Efficient Search in the Adaptive Codebook for ITU-T G.729 Codec

Thiago de M. Prego and Sergio L. Netto, *Senior Member, IEEE*

***Abstract*—This paper describes computationally efficient implementations for the ITU-T G.729 speech codec. Focus is given to the adaptive codebook search, more specifically in the open-loop stage, which first estimates the pitch period of the speech frame being coded. Different strategies are discussed to achieve an excellent compromise between computational complexity and signal quality. The result is an accelerated procedure for the G.729 and G.729A versions, reducing encoding time in about 12% and 9%, respectively, while sustaining original signal quality, as verified by objective and subjective measurements.**

***Index Terms*—Adaptive codebook, ITU-T G729, open loop.**

I. INTRODUCTION

THE ITU-T G.729 [1] (conjugate-structure algebraic code-excited linear prediction, CS-ACELP) and its accelerated version, the G.729 Annex A (G.729A) [2], are speech codecs with 8 kbps transmission rate and mean opinion score (MOS) of approximately 3.84 and 3.75, respectively. CELP-based coders [3] achieve such an excellent compromise between transmission rate and coding quality by employing the so-called analysis-by-synthesis scheme to determine the excitation signal for the linear prediction (LP) vocal-tract model. Implementing such a procedure, however, demands intense computational effort, motivating several attempts to accelerate this process [4]–[9], most of them prior to the development of ITU-T P.862 (perceptual evaluation of speech quality, PESQ) recommendation [10].

The present paper presents a modification in the adaptive-codebook search algorithm of the G.729 and G.729A encoders, more specifically in the open-loop stage, which estimates the pitch period of a particular speech frame. The result is a reduction in the order of 90% of the computational burden required by this stage, corresponding to a reduction of approximately 10% of the overall processing time, without affecting the signal quality of the decoded signal.

To introduce the proposed technique, this paper is organized as follows. In Section II, the search algorithm for the adaptive codebook in the G.729 and G.729A codecs is described. In Section III, the accelerated computation of the pitch period for the adaptive codebook is presented. Section IV demonstrates the computational savings achieved by the proposed methods and Section V validates the results with respect to overall signal quality; Section VI then closes the paper emphasizing its technical contributions.

Manuscript received April 01, 2009; revised June 22, 2009. First published June 26, 2009; current version published July 29, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Stephen M. McLaughlin.

The authors are with the Electrical Engineering Program, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil (e-mail: thprego@lps.ufrj.br; sergioln@lps.ufrj.br).

Digital Object Identifier 10.1109/LSP.2009.2026204

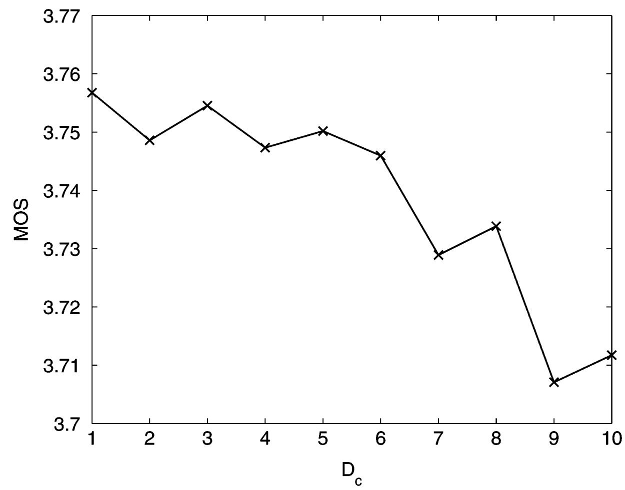


Fig. 1. PESQ-MOS of modified G.729A codec using time decimation D_c.

II. ADAPTIVE CODEBOOK SEARCH

The G.729 and G.729A adaptive codebook contains samples of the LP excitation signal for previous speech frames. In this codebook, the excitation index search is performed according to three steps.

In the first step, also known as open loop, the autocorrelation $R(\tau)$ function of the target signal $sw(n)$ is determined as

$$R(\tau) = \sum_{n=0}^{39} sw(D_c n - \tau), \quad 20 \leq \tau \leq 143 \qquad (1)$$

where the decimating factor $D_c = 2$ is used to reduce computational cost. The first estimate $T_{\rm op}$ for the pitch period is then determined as the value of τ which maximizes $R(\tau)$. In the G.729A implementation, for the interval $80 \leq \tau \leq 143$, (1) is determined only for even values of τ, to further reduce computational effort.

