

Residual Excited Linear Predictive Coding

Mr. Santosh F. Bankar

Electronics & Telecommunication department
Government Engineering College
Aurangabad, India

Prof. N. R. Kolhare

Electronics & Telecommunication department
Government Engineering College
Aurangabad, India

Abstract - In this paper we present a low bit rate voice coding technique called the residual-excited linear prediction (RELP) coding. It uses 10th order Levinson-Durbin Recursive algorithm. It provides very good and accurate estimates of speech parameters and is relatively efficient for computation. In the RELP system, vocal tract modeling is done by the LPC technique, and the LPC residual signal is used as the excitation signal.

The range of the transmission rate is reduced to 9.6 kbits/s the synthetic speech in this range is quite good. As the transmission rate is lowered, the synthetic speech quality degrades very gradually. Since no pitch extraction is required, it is robust in any operating environment. The speech signal of males and females were coded and the results showed that the coding technique gives good speech quality with low complexity.

Keywords— Residual excited linear predictive coding.

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. Speech compression is required in long distance communication, high quality speech storage, and message encryption. For example, in digital cellular technology many 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 LPC, wave form coding and Sub band coding exist. The speech signals that need to be coded are wideband signals with frequencies ranging from 0 to 8 KHz.

The LPC approach is an analysis-synthesis method of coding in which the excitation (pitch) and the vocal tract modeling are treated separately. In the LPC approach the vocal tract is modeled by a time-invariant, all-pole recursive digital filter over a short time segment (typically 10 to 30 ms). The time variant character of speech is handled by a succession of such filters with different parameters. The excitation is modeled either as a series of pitch pulses (voiced) or as white noise (unvoiced).

There are, two types of linear prediction coding techniques: pitch-excited and residual-excited. The major difference

between these two types lies in how the excitation signal for the synthesizing filter is characterized. In a pitch excited LPC coding the vocal tract, glottal flow, and radiation are represented by the prediction coefficients. Those coefficients are transmitted together with the information regarding excitation of speech, that is, the fundamental frequency or pitch F_0 , the voiced/unvoiced (V/UV) decision, and a gain A extracted from either the residual signal or the speech input.

In the residual-excited LPC coding the vocal tract is characterized in the same way as in the pitch-excited one. However, instead of the excitation feature properties (V/UV, and A) being extracted and transmitted, the residual signal (or the prediction error signal) is encoded and transmitted.

2. HUMAN SPEECH PRODUCTION

Speech coding algorithms can be made more efficient by removing the irrelevant information from speech signals. In order to design a speech coding algorithm, it is necessary to know about the production of human speech, its properties and human perception of speech signal, so that the redundancies and the irrelevant parts of these signals can be identified.

A speech signal is produced in three stages: first of all, air flows outwards from the lungs; then the air flow is modified at the larynx; and finally further constriction of the airflow occurs by varying the shape of the vocal tract. The simplified diagram of vocal tract is shown below.

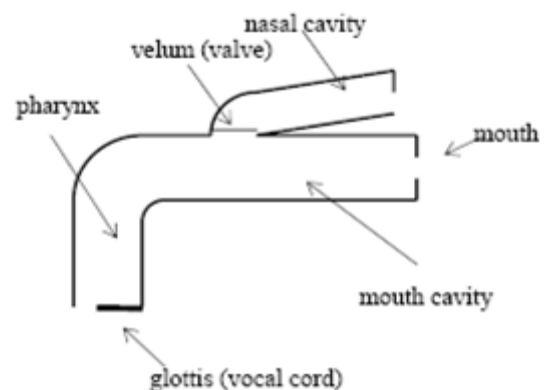


Figure 1. Vocal tract

For certain voiced sound, vocal cords vibrate. The rate at which the vocal cords vibrate determines the pitch of voice. For certain fricatives and plosive(or unvoiced) sounds,

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 sounds.

3. RELP CODER

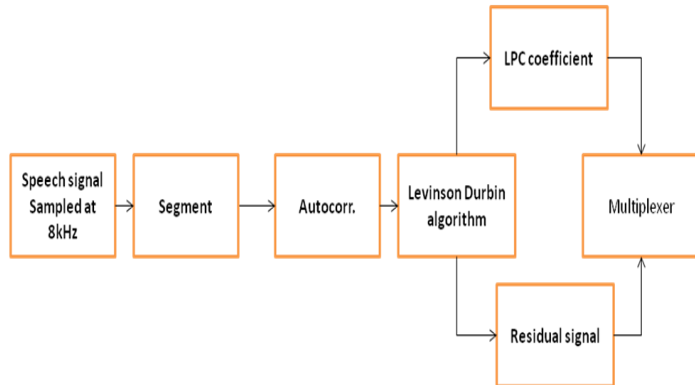


Figure 2. RELP analysis (Tx)

