AN IMPROVED METHOD OF SPEECH COMPRESSION USING WARPED LPC AND MLT-SPIHT ALGORITHM

MAITREYEE DUTTA[1], DR. RENU VIG[2]
[1] Computer Science Department
National Institute of Technical Teachers' Training and Research
Sector-26
Chandigarh, INDIA
http://www.nitttrchd.ac.in
[2]Department of Information Technology
Panjab University, Chandigarh, INDIA

Abstract: Frequency –warped signal processing techniques are attractive to many wideband speech and audio applications since they have a clear connection to the frequency resolution of human hearing. A warped version of the linear predictive coding (LPC) for speech compression is implemented in this paper and an analysis of the application of Set Partitioning In Hierarchical Trees (SPIHT) algorithm to the compression of speech signals is performed. It has been shown that the proposed scheme i.e. Warped LPC with MLT_SPIHT algorithm produces an enhancement in speech quality . The proposed scheme is based on the combination of the Modulated Lapped Transform(MLT) and SPIHT. Comparisons are made with Plain LPC Coder, Voice Excited LPC Coder with the coding of the residual signal with DCT, Voice Excited LPC Coder with the coding of the residual signal with MLT and SPIHT. The performance of the coders described has been assessed by computer simulation in terms of

a) Signal –to –noise ratio (SNR)
b) Compression ratio
c) Informal subjective listening test.

Keywords:
Plain LPC coder, Voice-Excited LPC coder, DCT, MLT, SPIHT, Warped LPC, SNR

1 INTRODUCTION

Parametric representation of a spectrum by means of linear prediction (LP) is a powerful technique in speech and audio signal processing. For speech coding applications, it was introduced more than 30 years ago and it is still the main tool in that field. The main advantage in speech applications is usually attributed to the all-pole characteristics of vowel spectra. However , it also has clear advantage in terms of human hearing because the ear is obviously more sensitive to spectral poles than zeros[13]. Linear Predictive Coding (LPC) is also more powerful in compressing the spectral information into few filter coefficients for which very efficient quantisation techniques are readily available [14]. The compression of Speech Signals refers to the reduction of the bandwidth required to transmit or store a digitized speech signal. Ideally, the digital representation of a speech signal is coded using a minimum number of bits to achieve a satisfactory quality of the synthesised signal whilst maintaining a reasonable computation complexity.

Most speech coding methods have been designed to remove redundancies and irrelevant information contained in speech, thus aiming toward producing high quality speech with low bit-rates[1]. The optimization of the bit-rate and quality of the synthesised signal is closely related, where an improvement of one aspect compensates to the degradation of the other. Hence, the main development issue usually evolves around the compromise between the need for low rate digital representation of speech and the demand for high quality speech reconstruction[2].
In warped signal processing techniques , the spectral representation of a system is modified. This is typically done by replacing the unit delay elements of a conventional structure by first order all pass filters. Warped linear predictive coding, WLPC , is a clear step forward in the utilization of characteristics of human hearing in designing coding algorithms since a WLPC system can be adjusted so that the spectral resolution closely approximates the frequency resolution of human hearing.

The Set Partitioning In Hierarchical Trees (SPIHT) algorithm [3] is a coding algorithm that allows the transmission of coefficients in a pseudo-sorted fashion where the most significant bits of the largest coefficients are sent first. The sorting is carried out according to the relative importance of the coefficients, determined by Coefficient amplitude, and transmits the amplitudes Partially refining the transmitted Coefficients continuously until the bit limit is reached[3]. The original design of SPIHT was aimed at image compression, and the intent was to use the algorithm in the frequency domain. However the algorithm may be used in the time domain. The SPIHT has been combined with the Modulated Lapped Transform(MLT) to code the LPC residual signal and the results have been compared with DCT based scheme[15]. This work has been published in **WSEAS Transactions on systems**[15]. The performance of **warped LPC technique** with MLT-SPIHT algorithm has been discussed and analysed in this paper.

2. SET PARTITIONING IN HIERARCHICAL TREES

The Set Partitioning In Hierarchical Trees Algorithm (SPIHT) was introduced by Said and Pearlman [3]. It is a refinement of the algorithm presented by Shapiro in [9]. The algorithm is built on the idea that spectral components with more energy content should be transmitted before other components, allowing the most relevant information to be transmitted using the limited bandwidth available[5]. The algorithm sorts the available coefficients and transmits the sorted coefficients as well as the sorting information. The sorting information transmitted modifies a predefined order of coefficients. The algorithm tests available coefficients and set of coefficients to determine if those coefficients are above a given threshold. The coefficients are thus deemed significant or insignificant relative to the current threshold. Significant coefficients are transmitted partially in several stages, bit plane by bit plane.

