Algoritmo de Busca Eficiente no Dicionário Adaptativo para o Codec ITU-T G.729

Thiago de M. Prego e Sergio L. Netto

***Resumo*— Este artigo descreve implementações computacionalmente eficientes do codificador de voz CS-ACELP em suas versões original (ITU-T G.729) e acelerada (ITU-T G.729 Anexo A). O foco reside na busca do dicionário adaptativo, mais especificamente na etapa de *open loop*, em que se estima o pitch de um segmento do sinal codificado. Diferentes estratégias são testadas para ambas versões, procurando-se atingir excelente compromisso de complexidade computational e qualidade do sinal. O resultado é uma busca acelerada, com cerca de 15%–30% da complexidade original, com a mesma qualidade medida pelo avaliador objetivo PESQ (ITU-T P.862).**

***Palavras-Chave*— Codificação de voz; ITU-T G.729; Algoritmo eficiente.**

***Abstract*— This paper describes computationally efficient implementations for both standard (ITU-T G.729) and accelerated (ITU-T G.729 Annex A) versions of the speech codec CS-ACELP. Focus is given to the search in the adaptive codebook, more specifically in the open loop stage, which estimates the pitch of the speech frame being coded. Different strategies are discussed in an attempt of reaching an excellent compromise between the overall computational complexity and signal quality. The result is an accelerated procedure, with about 15%–30% of the original burden, with the same signal quality estimated by the PESQ (ITU-T P.862) recommendation.**

***Keywords*— Speech coding; ITU-T G.729; Efficient algorithm.**

I. INTRODUÇÃO

Os codificadores de fala da família CELP (*Code-Excited Linear-Prediction*) [1] são amplamente utilizados em telecomunicações, principalmente em telefonia móvel e em sistemas de VoIP (*voice over Internet protocol*). Este tipo de codificador visa gerar *bitstreams* com taxas entre 2 kbps e 16 kbps e com uma qualidade MOS (*mean opinion score*) entre 3,0 e 4,0, como indicado na Figura 1 [2].

A recomendação G.729 CS-ACELP (*conjugate-structure algebraic-CELP*) da ITU-T [3] contém a descrição e implementação de um algoritmo para a codificação de sinais de fala com uma taxa de transmissão de 8 kbps. Os sinais de entrada deste codificador são sinais de fala com banda telefônica (segundo a recomendação ITU-T G.712), com freqüência de amostragem de 8 kHz e codificação PCM linear com 16 bits por amostra. A recomendação G.729 foi primeiramente definida para ponto fixo e posteriormente para ponto flutuante no anexo C [4]. Este anexo descreve a implementação das versões original e acelerada (definida no anexo A [5] e a partir daqui denotada por G.729A) do codec G.729.

Thiago de M. Prego e Sergio L. Netto, Programa de Engenharia Elétrica/COPPE, Universidade Federal do Rio de Janeiro, CP 68504, Rio de Janeiro, Brasil, 21945-972. E-mails: thprego@lps.ufrj.br, sergioln@lps.ufrj.br.

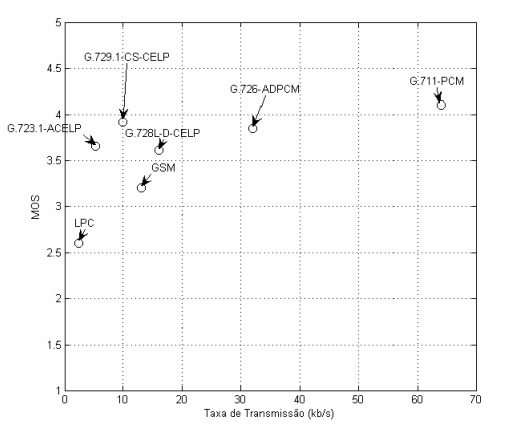

Fig. 1. Relação MOS × taxa para diversos codificadores de voz.

