LINEAR PREDICTIVE CODING

[1]RANJITHA G S, [2]PALLAVI S M, [3]SAMEER R KASHYAP, [4]NAMRATA PURVIMATH, [5]SPOORTHI PUJARI

[1,2]Department of Electronics, R.V College of Engineering, Bangalore, India
[3]Department of Electronics, BMS College of Engineering, Bangalore, India
[4,5]Department of Information Science, R.V College of Engineering, Bangalore, India

Abstract— Linear predictive coding (LPC) is a technique used to compress audio and speech signals in digital processing. Subjective analysis using LPC-10 technique has been adopted in this paper. The voice signal which is given as the input is compressed and the output with low bit rate has been simulated in Matlab.

Keywords— Linear predictive coding, Subjective analysis, MATLAB.

I. INTRODUCTION

Speech coding is the technique of compressing audio signals using various compression algorithms. The algorithms are Linear Predictive Coding, Waveform Coding and Sub-band Coding. Waveform coding algorithm consists of sampling and quantizing the input signal. Sub-band Coding decomposes the input signal into different frequencies and encodes them individually. Linear Predictive Coding, which is used in this paper employs Envelope Spectrum to compress the input signal at low bit rate.
The four main attributes of LPC vocoders are bit rate, delay, complexity and quality. LPC takes advantage of these characteristics and enables transmission of data over low bandwidth using less information. This property finds applications in military where secure encryption of data is a priority over high quality of speech. It is also used in Text-to-speech conversion systems in which speech is obtained from text.

II. LITERATURE SURVEY

In this section, a general overview of existing methodologies for Linear Predictive Analysis is elucidated. Basically, GSM supports three standard coding formats which are referred to as Full rate, Half rate and Enhanced Full rate. Their corresponding European Telecommunications standards are the GSM 06.10, GSM 06.20 and GSM 06.60,respectively. These coders work on a 13 bit uniform PCM speech input signal, sampled at 8 kHz. Hence the sampling rate would be 16 kHz. In the paper titled" *Gsm Speech Coding And Speaker Recognition"[1]*, by, *L. Besacier* , whole TIMIT database was passed through these coders, obtaining three trans-coded databases for speaker recognition application. Full rate coders support a data rate of 13kbps, half rate support a data rate of 11.4kbps and Enhanced Full rate support a data rate of 12.22kbps. All three coding try to improve their performance in-order to make the data rate to 16kbps. In this regard the author performs three experiments on the TIMIT database using these coding techniques. 1) Using *long training /short test protocol. 2)* The purpose of this experience is twofold, namely to find outwhich portions of the encoder are responsible for majordegradations and to improve the performance with respect to theresults obtained by extracting the features from resynthesizedspeech.
In the first experiment, it was found that usage of GSM coding degrades significantly the identification and verification performance. The second experiment provides a measurement of the different performance degradation sources within the FR coder. Moreover, it enlightens to directly exploit the coder output parameters instead of decode and reanalyse speech.In the paper titled “Speech compression using Linear Predictive Coding (LPC)” by Nikhil Sharma[2] subjective analysis has been used to analyse the speech coder. Subjective analysis involves determining quality based on encoded speech signal. LPC-10 technique was used for speech coding which has a low bit rate of 2.4 kbps. In LPC-10 vocal tract is represented as a time-varying filter and speech is windowed about every 30ms. For each frame, the gain and only 10 of the coefficients of a linear prediction filter are coded for analysis and decoded for synthesis. The steps involved in LPC implementation are-

1. *Sampling:* Sampling frequency is chosen such that all the necessary components are retained. Generally, the sampling frequency for voice transmission is selected as 10 kHz because it can encapsulate all the speech energy.
2. *Segmentation:* The speech is divided into blocks and is processed as frames which consists of a number of samples. The length of the frame is selected between 10 and 30 ms or 80 and 240 samples.
3. *Pre-emphasis:* Speech signal has a high frequency roll-off with low amplitude. However LP analysis of such a signal requires high precision for computation and hence the speech is filtered to flatten the signal boosting high frequencies.
4. *Voicing detector:* Voice signals have periodic waveform as opposed to unvoiced signals. To separate voice signals from the unvoiced signals, the parameters that are used are:

- Energy
- Zero crossing rate
- Pitch period

5. *Pitch period estimation:* Pitch of voiced speech is one of the most vital parameters in speech synthesis which is related to the speaker directly.
6. *Coefficient determination:* The linear predictive coefficients from a speech waveform are evaluated using the Levinson - Durbin algorithm.
7. *Gain calculation:* The gain is calculated by taking the square root of power in case of the unvoiced signal.
8. *Quantization:* The intermediate values of the predictive coefficients are called as line spectral frequency coefficients which regulate changes in the waveform which in turn has an effect on the spectrum.The quality of the sound is enhanced with the help of voice excited LPC vocoders.