As SPIHT includes the sorting information as part of the partial transmission of the coefficients, an embedded bit stream is produced, where the most important information is transmitted first. This allows the partial reconstruction of the required coefficients from small sections of the bit stream produced.

3 THE COMPRESSION SCHEME

In this paper, first of all speech signal has been compressed with Plain LPC-10 Vocoder and also with voice excited LPC coder with DCT of the residual

signal. A block diagram of Plain LPC Vocoder is shown in Figure 1[4].

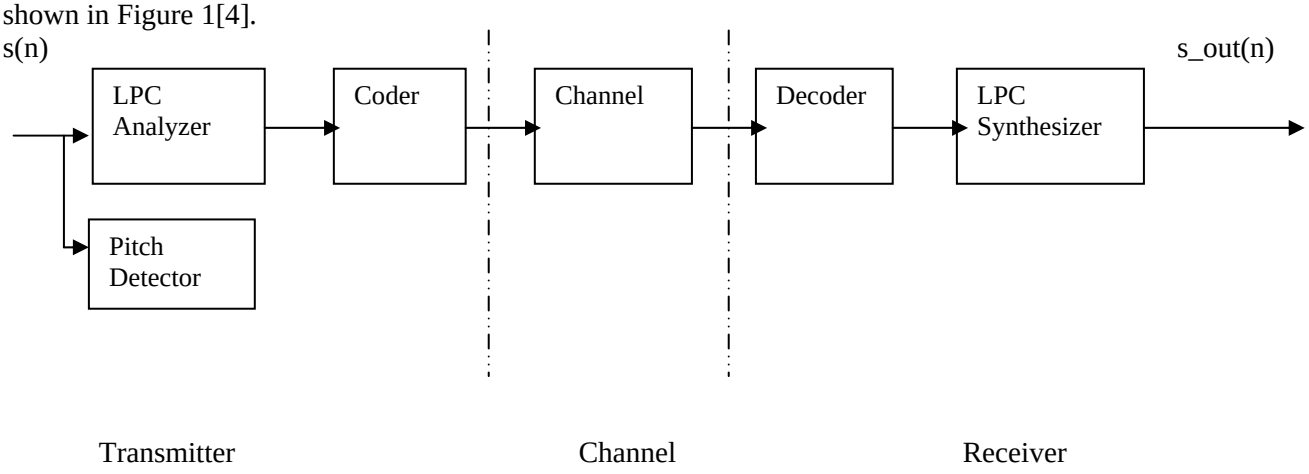


Figure 1: Block Diagram of an LPC vocoder

The principle behind the use of LPC is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. This could be used to give a unique set of predictor coefficients. These predictor coefficients are normally estimated every frame, which is normally 20 ms long. The predictor coefficients are represented by a_k. Another important parameter is gain G. The transfer function of the time varying digital filter is given by :

$$H(z) = \frac{G}{1 - \sum a_k z^{-k}}$$

The two most commonly used methods to compute the coefficients are, but not limited to, the covariance method and the auto-correlation method. For our implementation, we have used the autocorrelation method. Levinson-Durbin recursion will be utilised to compute the required parameters for the auto-correlation method.
The LPC analysis of each frame also involves the decision regarding if a sound is voiced or unvoiced. If a sound is decided to be voiced, an impulse train is used to represent it, with nonzero taps occuring every pitch period. A pitch-detecting algorithm is employed to determine correct pitch period/frequency. We used the autocorrelation function to estimate the pitch

period as proposed in[7]. However, if the frame is unvoiced, then white noise is used to represent it and a pitch period of T=0 is transmitted.From the speech production model it is known that the speech undergoes a spectral tilt of –6dB/oct. To counteract this fact a pre-emphasis filter is used.

