Reduction of Complexity for Estimating the Open Loop Pitch of the CS-ACELP Codec

Abhijit Maidamwar, D.Marotakar, Manisha Khorgade, Swati Sorte

Abstract— G.729 or Conjugate structure algebraic CELP is a audio voice codec that compresses speech signal based on model characteristics of human voice. This paper deals with the reduction of the computational complexity for estimating the open loop pitch of the CS-ACELP codec, described in ITU recommendation G.729. For reduction in computation of open loop pitch analysis using Matlab 7.4, the weighted delta-LSP function is used. This depth first tree search is also used in G.729 for reducing the search complexity with minimum effort. In experimental study of our paper we are showing the comparing graphical result of Open Loop Pitch in Matlab 7.4, we are trying to prove that our proposed method save the computational time for calculation of open loop pitch

Index Terms— Open loop pitch analysis of G.729, Graphical result of open loop pitch, A-CELP, bit allocation of 8 kbps in G.729

I. INTRODUCTION

The ITU-T standardized 8 kbits/s speech codec to operate with a discrete-time speech signal. G.729 provides coding of speech signals used in multimedia applications at 8 kbits/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP) [1][2]. The quality of this 8 kbits/s algorithm is equivalent to that of a 32 kbits/s ADPCM under most operating conditions. Typical input rates include mu-law or A-law **64** kbits/s PCM or 128 kbit/s linear PCM providing a compression ratio of 16:l. These coders are all based on a model of the human vocal system. In that model, the throat and mouth are modelled as a linear filter, and voice is generated by a periodic vibration of air exciting this filter. In the frequency domain, this implies that speech looks somewhat like a smooth response (called the envelope), modulated by a set of discrete frequency components. CELP coders all vary in the manner in which the excitation is specified, and the way in which the coefficients of the filter are represented. All of them generally break speech up into units called frames, which can be anywhere from 1ms to 100ms in duration. For each frame of speech, a set of parameters for the model are generated and sent to the decoder. This implies that the frame time represents a lower bound on the system delay; the encoder must wait for at least a frames worth of speech before it can even begin the encode process.

Manuscript received on May, 2013.

Abhijit Maidamwar, ETC Department, RTMNU University/ RGCER, Nagpur, India.

Swati Sorte, ETC Department , RTMNU University/RGCER, Nagpur , India.

D.Marotkar, ETC Department, RTMNU University/RGCER,Nagpur, India.

Manisha Khorgade, ETC Department, RTMNU University/ RGCER, Nagpur, India.

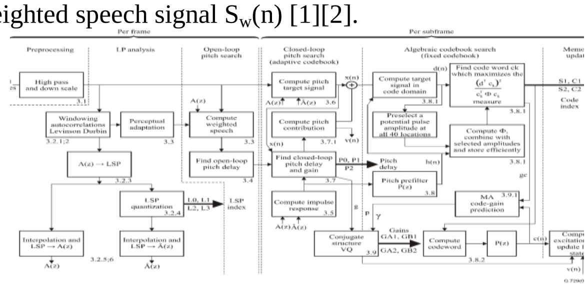
G.729 will be used for Voice over IP (VoIP), Videophones, Digital Satellite Systems, Integrated Services Digital Network (ISDN), Land-Digital Mobile Radio, Future Public Land Mobile Telecommunication Systems (FPLMTS), Digital Circuit Multiplication Equipment (DCME), Digital Simultaneous Voice and Data (DSVD), and other applications. This speech codec's relative low complexity makes it an attractive choice for Internet telephony.

For the real-time implementation of G.729 CS-ACELP codec, computational complexity is required to be small and its algorithmic delay has to be small. If we check the computations required in calculating the open loop pitch, it is found that more than 25 percent of the total time is consumed on open loop pitch analysis.

II. STUDY ON OPEN LOOP PITCH ANALYSIS

The encoding principle of G.729 is shown in Figure 1. The input signal is high-pass filtered and scaled in the pre-processing block. The pre-processed signal serves as the input signal for all subsequent analysis. LP analysis is done once per 10 ms frame to compute the LP filter coefficients. These coefficients are converted to Line Spectrum Pairs (LSP) and quantized using predictive two-stage Vector Quantization (VQ) with 18 bits [3][4]. The excitation signal is chosen by using an analysis-by-synthesis search procedure in which the error between the original and reconstructed speech is minimized according to a perceptually weighted distortion measure. This is done by filtering the error signal with a perceptual weighting filter, whose coefficients are derived from the unquantized LP filter. The amount of perceptual weighting is made adaptive to improve the performance for input signals with a flat frequency response.

