

Nova Aceleração da Busca no Dicionário Adaptativo do Codificador ITU-T G.729 Anexo A

Thiago de M. Prego e Sergio L. Netto

Resumo—Este artigo descreve uma aceleração no algoritmo de busca pela excitação do dicionário adaptativo no codificador ITU-T G.729 Anexo A. As modificações são aplicadas no estágio de *open loop* da busca no dicionário adaptativo, que fornece como saída uma primeira estimativa do período de *pitch* do segmento a ser codificado. São apresentadas duas estratégias que objetivam diminuir a complexidade computacional desta etapa, mantendo uma alta qualidade da codificação resultante. Os resultados incluídos neste trabalho indicam que estas duas técnicas podem ser combinadas de uma maneira complementar, resultando em uma redução de aproximadamente 90% da complexidade computacional da etapa de *open loop*, sem afetar a qualidade percebida do sinal resultante de forma significativa.

Palavras-Chave—ITU-T G.729 Anexo A, busca no dicionário adaptativo, estágio de *open loop*.

Abstract—This paper describes accelerating strategies for the adaptive-codebook excitation search of ITU-T G.729 Annex A encoder. The modifications are applied to the open-loop stage, which gives a first estimate for the pitch period of the frame being coded. Two strategies are presented reducing the computational complexity of the open-loop stage while sustaining the quality of the resulting speech signal. The results presented indicate that these two techniques can be combined in a complementary manner, reducing the open-loop computational complexity in approximately 90%, without affecting significantly the perceptual quality of the decoded signal.

Keywords—ITU-T G.729 Annex A encoder, adaptive codebook search, open-loop stage.

I. INTRODUÇÃO

A Recomendação ITU-T G.729 Anexo A [1] (G.729A) descreve um algoritmo para um codificador de fala utilizando a técnica de *conjugate-structure* (CS) *algebraic code-excited linear prediction* (ACELP) [2], [3]. A fim de obter um excelente compromisso entre taxa de transmissão e qualidade de codificação percebida, os codificadores ACELP se baseiam na técnica de análise-por-síntese (*analysis-by-synthesis*, AbS) para determinar o sinal de excitação para o modelo LP do trato vocal. O esquema AbS requer um grande esforço computacional, razão pela qual muitas trabalhos na literatura [4]–[10] buscam acelerar tal processo.

O presente trabalho investiga um novo esquema de aceleração da etapa de *open-loop* de busca da melhor excitação do dicionário adaptativo do codificador G.729A, responsável por aproximadamente 26% do tempo total de codificação [7]. A seguir, é analisada a combinação desta nova técnica com outras propostas em [10]. O avaliador objetivo de qualidade de

sinais de fala PESQ (*perceptual evaluation of speech quality*), descrito na Recomendação ITU-T P.862 [11], é utilizado para estimar a qualidade percebida do codificador resultante. Resultados indicam que é possível acelerar a implementação do codificador G.729A de forma significativa, sem alterar a qualidade do sinal de voz avaliada pelo algoritmo PESQ.

Este artigo está organizado da seguinte forma: a Seção II descreve o algoritmo padrão da busca no dicionário adaptativo do codificador G.729A; a Seção III descreve simplificações propostas em [10] enquanto que a Seção IV apresenta uma nova simplificação para a etapa de *open loop* do G.729A; a Seção V então analisa a combinação das modificações vistas nas seções anteriores com respeito à complexidade computacional resultante e qualidade percebida de fala, como avaliado pelo PESQ; a Seção VI fecha o artigo resumindo suas principais contribuições.

