National Conference on "Advanced Technologies in Computing and Networking"-ATCON-2015
Special Issue of International Journal of Electronics, Communication & Soft Computing Science and Engineering, ISSN: 2277-9477

A Study of LPC: Speech Coding Compression Method

Nilesh S. Bhoyar **Atul K. Shingarwade** **Harshalata Petkar**

Abstract — **One of the most powerful speech analysis techniques is the method of linear predictive analysis. This method has become the predominant technique for representing speech for low bit rate transmission or storage. The importance of this method lies both in its ability to provide extremely accurate estimates of the speech parameters and in its relative speed of computation. The basic idea behind linear predictive analysis is that the speech sample can be approximated as a linear combination of past samples. In this paper We describe the necessary background needed to understand data decompression method, speech coding and linear predictive coding model. We also explained how linear predictive coding work. We conclude by discussing possible disadvantages and applications of the LPC model.**

Key Words — **LPC, Speech Coding, Compression Model,Data Compression.**

I. INTRODUCTION

Data compression method reduces the storage cost. Data compression is a common requirement for most of the computerized application [1]. Compression method classified in two category lossy or lossyness. Lossy method reduces data size more than lossyness because it removes redundancy part. This methods useful when a perfect consistency with the original data is not necessary after decompression. The algorithm which removes some part of data is called lossy data compression and the algorithm that achieve the same what we compress after decompression is called lossyness data compression [2].

Speech coding is an application of data compression of digital audio signals containing speech. Speech coding uses speech - specific parameter estimation using audio signal processing techniques to model the speech signal, combined with generic data compression algorithms to represent the resulting modeled parameters in a compact bit stream[3]. Speech coder can be classified in two category waveform coders (Time domain, frequency domain) and Vocoders (Linear predictive coders, formant coders)[4].

Linear Predictive Coding (LPC) is the method of data compression. Specifically, LPC models process is a linear sum of earlier samples using a digital filter input as an excitement signal. An alternate explanation is that, linear prediction filters attempt to predict future values of the input signal based on past signals. LPC "...models speech as an autoregressive process, and sends the parameters of the process as opposed to sending the speech itself"[5] . It was first proposed as a method for encoding human speech by the United States Department of Defence in federal standard 1015, published in 1984. Another name for federal standard 1015 is LPC-10 which is the method of Linear predictive coding that will be described in this paper.

Speech coding or compression is usually conducted with the use of voice coders or vocoders. There are two types of voice coders: waveform-following coders and model-base coders[4]. Waveform following coders will exactly reproduce the original speech signal if no quantization errors occur. Model-based coders will never exactly reproduce the original speech signal, regardless of the presence of quantization errors, because they use a parametric model of speech production which involves encoding and transmitting the parameters, not the signal. LPC vocoders are considered model-based coders which means that LPC coding is lossy even if no quantization errors occur.

Lossy audio compression is used in a wide range of applications. In addition to the direct applications (mp3 players or computers), digitally compressed audio streams are used in most video DVDs, digital television, streaming media on the internet, satellite and cable radio, and increasingly in terrestrial radio broadcasts. Lossy compression typically achieves far greater compression than lossless compression (data of 5 percent to 20 percent of the original stream, rather than 50 percent to 60 percent), by discarding less-critical data.

Linear predictive coding (LPC), the lossy compressors used with speech are source-based coders. These coders use a model of the sound's generator (such as the human vocal tract with LPC) to whiten the audio signal (i.e., flatten its spectrum) before quantization. LPC (Linear predictive coding) is the most useful method for encoding good quality speech at a low bit rate[6,7]. LPC may be thought of as a basic perceptual coding technique: reconstruction of an audio signal using a linear predictor shapes the coder's quantization noise into the spectrum of the target signal, partially masking it.

The general algorithm for linear predictive coding involves an analysis or encoding part and a synthesis or decoding part. In the encoding, LPC takes the speech signal in blocks or frames of speech and determines the input signal and the coefficients of the filter that will be capable of reproducing the current block of speech. This information is quantized and transmitted. In the decoding, LPC rebuilds the filter, based on the coefficients received. The filter can be thought of as a tube which, when given an input signal, attempts to output speech.