In the second step, the so-called closed loop, the cross-correlation between the target signal and the LP-filter output is maximized. In this search, only the excitations associated to the interval $(T_{\rm op} - 3) \leq \tau \leq (T_{\rm op} + 3)$ are taken into account, and the adaptive codebook delay for the first and second subframes is referred to as T_1 and T_2, respectively.

In the third and final step, an interpolator refines the values of T_1 and T_2 by testing fractional delay times between the optimal excitation and the target signal.

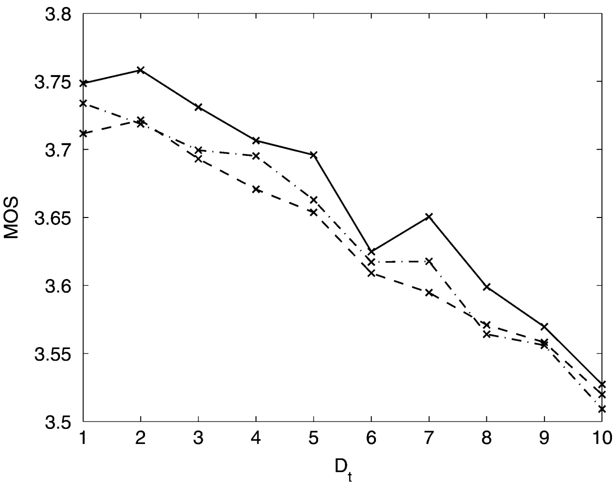


Fig. 2. PESQ-MOS of modified G.729A codec using lag decimation D_t, for $D_c = 2$ (solid line), $D_c = 8$ (dash-dotted line), $D_c = 10$ (dashed line).

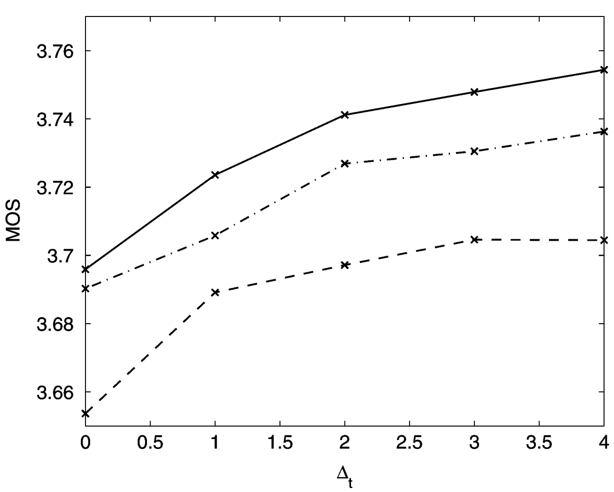


Fig. 3. PESQ-MOS of modified G.729A codec using neighboring Δ_t search, for $D_t = 5$ and $D_c = 2$ (solid line), $D_c = 5$ (dash-dotted line), and $D_c = 10$ (dashed line).

III. Proposed Modifications

The section describes four modifications to the open-loop stage of the G.729 and G.729A codecs. These proposals, when combined in a proper way, can reduce significantly the overall computation burden for that stage.

A first and natural acceleration proposal is to consider (1) with larger values of D_c. For instance, by using $1 \leq D_c \leq 10$, the resulting average PESQ-MOS [10] of a set of 40 speech samples are shown in Fig. 1 for the modified G.729A codec. From this figure, one notices that larger values of D_c simplify the open-loop stage but also reduce the final speech quality. Similar results are achieved by modifying all G.729 variations.

A second proposal is the decimation of τ by a factor D_t in the computation of $R(\tau)$, for the whole interval $20 \leq \tau \leq 143$. Naturally, this simplification also reduces the quality of the resulting signal, as shown in Fig. 2, for $1 \leq D_t \leq 10$ and several values of D_c.

