Performance Analysis of CELP Codec for Gaussian and Fixed Codebooks

Sooraj S
Dept. of Electronics and Communication Engineering
MBCET, Trivandrum, India
sooraj_sureshkumar@ymail.com

Ancy S Anselam
Dept. of Electronics and Communication Engineering
MBCET, Trivandrum, India
ancy_anselam@yahoo.co.in

Sakuntala S Pillai
Senior Member IEEE
Dept. of Electronics and Communication Engineering
MBCET, Trivandrum, India
Sakuntala.pillai@gmail.com

***Abstract*—The design of high quality speech coders with low data rate is a very challenging task in speech processing. CELP provides good quality coded speech at a very low bit rate of 4.8kbps. Two types of codebooks namely Gaussian codebook and Fixed codebook are used in this analysis. The changes in the performance of the CELP codec is evaluated for theses two codebooks. Optimization of the codebook can be done by improving the training process in the codebook generation stage. Two different types of codebooks, the Gaussian codebook and the fixed codebook is generated to be implemented in the G723.1 CELP codec and the variation in the performance is evaluated.**

***Index Terms*—CELP, Gaussian Codebook, Fixed Codebook, PSVQ, K-Means.**

I. Introduction

In various applications there is a need for low bit rate and high quality speech codecs to facilitate better storage and transmission capabilities. An efficient speech coding algorithm which maintains the speech quality while achieving a very low bit rate of 4.8kbps is the Code Excited Linear Prediction(CELP) [1]. In CELP an Analysis-By-Synthesis [1] approach is used where an excitation signal is selected by a closed loop search and it is given to the synthesis filter. For this a codebook is used which is nothing but collection of the possible excitation signals. A comparison of the performance of the G723.1 CELP codec for a Gaussian codebook and a fixed codebook is done. In order to reduce the searching delay, the fixed codebook is divided into two smaller codebooks and the searching is done on these codebooks simultaneously. The performance is measured in terms of various parameters such as PESQ, PSNR, SNR, MSE.

II. CODE EXCITED LINEAR PREDICTION

Speech coders having low bit rate are used in many applications such as mobile communication for transmitting digital speech through analog channels. There is a need for reducing the data rate of the speech without affecting the quality of the speech for the purpose of storage and transmission. Code Excited Linear Prediction is one of the most efficient speech coding algorithms which can achieve a compression rate of up to 4.8kbps while maintaining the quality of the speech signal. In CELP the Analysis-by-Synthesis approach is implemented where the excitation is got from the closed loop search and then it is given to the synthesis part. The synthesized wave is compared with the input to reduce the error and to find the best excitation vector from the codebook. Then an index corresponding to that code vector is transmitted to the decoder. At the decoder end this index is used to retrieve the excitation code vector and reconstruct the speech using this code vector.

A. CELP ENCODER

The CELP coder is designed to provide the best possible speech quality while maintaining the computational complexity as low as possible at a rate of 4.8kbps. The LPC coefficients are calculated and further computations are done directly on them. The block diagram of CELP encoder [1] is shown in Fig 1. The input speech is segmented into frames and then to sub frames. The frame length is 20 msec containing 160 samples and these 20 msec frames are further divided to four subframes having length of 5 msec having 40 samples each. These sub frames are now used in the further stages. After converting to sub frames the analysis becomes easier since speech is non-stationary and quasi-periodic. Linear Prediction coefficients are extracted from each sub frame by short term linear prediction analysis. A total of 40 coefficients are generated from a single frame. These 40 coefficients are then used for further processing in the upcoming stages. The long term prediction is done on the input signal. Autocorrelation is applied to speech signal first and then the Livenson-Durbin [2] algorithm is applied to obtain the values of the LPC coefficients.

