Linear Predictive Coding for Speech Compression

YASIR SALEEM*, FAISAL HAYAT*, AND ARSLAN IQBAL*

RECEIVED ON 20.02.2013 ACCEPTED ON 11.09.2013

ABSTRACT

Telecommunication industry is growing and different services are rapidly introduced by different competitors to attract the users. Speech communication and its quality conservation is the most prevalent and common service provided by almost all companies. The objective of this project is the development of a LPC (Linear Predictive Coding) based voice coder. Attributes for speech like pitch, voiced and unvoiced decision and silence were extracted and speech was modeled using LDR (Levinson Durbin Recursion) and SDA (Steepest Descent Algorithm). LPC filter is analyzed and its model is implemented. LPC's different attributes complexity, delay and bitrate are deliberated and tradeoffs are highlighted. The results were analyzed and quality of speech was determined using spectrograph and by listening to the synthesized speech. At the end quality of original and synthesized speech is discussed and shown graphically and a soft comparison between both above mentioned technique is also added.

Key Words: **Speech Compression, Speech Encoding, Linear Filter, Predictive Filter.**

1. INTRODUCTION

Although data links in terms of transfer rate and capacity have been increased and still increasing in bandwidth and speed, speech communication is the most prevailing and general service in today's telecommunication industry. The increase in use of telephony communication in its various forms in commercial and private firms lead to a rapid development in this field [1]. The traditional communication which is analogue has served remarkably due to its simplicity in the last century. However the requirements of modern information technology have introduced and shown us the more robust and rapid substitute of the analogue systems. The encoding of speech has shown significant prominence to quality and bandwidth issues. Waveform encoders have been employed in industry in form of PCM (Pulse Code Modulation) and APCM (Adaptive PCM) and these techniques have evolved many new application areas like voice coders [1-3].

The attraction towards voice coders is obvious. Speech is reduced to a few bytes with all of the general advantages offered by waveform encoders. It has removed security issues, rapid regeneration is available, and it is best option for bandwidth limited communication systems and mobile networks. After the digitization of the analog signal, it is encoded with the help of a method which is primarily known as "LPC". A particular value is predicted in linear fashion with respect to the past values of the signal [4]. The human speech is usually produced in vocal tract which can be roughly estimated by the variable diameter tube and it

* Assistant Professor, Department of Computer Science & Engineering, University of Engineering & Technology, Lahore.

requires mathematical approximation of the vocal tract for the LPC model. The speech sample at a particular time represents the linear sum of past samples. Core idea behind the selection of LPC filter is the prediction of next samples using linearly combined previous samples [3-5].

Conventionally speech is sampled at 8k samples/sec in which one byte is used to represent each sample at a rate of 64k bits/sec. LPC can be used for reduction in this data rate up to 2.4 bits/sec. This reduction in the rate of speech has a visible quality loss as it costs a unique sound imitation but still the speech is audible. It has been notified that this form of compression is not lossless because it contains information loss which can easily be observed in the LPC [4]. A good strength of speech compression methods takes advantage of the fact that speech production is a result of deliberate anatomical actions with narrow range frequency. The frequency with which human speech is produced has a range from 300-3.4KHz. The speech coding algorithms exploits many characteristics, like silence period covers more than half (50%) of conversations. Authors mostly have recommended and highlighted that easiest tool to reduce the needed amount of information alongside the saving of bandwidth is by stopping the transmission of silence in speech signal. An important aspect of speech as suggested by many authors [5-11] is the strong correlation among the contiguous samples. In most forms of speech compression, the process of speech production is modeled as a linear digital filter. The reason behind the encoding of digital filter parameters (slow varying) is to achieve compression resulting from the speech signal.