To reduce the effect of decimating the lag value τ, a third modification contemplates the additional computation of $R(\tau)$ in an interval of length $0 \leq \Delta_t \leq (D_t - 1)$ around the initial delay estimate. This process slightly increases the compu-

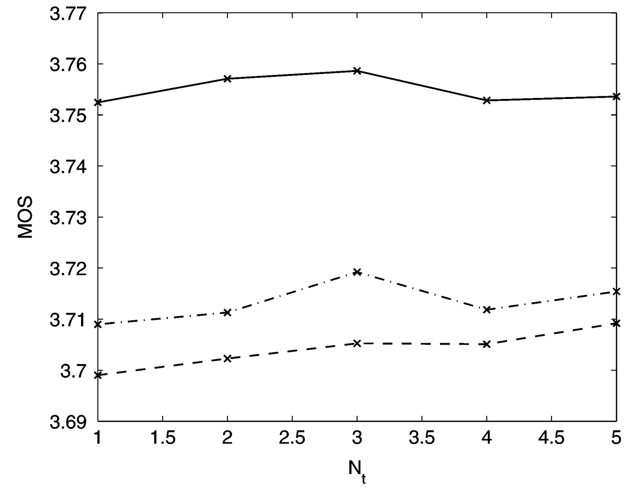
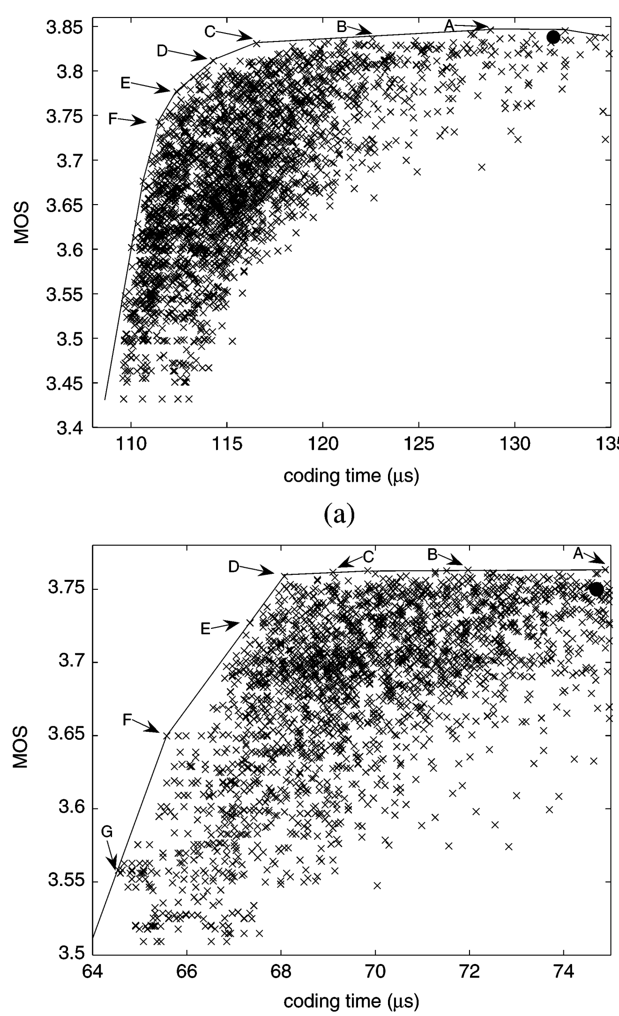


Fig. 4. PESQ-MOS of modified G.729A codec using N_t initial estimates of T_{op}, for distinct values of D_c, D_t and Δ_t : $\Delta_t = 1$, $D_t = 2$, $D_c = 2$ (solid line), $\Delta_t = 5$, $D_t = 7$, $D_c = 8$ (dash-dotted line), $\Delta_t = 7$, $D_t = 8$, $D_c = 10$ (dashed line).



(b)

Fig. 5. PESQ-MOS × coding time of different configurations of modified codecs. The marked points define the concave closure for all parameter sets and the large dot represent the original codec: (a) G.729; (b) G.729A.