II. BUSCA NO DICIONÁRIO ADAPTATIVO

O dicionário adaptativo do codificador G.729A é constituído por excitações do filtro LP de segmentos anteriores. A busca pela melhor excitação é feita em três etapas, como mostra a Figura 1, descritas por:

- *Open loop*: Nesta etapa, a função de autocorrelação $R(\tau)$ é determinada para cada segmento de 10 ms do sinal de fala ponderado pelo filtro perceptual $sw(n)$, segundo

$$R(\tau) = \sum_{n=0}^{39} sw(2n)sw(2n - \tau). \quad (1)$$

Três intervalos para τ são considerados: (a) $20 \leq \tau \leq 39$; (b) $40 \leq \tau \leq 79$; (c) $80 \leq \tau \leq 143$. Para o G.729A, apenas os valores pares de τ são considerados no intervalo (c). Para cada intervalo, um candidato T_i , com $i = a, b, c$, é determinado como o valor de τ que maximiza $R(\tau)$, como dada na equação (1). Para estes candidatos a função de autocorrelação é normalizada pela energia total do segmento em questão, sendo modificada para:

$$R'(T) = \frac{R(T)}{\sum_{n=0}^{39} sw^2(2n - T)}, \quad (2)$$

e a estimativa do atraso de *pitch* T_{op} é selecionada dentre estes três candidatos através de uma comparação ponderada, favorecendo atrasos menores. Esta ponderação tem a finalidade de evitar a seleção de múltiplos do período de *pitch*, que estão associados a harmônicos da frequência fundamental.

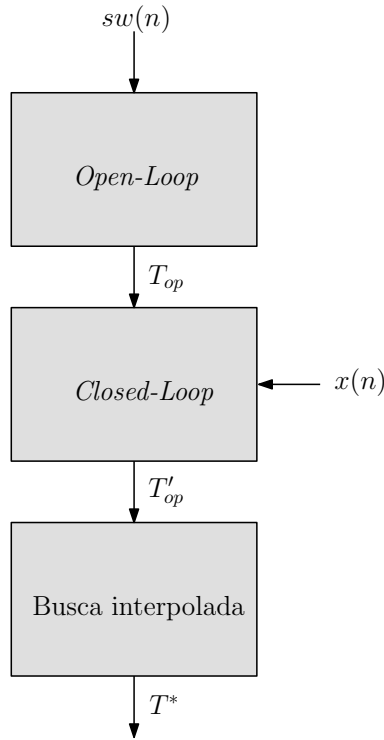


Fig. 1. Diagrama de blocos da busca no dicionário adaptativo do codificador G.729A.

Neste contexto do G729A, o cálculo de cada uma das 92 autocorrelações $R(\tau)$ para diferentes valores de τ demanda 40 multiplicações, totalizando $M = 3680$ multiplicações para a etapa de *open loop*.

- *Closed loop*: Nesta etapa, a estimativa refinada do atraso T'_{op} é determinada como o valor de k que maximiza a função de correlação cruzada

$$R_{xy}(k) = \frac{\sum_{n=0}^{39} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{39} y_k(n)y_k(n)}}, \quad (3)$$

entre o sinal alvo $x(n)$ e a saída do filtro LP $y_k(n)$ para a entrada da excitação do dicionário adaptativo com um atraso k . Para reduzir a complexidade computacional nesta etapa, apenas os valores de k no intervalo $(T_{op} - 3) \leq k \leq (T_{op} + 3)$ são considerados.

- *Busca interpolada*: Nesta etapa, um interpolador é utilizado para detectar possíveis valores fracionários ao redor de T'_{op} , com uma resolução de $\frac{k}{3}$, entre a excitação ótima e o sinal alvo.

III. PROPOSTAS ANTERIORES DE BUSCA SIMPLIFICADA

Esta seção lista um conjunto de modificações propostas em [10] para o codificador G729 [2] mas que podem também ser adaptadas para o G.729A.

A. Fator de dizimação no tempo

Uma extensão natural da aceleração utilizada pelo G.729A é considerar diferentes fatores $D_c \geq 2$ de dizimação no tempo n na equação (1), fazendo

$$R(\tau) = \sum_{n=0}^{\lfloor \frac{79}{D_c} \rfloor} sw(D_c n)sw(D_c n - \tau). \quad (4)$$

que requer apenas $(\lfloor \frac{79}{D_c} \rfloor + 1)$ multiplicações para cada valor de τ .