In the paper titled "Speech compression using Linear Predictive Coding" [3], speech compression is done using the normal LPC method and the voice excitation method. Voice excitation is performed on the speech signal using the 10th order Levinson-Durbin recursion algorithm. Speech coding is basically converting speech signal to a more transmission friendly form. Speech signals have a frequency range of 0 – 8 KHz. Hence, the sampling frequency used must be around 16 KHz.

A speech coder is developed and the output is analysed using both subjective analysis (physically listening to the signal) and objective analysis (performed by calculating the segmental signal to noise ratio). The whole principle is to minimize the sum of square differences between the original speech signal and the reconstructed speech signal, also known as the mean square error. The transfer function of the digital filter used is

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

Where G is the gain of the filter and $a_k$is the predictor coefficient. Autocorrelation formula is used to compute the coefficients because the roots of the denominator always lie within the unit circle, thereby ensuring stability.

If the sound is voiced, then it is represented using an impulse train with non-zero taps after every pitch period. If the sound is unvoiced, then white noise is used to represent it. Thus, either of the two becomes the input to the LPC synthesis filter.

Small changes in the predictor coefficients gives rise to large changes in the pole positions of denominator. To overcome this, a high accuracy Levinson-Durbin recursion algorithm is used. The main idea of voice excitation is to filter the input speech signal by suing the estimated transfer function of LPC analyzer, also known as residual. Only the lower frequencies of the residual signals are required for a proper reconstruction of input.

Discrete Cosine Transform (DCT) is used to ensure high compression rate. The first few coefficients constitute most of the signal energy. Thus, the signal can be compressed by transferring only these coefficients. Receiver meanwhile performs inverse DCT and this signal is used to excite the voice.

It is found that the LPC reconstructed speech has a lower quality than the voice excited LPC signal, whose waveforms closely matches with that of the original input. The sound due to the normal LPC was much noisier than the voice excited LPC. One drawback of the voice excited LPC was that the bits per sample increased thereby increasing the bandwidth required for transmission.

III. IMPLEMENTATION

A. MATLAB Implementation

We are implementing the speech codec which is used in GSM cellular networks. When we speak into the microphone on a GSM phone, the speech is converted into a digital signal with a resolution of 8 bits. The audio signal of a single phone call is encoded as 8000samples/second (that is 8 kHz sampling frequency) of 8 bits each giving a 64kbps digital signal. This forms the input to all GSM speech codecs.

The codec analyses the voice and builds up a bit stream composed of a number of parameters that describe the aspects of the signal.

An input speech file which is a .wav file which is sampled at a frequency of 8 kHz and is a 8 bit mono audio file. It is segmented into 20ms speech frames of 160 samples using a Hamming window. The signal is pre-emphasised using a high pass filter and the LPC filter coefficients are calculated using Levinson Durbin algorithm. These coefficients are transmitted and are decoded at the decoder end. The output signal is reconstructed using the transmitted coefficients and is deemphasized.

Output

A spectrum of the output speech signal is plotted as shown below. The reconstructed output audio is also played using the audio system of the PC/laptop and the waveform is verified by comparing to the input waveform.

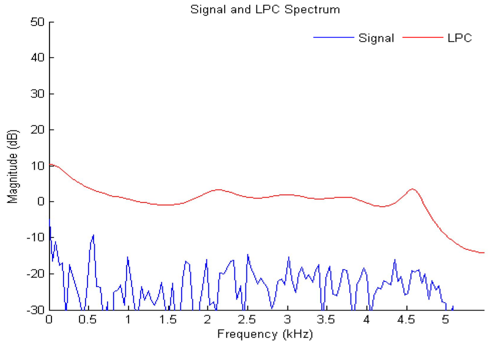

Fig.1: Output speech signal

B. SIMULINK Implementation

The block diagram of the encoding and decoding blocks of Linear Predictive coding technique are simulated on Simulink as shown in the figure below.

