RELP Coding for Voice Communication

Norman A. Lopez
Advisor: Dr. Shawn Hunt

Digital Signal Processing Laboratory
Electrical and Computer Engineering Department
University of Puerto Rico, Mayagüez Campus
Mayagüez, Puerto Rico 00681-5000
nlopez@ece.uprm.edu

Abstract

In mobile communication systems, bandwidth is a precious commodity, and service providers are continuously faced with the challenge of accommodating more users within a limited allocated bandwidth. Linear Predictive Coding (LPC) offers low bit-rate speech coding that can be used to meet this challenge. The lower the bit rate at which the coder can deliver toll quality speech, the more speech channels can be compressed within a given bandwidth. Work done in a class of LPC, Residual Excited LPC (RELP), is presented in this paper. The RELP coding of speech and its transmission were implemented using the MATLAB numerical computation software package.

1. INTRODUCTION

The design and subjective test of speech coders for mobile communications has been extremely difficult. Without low data rate speech coding, digital modulation schemes offer little in the way of spectral efficiency for voice traffic. The goal of all speech coding systems is to transmit speech with the highest possible quality using the least possible channel capacity. Speech coders differ widely in their approaches to achieving signal compression. Based on the means by which they achieve compression, speech coders are broadly classified into two categories: Waveform Coders and Vocoders [1]. LPC coding, specifically RELP, falls under the category of Vocoders.

Vocoders are a class of speech coding systems that analyze the voice signal at the transmitter, transmit parameters derived from the analysis, and then synthesize the voice at the receiver using those parameters. All vocoder systems attempt to model the speech generation process as a dynamic system and try to quantify certain physical constraints of the system [1]. Vocoders are much more complex than Waveform Coders, but they are less robust and their performance tends to be talker dependent. The most popular among the vocoding systems is LPC.

LPC belongs to the time domain class of vocoders. It provides an accurate and economical representation of relevant speech parameters that can reduce transmission rates in speech coding, increase accuracy and reduce calculation time in speech recognition, and generate efficient speech synthesis. The popularity of LPC derives from its compact yet precise representation of the speech spectral magnitude as well as its relative simplicity of computation [2]. In this paper, work done using one type of LPC: the Residual Excited LPC (RELP), is presented.

2. SYSTEM CONCEPTUALIZATION

In the RELP approach, standard LPC analysis yields the spectral coefficients. The number of spectral coefficients, L, is pre-determined. These spectral coefficients are used to synthesize the original speech. The residual sequence is then calculated and the spectral coefficients are transmitted as side information. A diagram of this operation is shown in figure 1.

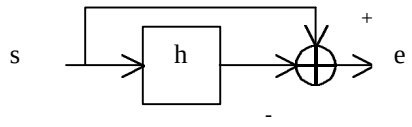

Figure 1: Diagram of LPC analysis

The signal, s, represents the original speech, is the synthesized speech signal, which is obtained by

$$= \mathbf{s} \times \mathbf{h},$$

and e is the residual sequence which is the subtraction of the original speech with the synthesized speech,

e = s – .

After the LPC analysis, the residual sequence is low-pass filtered. This is done to extract a baseband signal from the residual sequence. This baseband residual sequence is then decimated. The decimation factor depends on how much compression is desired and on how much aliasing is perceptually acceptable in the baseband of the output speech. The baseband residual sequence, the LPC coefficients and the first L original speech values are then coded to bits and transmitted. A diagram of the transmitter part of the RELP system is shown in figure 2.

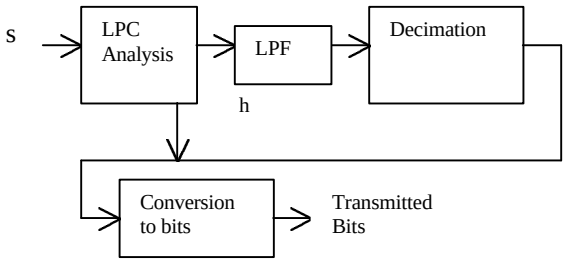

Figure 2: Transmitter part of RELP system

At the receiver, the received signal r, is first interpolated back to its original sampling rate. This interpolated signal is fullband reconstructed through nonlinear distortion. Only the high frequencies need to be reconstructed since the baseband signal is sent intact. To accomplish this high frequency regeneration, the interpolated signal is full-wave rectified followed by a double-differencing operation. The resulting signal is high-pass filtered using the same cut-off frequency as the low-pass filter. Now the baseband interpolated signal is added with the high-frequency signal to produce the residual sequence. Using the spectral coefficients that were transmitted, the L original values and the residual sequence the speech is reconstructed. This reconstructed speech, ø, is calculated using the following equation:

$$ø = \mathbf{y} \times \mathbf{h} + \mathbf{e},$$

where **y** are the L original speech values, **h** are the spectral coefficients, **e** is the reconstructed residual sequence and ø is the reconstructed speech signal. The receiver part of the RELP system is shown in figure 3.

