IJCSN International Journal of Computer Science and Network, May 2014
ISSN (Online) : 2277-5420
www.IJCSN.org
Conference Proceeding – RDFCORDGE - 2014

Speech Compression using LPC

[1]Chhayanka Sharma,[2]Daisy Rani Sahu,[3]Priyanka Gakre, [4] Ivneet Kaur Kaler

[1-3]E&TCDepartment (8th sem) ,
MM College of Technology, Raipur-492001,(C.G.), India

[4] Asst. Prof. E&TC Department ,
MM College of Technology, Raipur-492001,(C.G.), India

Abstract- One of the most powerful speech compression 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. The linear predictor model provides a robust, reliable and accurate method for estimating parameters that characterize the linear, time varying system.

In this project, we implement a voice excited LPC-10 vocoder for low bit rate speech compression. Levinson Durbin algorithm is used to compute LP coefficients, reflection coefficients and predictor error. The compress files contain LP coefficients and previous sample. These files are very small in size compared to the size of the original signals. Compression ratio is calculated from the size of the compressed signal relative to the size of the uncompressed signal. The proposed algorithms were fulfilled with the use of Mat lab package.

Keywords- *Speech signal, LPC, Speech compression, Pitch*

1. Introduction

Speech is the vocalized form of human communication and it is based upon the syntactic combinations of lexical and name that are drawn from very large vocabularies. Speech compression is the technology of converting human speech into an efficiently encoded representation that can later be decoded to reproduce an approximation of the original signal. Speech compression uses speech specific parameters and resulting parameters represent in a compact bit streams. Since it is not possible to access unlimited bandwidth, therefore there is need to code and compress speech signals. Speech coding has been and still is a major issue in the area of digital speech processing in which speech compression is needed for storing digital voice and it requires fixed amount of available memory and compression makes it possible to store longer messages. Several techniques of speech coding such as Linear Predictive Coding (LPC), Waveform Coding and Sub band Coding exist. This is used to characterize the vocal track and inverse filter is used to describe the vocal source and therefore it is used as the input for the coding. The speech coder that will be developed is going to be analyzed using subjective analysis. Subjective analysis will consist of listening to the encoded speech signal and making judgments on its quality. There are several different methods to successfully accomplish and Sub band coders. The speech coding in this Project will be accomplished by using a modified version of LPC-10 technique. Linear Predictive Coding is one possible technique of analyzing and synthesizing human speech. The exact details of the analysis and synthesis of this technique that was used to solve our problem will be discussed in the methodology section.LPC makes coding at low bit rates possible. For LPC-10, the bit rate is about 2.4 kbps. Even though this method results in an artificial sounding speech, it is intelligible. This method has found extensive use in military applications, where a high quality speech is not as important as a low bit rate to allow for heavy encryptions of secret data.

However, since a high quality sounding speech is required in the commercial market, engineers are faced with using other techniques that normally use higher bit rates and result in higher quality output. 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. In 1996, LPC-10 was replaced by mixed-excitation linear prediction (MELP) coder to be the United States Federal Standard for coding at 2.4 kbps. This MELP coder is an improvement to the LPC method, with some additional features that have mixed excitation, a periodic pulses, adaptive spectral enhancement and pulse dispersion filtering. Waveform coders on the other hand, are concerned with the production of a reconstructed signal whose waveform is as close as possible to the original signal, without any information about how the signal to be coded was generated. Therefore, in theory, this type of coders should be input signal independent and work for both speech and no speech input signals.

2. Speech Production

Speech production is the process by which spoken words are selected to be produced, have their phonetics formulated and then finally are articulated by the motor system in the vocal apparatus. Speech production can be spontaneous such as when a person creates the words of a conversation, reaction such as when they name a picture or read aloud a written word, or a vocal imitation such as in speech repetition. Speech production is not the same as language production since language can also be produced manually by signs. In ordinary fluent conversation people pronounce each second roughly four syllables, ten or twelve phonemes and two to three words out of a vocabulary that can contain 10 to 100 thousand words. Errors in speech production are relatively rare occurring at a rate of about once in every 900 words in spontaneous speech. Words that are commonly spoken or learned early in life or easily imagined are quicker to say than ones that are rarely said, learnt later in life or abstract.

Normally speech is created with pulmonary pressure provided by the lungs that generates sound by phonation in the glottis in the larynx that then is modified by the vocal tract into different vowels and consonants. However speech production can occur without the use of the lungs and glottis in a laryngeal speech by using the upper parts of the vocal trait. An example of such a laryngeal speech is Donald Duck talk. The vocal production of speech can be associated with the production of synchronized hand gestures that act to enhance the comprehensibility of what is being said. Before proceeding with the handling of digitized speech, it is crucial to have a basic understanding of how speech is produced. Speech is produced when the lungs force the direction of airflow to pass through the larynx into the vocal tract. In normal speech production, the air that is driven up from the lungs is passed through the glottis and vocal tract narrowing resulting in periodic or non-periodic (noise) excitation.