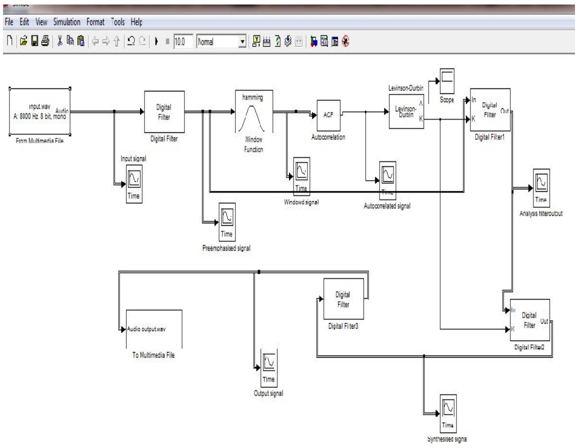

Fig. 2: LPC block diagram

The input to the encoder block is a .wav file which is a an uncompressed 8bit,Mono Audio speech file sampled at a rate of 8kHz frequency and has a bit rate of 64kbps.

The vector scopes are connected to each block to observe the waveform in various stages. The output audio file after decoding gets saved in the location given. The output of the LPC decoder has a bit rate of 2.4kbps.Thus it has a very high compression ratio. The GSM transmits this data at a rate of 13kbps. A screenshot of the Simulink model is taken during execution as shown below-

