An Efficient Time Domain Speech Compression Algorithm Based on LPC and Sub-Band Coding Techniques

P.Venkateswaran
Dept. of Electronics & Tele-Communication Engg., Jadavpur University, Kolkata-700032. (INDIA)
Email: pvwn@ieee.org

Arindam Sanyal[1], Snehasish Das[2], R.Nandi[3] , S.K.Sanyal[4]
Dept. of Electronics & Tele-Communication Engg., Jadavpur University, Kolkata-700032. (INDIA)
Email: arindam_3110@yahoo.co.in[1], snehasish.das@tcs.com[2], robnon@ieee.org[3], s_sanyal@ieee.org[4]

Abstract— **Speech compression is a mature technology with many applications. Over the past decade, huge advances have been made in the area of speech coding for reduced bit-rate transmission. With perceptual audio coding, the signal is coded efficiently using a psychoacoustic model, as in MPEG standards. In this paper, two well known algorithms, LPC (Linear Predictive Coding) and subband coding are combined to reduce data transmission rate from 128 Kbps to as low as 12 Kbps at minimum complexity and implemented on a digital signal processor. The performance of the proposed algorithm is almost same as the MPEG algorithm without its complexity.**

Index Terms— **Digital signal processors, Levinson-Durbin algorithm, Linear Predictive Coding, RELP, subband coding.**

I. Introduction

Efficient speech coding [1] at low bit rates find many applications for bandwidth conservation and enhanced privacy in wireless cellular and satellite communications. Digital speech coders attempt to reduce the bit rate while minimizing the quantization error. Time Domain coding methods such as delta modulation or sigma delta modulation operate across full bandwidth and can achieve compression ratios up to 2.5:1[2]. Frequency domain coding methods analyze the signal in frequency domain and reduce the quantization error using a psychoacoustic model.

The time domain and frequency domain methods can be combined to yield the best of both the methods. In this work we have studied the effects of combining a time domain method, LPC coding, with frequency domain method, subband coding, to achieve compression ratios of up to 7:1 at minimum complexity. Both the LPC and subband coding are well known algorithms. Subband coding is particularly suitable for speech compression as speech energy is mostly concentrated in the low frequency bands.

The main motivation of the present work is to develop an algorithm which can be easily embedded in low power, portable systems and hand held devices. As such, the algorithm has to be simple yet provide performances matching those of the other algorithms in use. In this present work, it has been shown that a combination of simple algorithms like LPC and subband coding can match the compression ratios of the other complex algorithms using psychoacoustic analysis. A digital signal processor is used to evaluate the performance of the algorithm in hardware, though the same could have been done on an FPGA platform.

The speech compressor used in the present work has a synthesis by analysis structure. The block diagram of the analysis and synthesis sections of the proposed speech compressor is shown in Fig.1.

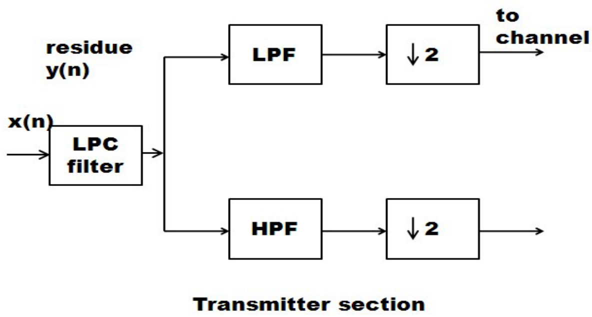

Fig.1.a. Transmitter (Analysis Section)

Based on the paper "An Efficient Time Domain Speech Compression Technique and Hardware Implementation on TMSC3205416 Digital Signal Processor", preliminary version published by Arindam Sanyal, Snehasish Das, P.Venkateswaran, S.K.Sanyal and R.Nandi which appeared in the Proceedings of the 1st IEEE International Conference on Signal Processing, Communications and Networking- ICSCN 2007, Anna University, Chennai , India, Feb. 2007.