2.1 Speech Production Model

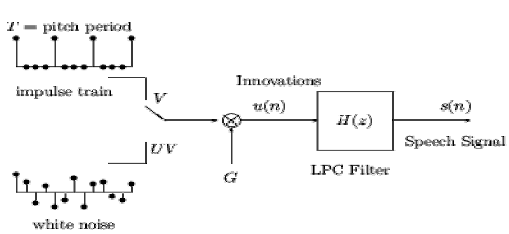

Figure 1:Speech Production

could be distinguished to be different. Several techniques of speech coding such as Linear Predictive Coding (LPC), Waveform Coding and Subband Coding

2.2 Speech Coding

Speech coding has been a common area of research in signal processing since the introduction of wire-based telephones. Numerous speech coding techniques have been thoroughly researched and developed, spurned further by the advances in internet, technology and wireless communication. Speech coding is a fundamental element of digital communications, continuously attracting attention due to the increase of demands in telecommunication services and capabilities. Applications of speech coders for signal processing purposes has improved at a very fast pace throughout the years in order to allow it to take advantage of the increasing capabilities of communication technology infrastructure and computer hardware.

This project focuses on the area of speech coding. This particular area has become a fundamental necessity due to the bandwidth limitation of most signal transmission systems. Ideally in speech coding, a digital representation of a speech signal is coded using a minimum number of bits to achieve a satisfactory quality of the synthesized signal whilst maintaining a reasonable computational complexity. Speech coding has two main applications: digital transmission and storage of speech signals. In speech coding, our aim is to minimize the bit-rate while preserving a certain quality of speech signal, or to improve speech quality at a certain bit rate. Currently, there are various kinds of coders being implemented. This project focuses on the design and implementation of a Code Excited Linear Predictive (CELP) coder. This Linear Predictive Coding (LPC) method performs LP analysis of speech by extracting the LP parameters and coefficients and employs a quantization method to search a codebook and compute the excitation signal. The quantization of the LP parameters, play an important role in the performance of the CELP coder. This project analyzes the performance of the CELP coder by using various quantization methods such as Scalar, vector, DPCM and TSVQ to quantize the LP parameters.

The speech coder that is developed is analyzed using both subjective and objective analysis. Subjective analysis will consist of listening to the encoded speech signal and making judgment on its quality. The quality of the played back speech will be solely based on the opinion of the listener. An objective analysis will be introduced to technically assess the speech quality and to minimize human bias. The objective analysis will be performed by computing segmental signal to noise ratio(SEGSNR) between the original and the coded speech signal. Speech coding is a lossy type of coding, which means that the output signal does not exactly sound like the input. The input and the output signal exist. The speech signal that need to be coded are wideband signal with frequencies ranging from 0 to 8KHz. LPC is used to estimate basic speech parameter

IJCSN International Journal of Computer Science and Network, May 2014
ISSN (Online) : 2277-5420
www.IJCSN.org
Conference Proceeding – RDFCORDGE - 2014

like pitch formants and spectra. The principal behind the use of LPC is minimize the sum of the squared differences between the original speech signal and estimated speech signal over a finite duration. This could be used to give a unique set of predictor coefficients. These predictor coefficient are normally estimated every frame, which is normally 20ms long and another parameter is the gain.

3. Methodology

As mentioned above the LPC coding function will take the speech audio signal and divide it into 30mSec frames. These frames start every 20mSec. Thus each frame overlaps with the previous and next frame. Shown in the figure below:

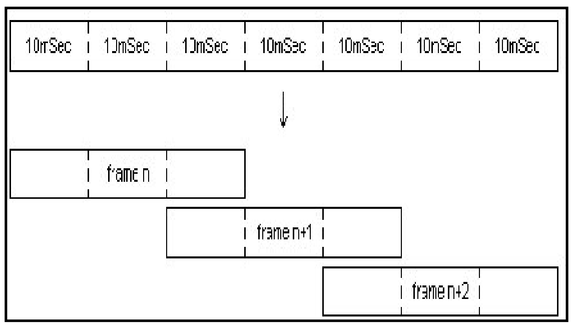

Figure2: Audio Signal to Separate Frames

After the frames have been separated, the LPC function will take every frame and extract the necessary information from it. This is the voiced/unvoiced, gain, pitch, and filter coefficients information. To determine if the frame is voiced or unvoiced we need to find out if the frame has a dominant frequency. If it does, the frame is voiced. If there is no dominant frequency the frame is unvoiced. If the frame is voiced you can find the pitch. The pitch of an unvoiced frame is simply 0. The pitch of a voiced frame is in fact the dominant frequency in that frame. One way of finding the pitch is to cross correlate the frame. This will strengthen the dominant frequency components and cancel out most of the weaker ones. If the 2 biggest data point magnitudes are within a 100 times of each other, it means that there is some repetition and the distance between these two data points is the pitch. The gain and the filter coefficients are found using Levinson's method. This code is already written here.

