Implementation of Low Complexity CELP Coder and Performance Evaluation in terms of Speech Quality

Nasir Saleem
Institute of Engineering and Technology, Gomal University, D.I.Khan, KPK, Pakistan

Usman Khan
University of Engineering and Technology, Peshawar, KPK, Pakistan

Imad Ali
University of Engineering and Technology, Peshawar, KPK, Pakistan

ABSTRACT
The critical issues that are serving as constraints in wireless communication particularly in mobile communication are bandwidth, storage memory and power. The speech transmission in wireless networks is associated with the reduction of extra information present in signal in such a way to preserve the quality and intelligibility of speech. To remove the redundancy and transmit the speech with acceptable quality, speech compression algorithms are deployed. Because of this reason the speech coding is and will be the most important research issue. This paper addresses the implementation of CELP coder having low computational complexity with acceptable speech quality and preserves the intelligibility. The coder is assessed in terms of quality for different kinds of speakers using PESQ, PSNR,Frequency Weighted SNRseg, and SNRseg.

Keywords
CELP, LPC, Speech Coder, PESQ, PSNR, FwSNRseg, SNRseg.

1. INTRODUCTION
Low bit rate speech codecs [1][2] are applied to many applications including mobile radio communication in order to provide the opportunity to transmit the digital speech over analog channels. So it is required to bring high quality speech down to low rates for better storage and transmission purposes. Code Excited Linear Prediction (CELP) [3][4][5][6] is one of the most efficient speech coding algorithms where the speech is compressed with rate of 4.8 kbps by preserving quality of speech. CELP coder provides the bridge among waveform coders and vocoders as it presents compression of speech comparable to medium bit rate waveform coders. Basically the CELP is Analysis-by-Synthesis (AbS) sort of algorithm where the excitation signal is selected from the Closed loop search method and this excitation is than introduced to the synthesis filter. The synthesized waveform is judged against to novel segment and this progression is repetitive for all excitation code vectors present in code book. The index for the best excitation is selected and transmitted to decoder where it retrieves the excitation vector from code book and synthesized the speech. The name code excited is derived from the fact that code book contains the codes which excites the synthesis filter. Figure 1 symbolizes CELP speech generation representation. The CELP coder mainly depends upon Short and long term linear prediction models. The function of the "pitch synthesis filter" is to create periodicity in speech signal that is linked with the fundamental pitch frequency where the formant synthesis filter produces the spectral envelope. The post filtering enhances the synthesized speech. The CELP codebooks are either adaptive or fixed and hold deterministic pulses or random noise.

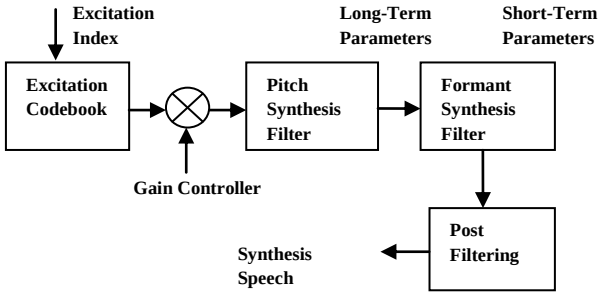

Fig.1.1: CELP Speech Generation

As the synthesized speech is used for the analysis principle, therefore, CELP is called as AbS. All the parameters are optimized together in order to obtain better results. But this process required more computation and is theoretical in nature. In practical manners, sub-set of parameters are preferred for closed loop optimization where the rest of parameters are obtained by open loop analysis. The parameters updating process is frequent in order to get best matching value, so all these process are held in chunks.

2. CELP CODER
TheCELP coder is employed to accomplish best quality speech with low computation complexity at 4.8kbps rate. In this execution several major techniques are utilized to guarantee low computational complexity that makes it appropriate for Digital Signal Processor real time processing. For low computation, the sub-routines for quantization of coefficients are not written. Similarly the computations are directly carried out on LPC coefficients. Delta search technique is used to minimize the computation.