Manuscript received July 26, 2008; revised December 30, 2008; accepted January 21, 2009.

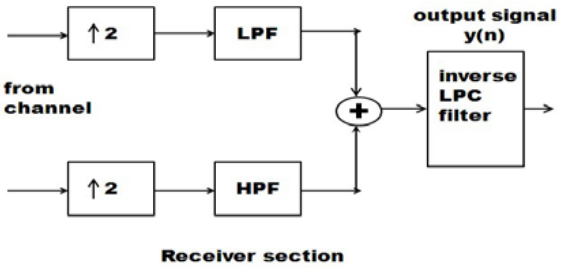

Fig. 1.b. Receiver (Synthesis Section)

The paper is organized as follows: the RELP coder is described in Section II, the subband coding principle is described in Section III, and the hardware implementation and results are discussed in Section IV. Section V brings up the conclusion and possible extensions to the present work.

II. RELP CODER

Linear prediction [3] is an important topic in digital signal processing with many practical applications. In our proposed speech compressor, we have used a one-step forward linear predictor, which forms the prediction of the value x(n) by a weighted linear combination of the past values x(n-1), x(n-2), …., x(n-p) where p is the order of the predictor. Hence the linearly predicted value of x(n) is

$\tilde{x}(n) = -\sum_{k=1}^{p} a_p(k)\; x(n-k)$ where $\{a_p(\mathrm{k})\}$ represent the prediction coefficients of the linear predictor. The difference between the value x(n) and the predicted value is called the forward prediction error, denoted as

$$f_p(n) = x(n) - \tilde{x}(n)$$

$$= x(n) + \sum_{k=1}^{p} a_p(k)\, x(n-k)$$

$$= \sum_{k=0}^{p} a_p(k)\, x(n-k)\ ,\ a_p(0) = 1$$

An equivalent realization of the prediction error filter is shown in Fig. 2.

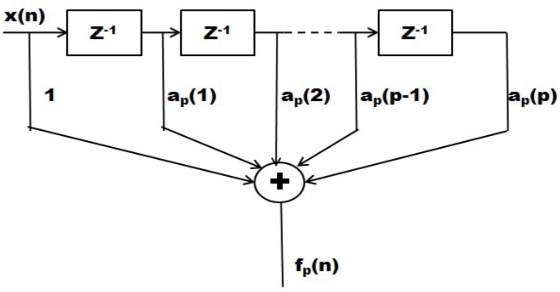

Fig. 2. Prediction-error filter

Thus, we can model the prediction error filter as an all-zero FIR filter whose z-transform relationship is given by $F_p(z) = A_p(z)\ X(z)$. At the receiver side, we have an inverse filter with transfer function $(1/A_p(z))$,which is an all-pole filter, to recover the original signal. The minimization of the mean-square value of the forward prediction error results in a set of normal equations, expressed in compact form as

$$\sum_{k=0}^{p} a_p(k)\,\gamma_{xx}(l-k) = 0 \quad l = 1, 2, ..., p\ ;\ a_p(0) = 1$$

where $\gamma_{xx}(l) = \sum_{n=i}^{N-|k|-l} x(n)\, x(n-l)$, N is the length of the causal sequence x(n), i = l, k = 0 for $l \geq 0$, and i =0 , k = l for $l < 0$

The Levinson-Durbin algorithm is a computationally efficient algorithm for solving the normal equations. This algorithm exploits the special symmetry in the autocorrelation matrix

$$\begin{bmatrix} \gamma_{xx}(0) & \gamma^*_{xx}(1) & \cdots & \gamma^*_{xx}(p-1) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \cdots & \gamma^*_{xx}(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{xx}(p-1) & \gamma_{xx}(p-2) & \cdots & \gamma_{xx}(0) \end{bmatrix}$$

The desired recursion for the predictor coefficients in the Levinson-Durbin algorithm is

$$\mathbf{a}_m(m) = -\frac{\gamma_{xx}(m) + \gamma^{bt}_{m-1}\,\mathbf{a}_{m-1}}{\gamma_{xx}(0) + \gamma^{bt}_{m-1}\,\mathbf{a}^{b*}_{m-1}}$$

$$\mathbf{a}_m(k) = \mathbf{a}_{m-1}(k) + \mathbf{a}_m(m)\,\mathbf{a}^*_{m-1}(m-k)\ \text{,k=1,2,…..,m-1;}$$

m=1,2,…., p.

$$\gamma^{bt}_{m-1} = [\gamma_{xx}(m-1) \quad \gamma_{xx}(m-2) \ \cdots \ \gamma_{xx}(1)]$$

and γ_m^t denotes the transpose of γ_m . Thus, using the Levinson–Durbin algorithm, we can construct the all-zero prediction error filter at the transmitter and subsequently, the all-pole filter at the receiver to re-construct the original signal. The residual signal essentially carries all the information that has not been captured by the LP analysis, e.g., phase and pitch information, zeroes due to nasal sounds etc. In RELP, the residual encoding is based on spectral rather than waveform matching. In addition, RELP coders rely on the fact that the low-frequency components of speech are perceptually important. This residual signal is then subsequently divided into two subbands for preferential encoding. Since only the residual is processed at the subsequent stages, RELP allows sufficient bit compression.

III. SUBBAND CODING

Subband coding [4] is an effective method to achieve data compression in speech signals, where the signal energy is concentrated mainly in the low frequency region. We have divided the 8 KHz sampled speech signal into two frequency regions – below 2 KHz and, above 2 KHz. Further, sub-division of the speech signal yields no comparable increase in speech compression ratio. The decimated signal is then preferentially encoded by allocating larger number of bits to the low frequency region than the high frequency region. We have used QMF (Quadrature Mirror Filters) [5] to decimate and subsequently interpolate the signal. The filter characteristics of the low pass and high pass QMFs as created using the C5416 DSP chip are given in Fig. 3

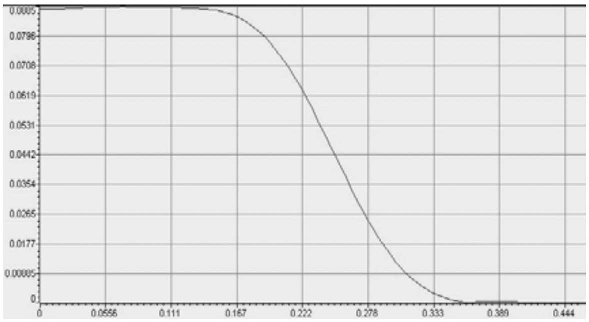

Fig. 3.a. Filter characteristics of low pass filter created with C5416 Processor

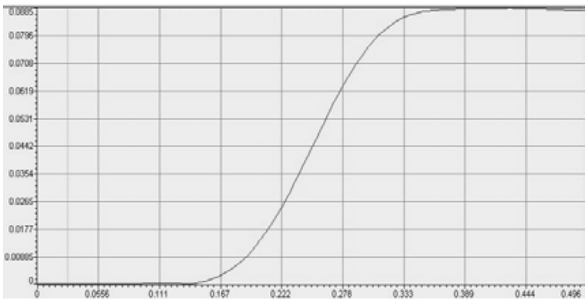

Fig.3.b. Filter characteristics of high pass filter created with C5416 Processor

With the choice of the above filter characteristics, the component due to aliasing vanishes. Thus the aliasing resulting from decimation in the analysis section of the transmitter is perfectly cancelled by the image signal spectrum that arises due to interpolation at the receiver. To save computations, we performed the decimation and interpolation using polyphase filter structures as shown in Fig. 4.

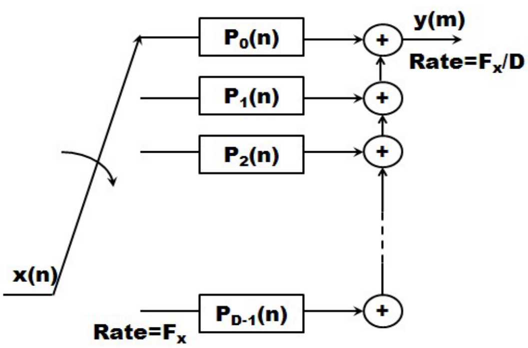

Fig. 4.a. Decimation by use of Polyphase filters

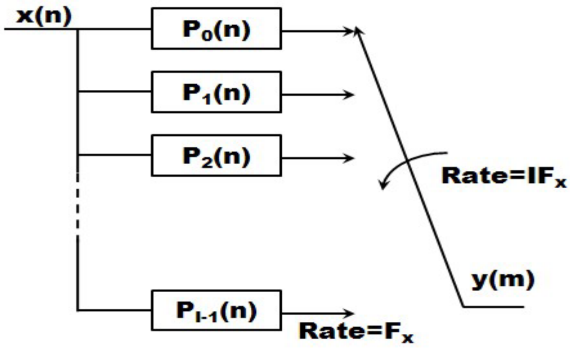

Fig. 4.b. Interpolation by use of Polyphase filters

IV. HARDWARE IMPLEMENTATION AND RESULTS

The incoming signal, in the form of 16 bit PCM samples, is passed through a predictor block which uses the history of previous samples to predict the next signal. This prediction is compared with the actual signal and the difference is measured. The encoder in the transmitting block takes this difference signal and passes this through a QMF (Quadrature Mirror Filter) bank. This specially designed filter splits the signal into two equally divided subbands [6] according to the frequency content. Through this process, the QMF also down samples the clock rate to half the original clock. In the subband processor, the 16-bit word is then passed through a quantizer routine which allocates greater number of bits to the low frequency region than the high frequency region. The decoding process is the inverse of the encoding process, where the bit stream is split into two subband signals and the inverse quantizer and predictor is used to regenerate the original signal.

A 1-minute speech was recorded using a microphone and the voice recorder provided with Windows XP in 16-bit mono mode and sampling frequency 8 KHz, for our experiments. This file recorded in .wav format was then provided as an input to the C5416 processor [7], in frames of 320 samples. The compressed file as well as the recovered file was stored in the hard disk of a computer which was interfaced with the C5416 processor. The setup is shown in Fig. 5 below.

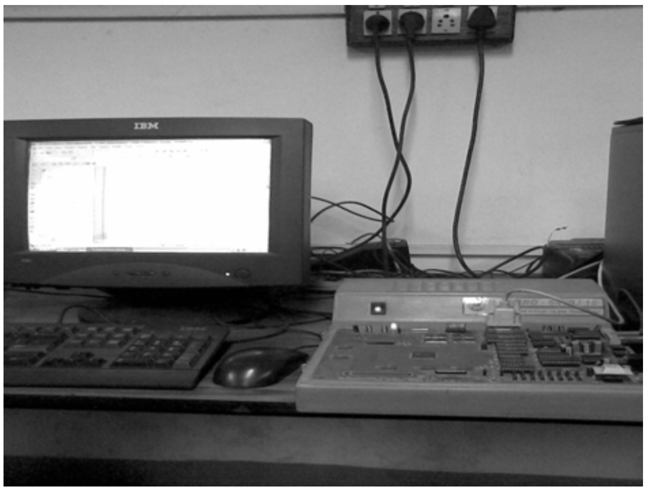

Fig. 5. The setup used in the present work

The spectra of the original as well as the recovered signals were then viewed using a spectrum analyzer. While processing the original file, the first 44 samples, which form the header, were first extracted and later were added to the recovered file. For every 320 input samples, a corresponding frame containing the encoded data was written to the compressed file. The decoder at the receiving side generated the 320 samples from this frame. The first six samples, each of 16-bits, contained the LPC parameters. The next two 16 bit samples contained the gains in the two frequency subbands. The remaining samples contained the encoded speech data. Table I shows the number of bits with which a 16- bit sample in low frequency band and a 16-bit sample in high frequency band are encoded , and the corresponding compression ratios achieved.

TABLE I

VARIOUS COMPRESSION RATIOS ACHIEVED

Number of bits used in encoding the low frequency samples	Number of bits used in encoding the high frequency samples	Original file size	Compressed file size	Compression achieved
4	2	937 kB	223 kB	76.2% (4.2 : 1)
4	1	937 kB	194 kB	79.3% (4.8 : 1)
2	1	937 kB	135 kB	85.6% (6.9 : 1)

For comparison of the compressed signals with the original signal, we have given the signal spectrum and the time domain representations of the original as well as the recovered signals (Fig. 6.a – 6.h) for various values of the number of bits used to encode the high and low frequency regions respectively. The time domain representations show the first 2 ms of the speech signals. As can be seen from the time-domain representations, the recovered signals, and the original signal are identical in the location of the peaks and the energy distribution.

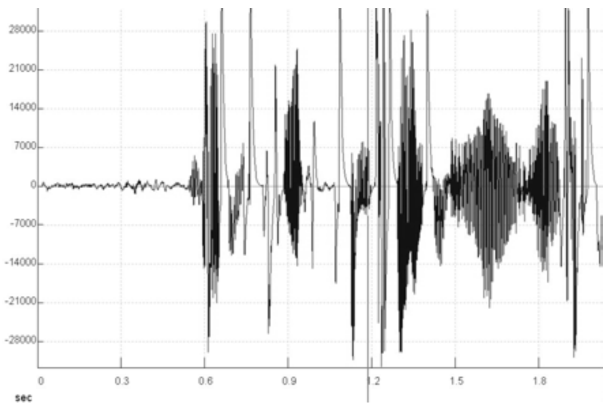

Fig. 6 b. Time –domain representation of the original signal (from 0 – 2s)