B. Fator de dizimação do atraso

Uma segunda proposta é decimar a variável τ no cálculo da função de autocorrelação $R(\tau)$ por um fator D_t em todo o intervalo $20 \leq \tau \leq 143$, e não somente no intervalo (c), como utilizado pelo G729A. Esta combinação de fatores de dizimação leva o número total de multiplicações para o cálculo de $R(\tau)$ para $M = (\lfloor \frac{123}{D_t} \rfloor + 1) \times (\lfloor \frac{79}{D_c} \rfloor + 1)$.

C. Análise local do atraso

A introdução do parâmetro D_t evita o cálculo de $R(\tau)$ para alguns valores de τ . Isto faz com que a qualidade percebida do sinal de fala reconstruído diminua conforme o valor de D_t aumenta. Para minimizar esta perda de qualidade, propõe-se o cálculo adicional de $R(\tau)$ em um intervalo de comprimento Δ_t em torno da primeira estimativa de T_{op} , introduzindo $2\Delta_t \times (\lfloor \frac{79}{D_c} \rfloor + 1)$ multiplicações adicionais.

IV. NOVAS PROPOSTAS DE BUSCAS SIMPLIFICADAS

Nesta seção, apresentamos uma nova técnica simplificada de busca no dicionário adaptativo do G.729A. Mais adiante, esta nova proposta será combinada às propostas vistas na Seção III e o efeito disto na complexidade computacional e qualidade resultantes será investigado.

Em um experimento inicial, foi verificada a importância de cada intervalo de τ , como definido na Seção II, no cálculo da função de autocorrelação $R(\tau)$. Neste estudo, usou-se $T_i = \bar{\tau}_i$ em um dado intervalo $i = a, b, c$, e a busca por T_{op} foi feita de forma padrão nos outros dois intervalos. Assim foram definidas três configurações do G729A modificado:

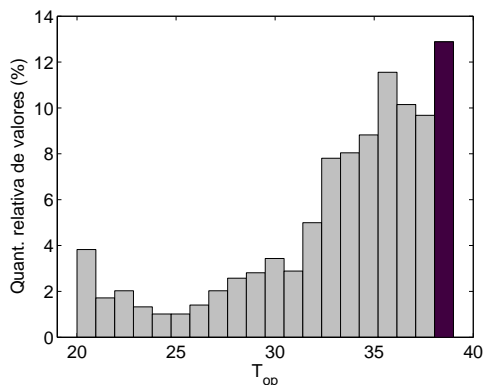
- Configuração I: T_a é mantido fixo e T_b e T_c são determinados da forma padrão;
- Configuração II: T_b é mantido fixo e T_a e T_c são determinados da forma padrão;
- Configuração III: T_c é mantido fixo e T_a e T_b são determinados da forma padrão.

A importância de cada intervalo foi avaliada pela perda de qualidade do sinal de voz quando T_i foi pré-determinado no intervalo i : quanto maior a perda de qualidade do sinal, maior a importância do intervalo.

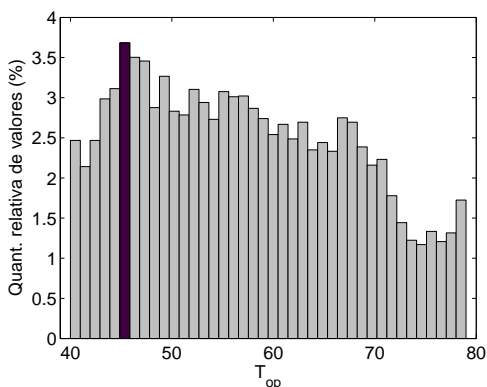
Neste experimento, o valor fixo $T_i = \bar{\tau}_i$ foi escolhido como a moda de T_{op} neste intervalo i , quando usamos o codificador G729A em um banco de dados BD1 composto por 40 sinais de voz no formato PCM 16-bits, 8 sinais para cada um de 5 idiomas: chinês mandarim (1 locutor masculino), inglês

britânico (1 locutor masculino), indiano (1 locutor masculino), francês (1 locutor masculino), inglês americano (6 frases de 1 locutora feminina e 2 frases de 1 locutor masculino). Os histogramas de cada T_{op} para cada sub-intervalo são mostrados na Figura 2, de onde se determina que