The autocorrelation function [1] is applied to speech is:

$$Arc(1) = \sum_{i=0}^{N-l-i} s(i) * s(i-l) \qquad (1)$$

The resultant LP coefficients [1] are given as:

$$LP_{A(Z)} = 1 - \sum_{i=1}^{k} aiz^{-1} \qquad (2)$$

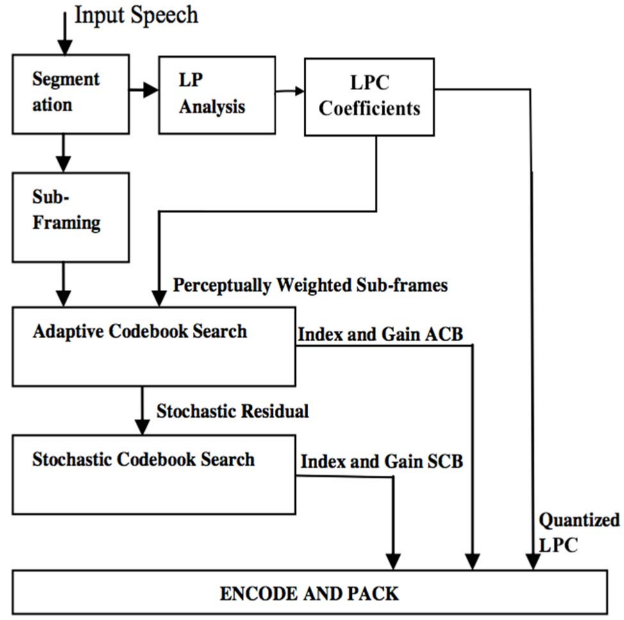

Fig.1: The CELP Coder Block Diagram

B. CELP DECODER

The CELP decoder [1] decodes the parameters from the CELP received bit stream. The same codebook that is present in the encoder is also available at the decoder side as well and the index number is used to select the stochastic code vector from the codebook. The weighted code vectors are filtered using the synthesis filter to reconstruct the input speech signal. Post filtering is done to reduce the quantization noise in the reconstructed speech and to improve some quality. The figure 2 shows the block diagram of the CELP decoder.

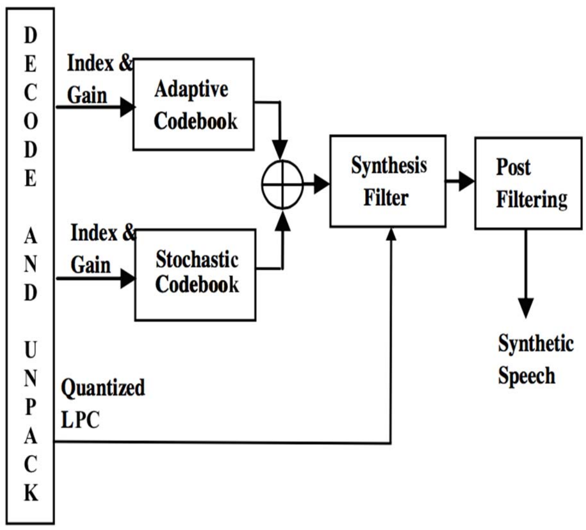

Fig.2: The CELP Decoder Block Diagram

III. CODEBOOK GENERATION

In this work two types of codebook were generated and was used in the G723.1 CELP codec. A Gaussian codebook was generated using the basic K-means algorithm [3]. Next two fixed codebooks having smaller size was generated using the PSVQ [3] method. Various speech samples were used for extracting the feature vectors that are used for generating the codebooks. All the operations were done directly on the LPC coefficients that were extracted from the speech samples.

Codebook generation is the process of building the codebook that is used in the CELP coder. The basic principle behind codebook generation is vector quantization [2]. In VQ the mapping of an N- dimensional continuous amplitude real valued vector $x = [x_1, x_2, ...x_N]$ to another real valued discrete amplitude vector $yi = [y_{i1}, y_{i2}, ...y_{iN}]$ is done. Then all possible values of the vector $y_i$is made into a codebook $Y = \{y_1, y_2....y_L\}$. If N is selected as the size of the codebook, it is a N-level codebook or N-level quantizer. For designing such a codebook, the M-dimensional space of the random vector x is partitioned into N regions or cells C_i and associate with each cell C_i a vector y_i. The quantizer then assigns the code vector y_i if x is in C_i.

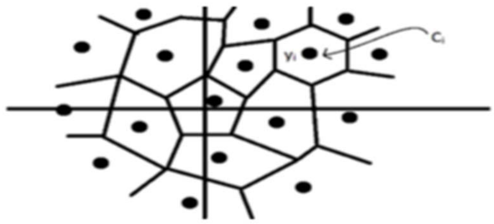

Fig.3: Partitioning the input space into different cells

A Gaussian Codebook

A Gaussian codebook having a size of 1024 having code vectors of length 40 samples each is generated using the K-means method. The K-means method is one of simplest methods of clustering available. The Linear combination of the coefficients yield a random vector that is Gaussian in nature. The various steps in the K-means algorithm is listed below:

- Elements of the initial codebook are selected randomly.
- Partition into nearest neighbor cells - Squared Distortion Criteria.
- Centroid calculation and codebook updation.
- Distortion sum calculation.
- If there is no change then stop, otherwise go to second step.

