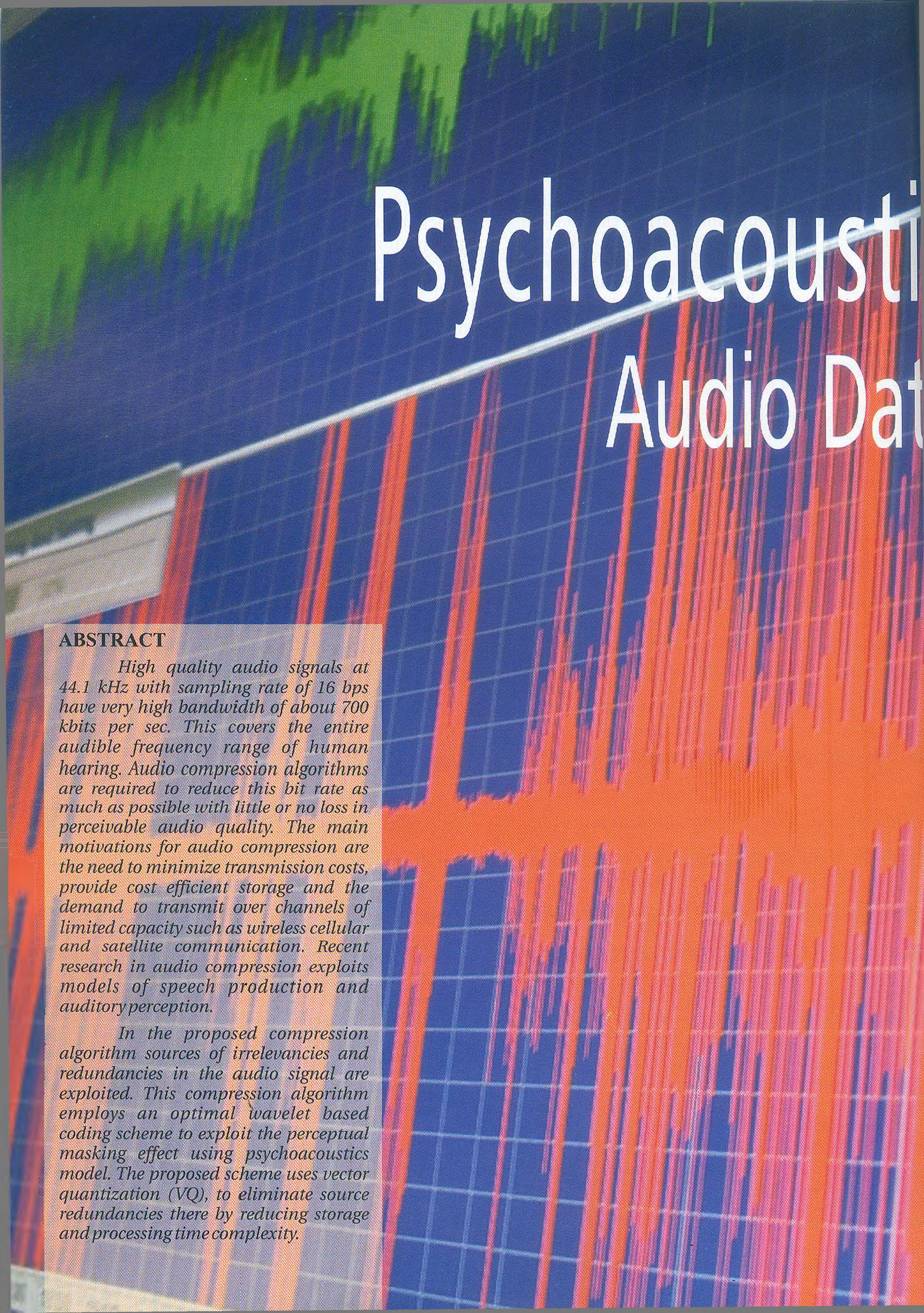

Psychoacousti
Audio Dat

ABSTRACT

High quality audio signals at 44.1 kHz with sampling rate of 16 bps have very high bandwidth of about 700 kbits per sec. This covers the entire audible frequency range of human hearing. Audio compression algorithms are required to reduce this bit rate as much as possible with little or no loss in perceivable audio quality. The main motivations for audio compression are the need to minimize transmission costs, provide cost efficient storage and the demand to transmit over channels of limited capacity such as wireless cellular and satellite communication. Recent research in audio compression exploits models of speech production and auditory perception.

In the proposed compression algorithm sources of irrelevancies and redundancies in the audio signal are exploited. This compression algorithm employs an optimal wavelet based coding scheme to exploit the perceptual masking effect using psychoacoustics model. The proposed scheme uses vector quantization (VQ), to eliminate source redundancies there by reducing storage and processing time complexity.

Model Based Compression Technique:

An Approach using DWT and VQ

Dr. Jyoti Singhai, Rakesh Singhai, Dr. R.P. Singh

INTRODUCTION

The well-known audio compression algorithms used in MPEG I and MPEG II standards are MUSICAM (Masking-pattern Universal Subband Integrated Coding and Multiplexing) and PAC (Perceptual Audio Coder) respectively.

MUSICAM uses Fast Fourier transform for power spectral estimation. Based on the estimated power spectrum, a masking curve is calculated. This masking curve is used for deciding quantization step size for each block. The information of the step size, number of steps etc. is also transmitted along with the quantized samples as overhead. This increases the number of bits transmitted [1]. In PAC system, the signal is synthesized through a filter bank based on perceptual model. Then the output is quantized using PCM quantization. The PAC system does not use entropy coding, the quantized values are sent directly. The maximum compression achieved in MUSICAM and PAC is 2.74 and 1.823 respectively. Further increase in the compression ratio, deteriorates the audio signal quality because of the constant bandwidth of the subbands used for computing masking threshold, which are not coincident with the critical bands of hearing. Also Fast Fourier transforms used for spectral estimation are unsuitable for non-stationary audio signals because they have fixed resolution and do not give the timing information. This timing information is essential for pitch of good quality of audio signal.