TABLE I
CHARACTERISTICS OF SELECTED CONFIGURATIONS OF MODIFIED G.729 WITH BEST COMPROMISE PESQ-MOS $\times T$ IN FIG. 5(A)

Points	D_c	D_t	Δ_t	N_t	MOS	M	$T(\mu s)$
G.729	1	1	0	1	3.84	9920	132.0
A	2	10	9	3	3.85	7000	128.7
B	2	6	5	3	3.84	4440	122.6
C	2	2	1	1	3.83	2720	116.6
D	2	4	3	1	3.81	1960	114.3
E	4	4	3	1	3.78	980	112.4
F	4	5	2	1	3.74	740	111.5

TABLE II
CHARACTERISTICS OF SELECTED CONFIGURATIONS OF MODIFIED G.729A WITH BEST COMPROMISE PESQ-MOS $\times T$ IN FIG. 5(B)

Points	D_c	D_t	Δ_t	N_t	MOS	M	$T(\mu s)$
G.729A	2	1	0	1	3.75	3680	74.7
A	2	7	4	3	3.76	3600	74.9
B	4	4	2	3	3.76	1340	72.0
C	3	3	2	1	3.76	1458	69.1
D	5	4	2	1	3.76	688	68.1
E	7	5	3	1	3.73	516	67.3
F	3	7	0	1	3.65	486	65.6
G	10	9	0	1	3.56	112	64.6

tational cost but also improves the signal quality, as shown in Fig. 3.

The previous improvement associated to parameter Δ_t is not able to determine the best value of $T_{\rm op}$ when it falls far away (within an interval greater than D_t) from its first estimate. This explains why the MOS value did not return to the 3.75 level in Fig. 3, even with the additional computations associated to the Δ_t parameter. To solve this problem, one can select the N_t best initial estimations of $T_{\rm op}$ for each of the three subintervals of τ, and then apply the Δ_t neighboring analysis around all of them to find the final delay estimation. If N_t is large enough, with the additional computations, the resulting MOS shall approximate its original value, as indicated in Fig. 4.

IV. COMPUTATIONAL COMPLEXITY AND QUALITY ANALYSES

Considering the contributions of all modifications presented above, the resulting open-loop stage requires M multiplications and A additions as given by

$$M = \left(\left\lfloor \frac{123}{D_t} \right\rfloor + 1 + 6N_t\Delta_t \right) \times \left(\left\lfloor \frac{79}{D_c} \right\rfloor + 1 \right), \quad (2)$$

$$A = \left(\left\lfloor \frac{123}{D_t} \right\rfloor + 1 + 6N_t\Delta_t \right) \times \left(\left\lfloor \frac{79}{D_c} \right\rfloor \right). \quad (3)$$

Therefore, the level of computation reduction depends on the combination of the four parameters working together. The four schemes presented above were incorporated simultaneously into the ITU-T G.729 and G.729A algorithms using the following ranges: $1 \leq D_c \leq 10, 1 \leq D_t \leq 10, 0 \leq \Delta_t \leq (D_t - 1), 1 \leq N_t \leq 5$. The resulting PESQ-MOS, for an ensemble of 40 speech signals, as a function of the frame coding time T (in an AMD Turion MK-36, 2.01 GHz, machine with 480 MB RAM) for the modified G.729 and G.729A codecs are shown in Fig. 5, in which some configurations of the vector $[D_c, D_t, \Delta_t, N_t]$ with the best quality/effort compromise are selected.

These G.729 and G.729A selected configurations are characterized in Tables I and II, respectively, which also includes the associated PESQ-MOS, number of multiplications M, and average frame coding time T. These results indicate that the proposed modifications can collectively reduce the overall coding time in about 12% (configuration C in Table I for the G.729) or 9% (configuration D in Table II for the G.729A) as compared to the original codec implementations, while sustaining the quality level of speech.

The quality of speech signals obtained by the modified codecs (G.729 configuration C and G.729A configuration D) was also assessed subjectively by a comparison category rating (CCR)

