Implementation of CELP CODER and to evaluate the performance in terms of bit rate, coding delay and quality of speech

Article · April 2011

DOI: 10.1109/ICECTECH.2011.5942056

3 authors, including:

Bhavina Patel
Sarvajanik College of Engineering and Technology

Implementation of CELP CODER and to evaluate the performance in terms of bit rate, coding delay and Quality of speech

Milind Tandel
Electronics Engineering Department
G H Patel College of Engineering and Technology
Vallabhvidhyanagar, India
milindtandel@yahoo.com

Vandana Shah
Electronics and Communication Engineering Department
Sarvajainik College of Engineering and Technology
Surat, India
vandana.shah@scet.ac.in

Bhavina Patel
Instrumentation and Control Engineering Department
Sarvajanik College of Engineering and Technology
Surat, India
bhavina.patel@scet.ac.in

Abstract—**Factors serving as constraints in today's wireless communication system include bandwidth and power. In wireless systems that require the transmission of speech, these goals are addressed by developing efficient methods of reducing the amount of information required to transmit and receive quality speech. For this reason, speech coding has been, and remains, the topic of aggressive research.**

This paper discusses the implementation of CELP CODEC and its analytical evaluation of performance in terms of bit rate, coding delay and Quality of speech. The CELP coder is one of the best methods for producing high quality speech at bit rates between 4.8 and 9.6 Kbps.

Keywords-CELP, CODEC, LPC.

I. INTRODUCTION

Medium or low bit speech coders have been researched for application to mobile radio communication. Code Excited Linear Prediction (CELP) coding is one of the most effective coding methods at low bit rates which was proposed by Schroeder and Atal. CELP algorithm can produce low bit-rate coded speech comparable to that of medium rate waveform coders thereby bridging the gap between waveform coders and vocoders. CELP is an analysis-by-synthesis method; the excitation signal is selected by a closed loop search procedure and applied to synthesis filter. The synthesized waveform is compared to the original speech segment, the distortion is measured, and the process is repeated for all excitation code vectors stored in a code book. The index of 'best' excitation sequence is transmitted to the decoder, which retrieves the excitation code-vectors from the code book identical to that at the encoder. The name of "code excited" comes from the excitation codebook, containing the "code" to "excite" the synthesis filters. Figure 1 shows the CELP speech production model. The CELP coder relies on the long term and short term linear prediction models. Pitch synthesis filter creates periodicity in signal associated with the fundamental pitch frequency and the formant synthesis filter generates the spectral envelope. The codebook can be fixed or adaptive and can contain deterministic pulses or random noise.

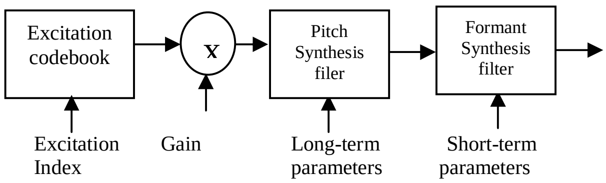

Figure 1. The CELP model of speech production

Since the signal is synthesized during encoding for analysis purposes, the principle is known as Analysis-by-Synthesis. Theoretically, all parameters of the speech coder can be optimized jointly so as to yield the best results. This approach however is too complex due to computation involved. In practice, only the subset of parameters is selected for closed loop optimization, while the rest are determined through an open-loop approach. A commonly used error criterion, such as a sum of squared error can be applied to select the final excitation sequence; hence waveform matching in the time domain is performed leading to partial preservation of phase information. Since the model requires frequent updating of the parameters to yield a good match to the original signal, the analysis procedure of the system is carried out in blocks.

II. CELP ENCODER

The detailed block diagram of generic CELP encoder is shown in Fig. 2. Input speech signal is segmented into frames and sub-frames. Length of frame is usually around 20 to 30 ms while that of sub-frame is in range of 5 to 7.5 ms. Short term LP analysis is performed on each frame to yield LPC. Afterwards, the long-term LP analysis is applied to each sub-frame. Input to short-term LP analysis is original speech; input to long-term LP analysis is short-term prediction error. Coefficients of perceptual weighting filter, pitch synthesis filter, and modified formant synthesis filter are known after this step.