De modo geral, para atingir um excelente compromisso de taxa e qualidade de codificação, a técnica CELP requer um grande esforço computacional. Neste sentido, diferentes trabalhos são encontrados na literatura associada [6], [7], [8], [9] procurando acelerar as implementações desta família de codificadores. Este artigo apresenta uma modificação no algoritmo de busca do dicionário adaptativo do codec G.729 e sua variante G.729A. Mais especificamente, é feita uma aceleração na etapa de *open-loop*, que estima o *pitch* do sub-bloco de voz em análise. A modificação proposta pode ser vista como uma extensão de uma das modificações implementadas no G.729A. Como resultado, atinge-se um excelente ganho em termos de complexidade computacional sem alterar a qualidade do sinal resultante.

Este artigo é estruturado da seguinte forma: A Seção II apresenta a recomendação G.729 e sua variante G.729A em suas características gerais; a Seção III descreve em detalhes a etapa de *open loop* destas duas versões na busca da melhor excitação do dicionário adaptativo; a Seção IV descreve as modificações propostas para esta etapa, ilustrando o efeito de cada uma nos aspectos de qualidade e complexidade computational; a Seção V apresenta os resultados gerais das modificações propostas no desempenho do codec G.729; na Seção VI os principais resultados deste artigo são resumidos.

II. CODECS G.729 E G.729A

O codificador CS-ACELP (*conjugate-structure algebraic-code-excited linear-prediction*) é descrito na recomendação G.729 da ITU-T [3] e é baseado no modelo de predição linear (LP, *linear prediction*) com entrada composta a partir

de dicionários de sinais. São usados dois dicionários: um de conteúdo adaptativo e outro de conteúdo fixo. A busca da melhor excitação em cada caso segue o procedimento de análise-por-síntese (*analysis-by-synthesis*, AbS) em que diferentes trechos de sinais são sintetizados e comparados com um dado sinal-alvo CELP [1]. O codificador G.729 trabalha com segmentos de fala de 10 ms, o que corresponde a 80 amostras do sinal. Os parâmetros do modelo LP são obtidos para cada segmento e os índices e ganhos dos dois dicionários são determinados a cada sub-segmento de 5 ms. Estes parâmetros são codificados e enviados ao decodificador que monta a excitação desejada e o filtro LP de síntese para recuperar o sinal de voz, que ainda é processado por um pós-filtro. O diagrama de blocos que representa o funcionamento geral do codificador G.729 é mostrado na Figura 2 [3].

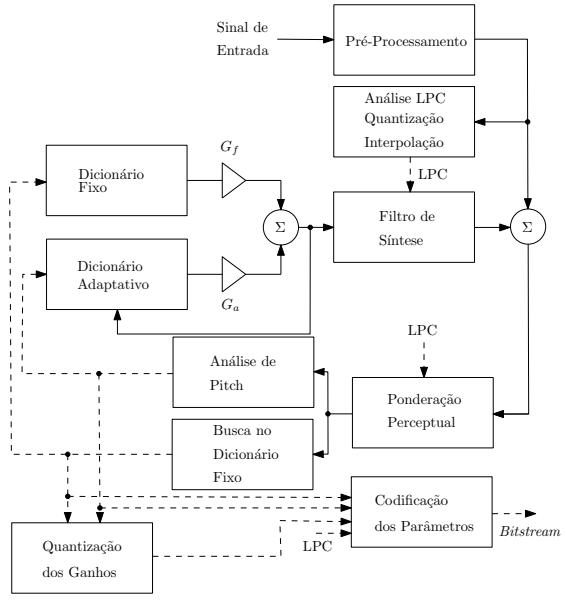

Fig. 2. Diagrama de blocos do codificador ITU-T G.729.

A versão acelerada G.729A possui, essencialmente, o mesmo funcionamento do G.729 básico. De modo geral, as seguintes simplificações das implementações dos blocos vistos na Figura 2 foram incorporadas à versão G.729A [4]:

- O filtro perceptual utiliza os parâmetros LP já quantizados e um peso $\gamma = 0{,}75$ fixo. Isto possibilita simplificações no cascateamento do filtro LP com o filtro perceptual;
- O cálculo de correlação, usado na etapa de *open loop* para estimar o *pitch* do segmento sendo codificado, é decimado por 2, como detalhado na Subseção IV-A;
- A busca do dicionário adaptativo é simplificada, não mais exigindo o cálculo da energia da excitação passada;
- A busca do dicionário fixo substituiu o sistema de laços aninhados por uma estrutura iterativa em árvore.