LPC Model

The particular source-filter model used in LPC is known as the Linear predictive coding model. It has two key components: analysis or encoding and synthesis or decoding. The analysis part of LPC involves examining the speech signal and breaking it down into segments or blocks. Each segment is then examined further to find the answers to several key questions:

• Is the segment voiced or unvoiced?
• What is the pitch of the segment?
• What parameters are needed to build a filter that models the vocal tract for the current segment?

LPC analysis is usually conducted by a sender, who answers these questions and usually transmits these answers onto a receiver. The receiver performs LPC synthesis by using the answers received to build a filter that when provided the correct input source will be able to accurately reproduce the original speech signal. Essentially, LPC synthesis tries to imitate human speech production.

LPC Analysis/Encoding

Input speech

According to government standard 1014, also known as LPC-10, the input signal is sampled at a rate of 8000 samples per second. This input signal is then broken up into segments or blocks which are each analysed and transmitted to the receiver.

Voice/Unvoiced Determination

According to LPC-10 standards, before a speech segment is determined as being voiced or unvoiced, it is first passed through a low-pass filter with a bandwidth of 1 kHz. Determining if a segment is voiced or unvoiced is important because voiced sounds have a different waveform then unvoiced sounds. The differences in the two waveforms create a need for the use of two different input signals for the LPC filter in the synthesis or decoding. One input signal is for voiced sounds and the other is for unvoiced. The LPC encoder notifies the decoder if a signal segment is voiced or unvoiced by sending a single bit.

Pitch Period Estimation

Determining if a segment is a voiced or unvoiced sound is not all of the information that is needed by the LPC decoder to accurately reproduce a speech signal. In order to produce an input signal for the LPC filter the decoder also needs another attribute of the current speech segment known as the pitch period. The period for any wave, including speech signals, can be defined as the time required for one wave cycle to completely pass a fixed position.

Vocal Tract Filter

The filter that is used by the decoder to recreate the original input signal is created based on a set of coefficients. Each speech segment has different filter coefficients or parameters that it uses to recreate the original sound. Not only the parameter themselves differs from segment to segment, but the number of parameters differ from voiced to unvoiced segment. Voiced segments use 10 parameters to build the filter while unvoiced sounds use only 4 parameters.

In order to find the filter coefficients that best match the current segment being analysed, the encoder attempts to minimize the mean squared error. The error signal will be as small as possible and represents the difference between the predicted signal and the original[8,9] . The mean squared error is expressed as:

$$e_n^2 = (y_n - \sum_{i=1}^{M} a_i y_{n-i} + G\epsilon_n)^2$$

where {yn} is the set of speech samples for the current segment and {ai} is the set of coefficients. In order to provide the most accurate coefficients, {ai} is chosen to minimize the average value of en^2 for all samples in the segment.

The first step in minimizing the average mean squared error is to take the derivative.

$$\frac{\delta}{\delta a_j} \mathrm{E}\,[(y_n - \sum_{i=1}^{M} a_i y_{n-i} + G\epsilon_n)^2] = 0$$

$$\rightarrow -2\mathrm{E}\,[(y_n - \sum_{i=1}^{M} a_i y_{n-i} + G\epsilon_n)\, y_{n-j}] = 0$$

$$\rightarrow \sum_{i=1}^{M} a_i \mathrm{E}[y_{n-i} y_{n-j}] = \mathrm{E}[y_n y_{n-j}]$$

$$(\text{Use fact that } \mathrm{E}[y_n y_{n-j}] = 0 \text{ if } j \neq 0)$$

Taking the derivative produces a set of M equations. In order to solve for the filter coefficients E[yn-iyn-j] has to be estimate. There are two approaches that can be used for this estimation: autocorrelation and autocovariance.

LPC Synthesis/Decoding

The process of decoding a sequence of speech segments is the reverse of the encoding process. Each segment is decoded individually and the sequence of reproduced sound segments is joined together to represent the entire input speech signal. The decoding or synthesis of a speech segment is based on the 54 bits of information that are transmitted from the encoder.