3.1 QUANTISATION OF LPC COEFFICIENTS

Usually direct Quantization of the predictor coefficients is not considered. To ensure stability of the coefficients (the poles and zeros must lie within the unit circle in the z plane) a relatively high accuracy (8-10 bits per coefficients) is required. This comes from the effect that small changes in the predictor coefficients lead to relatively large changes in the pole positions. There are two possible alternatives discussed in [7] to avoid the above problem. Only one of them is explained here, namely the partial reflection coefficients(PARCOR). These are intermediate values during

the calculation of the well known Levinson-Durbin recursion. Quantizing the intermediate values is less problematic than quantifying the predictor coefficients directly. Thus, a necessary and sufficient condition for the PARCOR values is $|K_i|<1$.

4 VOICE EXCITED LPC CODER

To improve the quality of the sound, the voice – excited LPC coders are used. Systems of this type have been studied by Atal et al. [8] and Weinstein [9]. Figure 2 shows a block diagram of a voice excited LPC vocoder.

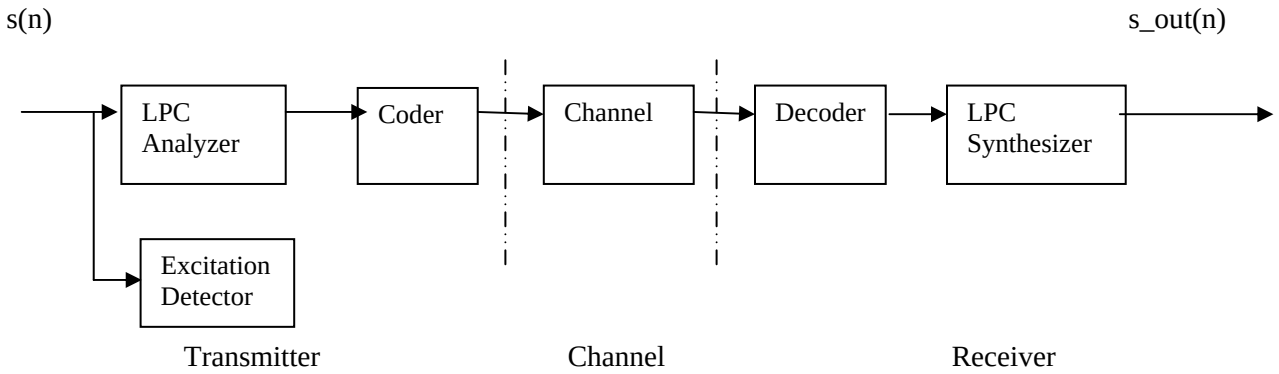


Figure 2: Block Diagram of a voice-excited LPC vocoder

The main idea behind the voice-excitation is to avoid the imprecise detection of the pitch and the use of an impulse train while synthesizing the speech. The input speech signal in each frame is filtered with the estimated transfer function of LPC analyzer. This filtered signal is called residual. If this signal is transmitted to the receiver one can achieve a very good quality.

4.1 DCT OF RESIDUAL SIGNAL

For a good reconstruction of the excitation only the low frequencies of the residual signal are needed. To achieve a high compression rate we employed the discrete cosine transform(DCT) of the residual signal[8][10]. It is known, that the DCT concentrates most of the energy of the signal in the first few coefficients. Thus one way to compress the signal is to transfer only the coefficients, which contain most of the energy. Our tests showed that these coefficients could even be quantized using only 4 bits. The receiver simply performs an inverse DCT and uses the resulting signal to excite the voice.

5 MLT-SPIHT CODER

In MLT-SPIHT coder, the desired order of the predictor can be selected. The residual signal obtained from Linear Predictor Coding method is passed through Modulated Lapped Transform method for compression.

The wavelet transform has been combined with SPIHT in [5] to compress audio. The attractive property of the wavelet transform is the fact that the transform is implemented in a tree structure and so the sets (or trees) originally developed in [4] could still be used. The filter pairs used in [5] were the 20 length Daubechies filter pairs. The sets that are required for SPIHT can be developed as given in [6].

The scheme based on the wavelet transform is diagrammatically represented by Figure 3.

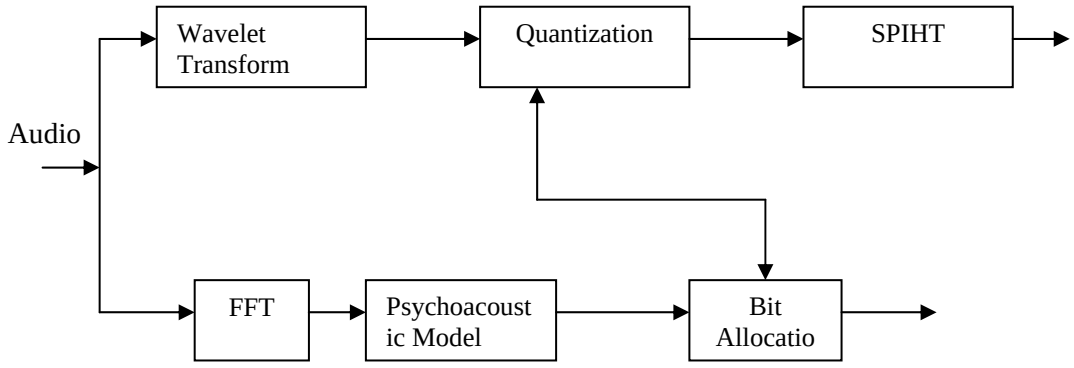


Figure 3: The wavelet based coding scheme

In the scheme shown, the Psycho acoustic model determines the bit allocation that should be used in the quantization of the wavelet coefficients. The results presented by Lu and Pearlman indicated that imperceptible distortion in the synthesized signal could be obtained at bit rates between 55-66 kbps[5].
In this thesis, this wavelet transform with SPIHT has been applied for the compression of the speech signal The results in Table 1 shows the SNR and the compression ratio of the 5 signals listed.

Wav File	Compression ratio	SNR(DB)
C.wav	9.0132	0.1231
John.wav	8.9845	0.0215
Mace.wav	8.9845	0.0225
Target.wav	9.2319	0.0866
E.wav	9.0132	0.0975

Table 1: Results with Wavelet-SPIHT

In figure 4, the codec based on the combination of the MLT with SPIHT is shown.

In this figure, the speech signal is divided into overlapping frames and the MLT is applied to the residual signal of each block. The obtained coefficients are quantized and transmitted by the SPIHT algorithm. At the decoder, SPIHT is used to decode the bit stream received and the inverse transform is used to obtain the synthesized speech signal.

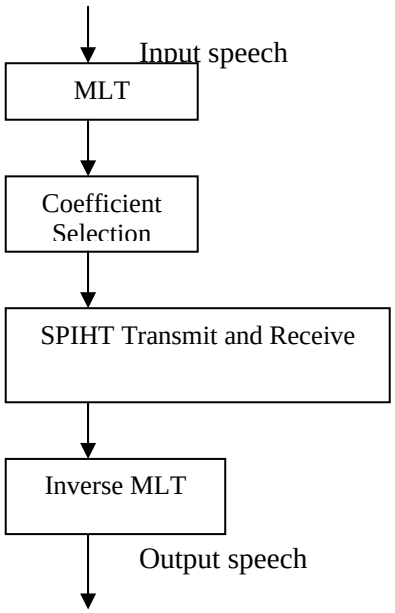


Figure 4: The MLT-SPIHT Coder

The Subband transform(eg. MLT) is one of widely used signal decomposition techniques for data compression as described in Chapter 3.In MLT, the length of the lowpass prototype is chosen as 2M (M is the total number of coefficients). So it has become possible to achieve not only aliasing cancellation, but also perfect reconstruction with identical analysis and synthesis filters. In this case, a window of 2M samples from two consecutive blocks undergoes a Cosine transform , called MLT. The basis functions of the MLT are given by:

$a_{n,k}$ =h(n) 2/M cos[(n+M+1/2)(k+1/2)pi/M]

Where k=0,…..M-1 and n=0,……., 2M-1.
The window chosen is h(n)=sin(n+1/2)pi/2M).

5.1 SETTING UP THE SPIHT SETS

In applying the MLT to an SPIHT based codec, the sets that were used for the wavelet based coding scheme no longer describe the relationship between the transform coefficients appropriately. In [4] sets are based on the tree structure organization of the coefficients, whereas the uniform M-band decomposition carried out by the MLT is a parallel operation.
In the following we define SPIHT sets that link together the frequency domain coefficients for a given frame. The roots of the used sets are at the low frequency end of the spectrum and the outer leaves are at the higher end of the spectrum. Thus the sets link together coefficients in the frequency domain in an order that fits the expectation that the lower frequency coefficients should contain more energy than the higher frequency coefficients. This ordering is similar to, although not the same as, the sets defined in [4].
In this implementation the sets are developed by assuming that there are N roots. One of the roots is the DC-coefficient and because it is not related to any of the other coefficients in terms of multiples of frequency, it is not given any offspring. Each of the remaining N-1 roots are assigned N offsprings. In the next step each of the offsprings is assigned N offsprings and so on, until the number of the available coefficientsis exhausted. The offsprings of any node (i) where (i) varies between 1 and M-1 (M is the total number of coefficients and I=0 is the DC coefficients), are defined as

O(i)=iN + {0, N-1}

Any offspring above M-1 are ignored. The descendants of the roots are obtained by linking the offsprings together. For example, if N=4, node number 1 will have offsprings {4,5,6,7}, node 4 will have offsprings {16,17,18,19} and the descendants of node 1 will include {4,5,6,7,16,17,18,19,……..}.

As part of the development of the M-band transform plus SPIHT coding system, a number of experiments were conducted to determine if the size of N affects the performance of the coder. Figure 3 shows the results of some of these experiments.

6 MASKED MLT-SPIHT CODER

Human auditory masking is a highly complex process which is only partially understood, yet we experience the effects in everyday life. In noisy environments, such as an airport or a train station, noise seems to have a habit of lowering intelligibility just enough so that you miss the last call for the flight or train you have to catch.

Audio codecs utilizing a psychoacoustic model generally calculate per critical band the amount of noise that can be masked. The samples are then quantized to the lowest bit-level allowed so that the introduced noise is still imperceptible. Utilising this system therefore reduces the overall bit-rate of the audio signal and provides the overall bit rate of the audio signal and provides compression[11].
Here in this thesis, the Psychoacoustic model has been used to determine the masked components . The coefficients that are found to be below the masking threshold are set to zero before the quantization is carried out on all coefficients.

7 WARPED LPC

The standard method of transforming a discrete signal precessing system to a warped system involves replacing the unit delays of the original system by first-order allpass filters[11]. For many algorithms, this can be done immediately. However, there are algorithms where the modification is not straightforward.
In classical forward linear prediction[12] an estimate for the next sample value x(n) is obtained as a linear combination of N previous values given by

$$\hat{x} = \sum_{k=1}^{N} a_k x(n-k)$$

$$\hat{X}(z) = [\sum_{k=1}^{N} a_k z^{-k}] X(z)$$

where a_k are fixed coefficients and X(z) is the Z transform of x(n). Here z-1 is a unit delay filter or a shift operator, which may be replaced by a first-order allpass filter D(z) to obtain

$$X(z)= [\sum_{k=1}^{N} a_k D(z)^{-k}] X(z)$$

In the time domain, we define a generalised shift operator $d_k[x(n)]=h(n)* h(n)* h(n)*\ldots.. h(n)*x(n)$, where the asterisk denotes convolution and h(n) is the impulse response of D(z). Furthermore, we denote $d_0[x(n)]=x(n)$. The inverse Z transform of $D(z)^k X(z)$ is $d_k[x(n)]$, and if $D(z)=z-1, d_k[x(n)]=x(n-k)$.

The mean square error of the estimate may now be written as

$$e=E\left[\left[x(n)- \sum_{k=1}^{N} a_k d_k[x(n)] \right]^2 \right] \quad (1)$$

where E[.] is expectation. A conventional minimization procedure leads to a system of normal equations

$$E[d_j[x(n)]d_0[x(n)]]- \sum_{k=1}^{N} a_k E[d_k[x(n)]d_j[x(n)]]=0 \quad (2)$$

With j=0,……N-1.

Where k,j are integers. This states that the same correlation values appear in both terms of the left-hand side of (2). Therefore, (2) can be seen as a generalised form of the Wiener-Hopf equations.

8 Performance Evaluation

The Results of the Four schemes (Plain LPC Coder, Voice Excited LPC Coder, MLT-SPIHT Coder, Masked MLT-SPIHT Coder) are shown in table 2 .

The Graphs are shown in the following figures(Figure 5,6,7) .

The No. (A)---Original Signal

(B)---Reconstructed Signal with Plain LPC Method

(C)--- Reconstructed Signal with Voice Excited LPC Method

(D)--- Reconstructed Signal with MLT-SPIHT Method

(E)--- Reconstructed Signal with Masked MLT-SPIHT Method

Wav Files	Plain LPC Coder		Voice Excited LPC Coder		MLT-SPIHT Coder		Masked MLT-SPIHT Coder	
	Compression Ratio	SNR	Compression Ratio	SNR	Compression Ratio	SNR	Compression Ratio	SNR
C.wav	12.0683	-0.076	12.0683	0.1717	45.3203	0.210	45.3203	0.371
John.wav	12.0704	-0.0208	12.0704	0.0248	45.3281	0.2779	45.3281	0.2110
Mace.wav	12.0704	0.02305	12.0704	0.0423	45.3281	0.1541	45.3281	0.3903
E.wav	12.0683	0.0975	12.0683	0.1295	45.3203	0.1062	45.3203	0.1520
Target.wav	12.0704	0.0701	12.0704	0.1743	45.3281	0.4796	45.3203	0.3961

Table 2: Results shown with four schemes

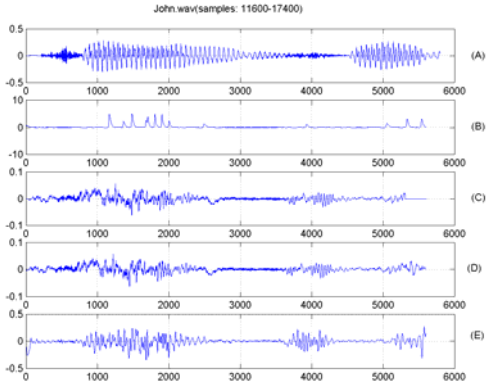


Figure 5 : John.wav

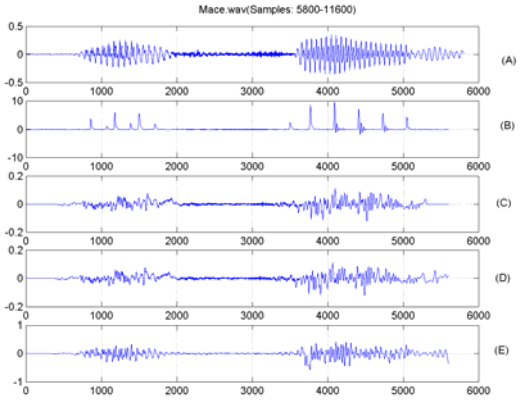
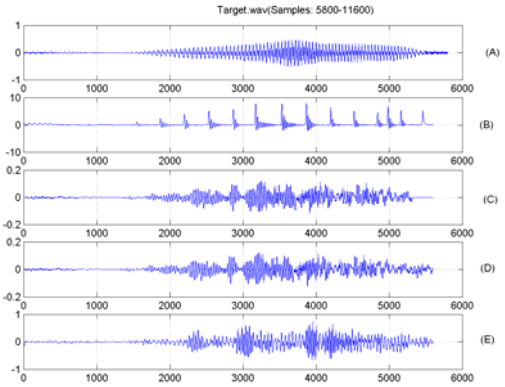


Figure 6 : Mace .wav



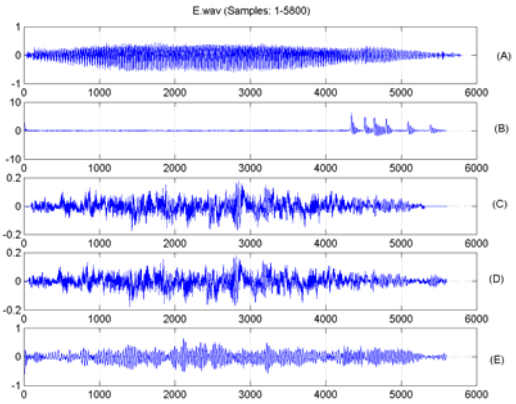


Figure 7: Target.wav and E.wav

As discussed before, the standard method of transforming a discrete signal processing system to a warped system involves replacing the unit delays of the original system by first-order allpass filters. The results in table 3 shows the Compression ratio and SNR with Warped LPC and Warped MLT-SPIHT method.

Wav Files	Warped LPC		Warped MLT-SPIHT	
	Compression Ratio	SNR	Compression Ratio	SNR
C.wav	12.0683	0.0946	12.0683	0.0822
John.wav	12.0704	-0.0584	12.0704	0.3132
Mace.wav	12.0704	0.1243	12.0704	0.3250
E.wav	12.0683	-0.5770	12.0683	-0.7705
Target.wav	12.0704	0.4969	12.0683	0.3715

Table 3: Results shown with Warped LPC and Warped MLT-SPIHT

The Figures(figure 8,9,10,11) Shows the Graphs of Warped LPC (B)and Warped MLT-SPIHT Method(C) with Original signal (A).