The proposed scheme explained in this paper employs an optimal Wavelet [2][3] based coding scheme to exploit the perceptual masking effect using psychoacoustics model and Vector Quantization to eliminate source redundancies. Psychoacoustics model is used to determine the portion of the audio signal that can be removed without loss of quality of sound to the human ear. The Discrete Wavelet Transform (DWT) used can decompose an input audio signal into non-overlapping frequency bands corresponding to critical bands of constant hearing [4]. This helps in employing the psycho-acoustic model easily for calculating the masking threshold in each frequency band [5]. The proposed technique achieves better compression ratio because with Vector quantization it is not required to store or to transmit extra information. This results in reduction in storage and processing complexity with better quality of sound signal. The proposed scheme is implemented using MATLAB 7. Its performance in terms of Signal to Noise ratio(SNR), Peak Signal to Noise Ratio (PSNR), Normalized Root Mean Square Error (NRMSE) and Compression Score is evaluated for different types of test signals.

AUDIO CODING USING DWT AND VQ

In order to compress data by a large factor, an algorithm must be lossy and it must throw out some of the less important information. In the case of an audio signal, it is assumed that throwing out portions would result in a noticeable degradation in sound quality. However, use of psychoacoustics model helps in minimizing the audible effects of lossy compression.

The proposed scheme is shown in Figure 1(a) and Figure 1 (b). In the scheme, an input audio signal is windowed and divided into shorter segments. The audio data is segmented using the square root of the Hann window $\omega(n) = 0.5 - 0.5\cos(2n / N)$. In the following experiment the segment length used was 1024 samples (23.2ms). The DWT of each segment is carried out thereby yielding the transform coefficients at different resolutions.

DISCRETE WAVELET TRANSFORM

The DWT coefficients are computed using the Wave packet representation [4][5], as shown in Figure 2. Since the audio signals have a higher bandwidth of 20 kHz as compared to speech, the task of data compression becomes easier if the coarser i.e. low frequency and finer i.e. high frequency details of the signal are separated out and then these parts are coded separately. Hence in such a case it is preferable to decompose both the high pass and the low pass bands as opposed to decomposing only the low pass band in the conventional wavelet transform representation. In this algorithm, each audio segment is passed through a bank of perfectly reconstructed low pass and high pass filters followed by decimators. The output of the high pass filter is sub sampled by two and collected as the DWT coefficients at one particular resolution. The output of the low pass filter is also sub sampled and again passed through the same filter bank to compute the DWT coefficients at a lower resolution. This recursive procedure is carried out till the lowest resolution coefficients are obtained and the desired decomposition is achieved. For the last stage of decomposition, the output of both the low pass and high pass filters are sub sampled and collected as the DWT coefficients at the lowest resolution. This can expressed as

$$y_{low}[k] = \sum_{n=-\infty}^{\infty} x[n].\,h[2k-n] \qquad (2.1)$$

$$y_{high}[k] = \sum_{n=-\infty} x[n].\,g[2k-n] \qquad (2.2)$$

where $y_{low}[k]$ and $y_{high}[k]$ are output of the low pass and high pass filters after sub sampling by 2. h[n] and g[n] are impulse response of low pass and high pass filters. The two filters are Quadrature Mirror Filters and odd index alternated reversed version of each other. They are related as

$$g[L-1-n] = (1)^n.h[n] \qquad (2.3)$$

L is the length of the filter in number of points [8].

** For all figures and tables please refer to the Appendix*

This Wavepacket representation is applied to the input audio signal of 22 kHz bandwidth and decomposed it into two 11 kHz bands. These two bands are further decomposed in four 5.5 kHz bands and this decomposition continues till critical bands of frequency are obtained, this is shown in Figure 3. In Figure 3 the figures in the brackets shown on the right side are critical bands in kHz. In the MATLAB simulation, the DWT coefficients for each separated critical band are extracted with the help of wavelet corresponds to Harr and Daubechies' 8 tap filter [7]. The shape of the wavelet chosen is decided by the coefficients of quadrature mirror decomposition filter and the wavelets Harr and Daubechies used provides orthogonal wavelet decomposition. The threshold levels are calculated using approximation results from psychoacoustics model.

PSYCHOACOUSTIC MODEL

The psychoacoustic model is based on many studies of human perception. Effects due to different sounds in the environment and limitations of the human sensory system lead to facts that can be used to remove portions of a signal that are inaudible and indiscernible to the average human being. The two main properties of the human auditory system that make up the psychoacoustic model are absolute threshold of hearing and auditory masking [10] [11].

Absolute Threshold of Hearing (ATH) in dB Sound Pressure Level is determined as :

$$ATH = 3.64 \times (f/1000)^{-0.8} - 6.5e^{(-0.6*((f/1000)-3.3)\wedge 2)} + 10^3 \times (f/1000)^4 \text{ (dB SPL)} \qquad (2.4)$$

Where f is the frequency in hertz and ATH is Absolute Threshold of Hearing in dB Sound Pressure Level. Thus, if a signal has any frequency components with power levels that fall below the ATH, then these components can be discarded, as the average listener will be unable to hear those frequencies of the signal anyway.

The concept of auditory masking [12] is that the stronger signal will mask the weaker signals at nearby frequencies and making them inaudible to the listener. For a masked signal to be heard, its power will need to be increased to a level greater than that of a threshold that is determined by the frequency of the masker tone and its strength. In the compression algorithm, therefore, tone masker, noise masker and masking effect due to these maskers is determined. If any frequency components around these maskers fall below the masking threshold, they can be discarded [13] [14].