Predictive coding is a model based approach. Here it is suggested that an estimate $\hat{x}(n)$ for a signal sample value $x(n)$ can be predicted from past sample values using a model P , such that

$$\hat{x}(n) = P[x(n-1), x(n-2), \dots].$$

Depending on the type of the predictor P , there is some estimation method, which can be used to find parameters or coefficients of the model such that mean square error is minimized. Typically, the error is the expectation of the energy of a prediction error signal, given by

$$e(n) = x(n) - \hat{x}(n).$$

If the model is simply a linear weighted combination of previous sample values given by,

$$\hat{x}(n) = \sum_{k=1}^{L_u} c_k x(n-k).$$

This formulation leads to a technique which is linear prediction LP. The coefficients of the filters are estimated from an input so that the energy of prediction error is minimized by using Levinson durbin algorithm. This is performed usually in frames and the filter is updated after each interval of an analysis frame which is also called as segments.

4. LEVINSON DURBIN ALGORITHM

An efficient algorithm known as Levinson Durbin algorithm is used to estimate the linear prediction coefficients from a given speech waveform.

Assume that the present sample of the speech is predicted by the past M samples of the speech such that

$$\hat{x}(n) = a_1 x(n-1) + a_2 x(n-2) + \dots + a_M x(n-M) = \sum_{i=1}^M a_i x(n-i)$$

Where $\hat{x}(n)$ is the prediction of $x(n)$, $x(n-i)$ is the i th step previous sample, and $\{a_i\}$ are called the linear prediction coefficients. The error between the actual sample and the predicted one can be expressed as

$$\varepsilon(n) = x(n) - \hat{x}(n) = x(n) - \sum_{i=1}^M a_i x(n-i).$$

The sum of the squared error to be minimized is expressed as

$$E = \sum_n \varepsilon^2(n) = \sum_n \left(x(n) - \sum_{i=1}^M a_i x(n-i) \right)^2.$$

We would like to minimize the sum of the squared error. By setting to zero the derivative of E with respect to a_i (using the chain rule), one obtains

$$2 \sum_n x(n-k) \left(x(n) - \sum_{i=1}^M a_i x(n-i) \right) = 0 \quad \text{for } k = 1, 2, 3, \dots, M.$$

It results in M unknowns in M equations such that

$$\begin{aligned} a_1 \sum_n x(n-k)x(n-1) + a_2 \sum_n x(n-k)x(n-2) + \dots + a_M \sum_n x(n-k)x(n-M) \\ = \sum_n x(n-k)x(n) \quad \text{for } k = 1, 2, 3, \dots, M. \end{aligned}$$

If there are N samples in the sequence indexed from 0 to $N-1$ such that $\{x(n)\} = \{x(0), x(1), x(2), \dots, x(N-2), x(N-1)\}$ it can be approximately expressed in terms of matrix equation.

$$\begin{bmatrix} r(0) & r(1) & \dots & r(M-2) & r(M-1) \\ r(1) & r(0) & \dots & r(M-3) & r(M-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r(M-2) & r(M-3) & \dots & r(0) & r(1) \\ r(M-1) & r(M-2) & \dots & r(1) & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{M-1} \\ a_M \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(M-1) \\ r(M) \end{bmatrix}$$

$$R a = r.$$

Where

$$r(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k).$$

The sum of squared errors of the M -th order prediction (or simply the M -th order prediction error) can be rewritten as

$$E_M = \sum_n x(n)\varepsilon(n) - \sum_n \left(\sum_{i=1}^M a_i x(n-i) \right) \varepsilon(n)$$

Equation can be rewritten as

$$\sum_n x(n-i)\varepsilon(n) = 0 \quad \text{for } i = 1, 2, 3, \dots, M.$$

Because of equation, the second summation is zero. Thus, the final expression of the prediction error becomes

$$E_M = \sum_n x(n) \left(x(n) - \sum_{i=1}^M a_i x(n-i) \right)$$

$$= r(0) - a_1 r(1) - a_2 r(2) - \dots - a_{M-1} r(M-1) - a_M r(M) = r(0) - \sum_{i=1}^M a_i r(i)$$

We now want to develop a recursive method to solve Toplez matrix.