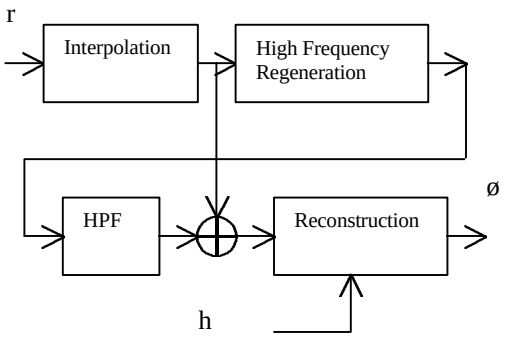

Figure 3: Receiver part of the RELP system

3. SYSTEM IMPLEMENTATION

The system was implemented in MATLAB using the transmitter and receiver parts shown in figures 2 and 3 respectively. The speech signal was generated using a sampling frequency of fs = 8 kHz and 8 bits per sample. The number of spectral coefficients was L = 10. The cutoff frequency for the low-pass and high-pass filter was fc = 800 Hz. The decimation factor, R, was obtained in the following manner:

$$R = fs / (2*fc) = 5.$$

At the transmitter, the residual sequence e, the first 10 original speech samples and the 10 spectral coefficients were converted to bits and transmitted through an ideal channel. At the receiver, the bits were decoded back to their corresponding values. The received baseband residual sequence was interpolated using a factor of 1/R. The high frequency regeneration was done by full-wave rectifying the interpolated signal, which basically means taking the absolute value of each sample. Then, the double-differencing operation was performed by subtracting a sample by its previous sample. The resulting high-pass signal was added to the interpolated baseband signal, and then used to reconstruct the original speech signal.

4. RESULTS

The plots of the original speech signal and the reconstructed speech signal are shown in figure 4.

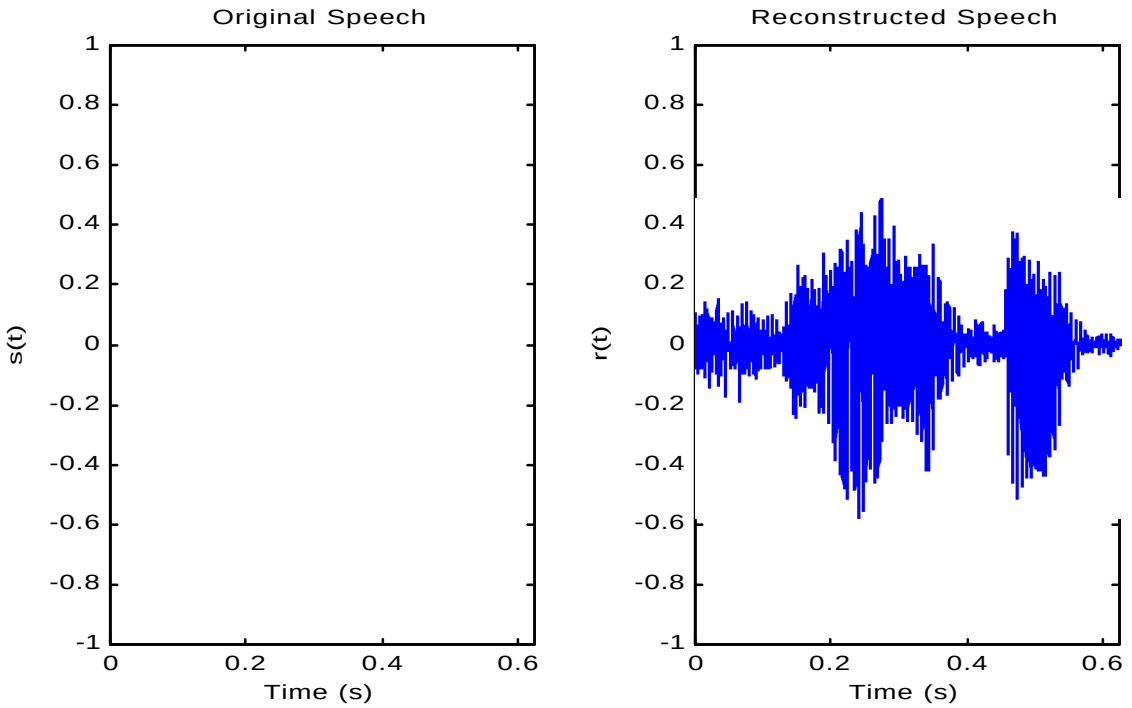

Figure 4: Plots of the original and reconstructed speech signals.

The Mean Square Error (MSE) between the original and the reconstructed speech signals was MSE = 0.0225. A plot of the error signal is shown in figure 5.

This system was also implemented in C – language and presently work is being done for its real-time implementation in a Texas Instrument DSP board.

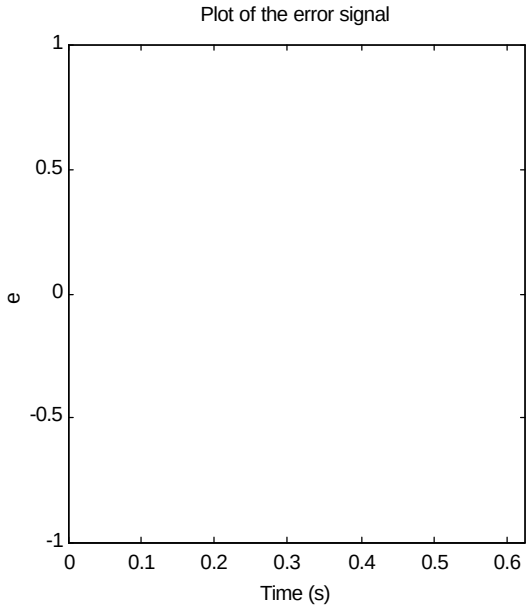

Figure 5: Plot of the error signal

REFERENCES

[1] Rappaport, Theodore S. , "Wireless Communications" , First Edition, Prentice Hall.

[2] O'Shaughnessy, Douglas, "Speech Communication" , First Edition, Addison and Wesley.