The excitation sequence can now be determined. The length of each excitation code-vector is equal to that of the sub-frame; thus, an excitation code-book search is performed once every sub-frame.

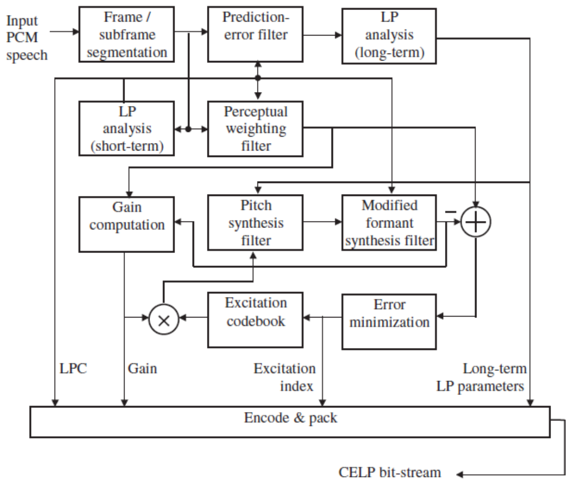

Figure 2. Detailed block diagram of generic CELP encoder

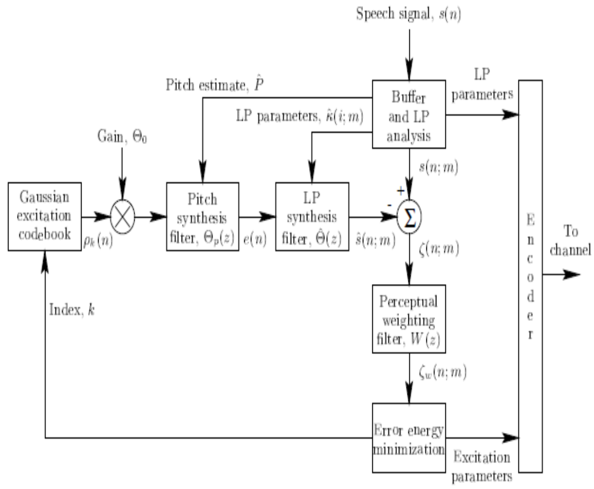

Figure 4. CELP analysis-by-synthesis coder.

The search procedure begins with the generation of an ensemble of filtered excitation sequences with the corresponding gains; mean squared error is computed for each sequence, and the code-vector and gain associated with the lowest error is selected. The index of excitation code-book, gain, long term LP parameters and the LPC are encoded, and transmitted as the CELP bit-stream.

III. CELP DECODER

The detailed block diagram of generic CELP Decoder is shown in Fig. 3.

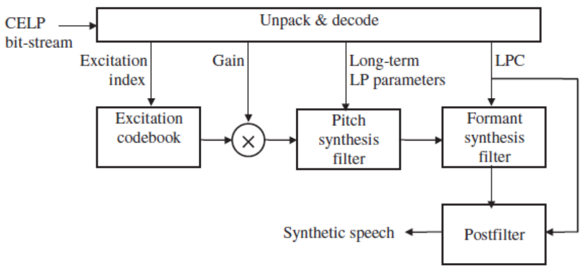

Figure 3. Detailed block diagram of generic CELP decoder

CELP decoder basically unpacks and decodes various parameters from the bit-stream, which are directed to corresponding block so as to synthesize the speech. The post filter is added at the end to enhance the quality of resultant signal.

IV. IMPLEMENTATION OF CELP CODER

This algorithm can be implemented in MATLAB by writing code and generating .m file.

First the LP analysis is used to estimate the vocal system impulse response in each frame. Then the synthesized speech ŝ (n; m) is generated at the encoder by exciting the vocal system filter model. The difference between the synthetic speech and the original speech s (name) constitutes an error signal ζ (n; m), which is spectrally weighted to emphasis perceptual important frequencies and then minimized by optimizing the excitation signal.

Optimal excitation sequences are typically computed over four blocks within the frame duration, meaning that the excitation is updated more frequently than the vocal system filter. Usually, a frame duration of 20 msec is used for the vocal tract analysis and 5 msec block duration for determining the excitation.