2.1 The CELP Encoder
The comprehensive block diagram for CELP coder is sketched in figure 3. (The RED blocks are not implemented in order to reduce complexity. we have used LPC directly). The segmentation of the input speech is carried out, speech is cracked to 30 msec frame contains 240 samples and these 30 msec frames are sub-divided into sub-frames having length 7.7 msec having 60 samples in each sub-frame. After this stage only sub-frames are required for each upcoming block. The reason of sub-framing is the non-stationary and quasi-periodic nature of speech. At sub-frame level the analysis become easier. The short-term linear prediction analysis is executed on each sub-frame to generate 10 LP coefficients so total 40 LP coefficients are generated at this stage. After that Long-term linear prediction analysis is performed on each sub-frame. The LP coefficients show the spectral shape of the input speech. Autocorrelation is applied to acquire the LP

coefficients and afterward applying Livenson-Durbin algorithm to autocorrelation values for LPC. The autocorrelation function is applied to speech is:

$$\text{Arc (l)} = \sum_{i=0}^{N-l-1} s(i) * s(i - l) \qquad (2.1)$$

The resultant LP coefficients are given as:

$$LP_{A(z)} = 1 - \sum_{i=1}^{k} a_i z^{-1} \qquad (2.2)$$

With the help of bandwidth expansion, all these coefficients are expanded with factor γ as $a_i = \gamma_i a_i$. With this process the poles of the filter will shift towards the center in Z-plane assuring that system is stable. The value for $\gamma = 0.994$. After LP analysis (Short and Long term) on speech frames, only random signals and periodic pitch information has been left. At this stage the coefficients of other filters including PWF, pitch synthesis filter and formants synthesis filters are computed

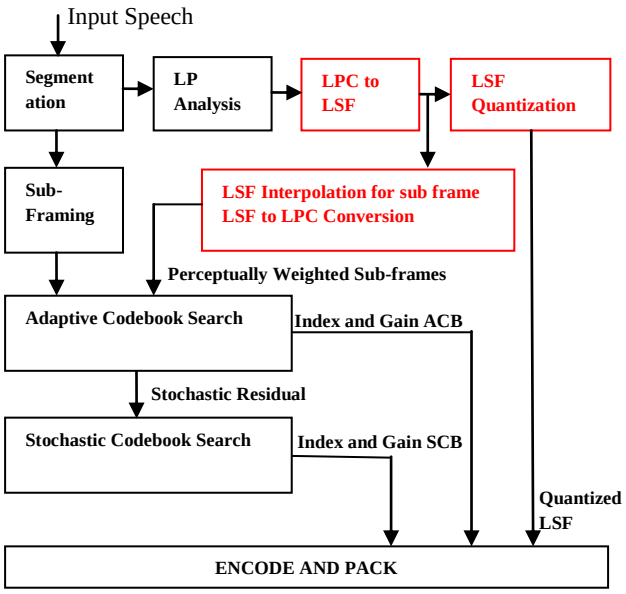

Fig.2.: The CELP Coder Block Diagram

The excitation sequence search is executed in order to get pitch information. Adaptive and stochastic code book search techniques are used. The key principle behind the adaptive codebook search is to eliminate pitch information form pitch residual. The codebook contains the 256 code words. For thepurpose of finest matching, the residual is compared to the code-words. These 256 codewordsare upgraded for each sub-frame. It contains 128-integers and 128 no integer delaysvarying from 20 to 127 samples. This delay in samples with respect to time is known as pitchdelay. These delays are useful in indexing the adaptive code-words. The range 20 to 127 isselected corresponds to 54 Hz to 400 Hz respectively which is typical range for pitch information in all kind of speakers. The adaptive code-book (ACB) is alinear vectorfor overlapped code-words. For the 1st sub-frame, the pitch search is not carriedout because at this stage the code-book vector is vacant having no entry. In order to createadaptive code-book Vector, the excitation vector of the first sub-frame is used. For the creation of the first integer code-word (60 samples), the first 20 samples are recurring thrice.The 20 samples are selected which corresponds to delay of 20. 21 samples are using till thenext code-word formed and the samples are recurring till the size of sub-frame reached. Thestochastic search procedure is very similar to the adaptive exploration scheme. Theperceptually weighted LP synthesis filter weights the code-words. As the pitch informationhas been subtracted by adaptive code-book searching from the input sub-frame to create theresidual. Those residuals are then convolved with the filtered code-words. At this stage theenergy is also calculated for the code-word. The convolution divided by the energy results inthe gain. The match score for the specific code-word is calculated by the multiplication ofgain with the convolution of residuals and filtered code-word. By this method the matchscore is calculated for all 512 code-words as well as for the gains. The highest match is tracedwith gain and is transmitted.