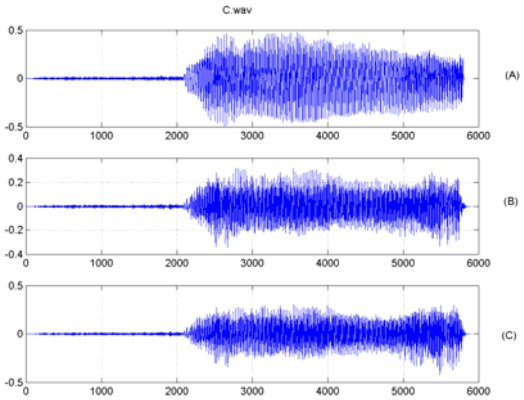


Figure 8 : C.wav

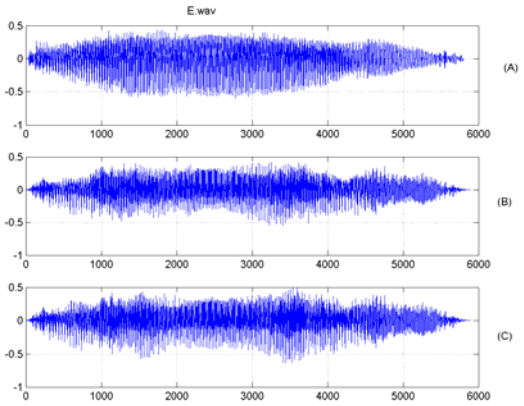


Figure 9 : E.wav

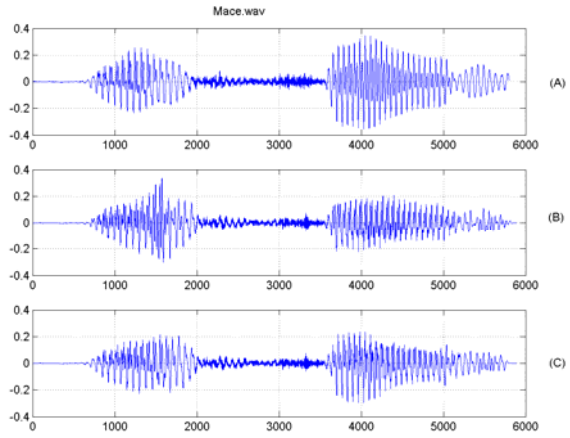


Figure 10 : Mace.wav

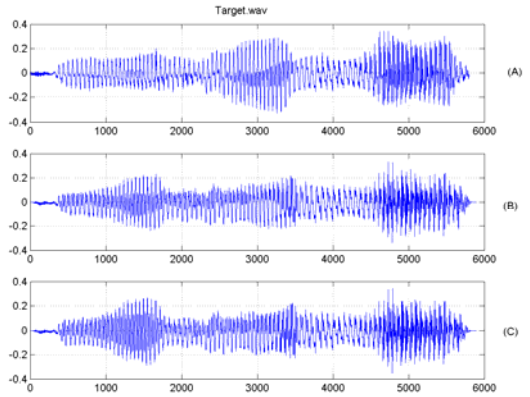


Figure 11 :Target.wav

9 ANALYSIS OF THE RESULT

A comparison of the original speech sentences against the LPC reconstructed speech and the voice –excited LPC methods were studied. In both cases, the reconstructed speech has a lower quality than the input speech sentences. Both of the reconstructed signals sound mechanized and noisy with the output of plain LPC vocoder being nearly unintelligible. The LPC reconstructed speech sounds guttural with a lower pitch than the original sound. The sound seems to be whispered. The noisy feeling is very strong. The voice –excited LPC reconstructed file sounds more spoken and less whispered. The guttural feeling is also less and the words are much easier to understand. The speech that was reconstructed using Voice-excited LPC with the MLT-SPIHT of the residual signal sounded better, but still sounded muffled.

Looking at the SNR computed in Table 2 , it is obvious that the first sound(with plain LPC) is very noisy, having a negative SNR. The noise in this file is even stronger than the actual signal . The voice excited LPC encoded sound is far better, and its SNR, although barely , is in the positive side. However, even the speech coded with the improved voice-excited LPC e.g. Voice-excited LPC with the MLT-SPIHT of the residual signal does not sound exactly like the original signal. The noise has been minimized and in turn an improvement has ben obtained in the Signal to Noise ratio with Masking algorithm. From the Table 3 , it is understood that with warped LPC , the SNR of the signals have been improved. The informal listening test implies that the quality has also been improved.

