Machine Intelligence Research Labs
MIR Labs Global Network for Innovation and Research Excellence

# Speech Compression Using Linear Predictive Coding

Amol R. Madane
*TCS, New Delhi*
*amol.madane@gmail.com*

Zalak Shah
*SPCE, Mumbai*
*zalakshah@yahoo.com*

Raina Shah
*SPCE, Mumbai*
*raina@yahoo.com*

Sanket Thakur
*SPCE, Mumbai*
*sanketthakur78@gmail.com*

## *Abstract*

*The aim of the project is to develop a system for encoding good quality speech at a low bit rate. To implement this we have used most powerful speech analysis technique called Linear Predictive Coding (LPC). It uses 10th order Levinson-Durbin Recursion algorithm to accomplish the task. It provides extremely accurate estimates of speech parameters, and is relatively efficient for computation.The speech signal of males and females were coded. The tradeoffs between the bit rate, end-to-end delay, speech quality and complexity were analyzed. The results show that project was successful in coding the speech signal at relatively low bit rates with good quality.*

## 1. Introduction

Speech coding has been and still is a major issue in the area of digital speech processing. Speech coding is the act of transforming the speech signal to a more compact form, which can then be transmitted with a considerably smaller memory. It is not possible to access unlimited bandwidth Therefore, there is a need to code and compress speech signals. Speech compression is required in long-distance communication, high-quality speech storage, and message encryption. For example, in digital cellular technology many users need to share the same frequency bandwidth. Utilizing speech compression makes it possible for more users to share the available system. Another example where speech compression is needed is in digital voice storage. For a fixed amount of available memory, compression makes it possible to store longer messages.

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 could be distinguished to be different. Several techniques of speech coding such as Linear Predictive Coding (LPC), Waveform Coding and Subband Coding exist.. The speech signals that need to be coded are wideband signals with frequencies ranging from 0 to 8 kHz. The sampling frequency should be at 16 kHz with a maximum end-to-end delay of 100 ms. Different types of applications have different time delay constraints. For example in network telephony only a delay of 1ms is acceptable, whereas a delay of 500 ms is permissible in video telephony. Another constraint at hand is not to exceed an overall bit rate of 16 kbps. When all is said and done, the system must have less than 20 million operations per second (MOPS).

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 judgments 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. The report will be concluded with the summary of results and some ideas for future work.

## 2. Technical Work Preparation

In this section an explanation of the LPC speech coding technique will be given.

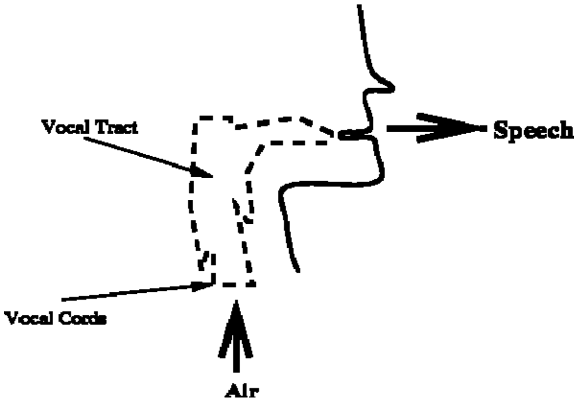

Fig.1 Physical Model

- For certain *voiced* sound, vocal cords vibrate . The rate at which the vocal cords vibrate determines the *pitch* of voice.

[1] Amol Madane is a Researcher, Multimedia Research Group, Innovation Labs, Tata Consultancy Services Ltd., New Delhi. Author had worked on Project during his Post Graduate course in Electronics Engineering Department, Sardar Patel College of Engineering (unaided), Andheri (w), Mumbai.

[2] Zalak Shah, [3] Raina Shah, [4] Sanket Thakur is a student of third year Electronics Engg.,S.P. College of Engg.,Mumbai,India.

- For certain *fricatives and plosive (or unvoiced)* sound, vocal cords do not vibrate but remain constantly opened.
- The shape of vocal tract determines the sound.
- As one speaks, vocal tract changes its shape producing different sound.

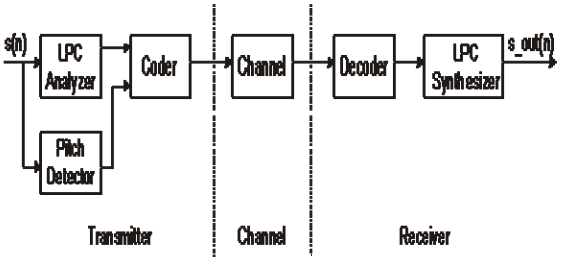

Fig. 2 Block diagram of an LPC vocoder

LPC is used to estimate basic speech parameters like pitch, formants and spectra. The principle behind the use of LPC is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. This could be used to give a unique set of predictor coefficients. These predictor coefficients are normally estimated every frame, which is normally 20 ms long. The predictor coefficients are represented by ak. Another important parameter is the gain (G). The transfer function of the time-varying digital filter is given by:

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

The summation is computed starting at k=1 up to p, which will be 10 for the LPC-10 algorithm, This means that only the first 10 coefficients are transmitted to the LPC synthesizer. The two most commonly used methods to compute the coefficients are, the covariance method and the auto-correlation formulation. For our implementation, we will be using the auto-correlation formulation. The reason is that this method is superior to the covariance method in the sense that the roots of the polynomial in the denominator of the above equation is always guaranteed to be inside the unit circle, Hence guaranteeing the stability of the system H (z). Levinson -Durbin recursion will be utilized to compute the required parameters for the auto-correlation method.

The LPC analysis of each frame also involves the decision-making process of concluding if a sound is voiced or unvoiced. If a sound is decided to be voiced, an impulse train is used to represent it, with nonzero taps occurring every pitch period. A pitch-detecting algorithm is employed to determine to correct pitch period /frequency. We used the autocorrelation function to estimate the pitch period. However, if the frame is unvoiced, then white noise is used to represent it and a pitch period of T=0 is transmitted. Therefore, either white noise or impulse train becomes the excitation of the LPC synthesis filter. It is important to re-emphasize that the pitch, gain and coefficient parameters will be varying with time from one frame to another.

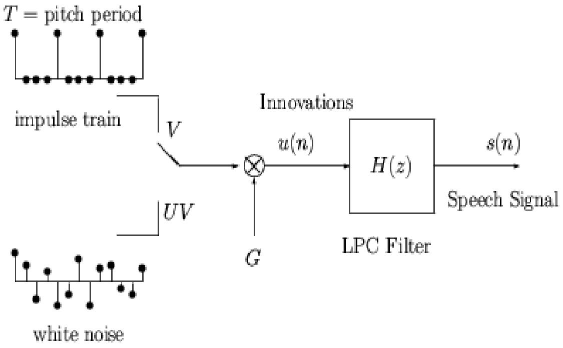

Fig. 3 Mathematical model for speech production

## 3. Quantization of LPC Coefficients

Usually direct Quantization of the predictor coefficients is not considered. To ensure stability of the coefficients (the poles and zeros must lie within the unit circle in the z-plane) a relatively high accuracy (8-10 bits per coefficients) is required. This comes from the effect that small changes in the predictor coefficients lead to relatively large changes in the pole positions. These are intermediate values during the calculation of the well-known Levinson-Durbin recursion. Quantizing the intermediate values is less problematic than quantifying the predictor coefficients directly.

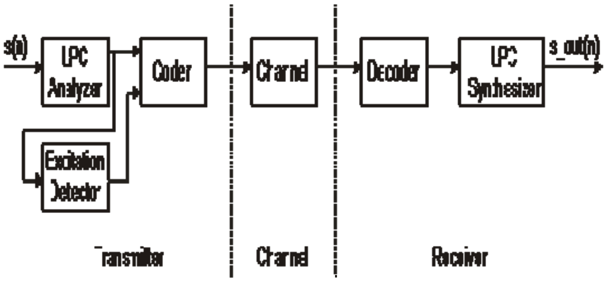

Fig. 4 Block diagram of a voice-excited LPC vocoder

The main idea behind the voice-excitation is to avoid the imprecise detection of the pitch and the use of an impulse train while synthesizing the speech Thus the input speech signal in each frame is filtered with the estimated transfer function of LPC analyzer. This filtered signal is called the residual. If this signal is transmitted to the receiver one can achieve a very good quality.

First of all, for a good reconstruction of the excitation only the low frequencies of the residual signal are needed. To achieve a high compression rate we employed the discrete cosine transform (DCT) of the residual signal. It is known, that 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. Our tests and simulations showed that these coefficients could even be quantized using only 4 bits. The receiver simply performs an inverse DCT and uses the resulting signal to excite the voice.

## 4. Mean Square Error

The formula for MSE is given by,

$$\text{MSE} = \{ \sum \text{err}^2 \} / N \qquad (2)$$

The difference between the original signal and the Reconstructed signal is Error signal, which is denoted as 'err'. Mean Square Error is calculated by taking the average of squares of sample values of the 'err'. The value of MSE should be as low as possible.

## 5. Comparative Analysis of LPC methods

A comparison of the original speech sentences against the LPC reconstructed speech and the voice-excited LPC method is done. In both cases, the reconstructed speech has a lower quality than the input speech sentences. The LPC reconstructed speech has a lower pitch than the original sound. The sound seems to be whispered. The voice-excited LPC reconstructed file sounds more spoken and less whispered. Overall the speech that was reconstructed using voice-excited LPC sounds better, but still sounds muffled. The waveforms in Fig 5 give the same idea. The voice-excited waveform looks closer to the original sound than the plain LPC reconstructed one.

### 5.1. Power Signal to Noise Ratio

$$\text{PSNR} = 10\log_{10} \{[\max (A)]/\text{MSE}\} \qquad (3)$$

where A=samples of original signal.

Power signal to noise ratio compares the level of a desired signal to the level of background noise. It is obvious that the first sound is very noisy, having a negative PSNR. The noise in this file is even stronger than the actual signal. The voice-excited LPC encoded sound sounds far better, and its PSNR, although barely, is in the positive side. However, even the speech coded with the improved voice-excited LPC does not sound exactly like the original signal.

### 5.2. Bit rate performance

The achieved bit rate in both method are quite low, both under the required 16kbps. However, the voice-excited LPC coding requires a bandwidth twice as large as the plain LPC coding. This huge increase ends up with a better sound, but still not perfect.

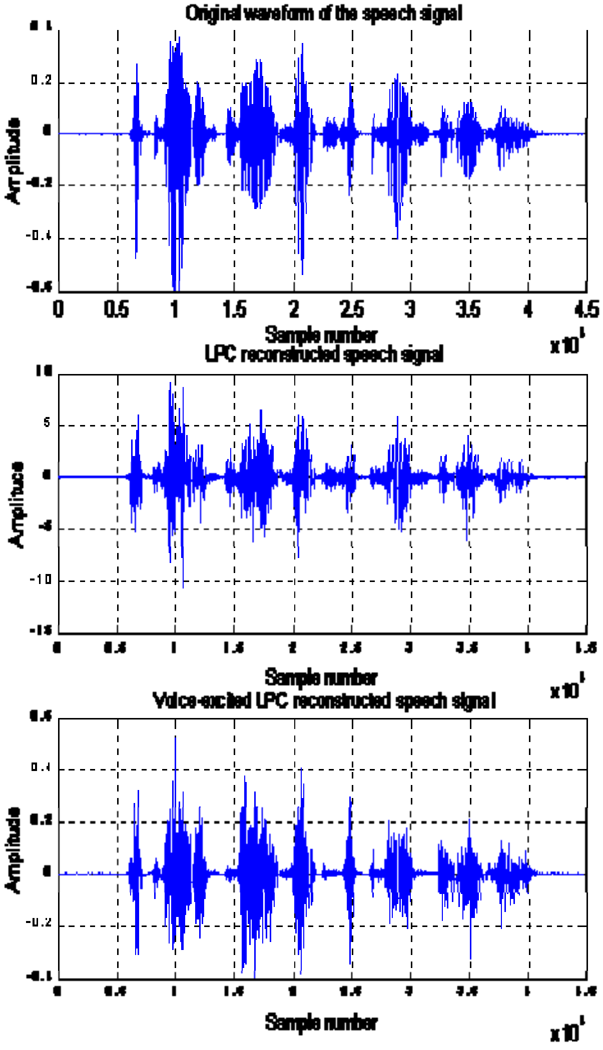

Fig. 5: a) original speech signal, b) LPC reconstructed speech signal, c) voice-excited LPC reconstructed speech signal

## 6. Conclusions

We carried out speech compression using two methods of LPC: Plain LPC and Voice-excited LPC. The quality of compressed signal obtained in case of Voice-excited LPC method is better than Plain LPC method. Though there is an improvement in quality when we use Voice-excited method, the bits per sample increases causing an increase in Bandwidth of the signal. But at the same time when SNR for both cases were compared it was observed that the sound due to Plain LPC was found to be more noisy, having a negative SNR. The noise in this is even stronger than the actual signal. The voice-excited LPC encoded sound sounds far better, and its SNR is positive. When the signal level is weak, system performance degrades.

## References

[1] M. H. Johnson and A. Alwan, "Speech Coding: Fundamentals and Applications", to appear as a chapter in the Encyclopedia of Telecommunications, Wiley, December 2002.

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

[3] B. S. Atal, M. R. Schroeder, and V. Stover, "Voice-Excited Predictive Coding Systetm for Low Bit-Rate Transmission of Speech", Proc. ICC, pp.30-37 to 30-40, 1975.

[4] C. J. Weinstein, "A Linear Predictive Vocoder with Voice Excitation", Proc. Eascon, September 1975.

[5] Orsak, G.C. et al. "Collaborative SP education using the Internet and MATLAB" IEEE Signal processing Magazine, Nov. 1995. vol.12, no.6, pp.23-32.