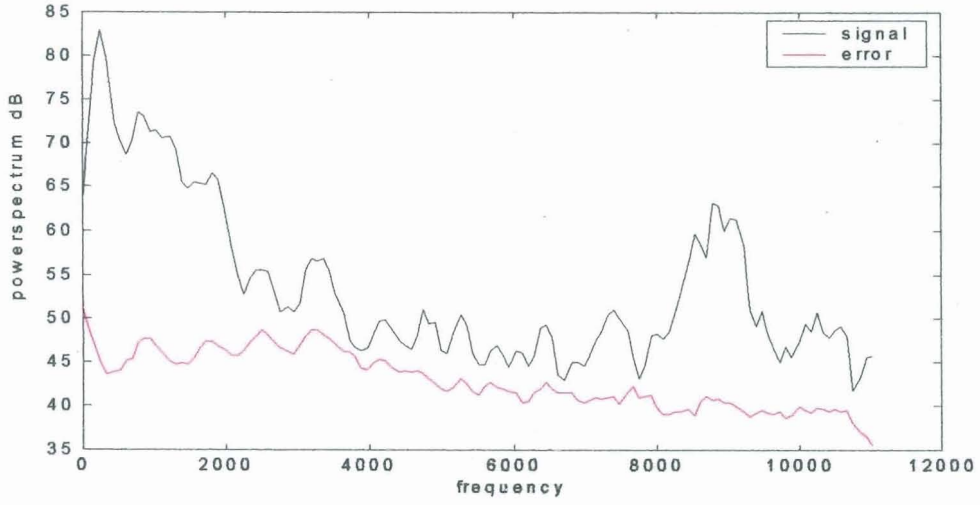


Fig. 6.7: Power spectra plot (female2, CR=6)

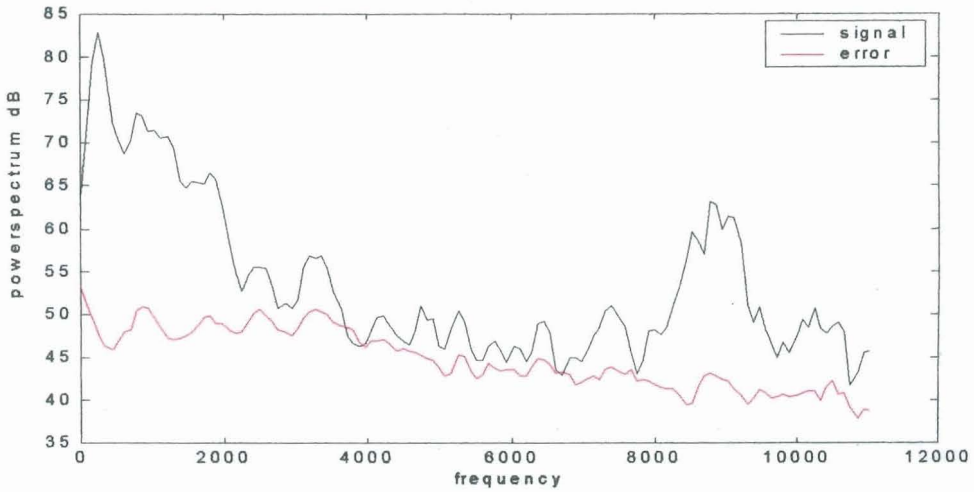


Fig. 6.8: Power spectra plot (female2, CR=10)

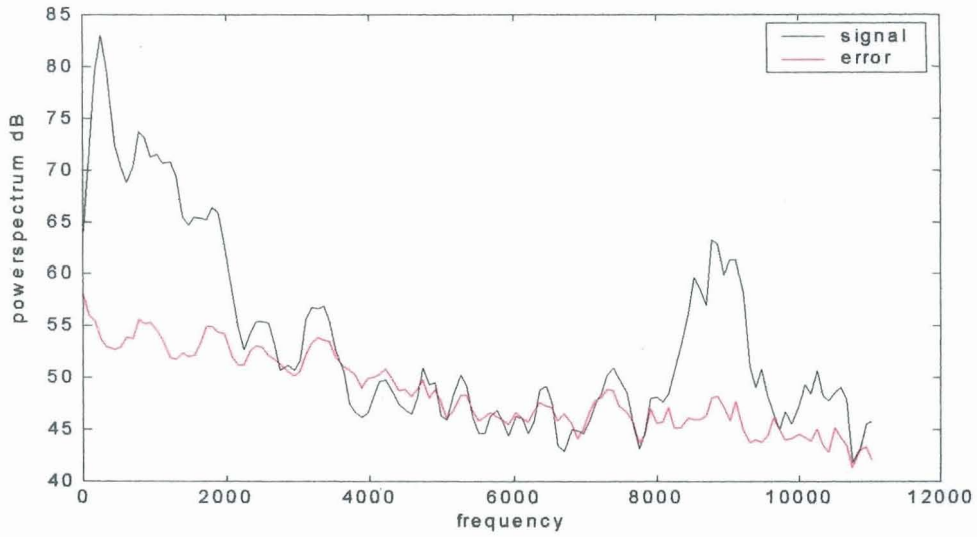


Fig. 6.9: Power spectra plot (female2,CR=15)

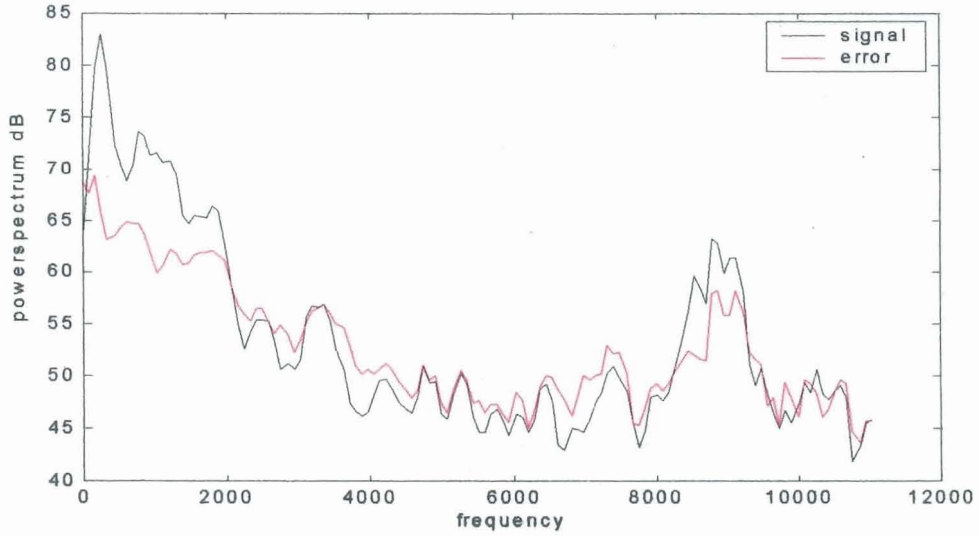


Fig. 6.10: Power spectra plot (female2, CR=25)

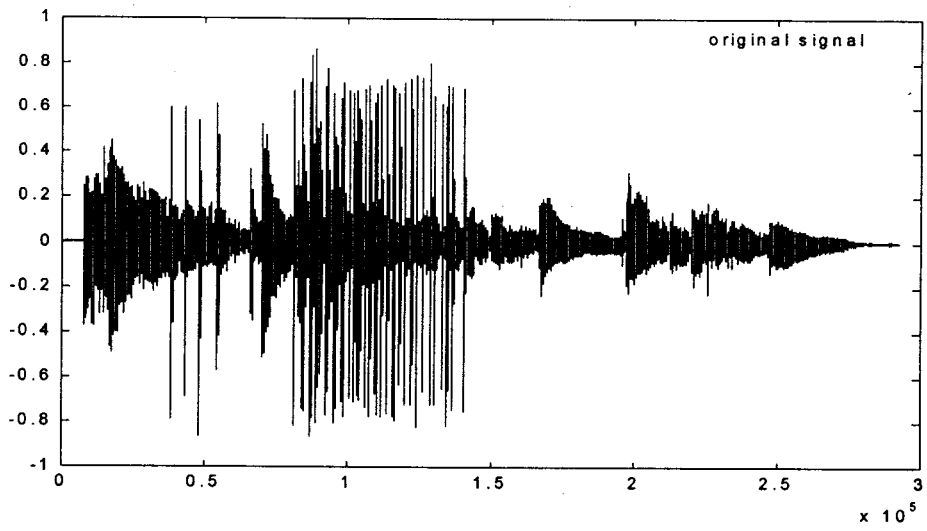


Fig.6.11: (a) Original signal ('castanets')

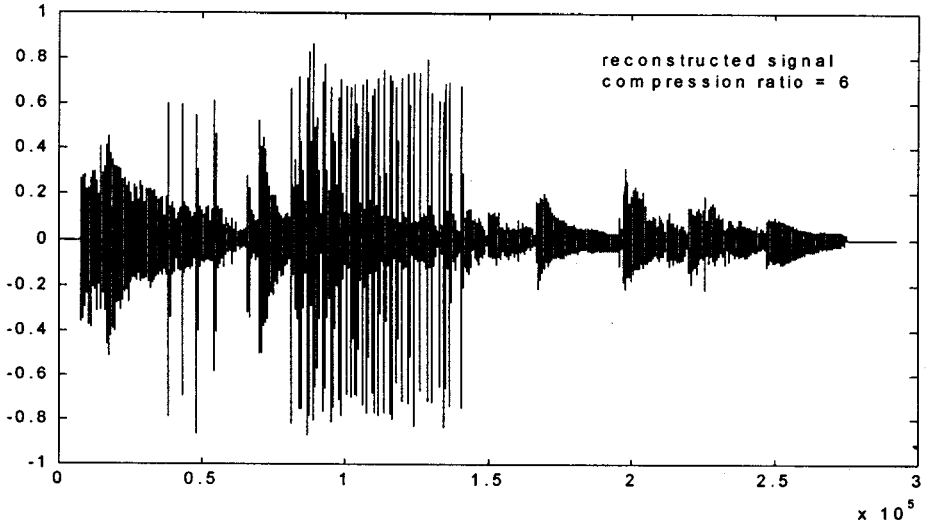


Fig.6.11: (b) Reconstructed signal (CR=6)

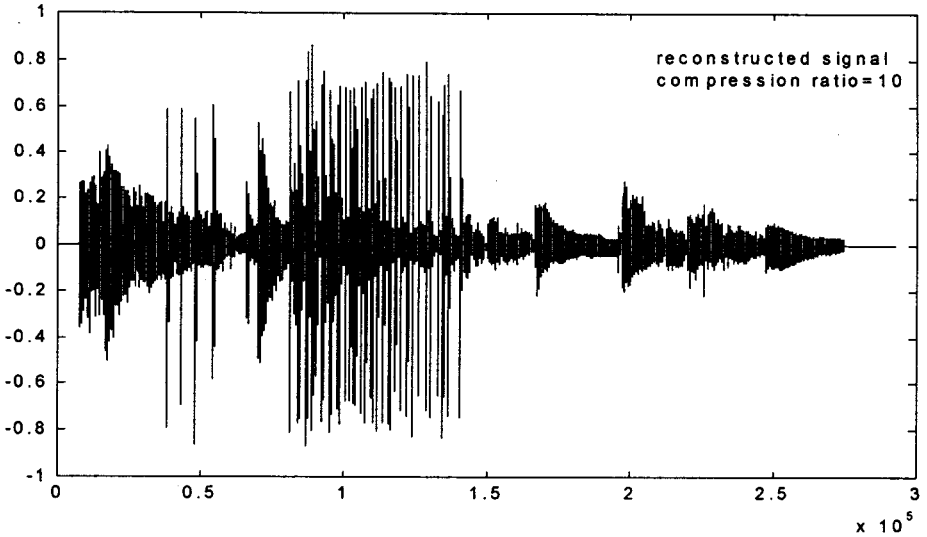


Fig.6.11: (c) Reconstructed signal (CR=10)

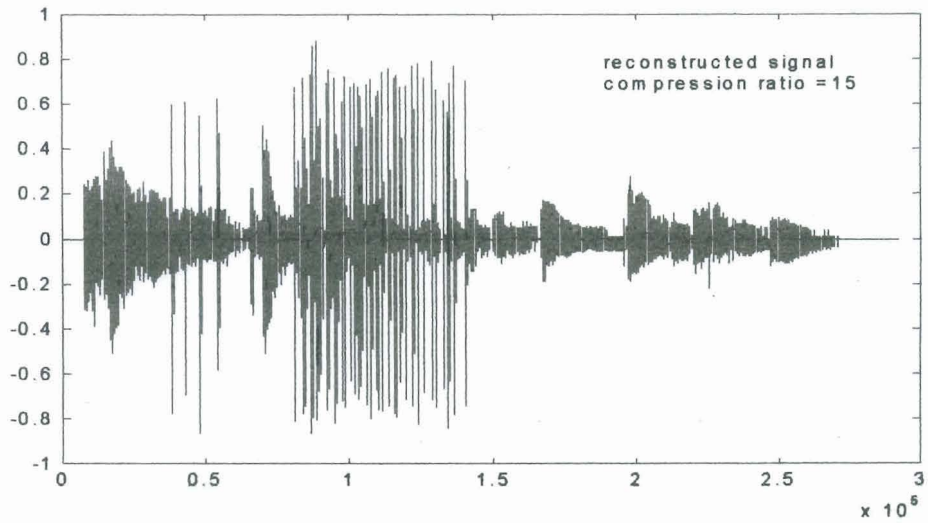


Fig.6.11: (d) Reconstructed signal (CR=15)

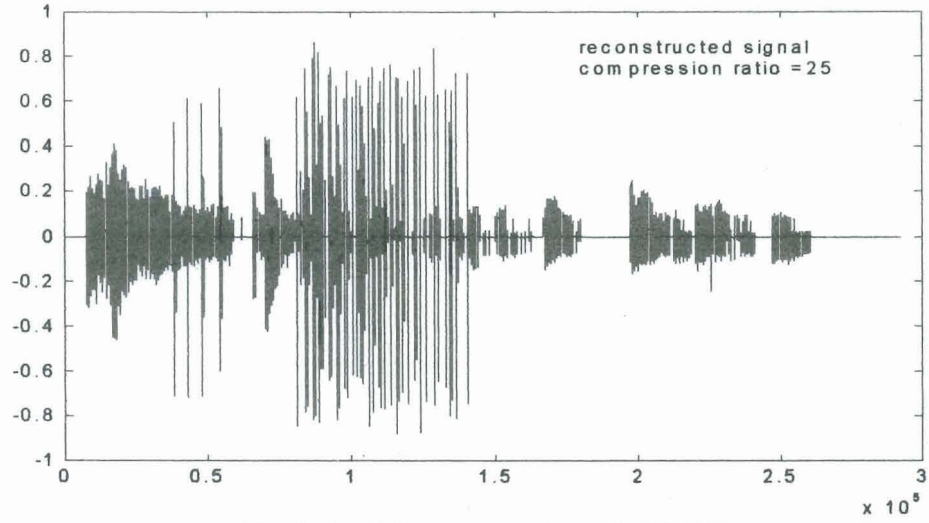


Fig.6.11: (e) Reconstructed signal (CR=25)

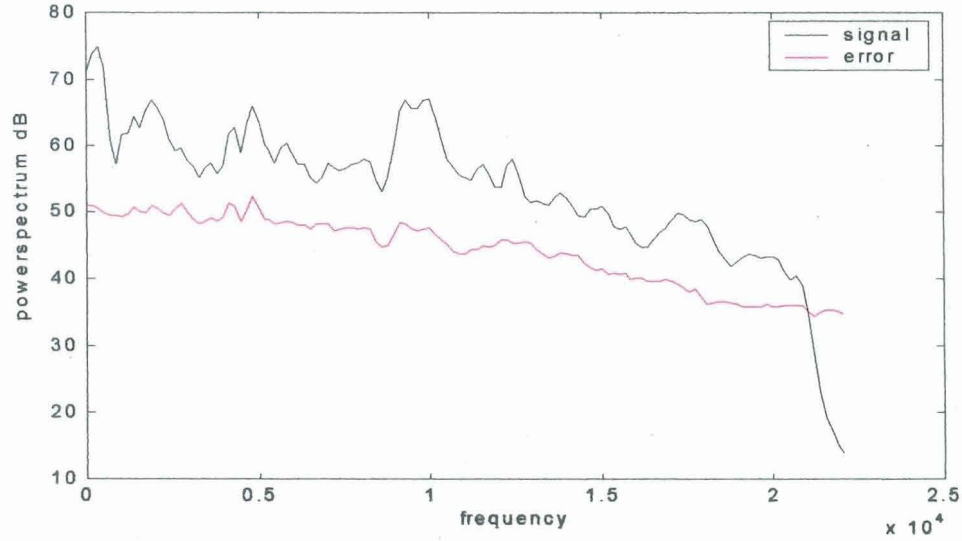


Fig.6.12: Power spectra plot (castanets, CR=6)

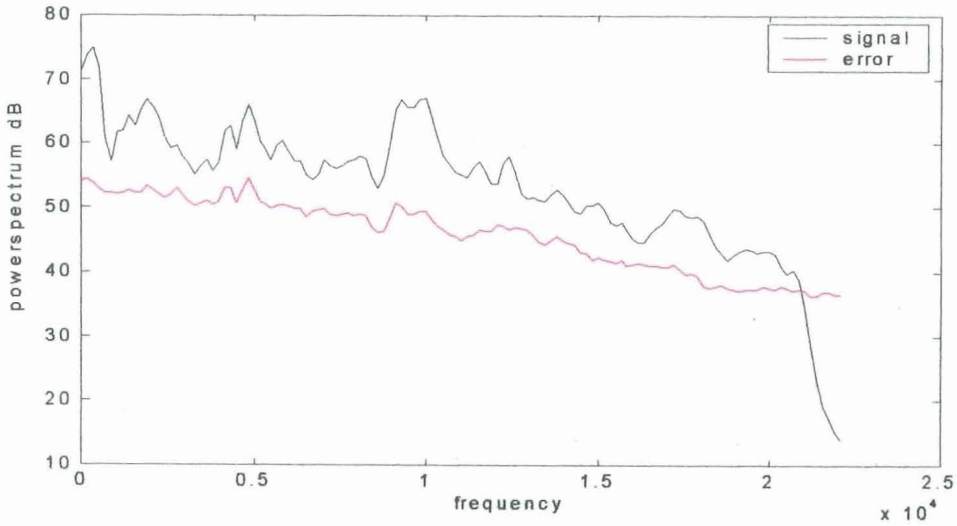


Fig. 6.13: Power spectra plot (castanets, CR=10)

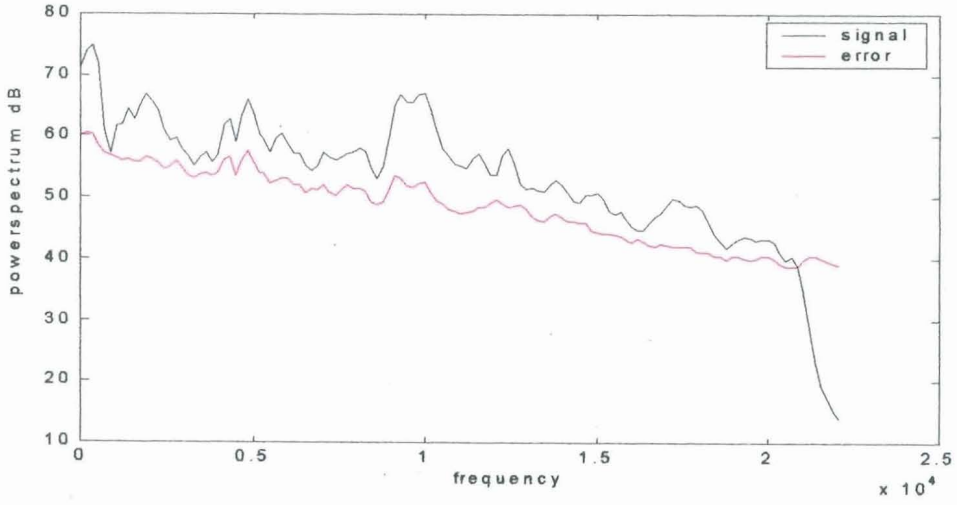


Fig.6.14: Power spectra plot (castanets, CR=15)

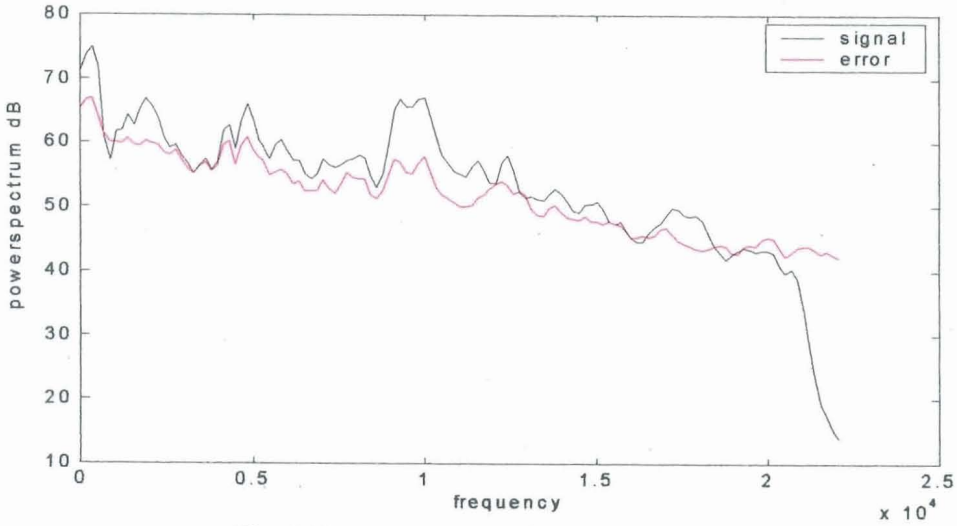


Fig. 6.15: Power spectra plot (castanets, CR=25)

6.4 Scalability

Scalability is the property of a coded signal that, part of the coded bit stream can be decoded in isolation. The fact that a subset of the coded bits is sufficient for generating a meaningful audio signal is important in at least two contexts. A first is where both cheap-simple and expensive-complex decoders are envisaged as being receivers of the signal: it is as if both AM and FM radio quality are available in the same transmitted signal dependent on an appropriate decoder. A second is when the transmission channel cannot guarantee the full necessary bandwidth to handle the complete bit stream, for example: internet radio. There are several types of scalability, in terms of bandwidth; number of channels ; and the most important, the so called SNR scalability. Therefore, in this section, a two-stage wavelet packet based SNR scalable perceptual audio coding scheme is proposed.

6.4.1 Optimum Wavelet Packet Based SNR Scalable Audio Coder

Scalable audio coding is usually done by coding a low bit-rate version of the original signal first. The difference between the original and the reconstructed signal is then coded using a second stage coding system.

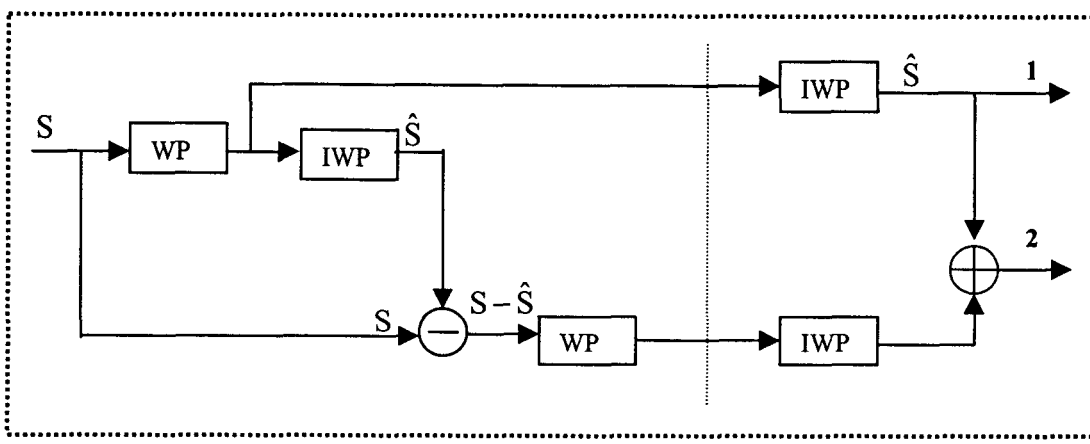


Fig.6.16: Block diagram of the proposed two stage SNR scalable audio codec

The block diagram of the proposed audio coder is shown in Fig.6.16. The first stage (low quality signal) of the proposed scheme is the wavelet packed based coding scheme using optimum wavelet basis, described in the previous section. The second stage (high quality signal) uses the same components as the first stage, but the difference between the original and the low quality reconstructed signal is coded.