For the purposes of this project, a tone is determined as a frequency, which has a sharp peak in the frequency spectrum. To determine whether a certain frequency is a tone or masker is done with the help of following definition:

A frequency f is a tone if its power P[k] is:

1) Greater than P[k-1] and P[k+1], i.e., it is a local maxima.
2) 7dB greater than the other frequencies in its neighborhood, where the neighborhood is dependent on f:

(i) If 0.17 Hz < f < 5.5 kHz, the neighborhood is [k-2...k+2].
(ii) If 5.5 kHz =< f < 11 kHz, the neighborhood is [k-3...k+3].
(iii) If 11 kHz =< f < 20 kHz, the neighborhood is [k-6...k+6].

VECTOR QUANTIZATION (VQ)

Vector quantization is a pattern matching procedure. In VQ source output is grouped into blocks or vectors. For example, L consecutive samples of speech are treated as the element of an L dimension vector. This vector of source output forms the input to the vector quantizer. At the encoder end and decoder end, the vector quantizer consists of a set of L-dimensional vectors, which is called the codebook of vector quantizer. Best-matched representative code vector of the L dimensional input vector is chosen from the codebook at the encoder and its address, is sent to the receiver where an exact replica of the codebook is stored. The decoder on receiving the address, outputs the reconstructed vector from its codebook [15][16]. Thus VQ saves the number of transmitted bits and also simplifies the decoding process. VQ is hence ideal for single encoder and multiple decoder scenarios, for example, in multimedia applications. In multimedia, considerable computational resources are available for encoding operation. However as the decoding is to be done by software the required amount of computational resources at the decoder is very less [18]. The most prevalent technique for codebook design in vector quantization is LBG algorithm generalized by Linde, Buzo and Gray [17].

The algorithm developed for audio data compression using DWT and VQ based on LBG algorithm is defined as follows:

1. Segment the audio data using the square root of Hann window (n). The Hann window for length N is ω(n) = 0.5 0.5 cos (2n/N).
2. Extract the coefficients of audio signal by Discrete Wavelet Transform using filter bank of low pass filter h[n] and high pass filter g[n], $g[n] = (-1)^n h[L-1-n]$.
3. Determine the threshold level "θ" based on psychoacoustic model and neglect DWT coefficients with values below global threshold level and assign them zero value.
4. Encode consecutive zero valued coefficients with two bytes. One byte is used to specify a starting string of zeros and the second byte keeps track of the number of successive zeros.
5. For quantization, start with an initial set of DWT reconstruction values $\{Y_i^{(0)}\}_{i=0}^{M}$ and a set of DWT training vectors $\{X_n\}_{n=1}^{N}$. Set k=0, $D^{(0)} = 0$. Select Threshold ∈.
6. Find the quantization regions $\{V_i^{(k)}\}_{i=1}^{M}$ are given by,
$$V_i^{(k)} = \{X_n : d(X_n, Y_j) < d(X_n, Y_j) \forall j \neq i\}$$
i = 1, 2, ..., M. We assume that none of the quantization regions are empty.
7. Compute the distortion $D^{(k)}$ between the training vectors and the representative reconstruction values.

i. $$D^{(k)} = \sum_{i=1}^{M} \int_{V_i^{(k)}} \|X - Y_i^{(k)}\|^2 f_x(X)dX$$

8. If $\dfrac{(D^{(k)} - D^{(k-1)})}{D^{(k)}} < \in$, stop; otherwise continue.
9. For k = k+1. Find new reconstruction values $\{Y_i^{(k)}\}_{i=1}^{M}$ that are the average value of the elements of each of the quantization regions. $\{V_i^{(k-1)}\}$ Go to step 6.

PERFORMANCE OF THE PROPOSED ALGORITHM

The proposed audio coding algorithm uses DWT to exploit the perceptual masking effect in human hearing process and uses vector quantization to eliminate source redundancies. The proposed algorithm is implemented using MATLAB [19] and various experiments are conducted to evaluate the performance for three different types of test signal viz, a male voice test signal, a female voice test signal and an instrumental sound test signal. MATLAB Files used in Simulation are divided into different stages as follows:

A. *Windowing and Segmenting*: It converts the audio data in column vector and segment it into frames of given size using matfiles Openfile.m, FileSize.m, FrameSelect.m.
B. *Psychoacoustic Model*: It normalized power spectrum of the new signal and determines the PSD of tones (Ptm), PSD of noise at location and location within critical band (Pnm_at_loc, loc), Masking threshold, Global masking threshold using Matfiles Dbinv.m, hz2bark.m, normalize.m, psd.m, findtones.m, noise_masker.m, mask_threshold.m, global_threshold.m.
C. *Extraction of Coefficients Using DWT*: It extracts the coefficients using DWT and compresses them using threshold level calculated. It encodes the coefficients of DWT to eliminate the consecutive zeroes.
D. *Quantization*: Quantize the encoded coefficients using vector quantization.
E. *Inverse Quantization*: Inverse quantize the quantized coefficients to get the encoded DWT coefficients.
F. *Decompression*: Decode the inverse quantized coefficients and reconstruct the signal.
G. *Performance measurement*: Performance of the proposed system is calculated in terms of Compression Score, Signal to noise ratio (SNR), peak SNR (PSNR), normalized root mean square error (NRMSE) plots are shown in Figure-4, 5 and 6.