LP analysis estimates the all pole (vocal tract) filter in each frame, used to generate the spectral envelope (formants) of the speech signal. The function performs the LP analysis on a speech frame x(n;m) using Levinson- Durbin recursion and obtain residual error sequence ê(n;m) through inverse filtering. The input parameters are the signal frame x and prediction order M. The output parameters are estimated LP parameters â(i,m) in vector ar, the prediction error energies ξ(m) in vector xi, the estimated reflection coefficients k(i;m) in kappa and residual signal in that. The autocorrelation function r(η;m) of the residual error sequence ê(n;m) is used to estimate the pitch period P for voiced frame as

$$P = \arg\max r(\eta;m) \qquad \ldots\ldots\ldots\ldots \ (1)$$

If the peak value is below some threshold based on total residual energy, the frame is unvoiced.

Another function implements pitch estimation whose input parameters are the residua frame rhat, the threshold (th) and the lag range minlag and maxlag for peak search. The estimated LP synthesized filter Θ(z) is usually realized as a lattice filter. The output of the filter is the synthetic speech frame ŝ(n;m), which is subtracted from the original speech frame to form the residual error sequence.

$$\zeta(n;m) = s(n;m) - \hat{s}(n;m) \qquad \ldots\ldots\ldots\ldots\ldots \ (2)$$

The error sequence is passed through a perceptual error weighting filter with the system function

$$W(z) = \Theta(z/c) / \Theta(z) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Where c is a parameter, in the range $0 < c < 1$, that is used to control the noise spectrum weighting. In practice the range $0.7 < c < 0.9$ has proved more effective. The signal e(n) is used to excite the LP synthesis filter is determined dynamically every few milliseconds within the frame under analysis. First, an excitation sequence is selected from a Gaussian codebook of stored sequences. If the sampling frequency is 8 kHz and the excitation selection is performed every 5 msec, then the codebook word size is 40 samples. A codebook of 1024 sequence has been found to be sufficiently large to yield good quality speech, and require 10 bits to send the index. Gaussian codebook of 1024 sequences of length 40 is generated.

For voiced speech, the excitation sequence shows a significant correlation from one pitch period to the next. Therefore, a long delay correlation filter is used to generate the pitch periodicity in voiced speech. This filter typically has the form

$$\Theta_p(z) = 1/(1 - bz^{-P}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

where $0 < b < 1.4$ and P is an estimate of the number of samples in the pitch period. Thus the excitation sequence e(n) is modeled as a sum of a Gaussian codebook sequence and a sequence selected from an interval of past excitation.

Figure 5 shows the block diagram of CELP synthesizer. It consists of cascade of the two all-pole filters with that are updated periodically, and the excitation signal is taken from the codebook. This function is applied to the coded speech frame, and compared with the original frame. Observe the similarities between synthesized frame and the original speech frame.

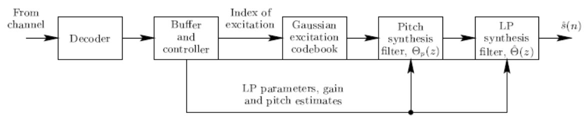

Figure 5 Block diagram of CELP synthesizer

Finally the frame based CELP coding and synthesis is performed on the speech signals i.e. the speech signal is segmented into the frames of size N.

V. OBJECTIVE EVALUATION OF CELP CODER

An objective criterion of evaluation on various speech and non-speech files includes calculation of different parameters like Absolute Error, Mean Square Error, Root Mean Square Error and Signal to Noise Ratio.