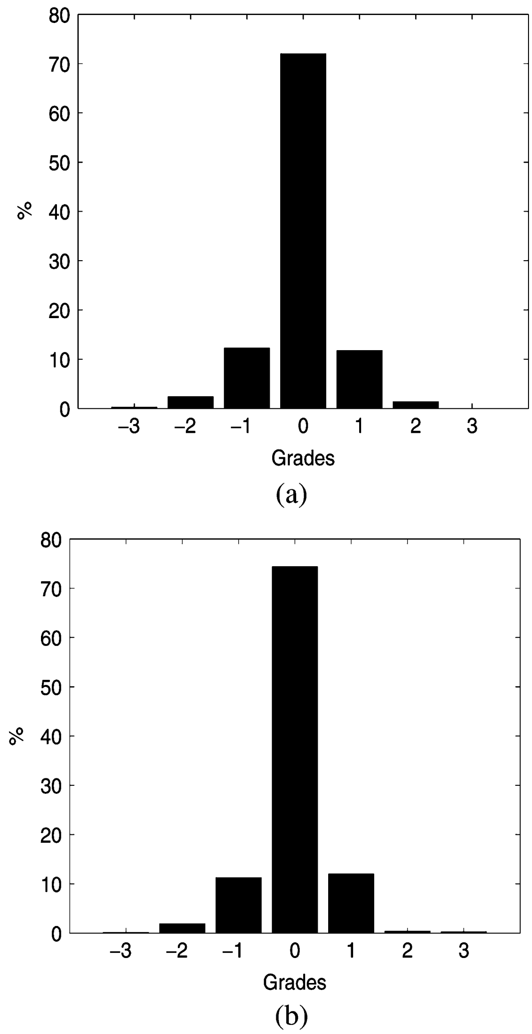


Fig. 6. Histograms of comparative grades (positive scores favor the modified codec): (a) original $\times$ modified G.729 (configuration C in Table I); (b) original $\times$ modified G.729A (configuration D in Table II).

TABLE III
GRADE SCALE DESCRIPTION OF CCR TEST [11]: POSITIVE SCORES FAVOR THE MODIFIED CODEC

CCR Grade	Meaning
$+3$	Much Better
$+2$	Better
$+1$	Slightly Better
0	About the Same
-1	Slightly Worse
-2	Worse
-3	Much Worse

test, as described on the ITU-T P.800 recommendation [11]. In this test, 32 different speech signals were encoded/decoded by the original and modified versions of the G.729 and G.729A codecs. Then, 25 listeners were asked to compare the modified signals directly to their original counterparts following a random listening order. In each comparison, a grade was given in response to the enquiries "Which sample has better quality? By how much?," following the scale shown in Table III, [11], in which the $+$ and $-$ signs favor the modified or original signals, respectively. Results for this test are seen in Fig. 6, and indicate, for all practical purposes, complete equivalence between the modified and original versions of both codecs.

V. Conclusion

Four modifications were proposed for accelerating the open-loop stage in the adaptive codebook search for the ITU-T G.729 and G.729A encoders, keeping complete compatibility with respective original decoders. If put together, the proposed simplifications can reduce the processing time required by these codecs in approximately 12% and 9%, respectively, without affecting the quality of the resulting speech signal, as confirmed by both objective and subjective measurements.

References

[1] Coding of Speech at 8 Kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP) 1996, ITU-T Rec. G.729.

[2] Reduced Complexity 8 Kbit/s CS-ACELP Speech Codec 1996, ITU-T Rec. G.729 Annex A.

[3] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High quality speech at very low bit rates," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tampa, USA, 1985, vol. 2, pp. 437–440.

[4] H. K. Kim, "Adaptive encoding of fixed codebook in CELP coders," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Seattle, USA, May 1998.

[5] N. K. Ha, "A fast search method of algebraic codebook by reordering search sequence," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, AZ, May 1999.

[6] T. Amada, K. Miseki, and M. Akamine, "CELP speech coding based on an adaptive pulse position codebook," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, AZ, May 1999.

[7] S.-H. Hwang, "Computational improvement for G.729 standard," *Electron. Lett.*, vol. 36, no. 13, pp. 1163–1164, Jun. 2000.

[8] M. A. Ramirez and M. Gerken, "Joint position and amplitude search of algebraic multipulses," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 633–637, Sep. 2000.

[9] E. D. Lee, S. H. Yun, S. I. Lee, and J. M. Ahn, "Iteration-free pulse replacement method for algebraic codebook search," *Electron. Lett.*, vol. 43, no. 1, pp. 59–60, Jan. 2007.

[10] Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs 2001, ITU-T Rec. P.862.

[11] Methods for Subjective Determination of Transmission Quality 1996, ITU-T Rec. P.800.