2.2 The CELP Decoder

The CELPdecoderdecodes the parameters and is sketched in figure 4. (RED blocks are not implemented) Thequantized LSP are interpolated and retransformed to the LPC coefficients. Same versions ofthe stochastic code-word are present on the receiving side as well and the index number isused to select the stochastic code-word. Same process is carried out for the adaptive codebook investigation. The weighted code-words are filtered with the help of LPsynthesis filterto regenerate the synthetic speechsignals. In order to minimize thequantization noise, post-filtering is used.

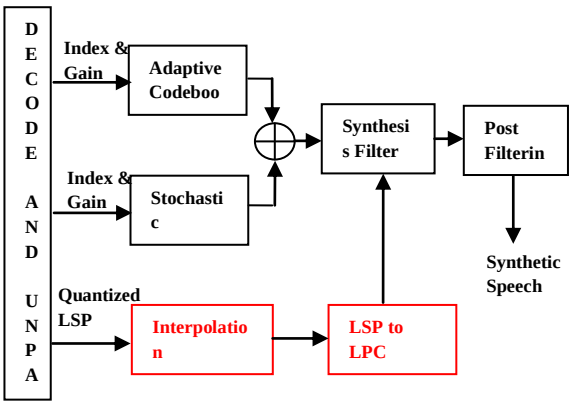

Fig. 2.2: The CELP Decoder Block Diagram

2.2.1 The Post Filtering

The aim of the post-filtering is to minimize quantization distortion in regenerated syntheticspeech as well as to improvethe quality.The postfiltering procedure utilizes similar LP parameters in current sub-frame. Transfer function for postfilter is:

$$H(z) = LP_{A(z/\beta)} / LP_{A(z/\alpha)} \qquad (2.3)$$

$$LP_{A(z)} = 1 - \sum_{i=1}^{k} a_i z^{-1} \qquad (2.4)$$

The post-filter achieved the noise cutback by repressing the noise in the region of the valleys and sharpening the formants peaks.Because of this property of the post-filter, the peaks sound more loudly as compared valleysleading the suppressed noise.

3 MATLAB IMPLEMENTATION

Open loop linear prediction is applied in order to extract the vocal tract parameters from input speech. The residual signal is acquired that is error signal extracted by subtracting the vocal tract envelope from the input speech. The pitch search routine investigates the error (residual) signal. The extracted residual signal is perceptually weighted and judged against all the code words in codebook which is basically the adaptive

codebook or adaptive pitch buffer. Because of the overlapping of the codebook, end point correction (EPC) is applied in order to minimize the working out. As the pitch information are short-term stationary, delta search technique is applied so that the computations are made simpler. In primary sub-frame complete search is executed while the rest of 3 sub-frames, searching are carried out only in neighboring 64 pitch delay of last delay. For high pitch speakers, the pitch prediction resolution must be increased so the sampling rate must also be increased with following equation:

$$S(t) = \sum_{-\infty}^{+\infty} x(n)\text{sinc}[(t - nT)/T] \qquad (3.1)$$

Sinc function is hamming windowed in order to trim down the band limited signal and further minimize the aliasing effect.