6.5 Experiments and Results with two coding stages

The results of the two stage coder for various compression ratios are shown in Tables 6.5-6.8 and Figs.6.17-6.22. These results reveal that two stages of compression improve the SSNR values of the signal. Hence, high quality signal can be decoded from 2 and low quality version of the same signal can be decoded from 1. Original signal, reconstructed signals at 1 & 2 for various compression ratios and error power spectrum plots at 1 & 2 in the case of some typical audio signals are shown in Figs.6.17-6.22. From these figures it can be seen that error at 2 is less than that of error at 1 in all the cases. The error power spectra plots obtained are also in close agreement with the MOS and SSNR values given in Tables 6.5-6.8.

6.6 Summary

A constant bit rate discrete wavelet packet based audio coder has been designed and implemented in this chapter. Performance of the proposed codec is validated through subjective listening tests at different compression ratios. The perceptual quality of the coded signal is not constant here. It varies with the available bandwidth (bit rate) or compression ratio. In this chapter, a two stage SNR scalable wavelet packet based audio coding scheme using optimum wavelets has also been proposed. Results at different compression ratios for the low and high quality decoded bit streams have been quantified.

Table 6.5 Performance of Two Stage SNR Scalable WP based Coder (Compression Ratio = 6, Optimisation Method =3, Scalar Quantization)

Audio signal (.wav)	Sampling frequency (kHz)	SSNR at 1 (dB)	MOS	SSNR at 2 (dB)	MOS
male5a316	11.025	18.3	4.1	22.01	4.2
female5a316	11.025	15	4.2	17.1	4.2
clap5a316	11.025	16	4.8	19.2	4.8
ring5a316	11.025	3.7	1.0	4.5	1.0
clarinet5a316	11.025	11.3	3.5	12.8	3.7
pup5a316	22.05	13.2	4.1	15.1	4.3
whistle5a316	22.05	4.97	2.8	5.36	2.8
drums5a316	22.05	41.2	4.5	44.5	4.5
female25a316	22.05	24.1	4.3	28.2	4.5
crow5a316	22.05	16.5	4.7	20.5	4.7
cast5a316	44.1	14.7	4.6	17.8	4.6
mpegtest5a316	44.1	24.7	4.6	27.8	4.6
kadal5a316	44.1	22.5	4.2	26.3	4.3
else5a316	44.1	18.4	4.4	21.2	4.4
sitar5a316	44.1	18.5	4.2	21.6	4.6

Table 6.6 Performance of Two Stage SNR Scalable WP based Coder (Compression Ratio = 10, Optimisation Method =3, Scalar Quantization)

Audio signal (.wav)	Sampling frequency (kHz)	SSNR at 1 (dB)	MOS	SSNR at 2 (dB)	MOS
male5a3110	11.025	11.5	4.0	14.1	4.2
female5a3110	11.025	13.0	4.2	16.0	4.2
clap5a3110	11.025	14.3	4.8	17.1	4.8
ring5a3110	11.025	3.2	1.0	5.1	1.0
clarinet5a3110	11.025	8.3	3.2	9.7	3.2
pup5a3110	22.05	9.8	4.3	11.6	4.3
whistle5a3110	22.05	3.8	2.7	4.7	2.7
drums5a3110	22.05	30.1	4.3	35.6	4.4
female25a3110	22.05	20.5	4.1	24.6	4.2
crow5a3110	22.05	14.7	4.3	16.2	4.3
cast5a3110	44.1	12.73	4.2	13.8	4.2
mpegtest5a3110	44.1	23.2	4.6	25.8	4.6
kadal5a3110	44.1	17.2	4.0	20.7	4.0
else5a3110	44.1	15.1	4.4	18.3	4.4
sitar5a3110	44.1	13.5	4.2	15.1	4.5

Table 6.7 Performance of Two Stage SNR Scalable WP based Coder (Compression Ratio = 15, Optimisation Method =3, Scalar Quantization)

Audio signal (.wav)	Sampling frequency (kHz)	SSNR at 1 (dB)	MOS	SSNR at 2 (dB)	MOS
male5a3115	11.025	8.9	3.2	9.7	3.3
female5a3115	11.025	9.5	3.6	10.2	3.6
clap5a3115	11.025	10.1	4.4	12.1	4.4
ring5a3115	11.025	2.4	1.0	2.7	1.0
clarinet5a3115	11.025	6.8	2.9	7.3	3.0
pup5a3115	22.05	6.2	3.9	7.8	3.9
whistle5a3115	22.05	2.5	1.9	3.1	1.9
drums5a3115	22.05	22.5	4.1	24.7	4.2
female25a3115	22.05	16.5	3.8	18.1	3.8
crow5a3115	22.05	11	4.1	12.9	4.2
cast5a3115	44.1	9.4	3.9	10.5	3.9
mpegtest5a3115	44.1	13.4	3.5	14.7	3.6
kadal5a3115	44.1	12.8	3.8	14.1	3.8
else5a3115	44.1	13.7	4.0	15.2	4.0
sitar5a3115	44.1	11	4.1	12.7	4.1

Table 6.8 Performance of Two Stage SNR Scalable WP based Coder
(Compression Ratio = 25, Optimisation Method =3, Scalar Quantization)

Audio signal (.wav)	Sampling frequency (kHz)	SSNR at 1 (dB)	MOS	SSNR at 2 (dB)	MOS
male5a3125	11.025	6.6	2.8	7.1	2.9
female5a3125	11.025	8.1	3.0	9.3	3.2
clap5a3125	11.025	7.7	3.6	8.8	3.8
ring5a3125	11.025	1.22	1.0	1.34	1.0
clarinet5a3125	11.025	2.5	2.0	3.2	2.1
pup5a3125	22.05	3	2.8	4.1	2.9
whistle5a3125	22.05	1	1.3	1.3	1.3
drums5a3125	22.05	11.2	3.7	12.8	3.8
female25a3125	22.05	9.06	3.0	10.6	3.2
crow5a3125	22.05	6.1	3.7	7.1	3.8
cast5a3125	44.1	5.4	3.1	5.9	3.4
mpegtest5a3125	44.1	9.3	2.9	10.4	3.3
kadal5a3125	44.1	6.9	3.0	8.3	3.4
else5a3125	44.1	6.7	3.1	7.8	3.4
sitar5a3125	44.1	5.4	3.2	6.1	3.5

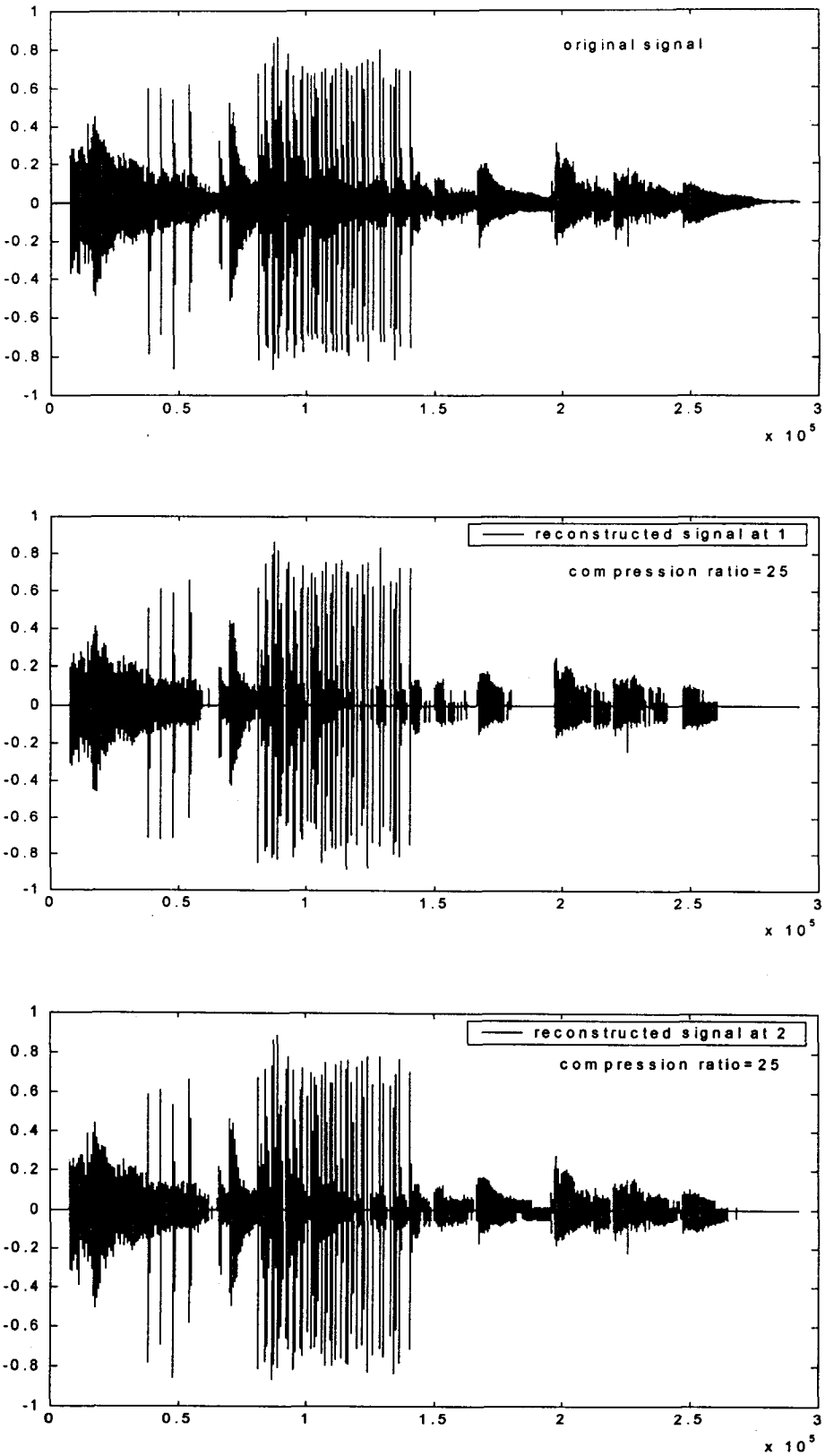


Fig. 6.17: Performance of two stage SNR Scalable Audio Codec ('castanets', 'CR=25)

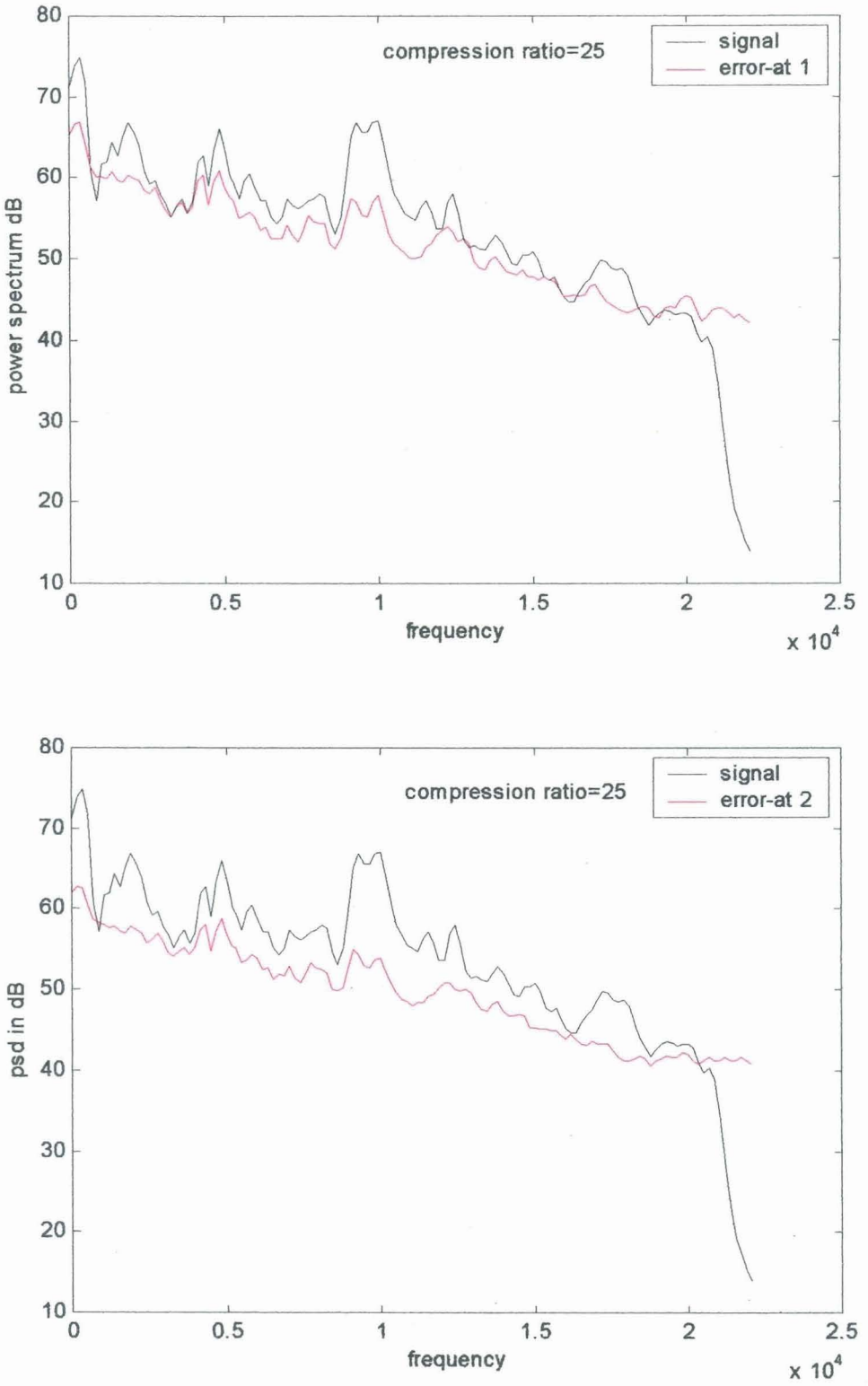


Fig.6.18: Power spectra plots at 1 and 2 for two stage SNR Scalable Audio Codec ('castanets', CR=25)

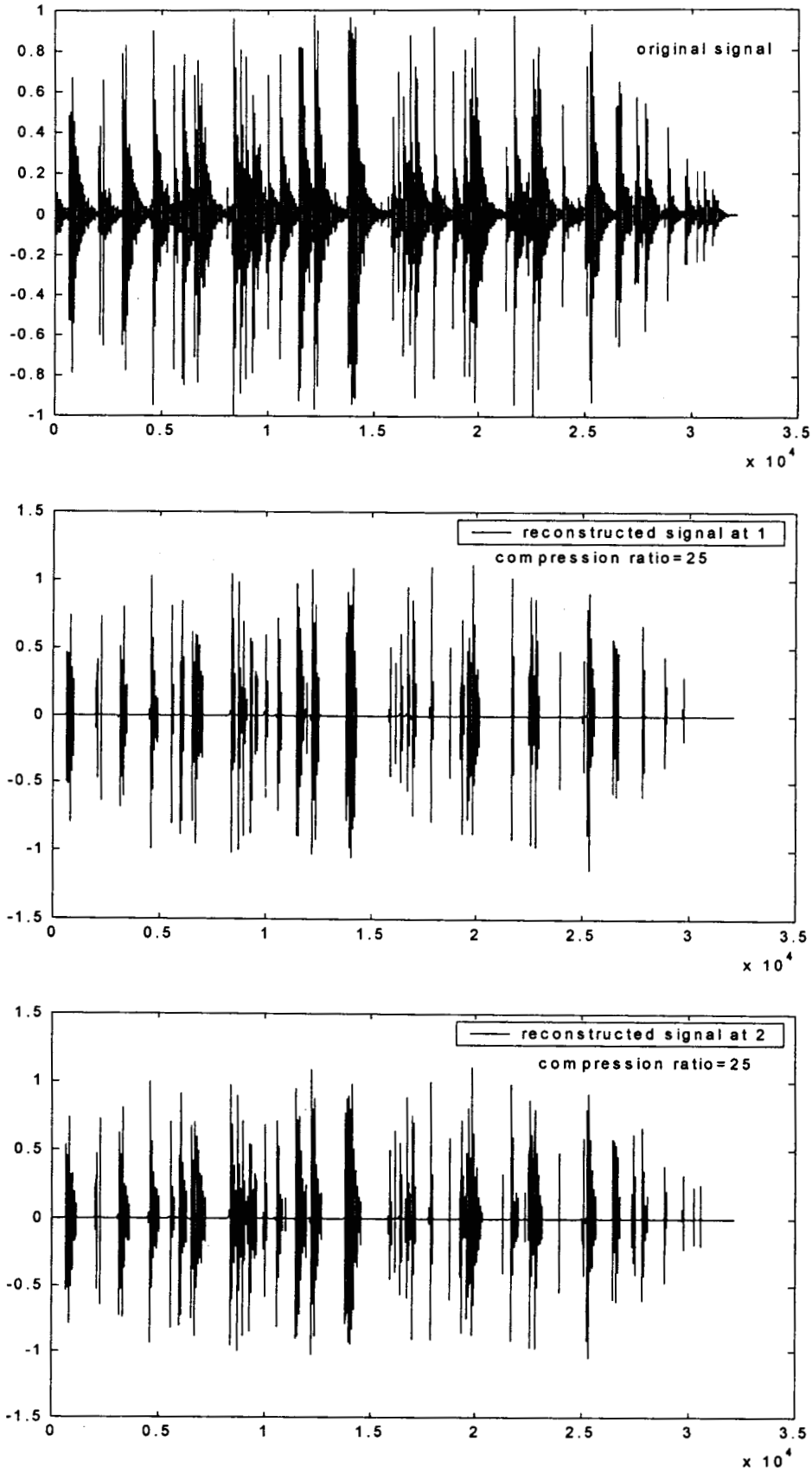


Fig.6.19: Performance of two stage SNR Scalable Audio Codec ('clap', CR=25)

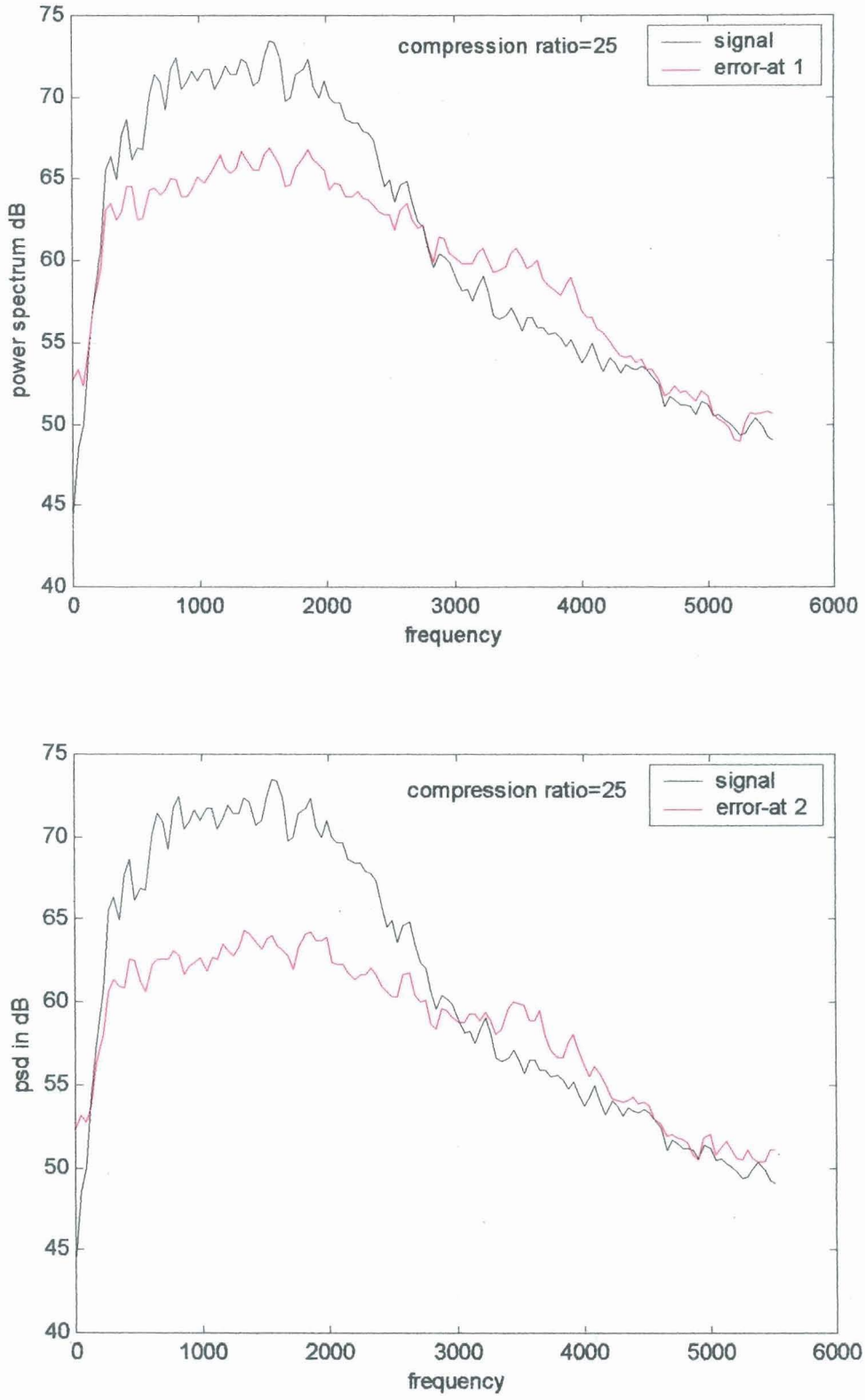


Fig. 6.20: Power Spectra Plots at 1 and 2 for two stage SNR Scalable Audio Codec ('clap', CR=25)