In the proposed audio signal compression algorithm, the compression score for different test signals is calculated as

Compression Ratio (C) = Length (x(n)) / Length(CWC) (3.1)

where x(n) is the speech signal and CWC is the Compressed Wavelet transform Code. Compression score for different types of wavelets for three best score male voice signal, female voice signal and instrumental sound signal are obtained from source code developed in MATLAB and are tabulated in Table 1. The accuracy of reconstructed signal is determined by measuring distortion in the decoded signal in terms of SNR, PSNR and NRMSE. These related measures of SNR, PSNR and NRMSE are defined as

Signal to Noise Ratio $(SNR) = 10\log_{10}\left(\dfrac{\sigma_x^2}{\sigma_e^2}\right)$ (3.2)

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

Peak Signal to Noise Ratio $(P \quad) = 10\log_{10}\left(\dfrac{NX^2}{|x-r|^2}\right)$ (3.3) where N is the length of the reconstructed signal, X is the maximum absolute value of the signal x, $|x-r|^2$ is the energy of the difference between original and reconstructed signals,

Normalized Root Mean Square Error

$$(\quad) \quad \sqrt{\frac{(x(n) \quad r(n))^2}{(x(n) \quad (n))^2}} \qquad (3.4)$$

where x(n) is the original speech signal r(n) is reconstructed speech signal and $\mu_x(n)$ is the mean of the speech signal. These measures vary with the level of truncation and thus different applications for different quality of signals for different compression can be achieved. Plot for SNR, PSNR, NRMSE with respect to level of truncation are shown in Figure 4, 5 and 6. These measures SNR, PSNR, NRMSE for different DWT with constant truncation are calculated for three best score male voice test signal, for three best score female voice test signal and for three best score instrumental sound test signal and extra measured & NR, PSNR, NRMSE for different DWT with constant truncation are arranged in Table2 (a), Table2 (b) and Table2(c) respectively.

Plots for original and compressed signal for best score male voice signal, female voice signal and instrumental sound signal are shown in Figure 7, Figure 8 and Figure 9 respectively. In these figures plot in red is for original signal and plot in yellow is for compressed signal. These plots are plotted with reference to sample on x-axis. Figure 7 shows plots of original and compressed signal for male voice for 71680 samples. Figure 8 shows plots of original and compressed signal for female voice for 51200 samples. Figure 9 shows plots of original and compressed signal for instrumental sound for 40960 samples. The proposed audio signal compression algorithm achieves signal to noise ratio of 17.45 db and Peak signal to noise ratio of 98.45 db at a compression ratio of 3.98 using the Daubechies-10 wavelet transform with level 5 decomposition for male voice test signal. For female voice test signal it achieves signal to noise ratio of 14.39 db and Peak signal to noise ratio of 96.24 db at a compression ratio of 3.65 using the Daubechies-10 wavelet transform with decomposition at 5 level, and for instrumental sound test signal it achieves signal to noise ratio of 19.53 db and Peak signal to noise ratio of 98.98 db at a compression ratio of 3.5 using the Daubechies-10 wavelet transform with decomposition at 5 level.

CONCLUSION

The performance of the proposed scheme in terms of compression scores and signal quality is comparable with other good techniques such as MUSICAM and PAC. The MUSICAM compresses audio signals of 700 kbps down to around 256 kbps i.e. compression score of about 2.74 without audible impairments where as PAC coder provides compression ratio of about 1.823 for transparent or near transparent quality coding, where as the proposed scheme has compression score of 3.65, which is higher than the existing schemes [1][2]. In addition, in this scheme the compression ratio is variable, while most other compression techniques have fixed compression ratios. The VQ used reduces storage and searching complexity. Thus the proposed scheme is more suitable for multimedia applications as it provides better compression with better quality of signal and reduced storage and searching complexity. This compression approach can be used in number of applications such as transmission of speech over mobile satellite communication, cellular telephony, voice mail systems and synthetic voice in toys that speak.

REFERENCES

1. A.H. Tewfik, D. Sinha, P. Jorgensen, *"On Optimal Choice of a Wavelet for Signal Representation"*, IEEE Transactions on Information Theory, vol.38, No.2, pp 747-765, (Mar,1992).
2. A.H. Tewfik et al., *"Low Bit Rate Transparent Audio Compression Using Adapted Wavelets"*, IEEE Transactions on Signal Processing, vol 41, No.12, pp 3463-3479, (Dec. 1993).
3. A.S. Spanias, *"Speech Coding: A Tutorial Review"*, Proc, IEEE vol 82, pp.1541-1582, (Oct 1994).
4. Chen A., Shehad N., et al, *"Discrete Wavelet Transform for Audio Compression"*, http://is.rice.edu/~welsh/elec431/wavelets.html (current April 16, 2005)
5. E. Zwicker and H. Fastl, *"Psychoacoustics: Facts and Models"*, Berlin, Germany: Springer (1990).
6. Graps A., *"An Introduction to Wavelets"*, IEEE Computational Science and Engineering, http://www.amara.com/IEEEwave/IEEEwavelet.html (current Mar 15,2005)
7. H. Abut (Editor), *"Vector Quantization"*, New york, NY: IEEE Press, (1990).
8. Hans M., Schafer R.W., *"Lossless Compression of Digital Audio"*, IEEE Signal Processing magazine, Vol.18, July 2001, pp 21-32.
9. Homayoumfar K., *"Rate Adaptive Speech Coding for Universal Multimedia Access"*, IEEE Signal Processing magazine, issue 5, Sept 2003.
10. J.D. Johnston, *"Transform Coding of Audio Signals Using Perceptual Noise Criterion"*, IEEE Journal on Selected Areas of Communication, vol.6, No.2, pp 314-323, (Feb.1988).
11. K. Brandenburg and G. Stoll, *"ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio"*, Journal of the Audio Engineering Society, 42, 780-792, (October, 1994).
12. M. Vetterli and J. Kovacevic, *"Wavelets and Subband Coding"*, Prentice Hall PTR Englewood cliffs, N.J., (1995).
13. Mistri M., Mistri Y., Oppenham G., *"Matlab Wavelet Toolbox"*, The MathWorks Inc., 2003.
14. S.G.Mallat, *"A Theory of Multiresolution Signal Decomposition: The Wavelet Representation"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.11, No.7, pp 674-693, (July1989).
15. Sinha, D., and J.D. Johnston, *"Audio Compression at Low Bit Rates Using a Signal Adaptive Switched Filter Bank"*, IEEE International Conference on Acoustics, Speech and Signal Processing, Munich, (April 1997).
16. T. Parsons, *"Voice and Speech Processing"*, McGraw Hill Publications, (Edition 1987).
17. Wilson B., *"Digital Signal Processing Application for Hearing Accessibility"*, IEEE Signal Processing magazine, issue 05, Sept 2003.
18. X.Wu and L.Guan, *"Acceleration of the LBG Algorithm"*, IEEE Trans. Commun., vol 42, pp 1518-1523, (Apr.1994).
19. Y.Linde, A.Buzo, R.Gray, *"An Algorithm for Vector Quantization Design"*, IEEE Transactions on Communications, vol.Com-28, No.1, pp 84-95, (Jan, 1980).