We can define the centroid cent(R_o), of any nonempty set $R_o \in R^M$ as the vector $y_o$that minimizes the distortion between a pointX $\in R_o$ and y_o, averaged over the probability distribution of X given X $\in R_o$.

The centroid [3] is calculated as follows

$$Cent(R_0)=\{y_0:E\{d(X,y_0)|X\in R_0\}\leq E\{d(X,y)|X\in R_0\}\} \quad (3)$$

For a given partition, $\{R_i; i=1,2,...,N\}$, the optimal code vectors satisfy

$$Y_i = Cent(R_i) \quad (4)$$

A widely used distortion measure is the squared error which is given by

$$d(x,x_1) = ||x-x_1||^2 = \sum_{j=1}^{M}(xj - x1j)2 \quad (5)$$

The computational cost [3] for training an M dimensional vector is given by,

$$C_T=MLNI \quad (6)$$

The storage cost [3] is given by,

$$M_T=ML \quad (7)$$

Where,

N- Size of the codebook
L-Number of training vectors
I- Number of iterations

B. Fixed Codebook

Instead of one Gaussian codebook two fixed codebooks having each half the length of the Gaussian codebook was generated in the second stage. The fixed codebooks have vectors of length 20 samples. The size of the codebook is same as before i.e., 1024. The first codebook is used for coding the narrow band speech signals and the second codebook is used for the wide band speech signals.

The method used for generating the fixed codebook is the Predictive Split Vector Quantization (PSVQ) method. In this method the higher dimensional vectors are spilt into smaller dimension and then quantization is performed. By splitting the higher dimensional vector into two or more sub vectors with lower dimension the storage and search complexity can be reduced compared to coding a higher dimensional vector. The vector space is structured in terms if two lower dimensional vectors with dimensions N_1 and N_2. These subspaces are the quantized with codebooks of size L_1 and L_2. The codebook indices are selected such that the concatenation of the two smaller vectors will yield vectors with the smallest possible distortion compared to the input speech vectors.

The memory cost [3] associated with storage of split vector quantization would be

$$M_T=N_1L_1+N_2L_2 \quad (8)$$

The computational cost [3] associated with usage of independent split vector quantization if I_i number of iterations are required to quantize each sub vector is,

$$C_T = M\,(N_1L_1I_1 + N_2L_2I_2) \quad (9)$$

IV.EXPERIMENTAL RESULTS

Both the Gaussian codebook and Fixed codebook were implemented in the G723.1 CELP codec and the results were observed. The reconstructed signals had very good quality compared to the original speech signal. Both these reconstructed signals were enhanced using Phase Spectrum compensation technique. Then objective quality measures were calculated for performance evaluation. The PSNR, PESQ, SNR, MSE of the signals were evaluated and tabulated. Perceptual Evaluation of Speech Quality (PESQ), is a family of standards comprising a test methodology for automated assessment of the speech quality as experienced by a user of a telephony system. The spectrogram of the different signals was plotted to observe the reconstruction efficiency. The fig 4 and 5 shows the waveform of the different signals that are generated along with the PESQ and SNR values.

The MSE [1] and PSNR [1] of the reconstructed signals when compared with the input signal is computed using the following equations,

The MSE is given by,

$$\text{MSE} = (1/XY)\sum_{i=0}^{X-1}\sum_{j=0}^{Y-1}[\text{I}(i,j) - \text{K}(i,j)]^2 \quad (10)$$

Where 'I' is the Input signal and 'K' is the Reconstructed signal.

The PSNR is given by,

$$\text{PSNR (dB)} = 10 \text{ Log}_{10}(\text{Max}^2\text{I/MSE}) \quad (11)$$

The different objective quality measures for speech are evaluated and is tabulated as shown below.

