Speech Encoding using LPC Method

Thimmaraja Yadava G[1] Prem Narayankar[2] Rakesh Patil M S[3]
[1]Digital Electronics, GM Institute of Technology, Davangere, India
[2]Digital Electronics and Communication, SSE, Surathkal, India
[3]Network and Internet Engineering, JNNCE, Shimoga, India
Email: thimmrajyadav@gmail.com[1], msrakeshpatil@gmail.com[3]

Abstract—**Today a lot of research work is going on in industry and educational institutes on how to efficiently encode and transmit the speech signal from the transmitter and retrieve it back with high accuracy at the receiver. Shortly speech encoding is a way to transmit speech using lower data rate. The past has witnessed tremendous progress in the application of low rate speech coders to civilian and military communications as well as computer related applications. Central to this, progress has been the development of new speech coders capable of producing high quality speech at low data rates.**

Most of these coders incorporate mechanisms to represent the spectral properties of speech, provide for speech waveform matching and optimize the coders' performance for the human ear. A number of these coders have already been adopted in international cellular telephony standards. In cellular telephony systems, service providers are continuously met with challenge of accommodating more users within a limited allocated bandwidth. For this reason, manufacturers and service providers are continuously in search of low bit rate speech coders that deliver toll-quality speech. LPC is such a low bit rate speech coder which enables the service provider to share a limited bandwidth among more users.

Keywords—**Linear Predictive coding (LPC).**

I. Introduction

The studies of how human speech sounds are produced and how they are used in form of a language is an established scientific discipline, with well-developed theoretical background. However, even today, lot of research work is going on in industry and educational institutes on how to efficiently encode and transmit the speech signal from the transmitter and retrieve it back with high accuracy at the receiver. Advancements in field of Digital Signal Processing have contributed significantly to address this issue.

Today, many communication systems and networks are digitized - movies, televisions, music, images and voice. Digital signals are used because of ease in storage and long-distance transmission without accumulating distortion; and the digital representations are highly resistant to minor degradation. However, there is a downside too. The bandwidth occupied by the un-compressed digital data is very high. To address this pitfall, various data compression techniques are used in the industry these days based on requirement and cost factor. There are enormous activities recently in establishing speech coding standards both nationally and internationally.

Although with the emergence of optical fiber bandwidth in wired communications has become inexpensive, there is a growing need for bandwidth conservation and enhanced privacy in wireless cellular and satellite communications. Most of these applications require that the speech signal is in digital format so that it can be processed, stored, or transmitted under software control. Although digital speech brings flexibility and opportunities for encryption, it is also associated with a high data rate and hence high requirements of transmission bandwidth and storage. Speech Coding is the field concerned with obtaining compact digital representations of voice signals for the purpose of efficient transmission or storage.

Speech is generally band limited to 4kHz and sampled at 8kHz.The simplest nonparametric coding technique is Pulse-Code Modulation which is simply a quantizer of sampled amplitudes.

II. Speech Properties and Historical Perspective

Before beginning the presentation of the speech coding methods, it would be useful if we briefly discussed some of the important speech properties. First, speech signals are non stationary and at best they can be considered as quasi-stationary over short segments, typically 5-20ms.The statistical and spectral properties of speech are thus defined over short segments.

Speech can generally be classified as voiced (eg., /a/, /i/, etc), unvoiced (eg., /sh/), or mixed. The energy of voiced segments is generally higher than the energy of unvoiced segments. The formants are the resonant modes of the vocal tract.

The properties of speech are related to the physical speech production system as follows. Voiced speech is produced by exciting the vocal tract with quasi-periodic glottal air pulses generated by the vibrating vocal chords. The frequency of the periodic pulses is referred to as the fundamental frequency or pitch. Unvoiced speech is produced by forcing air through a constriction in the vocal tract, and plosive sounds (e.g., /p/) are produced by abruptly releasing air pressure which was built up behind a closure in the tract. Speech coding research started over fifty years ago with the pioneering work of Homer Dudley of the Bell Telephone Laboratories. The motivation for speech coding research at that time was to develop systems for transmission of speech over low-bandwidth telegraph cables. The emerge of VLSI technologies along with advances in the theory of digital signal processing during the 1960's and 1970's provided even more incentives for getting new and improved solutions to the speech coding problem.