Appendix

Type of wavelet	Male voice			Female voice			Instrumental sound		
	M1	M2	M2	F1	F2	F3	I1	I2	I3
Harr	3.4583	3.2987	3.0651	3.3658	3.1532	3.0032	3.2458	3.0683	2.9742
Db4	3.7724	3.5021	3.3345	3.4765	3.2670	3.0975	3.3657	3.1582	3.0432
Db6	3.8802	3.5531	3.3980	3.5988	3.3543	3.1562	3.4660	3.2741	3.1476
Db8	3.8950	3.5910	3.4004	3.6352	3.4521	3.2374	3.4958	3.3058	3.1998
Db10	3.9753	3.7034	3.5145	3.6442	3.4983	3.2879	3.5013	3.3996	3.1145

Table1: Compression Score for 09 different types of audio signals for different wavelet

Type of wavelet	Measures for Male Voice			Measures for Male Voice			Measures for Male Voice		
	Test Signal M1			Test Signal M2			Test Signal M3		
	SNR	PSNR	NRMSE	SNR	PSNR	NRMSE	SNR	PSNR	NRMSE
Harr	11.7428	92.7803	0.2587	10.5382	90.4738	0.4372	9.5328	88.1630	0.5281
Db4	16.5870	97.6245	0.1481	15.4092	95.3896	0.3762	14.2062	93.0752	0.4997
Db6	16.8722	97.9.98	0.1433	15.6583	95.6693	0.3502	14.4837	934078	0.4505
Db8	17.1963	98.0138	0.1399	16.0769	95.9968	0.2979	15.6729	93.8085	0.3989
Db10	17.4502	98.4877	0.1341	16.3654	96.4237	0.2078	15.9879	94.2628	0.2958

TABLE 2(a): Measures for Male Voice Test Signal for different DWT with constant truncation

Type of wavelet	Measures for Female Voice			Measures for Female Voice			Measures for Female Voice		
	Test Signal F1			Test Signal F2			Test Signal F3		
	SNR	PSNR	NRMSE	SNR	PSNR	NRMSE	SNR	PSNR	NRMSE
Harr	10.0972	91.9450	0.3127	9.8593	89.6493	0.5318	8.6592	87.4392	0.6884
Db4	13.4386	95.2864	0.2128	12.2749	93.0421	0.4295	10.4923	91.5286	0.6052
Db6	13.9378	95.7947	0.2007	12.7328	93.4859	0.4012	10.8402	91.8740	0.5755
Db8	14.2855	96.1318	0.1931	13.0542	93.9207	0.3917	11.5287	92.0076	0.5003
Db10	14.3866	96.2324	0.1909	13.1657	94.1382	0.3583	11.8904	92.1152	0.4729

TABLE 2(b): Measures for Female Voice Test Signal for different DWT with constant truncation

Type of Wavelet	Measures for Instrumental Sound Test I1			Measures for Instrumental Sound Test Signal I2			Measures for Instrumental Sound Test Signal I3		
	SNR	PSNR	NRMSE	SNR	PSNR	NRMSE	SNR	PSNR	NRMSE
Harr	12.0087	93.9877	0.2133	10.9803	91.6729	0.4287	8.9367	89.3572	0.59983
Db4	14.8465	96.8098	0.1901	12.3825	94.5392	0.3979	10.5393	92.6720	0.5037
Db6	15.0792	97.0424	0.1499	13.7639	95.4297	0.3692	11.6392	93.7820	0.4869
Db8	18.7398	97.7387	0.1413	16.4826	95.9845	0.3003	14.0639	93.7892	0.4148
Db10	19.5240	98.9875	0.1298	17.6793	96.5620	0.29683	15.6294	94.5620	0.3751

TABLE 2(c): Measures for Instrumental Sound Test Signal for different DWT with constant truncation