LPC vocoders consist of four main characteristics: quality, bit rate, complexity and delay. Any voice coder should have ability to make tradeoffs between different attributes, regardless of the applied algorithm. The common algorithm for LPC is based on an encoding (analysis) part and a decoding (synthesis) part. The frames and blocks of speech signal are taken as input to the encoder to build the input signal. The current block of speech can be regenerated by using the coefficients of this filter. After transmitting the decoding section reproduces the filter using the coefficients received. Decoder uses the speech signal as an additional source of information about the absolute speech signal to predict the input being transmitted [5-6].

2. PROCEDURE

This specific type of input filter model based on LPC technique is termed as LPC model. The two parts of it are encoding and decoding. The tiny segments which results by splitting the speech signal and extracting the speech attributes from the segments are studied in encoding like:

- The segment contains silence or voice.
- Pitch of the segment.
- Gain of the Segment.
- Factors required for building a filter (Linear predictive filter coefficients).

These attributes are extracted by the sender and sent to the receiver, which uses these to build a filter that reproduces the speech signal as shown in Figs. 1-2.

3. IMPLEMENTATION OF PITCH ESTIMATION

Numerous algorithms have been proposed for estimating the pitch period from short time Autocorrelation Function. The algorithm-I used for estimating the pitch is summarized as:

(1) The speech signal is filtered with a 900Hz low pass filter sampled at a rate of 8 KHz/s.

(2) Segments of 30 msec are selected at 10 msec of interval. Thus segments overlap with 20 msec.

(3) The variance of segment is compared with the variance of noise for categorizing the segment as voiced or silence.

(4) The samples corresponding to one segment are clipped to 3 values: 1, 0,-1. The clipping threshold is adaptive, it is 30 % of the max value in the voice segment.

(5) The usage of this clip level facilitates the processing of the speech signal with a center clipper which consists of degree 3.

(6) The largest peak of the autocorrelation function is located and the peak value is compared to a fixed threshold 30% of Rn(0).If the peak falls below threshold, the segment is classed as unvoiced and it is above, the pitch period is defined as the location has the largest peak.

There is one problem with the first implementation that when the signal has a low frequency fluctuation this leads to high value of autocorrelation function that produce an annoying synthesized voice. The remedy of this problem is that if the minimum value autocorrelation function in the interval of lags (1,23) is greater than the maximum value in the interval of interest (24,160) the frame is considered as unvoiced. This small modification leads to much better speech quality. Next section highlights some factors which play an important role in determination of speech quality.

4. FACTORS EFFECTING THE SPEECH QUALITY

- Overlapping of Speech Segments affects Speech Quality, that it reduces the synthetic metal sound from the Speech.
- Median Filter of Order 3 discussed and used improves the speech Quality but it also increases the delay of the System.
- If the minimum value autocorrelation function in the interval of lags (1,23) is greater than the maximum value in the interval of interest (24,160) the frame is considered as unvoiced. This small modification leads to much better speech quality.