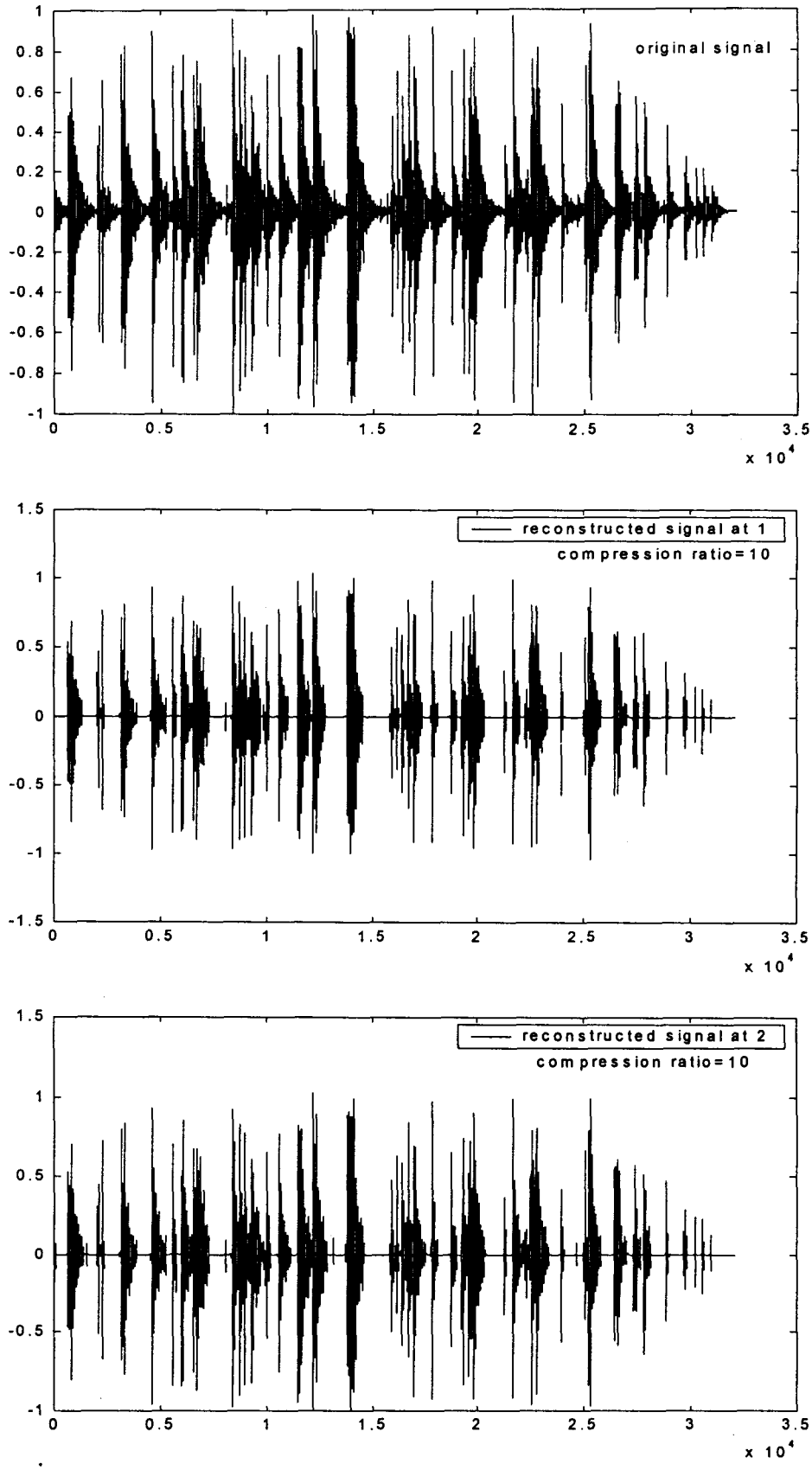
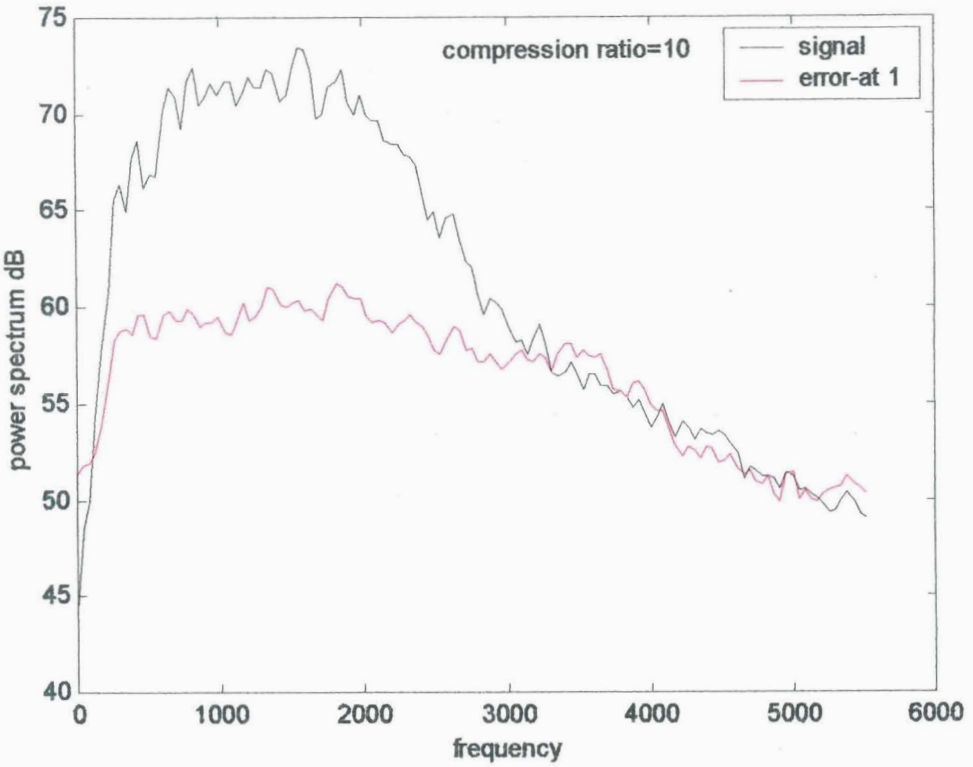


Fig. 6.21: Performance of two stage SNR Scalable Audio Codec ('clap', CR=10)



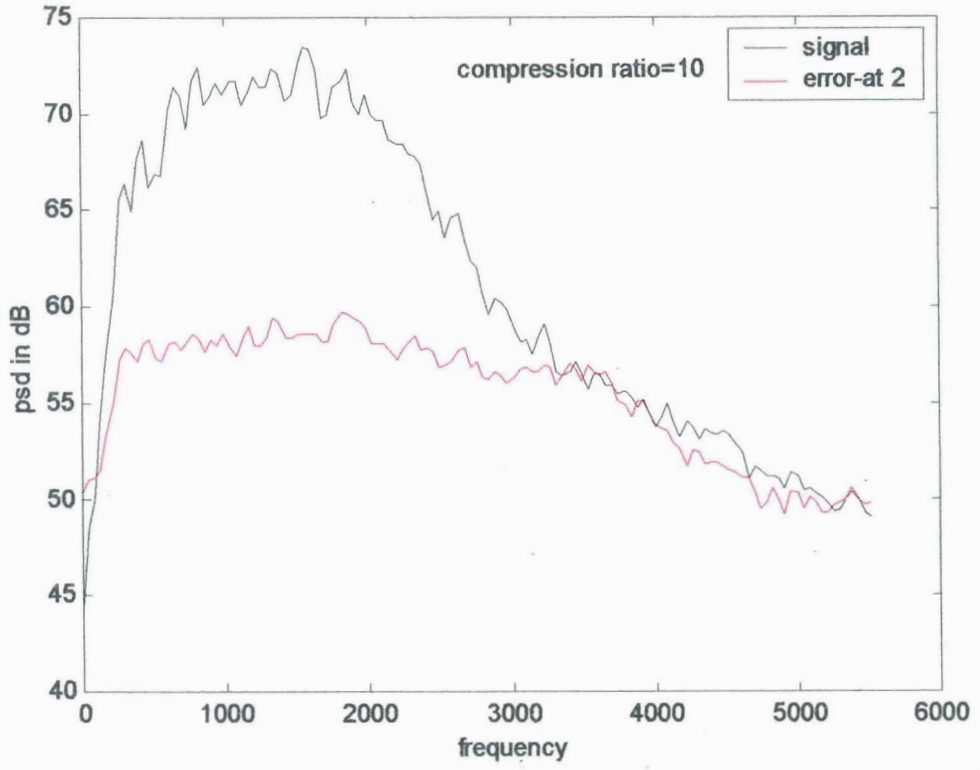


Fig. 6.22: Power Spectra Plots at 1 and 2 for two stage SNR Scalable Audio Codec ('clap', CR=10)

Chapter 7

ENHANCEMENT AND MODIFICATION OF SPEECH AND AUDIO FOR THE HEARING IMPAIRED

7.1 Introduction

Approximately 7.5 percent of the population has some degree of hearing loss and about 1.0 percent has a loss that is moderately severe or greater. A majority of the hearing impaired population would benefit significantly from improved methods of acoustic amplification. Two major types of hearing impairments are *conductive deafness* and *sensorineural deafness*. Conductive deafness is caused by the degraded transmission of the acoustic energy to the sense organ (cochlea) and can be modeled by a linear distortion. Hence this can be adequately compensated by using normally available analog hearing aids (in which linear amplification is used). On the other hand, sensorineural deafness is caused by abnormal function of the cochlea, the auditory nerve or both. This impairment is mainly caused by prolonged or excessive exposure to noise, aging, hereditary factors etc.

Considerable research has been devoted to the development of various types of hearing compensation techniques/algorithms. Perception of loudness and its abnormalities in impaired hearing are briefly discussed in this chapter. Various hearing compensation algorithms/techniques are also reviewed. New algorithms for digital hearing aids using Discrete Wavelet Transforms (DWT) and Discrete Wavelet Packets (DWP) are proposed here. The clinical test results of the proposed hearing aid revealed that wavelet analysis is a promising tool for designing efficient digital hearing aids.

7.2 The Perception of Loudness

The human ear is remarkable both in terms of its absolute sensitivity and in terms of the range of sound intensities to which it can respond. The most intense sound we can hear without damaging our ears has a level about 120 dB above that of the faintest sound we can detect. This corresponds to a ratio of intensities of 1×10^{12} :1. Loudness [65] is defined as that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud.

7.2.1 Absolute Thresholds

The absolute threshold of a sound is the minimum detectable level of that sound in the absence of any other external sounds. This is shown in Chapter 2 [Fig.2.7]. The highest audible frequency varies considerably with the age of the subject. Young children can often hear tones as high as 20kHz, but most adults cannot hear tones above about 15kHz. The loss of sensitivity with increasing age is much greater at high frequencies than at low frequencies. Sounds below about 16 Hz are not heard in the normal sense, but are detected by virtue of the distortion products (harmonics) which they produce after passing through the middle ear. The low-frequency limit for the 'true' hearing of pure tones probably lies around 16 Hz.

In many practical situations our ability to detect faint sounds is limited not by our absolute sensitivity to those sounds but by the level of ambient noise. In other words, detection depends upon the masked threshold rather than the absolute threshold; information about masking and masked thresholds are given in Chapter 2. In the clinical assessment of hearing, thresholds are usually specified relative to the average threshold at each frequency for young, healthy listeners with 'normal' hearing. Thresholds specified in this way have units dB HL (hearing level) in Europe or dB HTL (hearing threshold level) in the USA. Thus for example, a threshold of 40dB HL at 1 kHz means that the patient has a threshold which is 40 dB higher than 'normal' at that frequency. Threshold in this case correspond to 46 dB SPL (Since for normal persons threshold is 6 dB SPL at 1 kHz).

In psychoacoustic work, thresholds are normally plotted with threshold increasing upwards. However in clinical situations, threshold elevation are shown as hearing losses, plotted downwards. The average normal' threshold is represented as a horizontal line at the top of the plot and the degree of hearing loss is indicated by how much the threshold falls below this line. This type of plot [98] is called an 'audiogram'. Audiogram for a patient is obtained by using a pure-tone audiometer. This is an instrument which delivers tones of variable frequency and intensity to the ear by earphones. The frequencies usually tested are in octave steps, i.e., 250, 500, 1000, 2000, 4000 and 8000 Hz. Occasionally half octave steps, i.e., 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000 and 8000 Hz, are also used. The intensity can be increased or decreased for each frequency and can vary from 10dB to 120dB. Most audiometers used today are calibrated to the International (ISO) standard zero level. The sample audiograms of a normal person and a hearing impaired person with sensorineural deafness.are shown in Figs.7.1 and 7.2 respectively.

7.2.2 Abnormalities of Loudness Perception in Impaired Hearing

As mentioned in the introduction, hearing losses may be broadly categorized into two main types. The first type, *conductive hearing loss* occurs when there is a defect, usually in the middle ear, which reduces the transmission of sound to the inner ear. For, example, viscous fluid may build up in the middle ear as a result of infection or the stapes may be immobilized as a result of growth of bone over the oval window. Sometimes a conductive loss is produced by wax in the ear canal. In general, a conductive loss results in a more or less uniform hearing loss as a function of frequency; it can be regarded as resulting in a simple attenuation of the incoming sound. The difficulty experienced by the sufferer can be well predicted from the elevation in absolute threshold. A simple hearing aid is usually quite effective in such cases and surgery can also be effective.

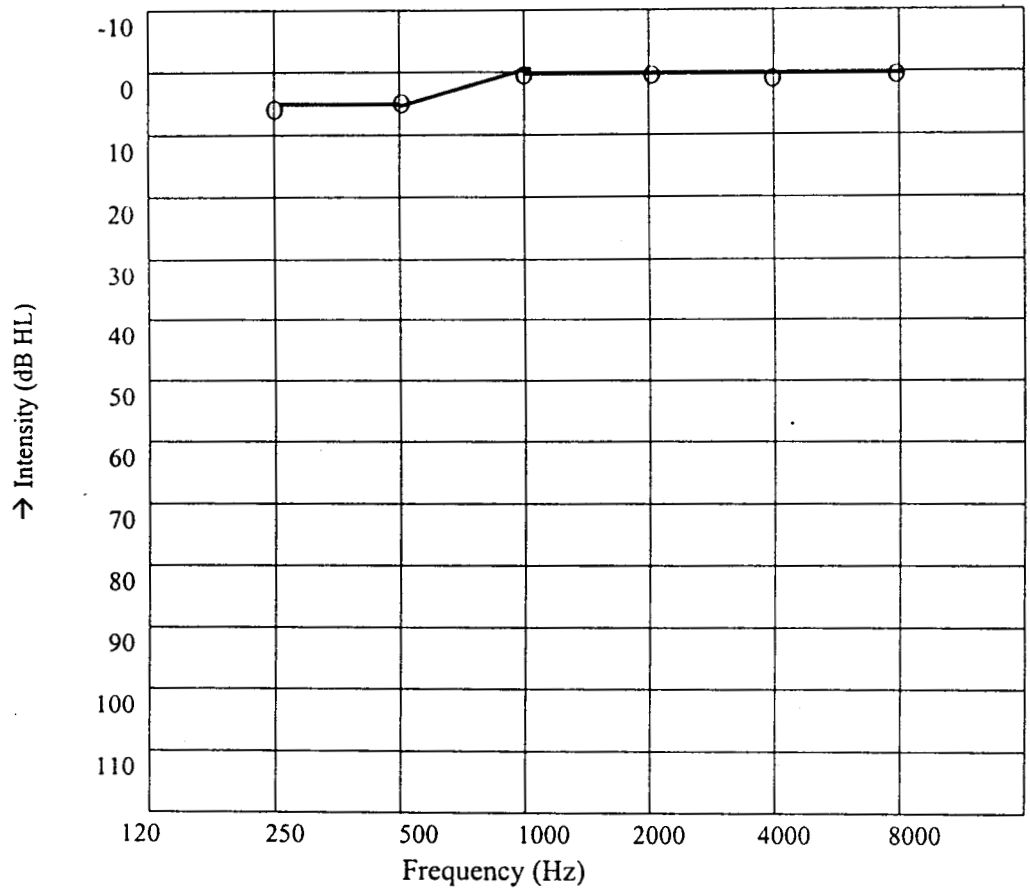


Fig. 7.1: Sample audiogram of a normal person

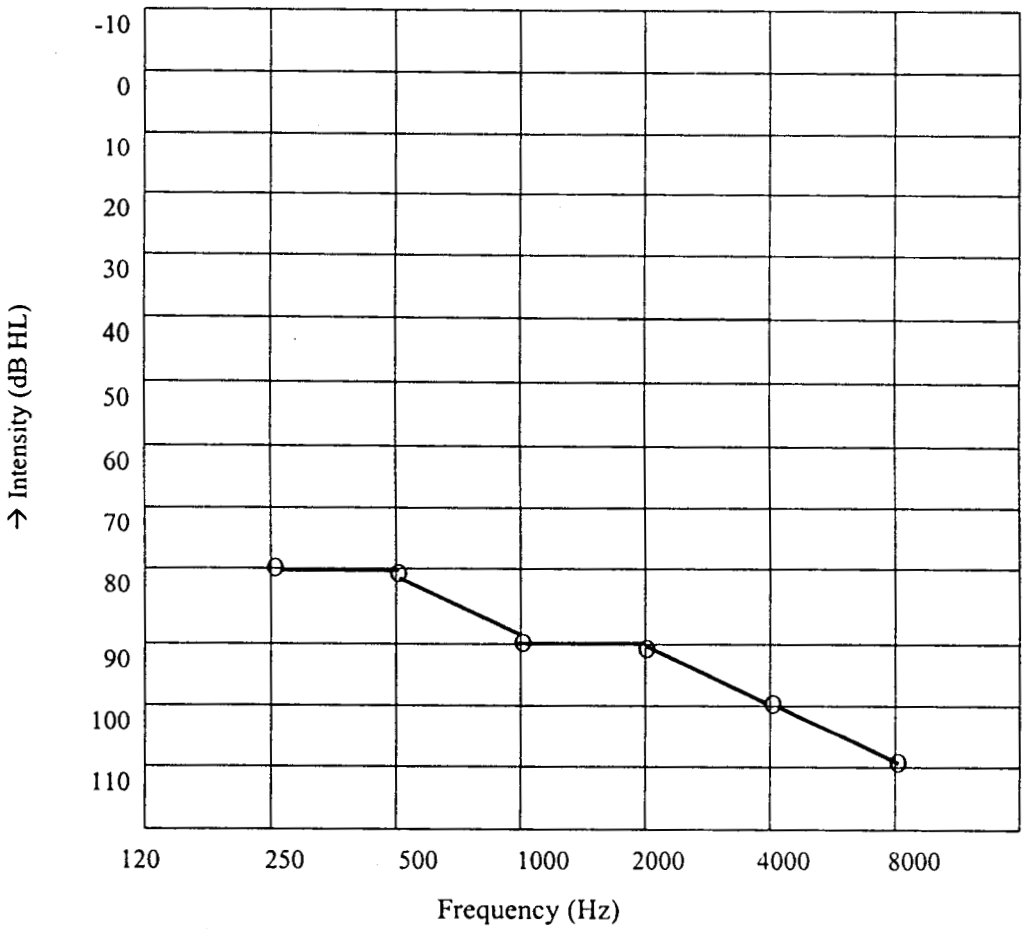


Fig 7.2: Sample audiogram of an impaired person with sensorineural deafness

The second type of hearing loss is called *sensorineural hearing loss*, although it is also inaccurately known as 'nerve deafness'. Sensorineural hearing loss most commonly arises from a defect in the cochlea and is then known as a cochlear loss. This is mainly caused by prolonged or excessive exposure to noise, aging, hereditary factors etc. However, sensorineural hearing loss may also arise as a result of defects in the auditory nerve or higher centres in the auditory system. Hearing loss due to neural disturbances occurring at a higher point in the auditory pathway than the cochlea is known as retrocochlear loss. The particular difficulties experienced by the sufferer and the types of symptoms exhibited, depend on which part of the system is affected. Often the extent of difficulty experienced by the sufferer is not always well predicted from the audiogram. People with sensorineural hearing loss often have difficulty in understanding speech in noisy environments and the condition is usually not completely alleviated by a hearing aid. Most sensorineural losses cannot be treated by surgery.

7.3 Loudness Recruitment

One phenomenon which often occurs when there are defects in the cochlea (sensorineural deafness) is loudness recruitment. This refers to an unusually rapid growth of loudness as the level of a tone is increased. The phenomenon of loudness recruitment probably accounts for a statement which is often heard from people with this type of hearing loss: *'Don't shout; you're talking loud enough, but I can't understand what you are saying!'*. The person may not be able to hear very faint sounds, but sounds of high intensity are just as loud as for a normal listener. However, sounds which are easily audible may not be easily intelligible.

Loudness recruitment occurs consistently in disorders of the cochlea and is usually absent in conductive deafness. It is usually connected with hair-cell damage and in particular with damage to the outer hair cells. It is certainly the case that the outer hair cells are highly susceptible to damage (e.g. from intense auditory stimulation). Recruitment can be caused by brain stem disorders also.

Kiang and Evans suggested that reduced sharpness of tuning might be the main factor for contributing to loudness recruitment [64]. For a sinusoidal stimulus, this

leads to an excitation pattern which is broader (spreads over a greater range of frequencies) in an impaired than in a normal ear. They suggested that, once the level of a sound exceeds threshold, the excitation in an ear with cochlear damage spreads more rapidly than normal, across the array of neurons and this leads to the abnormally rapid growth of loudness with increasing level.

An alternative explanation for loudness recruitment is that it results from damage to or loss of the active process in the cochlea which enhances sensitivity for low-input sound levels. This process is non linear and results in an amplification of the Basilar Membrane (BM) response to low level sounds, while leaving the response to high level sounds relatively unamplified. If this active process is lost, then the response to low level sounds is not amplified and the absolute threshold is elevated. However, the response to high level sounds remains roughly the same as normal.

The thresholds of hearing and pain bound the range of sounds that a person can hear. This range is called *dynamic range of hearing.* For normal listeners, equal increments in sound intensity produce equal increments in loudness perception uniformly across all frequencies. The threshold of hearing is raised with recruitment of loudness and the threshold of pain may remain constant or even may be lowered. So the dynamic range of hearing is reduced, causing relatively small changes in intensity to give larger corresponding changes in perceived loudness.

7.4 Frequency Selectivity in Impaired Hearing

There is now considerable evidence that frequency selectivity is impaired by damage to the cochlea. It is worth considering the perceptual consequences of a loss in frequency selectivity. The first major consequence is a greater susceptibility to masking by interfering sounds. When we are trying to detect a signal in a noisy background we use the auditory filter(s) giving the best signal-to-noise ratio. In a normal ear, where the auditory filters are relatively narrow, all of the background noise, except a narrow band around the signal frequency is attenuated at the filter output. In an impaired ear, where the filters are broader, much more of the noise gets through the filter, so the detectability of the signal is reduced. Thus background noise severely

disrupt the detection and discrimination of sounds, including speech. This may partly account for the great difficulties experience by those with cochlear impairment in following speech in noisy situations such as in public places or at parties.

A second difficulty arises in the perceptual analysis of complex sounds such as speech or music. When frequency selectivity is impaired, the ability to detect differences in the spectral composition of sounds is reduced. Thus it may be more difficult for the impaired listener to tell the difference between different vowel sounds or to distinguish different musical instruments. We should note that the provision of a hearing aid which simply amplifies sound will not overcome any of the above difficulties. Such an aid may help to make sounds audible, but it does not correct the impaired frequency selectivity.

7.5 Hearing Model

Based on numerous studies of the physiology of the ear as well as results from psychoacoustical tests, David.V.Anderson [99] developed a hearing model which represents the major points of the auditory system. This model is complete in the sense that it takes sound as an input and provides as an output a loudness estimate for each frequency band. This model is relatively simple to analyze. and it mimics four functions of the auditory system as illustrated in Fig 7.3.

1. Frequency separation or filtering (performed primarily by the basilar membrane).
2. Automatic Gain control (AGC) provided by variable feedback from the Outer Hair Cells (OHC).
3. Transduction of incoming sound pressure signals to neural signals (by the inner hair cells).
4. Loudness Perception.

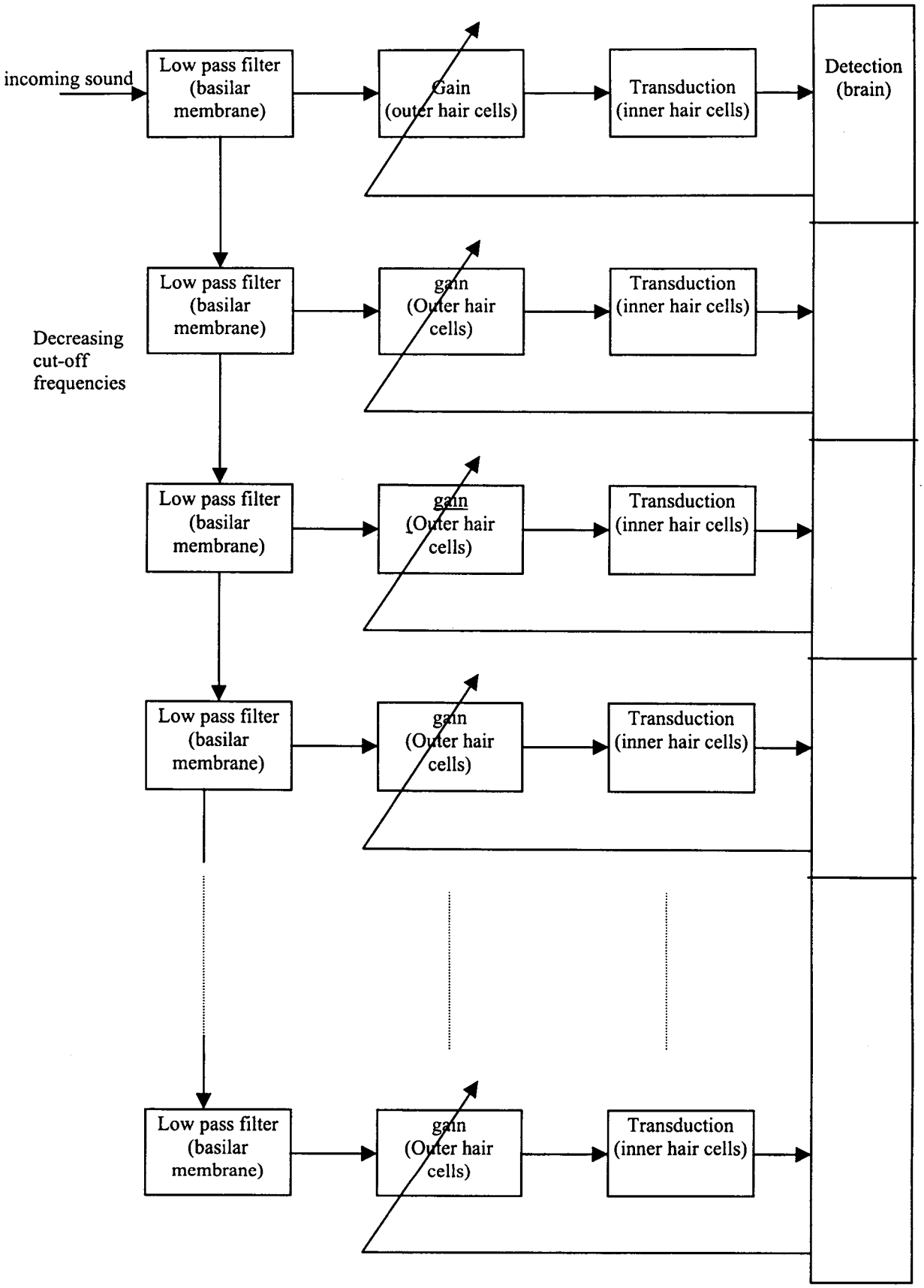


Fig. 7.3: Basic hearing model

In the human ear, the most physiologically vulnerable components are outer hair cells (OHC). The OHCs provide the positive feedback necessary for detecting low level signals. This feedback also sharpens the tuning characteristics of the cochlea. If OHCs are damaged or destroyed, the ability of the ear to hear low level sounds is lost. The dynamic range of sounds which the ear may hear is decreased. Damage may occur for a number of reasons including noise exposure, aging, ototoxic drugs etc. Hence automatic gain control function will be lost and broadening of the tuning characteristics of basilar membrane will take place due to damage of OHC. For conductive loss, the coupling of sound in the middle ear is severely impaired by missing or damaged middle ear bones, fluid build-up or a damaged ear drum. In the case of conductive loss, all frequencies are attenuated; which may be modelled as simple attenuation of the input signal by some fixed amount. This loss can be compensated by linear amplification. But, loss due to OHC damage (sensorineural deafness), cannot be compensated by a linear amplification. In this type of loss, hearing compensation methods should provide automatic gain control according to the signal intensity at various frequency bands so that a hearing impaired individual will hear a sound just as a normal hearing individual would hear it.