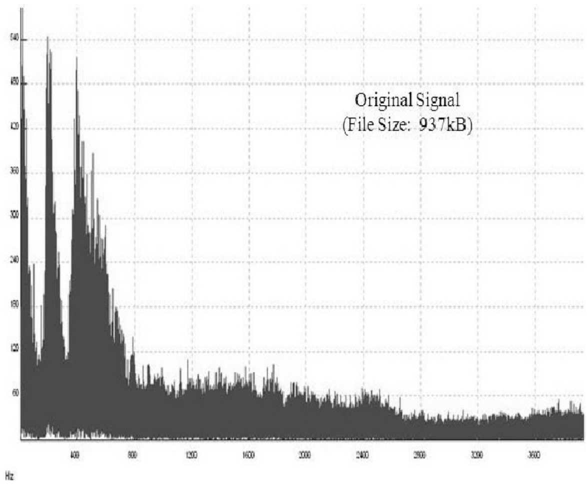

Fig. 6 a. Spectrum of the original signal

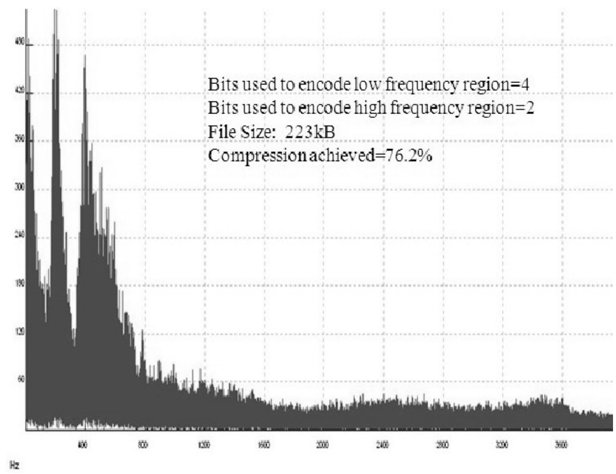

Fig. 6 c. Spectrum of the 76.2% compressed signal

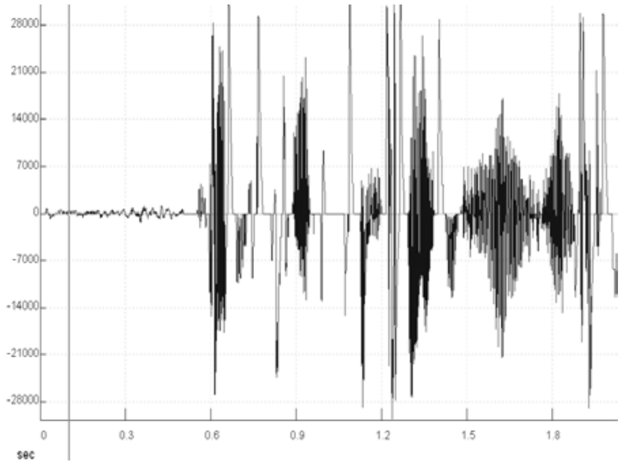

Fig.6 d. Time –domain representation of the 76.2% compressed signal (from 0 – 2s)

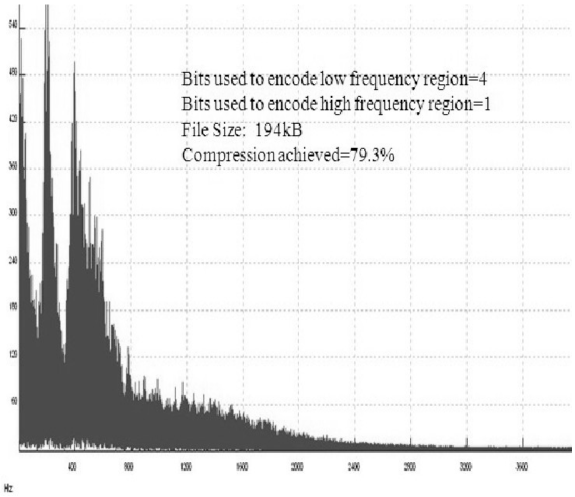

Fig. 6 e. Spectrum of the 79.3% compressed signal

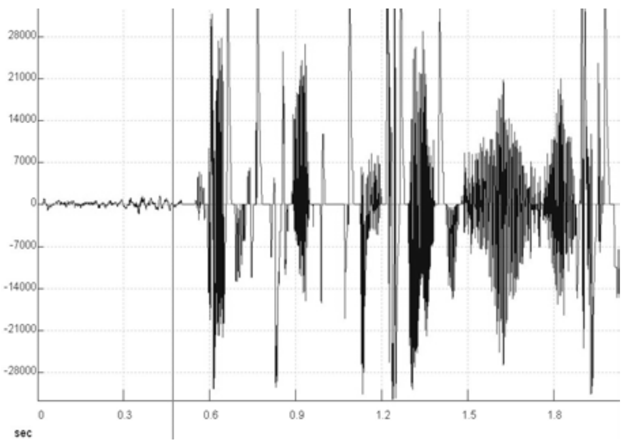

Fig. 6 f. Time –domain representation of the 79.3% compressed signal (from 0 – 2s)