Analysis of speech using the Short-Time Fourier Transform (STFT) was proposed by Flanagan and Golden under "Phase Vocoder". In addition, Schafer and Rabiner designed and simulated an analysis-synthesis system based on the STFT and Port off provided a theoretical basis for the time-frequency analysis of speech using the STFT. In the mid- to late 1970's there was also a continued activity in linear prediction, transform coding and sub-band coding.

III. TYPES OF SPEECH AND ITS PROPERTIES

Basically there are two types of Speech namely,

a. Voiced
b. Un- Voiced

Voiced sounds are usually vowels that usually have high average energy levels and very distinct resonant or formant frequencies. Air from within the lungs generates voiced sounds by forcing the vocal cords. Due to vibration of vocal folds, seemingly periodic patterns in form of series of air pulses are produced that are called glottal pulses. These glottal pulses excites a vocal tract cavity and produces a vowel (or voiced) sound. The rate at which the vocal cords vibrate determines the pitch of the sound produced. These air pulses finally pass along the rest of the vocal tract where some frequencies resonate. Generally, women and children have higher pitched voices than men because of a faster rate of vibrations during the production of voiced sounds. It is therefore important to include the pitch period in the analysis and synthesis of speech of the final output in order to accurately represent the original input signal.

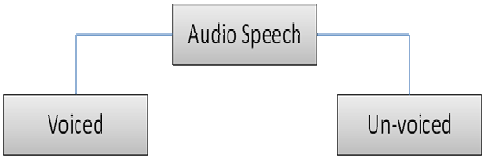

Fig.1. Types of Audio Speech.

On the other hand, unvoiced sounds are usually consonants having comparatively less energy at higher frequencies than voiced sounds. Air from vocal folds, in a turbulent flow, is the source of unvoiced sound generation. During this process, the vocal cords do not vibrate, but instead they stay open until the sound is produced. Since there is no vibration of the vocal cords, no glottal pulses and since unvoiced sounds are not periodic, pitch is an unimportant attribute of unvoiced speech.Air from within the lungs generates voiced sounds by forcing the vocal cords.

The following are some of the important properties of speech:

a) Fricatives (s, sh, f, th) - Produced when the vocal tract is constricted at some location and air is forced through that constriction.
b) Plosives (p, k, t) - Produced when the end of the vocal tract is constricted or closed momentarily while air pressure built up, then pressure is suddenly released.
c) In English, there are about 40 phonemes (sound elements - 16 vowels, 24 consonants).
d) In normal speech, 10 to 15 phonemes are spoken in one second.

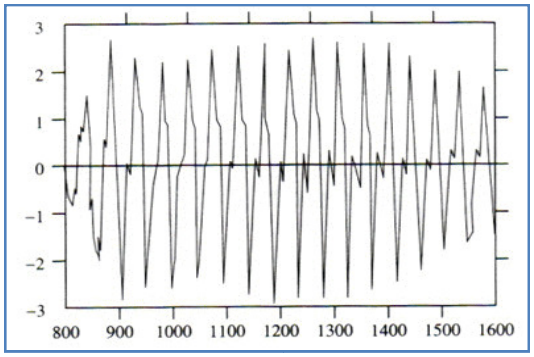

Fig. 2. Voiced Sound for letter "E"

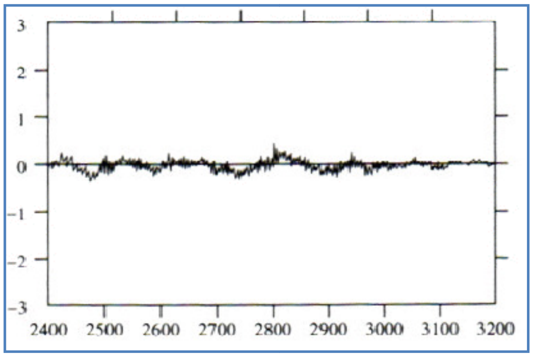

Fig. 3. Unvoiced Sound for letter "S"

IV. LINEAR PREDICTIVE CODING AND ITS ALGORITHM

Linear Predictive Coding is analytical/synthesis method, introduced in the sixties, for predicting a present sample of speech based on several previous samples. LPC is an efficient way of getting synthesized speech signal. The efficiency of this method is due to the speed of the analysis algorithm and low bandwidth required for the encoded signals.

There are two ways of measuring/estimating the spectrum/spectral envelope, of a sound. First way is through Fast Fourier Transform (FFT), which measures the spectrum of a sound by sampling amplitude values at equally spaced frequency points in the given range. This method provides an accurate estimation of the spectrum. The other way is to use Linear Predictive Coding. This method measures the overall spectral envelope to create a linear image of the sounds' spectrum. Both have their strengths, and weaknesses, but LPC is particularly effective with manipulating speech.

LPC generally deals with modeling and FFT makes the spectrum estimation. LPC is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate. It provides extremely accurate estimates of speech parameters, and is relatively efficient for computation. The basic assumption behind LPC is, one can predict nth sample in a sequence of speech samples and represent it by the weighted sum of the k previous samples of the target signal.

$$\hat{S}[n]=\sum_{k=1}^{p}(a_k * S[n-k]) \qquad (1)$$

wherep indicates the order of the LPC. As p approaches infinity, one should be able to predict the exact value of nth sample.

However, with the limitation in computation, p is usually of the order of 10-20, such that it still produces the accurate results. The equation below represents the error signal, referred to as LPC residual.

$$e[n] = S[n] - \hat{S}[n]$$

$$e[n] = S[n] - \sum_{k=1}^{p}(a_k * S[n-k]) \qquad (2)$$

Taking Z- transform of equation (2):

$$E[z] = S[z] - \sum_{k=1}^{p}\left(a_k * S[z] * z^{-k}\right) \qquad (3)$$

$$E[z] = S[z] * [1 - \sum_{k=1}^{p}(a_k * z^{-k})] \qquad (4)$$

$$E[z] = S[z] * A[z] \qquad (5)$$

Thus, one can represent the error signal as the product of original speech sample, S[z] and the transfer function, A[z]. A[z] represents an all-zero digital filter, where the a_k coefficients correspond to zeros in the filter's z-plane. Similarly, one can retrieve the original speech signal S[z] as the product of the error signal E[z] and the transfer function 1/A[z]:

$$S[z] = E[z] * 1/A[z] \qquad (6)$$

The transfer function 1/A[z] represent an all-pole digital filter, where the a_k coefficients correspond to the poles in the filter's z-plane. For stability reasons, the roots of transfer function, A[z] must lie within the unit circle.

The spectrum of the error signal E[z] is different for voiced and unvoiced sound. Vibrations of the vocal chords produce voiced sounds, while the unvoiced sounds are usually consonants and generally have less energy and higher frequencies like white noise. The Spectrum of voiced sounds are periodic with some fundamental frequency called pitch e.g. all vowels. The unvoiced signals, however don not have any fundamental frequency or a harmonic structure.

Usually, speech is sampled at 8 KHz with sample size of 8 bits. Therefore, the data processing rate would be 64000 bits/second. LPC algorithm uses compression algorithm to reduce the data rate to 2400 bits/second. LPC does so by breaking the speech into segments and then sending them as voiced/unvoiced information, the pitch period, and the coefficients for the filter that represents the vocal tract for each segment.