7.6 Hearing Aid Design

Psychoacoustic research helps in the design of various types of hearing aids to compensate for some of the abnormalities of perception which occur in the impaired ear. Psychoacoustics has two roles to play in this. The first is in the characterization of the perceptual abnormalities. This is important for deciding the type of compensation to be used. For example, conductive deafness can be compensated by linear amplification. The second is in the design and evaluation of particular compensation schemes.

An example of this approach is found in attempts [100-115] to compensate loudness recruitment. As already mentioned, in an ear with recruitment, the absolute threshold is elevated, but the level at which sounds become uncomfortably loud may be almost normal. That is, the growth of loudness with increasing intensity is more rapid than normal. A hearing aid which amplifies all sounds equally is not satisfactory in

such cases. If the gain of the aid (the amount by which it amplifies, usually specified in decibels) is set so as to make faint sounds audible, then more intense sounds will be over amplified and will be uncomfortably loud.

One method of compensating for loudness recruitment is to use a hearing aid which compresses the dynamic range of the input, so that low level sounds are amplified more than high level sounds. This is sometimes done in a crude way in hearing aids by limiting the maximum voltage that can be generated by the output amplifier, but this introduces a considerable amount of distortion. It sounds very unpleasant and it reduces speech intelligibility. An alternative method is to use an amplifier whose gain is adjusted automatically according to the level of the input (AGC). But the design of effective AGC systems for hearing aids is difficult. One reason for this is, that, the average level of speech varies from one situation to another over a range of at least 30dB. The person with recruitment may only be able to hear comfortably over a much smaller range of sound levels. This problem can be dealt with by using AGC in which the gain changes slowly from one situation to another. Unfortunately, it is also necessary for the gain to be rapidly reduced in response to a sudden intense sound, such as a door slamming, or a cup being dropped. With slow acting AGC, this means that the aid goes 'dead' for a certain time after the intense sound, which is very annoying for the user.

Moore et al. have described an AGC system which changes gain slowly from one situation to another, but which can rapidly reduce the gain for sudden intense sounds. After cessation of a brief intense sound, the gain returns quickly to what it was before the sound occurred. This type of AGC is very effective in dealing with variations in overall sound level from one situation to another [64].

A second reason is that, even for speech at a constant average level, individual acoustic elements may vary over a range of 30dB or more. Typically, the acoustic elements associated with vowels are more intense than those associated with consonants. Thus a person with recruitment may be able to hear the vowel sounds, but weaker consonants may be inaudible. To deal with this, it is necessary to use AGC whose gain changes rapidly, so that the levels of weak consonants can be increased relative to those of vowels.

David V. Anderson et al. developed a hearing compensation algorithm based on a homomorphic multiplicative AGC and implemented on a real time DSP with a 16 bit ADC and DAC and a sampling frequency of 21.333 kHz. The sound was filtered into 12 one third octave bands, the loudness of each band was compressed in such a way as to greatly increase the intensity of soft sounds but not loud sounds, and the bands were each recombined and the sound was played through a standard hearing aid receiver (speaker). Here, compression refers to mapping speech/audio from the normal range into the reduced range of an impaired listener. Clinical tests were conducted on eight subjects and compared against commercially available digitally programmable analog hearing aid. The authors have reported that the new algorithm had made significant progress in restoring normal or near normal hearing for hearing impaired individuals [64]. Major drawback of this algorithm is that, subbands are not according to the common audiogram standards .

Sigisbert Wyrch and August Kaelin, presented a hearing aid concept working in subbands in [112]. Consideration was given to the compensator for recruitment of loudness. The transformation into subbands is based on the modulated lapped transform (MLT). In this the audible frequency range from 20 Hz to 16 kHz is subdivided into 24 critical bands. A frequency and magnitude dependent compensator is provided here. The draw backs of this scheme are:

- Auditory filter band widths for the hearing impaired people are wider than the critical bands. But, in this scheme, signal is divided into 24 critical bands (narrow bands).
- Subbands are not according to the common audiogram standards.

Janet C. Rutledge and Mark A. Clements [100], explored the problem of compensation of recruitment of loudness in sensorineural hearing impairment. The technique used is based on a digital sinusoidal analysis/synthesis model of speech. Time Varying, Frequency Dependent (TVFD) non-linearity are applied between the analysis and synthesis modules to perform the compensation. This system differs from traditional multichannel filtering techniques by incorporating a model of the interaction of sinusoidal maskers in both normal hearing and hearing impaired individuals. Psychoacoustic properties like spread of masking, elevated thresholds, and reduced

dynamic range are taken into account for compensation. A system of simultaneous equations is solved to determine the proper gains for each sinusoid. This scheme is a combination of multichannel amplitude compression and automatic gain control since the compressive gains calculated separately for each frame of speech "automatically" adjust to the level of the speech components in that frame. Speech is represented as a sum of sinusoids with various amplitudes, frequencies and phases. These components are calculated from STFT using a 20 ms window. Calculation of gain for each frequency component is based on preserving the relative distance above masked threshold which allows the algorithm to adjust to the time varying qualities of the input speech.

Laura A. Drake and J.C.Rutledge and Jonathan Cohen [101] have developed a wavelet based multiband dynamic range compression technique to compensate for recruitment loudness. This algorithm combines standard compression with intensity level dependent gain calculation. The calculation of the gain is based on TVFD method described earlier. That is, a gain is calculated for each wavelet coefficient in each frequency band, such that the ratio of log intensity above hearing threshold to dynamic range of hearing is the same for the hearing impaired listener as the corresponding ratio is for the normal hearing listener. In this method only three bands namely: 0-1000 Hz, 1000Hz - 2000 Hz, 2000 Hz – 4000 Hz: were used. The compression algorithm used 2:1 compression in each band. The authors have reported that the calculation of the gains will keep the signal within the impaired listener's dynamic range regardless of the intensity level of the input speech, compared to the multiband filter compression systems which are time- invariant. This scheme is also not according to the common audiogram standards.

Multiband syllabic compression systems [64] reduce the variation in speech level in each frequency band according to the subject's reduced dynamic range in that band. Single channel (wide band) systems process the entire speech signal on the basis of the overall level. Although wide band processing can not match a person's hearing profile as well as multiband processing can, wideband processing does not distort the short-tem spectral shape.

The experimental results of previously tested compensation techniques have generally shown that multichannel compression has not improved speech intelligibility relative to well chosen linear amplification. Villchur [111] was among the first researchers to report positive results using multichannel compression in compensating for recruitment. He used a two channel system with the compression ratio and gain in each channel adjusted to compensate for the recruitment in each subject tested. Lippman et al. [104] devised a sixteen channel analog filter system with centre frequencies ranging from 160 – 8 kHz. In all cases, the filter gains were fixed based on the long term comfort levels for speech filtered through each band in isolation. For the most part, the linear systems led to better performance when there was significant word-to-word level variation presented to subjects with more severe losses, and when the input level was much lower than the desired level for all subjects.

Bustamante [114] proposed compression of the principal components of the short term speech spectrum. Results indicated again that linear amplification produced higher intelligibility scores except when the input was presented at a level 15 dB below the most comfortable level.

Quatieri and McAulay [115] reduce the amplitude level variations in speech by dispersing the phase parameter in the sinusoidal model of speech. The output of the system is a waveform with reduced dynamic range during voiced segments and amplified unvoiced regions. Experiments performed on speech in noise showed increased intelligibility using this system.

7.6.1 Conventional Hearing Aids

Block diagram of a generic electronic hearing aid is shown in Fig 7.4. As shown in the figure, conventional hearing aids are composed of a microphone and associated preamplifier, an active filter and a power amplifier, an output transducer or receiver and a power supply or battery. These components are packaged in one of several housings: a body-worn aid, which places all of the component except the output transducer in a shirt-pocket-sized enclosure and features a wire connecting this module to the output transducer module worn in the ear; a BTE (Behind the Ear) aid which houses all of the components in a curved module designed to fit comfortably behind the

ear; and ITE (In the Ear) design which is the least objectionable aid from a cosmetic viewpoint since it fits completely inside of the outer ear and in the case of a canal aid, completely inside the canal.

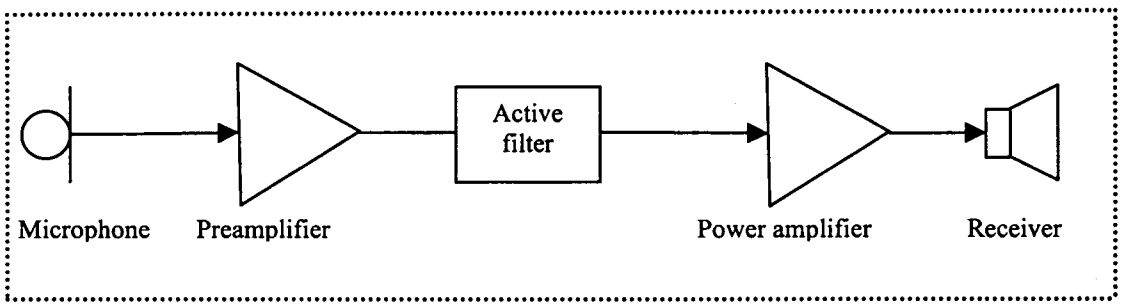


Fig. 7.4: Block diagram of generic electronic hearing Aid

7.6.2 Digital Hearing Aids

The block diagram of a generic digital hearing aid is shown in Fig.7.5. Sound waves picked up by the microphone and transformed into electrical signals are converted to a digital representation by an ADC for processing by a Digital Signal Processor (DSP). The processed signals are converted back to analog by the D to A converter whose output drives the power amplifier. The advantage of this is that hearing compensation algorithms can be implemented on a fast powerful digital signal processor.

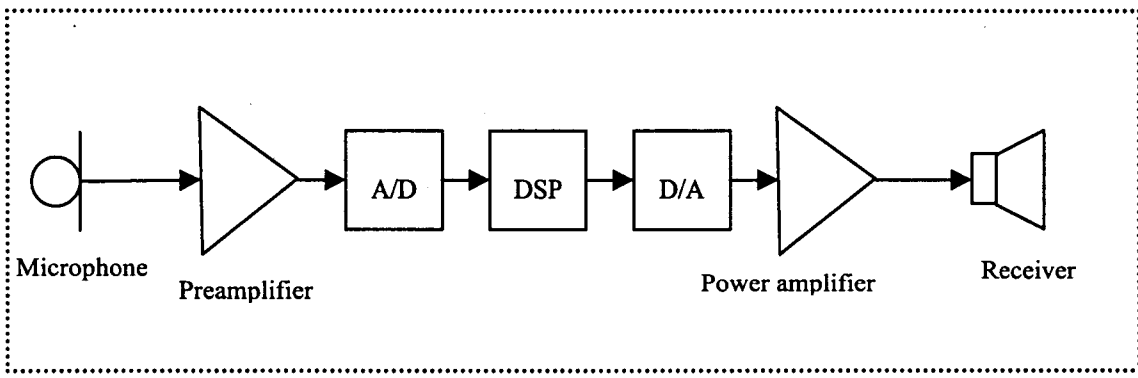


Fig. 7.5: Block diagram of a generic digital hearing aid

The studies on various hearing compensation techniques revealed that hearing compensation techniques can be generally classified into three broad categories:

1. Linear Amplification with AGC : The basic amplifier raises the signal above the threshold of hearing and automatic gain control keeps the signal below the threshold of pain.
2. Amplitude Compression : Based on mapping speech/audio from normal range into the reduced dynamic range of an impaired listener so that the impaired listener will hear a sound at a comparable loudness to that at which the normal listener hears the compressed sound. In this type of compression scheme, low level input may not receive enough boost, and high intensity sounds may be severely clipped.
3. Parametric Compression: TVFD (Time Varying Frequency Dependent) processing technique is based on sinusoidal modelling of speech. Here speech is represented as the sum of sinusoids with various amplitudes, frequencies and phases. These are calculated by STFT over a 20msec window. The sinusoidal amplitude parameters are given compression gain based on listener's thresholds of hearing and pain. Gain is calculated for each frequency components such that the ratio of log intensity above hearing threshold to dynamic range of hearing is the same for the normal and hearing impaired listener. So this is a multichannel amlpitude compression scheme. Wavelet based technique developed by [87] is also same as TVFD discussed above, only difference is that parameters are wavelet coefficients.

Psychoacoustics research revealed that various frequency ranges require amplification in varying degrees for hearing impaired people, especially with sensorineural deafness. Hence an adaptive amplification (non-linear) according to the signal intensity at various frequencies is to be achieved by a hearing aid. The studies on various hearing compensation methods, revealed the need of a compensation algorithm which takes audiogram details (i.e., hearing loss of the patient at various frequencies) as the inputs and then modifies the signal according to the signal intensity and patient's hearing loss at those frequencies. Hence an attempt is made here to design and develop algorithms for enhancement and modification of speech/audio for

the hearing impaired people, using wavelet analysis. The performance of the hearing compensation algorithm using DWT representation for speech/audio signals, is validated through clinical tests on 12 hearing impaired subjects (age 17 years to 78 years) with different types of hearing losses and performance is compared with the performance of conventional hearing aids.

7.7 Speech/Audio Enhancement and Modification using Discrete Wavelet Transform (DWT)

A new hearing aid algorithm is developed in the present work which provides compensation for several types of hearing losses. This algorithm is based on the decomposition of speech/audio signal into various frequency bands matching to the common Audiogram Standard [Fig 7.1] using octave band steps (i.e., 250, 500, 1000, 2000, 40000 and 8000 Hz.). The signal is sampled at 22.05 kHz. Signal decomposition is done using 5 level Discrete Wavelet Transform. The QMF tree structure used for the implementation of DWT is shown in Fig.7.6 The wavelet coefficients cd1, cd2, cd3, cd4, cd5, and ca5 represent the subband samples in the frequency range as shown in the figure. Pure tones marked in the audiogram are 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 40000 Hz and 8000 Hz. In this proposed technique, hearing loss at 250 Hz is taken as the hearing loss at all frequencies in the range 0-350 Hz, hearing loss at 500 Hz is taken as the hearing loss at all frequencies in the range 350-700 Hz and so on.

Gain/attenuation factors required for each band are calculated depending upon the intensity of the signal in that band and hearing loss of the patient in that band. DWT coefficients are modified accordingly. Inverse wavelet transform of the modified DWT coefficients are then taken to reconstruct the modified signal for the impaired. Inputs to this algorithm are the hearing losses at various frequencies as marked in the audiogram of the patient and Un-comfortable level (UCL) of the patient. The proposed algorithm can compensate many types of hearing losses. A major complaint with normal hearing aids is that speech is especially difficult to understand in noisy environments. Hence denoising steps are also included in this algorithm. The new hearing compensation algorithm developed here is briefly given in section 7.7.1.

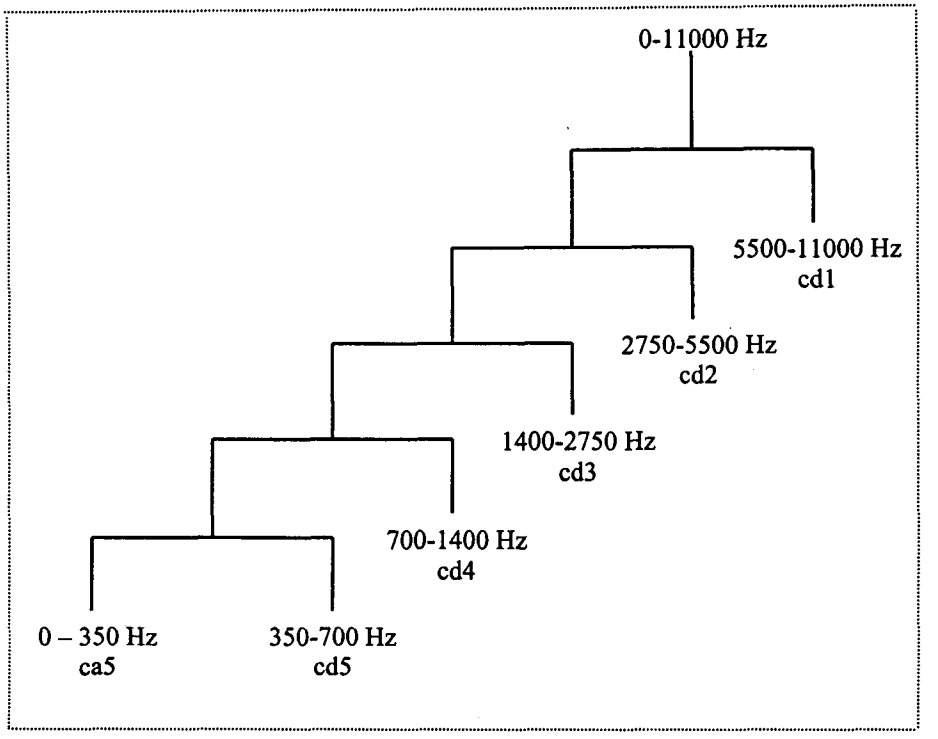


Fig.7.6: DWT tree structure according to the audiogram standard (with octave steps)

7.7.1 Algorithm

1. Read Audiogram Details of the patient (Hearing Losses at various frequencies).
2. Read UCL (Un Comfortable Level or Threshold of Pain).
3. Absolute threshold for the impaired = Absolute threshold for the normal + hearing loss.

   (Absolute threshold for the normal = $T_q(f)= 3.64(f/1000)^{-0.8}- 6.5\exp(-0.6(f/1000-3.3)^2) + 10^{-3}(f/1000)^4$ (dB SPL) at frequencies 250, 500, 1000, 2000, 4000 and 8000 Hz)
4. Perception level = absolute threshold for the impaired + 15 dB.
5. Perform 5 level DWT on the input speech/audio to get approximation coefficients ca5 and detail coefficients cd5, cd4, cd3, cd2 & cd1.
6. Remove noise by thresholding of detail coefficients.(denoising steps are given separately).
7. Calculate the actual Sound Pressure Level (SPL) needed at each frequency band for the impaired (by considering perception level calculated in step 4 and basic amplifier gain).

8. If SPL needed > UCL, then set S P L needed = UCL.
9. Calculate signal intensity in each frequency band from the denoised wavelet coefficients :ca5, cd5,cd4, cd3,cd2 and cd1.
10. Obtain gain / attenuation factors ga5, g5, g4, g3, g2 and g1, to be applied to ca5, cd5,cd4,cd3,cd2 and cd1 to get the actual S P L required.
11. If the intensity of the signal in a band < threshold of the normal hearing threshold, no modification is done on those coefficients (Corresponding gains are kept as 1).
12. Modify ca5, cd5, cd4, cd3, cd2 and cd1 by multiplying with the gains calculated in step 10.
13. Let modified coefficients = mc = [ga5×ca5; g5×cd5 ; g4×cd4 ; g3×cd3; g2×cd2 g1×cd1]
14. Find inverse DWT of the modified coefficients to get the reconstructed signal (modified signal for the patient).

Denoising by Soft Thresholding

The method developed here for the removal of noise is a modified form of the "Denoising method" developed by Donoho [116]. This is a simple thresholding procedure for recovering signals from noisy data. It has three steps:

1. Compute the N level wavelet decomposition of the noisy signal using Mallat algorithm.
2. Thresholding of Wavelet Coefficients: Apply the soft thresholding non-linearity $\eta_t(y) = \text{sgn}(y)(|y| - t)_+$ on the detail coefficients with a threshold value $thr = \sigma\sqrt{2\log(n)/n}$, where σ^2 is the variance of detail coefficients in a band.
3. Reconstruction : Compute inverse wavelet transform based on the original approximation coefficients at level N and the thresholded detail coefficients at all levels. Soft threshold characteristics is shown in Fig.7.7.

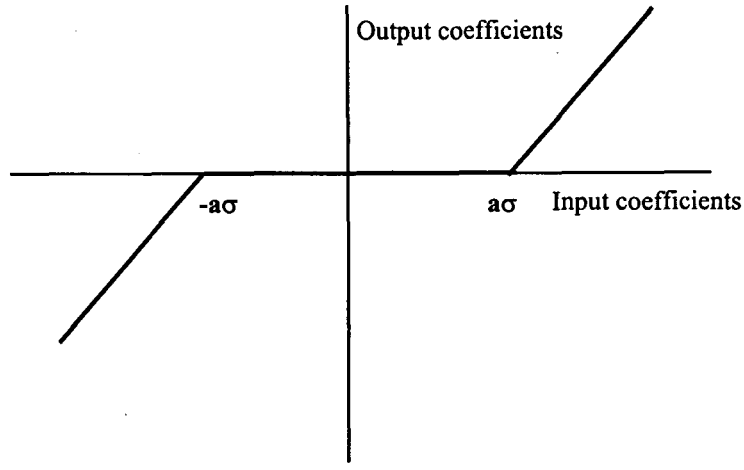


Fig.7.7: Soft thresholding Characteristics

Modification of the above algorithm and its implementation

In the proposed digital hearing aid, signal is decomposed using 'db4' wavelet with 5 levels and the coefficients are obtained as given below.

1. cd1- detail coefficients at level 1.
2. cd2- detail coefficients at level 2.
3. cd3- detail coefficients at level 3.
4. cd4- detail coefficients at level 4.
5. cd5- detail coefficients at level 5.
6. ca5- approximation coefficients at level 5.

σ of the detail coefficients at level 1 is calculated. Threshold is fixed as $thr = \sigma\sqrt{2\log(n)/n}$. All detail coefficients at levels 1,2,3,4,& 5 are soft thresholded with this threshold value. Modification of these denoised coefficients are done according to the algorithm given earlier.

Justification: Sampling frequency of the signal is 22.05 kHz. Since the noise assumed here is white Gaussian type, most of the noise tends to be represented by wavelet coefficients at the finer scales. Level 1 detail coefficients will mainly contain this noise information. Hence the threshold value for soft thresholding is estimated from the first level. No thresholding is done on approximation coefficients since it contains most of the signal information.

The hearing compensation algorithm developed in this work is validated by clinical testing on 12 impaired subjects (age ranging from 17 yrs to 78 yrs). No standard exists for testing the hearing compensation methods. Hence to test the validity of the algorithm developed, each subject was asked to rate the perception and intelligibility of the speech/audio signal with the new hearing compensation algorithm in quiet as well as in the presence of white noise. Analog hearing aid performs linear amplification and the gain required is calculated by taking the average of hearing losses at 1 kHz, 2 kHz, and 4 kHz. Hence, to simulate an analog hearing aid , digitised samples are modified by the calculated gain and then reconstructed and played back using a DAC and a headphone. The tests were conducted at the *Audiology Department, All India Institute of Speech and Hearing (AIISH), Mysore, Karnataka, INDIA* and the results are presented in Table 7.1.