The excitation parameters (fixed and adaptive codebook parameters) are determined per sub-frame of 5 ms **(**40 samples) each. The quantized and un-quantized LP filter coefficients are used for the second sub-frame, while in the first sub-frame interpolated LP filter coefficients are used (both quantized and un-quantized). An open-loop pitch delay T_{OP} is estimated once per 10 ms frame using the perceptually weighted speech signal $S_w(n)$ [1][2].



The open loop pitch lag estimation uses the weighted speech signal $S_w(n)$:

$$R(k) = \sum_{n=0}^{79} s_w(n)s_w(n-k)$$

The three maxima of the correlation are found in following three ranges; (20:39), (40:79), (80:143). The open loop pitch is obtained by taking the maxima of the three ranges from the normalized autocorrelation function.

$$R'(t_i) = \frac{R(t_i)}{\sqrt{\sum_n s_w^2(n-t_i)}}, \quad i = 1, \cdots, 3.$$

For one frame, total operations required are 10160 multiplications, 10033 additions, 123 comparisons, 3 radical and 3 division operations to estimate open loop pitch.

The computation of the pitch is dependent to the voice and unvoiced signal. The pitch contour lies in the voice signal only. The weighted delta-LSP function (Wd) is employed to differentiate between voice and unvoiced signal. The function Wd is given by:

$$Wd = \sum_{k=1}^{10} w_k * [LSP_i(k) - LSP_{i-1}(k)]^2$$

If the function value Wd is greater than some pre-defined threshold, the open loop pitch lag is estimated otherwise the pitch value is taken as same as that of previous frame. The $LSP_i(k)$ is the LSP coefficient of the k^{th} order at the i^{th} frame and w_k is the weighted coefficient [5]. Hence the calculation required in this automatically reduced.

III. DEPTH FIRST TREE APPROACH

G.729a is a compatible extension of G.729, but requires less computational power. This lower complexity, however, bears the cost of marginally reduced speech quality. It also contained the same delay. The fixed codebook search is simplified, it is not used nested loop focused search. Basically it has used depth first tree search. In this new approach a smaller number of pulse position combinations is tested and it has fixed complexity.

IV. BIT ALLOCATION OF THE 8 KBIT/S CS-ACELP ALGORITHM

The CS-ACELP coder is based on the code-excited linear prediction (CELP) coding model. The coder operates on speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8000 samples per second. For every 10 ms frame, the speech signal is analysed to extract the parameters of the CELP model (linear prediction filter coefficients, adaptive and fixed-codebook indices and gains). These parameters are encoded and transmitted. The bit allocation of the coder parameters is shown in Table 1. At the decoder, these parameters are used to retrieve the excitation and synthesis filter parameters. The speech is reconstructed by filtering this excitation through the short-term synthesis filter, as is shown in Figure 1. The short-term synthesis filter is based on a 10th order linear prediction (LP) filter. The long-term, or pitch synthesis filter is implemented using the so-called adaptive-codebook approach. After computing the reconstructed speech, it is further enhanced by a postfilter.

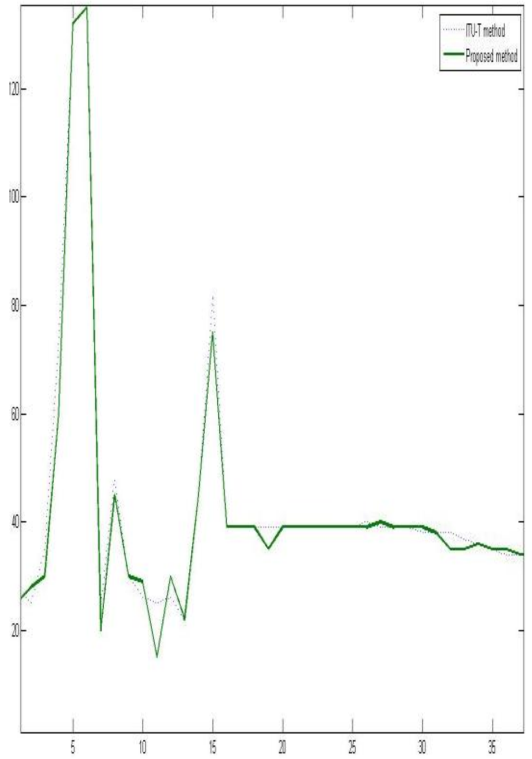
Table 1 – Bit allocation of the 8 kbit/s CS-ACELP algorithm (10 ms frame)