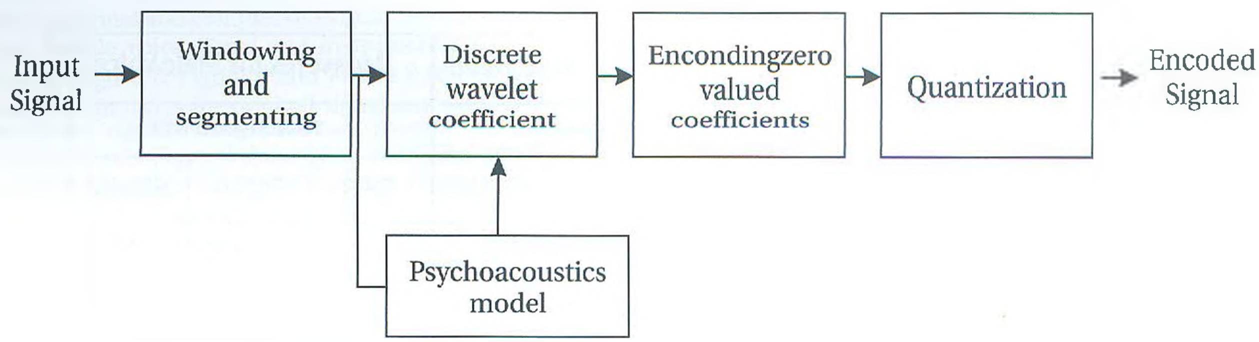

Figure 1(a) Block Diagram of Encoder Circuit

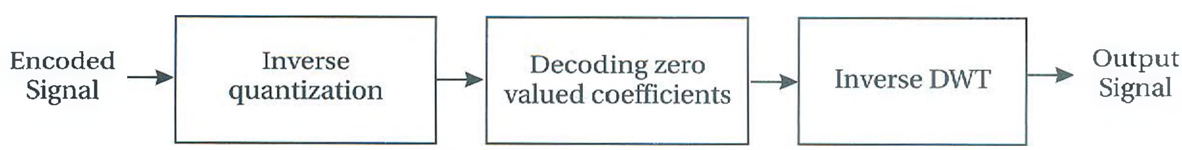

Figure 1(b) Block Diagram of Decoder Circuit

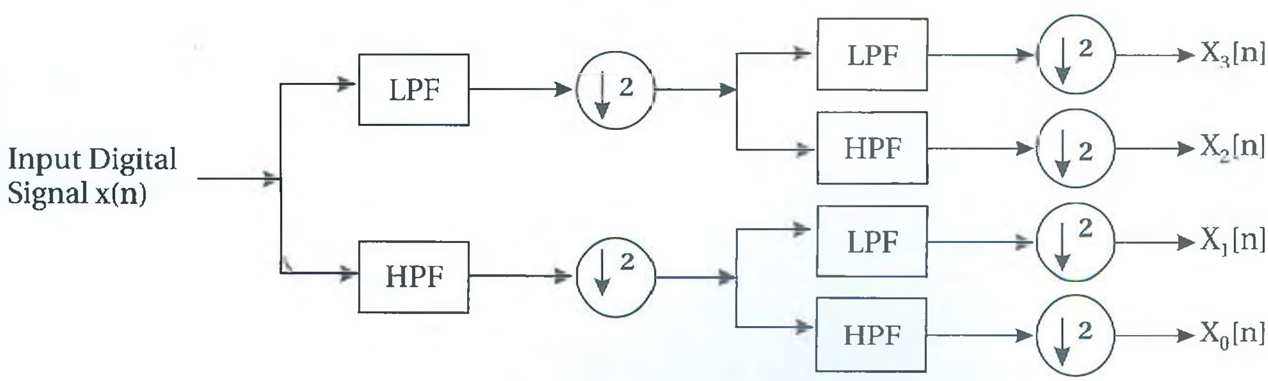

Figure 2: Wavepacket Decomposition

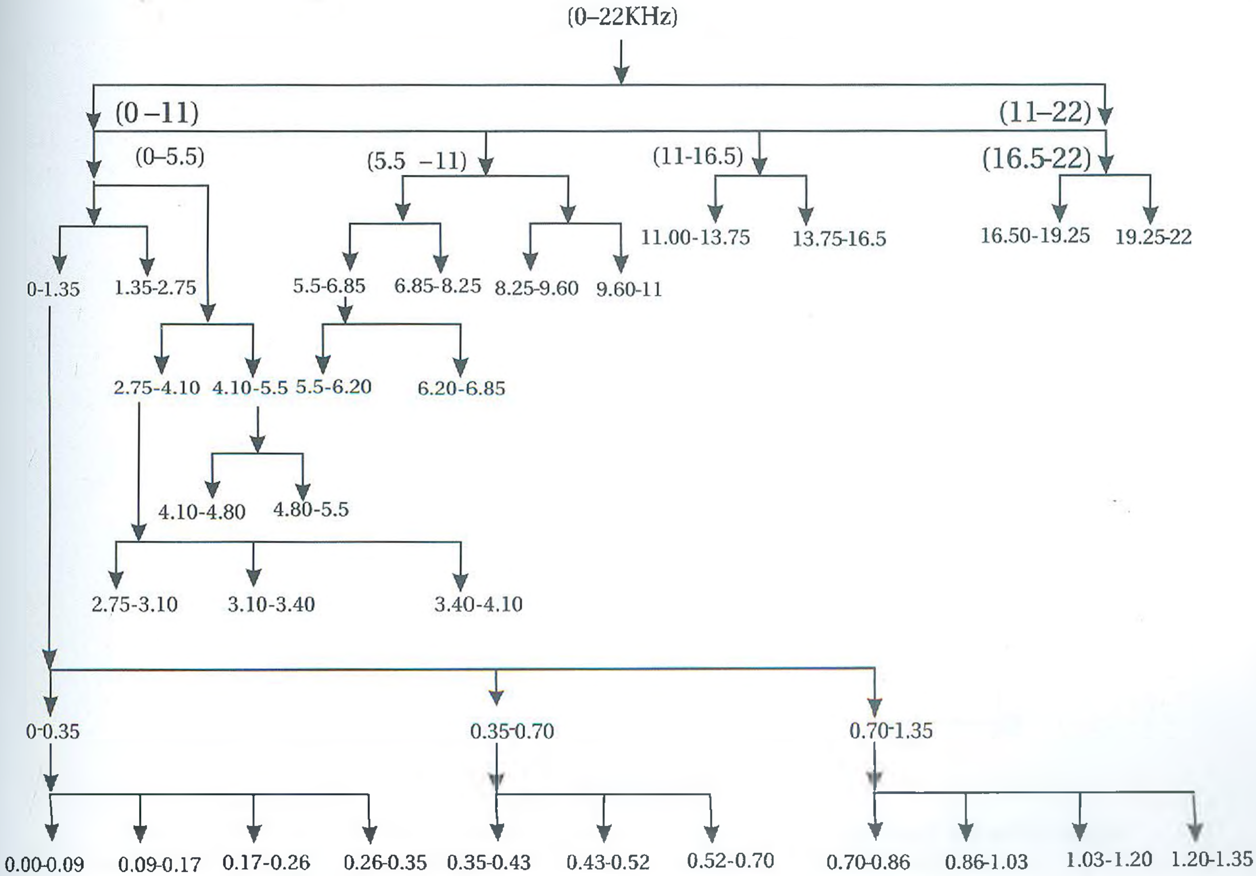

Figure 3: Wavelet decomposition tree used in the audio code. The numbers in the figure refer to the lower and higher cutoff frequencies in kHz of each band.