Advantages of the algorithm developed in the present work are:

(i) The calculation of the gain is customised to the common audiogram standard with octave steps.

(ii) The gain for consonants and stops, which contain predominantly high frequencies will be calculated from a smaller window than vowels, which contain more low frequencies, while other multichannel compression schemes use the same window size for each frequency band.

(iii) The gain for each frequency band is calculated according to the intensity of the signal in each frequency band and the hearing loss of the patient in the respective frequency band.

(iv) This method of hearing compensation ensures that the intensity of the modified speech/audio signal always lies within the dynamic range of hearing of the patient.

Original signal, modified signal, DWT coefficients of the original signal, and modified DWT coefficients, for some speech/audio signals in quiet as well as in the presence of white gaussian noise, for two patients, are presented in Figs.7.8-7.23. Corresponding audio signals in *.wav* format, for one patient, are provided in the attached compact disc.

Table 7.1 Test results of the hearing compensation method

Name of the patient	Age (yrs)	Audiogram details		Signal	In quiet		With noise	
		frequency (Hz)	hearing loss(dB)		Analog	New aid	Analog	New aid
Bhagya Lakshmi	63(F)	250 500 1000 2000 4000 8000	55 50 50 55 70 90	Clap Female voice1	Good Good	Good Good	Poor Poor	Good Good
Vijaya Lakshmi amma	70(F)	250 500 1000 2000 4000 8000	50 45 70 95 90 95	Clap Crow Female Voice1	Fair Fair Good	Good Good Good	Poor Poor Poor	Good Good Good
Vishalak-shamma	75(F)	250 500 1000 2000 4000 8000	35 45 50 45 55 85	Clap Crow Female Voice1	Good Good Good	Good Good Good	Poor Poor Poor	Good Good Good
Ananda ram	41(M)	250 500 1000 2000 4000 8000	60 40 50 50 75 60	Clap Crow Female voice1 Evil laugh music	Good Good Good Good Good	Good Fair Good Fair Fair	Poor Fair Poor Poor Poor	Good Good Good Good Good

(Contd..)

Name of the patient	Age (yrs)	Audiogram details		Signal	In quiet		With noise	
		frequency (Hz)	hearing loss (dB)		Analog	New aid	Analog	New aid
Venkate-sh	47(M)	250 500 1000 2000 4000 8000	100 105 80 90 95 100	Female voice1	Good	Good	Poor	Good
				Crow	Good	Good	Good	Good
				Ring	Good	Fair	Good	Good
				Music	Good	Good	Fair	Good
Papanna	54(M)	250 500 1000 2000 4000 8000	65 65 60 80 75 95	Female voice1	Good	Good	Poor	Good
				Crow	Good	Good	Poor	Good
				Ring	Good	Good	Poor	Fair
				Music	Good	Good	Poor	Good
				Clap	Good	Excellent	Poor	Good
Robert Gnanao-liyu	78(M)	250 500 1000 2000 4000 8000	35 35 40 55 65 75	Clap	Good	Good	Poor	Good
				Female Voice1	Good	Good	Poor	Good
				Female Voice2	Good	Good	Poor	Good
				Evil-Laugh	Good	Good	Poor	Good
				Ring	Good	Good	Good	Poor
				Pup	Good	Good	Poor	Good

(Contd..)

Name of the patient	Age (yrs)	Audiogram details		Signal	In quiet		With noise	
		frequenc (Hz)	hearing loss (dB)		Analog	New aid	Analog	New aid
H.S.Naga Raja Shetty	59(M)	250 500 1000 2000 4000 8000	65 70 80 80 70 50	Ring Female voice1 Clap Pup	Good Poor Poor Poor	Poor Good Good Good	Good Poor Poor Poor	Poor Good Good Good
Bilalbi	60(F)	250 500 1000 2000 4000 8000	55 65 60 60 55 80	Clap Pup Female Voice2 Male Voice	Good Good Good Good	Good Good Good Good	Poor Poor Poor Poor	Good Good Good Good
Anitha Reddy	33 (F)	250 500 1000 2000 4000 8000	60 75 90 80 75 65	Clap Pup Female Voice2 Male Voice	Good Good Good Good	Good Good Good Good	Poor Poor Poor Poor	Good Poor Good Good
Kamal-amma	49(F)	250 500 1000 2000 4000 8000	60 55 50 60 65 70	Clap Pup Male Voice	Good Good Good	Good Good Good	Poor Good Poor	Good Poor Good

(Contd..)

Name of the patient	Age (yrs)	Audiogram details		Signal	In quiet		With noise	
		frequenc (Hz)	hearing loss (dB)		Analog	New aid	Analog	New aid
Bharatha	17(M)	250 500 1000 2000 4000 8000	60 80 90 95 100 110	Clap Female Voicel Pup	Good Good Good	Good Good Good	Good Good Good	Good Good Good

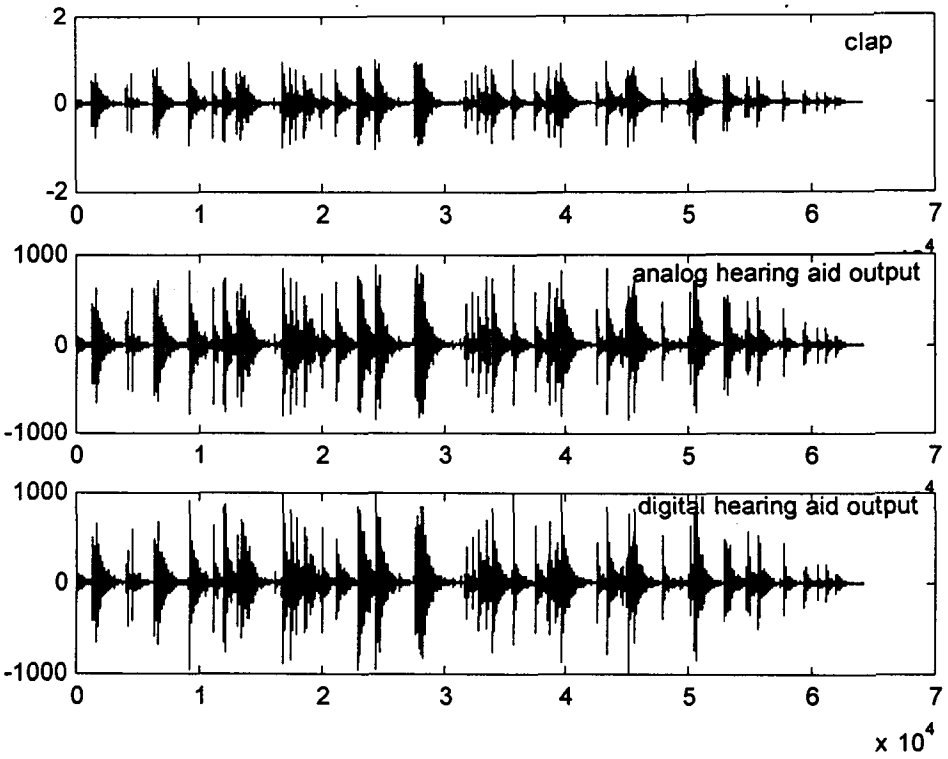


Fig.7.8: Hearing aid output for patient2
(Signal -clap without noise)

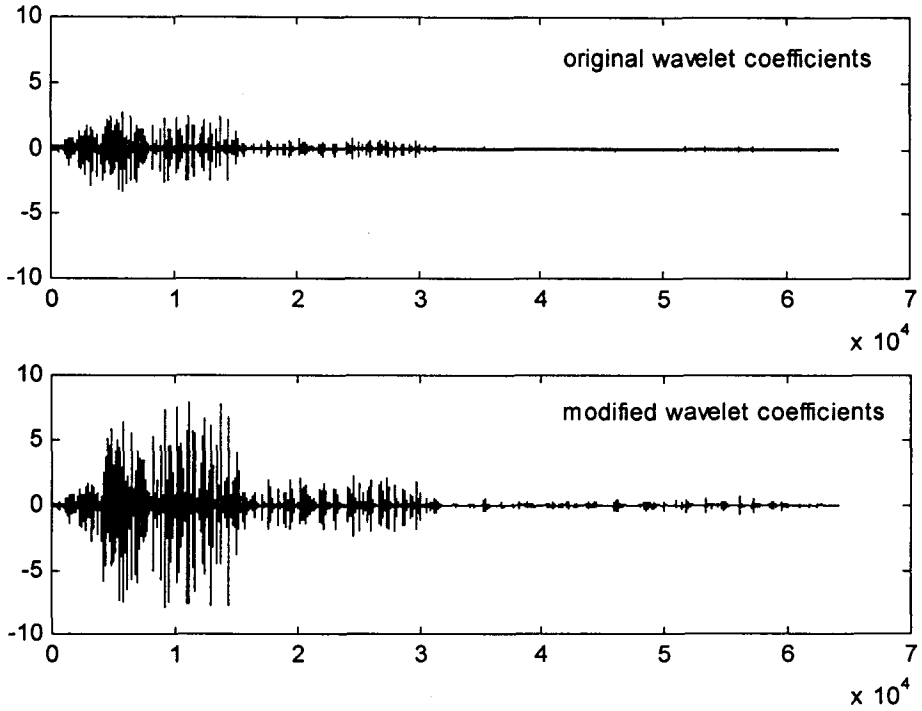


Fig.7.9: Wavelet coefficients before and after modification for patient2
(Signal-clap without noise)

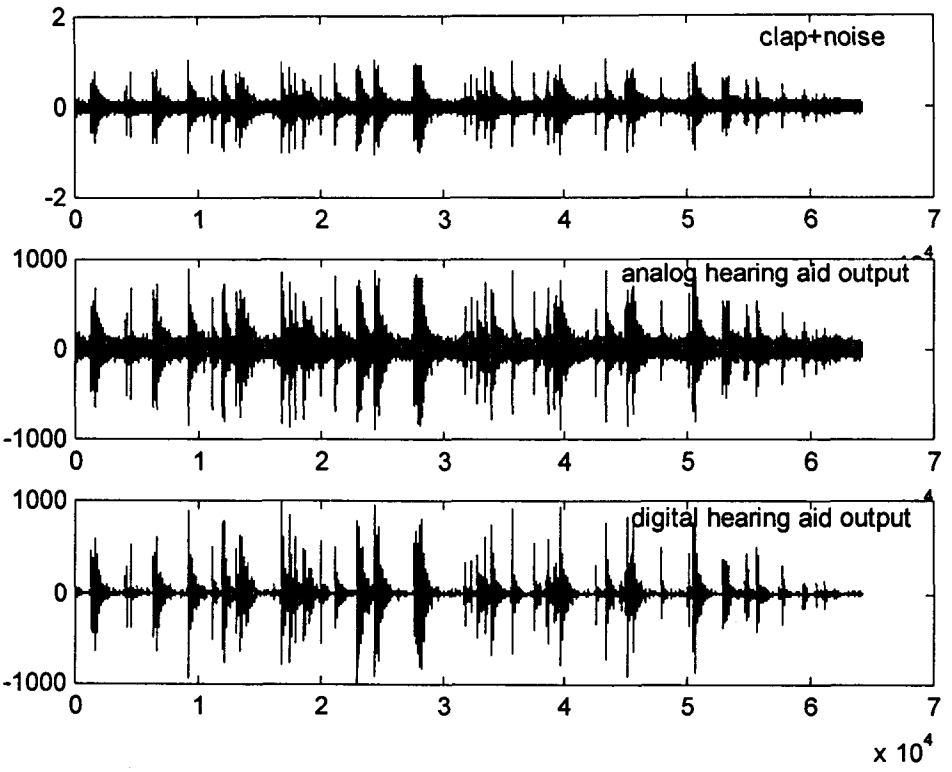


Fig.7.10: Hearing aid output for patient2
(Signal -clap with noise)

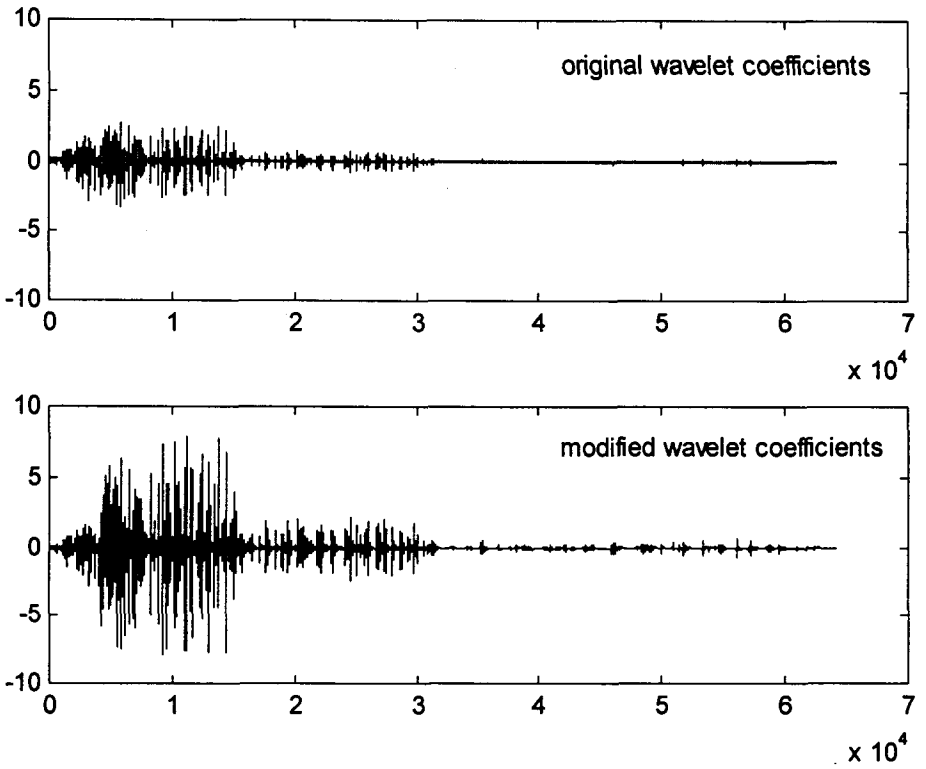


Fig.7.11: Wavelet coefficients before and after modification for patient2
(Signal-clap with noise)

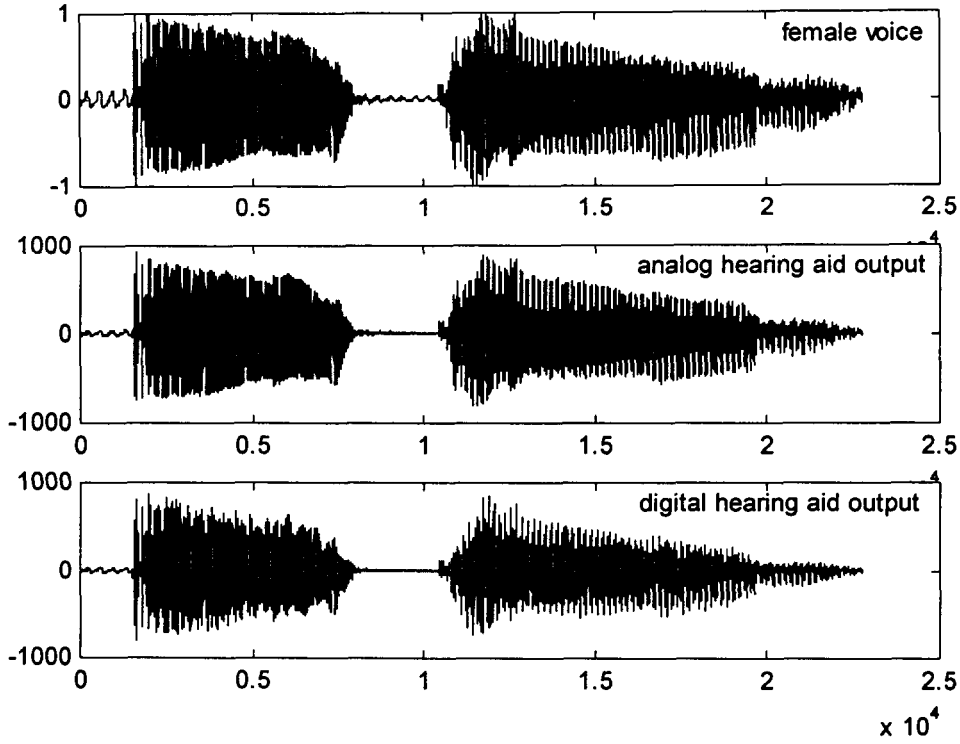


Fig.7. 12: Hearing aid output for patient2
(Signal –female voice without noise)

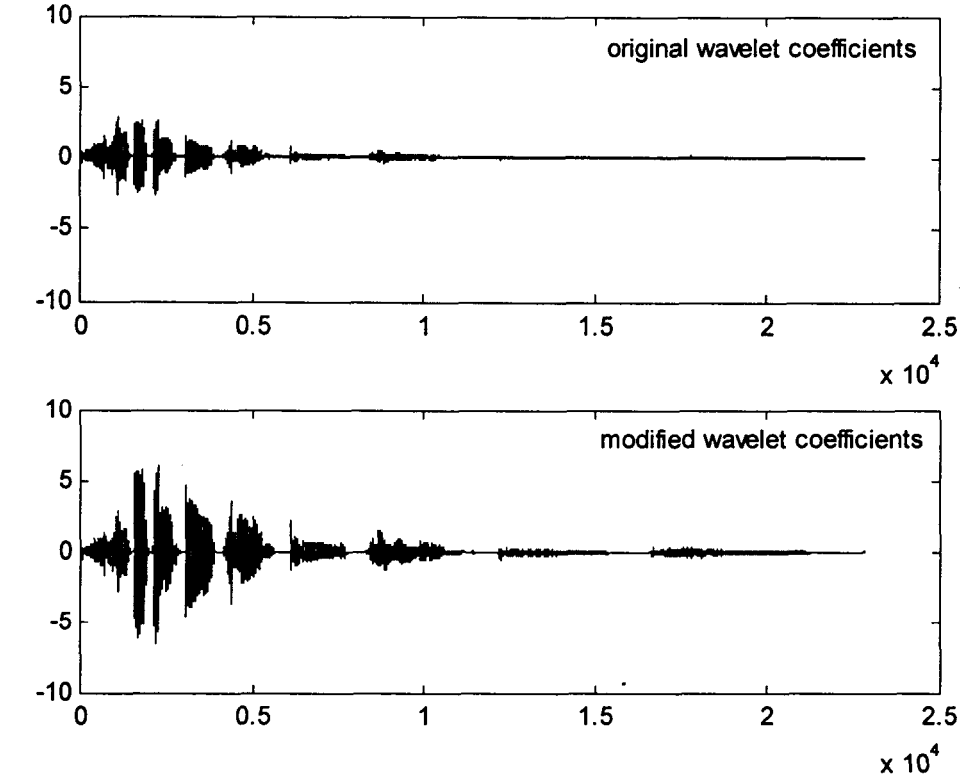


Fig.7.13: Wavelet coefficients before and after modification for patient2
(Signal-female voice without noise)

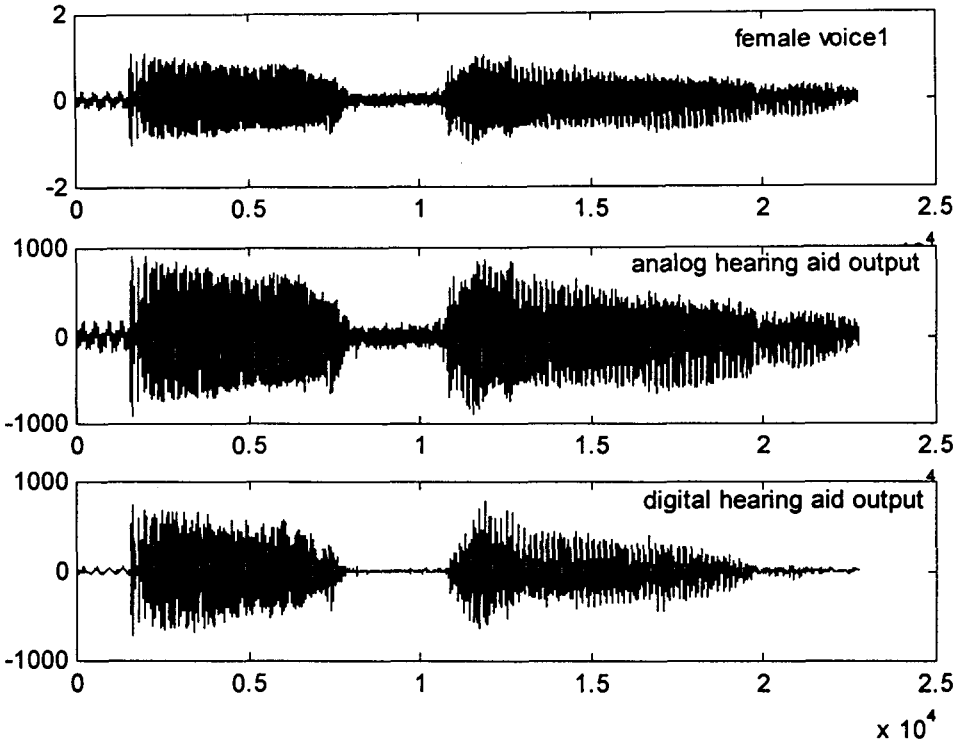


Fig.7.14: Hearing aid output for patient2
(Signal –female voice with noise)

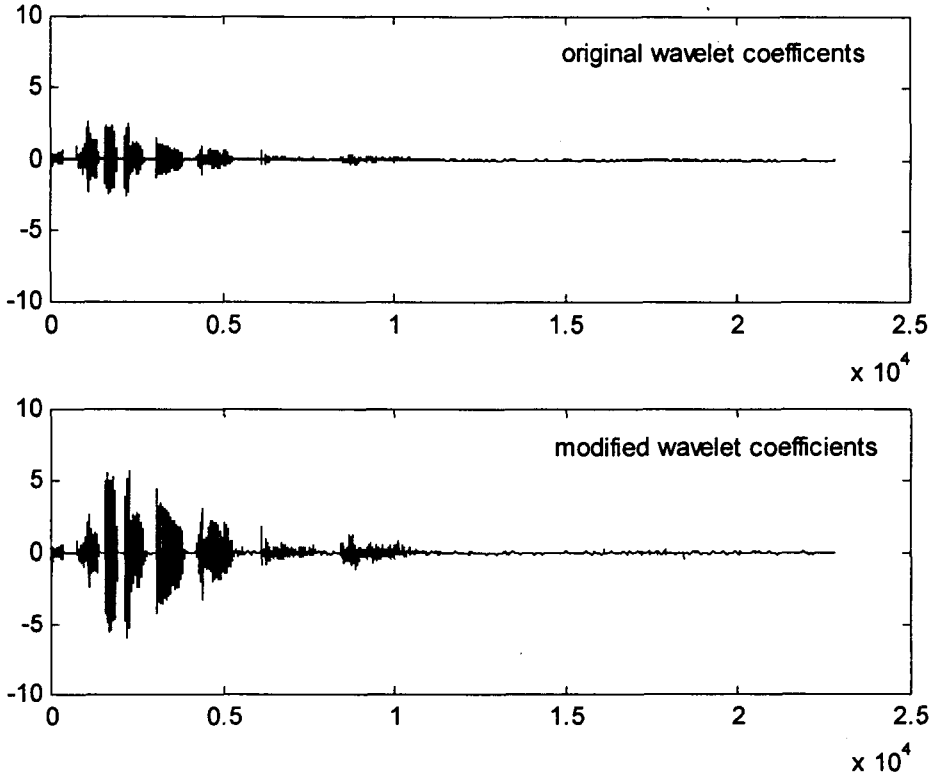


Fig.7.15: Wavelet coefficients before and after modification for patient2
(Signal-female voice with noise)