Signal to Noise Ratio (SNR) is calculated using following formulae :

$$sum = \sum (xhat[i])^2 \quad \ldots\ldots\ldots (5)$$

$$sum1 = \sum (xhat[i] - x[i])^2 \quad \ldots\ldots\ldots (6)$$

$$SNR = 10 \log (sum/sum1) \quad \ldots\ldots\ldots (7)$$

Mean squared error (MSE) is calculated as

$$MSE = sum1/duration \quad \ldots\ldots\ldots (8)$$

Root mean squared error (RMSE) is calculated as

$$RMSE = \sqrt{(MSE)} \quad \ldots\ldots\ldots (9)$$

Absolute error (ABBERR) = $\sum(xhat[i] - x[i])$…. … (10)

VI. RESULTS IN MATLAB WINDOW

Here, Co-efficient: 3-bits, 10^{th} order = 30-bits,
Gain: 4-bits, for 4 sub frames = 16-bits,
Pitch: 4-bits, for 4 sub frames = 16-bits,
Lag Pitch Filter: 4-bits, for 4 sub frames = 16-bits,
Codebook Index:5-bits, for 4 sub frames=20- bits,
Total: 98-bits,
Bit-rate: 4900 bits/second
Code-book Size: 40*1024

TABLE 1 OUTPUT OF VARIOUS FILES USING BIT RATE 4900 BITS/S

Parameters	clicks.wav	NO.wav	cricket.wav	horn.wav	brake.wav
Duration	8096	4598	6244	5320	5781
SNR	-1.0234	-0.2613	-0.4733	-0.1947	-0.1878
MSE	0.0202	0.1458	0.0819	0.4444	0.5844
RMSE	0.1420	0.3819	0.2862	0.6666	0.7645
ABSERR	66.8527	12.1723	50.4508	15.8345	52.6433

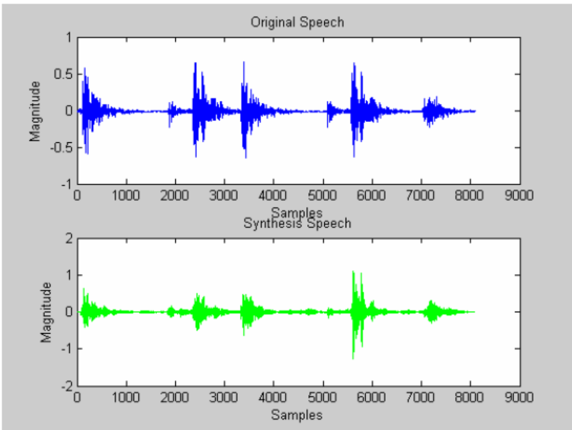

Figure 6 Input output waveform for click wave

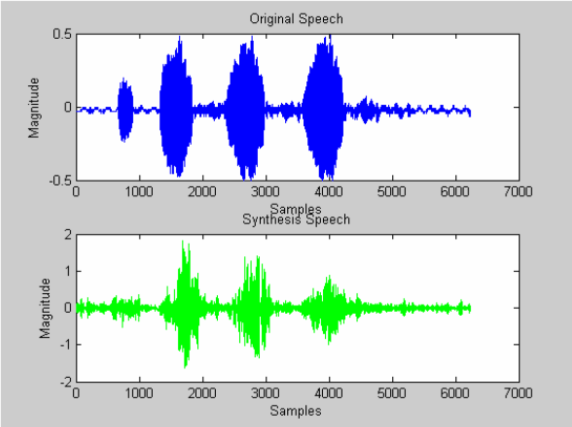

Figure 7 Input output waveform for cricket wave.

VII. CONCLUSION REMARK

First of all the CELP coder is studied and implemented in MATLAB and calculated various subjective and objective parameters like SNR, MSE, RMSE and ABSERR. The results are quite close to theoretical values and synthesized speech is similar to that of original one. The speech quality can still be improved by adding post filter at the end. From the waveforms of original and synthesized signal, it can be concluded that the synthesized speech is quite similar to original one.

REFERENCES

[1] Wai C. Cu – Speech Coding Algorithms Foundation and Evolution of Standardized Coder, John Wiley & sons, 2003 Edition.

[2] Theodore S. Rappaport – Wireless Communication Principle and Practice. 2005 Edition.

[3] A.M.Khamboh, Lewerence, University of Michigan- Project Report on 'Design of CELP Coder and analysis of various quantization techniques', Winter 2005.

[4] Z Yong Liu, M Ming Zhu, "Real Time Implementation Algorithm of CELP at 4.8 kb/s, Department of telecommunications Engg., Beijing University of posts and tele, P R. China, IEEE 1991.