LPC Applications

In general, the most common usage for speech compression is in standard telephone systems. In fact, a lot of technologies used in speech compression were developed by the phone companies. Linear predictive coding only has application in the area of secure telephony because of its low bit rate. Secure telephone systems require a low bit rate since speech is first digitalized, then encrypted and transmitted [6,7]. These systems have a primary goal of decreasing the bit rate as much as possible while maintaining a level of speech quality that is understandable. Other standards such as the digital cellular standard and the international telephone network standard have higher quality standards and therefore require a higher bit rate. In these standards, understanding the speech is not good enough, the listener must also be able to recognize the speech as belonging to the original source.

North American Telephone Systems
64 kb/s (uncompressed)
International Telephone Network
32 kb/s (can range from 5.3-64kb/s)

Digital Cellular standards
6.7-13 kb/s
Regional Cellular standards

3.45-13 kb/s
Secure Telephony
0.8-16 kb/s
(Bit Rates for different telephone standards)

A second area that linear predictive coding has been used is in Text-to-Speech synthesis. In this type of synthesis the speech has to be generated from text. Since LPC synthesis involves the generation of speech based on a model of the vocal tract, it provides a perfect method for generating speech from text.

Further applications of LPC and other speech compression schemes are voice mail systems, telephone answering machines, and multimedia applications. Most multimedia applications, unlike telephone applications, involve one-way communication and involve storing the data. An example of a multimedia application that would involve speech is an application that allows voice annotations about a text document to be saved with the document. The method of speech compression used in multimedia applications depends on the desired speech quality and the limitations of storage space for the application. Linear Predictive Coding provides a favorable method of speech compression for multimedia applications since it provides the smallest storage space as a result of its low bit rate.

Conclusion

Linear Predictive Coding is lossy speech compression that attempts to model the human production of sound instead of transmitting an estimate of the sound wave. Linear predictive coding achieves a bit rate of 2400 bits/second which makes it ideal for use in secure telephone systems. Secure telephone systems are more concerned that the content and meaning of speech, rather than the quality of speech, be preserved. The tradeoff for LPC's low bit rate is that it does have some difficulty with certain sounds and it produces speech that sound synthetic. Linear predictive coding encoders break up a sound signal into different segments and then send information of each segment to the decoder. The encoder send information on whether the segment is voiced or unvoiced and the pitch period for voiced segment which is used to create an excitement signal in the decoder. The encoder also sends information about the vocal tract which is used to build a filter on the decoder side which when given the excitement signal as input can reproduce the original speech.

References

[1] Introduction to Data Compression, Khalid Sayood, Ed Fox (Editor), March 2000.

[2] Burrows M., and Wheeler, D. J. 1994. A Block-Sorting Lossless Data Compression Algorithm. SRC Research Report 124, Digital Systems Research Center.

[3] M. Arjona Ramírez and M. Minami, "Low bit rate speech coding," in Wiley Encyclopedia of Telecommunications, J. G. Proakis, Ed., New York: Wiley, 2003, vol. 3, pp. 1299-1308.

[4] Soo Hyun Bae, ECE 8873 Data Compression & Modeling, Georgia Institute of Technology , 2004

[5] R. Sproat, and J. Olive. Text-to-Speech Synthesis, Digital Signal Processing Handbook (1999) 9-11.

[6] M. Ali Hossain, M.M.Rahman, U.K.Prodhan, M.F.Khan, Implementation of Back-Propagation Neural Network For Isolated Bangla Speech Recognition, IJIST,July 2013, Vol. 3.

[7] J.P.Campbell, Speaker Recognition: A tutorial, Proc IEEE, Sep. 1997,Vol.85,pp. 1437.

[8] G.K.Vallabha and B. Tuller, Systematic errors in the formant analysis of steady-state vowels, Speech Communication, Vol. 38, 2002,pp.141.

[9] L. Flanagan, Speech Analysis, Synthesis, and Perception, 2nd edition, New York, Springer-Verlag, 1972.

Author's Profile

Nilesh Suresh Bhoyar
nilesh.bhoyar@gmail.com

Atul K. Shingarwade
atul_shingarwade@rediffmail.com

Harshalata Petkar
harshalata@rediffmail.com