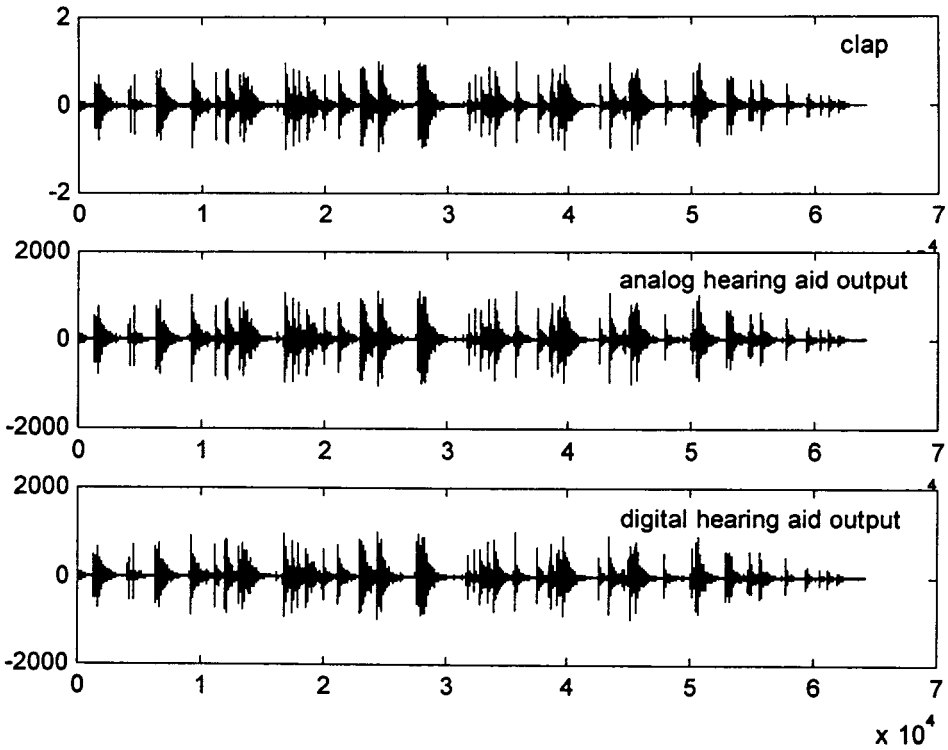


Fig.7.16: Hearing aid output for patient12
(Signal -clap without noise)

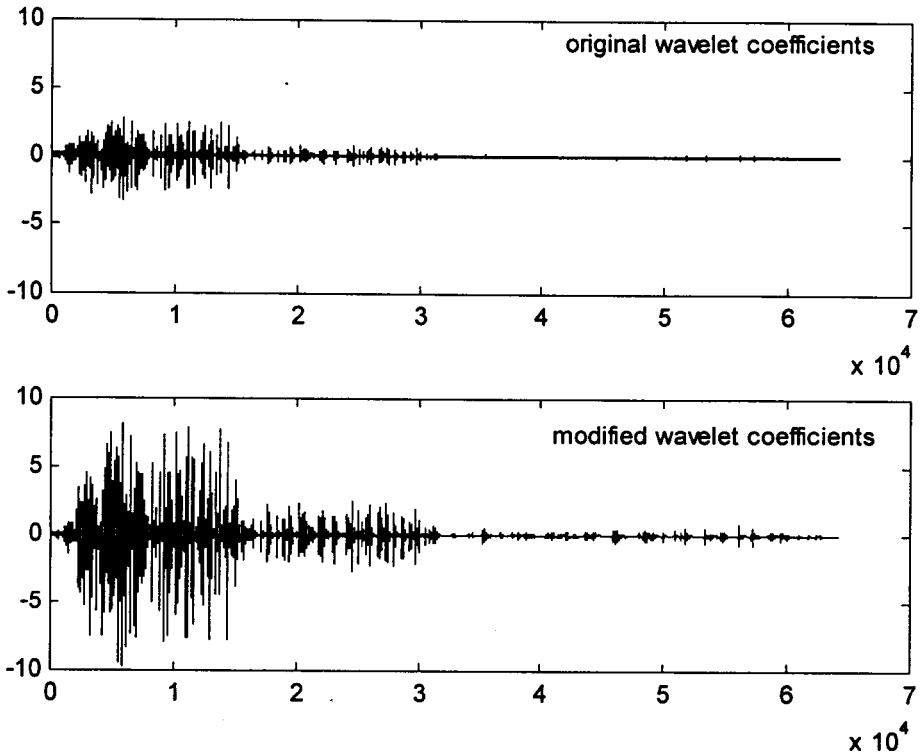


Fig.7.17: Wavelet coefficients before and after modification for patient 12
(Signal-clap without noise)

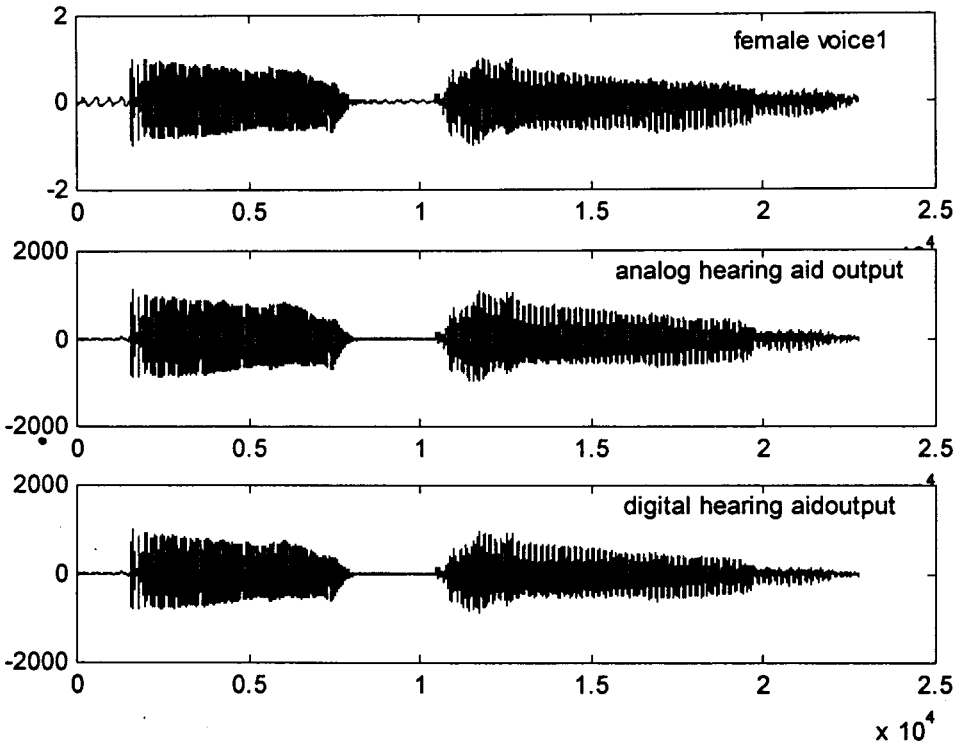


Fig.7.18: Hearing aid output for patient12
(Signal –female voice without noise)

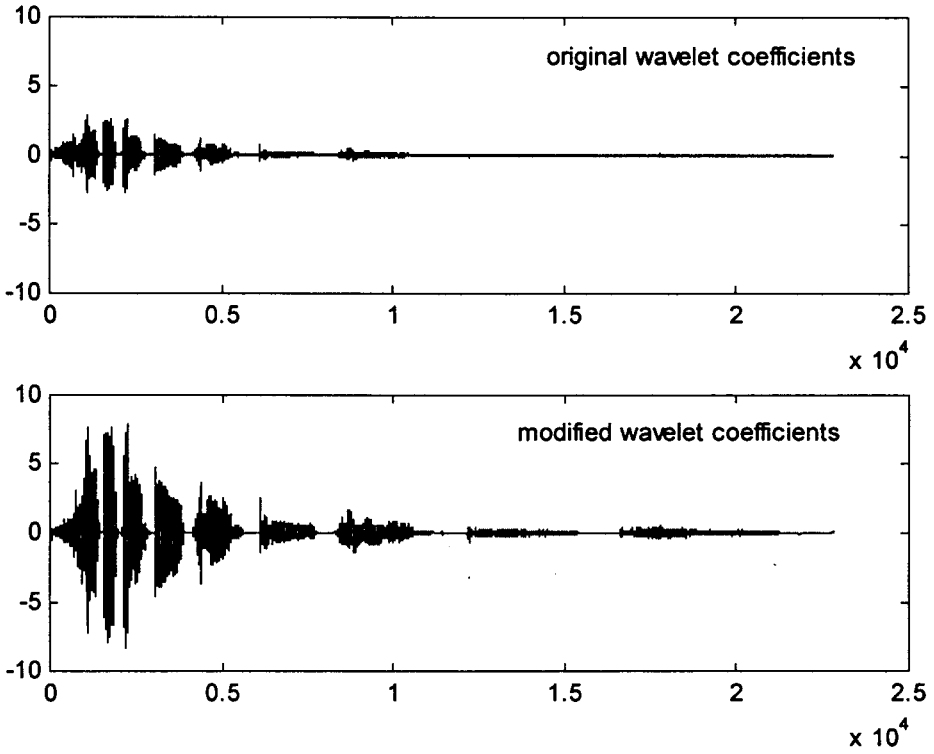


Fig.7.19: Wavelet coefficients before and after modification for patient2
(Signal-female voice without noise)

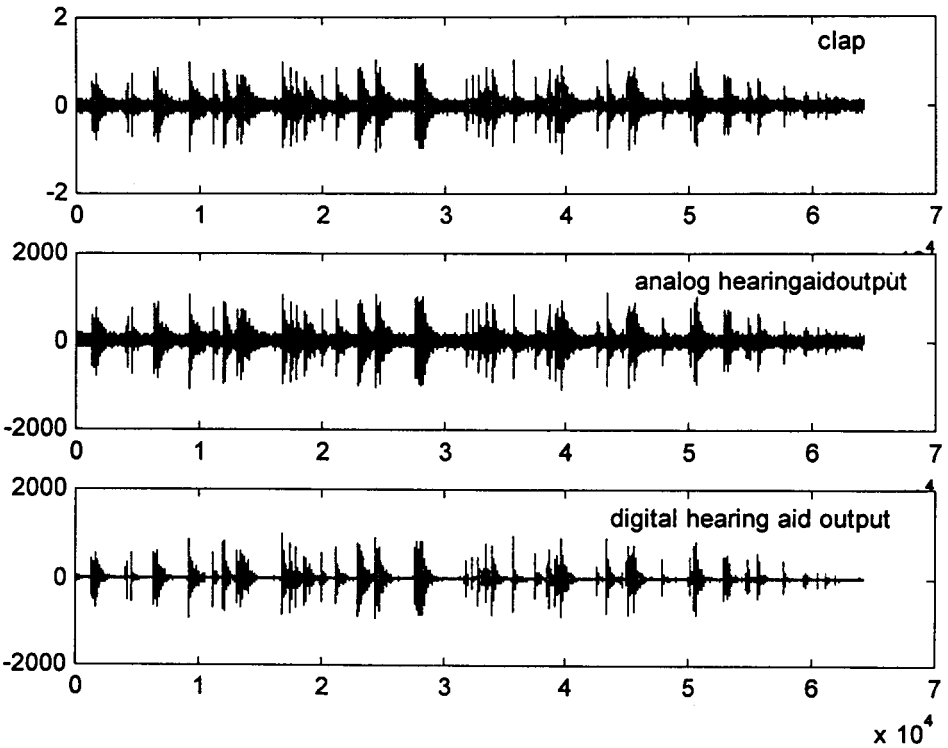


Fig.7.20: Hearing aid output for patient12
(Signal -clap with noise)

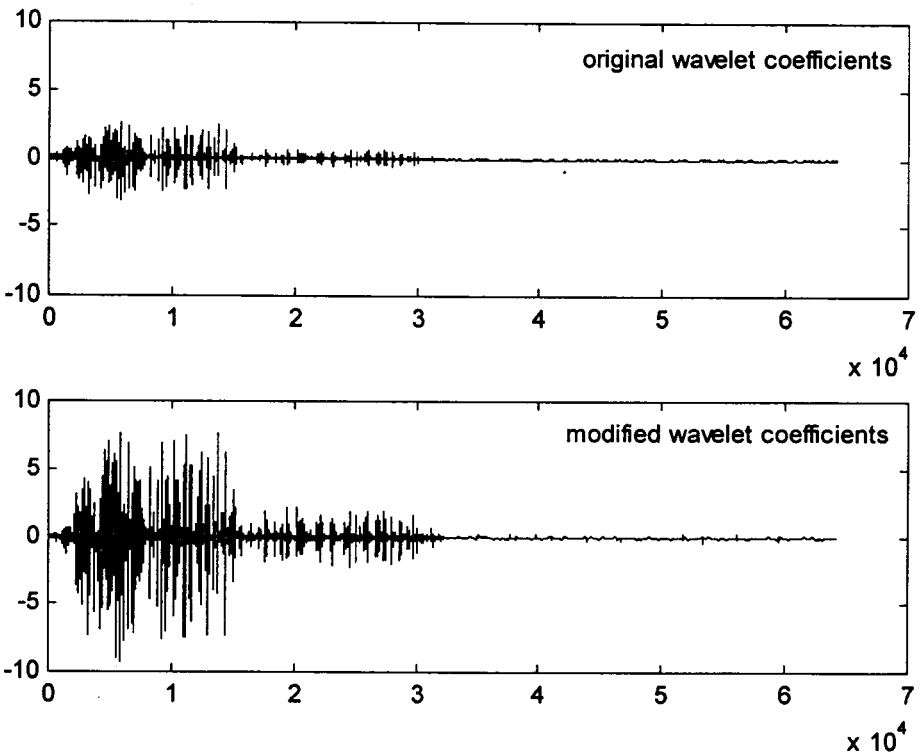


Fig.7.21: Wavelet coefficients before and after modification for patient 12
(Signal-clap with noise)

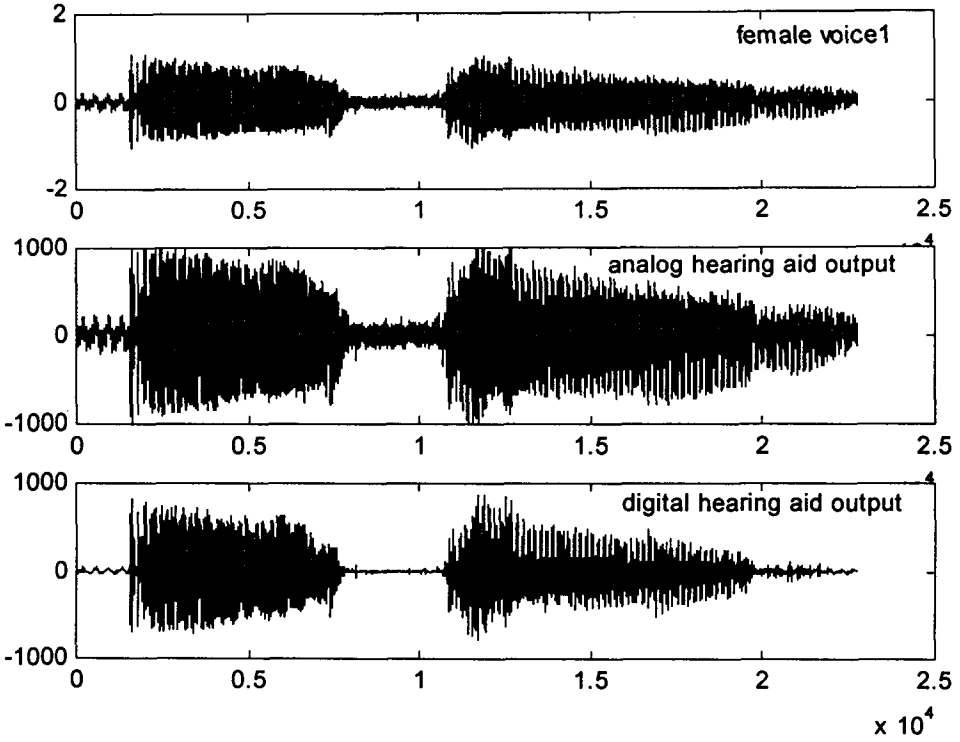


Fig.7.22: Hearing aid output for patient12
(Signal –female voice with noise)

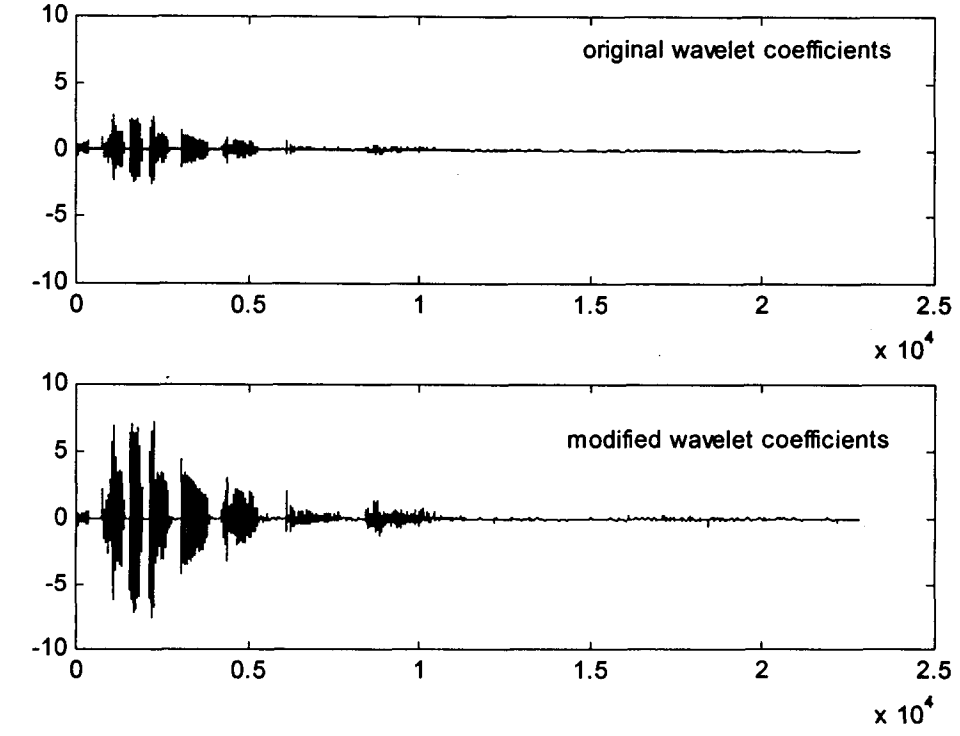


Fig.7.23: Wavelet coefficients before and after modification for patient 12
(Signal-female voice with noise)

7.8 DWP based decomposition

In the present work, a new hearing compensation algorithm using discrete wavelet packet is also developed. The signal is decomposed using DWP, into various frequency bands according to the audiogram standard (with half octave steps) shown in Fig 7.24.

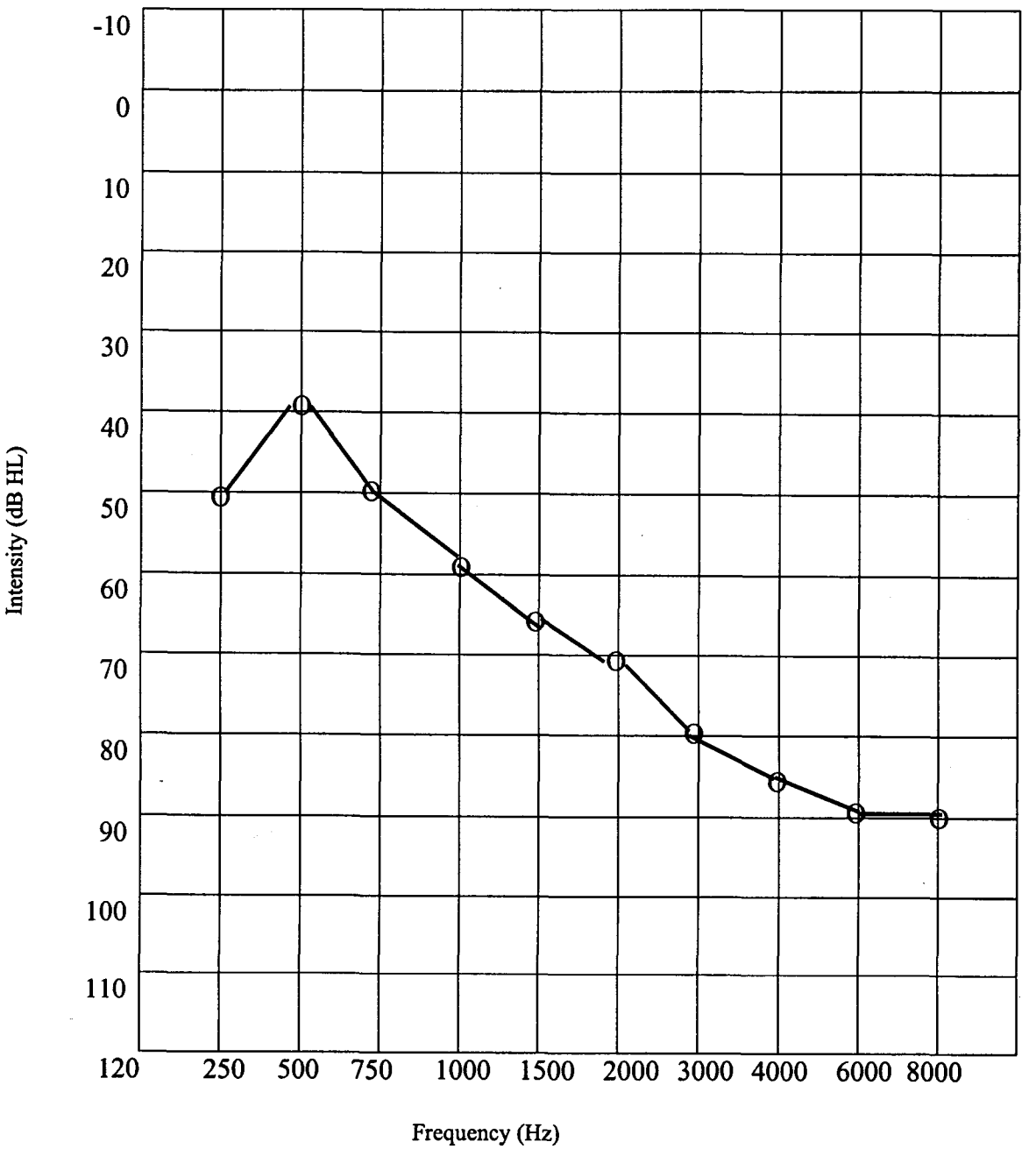


Fig.7.24: Audiogram Standard (Half Octave Steps)

The Wavelet Packet tree structure shown in Fig.7.25 is designed to meet the above audiogram specifications.

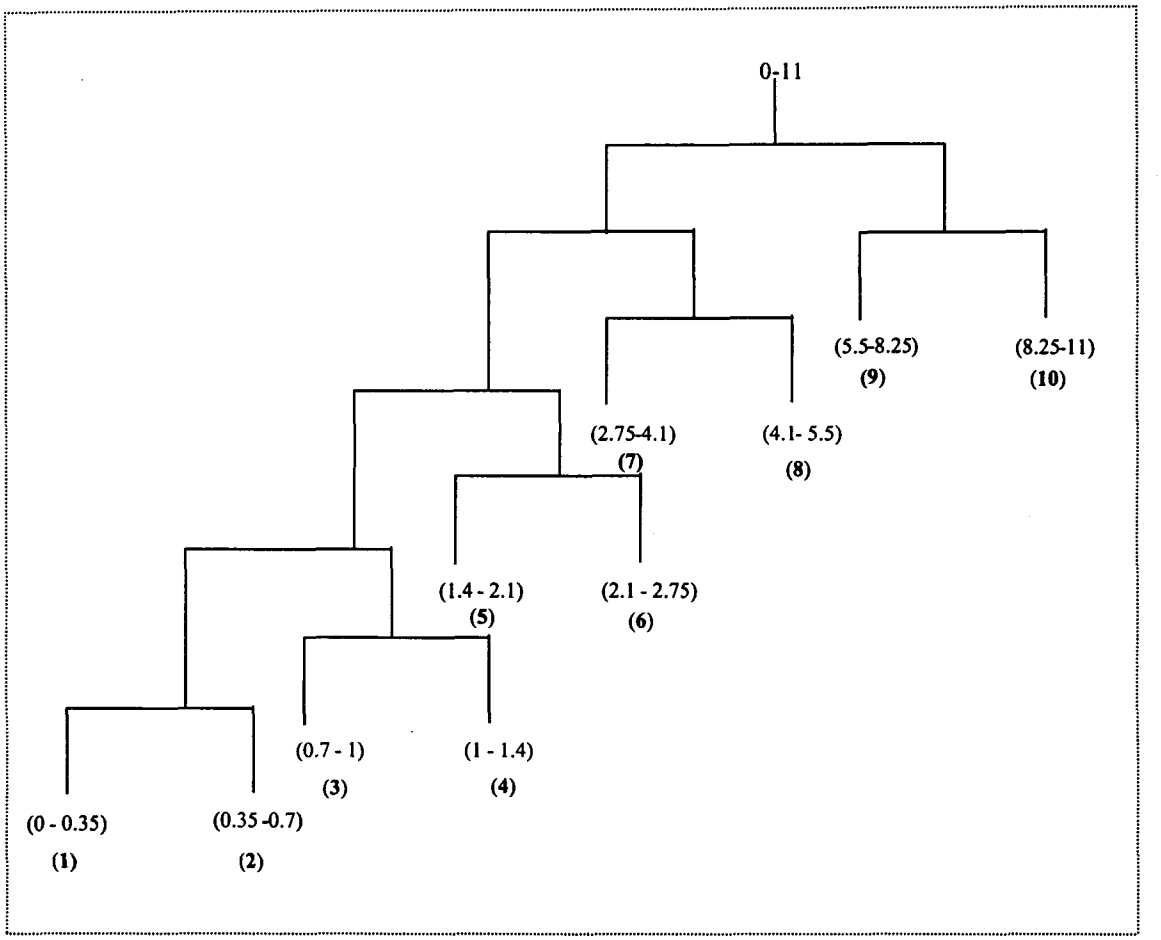


Fig.7.25: Wavelet packet tree structure for digital hearing aid II

Pure tones marked in the audiogram are 250 Hz, 500 Hz, 750 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz and 8000 Hz. In the present algorithm, hearing losses at these frequencies marked in the audiogram are taken as hearing losses in the frequency bands 1,2,3, 4,5,6,7,8,9 & 10 respectively as shown in the Fig.7.25. Gain/attenuation in each frequency band is calculated as explained earlier. Simulated results of the above digital hearing aid are shown in Fig.7.26 by assuming a sample audiogram with details as shown in Table 7.2.

Table 7.2 Audiogram details

Frequency (Hz)	Hearing loss (dB)
250	50
500	40
750	50
1000	60
1500	65
2000	70
3000	80
4000	85
6000	90
8000	90

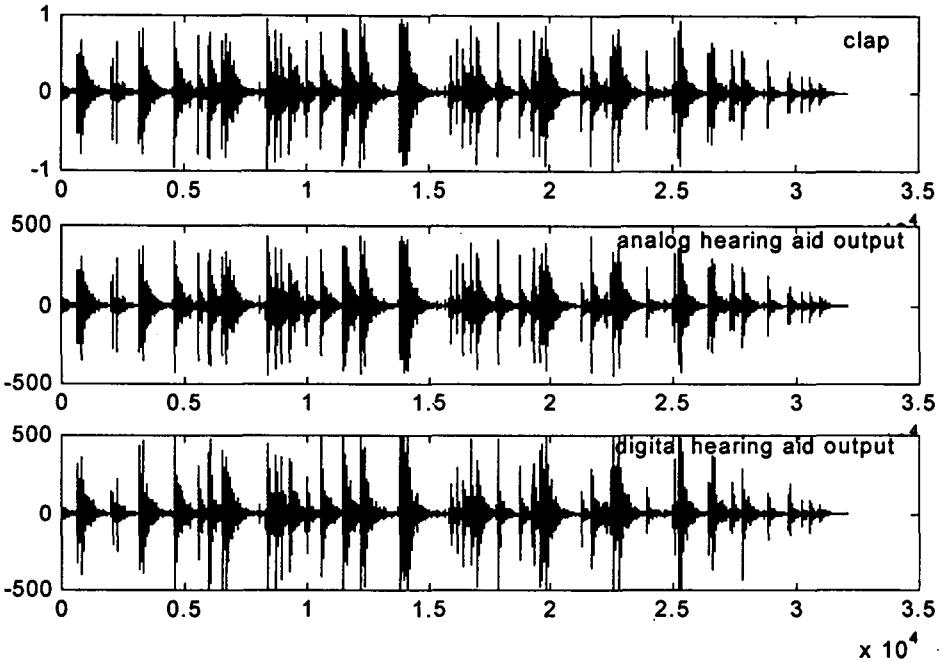


Fig. 7.26: Original signal and hearing aid outputs (Signal-clap without noise)

7.9 Summary

New hearing compensation algorithms using wavelet representation of speech/audio signals have been developed and implemented as a part of the present research work. A major advantage of these algorithms over existing algorithms is that the calculation of gains are customized to the hearing loss of the patient in each frequency band as marked in the standard audiogram and the intensity of the signal in each frequency band. The gain for consonants and stops, which contain predominantly high frequencies are calculated from a smaller window than vowels, which contain more low frequencies, while for other parameterization techniques, the window size is the same for each frequency band. This algorithm also ensures that amplified signal's level is always within the dynamic range of impaired listener.

Background noise is a major problem with hearing aids, because the noise not only masks consonants but also its amplification is distracting and often painful. Hence, a wavelet based denoising technique is also incorporated in this algorithm which serves as a pre-processor for the frequency dependent compensation for hearing impairments. The wavelet coefficients resulting from the denoising stage are used for the gain calculations.

The proposed wavelet approach to the combined problem of noise reduction and frequency dependent compensation for hearing impairments, offers high quality and flexibility since the parameters can be modified to fit the individual hearing loss and intensity of time varying signal characteristics.

Clinical Test results : All subjects rated the quality of the perceived signal with the new digital hearing aid, well above that of analog hearing aid, *in the presence of noise* for majority of the audio signals tested. The performance of this new aid *in quiet* was rated equally with that of analog hearing aid, by most of the subjects. But two subjects rated this new aid well above the analog hearing aid, *in quiet as well as in the presence of noise.* Hence, clinical test results reveal that wavelet analysis is a promising tool for the design of efficient Digital Hearing Aids.

Chapter 8

CONCLUSIONS

8.1 Conclusions

In this thesis, research has been carried out in the following two major aspects of digital speech processing: 1) Perceptual coding of speech and audio for teleconferencing and multimedia applications and 2) Enhancement and modification of speech and audio for the hearing impaired.

New perceptual audio coding schemes using discrete wavelet transforms and discrete wavelet packets have been presented in this thesis as alternatives to ISO/ MPEG international audio coding standards. In MPEG audio coding standards, a uniform filterbank is used for signal decomposition, which is not suitable for the encoding of non-stationary audio signals. Since human ear analyses signals using a non-uniform filterbank (critical bands), full exploitation of perceptual irrelevancies is not possible with uniform filterbanks. A major artifact of MPEG standard and other subband coders using uniform filter banks is "Pre-echo" distortion. However, wavelet based audio coders developed in this research work, offer efficient representation of non-stationary audio segments such as attacks or transients. Further more they allow for efficient exploitation of perceptual irrelevancies also. Hence "Pre-echo" distortion is almost absent in all the proposed coding schemes. Experimental work has included psychoacoustic modelling, bit allocation, various quantization schemes and entropy coding.

Performance of the proposed schemes has been studied with various wavelet families. The properties of wavelets like regularity, orthogonality, frequency selectivity, linear phase property *etc.* have been found to be less relevant in the case of perceptual audio coding. A novel theoretical result, "Coding efficiency or compression

ratio not only depends on the transform or basis function but also depends on the properties/statistical features of the source signal itself" has been experimentally proved here. Three optimisation methods have been developed as part of the present research work, to select the optimum wavelet basis from a predefined library of wavelets, for each frame (512 samples) of the audio source, according to the statistical features or local properties of the signal. Stationary audio segments have been more compactly represented using sinusoidal basis (DCT) than wavelet basis (DWT and DWP). This is due to high frequency resolution of polyphase filterbank used for the implementation of DCT. Hence, a novel switching scheme has also been developed to switch between sinusoidal and optimum wavelet basis according to the time varying characteristics of the audio source.

Performance of the proposed schemes have been tested for a number of audio sources covering a wide range of audio signals like male speech, female speech, instrumental music, vocal music, sounds of animals, birds, machines and other nature sounds. All the schemes have been implemented with 1) Scalar quantization 2) A new vector quantization scheme (named as 'hit book method') and 3) Combined scalar and vector quantization. The Compression Ratio (CR) and Mean Opinion Score (MOS) obtained are comparable to MPEG audio coding standard. By the introduction of the novel switching scheme and optimal wavelet basis selection, each audio frame has been represented using an optimum basis to achieve maximum compression ratio with almost transparent quality.

Wavelet based hearing compensation techniques (enhancement and modification of speech and audio for the hearing impaired people) have also been developed as a part of the present research work. New algorithms for digital hearing aids using Discrete Wavelet Transforms and Discrete Wavelet Packets matching to the common audiogram standards (ISO) (using octave steps and half octave steps) have been developed, implemented, and validated by clinical testing on a number of patients. Actual gain needed for each frequency band is calculated; according to the hearing loss of the patient and intensity of the audio signal in that band. The proposed algorithms ensure that intensity of the modified speech and audio signals always lies within the dynamic range of hearing of the patient.

8.2 Summary of Results

Discrete Wavelet Transform based audio coder, developed in Chapter 4 is less complex, and its performance is comparable to MPEG audio coding standards (Layers I & II). Discrete wavelet transform is implemented using the fast Mallat algorithm. Signal is decomposed into nine subbands using 8 level DWT. Bandwidth of these subbands increases as the frequency increases (in the same way as in the human ear). In this scheme, some of the subbands cover more than one critical band and the masking threshold for each subband is taken as the minimum of masking thresholds calculated for the various critical bands in that subband. Hence, efficient exploitation of perceptual irrelevancies could not be achieved to attain maximum compression. This drawback has been eliminated in Discrete wavelet packet based coding schemes, developed in Chapter 5. However, the computational complexity of the analysis filterbank is more in these schemes. In the first scheme, the signal is decomposed using DWP into 27 subbands closely mimicking the critical bands of the human ear. Bits are allocated to various subbands according to the masking thresholds calculated by the psychoacoustic model. If computational complexity and hence encoding delay is the major constraint, DWT based audio coder developed in Chapter 4 is preferred over this scheme using wavelet packet transform.

MPEG audio coding scheme, DWT based coding scheme proposed in Chapter 4 and the first scheme using wavelet packet analysis use separate high resolution FFT stage for psychoacoustic model implementation. When these schemes are to be implemented in real time, two Digital Signal Processors need to be employed. Hence, a low complexity (computationally more efficient) Discrete Wavelet Packet based audio coding scheme, in which psychoacoustic model is directly driven by DWP coefficients from analysis filterbank, has been developed and implemented. In this scheme, Mean Opinion Score is slightly less than that of the previous scheme since the frequency resolution of DWP is less than that of FFT.

Perceptual audio coding schemes, namely, DWT based coder, schemes 1 and 2 using discrete wavelet packets, employ analysis filterbank suitable for audio signals sampled at 44.1 kHz and hence mainly applicable for coding hi-fi music. Hence, a

flexible DWP based audio coder has been also developed in Chapter 5 to support three industrial audio sampling frequencies namely, 11.025 kHz, 22.05 kHz and 44.1 kHz. This codec is suitable for an advanced user. The codecs discussed earlier are of variable bit rate type. A constant bit rate discrete wavelet packet based audio coder has been developed and implemented in Chapter 6. SNR scalability feature has also been incorporated into this scheme by using two stages of compression. A user can select any one of the above coding schemes according to the available bandwidth of the channel, preferred application, the processing power available, available storage, quality to be met, price that can be paid *etc.*

Performance of all the proposed audio coding schemes has been validated through subjective listening tests. Attached Compact Disc (CD) contains some of the audio signals used in the present work as well as coded and reconstructed signals using the proposed schemes. All audio files are stored in *.wav* format and the details of the signals are given in Appendix B.

New hearing compensation methods developed in Chapter 7 is based on signal decomposition using wavelet analysis. The first method, using discrete wavelet transform, splits the signal into six frequency bands matching with the common audiogram standard (ISO) with octave steps. The second method, using discrete wavelet packets, splits the signal into ten subbands matching with the audiogram standard using half octave steps. The advantage of these methods is that, the only input to be fed by an audiologist are hearing losses of a patient at various frequencies as marked in the audiogram and threshold of pain or Un-Comfortable Level (UCL) of the patient. The algorithms will calculate the amplification needed for that patient according to the intensity of the signal in each frequency band and hearing loss of the patient in that band. Steps for removing noise have also been incorporated in these algorithms. Hence, the proposed wavelet approach to the combined problem of noise reduction and frequency dependent compensation for hearing impairments; offers high quality and flexibility since the parameters can be modified to fit the individual hearing loss, and intensity of time-varying-signal characteristics. Clinical test results reveal that wavelet analysis is a promising tool for the design of efficient digital hearing aids.

8.3 Major Contributions of the Thesis

- Replacement of Time/ frequency block (subbands with uniform bandwidth) of MPEG international standard by non –uniform filterbanks matching to both the signal characteristics and the perceptual properties of human auditory system to achieve compact representation and to eliminate "Pre-echo distortion".
- Development of discrete wavelet transform based audio coder using nine subbands emulating the analysis property of human auditory system.
- Development of discrete wavelet packet based audio coder using twenty seven subbands closely mimicking the human auditory system.
- Design of a flexible wavelet packet tree structure to incorporate three industrial audio sampling frequencies namely 11.025 kHz, 22.05 kHz and 44.1 kHz.
- Psychoacoustic model implementation using discrete wavelet packet to reduce the computational complexity of the wavelet packet based audio coders.
- Selection of optimum wavelet basis from a predefined library of wavelets, for perceptual audio coding, according to the local characteristics or statistical features of the audio signal.
- Design of a new vector quantization scheme (named as "Hit book method") in which length of the code book can be adaptively changed according to the masking thresholds calculated by the psychoacoustic model.
- New hearing compensation techniques using discrete wavelet transform and discrete wavelet packet both conforming to the common audiogram standards (ISO) (octave steps and half octave steps).

8.4 Scope for Future Work

- In the design of psychoacoustic model, temporal masking properties of the human ear can be incorporated to achieve higher degree of compression.
- The psychoacoustic model and the bit allocation part of the encoder can be substituted by a look-up table. The entries in the look-up table can be based on the mean power of each subband for a variety of audio signals and masking

threshold in each subband, such that the look-up table stores the number of bits allocated to each subband. Thus the whole psychoacoustic model calculations and bit allocation process can be simplified. This will reduce the computational complexity and processing time of the encoder significantly.

- In the case of digital hearing aid algorithms, steps for removing speech babble noise can also be incorporated. Speech babble noise occurs when a number of persons speak simultaneously and continuously.

BIBLIOGRAPHY

1. Ted Painter, Andreas Spanias, "Perceptual Coding of Digital Audio", Proceedings of the IEEE, Vol.88, No.4, pp.449-513, April 2000.

2. Peter Noll, "Wide band speech and audio coding", IEEE communication magazine, November 1993.

3. Tsuban Chen, " The Past, Present and Future of Audio Signal Processing", IEEE Signal Processing Magazine, pp.30-57, September 1997.

4. Davis Pan, " Digital Audio Compression", Digital Technical Journal,, Vol.5, No.2, Spring 1993.

5. Allen Gersho, "Advances in speech and audio compression", Proc.of IEEE, vol.82, No.6, pp.900 – 918, June 1994.

6. Richard V.Cox and Peter Kroon, "Low bit rate speech coders for multimedia communication", IEEE communication magazine, pp.34-40, December 1996.

7. A.M.Kondoz, "Digital Speech : Coding for Low bit rate communication system", John Wiley & Sons, 1994.

8. Thomas Parsons " Voice and Speech Compression", McGraw Hill Book Company, 1986.

9. Lawrence R.Rabiner and Ronald W.Schafer, "Digital Processing of speech signals", 1978, Prentice Hall Inc.

10. Sadaoki Furui, M.Mohan Sondhi, "Advances in Speech Signal Processing", Marcel Dekker, Inc., 1991.

11. Panos E.Papamichalis, "Practical Approaches to speech coding", Prentice-Hall, Inc.,1987.

12. N.S.Jayant, Peter Noll, "Digital Coding of Waveforms: Principles and applications to speech and video", Prentice – Hall, Inc., 1984.

13. Andreas S. Spanias, "Speech Coding: A tutorial Review", Proc. IEEE, vol.82, No.10, pp.1541-1582, October 1994.

14. Nikil Jayant, "Signal Compression : Technology, Targets Research directions",IEEE journal on Selected Areas in Communications, vol.10, No.5, pp.796 – 818, June 1992.

15. Nikil Jayant, James Johnston and Robert Safranek, " Signal Compression based on Models of Human Perception", Proc.IEEE, vol.81, No.10, pp.1385 – 1444, October 1993.

16. Allen Gersho, Robert M.Gray, "Vector quantization and signal compression", Kluwer Acaemic Publishers,1992.

17. Mark Nelson, Jean – Loup Gailly, "The Data Compression Book", BPB Publications, 1996.

18. Martin Cooke, Steve beet and Makolm Crawford, "Visual representations of speech signals",, John Wiley & sons,1993.

19. David J.M.Robinson, "The human auditory system", *www.essex.ac.uk*

20. E.Swicker, H.Fastl, "Psychoacoustics – Facts and Models", Springer Verlag, 1999.

21. Talbot Smith , "Audio Engineering Reference Book".

22. J.Johnston, " Transform Coding of Audio Signals using Perceptual Noise Criteria", IEEE Journal on Selected Areas on Communications, Vol.6, pp.314 – 323, Feb.1988.

23. ISO/IEC 11172-3, " Information Technology – Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbits/s – Part 3: Audio", August 1993.

24. D.Krahe,,"New Source coding method for high quality digital audio signals", NTG Fachtagung Hoerrundfunk, pp.S.371-S.381, 1985.

25. E.F.Sschroder and W.Voessing," High quality digital audio encoding with 3.0 bits/sample using adaptive transform coding", in Proc. 80th Conv. Aud. Eng. Soc., Mar.1986, preprint 2321.

26. K.Brandenburg, " OCF – A new coding algorithm for high quality sound signals", Proceedings of ICASSP, pp.5.1.1 – 5.1.4,May 1987.

27. K.Brandenburg,"High quality sound coding at 2.5 bits/sample", in Proc.84th Conv.Aud. Eng. Soc., March 1988, preprint 2582.

28. K.Brandenburg, "OCF; Coding high quality audio with data rates of 64 kbit/sec," in Proc.85th Conv.aud.Eng.Soc., March 1988, preprint 2723.

29. K.Brandenburg and J.D.Johnston, " second generation perceptual audio coding: The hybrid coder", Proc.88th Conv. Aud. Eng. Soc., March 1990, preprint 2937.

30. Y.Mahieux, J.Petit and A.Charbonnier, " Transform Coding of Audio Signals using correlation between successive transform blocks", Proc. Int.Conf. Acoustics, Speech and Signal Processing, ICASSP, May 1989, pp.2021-2024.

31. Y.Mahieux and J.Petit, " Transform coding of Audio signals at 64 kbits/sec", Proc. Globecom, Nov.1990, pp.405.2.1 – 405.2.5.

32. M. Paraskevas and J.Mourjopoulos, " A differential perceptual audio coding method with reduced bit rate requirements", IEEE Transactions on Speech and Audio Processing, pp.490 – 503, Nov.1995.

33. W.Y.Chan and A.Gersho, " High fidelity audio transform coding with vector quantization", Proceedings of ICASSP, pp.1109 – 1112, May 1990.

34. W.Chan and A.Gersho, "Constrained storage vector quantization in high fidelity audio transform coding", Proceedings of ICASSP, pp.3597 – 3600,,, May 1991.

35. N.Iwakami, T.Moriya and S.Miki, "High quality audio coding at less than 64 kbit/s by using transform domain weighted interleave vector quantization (TWINVQ),", Proc. ICASSP, pp.3095 – 3098, May 1995.

36. G.Theile, G.Stoll, and M.Link, " Low- bit rate coding of high quality audio signals", Proc.82nd Cov.Aud.Eng.Soc., Mar.1987, preprint 2432.

37. G.Stoll,,, M.Link and G.Theile, "Masking-pattern adapted subband coding : Use of the dynamic bit-rate margin ," Proc.84th Cov.Aud.eng.Soc., Mar.1988, preprint 2585.

38. R.N.J.Veldhuis, " Subband coding of digital audio signals without loss of quality", Proc.of ICASSP, pp.2009 – 2012, May 1989.

39. Y.F.Dehery, M.Lever and P.Urcun, "A MUSICAM source codec for digital audio broadcasting and storage", Proc.ICASSP, pp.3605 – 3608, May 1991.

40. Davis Pan, " A Tutorial on MPEG Audio Compression", *http://bok.net/~pan/*

41. Peter Noll, " MPEG Digital Audio", IEEE Signal Processing Magazine, pp.59-80, September 1997.

42. Sinha Deepen, Tewfik Ahmed H, " Low Bit Rate Transparent Audio Compression using Adapted Wavelets", IEEE Transactions on Signal Processing, Vol.41, No.12, pp.3463 – 3479, December1993.

43. Pramila Srinivasan, Leah H.Jamieson, "High Quality Audio Compression using an adaptive wavelet packet decomposition and psychoacoustic modelling", IEEE Transactions on Signal Processing, Vol.46, No.4, pp.1084 – 1093, 1998.

44. P.Philippe, F.Moreau de Saint Martin and M.Level, 'Wavelet filterbanks for low time delay audio coding", IEEE Transactions on Speech Audio Processing, Vol.7, pp.310 – 322, May 1999.

45. K.Hamdy, M.Ali and A.Tewfik, "Low bit rate high quality audio coding with combined harmonic and wavelet representations", Proc. ICASSP, pp.1045-1048, May 1996.

46. J.Princen and J.Johnson, "Audio Coding with signal adaptive filter banks", Proc.ICASSP, pp.3071 – 3074, May 1995.

47. B.Edler, "Technical description of the MPEG-4 audio coding proposal from University of Hannover and Deutsche Bundespost Telekom", ISO/IEC, JTC1/SC29/WG11 MPEG 96/MO632, Jan.1996.

48. H.Purnhagen, B.Edler and C.Ferekidis, "Object based Analysis/synthesis audio coder for very low bit rates", Proc. 104th Conv. Aud. Eng. Soc., May 1998, preprint 4747.

49. S.Singhal, "High quality audio coding using multipulse LPC", Proc. ICASSP, pp.1101-1104, 1990.

50. S.Boland and M.Deriche, " Hybrid LPC and Discrete Wavelet Transform audio coding with a novel bit allocation algorithm", Proc. ICASSP, pp.3657 – 3660, May 1998.

51. T.Yoshida, " The Rewritable Mini Disc System", Proc.IEEE, pp.1492 – 1500, Oct 1994.

52. R.McAulay and T.Quatieri, "Speech Analysis Synthesis based on a sinusoidal representation", IEEE Transactions on ASSP, vol.34, pp.744 – 754, Aug1986.

53. T.Verma and T.Meng, " Sinusoidal modelling using frame based perceptually weighted matching pursuits", Proc.ICASSP, pp.981 – 984, March 1999

54. M.Davis," The AC-3 multichannel coder", Proc.95th Conv.Aud.Eng.Soc., Oct.1993, preprint 3774.

55. J. Johnston, D.Sinha, S.Dorward and S.Quackenbush, "AT&T perceptual audio coding (PAC)", Collected papers on Digital Audio Bit-rate Reduction, pp.73-81,1996.

56. P.Duhamel, Y.Maheiux, J.P.Petit, "A fast algorithm for the implementation of filter banks based on time domain aliasing cancellation", Proc. of ICASSP, 1991.

57. Emmanuel C. Ifeachor, Barrie W.Jervis, "Digital Signal Processing – A practical approach", Addison – Wesley, 1993.

58. Ali.N.Akansu, "Multiresolution Signal Decomposition", Academic Press, Inc.
59. Gilbert strang, Truong Nguyen, "Wavelets and Filter Banks", Wellesley – Cambridge Press, 1996.

60. J.Princen, J.Johnston and A.Bradley, " Subband / Transform Coding using filterbank designs based on time domain aliasing cancellation",Proc.ICASSP, pp.50.1.1 – 50.1.4, May 1987.

61. Karlheinz Brandenburg, "MP3 and AAC explained", AES 17th Int.conf.on High quality audio coding.

62. B.G.Lee, "A new algorithm to compute the discrete cosine transform", IEEE Trans. On ASSP, Vol. ASSP –32, No.6, pp.1243-1245, Dec.1984.

63. William A.Yost and Donald W.Nielson, "Fundamentals of Hearing", Holt.Inc.,1985.

64. Brian C.J.Moore, "An introduction to the Psychology of Hearing", Academic Press, 1997.

65. Leo L.Beranek, " Acoustics", McGraw Hill Book company, 1954.

66. Jaideva. C.Goswamy, Andrew K.Chan, "Fundamentals of Wavelets", John Wiley & Sons, Inc., 1999.

67. C.Sydney Burrus, Ramesh A.Gopinath and Haitao Guo, "Introduction to Wavelets and Wavelet Transforms", Prentice Hall International, Inc.,1998.