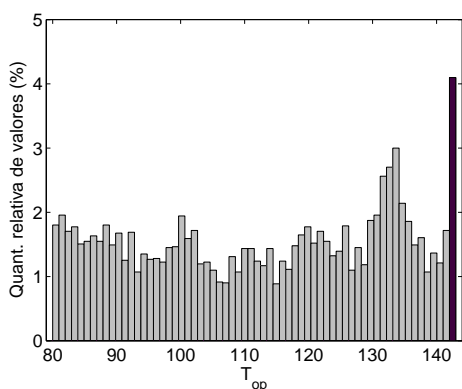
$$\begin{cases} \bar{\tau}_a = 39 \\ \bar{\tau}_b = 45 \\ \bar{\tau}_c = 143 \end{cases} \quad (5)$$



(a)



(b)



(c)

Fig. 2. Histograma (com destaque para a moda) da variável T_{op} para o codificador G.729A nos intervalos: (a) $20 \leq T_{op} \leq 39$; (b) $80 \leq T_{op} \leq 79$; (c) $80 \leq T_{op} \leq 143$.

A Tabela I mostra a distribuição de T_{op} em cada intervalo e a avaliação PESQ média para todo o banco BD1 para cada

configuração do G729A modificado. Com base nesta tabela, é possível concluir que, mantendo-se $T_c = \bar{\tau}_c = 143$ fixo para todo o intervalo (c), a nota PESQ para o banco BD1 permanece inalterada em relação à nota do G729A padrão, enquanto a complexidade computacional requerida pelo cálculo de $R(\tau)$ é decrescida de 35%. Neste caso, o intervalo (b) parece suprir toda a deficiência provocada no G729A modificado ao se fazer T_c fixo.

TABELA I
DISTRIBUIÇÃO DE T_{op} E AVALIAÇÃO PESQ PARA O BANCO DE SINAIS DE VOZ BD1 PARA CADA CONFIGURAÇÃO DO CODIFICADOR G729A MODIFICADO.

Configuração	(a)%	(b)%	(c)%	PESQ
G.729A	6,60	56,80	36,60	3,75
I	2,17	59,35	38,48	3,70
II	8,08	14,00	77,92	3,59
III	12,88	79,65	7,47	3,75

V. COMBINAÇÃO DAS PROPOSTAS

Nesta seção, investiga-se a combinação das três técnicas apresentadas na Seção III com a nova proposta vista na Seção IV. Verifica-se a maneira com que estes métodos todos trabalham em conjunto para reduzir a complexidade computacional do codificador G.729A e como eles afetam a qualidade percebida do sinal resultante de fala. Para isto, o algoritmo PESQ foi utilizado para estimar a qualidade percebida do codificador G.729A incorporando as quatro modificações vistas anteriormente com:

$$\begin{cases} 1 \leq D_c \leq 10 \\ 1 \leq D_t \leq 10 \\ 0 \leq \Delta_t \leq D_t - 1 \\ T_c \text{ variável ou } T_c = 143 \text{ (fixo)} \end{cases} \quad (6)$$

As notas PESQ médias para as diferentes versões do G729A são vistas na Figura 3 em função do número de multiplicações M correspondentes, que atua como uma medida da complexidade computacional. Os fechos côncavos formados pelas melhores relações PESQ $\times M$ da Figura 3 são descritos nas Tabelas II e III. Os resultados destas tabelas indicam que a combinação das técnicas permite reduzir a complexidade do estágio de *open loop* do G729A em aproximadamente 90% do valor original, sem afetar a qualidade do sinal resultante.