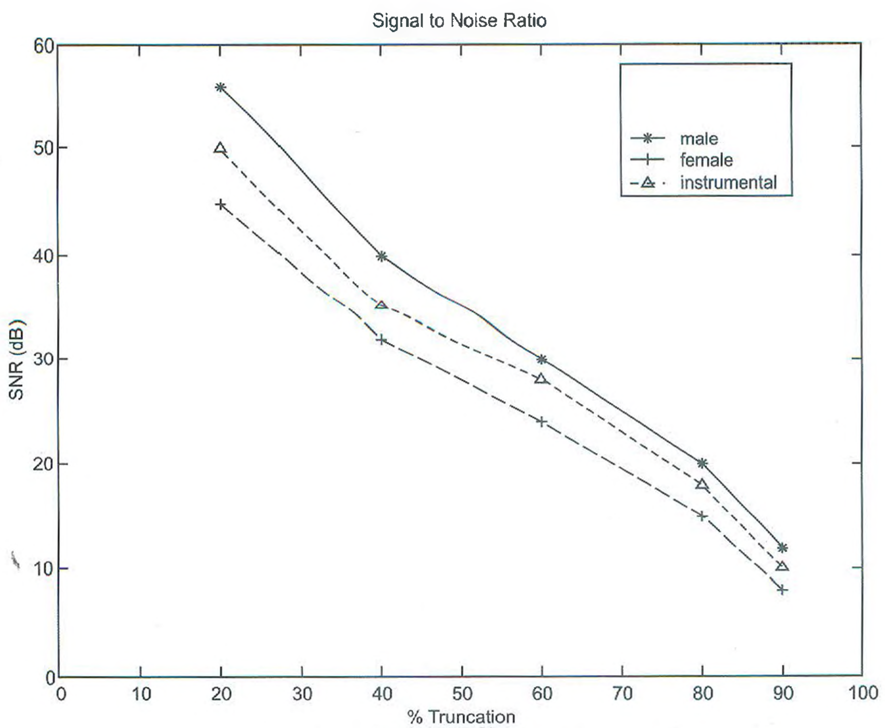

Figure 4: Plot for Signal to Noise Ratio Vs Truncation for Db10 for different test signals

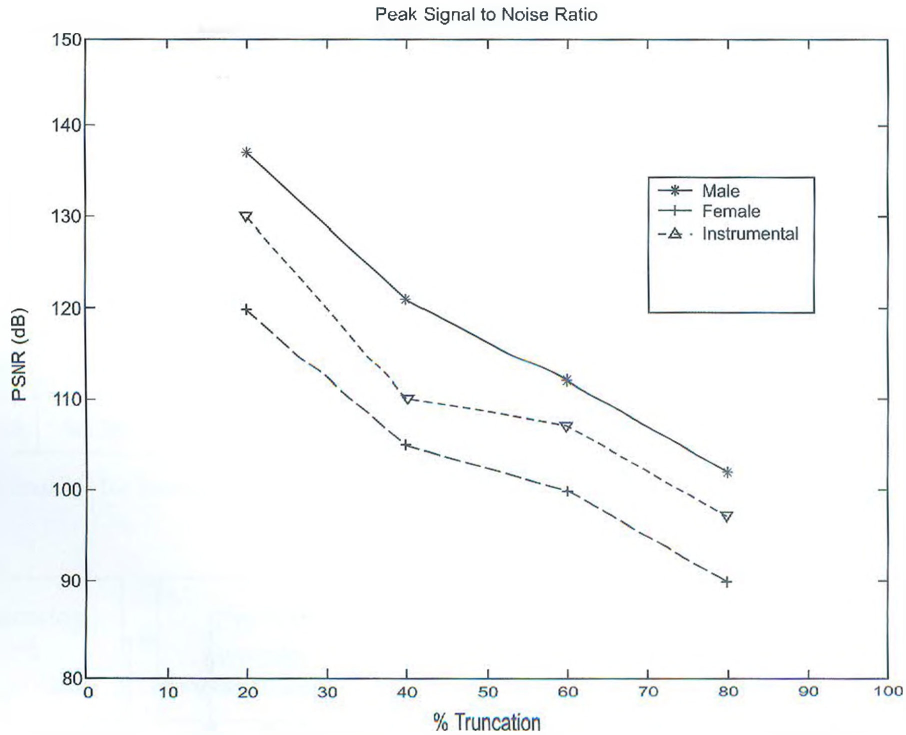

Figure 5: Plot for Peak Signal to Noise Ratio Vs Truncation for Db10 for different test signals