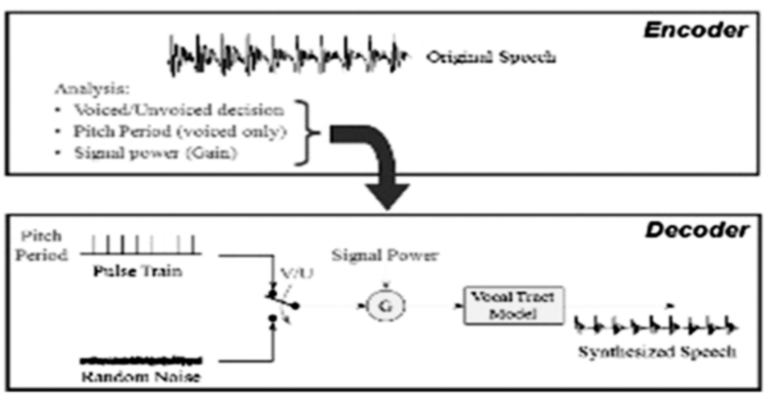

FIG. 1. LPC MODEL

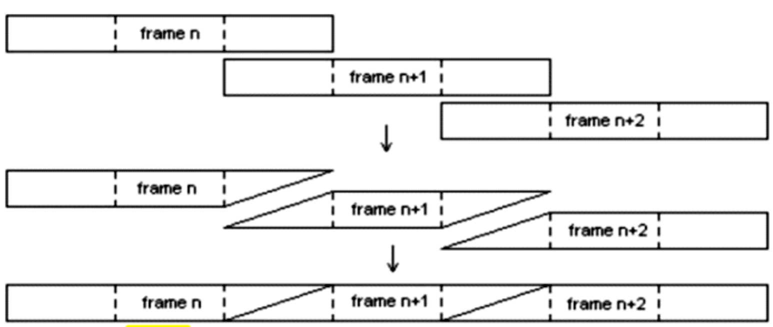

FIG. 2. FRAME DISTRIBUTION

- Transmitting the Silence Period of Speech reduces the Work Load of decoder, and helps removing a buzz sound from the synthesized speech.

5. IMPROVEMENTS ON THE PITCH ESTIMATION

The improvements implemented were Pitch Smoothing using Median Filter. Using the median Filter we can achieve much better synthesized speech.

The Filter Coefficients are computed by LDR relation and SDA. These Coefficients are computed for each Segment but before that segment is treated as shown in Figs. 3-4.

For the LPC analysis the speech is passed through a pre-emphasis filter with Coefficients equal to 0.9 and multiplied by a Hamming window. The order of LPC is 10 and Coefficients are computed by LDR relation and SDA [10]. The Stopping Criteria for the SDA was based on Norm and tolerance, that if our filter coefficients are within a specified tolerance than stop the recursive relation.

These attributes are transmitted to the receiver where it synthesized the speech from these for voiced signal a pulse train is generated and for unvoiced part a random signal is generated. These signals are multiplied with a gain and then pars through the filter realize with the filter coefficients received from the sender. Frames are arranged together back in a special way and the reason is clear as

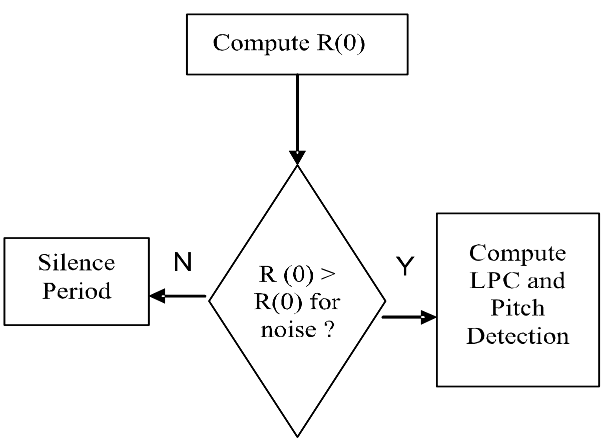

FIG. 3. LPC DECISION MAKING

each frame has its own pitch gain and filter so if we simply put them back then there will be increased chopping in sound. The overlapping is required for smooth transition from one frame to another. Fig. 4 makes it clear how it is accomplished. The amplitude of the tip and tail of each frame's data is scaled and then simply added.

6. CALCULATIONS

This section contains the mathematical expressions used to implement the Levinson Durbin Recursion and Steepest Decent algorithms with their required suitable parameters.

Time Average Formula:

$$\hat{r}(k) = \frac{1}{N}\sum_{n=1+k}^{N} u(k)u^*(n-k) \qquad (1)$$

Levinson Durbin Recursion:

$$a_{m,k} = a_{m-1,k} + K_m a^*_{m-1,m-k} P_m = P_{m-1}\left(1-|K_m|^2\right) \qquad (2)$$

Steepest Descent Algorithm:

$$w(n+1) = w(n)+\mu\,[p - Rw(n)] \quad n=0,1,2,3\ldots \qquad (3)$$

Stopping Criteria for Steepest Descent Algorithm:

$$\frac{\left\|w(n+1)-w(n)\right\|_\infty}{\left\|w(n+1)\right\|_\infty} = Tolerance \qquad (4)$$

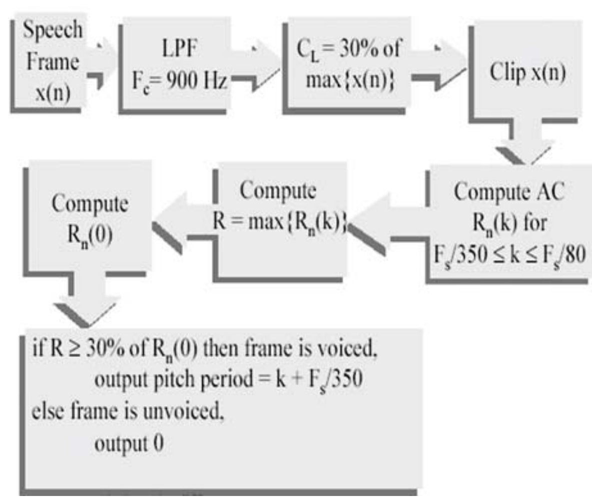

FIG. 4. LPC SYSTEM FLOW DIAGRAM

7. RESULTS

The results obtained from implementing the above mentioned scheme for LPC are presented below:

Fig. 5 shows the output of all operations performed on speech sample. Fig. 5(a) is the original speech sample and it is followed by synthesized sample using the SDA shown in Fig. 5(b). Fig. 5(c-d) shows the voiced/unvoiced section and unsmoothed pitch respectively. At this point median filter of order 3 is applied and smooth pitch can be observed in Fig. 5(d).

Fig. 5(e) shows the synthesized output of speech sample using LDR and it can be observed that coefficients determining the shape of output signal like minimum, maximum points, silence duration are similar as compared to the synthesized output signal shown in Fig. 5(b).

The Speech Compression we achieved can be calculated as: The original speech data rate was 64 Kbits/s and we have calculated some attributes of speech and our compressed speech data rate is 4.6Kbits/s as it is divided in following segments shown in Fig. 6.

We have 96 bits/frame. Fig. 5(f-g) gives the spectrogram and frequency analysis of synthesized speech respectively and showing the phase and magnitude plot once the frequency is normalized. In our case we had 50 speech segments sampled at 8 KHz; from calculation the data rate we achieved is 4.6 Kbits/s. So compression ratio is 1:14.

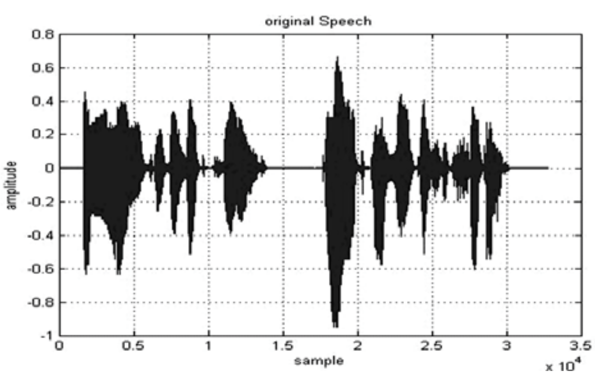

Fig. 5(a)

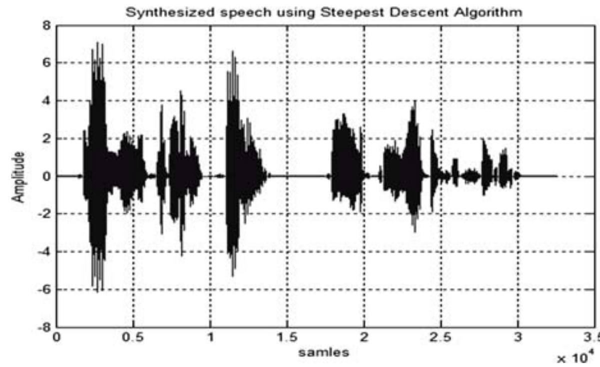

Fig. 5(b)