After finding these variables for all the frames the function will pass them back to the main file as seen below:

[Coeff, pitch, G] = proclpc (inspeech, Fs, Order);

Here,

. Coeff is a matrix of size (number of coefficients x number of frames) containing the filter coefficients to all the frames

· Pitch is a vector of size (number of frames) containing the pitch information to all the frames

· G is a vector of size (number of frames) containing the gain information to all the frames

· Inspeech is the input data

· Fs is the sample frequency of the input data

· Order is the order of the approximation filter

The general steps of algorithms are as follows-

Step1- Read the sound file.
Step2- Divide the sound file into different frame of size 20 ms.
Step3- Enter the desired number of predictive coefficient.
Step4- Compute the voice signal from the frame.
Step5-Compute the pitch of the frame.
Step6- Compute the gain of the voice signal.
Step7- Compute the LPC coefficient.

Once the LPC coefficient (usually 10), voice signal, pitch and gain of sound signal is computed then the sound signal can be represented by these 13 elements. With the help of these 13 elements, a sound signal similar to the original one can be reproduced but with degraded quality. Therefore compression using LPC method is called lossy compression. General block diagram of above steps is shown in below.

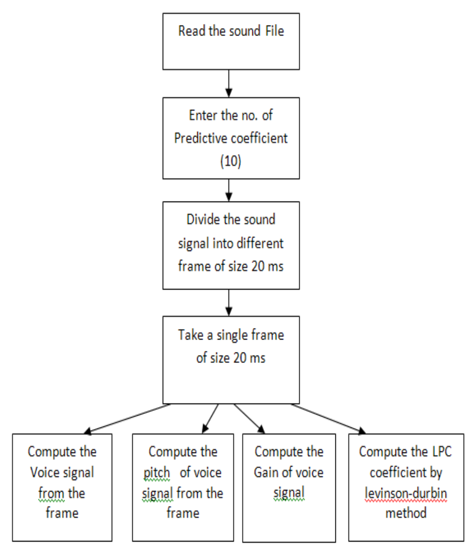

Figure 3: Block diagram of LPC analysis

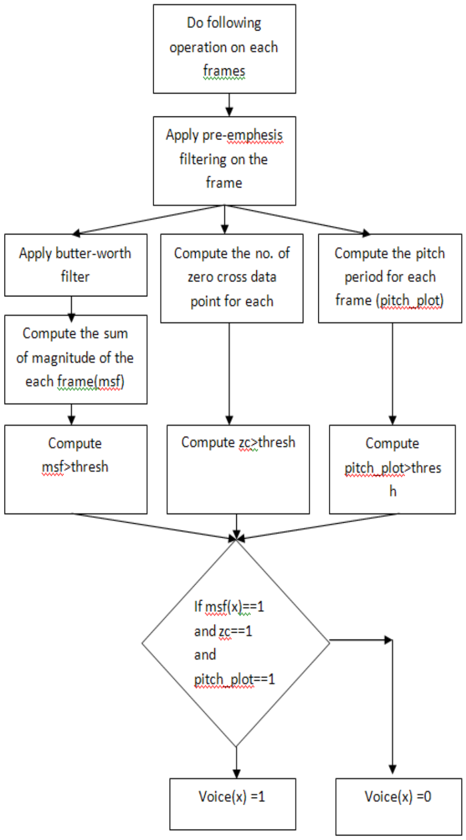

Figure 4: Block diagram of LPC analysis

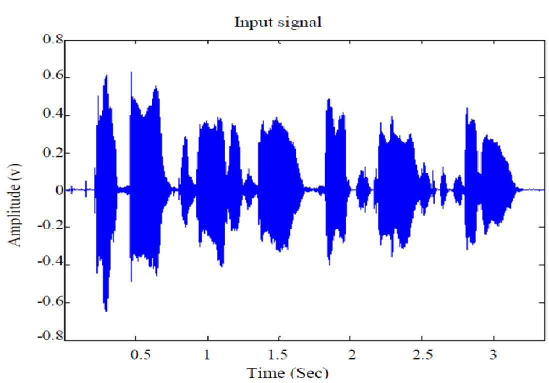

Figure 5: Input signal

Figure 6.1 shows the input signal at the encoder side. The signal is sampled at 16 KHz and the audio signal has audible quality. While the figure 5.2 shows the signal obtained by pre-emphasis filtering

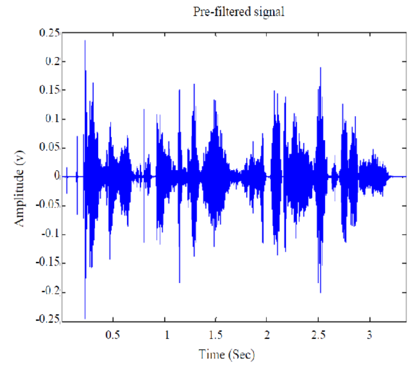

Figure 6:Pre-filtered signal

In any speech processing, detection of voiced signal from the sound signal is very important. In this work, the
method adopted to detect the voiced signal are as follows-

Step1- Extract the frame(of 20 ms size) from the sound signal.
Step2- Apply pre-emphesis filter on each frame to boost-up the high frequency component.
Step3-Apply butter-worth filter and then compute the sum of magnitude of all the data point of the frame.
Step4- Extract all the data point whose magnitude is greater then some predefined threshold.
Step5- Similarly extract the number of zero cross from the signal obtained after step2.
Step6- Extract all the data point whose zero cross is greater than the predefined threshold.
Step7- Compute the pitch of the frame.
Step8- all the data point for which magnitude, zero-cross, pitch==1,belong to the voice data point.

To verify the project design input signal is taken from the wave (Waveform Audio Format) file recorded in a quiet room with no noise effects.

4. Result and Conclusion

The implementation of audio coder based on perceptual linear predictive coding represents another way to code audio signal. The project is successfully implemented in MATLAB. Linear predictor coefficients are obtained using Levinson's theorem. After decoding, the reconstructed audio signal should have the same quality as the original audio signal. This project provided a great opportunity to study and learn the fundamentals of digital signal processing. During implementation, we have used these fundamentals to get the best results for this project. we have finished all the requirements for the success of this project. After developing this project, we have learned the fundamentals of a psychoacoustic model and how it is combined with a linear predictor filter to process the audio signal.

References

[1] News and Analysis of Speech Recognition Markets, Products and Technology, Num. 73 pp: 1-32, July 1999.
[2] Graps, A., An Introduction To Wavelets, IEEE Computational Sciences and Engineering, Volume 2, Number 2, pp: 50-61, Summer 1995.
[3] NIST, TIDIGITS, Speech Discs, Studio Quality Speaker-Independent Connected-Digital Corpus, NTIS PB91-506592, Texas Instruments, Feb.1991.
[4] NIST, Speech Discs 7-1.1, TI 46 Word Speech Database Speaker- Dependent Isolated- Digital Corpus, LDC93S9, Texas Instruments, Sep.1991.
[5] Rabiner, L.R. and Schafer, R.W., Digital Processing of Speech Signals, Prentice Hall, New Jersey, 1978.
[6] Artimy, M., Phillips, W.J. and Robertson, W., Automatic Detection Of Acoustic Sub-word Boundaries For Single Digit Recognition Proceeding IEEE Canadian Conference on Electrical and Computer Engineering,1999.
[7] Phillips, W.J., Tosuner, C. and Robertson, W., Speech Recognition Techniques Using RBF Networks, IEEE, WESCANEX, Proceedings,1995.
[8] Cooley, J.W. and Tukey, J.W., An Algorithm For The Machine Computation Of Complex Fourier series, Mathematics of Computation, Vol. 19. pp: 297-301, 1965.
[9] Oppenheim, A.V. and Schafer, R.W., Discrete-Time Signal Processing, Prentice Hall, Englewood Cliffs, New Jersey, 1989.
[10] Rabiner, L. Juang, B., Fundamental of Speech Recognition, Prentice Hall, New Jersey, 1993.
[11] Chan, Y.T., Wavelet Basics, Kluwer Academic Publisher, Boston, 1995.
[12] Reddy, D.R., Computer recognition of connected speech, Journal of the Acoustical Society of America, Vol. 42, pp:329-347, 1967.
[13] Rabiner, L. and Sambur, M.R., An algorithm for determining the end points of isolated utterances, Bell Systems Technical Journal, Vol.54, pp: 297-315, Feb. 1975.
[14] Strang, G. and Nguyen, T., Wavelets and Filter Banks, Wellesley MA, Wellesley–Cambridge Press, Wellesley, MA, 1996.
[15] Gabor, D., Theory of Communication, Journal of the IEEE No. 93, pp: 429-456, 1946.
[16] Picone, J.W., Signal Modeling Techniques in Speech Recognition, IEEE, Vol.81, No.9, September 1993.