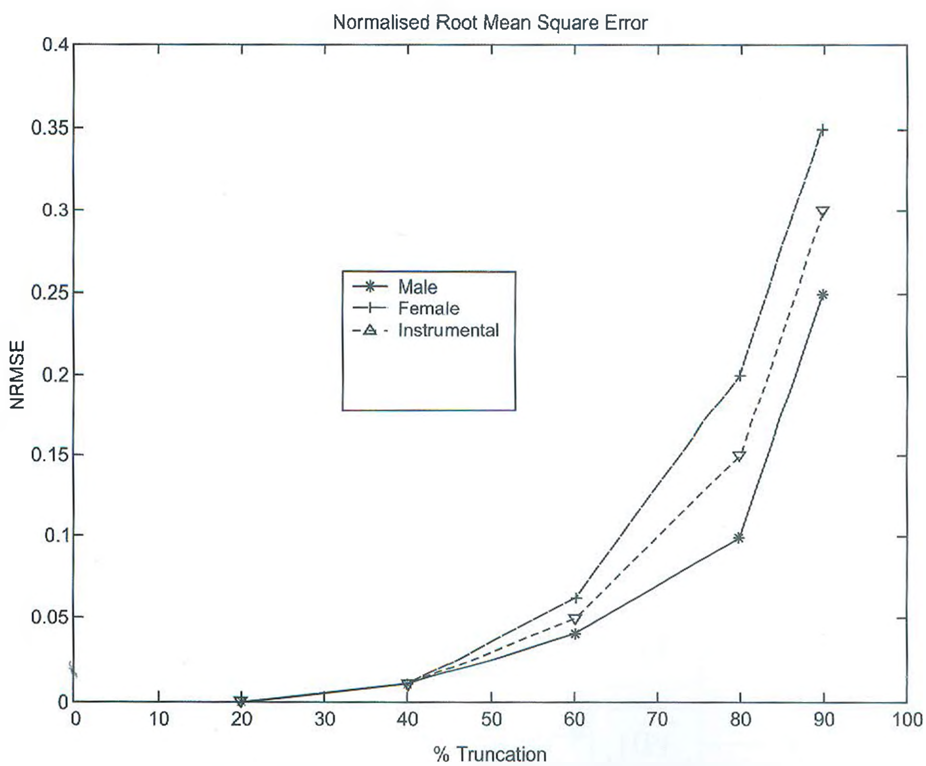

Figure 6: Plot for NRMSE Vs Truncation for Db10 for different test signals

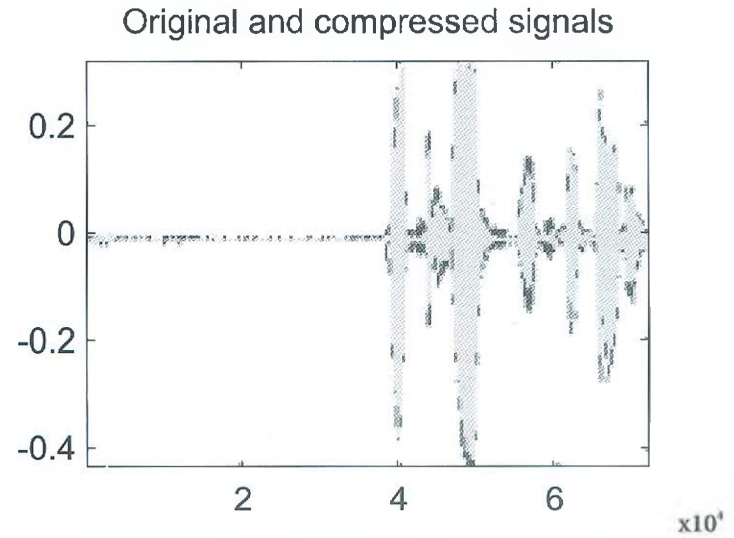

Figure 7: Original and Compressed Waveform for male voice test signal

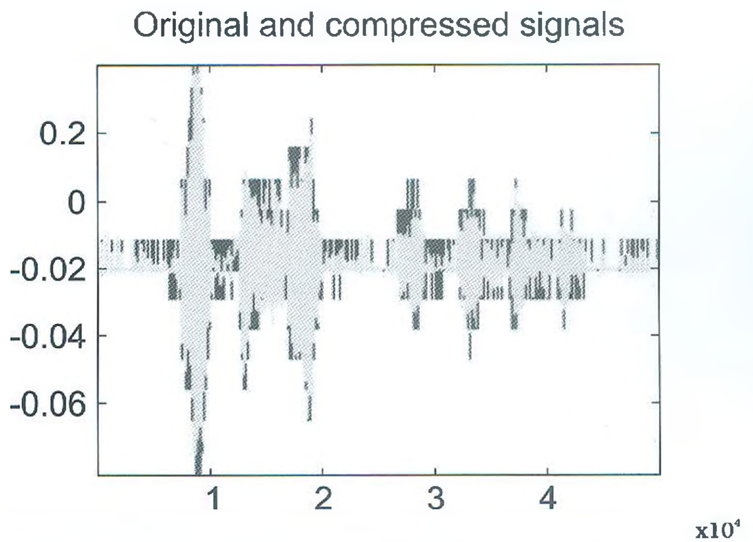

Figure 8 : Original and Compressed Waveform for Female voice test signal

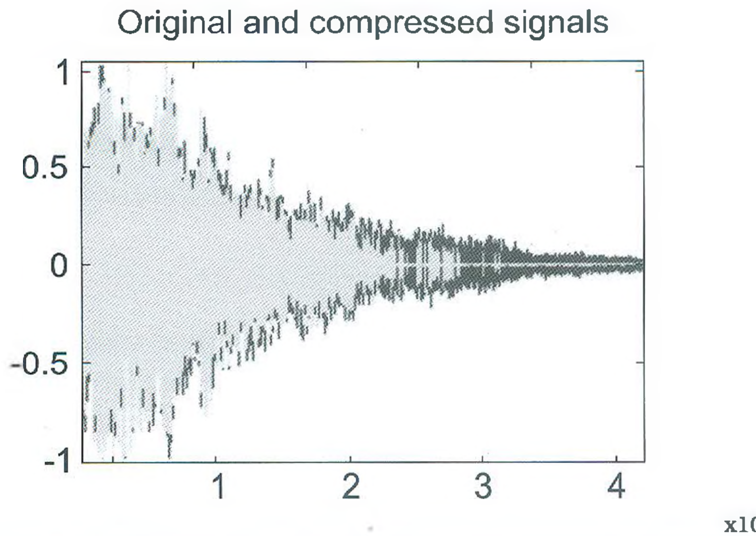

Figure 9: Original and Compressed Waveform for Instrumental sound test signal