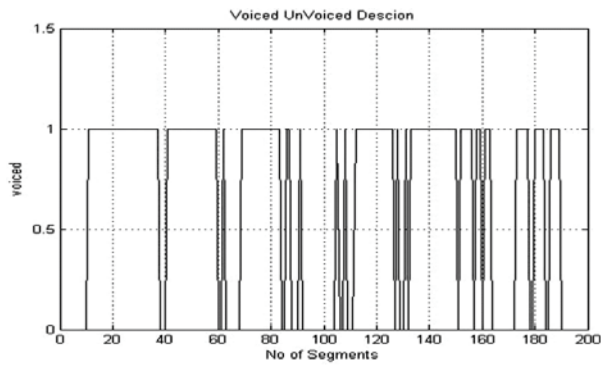

Fig. 5(c)

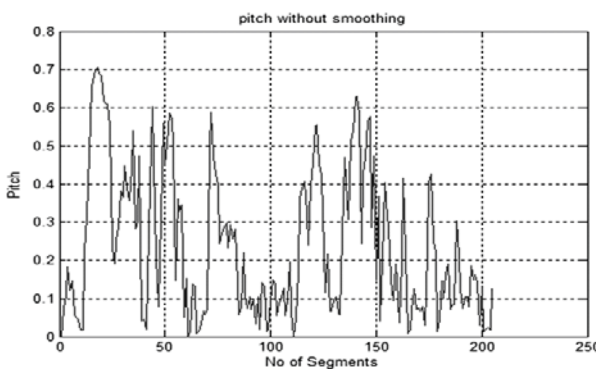

Fig. 5(d)

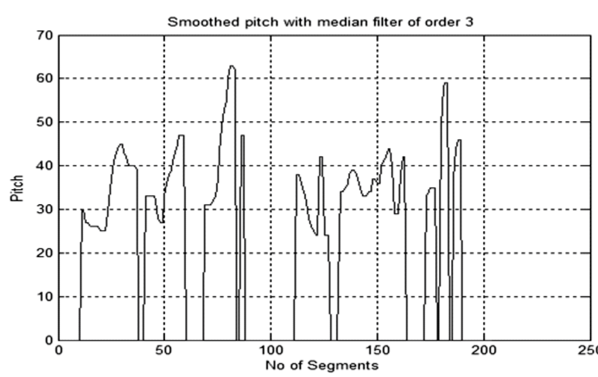

Fig. 5(e)

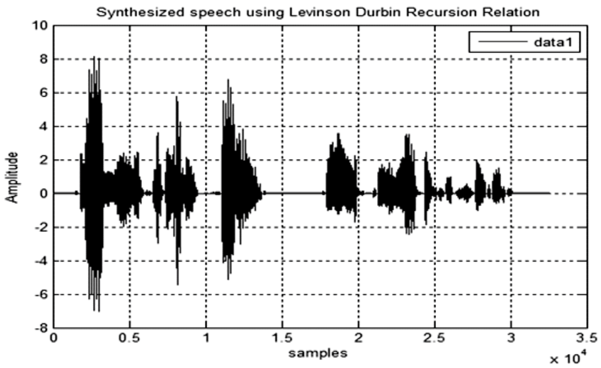

Fig. 5(f)

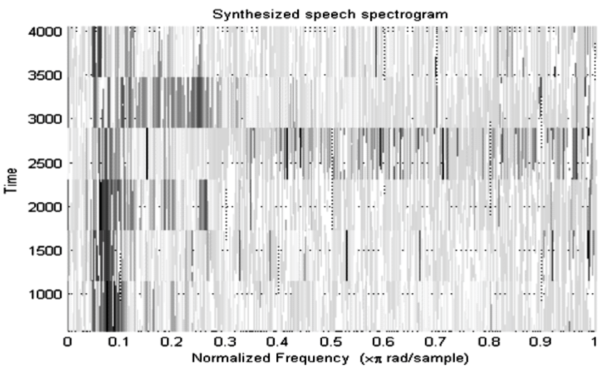

Fig. 5(g)