Estas simplificações reduzem consideravelmente a complexidade computacional do algoritmo, ao preço de uma pequena diminuição na qualidade do sinal de saída.

III. Busca no Dicionário Adaptativo

O dicionário adaptativo do codec G.729 contém amostras dos sinais de excitação do filtro LP relativos aos segmentos de voz anteriores. A busca do índice da excitação neste dicionário é feita em três etapas:

1) Primeiramente, na chamada etapa de *open loop*, calcula-se a autocorrelação do sinal-alvo para estimar o seu período de *pitch* $T_{\rm op}$. Nesta fase, determina-se a autocorrelação $R(\tau)$ com *lag* no intervalo $20 \leq \tau \leq 143$. O ponto deste intervalo onde a autocorrelação é máxima é usado como estimativa T_1 inicial do período de *pitch*;
2) Na etapa de *closed loop*, procura-se maximizar a correlação cruzada do sinal-alvo com as saídas do filtro LP para as possíveis excitações do dicionário adaptativo. Neste estágio, consideram-se apenas as excitações associadas a atrasos no intervalo $(T_1 - 3) \leq T_{\rm op} \leq (T_1 + 3)$;
3) Na última etapa, um interpolador é usado para se detectar possíveis valores fracionários para o atraso relativo entre a excitação ótima e o sinal-alvo.

Estas etapas ocorrem em seqüência e representam refinamentos sucessivos da estimativa do atraso-índice do dicionário adaptativo.

Na etapa de *open loop*, o intervalo de busca da estimativa do período de *pitch* é dividido em três sub-intervalos: (i) $20 \leq \tau \leq 39$; (ii) $40 \leq \tau \leq 79$; (iii) $80 \leq \tau \leq 143$. Se $s_w(n)$ é o sinal-alvo, calcula-se a autocorrelação

$$R(\tau) = \sum_{n=0}^{79} s_w(n) s_w(n-\tau), \tag{1}$$

e determina-se o valor máximo desta função para cada sub-intervalo de τ. Desta forma, no G.729, o cálculo desta função para $\tau \in [20, 143]$ requer $124 \times 80 = 9920$ multiplicações e $124 \times 79 = 9796$ adições. Vale lembrar que todos estes cálculos são realizados para cada segmento de 10 ms de voz.

Os valores máximos $R(\tau_i)$, para $i = 1, 2, 3$, de cada sub-intervalo são então normalizados da forma

$$R'(\tau_i) = \frac{R(\tau_i)}{\sqrt{\sum_{n=0}^{79} s_w^2(n-\tau_i)}}. \tag{2}$$

O maior valor $T_{\rm op}$ dentre as autocorrelações normalizadas é selecionado através de uma comparação ponderada, em que os menores atrasos são favorecidos. Esta ponderação é feita para se evitar a seleção de múltiplos do período de pitch, que estariam de fato associados a harmônicos da freqüência de pitch.

Para o codec G.729A, o cálculo da autocorrelação considera uma versão decimada por $D_c = 2$ do sinal-alvo, de modo que

$$R_A(\tau) = \sum_{n=0}^{39} s_w(D_c n) s_w(D_c n - \tau). \tag{3}$$

Esta modificação reduz, de um fator $D_c = 2$, o número de múltiplicações e adições realizadas neste estágio. A normalização subseqüente é feita da forma

$$R'_A(\tau_i) = \frac{R(\tau_i)}{\sqrt{\sum_{n=0}^{39} s_w^2(D_c n - \tau_i)}}, \tag{4}$$

novamente considerando-se apenas as amostras de índice par do sinal-alvo. Além disto, no sub-intervalo $80 \leq \tau \leq 143$, são considerados apenas valores pares de τ. Assim, na etapa de *open loop*, o G.729A requer $(124 - 32) \times 40 = 3680$ multiplicações e $(124 - 32) \times 39 = 3588$ adições, além do cálculo da normalização, que é ligeiramente mais simples que no caso do G.729 original.