TABELA II
AVALIAÇÃO DAS CONFIGURAÇÕES DO CODIFICADOR G.729A MODIFICADO FORMANDO O FECHO CÔNCAVO DA FIGURA 3 (A).

Configuração	D_c	D_t	Δ_t	M	PESQ
A	5	4	3	592	3,73
B	8	4	3	370	3,71
C	8	4	0	310	3,70
D	9	5	0	225	3,67
E	10	5	0	200	3,65

A comparação dos resultados obtidos para T_c variável e fixo é feita na Figura 4. Nesta figura, nota-se, claramente, um deslocamento do fecho côncavo para a esquerda quando

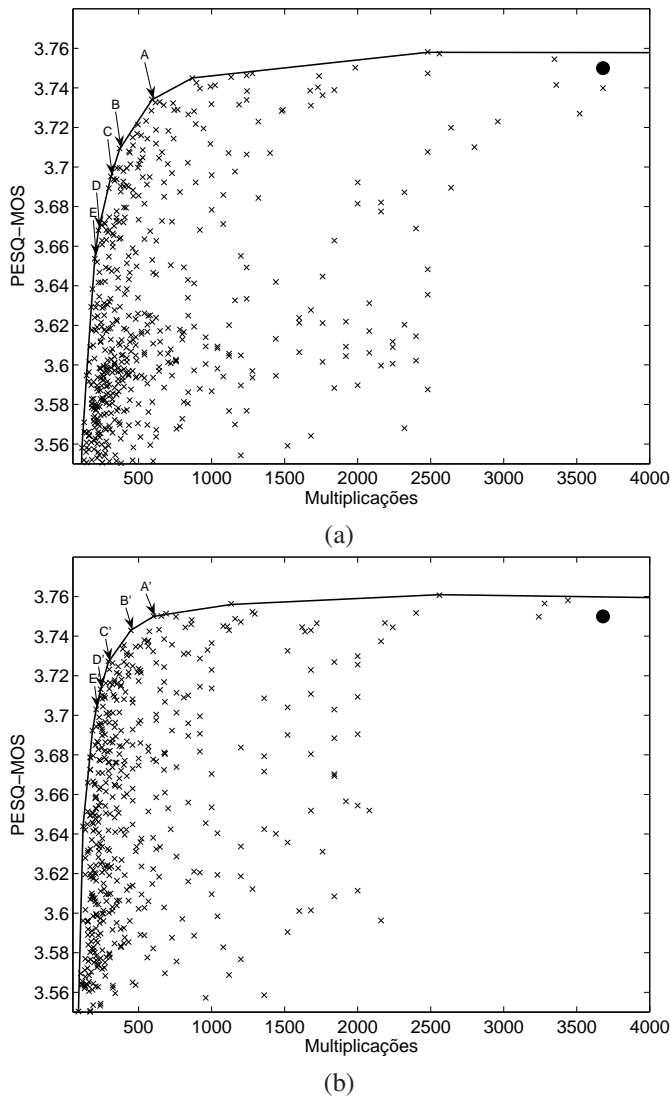


Fig. 3. Notas PESQ em função do número de multiplicações M para versões modificadas do codificador G.729A utilizando (O círculo cheio representa o codificador G.729A original): (a) T_c variável; (b) T_c fixo.

se matém T_c fixo. Isto indica que a nova proposta acarreta a diminuição da complexidade computacional enquanto mantém a qualidade percebida do sinal de fala reconstruído. Em termos quantitativos, para as configurações incluídas na Tabela III, o uso de T_c fixo diminui sensivelmente o número de multiplicações requeridas pela G.729A com as modificações de [10], sendo a redução mais relevante na faixa de 30–50% para uma nota PESQ acima de 3,70.