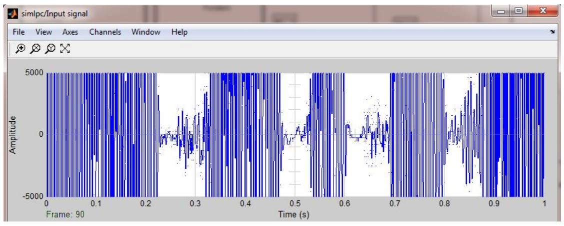

Fig. 3: Input signal

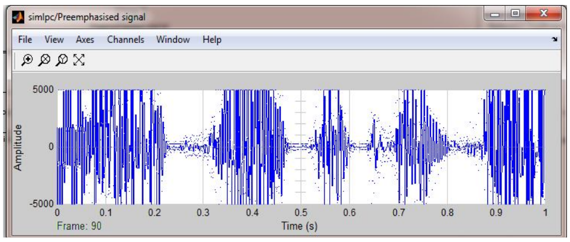

Fig. 4: Pre-emphasized signal

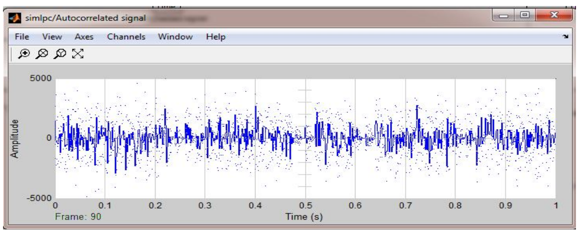

Fig. 5: Windowed signal

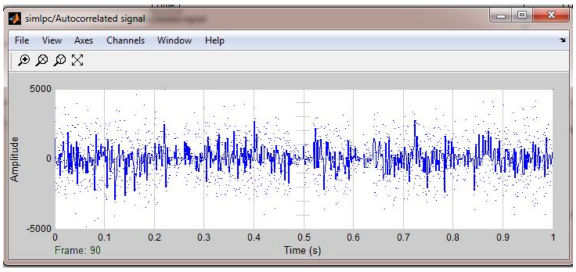

Fig 6: Autocorrelated signal

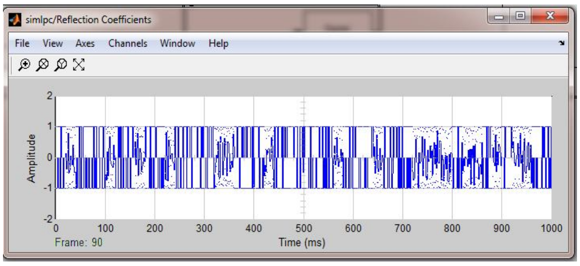

Fig 7: Relation coefficients

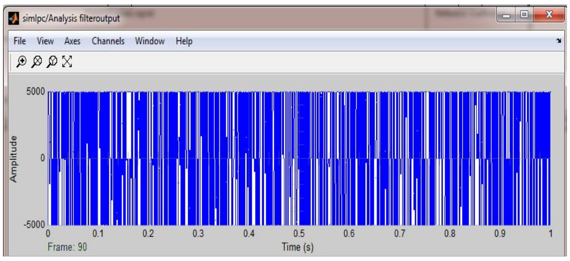

Fig 8: Analysis filter output

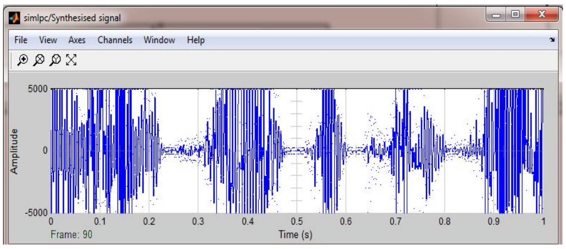

Fig 9: Synthesized signal

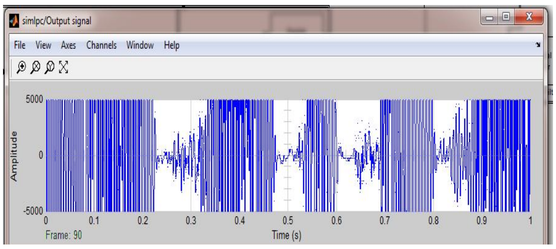

Fig 10: Output Signal

1. The output audio file gets saved in the set location. The reconstructed audio output is verified by comparing the input and output waveforms and also by playing the output audio file.
2. The bitrate of the compressed output file is found to be 13kbps. Thus it produces a 4:1 Compression Ratio
3. The difference in quality between the input signal and reconstructed signal are almost inaudible to the human ear.

IV. RESULT

An input speech file which is a wav file which is sampled at a frequency of 8 kHz and is a 8 bit mono audio file. It is segmented into 20ms speech frames of 160 samples using a Hamming window. The signal is pre-emphasized using a high pass filter and the LPC filter coefficients are calculated using Levinson Durbin algorithm. These coefficients are transmitted and are decoded at the decoder end. The output signal is reconstructed using the transmitted coefficients and is deemphasized.

A spectrum of the output speech signal is plotted as shown below. The reconstructed output audio is also played using the audio system of the PC/laptop and the waveform is verified by comparing to the input waveform.

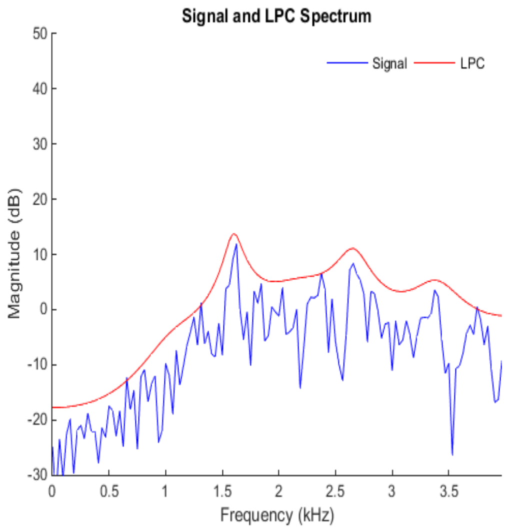

Fig. 11: Overall output graph

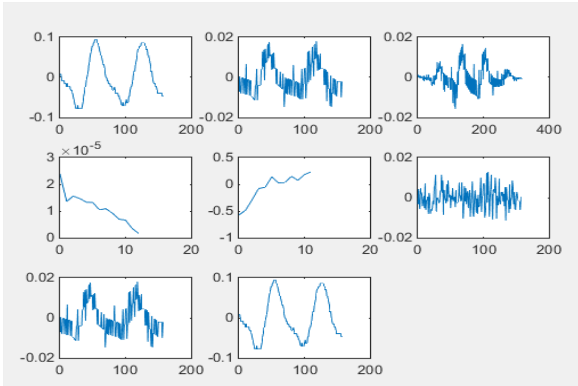

Fig. 12: Output graph at each stage

CONCLUSION

Matlab and Simulink simulations verify the Linear Predictive Coding Algorithm and show that the input speech can be reconstructed at the output by LPC technique

Linear Predictive Coding is an efficient speech compression algorithm as it produces high compression ratios with low bitrates. Though this algorithm is not very efficient, since the degradation in the quality of reconstructed output is not detected by the human ear, this technique is widely in GSM Cellular Networks.

REFERENCES

[1] L. Besacier, S. Grassi, A. Dufaux, M. Ansorge1, F. Pellandini, “GSM Speech Coding and Speaker Recognition”, Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP2000) 2, 1085-1088, 2000, pp 1-3

[2] Nikhil Sharma, “Speech Compression Using Linear Predictive Coding (LPC)”, International Journal of Advanced Research in Engineering and Applied Science, vol. 1, pp. 1-9, November 2012

[3] Amol R Madane, Zalak Shah, Raina Shah, Sanket Thakur, “Speech Compression Using Linear Predictive Coding”, International Workshop on Machine Intelligence Research, 2009, pp. 1-3.

★★★