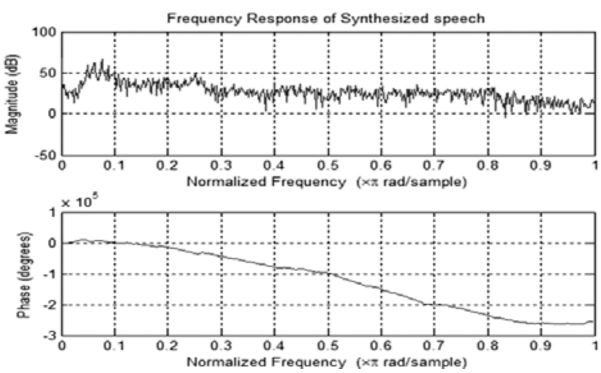

Fig. 5(h)

FIG. 5. LPC RESULTS

Voiced	Pitch	Gain	Silence	LPC Coefficients

FIG. 6. QUANTIZED AT 8BIT /SAMPLE

8. CONCLUSION

This paper elaborates the tradeoffs between quality of speech and its compression keeping the required transfer rate as less as 2.4 Kb/s. LPA was implemented with LDR and SDA. Voice quality of synthesized speech was simulated and result has shown satisfactory output as speech could easily be understood. In addition Spectrogram, demonstrating the signal spectral density variation versus time, of the original speech and synthesized speech are compared. The coefficients of LPC were calculated from the SDA and LDA (Levinson Durbin Algorithm) and these have same value. SDA is computationally heavy as compared to Levinson Durbin recursion Relation.

ACKNOWLEDGEMENT

The authors are thankful to the Department of Computer Science & Engineering, University of Engineering & Technology, Lahore, Pakistan, for providing the required guidance, environment and necessary Lab. facilities. This piece of literature could not be accomplished without said amenities.

REFERENCES

[1] Molisch, A., "Wireless Communications", Wiley-IEEE Communication, John Wiley and Sons Press, pp. 319-342, West Sussex, England, 2011.

[2] Hayes, H.M., "Statistical Digital Signal Processing and Modeling", John Wiley and Sons Press, pp. 225-235, West Sussex, England, 1996.

[3] Rabiner, L., and Schafer, R., "Digital Processing of Speech Signal", Prentice-Hall, 1978.

[4] Upadhya, S.S., and Wankhede, N., "Pitch Estimation using Autocorrelation Method and AMDF", International Conference on Advances in Computing and Management, pp. 249-253, 2012.

[5] "Linear Predictive Coding", Online Scholarly Article, Oregon State University, 2004. eecs.oregonstate.edu/education/docs/ece352/CompleteManual.pdf

[6] Klautau, Aldebaro, "LPC Vocoder", UFPA, Brazil, pp. 1-12, 2000. http://www.laps.ufpa.br/aldebaro/papers/LPCVocoder.pdf

[7] Zohar, E., Cidon, I., and Mokryn, O.O., "The Power of Prediction: Cloud Bandwidth and Cost Reduction", Proceedings of the ACM SIGCOMM Conference, pp. 86-97, USA, 2011.

[8] Bae, S.H., "Data Compression and Modeling", Georgia Institute of Technology, USA, 2004. http://www.users.ece.gatech.edu/~juang/8873/Bae-LPC10.ppt

[9] Frigon, J.F., and Telplitsky, V., "Implementation of Linear Predictive Coding of Speech", A Class Project, UCLA, California, Spring, 2000.

[10] Bradbury, J., "Linear Predictive Coding", Online Scholarly Article, Florida Institute of Technology, December, 2000. http://my.fit.edu/~vkepuska/ece5525/lpc_paper.pdf

[11] Garg, M., "Linear Predictive Algorithms", Online Scholarly Article, IIT, Bombay, pp. 1-15, India, 2003. http://www.ee.iitb.ac.in/uma/mohit