International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Advanced Voice Excited Linear Predictive Coding With Noise Reduction

Pretty Varghese[†*] and Ramesh P.[†]

[†]Department of Electronics and Communication, CUSAT, College of Engineering Munnar, Kerala, India

Accepted 31 May 2015, Available online 03 June 2015, **Vol.5, No.3 (June 2015)**

Abstract

Speech coding is a major issue in the area of digital signal processing. 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 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. The linear predictor model provides a robust, reliable and accurate method for estimating parameters that characterize the linear, time varying system. This project aims to implement a voice excited LPC vocoder for low bit rate speech compression.

Keywords: *Levinson Durbin Recursion, Linear Predictive Coding, Autocorrelation, Quantization.*

1. Introduction

There exist many different types of speech compression that make use of a variety of different techniques. The frequency of human speech production ranges from around 300 Hz to 3400 Hz. Most forms of speech coding are usually based on a lossy algorithm. Lossy algorithms are considered acceptable when encoding speech because the loss of quality is often undetectable to the human ear.

Most methods of speech compression exploit the fact that speech production occurs through slow anatomical movements and that the speech produced has a limited frequency range. Another fact about speech production that can be taken advantage of is that mechanically there is a high correlation between adjacent samples of speech. Most forms of speech compression are achieved by modeling the process of speech production as a linear digital filter. The digital filter and its slow changing parameters are usually encoded to achieve compression from the speech signal.

All vocoders have four main attributes: bit rate, delay, complexity, quality. Any voice coder, regardless of the algorithm it uses, will have to make tradeoffs between these different attributes.

First attribute of vocoders the bit rate, is used to determine the degree of compression that a vocoder achieves. Uncompressed speech is usually transmitted at 64 kb/s using 8 bits/sample and a rate of 8 kHz for sampling. Any bit rate below 64 kb/s is considered compression. The linear predictive coder transmits speech at a bit rate of 2.4 kb/s, an excellent rate of compression.

Delay is another important attribute for vocoders that are involved with the transmission of an encoded speech signal. Vocoders which are involved with the storage of the compressed speech, as opposed to transmission, are not as concern with delay. The general delay standard for transmitted speech conversations is that any delay that is greater than 300 ms is considered unacceptable.

The third attribute of voice coders is the complexity of the algorithm used. The complexity affects both the cost and the power of the vocoder. Linear predictive coding because of its high compression rate is very complex and involves executing millions of instructions per second. LPC often requires more than one processor to run in real time.

The final attribute of vocoders is quality. Quality is a subjective attribute and it depends on how the speech sounds to a given listener. One of the most common tests for speech quality is the absolute category rating (ACR) test. This test involves subjects being given pairs of sentences and asked to rate them as excellent, good, fair, poor, or bad.

2. Methodology

The speech coder that will be developed is going to be analyzed using subjective analysis. Subjective analysis

*Corresponding author: **Pretty Varghese**

will consist of listening to the encoded speech signal and making judgments on its quality. The quality of the played back speech will be solely based on the opinion of the listener. The speech can possibly be rated by the listener either impossible to understand, intelligible or natural sounding. Even though this is a valid measure of quality, an objective analysis will be introduced to technically assess the speech quality and to minimize human bias. The results achieved from the voice excited LPC are intelligible. On the other hand, the plain LPC results are much poorer and barely intelligible.

A) LPC System Implementation

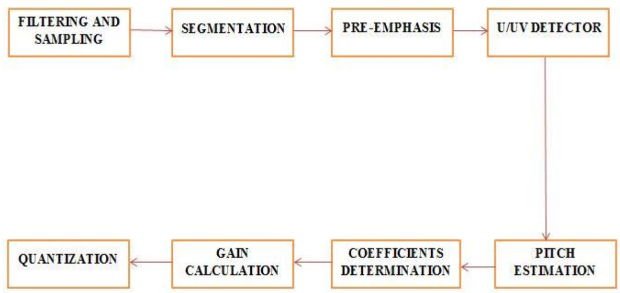


Fig. 1: LPC Encoder Block Diagram

1) Filtering and Sampling

First the speech is filtered by using a low pass filter. This will decrease the effect of high frequency noise in the speech signal. Usually a first order low pass filter is used for the purpose.

Then the speech is sampled at a frequency appropriate to capture all of the necessary frequency components important for processing and recognition. For voice transmission, 10 kHz is typically the sampling frequency of choice, it satisfies the nyquist theorem. For almost all speakers, all significant speech energy is contained in those frequencies below 4 kHz (although some women and children violate this assumption).

2) Segmentation

The speech is segmented into blocks for processing. Properties of speech signals change with time. To process them effectively it is necessary to work on a frame-by-frame basis .The actual duration of the frame is known as length. Length is selected between 10 and 30 ms or 80 and 240 samples. Within this short interval, properties of the signal remain roughly constant. Simple LPC analysis uses equal length blocks of between 10 and 30ms. Less than 10ms does not encompass a full period of some low frequency voiced sounds for male speakers. For certain frames with male speech sounded synthetic at 10ms sample windows, pitch detection became impossible. More than 30ms violates the basic principle of stationarity.

3) Pre-Emphasis

The spectral envelope of the speech signal has a high frequency roll-off due to radiation effects of the sound from the lips. Hence, high-frequency components have relatively low amplitude, which increases the dynamic range of the speech spectrum. As a result, LP analysis requires high computational precision to capture the features at the high end of the spectrum.

4) Voicing Detector

Voicing detector is used to classify a given frame as voiced or unvoiced. One or more speech parameters such as energy zero crossing rates is used for the purpose. The voiced speech has more energy and less zero crossing rates when compared to the unvoiced segment of speech.

5) Pitch Period Estimation

Pitch frequency is a unique characteristic of a person. Voiced sounds are generated when the airflow from the lungs is periodically interrupted by movements of the vocal cords. The time between these successive vocal cord openings is called the pitch period. For men, the possible pitch frequency range is usually between 50 and 250 Hz, while for women the range usually falls between 120 and 500 Hz.

6) Coefficients Determination

The coefficients of the difference equation characterize the formants. The estimate of these coefficients is done by minimizing the mean-square error between the predicted signal and the actual signal. The Levinson-Durbin algorithm is used to estimate the linear prediction coefficients from a given speech waveform.

7) Gain Calculation

The power of the prediction error sequence is used for calculating the gain of each segment.

For voiced case

$$P = 1/N\sum_{n=0}^{N-1} e^2[n] \qquad (1)$$

For unvoiced case

$$P = 1/[N/T]T\sum_{n=0}^{[N/T]T-1} e^2[n] \qquad (2)$$

This is based on the assumption that N>T, therefore the floor function assumes that the summation is always performed within the frame boundaries. Then the gain computation is done as:

$$G = \sqrt{P} \qquad (3)$$

8) Quantization

Direct quantization of the predictor coefficients is usually not considered during the quantization step. This is due to the effect that small changes in the predictor coefficient may lead to relatively large changes in the pole positions. Hence quantizing intermediate values is less problematic than quantifying the predictor coefficients directly.

B) Voice-excited LPC vocoder

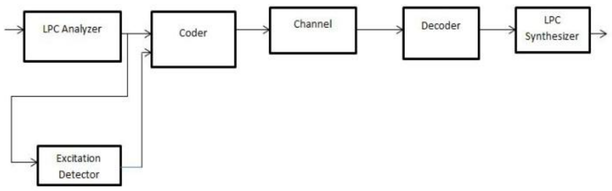


Fig.2: Block Diagram for Testing Sound Quality

First the speech is analyzed using the LPC analyzer. This will generate the residual signal that is needed as the excitation signal for the voiced segment. The coefficients and the speech parameters are also calculated during this phase. Then these signals are suitably encoded and transmitted.