Initial values:

$$E_0 = r(0)$$

$$a_{11} = \kappa_1 = r(1)/E_0$$

$$E_1 = E_0(1 - \kappa_1^2).$$

With $m \geq 2$, the following recursion is performed

$$(i) \quad q_m = r(m) - \sum_{i=1}^{m-1} a_{i(m-1)} r(m-i)$$

$$(ii) \quad \kappa_m = \frac{q_m}{E_{(m-1)}}$$

$$(iii) \quad a_{mm} = \kappa_m$$

$$(iv) \quad a_{im} = a_{i(m-1)} - \kappa_m a_{(m-i)(m-1)} \quad \text{for } i = 1, \dots, m-1$$

$$(v) \quad E_m = E_{m-1} [1 - \kappa_m^2].$$

5. RELP ANALYSIS MEAN SQUARE ERROR:

The MSE is given by,

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

The difference between the original signal and the reconstructed signal is Error signal, which is denoted as 'err'.

Power Signal to Noise Ratio:

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

Where A=samples of original signal.

Power signal to noise ration compares the level of a desired signal to the level of background noise.

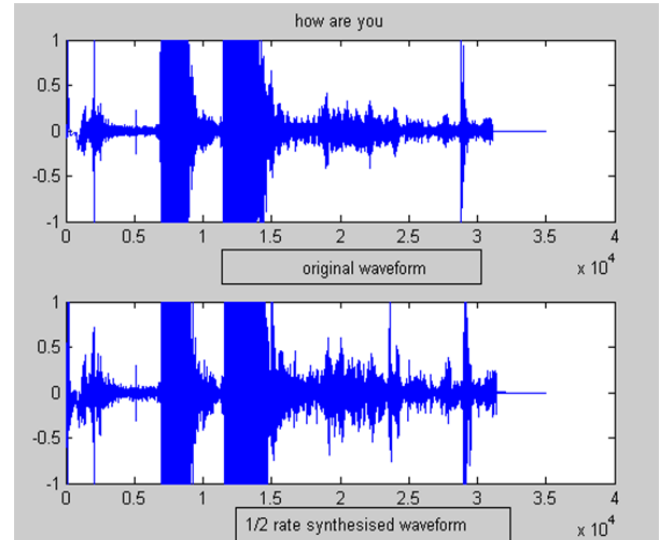


Figure 2. Original and 1/2 rate constructed signal

For 1/2 rates:

MSE=0.1891

PSNR=7.2336

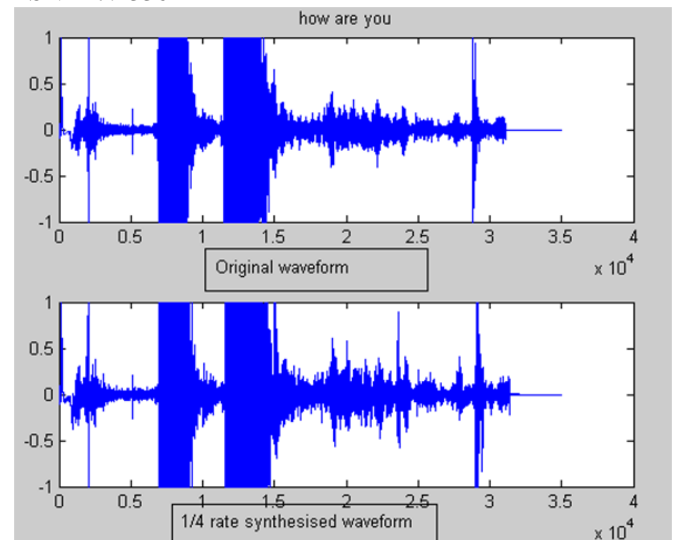


Figure 3. Original and 1/4 rate constructed signal

For 1/4 rate:

MSE=0.1898

PSNR=7.2163

6. CONCLUSION

We carried out speech compression for 1/2 rate and 1/4 rate in RELP the quality of compressed signal obtained was nearly same and the MSE and PSNR was also nearly same so with this we can conclude that with 1/4 rate of compression also the speech quality produced is very good.

REFERENCES

- (1) Zarkadis, D.J.; Evans, B.G, "Performance considerations of a 9.6kb/s RELP coder" IEEE Trans, pp.172-177, August 2002.
- (2) Katterfeldt, H., " A DFT-based residual-excited linear predictive coder" IEEE INFOCOM 2003, pp.824-827, January 2003.
- (3) Katterfeldt, H.; Behl, E., " Implementation of a robust RELP speech coder", IEEE 1983, pp.1316-1319.
- (4) Chong Un; Magill, D., " The Residual-Excited Linear Prediction Vocoder" IEEE comm., pp.1466-1474, jan.2003