68. Michael W.Frazier, " An Introduction to Wavelets through Linear Algebra", Springer- Verlag, New York, Inc., 1999.

69. Martin Vetterli, Jelena Kovacevic, "Wavelets aand Subband Coding", Prentice Hall PTR, 1995.

70. Y.T.Chan, "Wavelet Basics", Kluwer Academic Publications, 1995.

71. Stephane Mallat, "A Wavelet tour of signal processing", Academic Press, 1998.

72. Anthony Teolis, "Computational Signal Processing with Wavelets", Birkhauser, 1998.

73. Howard L.Resnikoff, Raymond O.Wells, "Wavelet Analysis", Springer – Verlag New York, Inc.,1998.

74. Raghuvir M.Rao and Ajit S.Bopardikar, "Wavelet Transforms", Addison – Wesley, Longman, Inc.,1998.

75. Robi Polikar, "Wavelet Tutorial Part I – IV", *http://www.public.iastate.edu/~rpolikar*, 1997.

76. Amara Graps, "An Introduction to Wavelets", *www.best.com/~agraps/*

77. Olivier Rioul and Martin Vetterli, " Wavelets and Signal Processing", IEEE SP magazine, pp.14 – 37, October 1991.

78. Lora G.Weiss, "Wavelets and wideband correlation processing", IEEE SP magazine, pp.13 – 32, January 1994.

79. Ingrid Daubechies, " Where do wavelets come from?- A personal point of view", Proceedings of the IEEE, vol.84, No.4, pp.510 – 513, April 1996.

80. Albert Cohen and Jelena Kovacevic, "Wavelets: The Mathematical Background", Proceedings of IEEE, vol.84, No.4, pp.514 – 522, April 1996.

81. Nikolaj Hess – Nielson and Mladen Victor Wickerhauser, " Wavelets and Time frequency analysis", Proc.IEEE, vol.84, No.4, pp.523-540, April 1996.

82. Kannan Ramachandran, Martin Vetterli and Cormac Herley, " Wavelets, Subbands Coding and Best Bases", Proc. IEEE, vol.84, no.4, pp.541 – 560, April 1996.

83. Ole Moller Nielsen, "Wavelets in Scientific Computing", Ph.D.Dissertation, *www.imm.dtu.dk/~omni*

84. C.Valens, "A really Friendly Guide to Wavelets", *www.mindless.com*

85. Norman Morrison, "Introduction to Fourier Analysis", John Wiley & Sons, Inc., 1994.

86. Gilbert Strang, "Linear Algebra and its applications", Academic Press, 1976.

87. D.Esteban and C.Galand, "Applications of QMF to split-band voice coding schemes", Proc. IEEE ICASSP, Hartford, pp.191-195,1977.

88. Olivier Rioul, "Regular Wavelets: A Discrete – Time Approach", IEEE Transactions on Signal Processing , Vol.41, No.12, pp.3572-3579, Dec.1993.

89. M.J.T.Smith, and T.P.Barnwell, "Exact reconstruction techniques for tree structured subband coders", IEEE Transactions on ASSP, 34,pp.434 –441, 1986.

90. M.Antonini, M.Barlandd, P.Mathieu, "Image coding using lattice vector quantization of wavelet coefficients", Proc. IEEE Int. conf. ASSP., pp.2273-2276,1991.

91. P.J.Burt and E.H.Andelson, "The Laplacian pyramid as a compact image code", IEEE Trans. Coomunications, Vol.31, pp.532 –540,1983.

92. M.D.Srinath, P.K.Rajasekharan, R.Viswanathan, "Introduction to statistical signal processing with applications", Eastern Economy Edition, Prentice Hall India, 1999.

93. Sophocles J.Orfanidis, "Optimum Signal Processing: An Introduction", McGraw-Hill International Editions, 1988.

94. N.K.Bose, P.Liang, " Neural Network Fundamentals with graphs, algorithms and applications", Tata Mcgraw-Hill,1996.

95. James A.Freeman, David M.Skapura, " Neural Networks: Algorithms, Applications and Programming Techniques", Addison – Wesley Publishing Company, July 1992.

96. Vikram Pudi, "Neural Networks Tutorial", *www.dsl.serc.iisc.ernet.in*

97. Sijo.N.Lukose, "Perceptual Coding of Digital Audio for multimedia applications using wavelet analysis", M.Tech Thesis, R.E.C, Calicut, 2001.

98. J.F.Birrell, "Diseases of the Nose, Throat and Ear", K.M.Varghese Company, Bombay, India, 1982.

99. David V. Anderson, " A Model Based Development of a Hearing Aid", M.S. Thesis, Department of Electrical and Computer Engineering, Brigham Young University, April 1994.

100. Janet C.Rutledge, Mark A.Clements , "Compensation for recruitment of loudness in sensorineural hearing impairments using a sinusoidal model of speech", Proc.ICASSP, pp.3641-3644,1991.

101. Laura A. Drake, Janet C. Rutledge, Jonathan Cohen, "Wavelet Analysis in Recruitment of loudness compensation", IEEE Trans. on signal processing, vol.41,No.12, pp.3306 –3312, December 1993.

102. Wayne J.Staab, "Digital Hearing Aids", Hearing Instruments journal, vol.36, No.11, pp.7 – 12,1985.

103. Larry E. Humes, Blas Espinozavaras and Charles S. Watson, "Modeling sensorineural hearing loss, I.Model and retrospective evaluation", Journal of Acoustics Society of America, 83(1), January 1988.

104. R.P.Lippman, 1.D. Braida and N. Durlach, "Study of multichannel ampltiude compression and linear amplification for persons with sensorineural hearing loss", J.Acoust.Soc.America, 69(2), Feb.1981.

105.Edger Villchur, "Simulation of the effect of recruitment of loudness relationships in speech", J.Acous.soc.Am.vol.56,No.5,November 1974.

106.Bryon Nielson, "Digital Hearing Aids: where are they?", Hearing Instruments journal, vol.37, No.2, pp.6 –10, 1986.

107.Harry Lewitt, Jean Sullivan, Jain – Yih Hwang, "A computerized hearing aid measurement/simulation system", Hearing Instruments journal, Vol.37, No.2, pp.16-18,1986.

108.James M.Kates, "Signal Processing for Hearing aids", Hearing Instruments journal, vol.37, No.2, 1986, pp.19-21.

109.Cristopher Schweitzer, " Complex Signals and Hearing Aids", Hearing Instruments, vol.37, No.2, pp.26 – 27,1986.

110 Michael Valente, "Hearing Aids: Standards, options and limitations", Thieme Medical Publishers, Inc,1996.

111.Edgar Villchur, "Signal Processing to improve speech intelligibility in perceptive deafness", Journal of Acoust.Soc.of Am., volume 53, No.6, pp.1646 – 1657,1973.

112.Sigibert Wyrsh, August Kaelin, "Subband signal processing for hearing aids", *www.ee.ethz.ch*

113.Osman Erogul, Arfan Karagoz, "Multiresolutional modification of speech signals for listeners with hearing impairment", Journal of Rehabilitation research and development, vol.36, No.3, July 1999.

114.D .K. Bustamante and L. D. Braida, "Principle - component amplitude compression for the hearing impaired", J.Acoust. Soc.Am., vol.82, no.4, pp.1227-1242, 1987.

115.T.F. Quatieri and R J. McAulay, " Sinewave- based phase dispersion for audio preprocessing", in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, vol.ASSP-34, pp.2558-2561, IEEE, 1988.

116.David L. Donoho, "Denoising by Soft Thresholding", IEEE Trans. on Information theory, vol.41,No.3, May 1995.

RESEARCH PUBLICATIONS

1. P.S.Sathidevi , Y.Venkataramani, "Secure Speech Codec", Proceedings of the Fifth National Conference on Communications (NCC-99), Indian Institute of Technology, Kharaghpur, 1999, pp.801-807.

2. P.S.Sathidevi & Y.Venkataramani, "Speech Compression and Encryption" , Proceedings of International conference on Robotics, Vision and Parallel Processing for Automation (ROVPIA 99), University Sains, Malaysia, 1999, pp .106 - 111.

3. P.S.Sathidevi , Y.Venkataramani, "Speech and Audio Coding using Wavelet Transforms", Proceedings of the National Seminar on Applied Systems Engineering and Soft Computing", (SASESC 2000), Dayalbagh Educational Institute, Agra, 2000, pp.242-246.

4. P.S.Sathidevi, Y.Venkataramani, "An Analysis - Synthesis method for transparent Speech and Audio compression using Wavelet transforms and Wavelet Packets", Proceedings of International Conference on Knowledge Based Computer systems -KBCS 2000", Mumbai, pp. 576-587, December 2000.

5. P.S.Sathidevi and Y.Venkataramani," Applying Wavelet Analysis in Coding of Speech and Audio Signals for Multimedia Applications", Published in the web site **www.black.csl.uiuc.edu/~meyn,** International Conference on Stochastic Optimization and Adaptation, ICSOA 2000, CUSAT, Cochin, December 2000.

6. Y.Venkataramani and P.S.Sathidevi," A Two Stage Optimal Wavelet Based Scalable Audio Coding Scheme" , Proceedings of 10^{th} year symposium of Faculty of Engg., European University of Lefke, Turkey, pp.147-153, 2000.

7. P.S.Sathidevi, Y.Venkataramani, " Perceptual coding of Speech and Audio Signals for Multimedia Applications using Optimal Wavelets " Proceedings of National Conference on Communications, NCC 2001, Indian Institute of Technology, Kanpur, pp.224-228, January 2001.

8. P.S.Sathidevi and Y.Venkataramani, "High Quality Low Complexity Audio Coding using Optimal Wavelet Packet Basis", Communicated to National System Conference (NSC 2001).

9. P.S.Sathidevi and Y.Venkataramani, "Perceptual Audio Coding using sinusoidal/optimum wavelet representation", Communicated to Circuits, Systems and Signal Processing (CSSP) Journal, Canada.

10. P.S.Sathidevi and Y.Venkataramani, "Low Complexity Scalable Perceptual Audio Coder using an Optimum Wavelet Packet representation and Vector Quantisation", communicated to IETE Journal of Research, New Delhi, India.

11. P.S.Sathidevi and Y.Venkataramani, "Wavelet Based Hearing Compensation algorithm", communicated to Hearing Instruments Journal.

12. P.S.Sathidevi and Y.Venkataramani, "High Quality Audio Coding using Signal Adaptive Wavelet Filters", Int.Conf. on Optimization Techniques and Its applications in Engineering and Technology, Raja Balwant Singh College, Agra (INDIA), September 2001, *accepted for publication.*

Appendix A

WAVELET FILTER COEFFICIENTS

Wavelet Name	LPF Decomposition	HPF Decomposition	LPF Reconstruction	HPF Reconstruction
db4	-0.0106	-0.2304	0.2304	-0.0106
	0.0329	0.7148	0.7148	-0.0329
	0.0308	-0.6309	0.6309	0.0308
	-0.1870	-0.0280	-0.0280	0.1870
	-0.0280	0.1870	-0.1870	-0.0280
	0.6309	0.0308	0.0308	-0.6309
	0.7148	-0.0329	0.0329	0.7148
	0.2304	-0.0106	-0.0106	-0.2304
coif4	0.0251	0.0561	-0.0561	0.0251
	0.0393	0.4153	0.4153	-0.0393
	-0.0962	-0.7822	0.7822	-0.0962
	-0.0666	0.4344	0.4344	0.0666
	0.4344	0.0666	-0.0666	0.4344
	0.7822	-0.0962	-0.0962	-0.7822
	0.4153	-0.0393	0.0393	0.4153
	-0.0561	0.0251	0.0251	0.0561

Wavelet Name	LPF Decomposition	HPF Decomposition	LPF Reconstruction	HPF Reconstruction
sym4	-0.0758	-0.0322	0.0322	-0.0758
	-0.0296	-0.0126	-0.0126	0.0296
	0.4976	0.0992	-0.0992	0.4976
	0.8037	0.2979	0.2979	-0.8037
	0.2979	-0.8037	0.8037	0.2979
	-0.0992	0.4976	0.4976	0.0992
	-0.0126	0.0296	-0.0296	-0.0126
	0.0322	-0.0758	-0.0758	-0.0322
bior4.4	0	0	0	0
	0.0378	-0.0645	-0.0645	-0.0378
	-0.0238	0.0407	-0.0407	-0.0238
	-0.1106	0.4181	0.4181	0.1106
	0.3774	-0.7885	0.7885	0.3774
	0.8527	0.4181	0.4181	-0.8527
	0.3774	0.0407	-0.0407	0.3774
	-0.1106	-0.0645	-0.0645	0.1106
	-0.0238	0	0	-0.0238
	0.0378	0	0	-0.0378
haar	0.7071	-0.7071	0.7071	0.7071
	0.7071	0.7071	0.7071	-0.7071

Appendix B

DETAILS OF THE ATTACHED COMPACT DISC

Audio signal (.wav)	Scheme	Optimisation method	Quantization
cast111	1 (Ch.4)	1	Scalar Quantization
cast112	1 (Ch.4)	1	Hit Book Method (VQ)
cast113	1 (Ch.4)	1	Scalar +VQ
kadal111	1 (Ch.4)	1	Scalar Quantization
kadal112	1 (Ch.4)	1	Hit Book Method (VQ)
kadal113	1 (Ch.4)	1	Scalar +VQ
mpegtest111	1 (Ch.4)	1	Scalar Quantization
mpegtest112	1 (Ch.4)	1	Hit Book Method (VQ)
mpegtest113	1 (Ch.4)	1	Scalar +VQ
else3111	1 (Ch.4)	1	Scalar Quantization
else3112	1 (Ch.4)	1	Hit Book Method (VQ)
else3113	1 (Ch.4)	1	Scalar +VQ
sitar111	1 (Ch.4)	1	Scalar Quantization
sitar112	1 (Ch.4)	1	Hit Book Method (VQ)
sitar113	1 (Ch.4)	1	Scalar +VQ
cast121	1 (Ch.4)	2	Scalar Quantization
cast122	1 (Ch.4)	2	Hit Book Method (VQ)
cast123	1 (Ch.4)	2	Scalar +VQ
kadal121	1 (Ch.4)	2	Scalar Quantization
kadal122	1 (Ch.4)	2	Hit Book Method (VQ)
kadal123	1 (Ch.4)	2	Scalar +VQ
mpegtest121	1 (Ch.4)	2	Scalar Quantization
mpegtest122	1 (Ch.4)	2	Hit Book Method (VQ)
mpegtest123	1 (Ch.4)	2	Scalar +VQ
else3121	1 (Ch.4)	2	Scalar Quantization
else3122	1 (Ch.4)	2	Hit Book Method (VQ)
else3123	1 (Ch.4)	2	Scalar +VQ
sitar121	1 (Ch.4)	2	Scalar Quantization
sitar122	1 (Ch.4)	2	Hit Book Method (VQ)
sitar123	1 (Ch.4)	2	Scalar +VQ
cast131	1 (Ch.4)	3	Scalar Quantization
cast132	1 (Ch.4)	3	Hit Book Method (VQ)
cast133	1 (Ch.4)	3	Scalar +VQ

Audio signal (.wav)	Scheme	Optimisation method	Quantization
kadal131	1 (Ch.4)	3	Scalar Quantization
kadal132	1 (Ch.4)	3	Hit Book Method (VQ)
kadal133	1 (Ch.4)	3	Scalar +VQ
mpegtest131	1 (Ch.4)	3	Scalar Quantization
mpegtest132	1 (Ch.4)	3	Hit Book Method (VQ)
mpegtest133	1 (Ch.4)	3	Scalar +VQ
else3131	1 (Ch.4)	3	Scalar Quantization
else3132	1 (Ch.4)	3	Hit Book Method (VQ)
else3133	1 (Ch.4)	3	Scalar +VQ
sitar131	1 (Ch.4)	3	Scalar Quantization
sitar132	1 (Ch.4)	3	Hit Book Method (VQ)
sitar133	1 (Ch.4)	3	Scalar +VQ
castdct11	DCT	--	Scalar Quantization
kadaldct11	DCT	--	Scalar Quantization
mpegtestdct11	DCT	--	Scalar Quantization
else3dct11	DCT	--	Scalar Quantization
sitardct11	DCT	--	Scalar Quantization
cast211	1(Ch.5)	1	Scalar Quantization
cast212	1(Ch.5)	1	Hit Book Method (VQ)
cast213	1(Ch.5)	1	Scalar +VQ
kadal211	1(Ch.5)	1	Scalar Quantization
kadal212	1(Ch.5)	1	Hit Book Method (VQ)
kadal213	1(Ch.5)	1	Scalar +VQ
mpegtest211	1(Ch.5)	1	Scalar Quantization
mpegtest212	1(Ch.5)	1	Hit Book Method (VQ)
mpegtest213	1(Ch.5)	1	Scalar +VQ
else3211	1(Ch.5)	1	Scalar Quantization
else3212	1(Ch.5)	1	Hit Book Method (VQ)
else3213	1(Ch.5)	1	Scalar +VQ
sitar211	1(Ch.5)	1	Scalar Quantization
sitar212	1(Ch.5)	1	Hit Book Method (VQ)
sitar213	1(Ch.5)	1	Scalar +VQ

Audio signal (.wav)	Scheme	Optimisation method	Quantization
cast311	2 (Ch.5)	1	Scalar Quantization
cast312	2 (Ch.5)	1	Hit Book Method (VQ)
cast313	2 (Ch.5)	1	Scalar +VQ
kadal311	2 (Ch.5)	1	Scalar Quantization
kadal312	2 (Ch.5)	1	Hit Book Method (VQ)
kadal313	2 (Ch.5)	1	Scalar +VQ
mpegtest311	2 (Ch.5)	1	Scalar Quantization
mpegtest312	2 (Ch.5)	1	Hit Book Method (VQ)
mpegtest313	2 (Ch.5)	1	Scalar +VQ
else3311	2 (Ch.5)	1	Scalar Quantization
else3312	2 (Ch.5)	1	Hit Book Method (VQ)
else3313	2 (Ch.5)	1	Scalar +VQ
sitar311	2 (Ch.5)	1	Scalar Quantization
sitar312	2 (Ch.5)	1	Hit Book Method (VQ)
sitar313	2 (Ch.5)	1	Scalar +VQ
clap421	3(Ch.5)	2	Scalar Quantization
clap422	3(Ch.5)	2	Hit Book Method (VQ)
clarinet421	3 (Ch.5)	2	Scalar Quantization
clarinet422	3 (Ch.5)	2	Hit Book Method (VQ)
crow421	3 (Ch.5)	2	Scalar Quantization
crow422	3 (Ch.5)	2	Hit Book Method (VQ)
crow423	3 (Ch.5)	2	Scalar +VQ
drums421	3(Ch.5)	2	Scalar Quantization
drums422	3 (Ch.5)	2	Hit Book Method (VQ)
drums423	3 (Ch.5)	2	Scalar +VQ
female421	3(Ch.5)	2	Scalar Quantization
female422	3(Ch.5)	2	Hit Book Method (VQ)
female2421	3 (Ch.5)	2	Scalar Quantization
female2422	3 (Ch.5)	2	Hit Book Method (VQ)
female2423	3 (Ch.5)	2	Scalar +VQ
male421	3(Ch.5)	2	Scalar Quantization
male422	3(Ch.5)	2	Hit Book Method (VQ)
ring421	3 (Ch.5)	2	Scalar Quantization
ring422	3 (Ch.5)	2	Hit Book Method (VQ)
pup421	3(Ch.5)	2	Scalar Quantization
pup422	3 (Ch.5)	2	Hit Book Method (VQ)
pup423	3 (Ch.5)	2	Scalar +VQ
whistle421	3(Ch.5)	2	Scalar Quantization
whistle422	3 (Ch.5)	2	Hit Book Method (VQ)
whistle423	3 (Ch.5)	2	Scalar +VQ

Audio signal (.wav)	Scheme	Optimisation method	CR	Quantization
male5316	5 (Ch.6)	3	6	Scalar Quantization
clap5316	5 (Ch.6)	3	6	Scalar Quantization
clarinet5316	5 (Ch.6)	3	6	Scalar Quantization
pup5316	5 (Ch.6)	3	6	Scalar Quantization
whistle5316	5 (Ch.6)	3	6	Scalar Quantization
drums5316	5 (Ch.6)	3	6	Scalar Quantization
female25316	5 (Ch.6)	3	6	Scalar Quantization
crow5316	5 (Ch.6)	3	6	Scalar Quantization
mpegtest5316	5 (Ch.6)	3	6	Scalar Quantization
kadal5316	5 (Ch.6)	3	6	Scalar Quantization
else35316	5 (Ch.6)	3	6	Scalar Quantization
sitar5316	5 (Ch.6)	3	6	Scalar Quantization
male53110	5 (Ch.6)	3	10	Scalar Quantization
clap53110	5 (Ch.6)	3	10	Scalar Quantization
clarinet53110	5 (Ch.6)	3	10	Scalar Quantization
pup53110	5 (Ch.6)	3	10	Scalar Quantization
whistle53110	5 (Ch.6)	3	10	Scalar Quantization
drums53110	5 (Ch.6)	3	10	Scalar Quantization
female253110	5 (Ch.6)	3	10	Scalar Quantization
crow53110	5 (Ch.6)	3	10	Scalar Quantization
mpegtest53110	5 (Ch.6)	3	10	Scalar Quantization
kadal53110	5 (Ch.6)	3	10	Scalar Quantization
else353110	5 (Ch.6)	3	10	Scalar Quantization
sitar53110	5 (Ch.6)	3	10	Scalar Quantization
male53115	5 (Ch.6)	3	15	Scalar Quantization
clap53115	5 (Ch.6)	3	15	Scalar Quantization
clarinet53115	5 (Ch.6)	3	15	Scalar Quantization
pup53115	5 (Ch.6)	3	15	Scalar Quantization
whistle53115	5 (Ch.6)	3	15	Scalar Quantization
drums53115	5 (Ch.6)	3	15	Scalar Quantization
female253115	5 (Ch.6)	3	15	Scalar Quantization
crow53115	5 (Ch.6)	3	15	Scalar Quantization
mpegtest53115	5 (Ch.6)	3	15	Scalar Quantization
kadal53115	5 (Ch.6)	3	15	Scalar Quantization
else353115	5 (Ch.6)	3	15	Scalar Quantization
sitar53115	5 (Ch.6)	3	15	Scalar Quantization

Audio signal (.wav)	Scheme	Optimisation method	CR	Quantization
female5a3125-1	5a (Ch.6)	3	25	Scalar Quantization
female5a3125-2	5a (Ch.6)	3	25	Scalar Quantization
female25a3125-1	5a (Ch.6)	3	25	Scalar Quantization
female25a3125-2	5a (Ch.6)	3	25	Scalar Quantization
clap5a3115-1	5a (Ch.6)	3	15	Scalar Quantization
clap5a3115-2	5a (Ch.6)	3	15	Scalar Quantization
clap5a3125-1	5a (Ch.6)	3	25	Scalar Quantization
clap5a3125-2	5a (Ch.6)	3	25	Scalar Quantization
kadal5a3110-1	5a (Ch.6)	3	10	Scalar Quantization
kadal5a3110-2	5a (Ch.6)	3	10	Scalar Quantization
kadal5a3125-1	5a (Ch.6)	3	25	Scalar Quantization
kadal5a3125-2	5a (Ch.6)	3	25	Scalar Quantization
Results of Hearing Compensation Method using DWT				
clap	Ch.7	-----	----	----------
clapanalog	Ch.7	-----	----	----------
clapdigi	Ch.7	-----	----	----------
clapn	Ch.7	-----	----	----------
clapnanalog	Ch.7	-----	----	----------
clapndigi	Ch.7	-----	----	----------
female	Ch.7	-----	----	----------
femaleanalog	Ch.7	-----	----	----------
femaledigi	Ch.7	-----	----	----------
femalen	Ch.7	-----	----	----------
femalenanalog	Ch.7	-----	----	----------
femalendigi	Ch.7	-----	----	----------
pup	Ch.7	-----	----	----------
pupanalog	Ch.7	-----	----	----------
pupdigi	Ch.7	-----	----	----------
pupn	Ch.7	-----	----	----------
pupnanalog	Ch.7	-----	----	----------
pupndigi	Ch.7	-----	----	----------

Note: ***clap*** – original signal, ***clapn*** – clap+noise, ***clapanalog*** - analog hearing aid output without noise, ***clapdigi*** – digital hearing aid output without noise, ***clapnanalog*** - analog hearing aid output with noise, ***clapndigi*** – digital hearing aid output with noise.