VI. CONCLUSÃO

Este artigo investigou o desempenho da combinação de quatro métodos de aceleração do algoritmo da etapa de *open loop* da busca no dicionário adaptativo do codificador ITU-T G.729 Anexo A. Os resultados indicam que a nova proposta acelera em até 50% os métodos anteriormente publicados em [10]. Além disto, foi visto que a combinação de todos estes métodos permite reduzir a complexidade computacional desta etapa em cerca de 90% do valor original, sem diminuir a qualidade do sinal de fala reconstruído (ponto B' da Tabela III).

TABELA III
AVALIAÇÃO DAS CONFIGURAÇÕES DO CODIFICADOR G.729A MODIFICADO FORMANDO O FECHO CÔNCAVO DA FIGURA 3 (B).

Configuração	D_c	D_t	Δ_t	M	PESQ
A'	6	2	1	602	3,75
B'	6	3	2	448	3,74
C'	6	4	0	294	3,73
D'	10	3	1	240	3,71
E'	8	4	0	210	3,70

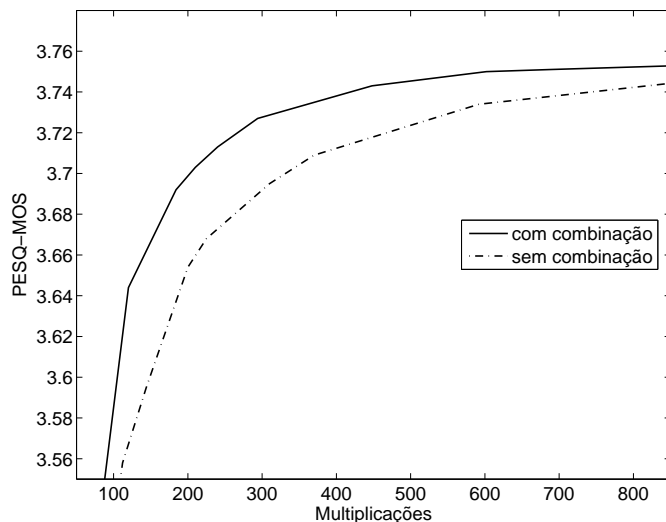


Fig. 4. Fechos cômavos da configuração referente a [10] (linha tracejada) e da configuração referente à combinação das propostas (linha contínua).

REFERÊNCIAS

- [1] ITU-T Rec. G.729 Annex A, *Reduced Complexity 8 kbit/s CS-ACELP Speech Codec*, 1996.
- [2] ITU-T Rec. G.729, *Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)*, 1996.
- [3] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High quality speech at very low bit rates," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tampa, USA, vol. 2, pp. 437–440, 1985.
- [4] H. K. Kim, "Adaptive encoding of fixed codebook in CELP coders," *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," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, USA, May 1999.
- [6] T. Amada, K. Miseki, and M. Akamine, "CELP speech coding based on an adaptive pulse position codebook," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, USA, May 1999.
- [7] S.-H. Hwang, "Computational improvement for G.729 standard," *IEE Electronics Letters*, vol. 36, no. 13, pp. 1163–1164, June 2000.
- [8] M. A. Ramirez and M. Gerken, "Joint position and amplitude search of algebraic multipulses," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 5, pp. 633–637, Sept. 2000.
- [9] E. D. Lee, S. H. Yun, S. I. Lee, and J. M. Ahn, "Iteration-free pulse replacement method for algebraic codebook search," *IEE Electronics Letters*, vol. 43, no. 1, pp. 59–60 Jan. 2007.
- [10] T. de M. Prego e S. L. Netto, "Algoritmo de Busca Eficiente no Dicionário Adaptativo para o Codec ITU-T G.729," *SBt - Simpósio Brasileiro de Telecomunicações*, Rio de Janeiro, Brasil, 2008.
- [11] ITU-T Rec. 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*, 2001.