IV. Modificações Propostas

Nesta seção, novas modificações são apresentadas e analisadas para o estágio de *open-loop* dos codecs G.729 e G.729A. Para ilustrar o efeito das simplificações propostas no sinal codificado/decodificado, avaliações objetivas usando a recomendação ITU-T P.862 (PESQ, *perceptual evaluation of speech quality*) foram realizadas em um banco composto de 40 sinais de voz de diversos idiomas (8 em chinês, 8 em francês, 8 em indiano, 8 em inglês britânico e 8 em inglês americano).

A. Decimação do Sinal-Alvo no Cálculo da Autocorrelação

Uma extensão natural da aceleração introduzida no G.729A na etapa de *open loop* é considerar, na equação (3), diferentes valores para o fator de decimação D_c. Desta forma, o cálculo da autocorrelação passa a requerer apenas $124 \times \lfloor \frac{80}{D_c} \rfloor$ multiplicações e $124 \times (\lfloor \frac{80}{D_c} \rfloor - 1)$ adições. O efeito do uso de diferentes valores $2 \leq D_c \leq 20$ para o G.729 e o G.729A é ilustrado na Figura 3. Claramente, é possível perceber que a redução na complexidade computacional vem acompanhada de uma queda da qualidade do sinal.

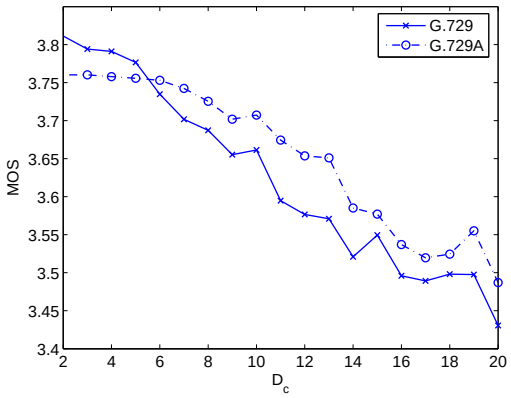

Fig. 3. Teste 1: PESQ-MOS dos codecs G.729 e G.729A em função do fator D_c.

B. Decimação do Lag no Cálculo da Autocorrelação

Uma segunda proposta é a decimação, de um fator D_t, da variável τ no cálculo da autorrelação. Desta forma, esta função passa a ser determinada apenas para os valores $R(D_t\tau)$, com $D_t = 2, 3, \ldots$. Isto reduz o número operações para aproximadamente $\lfloor \frac{124}{D_t} \rfloor \times \lfloor \frac{80}{D_c} \rfloor$ multiplicações e $\lfloor \frac{124}{D_t} \rfloor \times (\lfloor \frac{80}{D_c} \rfloor - 1)$ adições. Naturalmente que esta modificação também acarreta redução na qualidade do sinal de saída, como indicado na Figura 4 para o codec G.729 e diferentes valores de D_c.

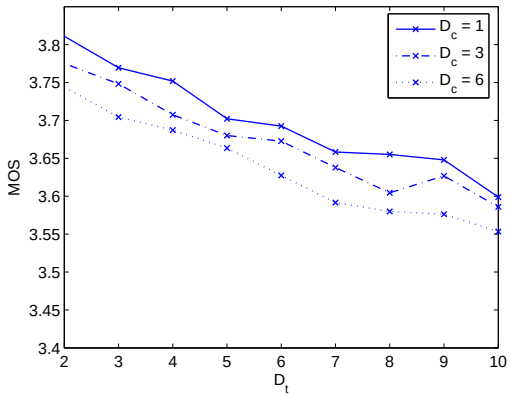

Fig. 4. Teste 2: PESQ-MOS do codec G.729 em função do fator D_t para diferentes valores de D_c.

C. Análise da Vizinhança de T_{op}

Com a introdução do parâmetro D_t, o cálculo da autocorrelação não é realizado para alguns valores de τ. Estes valores são assim automaticamente desconsiderados na estimativa do melhor atraso do dicionário adaptativo, o que causa a queda de qualidade ilustrada na Figura 4. Para atenuar este efeito, é proposto se recalcular $R(\tau)$ no intervalo de tamanho Δ_t em torno da estimativa inicial do atraso. Este processamento acrescenta $2\Delta_t \times \lfloor \frac{80}{D_c} \rfloor$ multiplicações e acarreta um aumento da qualidade do codificador, como ilustrado para o G.729 na Figura 5.

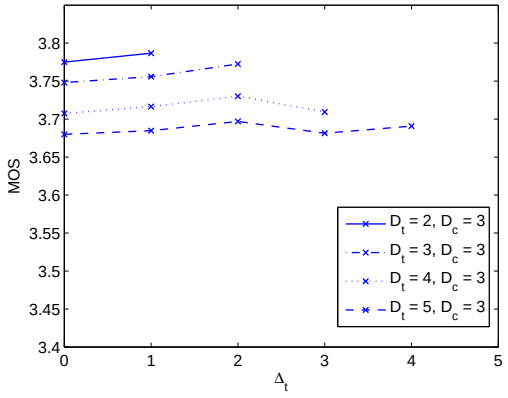

Fig. 5. Teste 3: PESQ-MOS do codec G.729 em função do parâmetro Δ_t para $D_c = 3$ e diferentes valores de D_t.

D. Análise de T_{op} Distantes

Analisando a Figura 5, nota-se que o parâmetro Δ_t não foi capaz de compensar pelas perdas decorrentes das decimações anteriormente introduzidas. Isto indica que o valor de T_{op} determinado originalmente está distante do valor indicado pelo algoritmo simplificado, já que se estivesse na vizinhaça imediata, este valor seria detectado com o uso do parâmetro Δ_t. Assim, uma nova modificação considera o recômputo da autocorrelação em um intervalo Δ_t em torno de diferentes

$T_{\text{op}}^{n_t}$, para $n_t = 1, 2, \ldots, N_t$. Estes valores seriam as N_t melhores estimativas selecionadas no método anterior. Este esquema introduz $2N_t\Delta_t \times \lfloor \frac{80}{D_c} \rfloor$ multiplicações ao esquema com as duas decimações, e causa um aumento da qualidade do sinal reconstituído como indicado na Figura 6.

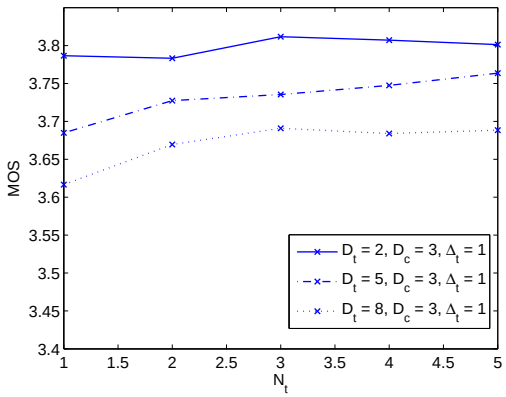

Fig. 6. Teste 4: PESQ-MOS do codec G.729 em função do parâmetro N_t para $D_c = 3$, $\Delta_t = 1$ e diferentes valores de D_t.

V. Comparações Gerais

Os quatro esquemas apresentados na Seção IV foram incorporados conjuntamente ao algoritmo G.729, com os seguintes intervalos de valores:

$$\begin{cases} 1 \leq D_c \leq 10, \\ 1 \leq D_t \leq 10, \\ 0 \leq \Delta_t \leq (D_t - 1), \\ 1 \leq N_t \leq 5. \end{cases} \tag{5}$$

Considerando todas as modificações propostas, a etapa de *open loop* requer os número M de multiplicações e A de adições respectivamente dados por

$$M = \left(\lfloor \frac{124}{D_t} \rfloor + 2N_t\Delta_t \right) \times \lfloor \frac{80}{D_c} \rfloor, \tag{6}$$

$$A = \left(\lfloor \frac{124}{D_t} \rfloor + 2N_t\Delta_t \right) \times \left(\lfloor \frac{80}{D_c} \rfloor - 1 \right). \tag{7}$$

A qualidade do codec G.729 modificado, com os parâmetros variando como definido na equação (5), foi aferida com o avaliador PESQ. O resultado é mostrado na Figura 7 em função do número M de multiplicações requeridas para a configuração em questão. Nesta figura, para uma melhor visualização dos resultados, são mostrados apenas os casos em que $M \leq 4960$.

Ainda na Figura 7, é possível definir um conjunto de configurações com melhor desempenho em termos de qualidade $\times$ complexidade computacional. Estas configurações são caracterizadas na Tabela I, de onde se conclui que as modificações propostas podem, em conjunto, gerar uma significativa redução da complexidade computacional da etapa de *open loop* sem afetar a qualidade do sinal de saída, o que é confirmado por testes subjetivos informais.

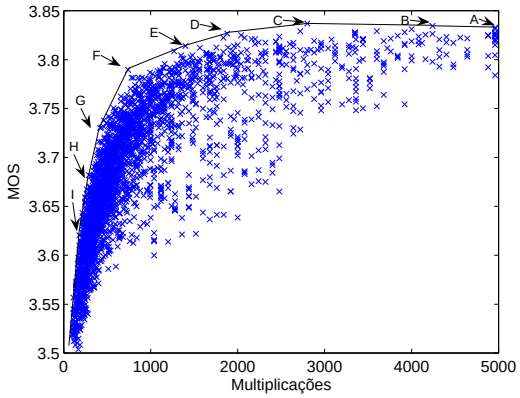

Fig. 7. PESQ-MOS de diferentes configurações do codec G.729 modificado em função do respectivo número de multiplicações. Os casos em destaque delimitam o fecho côncavo para o todo o conjunto de pontos.

TABELA I
CONFIGURAÇÕES DO G.729 MODIFICADO COM MELHOR COMPROMISSO PESQ $\times$ M.

Índice	D_c	D_t	Δ_t	N_t	PESQ	M
A	2	1	0	3	3.83	4960
B	1	6	4	4	3.83	4240
C	2	2	1	4	3.84	2800
D	2	4	2	4	3.83	1880
E	4	2	1	4	3.81	1400
F	4	4	1	3	3.79	740
G	7	5	1	5	3.73	420
H	8	6	1	4	3.68	290
I	8	7	0	1	3.62	180

VI. Conclusões

Quatro modificações foram propostas para acelerar a busca no dicionário adaptativo nas implementações dos codecs G.729 e G.729A. Em conjunto, estas simplificações são capazes de reduzir as operações aritméticas para cerca de 15%–30% na etapa de *open loop*, mantendo a qualidade do sinal codificado e a compatibilidade com um decodificador tradicional.

Referências

[1] 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, EUA, vol. 2, pp. 437–440, 1985.

[2] W. C. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*, Wiley, Hoboken, EUA, 2003.

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

[4] ITU-T Rec. G,729 Annex C, *Reference Floating-Point Implementation for G.729 CS-ACELP 8 kbit/s Speech Coding*, 1998.

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

[6] L. M. da Silva and A. Alcaim, "Modified CELP model with computationally efficient adaptive codebook search," *IEEE Signal Processing Letters*, vol. 2, no. 3, pp. 44–45, Mar. 1995.

[7] M. A. Ramírez and M. Gerken, "Efficient algebraic multipulse search," *Proc. International Telecommunications Symposium*, pp. 231–236, Brazil, Sept. 1998.

[8] L. M. J. Barbosa and L. G. Meloni, "A sequential search algorithm with signal-selected pulse amplitudes," *Proc. International Telecommunications Symposium*, Natal, Brazil, Sept. 2002.

[9] S.-H. Hwang, "Computational improvement for G.729 standard," *Electronics Letters*, vol. 36, no. 13, June 2000.