Study and Design of VoIP using G.729 Algorithm

Prof. Abhijit Maidamwar[1] Dr. Vivek Kapur[2]
[1]Assistant Professor [2]Professor
[1,2]Department of Electronics Telecommunication Engineering
[1,2]RGCER, Nagpur, Maharashtra 440002, India

Abstract— The algorithm described below is the 8 kb/s speech coding algorithm. The algorithm is based on a CS-ACELP (Conjugate-Structure Algebraic Code Excited Linear Prediction) coding technique. Short term synthesis filter is based on a 10th order Linear Prediction filter for every 10ms frame. This coder is designed to operate with a digital signal obtained by first performing telephone bandwidth filtering of the analogue input signal, then sampling it at 8000 Hz, followed by conversion to 16-bit linear PCM for the input to the encoder. The output of the decoder should be converted back to an analogue signal by similar means. The idea of the algorithm is to predict the next coming signals by means of linear prediction. It also uses statistical data to distinguish the resemblance of the signal to special signals in its codebook. It uses the fixed codebook and adaptive codebook. Long-term pitch synthesis filter is implemented using the adaptive-code book. The codec delivers quality of speech equivalent to that of 32 kb/s ADPCM (Adaptive Delta Pulse Code Modulation) for most operating conditions. This algorithm gives the reduced bit rates at the cost of lower quality. The coder structure will be discussed in detail and also the reasons behind certain design choices.

Key words: G.729, CS-ACELP

I. INTRODUCTION

Based on model parameter of human voice, CELP voice codec is used to compress the speech signal. It is a ITU-T standardized codec which works at 8 kbps. The speech quality produced by this coder is equivalent to that of a 32 kbps ADPCM under most operating conditions. Applications for this type of coder are personal communication systems, digital satellite systems and other applications such as packetized speech and circuit multiplexing equipment. The coder operates on frames of 10ms. The speech signal is analysed and the parameters of the CELP model are extracted ,i.e., linear prediction filter coefficients, adaptive and fixed codebook index .These parameters are then encoded and transmitted.

The coder is designed to operate with an appropriately band-limited signal which is sampled at 8000Hz.Figure shows the encoder and the processes involved while encoding.

II. ENCODER

A. Preprocessing

Preprocessing stage contain two steps: first scaling and second high pass filtering. The input sample to speech encoder is of 16 bit PCM. The input signal undergoes scaling and high-pass filtering process. Scaling means dividing the input by factor of two to avoid the possibility of overflow in fixed point implementation. The filtering process uses a second-order pole/zero high-pass filter with a cut-off frequency of 140 Hz. This high-pass filter precludes undesired low-frequency or DC components. The filtered and scaled signal is referred as s(n) and will be used in all subsequent encoder operations.