Parameter	Codeword	Subframe 1	Subframe 2	Total per frame
Line spectrum pairs	*L0, L1, L2, L3*			18
Adaptive-codebook delay	*P1, P2*	8	5	13
Pitch-delay parity	*P0*	1		1
Fixed-codebook index	*C1, C2*	13	13	26
Fixed-codebook sign	*S1, S2*	4	4	8
Codebook gains (stage 1)	*GA1, GA2*	3	3	6
Codebook gains (stage 2)	*GB1, GB2*	4	4	8
Total				80

V. CONCLUSIONS AND SIMULATION RESULT

The algorithm used in this paper was developed to reduce the computation required in finding the open loop pitch. The open loop pitch calculation time is saved by more than 80 percent. Hence the total computational saving is around 20-23 percent. This algorithm makes less computation and takes less time and hence this makes more practical in implementing G.729 for real-time systems. The ITU-T method and my my proposed method compared result is shown in figure 2.The graphical analysis has been perform in Matlab 7.4. In graph the dotted line represent the ITU-T method and the normal line represent the proposed method and it can be seen that both the method gives nearly the same result for calculation of open loop pitch.

For simulation we used a matlab Software. From fig 3 we displayed two graphs.graph 1 is shows the original speech and graph 2 shows reconstructed speech. Both the graph is nearly same and this proves that reconstructed speech gives the same toll quality



Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

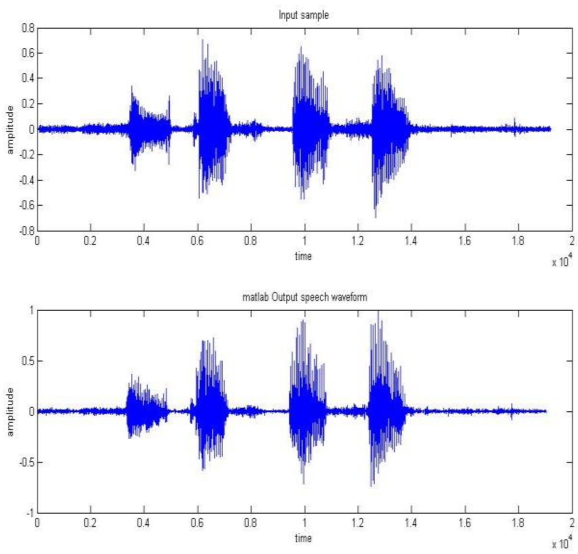


Figure 3:- original speech and reconstructed speech

thank ... ." Instead, write "F. A. Author thanks ... ." **Sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page**.

VI.ACKNOWLEDGMENTS

Special thanks to Mr.Salami, ITY-T, Mr. Kataoka who's paper help me to study for this topic

REFERENCES

[1] Salami et al: 'Design and Description of CS-ACELP: A toll quality 8kb/s speech coder', IEEE trans Speech Audio Process, 1996.

[2] ITU-T G.729: 'Coding of speech at 8 kb/s using CS-ACELP', 1996.

[3] Kataoka et al: 'An 8 kb/s speech coder based on conjugate structured CELP', IEEE int. conf. acoustic, speech, signal processing, 1993.

[4] kataoka et al: 'LSP and gain quantization for proposed ITU-T 8 kb/s speech coding standard', IEEE workshop on speech coding, 1995.

[5] Shaw Hwa Hwang: ' Comput tional improvement for G.729 standard', 2003.

[6] A. B. Roach, "Session Initiation Protocol (SIP) -specific event notification," RFC 3265, June 2002.

[7] A. Johnston, S. Donovan, R. Sparks, C. Cunningham, and K. Summers, "Session Initiation Protocol (SIP) Public Switched Telephone Network (PSTN) call flows," RFC 3666, December 2003.

[8] R. Sparks, "The Session Initiation Protocol (SIP) refer method," RFC 3515, April 2003.

[9] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Feb. 2001.

[10] ITU-T Recommendation P.862 Amendment 1, "Source code for reference implementation and conformance tests," March 2003.

[11] *A.* E. Conway, "Output-based method of applying PESQ to measure the perceptual quality of framed speech signals", in IEEE Wireless Communications and Networking Conference, Vol. 4, pp. 2521-2526,

[1] C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[2] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style)," *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301].

[3] M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.

[4] (Basic Book/Monograph Online Sources) J. K. Author. (year, month, day). *Title* (edition) [Type of medium]. Volume (issue). Available: http://www.(URL)

[5] J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com

[6] (Journal Online Sources style) K. Author. (year, month). Title. *Journal* [Type of medium]. Volume(issue), paging if given. Available: http://www.(URL)

[7] R. J. Vidmar. (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. *21(3).* pp. 876—880. Available: