

Speech Compression Using Voice Excited Linear Predictive Coding

Ms.Tosha Sen, Ms.Kruti Jay Pancholi

PG Student, Asst. Professor,
L J I E T, Ahmedabad

Abstract : The aim of the thesis is design good quality encode and decode transmission for long distance communication. 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. 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 decoder for low bit rate speech compression. **Index Terms:** Autocorrelation, Discrete Cosine Transform, Levinson Durbin Recursion, and Linear predictive coding (LPC).

Index Terms: Autocorrelation, Discrete Cosine Transform, Levinson Durbin Recursion, and Linear predictive coding (LPC).

INTRODUCTION

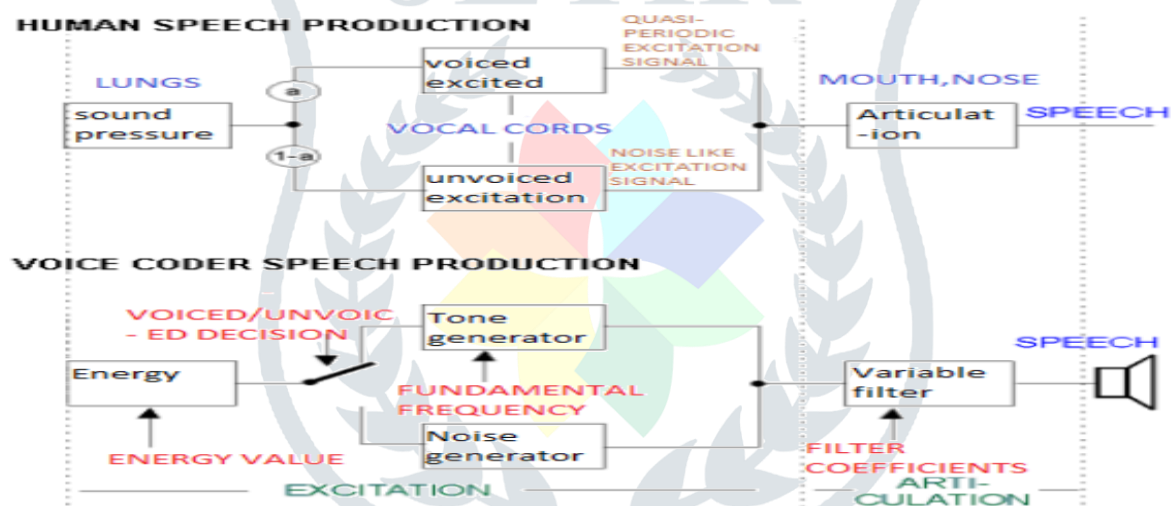


Fig1 human speech production

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. 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.

II. BACKGROUND

There are several different methods to successfully accomplish speech coding. Some main categories of speech coder are LPC decoders, Waveform 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, aperiodic 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 non-speech input signals.

III. METHODOLOGY

LPC System Implementation

1-Sampling

First, the speech is sampled at a frequency appropriate to capture all of the necessary frequency components important for processing and recognition. According to the Nyquist theorem, the sampling frequency must be at least twice the bandwidth of the continuous-time signal in order to avoid aliasing. For voice transmission, 10 kHz is typically the sampling frequency of choice, though 8kHz is not unusual. This is because, 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 then 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, where a frame consists of a certain number of samples. The actual duration of the frame is known as length. Typically, 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 stationary.

3- Pre-emphasis:

The typical 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. One simple solution is to process the speech signal using the filter with system function

$$H(z) = 1 - \alpha z^{-1}$$

This is high pass in nature. The purpose is to augment the energy of the high frequency spectrum. The effect of the filter can also be thought of as a flattening process, where the spectrum is „whitened“. Denoting $x[n]$ as the input to the filter and $y[n]$ as the output, the following difference equation applies:

$$Y[n] = x[n] - \alpha x[n]$$

The filter described in (1) is known as the pre-emphasis filter. By pre-emphasizing, the dynamic range of the power spectrum is reduced. This process substantially reduces numerical problems during LP analysis, especially for low precision devices. A value of α near 0.9 is usually selected. It is common to find in a typical speech coding scheme that the input speech is first pre-emphasized using (1). To keep a similar spectral shape for the synthetic speech, it is filtered by the de-emphasis filter with system function at the decoder side, which is the inverse filter with respect to pre-emphasis.

$$G(z) = 1 / (1 - \alpha z^{-1})$$

The main goal of the pre-emphasis filter is to boost the higher frequencies in order to flatten the spectrum. This pre-emphasis leads to a better result for the calculation of the coefficients using LPC. There are higher peaks visible for higher frequencies in the LPC spectrum. Clearly the coefficients corresponding to higher frequencies can be better estimated.

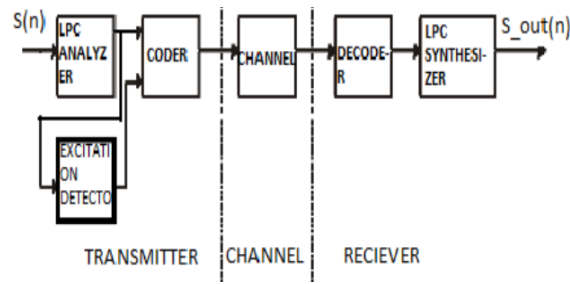
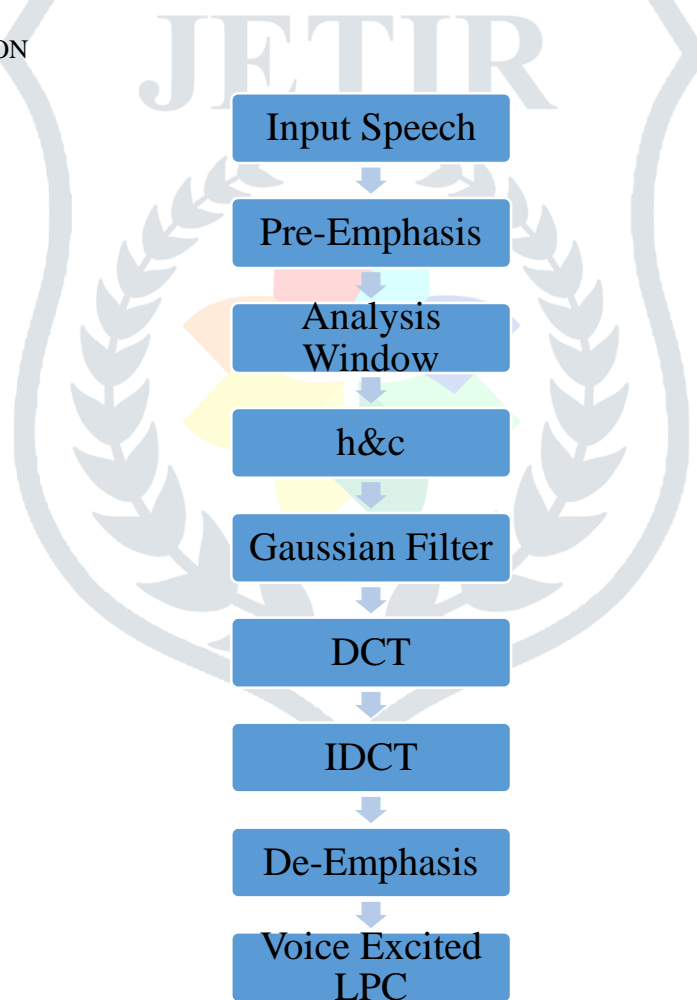


Fig 2 Tx and Rx CHANNEL

Voice-excited LPC Vocoder As the test of the sound quality of a plain LPC-10 vocoder showed, the weakest part in this methodology is the voice excitation. It is know from the literature that one solution to improve the qualityof the sound is the use of voice-excited LPC vocoders . Systems of this type have been studied by Atal et al. and Weinstein. Fig.3. shows a block diagram of a voice-excited LPC vocoder. The main difference to a plain LPC-10 vocoder, is the excitation detector, which will be explained in the sequel. To achieve a high compression rate ,the discrete cosine transform (DCT) of the residual signal could be employed. The DCT concentrates most of the energy of the signal in the first few coefficients. Thus one way to compress the signal is to transfer only the coefficients, which contain most of the energy. The tradeoff, however, is paid by a higher bit rate, although there is no longer a need to transfer the pitch frequency and the voiced /unvoiced information. We therefore looked for a solution to reduce the bit rate to 16 Kbits/sec.

I. IMPLEMENTATION



The project has been implemented in sMatlabR2009a. It has been divided into 3 parts namely basic LPC vocoder, Voice excited LPC model compressed using DCT, Voice excited LPC model compressed without using DCT. The waveform generated by each of these techniques have been plotted and analysed.

A) Original Speech

Basic LPC Vocoder: For implementing the basic LPC Vocoder, the pitch period is assumed to be 7.5ms. The filter coefficients have been evaluated using the Levinson-Durbin recursion algorithm. The original speech is recovered from the coefficients by passing it though a train of impulses which models the voiced sections of the speech. The plot of the generated waveform is shown in

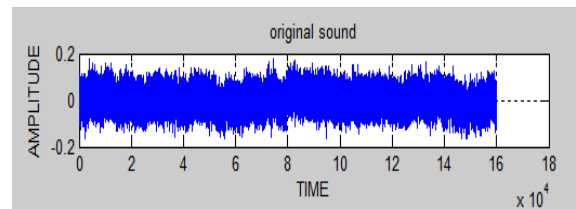


Fig 3 original signal

Voice excited LPC Model without Discrete Cosine transform: The quality of the compressed speech can be improved but at the cost of a higher bit rate. show in fig with LPC.

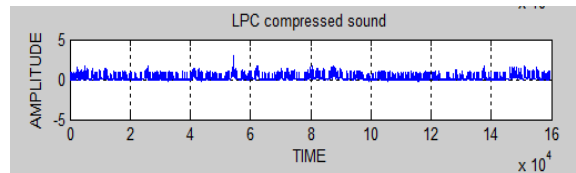


Fig 4 LPC Compression Sound

This is achieved by transmitting the encoding the residual signal as a whole without using DCT. This will help us achieve better reconstruction of the transmitted signal. The recovered signal from this method is shown in with voice exited signal.

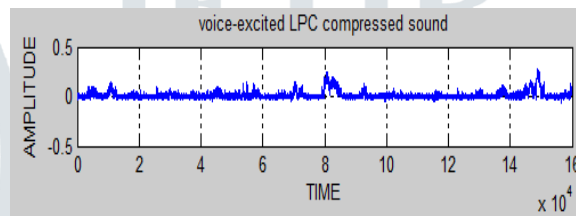


Fig 5: Voice Exited LPC Compression

CONCLUSIONS

The results achieved from the voice excited LPC are intelligible. On the other hand, the plain LPC results are much poorer and barely intelligible. This first implementation gives an idea on how a vocoder works, but the result is far below what can be achieved using other techniques. Nonetheless the voice-excited LPC used gives understandable results and is not optimized. The tradeoffs between quality on one side and bandwidth and complexity on the other side clearly appear here. If we want a better quality, the complexity of the system should be increased or a larger bandwidth has to be used. Since the voice-excited LPC gives pretty good results with all the required limitations of this project, we could try to improve it. A major improvement could come from the compression of the errors. If we can send them in a loss-less manner to the synthesizer, the reconstruction would be perfect. An idea could be the Use of Huffman code for the DCT.

Reference

- [1] Nikhil Sharma, Niharika Mehta, "Advanced Speech Compression VIA Voice Excited Linear Predictive Coding Using Discrete Cosine Transform(DCT)", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-2, Issue-3, February 2013, 144-148.
- [2] D.Ambika ,V.Radha, "A comparative study between Discrete Wavelet Transform and Linear Predictive Coding", IEEE, 978-1-4673-4805-8/12/\$31.00,2012.
- [3] J.Srinonchat, "New Technique to Reduce Bit Rate of LPC-10 Speech Coder", IEEE 1-4244-0549/\$20.00,2006 .
- [4] Mahmoud.A. Osman,Nasser Al, Hussein M. MagboubandS.AvAlfandi, "Speech compression using LPC and Wavelet", IEEE, 978-4244-6349-7/10/\$26.00,2010, V7-92—V7-99.
- [5] Rajesh G. ,Kumar A. , Ranjeet K. , "Speech Compression using Different Transform Techniques", IEEE , 978-4577-1386-611 \$26.00 2011, 146-151.

- [6] ArpanaMishra ,Javed Ashraf, “ *Speech Compression with Voice Excited Linear Predictive Coding*”, ISSN 2250-2459, Volume 2, Issue 6, June 2012, 306-309.
- [7] D. Deepa, C. Poongodi, A. Shanmugam, “*The Influence of Speech Enhancement Algorithm in Speech Compression with Voice Excited Linear Predictive Coding*”, Information Engineering (IE) Volume 2 Issue 4, December 2013, 68-72.
- [8] Sheetal D. Gunjal,Dr. Rajeshree D. Raut, “*Advance Source Coding Techniques for Audio/Speech Signal*”, IJCTA,Vol 3 (4),1335-1342, Aug-2012, 1335-1340.