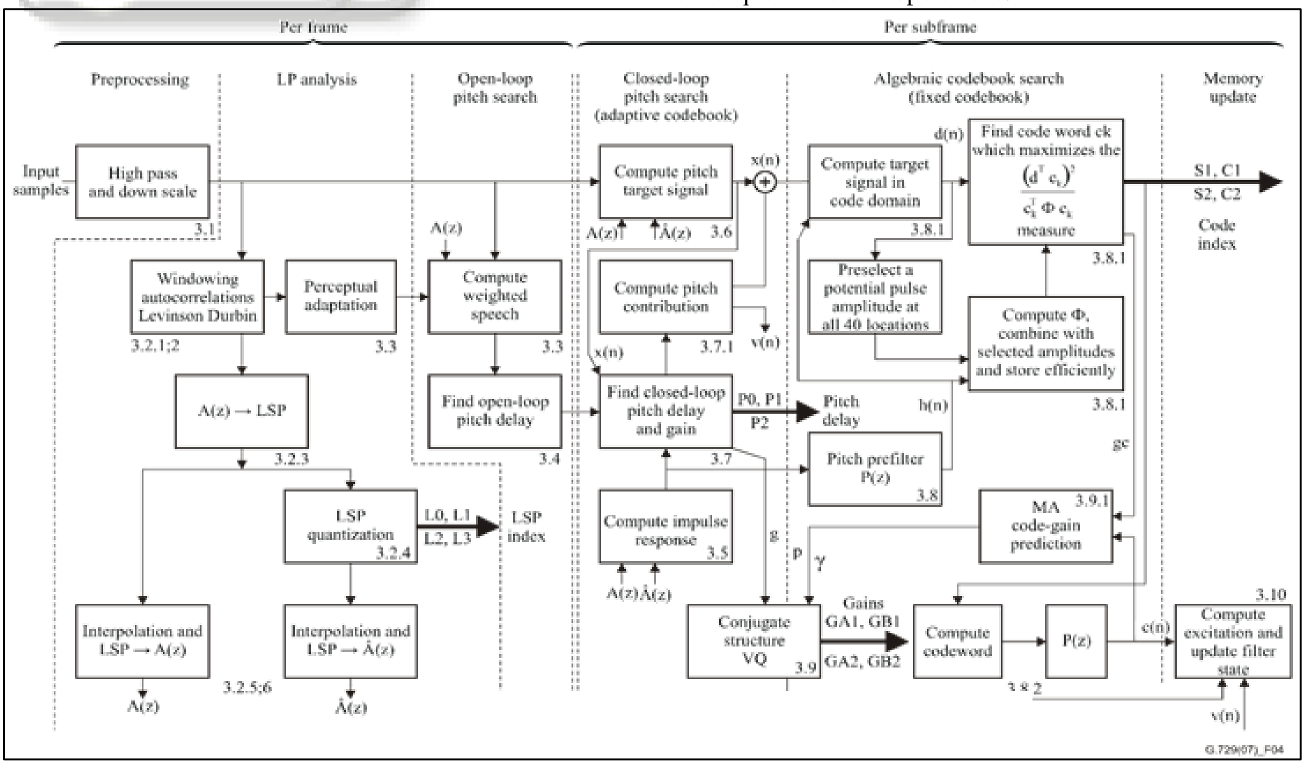

Fig. 1: Block diagram of CS-ACELP encoder

B. *LP Analysis and Quantization*

In LP analysis, the signal obtained after preprocessing has to be windowed using a asymmetric window that consists of half a hamming window and quarter of a cosine cycle. The window is of total duration 30ms and consists of total 240 samples out of which 120 samples are from the past speech frame, 80 samples from the present speech frame and 40 samples from the future speech frame. Every 80 samples, the autocorrelation coefficients of windowed speech are computed. The autocorrelation coefficients are used for the computation of LP coefficients using Levinson-Durbin algorithm. These LP coefficients are then further converted into LSP (linear spectral pair) which then undergo quantization and interpolation. The use of an asymmetrical window allows reduction in the look-ahead compromising quality. The interpolated quantized and unquantised LSF coefficients are converted back to LP coefficient are then further used to construct the synthesis and weighting filters for each subframe.

C. *Perceptual Weighting*

The weighted speech signal in a subframe is obtained by filtering the speech through a perceptual weighting filter. The use of the unquantized coefficients gives a weighting filter that matches better the original spectrum. For signals with a lot of low frequency energy, the amount of weighting is increased. An open-loop pitch delay Top is estimated once per 10 ms frame using the perceptually weighted speech signal Sw(n).For good performance of the CELP algorithm either a closed or a open pitch loop is essential at intermediate bit rates.

D. *Pitch Analysis*

The periodic component in the excitation signal is represented using adaptive codebook approach. For each sub frame the target signal x (n) and the impulse response of the weighted synthesis filter are computed. In first sub frame closed adaptive codebook search is performed around the index corresponding to the open-loop pitch lag estimate (±3). The first subframe of the adaptive codebook index is encoded with 8 bit. A fractional range of 1/3 is used as a sample resolution for the second subframe. The integer part of adaptive codebook lags in the first subframe straddles the boundaries of lag range is encoded with 5 bit differentiation. No noticeable degradation in the speech quality is introduced by differential coding as the open-loop pitch estimates provides pitch tracking. In open-loop pitch lag estimation the weighted speech signals are used. Three maxima of correlations are found in range of (80-143), (40-79) and (20-39) which are the values from previous frames. The maxima which has the lower range values is selected among the three normalized correlations by weighing them to the longer lag values.

E. *Algebraic Codebook Search*

1) *The Adaptive Codebook*

The adaptive codebook parameters are the indices corresponding to lag and gain. The excitation is repeated for lags less than the sub frame length to implement the pitch filter in adaptive codebook approach. This process is simpler as it produces almost same results when compare to using the adaptive codebook for complete sub frame but it is computationally costly due to use of fractional lags during the search stage.