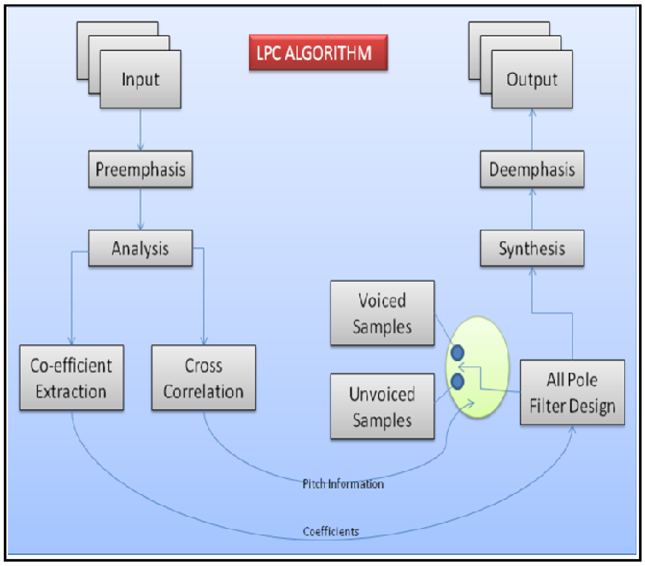

Fig. 4. Generalized LPC Algorithm

At 2400 bits/sec of bit-rate, the speech has a distinctive synthetic sound and there is a noticeable loss of quality of compressed sound. However, the speech is still audible and is understandable. Since there is information loss in linear predictive coding, it is a lossy form of compression. LPC analyses the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. Formants frequencies are the frequencies at which resonant peaks occur. The number that describes the formants and the residue can be stored or transmitted somewhere else.

In LPC, the input signal is sampled, broken into segments/blocks/frames, analysed and then transmitted to the receiver. Pre-emphasis refers to a process designed to increase, within a band of frequencies, the magnitude of some (usually higher) frequencies with respect to the magnitude of other (usually lower) frequencies in order to improve the overall signal-to-noise ratio by minimizing the adverse effects of such phenomena as attenuation distortion or saturation of recording media in subsequent parts of the system. In speech analysis, it is computationally intensive to determine the pitch period for a given segment of speech. There are several algorithms to compute this One such algorithm takes advantage of the fact that the autocorrelation of a period function, r(k), will have a maximum value when k is equivalent to the pitch period.

From equation (2),the sum of the squared error to be minimized is expressed as:

$$E = \{S[n] -\hat{S}[n]\}^{\wedge}2 \qquad (8)$$

Pitch Estimation: Pitch can be estimated by auto-correlation method, average magnitude difference method and cestrum. The autocorrelation of a stationary sequence $x(n)$ is defined as:

$$R_x(\tau) = [x(n) * x(n+\tau)\ ] \qquad (9)$$

Where τ is termed as lag. An autocorrelation is the average correlation between two samples from one signal and its τ samples delayed signal. In MATLAB, inbuilt function "xcorr" can be used either for cross correlation between two signals or auto correlation between a signal with itself.

The analysis/encoding part of LPC examines the speech signal by breaking it down into segments or blocks. Each segment is then examined further to find:
a) Voiced/unvoiced segment?
b) Is pitch information important for this particular segment?
c) What other information is required to build a filter that models the vocal tract for the current segment?

"A sender who answers these questions and usually transmits these answers to a receiver usually need to conduct LPC analysis. If both the linear prediction coefficients and the residual error sequence are available, the speech signal can be reconstructed using the synthesis filter. The receiver performs LPC synthesis by using the answers received to build a filter when provided the correct input source will be accurately reproduce the original speech signal. Essentially, LPC synthesis tries to imitate human speech production."

The below figure explains how the Predictor Filter is used in analyser of LPC systems. When the predictor filter has been adjusted to predict the input, at best it can do so from the immediate preceding samples.