TABLE I. Experimental Results

Signal	PESQ	PSNR (dB)	SNR(dB)	MSE
Input Signal	4.5	-------	-------	-------
Gaussian CB-Signal	1.94	121.938	2.14	0.052425
Fixed CB-Signal	2.04	122.125	2.73	0.04818
Enhanced Gaussian CB-Signal	2.59	121.6459	5.04	0.054223
Enhanced Fixed CB-Signal	2.63	122.672	8.10	0.056954

The fig 4 shows the Waveform of the different signals that are generated along with the PESQ and SNR values.

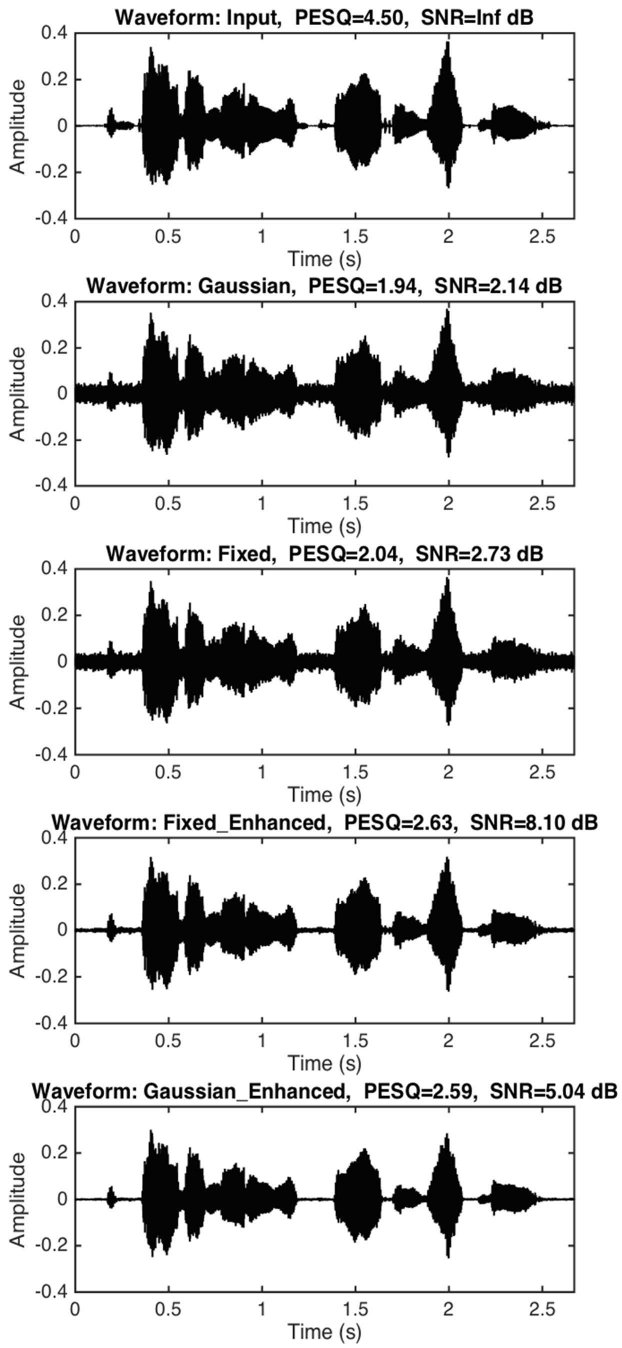

Fig.4: Different Speech Waveform Outputs

It can be seen that the Phase Spectrum compensation (PSC) enhanced version of the signal that has been reconstructed using the Fixed codebook closely matches the input signal. The SNR value and the PESQ value is better in the case of this signal as well. The MSE and PSNR values of both the Fixed codebook signals i.e. the PSC enhanced and the original reconstructed signal is very good compared to other signals.

The fig 5 shows the Spectrogram of the different signals that are generated along with the PESQ and SNR values.