At the decoder side these parameters are decoded and then are used by the inverse analyzer having following transfer function to generate the speech signal:

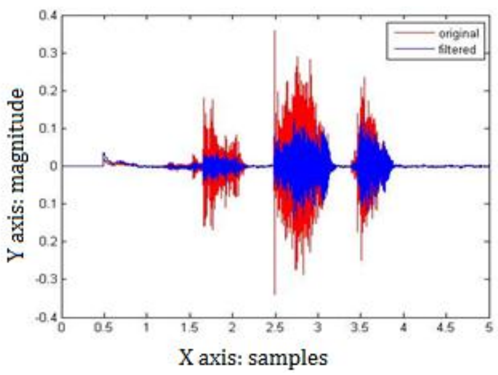
$$H(z) = \frac{G}{1 - \sum a_k z^{-k}} \quad (4)$$

The value of k varies from 1 to the number of coefficients.

3. Implementation

The Project is implemented in MatlabR2009a. And the project is divided in into two main sections namely the implementation with first order low pass filtering and without low pass filtering. The comparison of the output with and without filtering shows the need of the low pass filter at the input section.

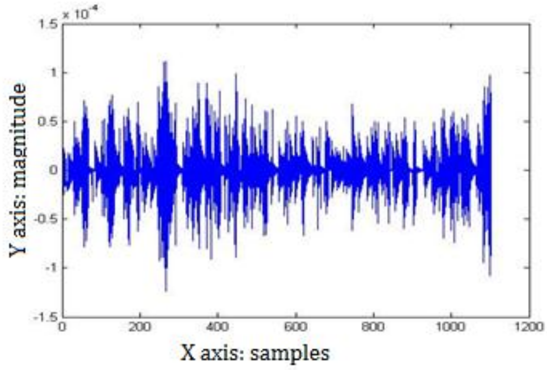
A) Input with and without filtering



This output waveform shows the need of the low pass filtering of input signal

B) LPC Vocoder

The filter coefficients were evaluated using the levinson durbin recursion algorithm. The original speech is recovered at the synthesizer section and its waveform is as shown below



Conclusions

The plain LPC results are much poorer and barely intelligible. Voice excited linear predictive type vocoders are more intelligible when compared with other systems. When a filter is employed in the input section we could decrease the effect of noise in the system. A low pass filter is usually used to decrease the effect of high frequency noises.

References

L. R. Rabiner and R. W Schafer, Digital Processing Of Speech Signals, Prentice- Hall, Englewood Cliffs, N.J.

K. Aizawa and T. S. Huang,(1995) Model-based image coding advanced video coding techniques for very low bit-rate applications, Proceedings of the IEEE, vol. 83, no. 2, pp. 259–271

B. S. Atal,(2006) The history of linear prediction, IEEE Signal Processing Magazine, vol. 23, no. 2, pp. 154–161.

A. Gersho, E. Shlomot, V. Cuperman, and C. Li, (May 15 2001)Method and apparatus for hybrid coding of speech at 4kbps,, uS Patent 6,233,550.

V. K. Goyal, Multiple description coding: Compression meets the network, Signal Processing Magazine, IEEE, vol. 18, no. 5, pp. 74 93, 2001.

D. W. Griffin and J. S. Lim,(1988) Multiband excitation vocoder, Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 36, no. 8, pp. 1223–1235.

N. Jayant, Digital coding of speech waveforms: Pcm, dpcm, and dm quantizers,(1974.) Proceedings of the IEEE, vol. 62, no. 5, pp. 611–632,

N. Jayant, J. Johnston, and R. Safranek,(1993) Signal compression based on models of human perception, Proceedings of the IEEE, vol. 81, no. 10, pp. 1385–1422,

N. S. Jayant and P. Noll,(1984.) Digital coding of waveforms: principles and applications to speech and video, Englewood Cliffs, NJ, pp. 115–251,

K. Kondo and T. Suzuki,(Feb. 20 1990) Method and system for transmitting variable rate speech signal,, uS Patent 4,903,301.