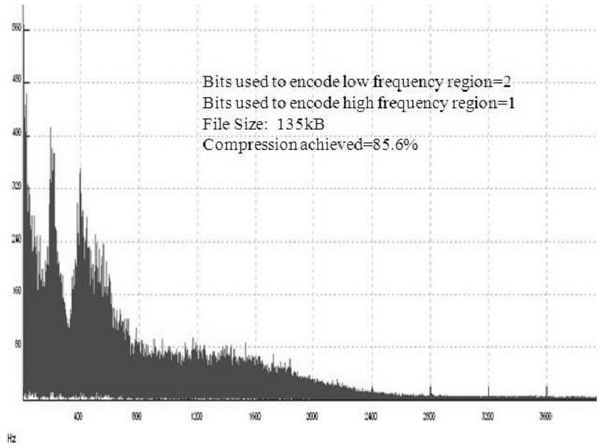

Fig. 6 g. Spectrum of the 85.6% compressed signal

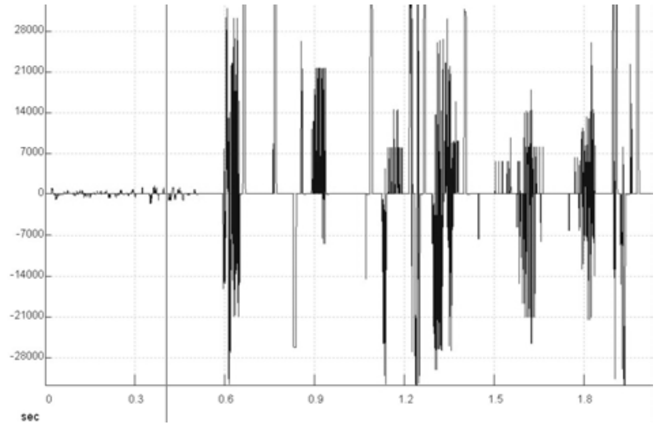

Fig. 6 h. Time –domain representation of the 85.6% compressed signal (from 0 – 2s)

From the spectra in Fig. 6, we see that the low frequency region of the speech signal containing most of the speech energy is almost unaffected due to bit compression. The high frequency spectrum gets distorted by the encoding of the speech signal. However, the information loss in the high frequency region does not affect the overall quality of the speech signal by a drastic amount, and the desired information can very well be extracted from the recovered signals. For processing the speech signal, we interpreted the 16-bit integer samples as q15 numbers, and then converted them to their equivalent floating point numbers. In q15 format, a number between -1 and ($1- 2^{-15}$) can be represented by 16-bits, in which the first bit represents the sign of the number and the next 15-bits represent the magnitude of the number. By converting the numbers to their equivalent floating point versions in this fashion, we restrict the numbers between -1 and ($1- 2^{-15}$) , thus reducing the quantization error. After extracting the residue through LP analysis, the range of the numbers is so much reduced that we can encode each 16-bit sample with less than 4 bits even while maintaining sufficient accuracy. Also, we calculate correlation between two samples bit-wise which results in greater accuracy, thus permitting us to compress the speech greatly. Also better results can be obtained if we apply a pre-emphasis filter to enhance the high frequency region of the speech before performing LP analysis. Such a signal would then have to be de-emphasized at the receiving end by a de-emphasis filter after performing inverse LP filtering.

V. Conclusion

Speech coding is an essential application of digital signal processing in modern day telephony and mobile communications, which employ high data compression ratios. In this paper, we have shown a simple method of effective speech compression technique combining the well known LPC analysis and subband coding. Using only RELP would achieve a compression ratio of 60% [8], while a combination of RELP and subband coding results in a compression ratio as high as 85.6%. We have reduced the transmission bandwidth to 12 Kbps from 128 Kbps, or a bandwidth reduction factor of 11:1, which is comparable to the MPEG (12:1 compression ratio) [9] without its complexity. The compression ratios achieved are also slightly better than a similar project undertaken at Communications University of Maryland, College Park, Epson Palo Alto, R&D lab, in which the compression ratios achieved are between 57.78% to 82.45% [10]. We have also presented the results of our experiments on speech compression which shows good improvement on Bandwidth with acceptable distortion level. Implementation of this system with a C5416 Digital Signal Processor is also given.