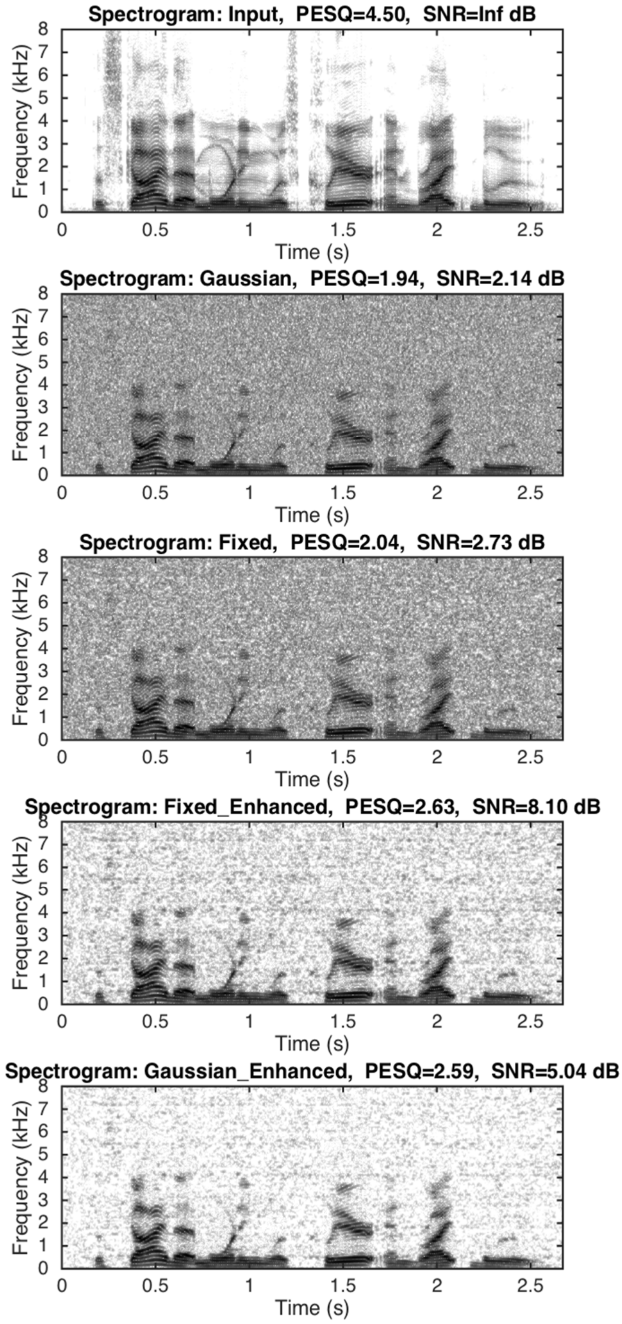

Fig 5.: Spectrogram of the different speech signals

V.CONCLUSION

The main aim of this work was to evaluate the performance of the CELP codec for two different codebooks. Two types of codebooks were generated, the Gaussian codebook and the Fixed codebook. The K-means method was used for generating the Gaussian codebook and the Fixed codebook was generated using the PSVQ method. The Gaussian codebook was a single codebook having a size of 1024 x 40. The Fixed code was actually two smaller codebooks having a size of 1024 x 20 each. Various speech samples were collected and the LPC coefficients were extracted from these speech samples. Further operations are directly done on these LPC coefficients to obtain the final codebook. The performance of the G723.1 CELP codec has been evaluated for these codebooks. Various objective quality measures were evaluated to analyze the performance of these two codebooks. Compared to the Gaussian codebook the Fixed codebook was observed to give improvement in performance.

REFERENCES

[1] Nasir Saleem, Usman Khan, Imad Ali, "Implementation of Low Complexity CELP Coder and Performance Evaluation in terms of Speech Quality", International Journal of Computer Applications, pp. 12-16, 2012

[2] Wai C. Chu, “Speech Coding Algorithms: Foundation and Evolution of Standardized Coders”. Wiley Interscience, 2003.

[3] Kavya Manohar, Premanand B, "Comparative Study on Vector Quantization Codebook Generation Algorithms for Wideband Speech Coding" 2012 International Conference on Green Technologies, pp. 82-88, 2012.

[4] Ming Yang, “Low bit rate speech coding”, IEEE Potentials, Volume:23, pp.32 - 36, 2004.

[5] Andreas S. Spanias, “Speech Coding - A tutorial review”, Proceedings of the IEEE, Vol. 82, No. 10, October 1994

[6] John Makhoul, Herbert Gish,“Vector Quantization in speech coding”, Procedings of the IEEE, Vol 73, 1985.

[7] M. K. Schroeder, B. S. Atal, "Code Excited LinearPrediction (CELP) Quality Speech at Very Low BitRates", ICASSP 1985, pp 25.I.1-25.I.4, 1985.

[8] Robert Gray et al. , “Vector Quantization of Speech and Speech-like Waveforms ”, IEEE transactions on acoustics, speech, and signal processing, vol. Assp-30, no, 3, june 1982