The difference between the input speech and the predictor output (known as residual) will have roughly flat spectrum. The spectral peaks caused by the resonance of speech production will have to be removed.

LPC synthesizes the speech signal by reversing the process: use the residue to create a source signal, use the formants to create an all-pole filter (which represents the tube), and run the source through the filter, resulting in speech. Since the speech signal varies with time, this process is done on short chunks of the speech signal called frames. Usually 30-50 frames per second give intelligible speech with good compression.

The below figure explains how the Predictor Filter is used in Synthesizer of LPC systems

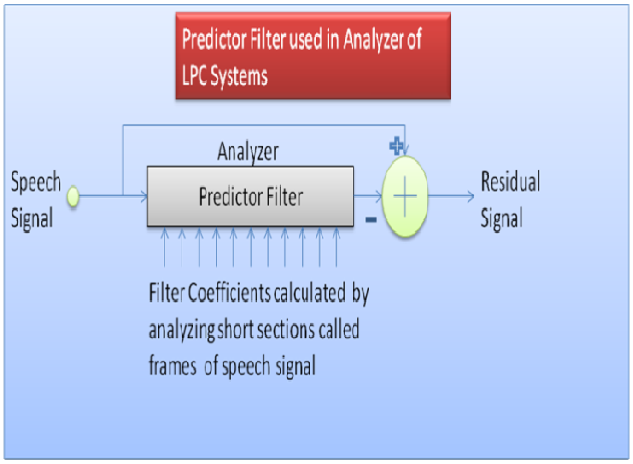
Predictor Filter used in Analyzer of LPC Systems
Speech Signal
Analyzer
Predictor Filter
+
-
Residual Signal
Filter Coefficients calculated by analyzing short sections called frames of speech signal

Fig.5. Predictor Filter used in Analyzer of LPC System

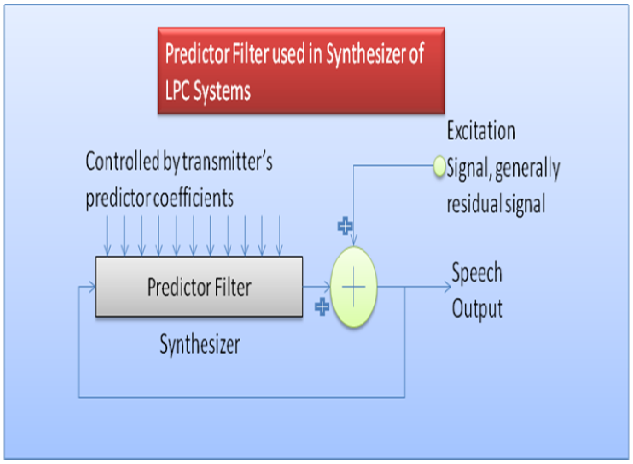
Predictor Filter used in Synthesizer of LPC Systems
Controlled by transmitter's predictor coefficients
Excitation Signal, generally residual signal
+
Predictor Filter
+
Speech Output
Synthesizer

Fig.6. Predictor Filter used in Synthesizer of LPC System

V. ANALYSIS AND SIMULATION RESULTS

Performance parameters:

- Compression ratio : Ratio of Compressed over uncompressed signal
- Speech signal size.
- Sampling frequency, fs
- Frame/Window length

Ex (a): The details of the simulation are given below:

```
Processing the wave file "s2ofwb"
The wave file is 3.01seconds long
Press a key to play your originally recorded wav file!
Press a key to play the LPC compressed sound!
```