The LPC method to transmit speech sounds has some very good aspects, as well as some drawbacks. The huge advantage of vocoders is a very low bit rate compared to what is achieved for sound transmission. On the other hand, the speech quality achieved is quite poor. The Voice-excited LPC with the MLT-SPIHT of the residual signal gives better compression ratio and SNR but even then the quality achieved is still not very good. The quality is proved with the application of Masking. The subject quality Mean Opinion Score also proves that The Warped LPC with MLT-SPIHT gives us the enhancement in speech quality.

10 CONCLUSION:

This paper has presented a new method of speech compression based on Modulated Lapped Transform and SPIHT and Warped LPC. The results show clearly that significant savings may be obtained if the MLT is used in place of DCT. The results presented have also highlighted the advantage of the SPIHT algorithm, combined with relevant transform coefficient relationships, to scalable coding of speech as the algorithm is designed with the aim of producing an embedded bit stream. The use of Warped LPC technique with MLT-SPIHT gives us the better quality of the signal. The reduction in bandwidth has been more significant because of the use of the masking model in the coding.

References:

[1] M.H.Johnson and A Alwan, " Speech Coding: Fundamentalsand Applications," to a chapter in the Encyclopedia of Telecommunication, Wiley, December, 2002

[2] Ozgu Ozun, Philipp Steurer and Deniel Thell," Wideband Speech Coding with Linear Predictive Coding (LPC) " in EE214A: Digital Speech Processsing , Winter 2002

[3]Amir Said and Willium A Pearlman, "A new, fast and efficient image codec based on Set Partitioning in hierarchical trees," IEEE Transactions on Circuits and Systems For Video Technology, vol. 6, no. 3, pp. 243-250, June 1996

[4] C.J.Weinstein, " A Linear Predictive Vocoder with Voice Excitation," Proc. Eascon, September 1975.

[5] Lu Z, Kim DY, Pearlman WA, " Wavelet Compression of ECG signals by the Set Partitioning in Hierarchical Trees Algorithm", in IEEE Transactions on Biomedical Engineering Vol. 47, no. 7 PP. 849-856, June 2000.

[6] M.Hans and R.W Schafer, " Lossless Compression of digital audio," IEEE Signal Proc. Mag. Vol. 18, PP 21-32, July 2001

[7] Changyou Jing and Heng-MingTai," A fast algorithm for computing modulated lapped transform , in Electronics Letters, 7th June 2001 vol.37 no.12

[8] De4bargha Mukherjee, Sanjit .K.Mitra, Image resizing in the Compressed domain using Subband DCT, in IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12 No. 7, July 2002

[9] Shapiro J.M. (1993), "Embedded image coding using zerotrees of wavelet coefficients " in IEEE Transactions on signal Processing, Vol. 41, No.12,pp.3445-3462.

[10] T.Liebchen, M.Purat, and P.Noll," Lossless transform coding of audio signals," Proceedings of the 102nd AES convention, AES Preprint 4414, March 1997.

[11] Aki Harma, and Unto K.Laine, "A Comparison of Warped And Conventional Linear Predictive Coding", IEEE Transactions on Speech and Audio Processing , Vol. 9, No.5, July 2001.

[12] R.Merced and A.H.Sayed, " Exact RLS Laguerre –lattice adaptive filtering", Proc.IEEE Int. Conf.Acoustics, Speech, Signal Proc.,vol. 1, pp. 456-459, June 2000

[13] M.R.Schroeder, " Linear prediction, external entropy and prior information in speech signal analysis and synthesis, " Speech Commun., vol. 1, no. 1, pp. 9-20, May 1982.

[14] K.K.Paliwal and W.B.Kleijin, " Quantisation of LPC parameters, " in Speech Synthesis and coding, Amsterdam, The Netherlands: Elsevire, 1995, pp. 433-466. In [46]

[15] Dutta Maitreyee and Dr. Vig Renu "Speech Compression With Masked Modulated Lapped Transform And SPIHT Algorithm " in WSEAS Transactions on systems , vol. 3 Issue 11, November,2005.