2) *The Fixed Codebook*

Fixed codebook is basically a 17 bit codebook here codebook vectors are determined using simpler algebra from the transmitted index rather than look up tables. The advantages of this structure are storage search complexity and robustness. The four nonzero pulses in each fixed codebook vector can assume the amplitudes and position are encoded separately using bit allocation. To evaluate all pulse position, a total of 2^{13} = 8192 combinations should be examined. Search is mainly focused on those combinations that are potentially a good match.

F. *Memory Update*

To compute the target signal in the next sub frame an update of the states of the synthesis filter and weighing filter is needed which can be performed by filtering the signals through the filter for the 40 sample subframe and saving the state of filters. The states of Synthesis filter and weighing filters are updated separately. Thus memory update is performed.

III. DECODER

The LP filter coefficients are generated in the decoding process from the transmitted information with the same producers as used in the encoder. The adaptive and fixed codebook vectors for each subframe are scaled and they are summed and filtered with the LP synthesis filter to generate the reproduced speech signal. Post filtering consists of adaptive post filtering and high-pass filtering. The adaptive post filter contain three filters and they are

- A long-term post-filter
- A short-term post-filter
- A tilt compensation filter

In post filter coefficients the speech which is reconstructed is filtered to produce the residual signal. This signal is used to compute the lag and gain of the long term post filter. The signal is then filtered through the long term post filter and the synthesis filter the output signal of the synthesis filter is given to the tilt compensation filter. To match the energy adaptive gain control is applied and the resulting signals are filtered through 100Hz high-pass filter and multiplied by 2 to produce the output signal of the decoder.

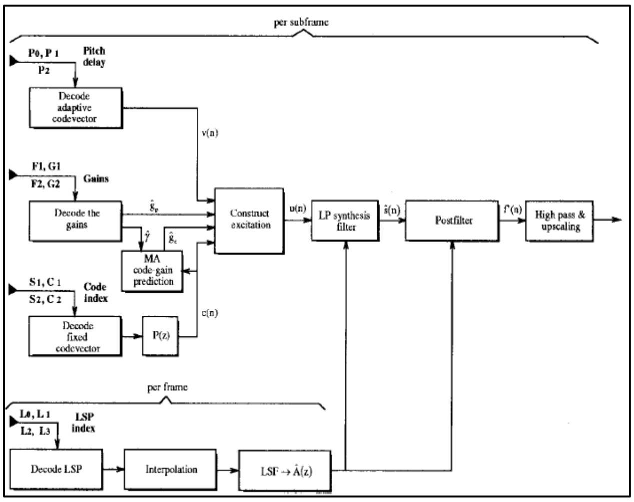

Fig. 2: Block diagram of CS-ACELP decoder

IV. ADVANTAGES

A parity bit P0 is computed to make the coder more robust against bit error on the six most significant bits of the lag index P1 of the first sub frame. This parity bit is recomputed at the decoder and if the recomputed value does not agree with the transmitted value, an error concealment procedure is applied. The indices of the LSF quantizer have been mapped such that single bit errors map to the most likely neighbouring vector. This is achieved by rearranging the order of the codebook vectors. A similar mapping is carried out for each codebook of the conjugate-structure gain quantizer. Since the element of this quantizer need to be ordered in value to allow an efficient search a separate table is used for mapping. The fixed codebook index describes the pulse signs and pulse position. Only the location of one single pulse is affected by single bit error leaving the remaining pulse positions intact. The use of Gray coding for the individual pulse position further minimized the effect of single bit errors. As Gray coding was done in the modified version of G.729 in the enhanced full rate coder for the U.S. cellular TDMA system. Although the coder uses prediction to quantize efficiently the spectral and gain parameters, these predictors are based on FIR filters such that propagation of channel error is controlled.

A. Concealment of Frame Erasures

To reduce the degradation in the reconstructed speech because of frame erasures in the bit-stream, an error concealment procedure has been incorporated in the decoder. This error concealment process is enabled when the frame of coder parameters has been identified as being erased. The concealment strategy reconstructs the current frame, based on the previously received information. This method replaces the missing excitation signal with one of the similar characteristics by using a voicing classifier based on the long-term prediction gain which is computed as part of the long term post-filter analysis. For the error concealment process, a 10ms frame is declared periodic if at least one 5ms sub frame has a long-term prediction gain of more than 3db. An erased frame inherits its class from the reconstructed frame. For an erased frame necessary steps required are as follows:

1) Repetition of the synthesis filter parameters

The synthesis filter in an erased frame uses the LP parameters of the last good frame. The memory of the MA LSF predictor contains the value of the received code-words. The codeword is computed from the repeated LSF parameters and predictive memory.