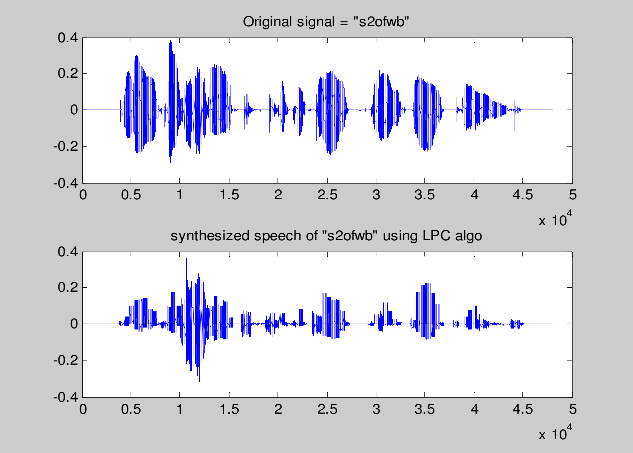

Fig. 7. LPC Vocoder Output: Original v/s Synthesized Speech Signal

Ex (b): The details of the simulation are given below:

Processing the wave file "s1ofwb"
The wave file is 2.92 seconds long
Press a key to play your originally recorded wav file!
Press a key to play the LPC compressed sound!

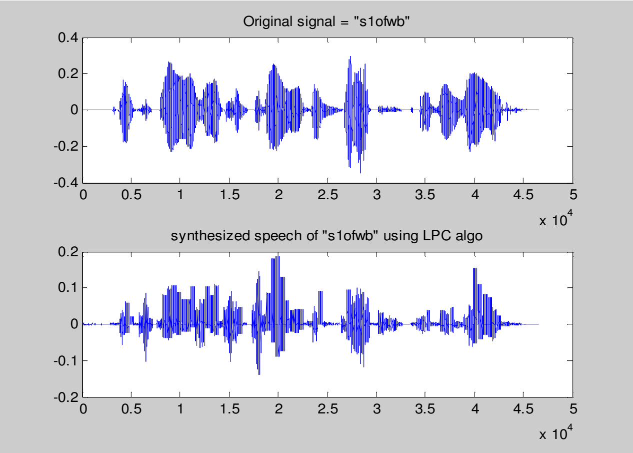

Fig. 8. LPC Vocoder Output: Original v/s Synthesized Speech Signal

The above figures 7 and 8 shows that when a given sample signal is encoded using the LPC algorithm, original signal can effectively be reconstructed. Since LPC is a lossy data compression technique, the quality of speech is degraded to some extent, but with a major advantage of considerably low bit rate.

VI. Conclusions

After thorough study of the LPC algorithms, now is possible to understand how Linear Predictive Coding is used as an analysis/synthesis technique.

LPC is a lossy speech compression technique that attempts to model the human production of sound instead of transmitting an estimate of the audio speech signal.

Linear predictive coding achieves much lower bit rate, which makes it ideal for use in secure control systems. Secure control systems are more concerned about the content and meaning of speech, rather than the quality of speech.

This project is 1D-signal based and LPC algorithm is used. By using this we have successfully studied and implemented LPC algorithm in MATLAB.

The work is implemented using $MATLAB^{®}$

The scope for future modifications:

i. In future, microphone and pre-amplifier input circuitry can also be used to implement this algorithm.
ii. This project could be extended to implement entire LPC algorithm on the micro-controller for speech feature recognition.
iii. It can also be used in the real time systems where compression of speech signals plays an important role.

Acknowledgment

It is a pleasure to recognize the many individual who have helped me in completing this technical paper.

Assistant Professor Shivkumar B R, GMIT, Davangere, helped me for all the technical guidance, encouragement and analysis of the data throughout this process.

References

[1] L R Rabiner and R W Schafer, "**Digital Processing of Speech Signals**", Prentice- Hall, Englewood Cliffs, NJ, 1978.

[2] Rabiner, Juang and Yegnanarayana, **"Fundamentals of Speech Recognition"**, Pearson Education, First edition, pp 113-119, 1993.

[3] Konduz.A.M **Digital Speech Coding for Low Bit Rate Communication System**. John Wiley & Sons Ltd.

[4] Speech signal processing by Gold and Morgan.