The number system used in this work is q15 format which is also suitable for FPGA (Field Programmable Gate Array) or ASIC (Application Specific Integrated Circuit) implementation. Also, the algorithm used can readily be implemented in VHDL or Verilog. Thus, this speech compression method can be extended to any low power hand held device, which can store a large number of audio files in compressed format, and decompress and play a file when required. The speech compression method is also far less complex than standard compression algorithms, thus requiring less gate count, and hence less area and power, during FPGA or ASIC implementation.

References

[1] J. D. Gibson : Speech Coding Methods, Standards, and Applications. IEEE Circuits and Systems Magazine. Vol. **5**, No. 4, Fourth Quarter 2005, pp. 3049.

[2] Tomas Kratchovil and Martin Sevcik : Frequency Analysis of Low Bit Rate Modern Digital Audio Compression Algorithms. Institute of Radio Electronics, Brno University of Technology.

[3] P. Strobach : Linear Prediction Theory-A Mathematical Basis for adaptive Systems. Springer-Verlag, 1990.

[4] John G. Proakis and D. J. Manolakis : Digital Signal Processing, PHI, 3/e, pp.864-868(July 2003).

[5] S. Sridharan.: Handbook of Neural Networks for Speech Processing. 1/e, Ed. S. Katagiri, Artech House, Chapter 4(199).

[6] Neeta Awasthy, J.P.Saini and D.S.Chauhan : Spectral Analysis of Speech: A New Technique. International Journal of Signal Processing, Vol. 2, no.1, 2005, ISSN 1304-4494.

[7] Texas Instruments.:TMS320C54X Optimizing C/C++ Compiler User's Guide., Literature Number SPRU103G.

[8] Biren Mohanty, "Some Studies On Excitation Methods Of RELP Vocoder For Speech Coding At Low Bit Rate", M.E Tel. E. Thesis Work, pp.62, Jadavpur University, Kolkata-700 032, India.

[9] Clark Novak, "Spectral Band Replication and AACPlus Coding-An Overview", Telos Systems, Cleveland, Ohio.

[10] Ali Zandifar and Shahrooz Shahparnia, "ENEE 6244 Project 2:Linear Prediction for Speech Coding and Recognition", Communications University of Maryland, College Park, Epson Palo Alto, R&D Lab – URL.

P. Venkateswaran is working as a Reader in the Dept. of Electronics & Tele-Communication Engg. (ETCE), Jadavpur University (JU), Kolkata, India. He has published over 30 papers in various National / International Journal / Conference Proceedings. His fields of interest are Computer Communication, Microcomputer Systems and Digital Signal Processing (DSP). He is a Member of IEEE (USA), and Present Secretary of Calcutta Chapters of IEEE Communications Society (ComSoc) & IEEE Circuits And Systems Society (CASS).

Arindam Sanyal has completed his Bachelor's degree from Dept. of ETCE, JU in 2007. He is currently pursuing his Masters degree at Indian Institute of Technology (IIT), Kharagpur in Microelectronics and VLSI discipline. His current research interest is RFIC.

Snehasish Das has completed his Bachelor's degree from Dept. of ETCE, JU in 2007. He is currently an Assistant Systems Engineer with Tata Consultancy Services (TCS) Limited, Kolkata.

Dr. Rabindranath Nandi is a Professor in the Dept. of ETCE, JU. His areas of interest are ASP, DSP, Computer Communication. He has authored more than 110 research papers in National / International Journals and some in Conferences / Seminars. He served as the Head of ETCE Dept., JU during 1999-2001 and served as the Chair of IEEE Calcutta Section during 2003-2005. He has taught in various Institutes abroad.

Dr. Salil Kumar Sanyal is a Professor and Head of the Dept. of ETCE, JU. He has published more than 100 research papers in International / National Journals / Conference Proceedings. His current research interests include Analog Signal Processing (ASP) & DSP, Tunable Microstrip Antenna, Communication Engineering, VLSI Circuits & Systems Design. He served as the Chair of IEEE Calcutta Section during 2006-2008.