2) Attenuation of the memory of the gain predictor

The gain predictor uses the previously selected fixed codebook vectors. The memory of the gain predictor is updated with an attenuated version of the codebook energy, once the good frames are received in order to avoid transitional effects at the decoder.

3) Generation of the replacement excitation

The excitation used depends on the periodicity classification. If the last reconstructed frame was classified as periodic, the current frame is considered to be periodic as well. In that case only the adaptive codebook is used whose lag is based on the integer part of the lag in the previous frame and is repeated for each successive frame.

If the last reconstructed frame was classified as non-periodic, the current frame is considered to be non-periodic as well, and adaptive codebook contribution is zero. The fixed codebook contribution is generated by randomly selecting a codebook index and sign index. The use of different decay factors for either type of excitation is allowed by the use of classification. The classifier would declare that the frame non-periodic, and the slow decay avoids the muting effects that would otherwise introduce annoying interruptions, if the frame erasures occur when no speech is active.

B. Use of Non-equal errors Protection

For many wireless applications it is common to use the source coder in combination with channel coding. The speech coder bits into groups of similar sensitivity to channel errors are an efficient way to do this and to apply different amounts of protection to each class. Hence it is necessary to analyze the sensitivity of each bit. The sensitivity analysis was used to rank the bit in terms of sensitivity. It was found that the first stage of the LSF quantizer and the absolute adaptive codebook index information are the most sensitive. The differential index and the codebook gains are of medium sensitivity and the pulse positions and pulse signs are the least sensitive. a division into two books of equal size can be made by assigning the codebook indices C1 and C2, the codebook signs S1 and S2, L3 which is the high part of the second LSF codebook and the least significant bit of the differential lag P2 to the low sensitivity group whereas all other group belong to the high sensitivity group. Error correction for the bit is the most sensitive group. By using more powerful redundancy check on the protected bits further improvement can be achieved and to trigger the frame erasure concealment procedure if this check does not match the receiver.

C. Taming procedure

Most CELP coder allows adaptive codebook gain larger than one to allow for rapid built up of an excitation signal during speech one-set. Consequently the decoder will be locally unstable and the encoder controls this instability without producing undesirable artefacts due to the analysis-by-synthesis mechanism. Due to the mismatch between encoder and decoder, this local instability could potentially lead to disaster. However, the time-varying nature of speech will in general introduce enough leakage that no adverse effects are encountered. The adaptive codebook gain is typically close to one, and frame erasures combined with the leakage can result in annoying effect at the decoder for stationary periodic signal such as sin wave. Limiting the adaptive codebook to value less than one would seriously affect the performance at speech onset hence it was decided to “tame” the coder by limiting adaptive codebook gain only in cases where it was deemed necessary. The taming procedure, in essence keeps track of each sample in the adaptive codebook contribution in terms of previous excitation samples and the adaptive codebook gain is used.

V. PERFORMANCE

During the study period of ITU-T various versions of this codec was tested by different organization. These organizations are involved in several other speech coder evaluations and they are the members of ITU-T SG12. Due to international character of ITU-T, an advantage is that testing is possible for a variety of languages, such as American English, French, Italian, and Japanese.

Degradation and comparison category rating tests are conducted to examine the performance with background noise. The DCR and CCR are only for information because they are more sensitive than ACR to small extent. ITU-t is required so that the coder produce a quality better than 32kbps ADPCM this condition is referred in all tests.

VI. CONCLUSIONS

In this paper, thus we described G.729 algorithm which delivers a quality of 8 kbps. This encoder has been recommended by ITU-T. It is robust against the channel errors and thus it can be used in numerous applications.

ACKNOWLEDGMENT

The authors acknowledge Prof. Abhijit Maidamwar who guided this project to a completion.

REFERENCES

[1] Coding of Speech At 8 Kbit/S Using Cs-Acelp Recommendation Itu-T Dated 06/2012.
[2] Performance Analysis of Voip Codecs for H323. Published In 2012.
[3] Evaluation of Hardware/Software Partitioning Of Cs-Acelp Algorithm Using Matlab. Published In 2007 Icict.
[4] Itu-T, “Recommendation G.729— Coding of Speech At 8kbit/S using Conjugate – Structure Algebraic - Code – Excited-Linear - Prediction (Csacelp) ,” March, 2006.
[5] R. Salami Et Al., “Description of tyhe Proposed Itu-T 8 Kb/S Speech Coding Standard,” In Proc.Ieee Workshop on Speech Coding, 2006, Pp.3–4.