3.1 Review of Computation Complexity in Pitch Searching

For prime code, The Convolution we need:
1+2+3+.....+10+10+....+10 = 555 MUL [60 Terms]
1+2+.....+9+9+.....+9 = 495 ADD [59 Terms]
For the upcoming codes we need only 9 MUL and 9 ADD which is significance of EPC. Therefore entirely we required following operations for 128 integer pitch search.
555+495+18*127 = 3336
For every Correlation, there is requirement of 60 MUL and 59 ADD so overall operations for 128 correlations are:
(60 + 59)*128 = 15230.
At this stage energy is calculated which is equal to the correlation operations.
Energy = 15230 Operations
All these operations are carried out per one sub-frame that is for 60 samples (7.5 msec), therefore the millions of instructions per second (MIPS) we need are:
(3336+15230+15230) / 7.5msec = 4.5 MIPS
Because of the use of delta search method the MIPS minimizes to the 3 MIPS.

3.2 Codebook Searching

The sub-frame is divided into three sub-vectors which signify 20*3. There are four non-zero entities in form of [-1 or +1]. These four entities are equivalently scattered and gives a looks like Y0000Y0000Y0000Y0000, in this representation the Y represents -1 or +1. To facilitate the codebook with 512entities, there is need of implementing the restriction on combination of -1 or +1. As we are allowing four 1's and four -1's equivalent to 1 combination and two 1's and two -1's (6 combinations) for every vector. Thus every sub-vector contains 8 combinations of signs ($8^3 = 2^9$).

3.3 Review of Computation Complexity in Codebook Searching

The convolution:
(Impulse responses * Y0000Y0000Y0000Y0000) = H1H2H3H4H5H1H2H3H4H5..... (60 H_i)
For computing of correlation of sub-vectors:
5*12= 60 MUL
4*12=48 ADD
2*512= 1024 ADD to calculate 512 inside products.
1228 operations for 512 inside products, thus 1228/7.5msec = 0.16 MIPS.

4 EXPERIMENTAL SETUP

4.1 Selection of Test Signals

Total 10 test signals are used in experiment and are divided into two main groups, the male speakers and female speakers. The signals are the mixer of voiced and unvoiced sounds. The detailed information about test signals is given in table 1.

Table 1: Detail Information of Speech Signals

Speech Types	Place of Recording	Sampling Rate	No. of samples	Speaker
Mixed Long	Laboratory	8000	21209	Male
Mixed Long	Laboratory	8000	27768	Male
Mixed Long	Laboratory	8000	28064	Male
Mixed Short	Laboratory	8000	5186	Male
Mixed Short	Laboratory	8000	5488	Male
Mixed Long	Laboratory	8000	21338	Female
Mixed Long	Laboratory	8000	18269	Female
Mixed Long	Laboratory	8000	19231	Female
Mixed Long	Laboratory	8000	23323	Female
Mixed Long	Laboratory	8000	23466	Female

4.2 Waveform Analysis

The waveforms for both male and female speakers are shown in figure 4.1 and 4.2 respectively.

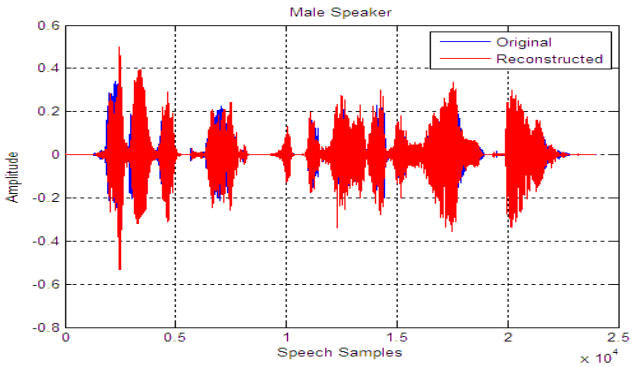

Fig.4.1: Male Speaker original and reconstructed waveforms

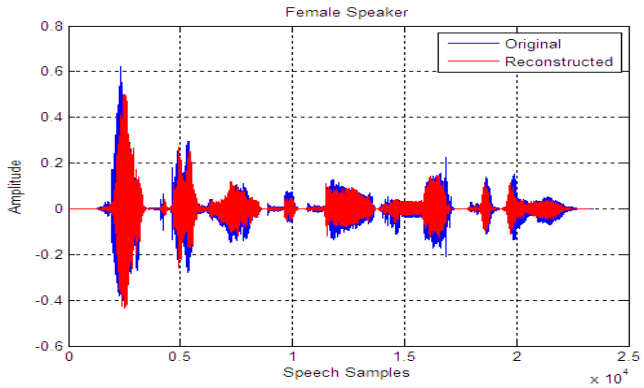

Fig.4.2: Female Speaker original and reconstructed waveforms

4.3 Objective Measurements

Objective evaluation techniques are put into operation to evaluate the compressed speech quality. PESQ [7][8], PSNR [9], SSNR [7][8], and FWSSNR [7][8] are computed with following formulas:

$$PESQ = A_o + A_1 Dave + A_2 Aave \quad (4.1)$$

$$PSNR = MSE = (1/XY) \sum_{i=0}^{x-1} . \sum_{j=0}^{y-1} .[I(i, j) - K(i, j)] \quad (4.2)$$

$$PSNR\ (dB) = 10\ Log_{10}\ (Max^2 I/MSE) \quad (4.3)$$

$$PSNR\ (dB) = 20\ Log_{10}\ (Max\ I) - 10\ Log_{10}\ (MSE) \quad (4.4)$$

$$SNRseg = (10/M) \sum_{M=0}^{M-1} . Log_{10} \sum_{i=Nm}^{Nm+N-1} . \left[\frac{\sum_{i=1}^{N} x2(i)}{\sum_{i=1}^{N} (x(i)-y(i))2}\right] \quad (4.5)$$

$$FWSNR = \frac{10}{M} \sum_{M=0}^{M-1} . \frac{\sum_{j=1}^{k} Bj\ Log\left[\frac{F2(m,j)}{(F(m,j) - F'(m,j))2}\right]}{\sum_{j=1}^{k} Bj} \quad (4.6)$$

4.3.1 PESQ

The Figure 4.3 represents that speech quality and intelligibility is degraded greatly in case of female speakers. While for the male speakers, PESQ values are close to theoretical range.

4.3.2 PSNR

TheFig. 4.4 represents that PSNR in dB, are higher for male speakers and vice versa. High PSNR value means better quality so male speakers have achieved better PSNR results.

4.3.3 FWSSNR

Fig. 4.5 shows the Frequency Weighted segmental SNR, represents that greater the computed value higher will be speech quality so from above computations it is concluded that male speakers have higher FWSSNR values as compared to female speakers.

4.3.4 SNRseg

TheFig. 4.6 represents the Segmental SNR in dB; the values are in negative region which shows more degradation. But among speakers, the SNRseg shows mix results for both kinds of speakers.

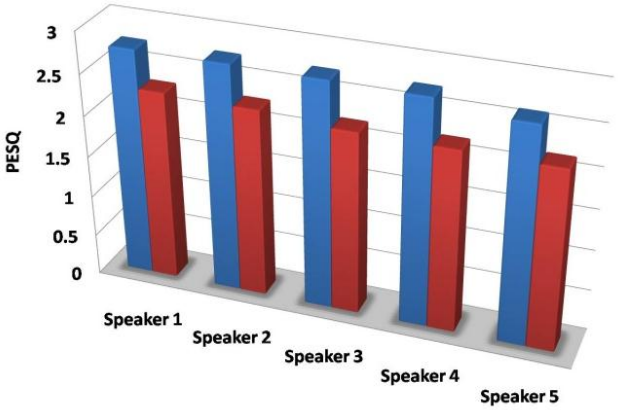

Fig. 4.3: PESQ: Male (Blue) and Female (Red) speakers

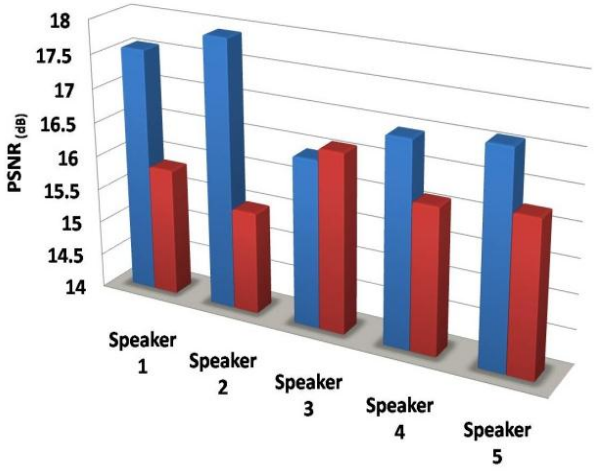

Fig.4 4: PSNR: Male (Blue) and Female (Red) Speakers

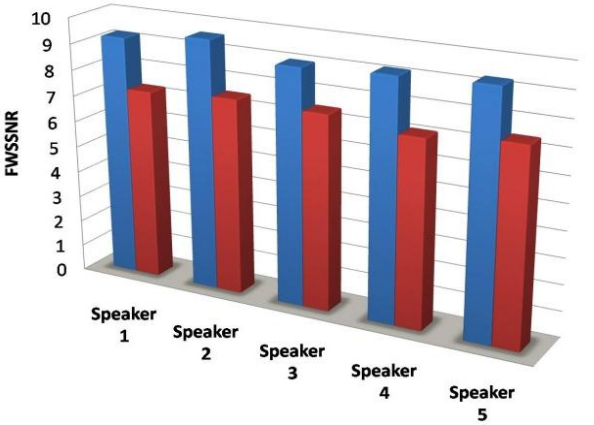

Fig.4.5: FWSSNR: Male (Blue) and Female (Red) speakers

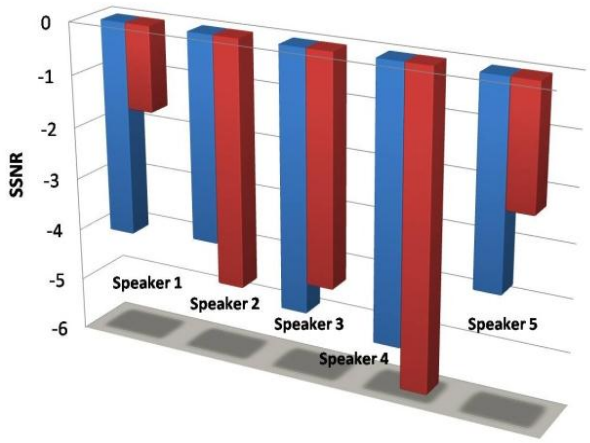

Fig.4.6: SNRseg: Male (Blue) and Female (Red) speakers

5 CONCLUSION AND FUTURE WORK

The CELP Coderis implemented in MATLAB (R2010a) and is investigated with a variety of objective measuring tools. The outcomes are reasonably close to theoretical assessments. The coder performed well for the male speakers but somewhat lower performance for female speakers. From waveforms of original and reconstructed speech sentences it can be concluded that synthesized speech is nearly the replica of novel speech. As there is tradeoff between quality and complexity, so we reduced the complexity but still the quality evaluation shows that quality is well within range. In future if we improve pitch prediction resolution for elevated pitch speakers, the coder will perform better for female speakers as well but overall computations will also be increased.

6. REFERENCES

[1] Ming Yang, "Low bit rate speech coding", IEEE, 2004, Volume:23, p32 - p36.

[2] Vuppala, A.K, Yadav, J, Chakrabarti.S, Rao, K.S, "Effect of Low Bit Rate Speech Coding on Epoch Extraction", (ICDeCom), 2011, p1 - p4.

[3] 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.

[4] Wai C. Cu, ''Speech Coding Algorithms Foundation and Evolutionof Standardized Coder'', John Wiley & sons, 2003 Edition.

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

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

[7] Jianfen Ma, Yi Hu, Philipos C. Loizou,"Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions", Acoustical Society of America, 2009, p3387 p3405.

[8] K. Kondo,"Subjective Quality Measurement of Speech", Signals and Communication Technology, Chapter 2, p8-p11.

[9] Dr. D.C. Dhubkarya, SonamDubey, "High Quality audio coding at low bit rate using wavelet and wavelet packet transform", Journal of theoretical and applied information technology, 2005-2009.