COPPE/UFRJ

ACELERAÇÃO DOS CODIFICADORES DE FALA G.729 E G.729A

Thiago de Moura Prego

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Sergio Lima Netto

Rio de Janeiro

Agosto de 2009

ACELERAÇÃO DOS CODIFICADORES DE FALA G.729 E G.729A

Thiago de Moura Prego

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Aprovada por:

Prof. Sergio Lima Netto, PhD

Prof. Eduardo Antônio Barros da Silva, PhD

Prof. Abraham Alcaim, PhD

RIO DE JANEIRO, RJ – BRASIL

AGOSTO DE 2009

Prego, Thiago de Moura

Aceleração dos codificadores de fala G.729 e G.729A/Thiago de Moura Prego. – Rio de Janeiro: UFRJ/COPPE, 2009.

XI, 59 p.: il.; 29, 7cm.

Orientador: Sergio Lima Netto

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2009.

Referências Bibliográficas: p. 55 – 56.

1. Codificador de fala. 2. G.729. 3. G.729A. I. Netto, Sergio Lima. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

A Deus pelas bênçãos de cada dia e à minha família pelo suporte dado em todos estes anos.

Agradecimentos

Meus sinceros agradecimentos:

- ao professor Sergio Lima Netto, pela orientação dada durante todo o período do mestrado, pelas oportunidades de participar de projetos bastante interessantes e pela paciência e sabedoria em momentos importantes.
- aos professores Eduardo Antônio Barros da Silva e Abraham Alcaim, por aceitarem o convite de participação na banca de examinação deste trabalho.
- a todas as pessoas que me ajudaram neste projeto por meio de dicas, orientação ou material de estudo.

Thiago de Moura Prego

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ACELERAÇÃO DOS CODIFICADORES DE FALA G.729 E G.729A

Thiago de Moura Prego

Agosto/2009

Orientador: Sergio Lima Netto

Programa: Engenharia Elétrica

Esta dissertação tem o objetivo de introduzir simplificações aos codificadores definidos pela Recomendação ITU-T G.729 (G.729) [1] e pela Recomendação ITU-T G.729 Anexo A (G.729A) [2], de forma a reduzir a complexidade computacional sem que ocorra uma diminuição da qualidade de codificação dos mesmos.

Para dar subsídios técnicos e teóricos ao leitor, fez-se uma descrição detalhada do codificador ITU-T G.729 e das alterações inseridas pelo Anexo A, com ênfase no bloco de busca pela excitação do dicionário adaptativo, pois é o bloco no qual as simplificações propostas neste trabalho foram feitas.

Inicialmente são propostas quatro simplificações na busca do dicionário adaptativo, que, quando combinadas, reduzem o tempo de codificação de um segmento de 30 ms de fala para os codecs G.729 e G.729 em 12% e 9%, respectivamente, sem que haja redução da qualidade de codificação. Em um segundo momento, propõe-se a combinação destas quatro simplificações com duas outras técnicas. A combinação com a primeira técnica resulta em uma redução de 12% e 11% do tempo de codificação e a combinação com a segunda técnica resulta em uma redução de 11% e 11% para os codificadores G.729 e G.729A, respectivamente. A combinação das seis técnicas apresentadas acarreta uma redução de 14% e 12% do tempo de codificação para os codificadores G.729 e G.729A, respectivamente. Para todos os casos não há redução da qualidade percebida do sinal de fala reconstruído.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

ACCELERATION OF SPEECH CODECS G.729 AND G.729A

Thiago de Moura Prego

August/2009

Advisor: Sergio Lima Netto

Department: Electrical Engineering

This thesis aims to introduce simplifications to coders defined by ITU-T Recommendation G.729 (G.729) [1] and the ITU-T Recommendation G.729 Annex A (G.729A) [2], to reduce the computational burden without a decrease in the coding quality.

To give technical and theoretical subsidies to the reader, a detailed description of the ITU-T G.729 encoder and the amendments included in Annex A was made, with emphasis on the adaptive codebook search block, as the simplifications proposed in this work were made at this block.

Initially four simplifications in the adaptive codebook search are proposed which, when combined, reduce the coding time of a 30 ms frame for speech codecs G.729 and G.729 by 12% and 9%, respectively, with no reduction in the coding quality. Then the combination of these four simplifications with two other techniques is proposed. The combination with the first technique results in a reduction of 12% and 11% of the coding time and the combination with the second technique results in a reduction of 11% and 11% for coders G.729 and G.729A, respectively. The combination of all the presented techniques results in a reduction of 14% and 12% in the coding time of a frame for coders G.729 and G.729A, respectively. In all cases there is no reduction in perceived quality of the reconstructed speech signal.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

A cada dia que passa, as pessoas sentem mais necessidade de falar umas com as outras, independente da distância física. Um dos objetivos da área de Telecomunicações é suprir esta necessidade a partir do desenvolvimento de mecanismos e dispositivos para tal necessidade. A evolução destes dispositivos é cada vez mais rápida, o que aumenta ainda mais o interesse das pessoas nesta área, o que estimula mais a evolução e assim sucessivamente. Um dos grandes motivos desta rápida evolução é a revolução digital, em que sinais de diversas naturezas podem ser tratados como sequências de bits, o que torna os computadores pessoais uma ferramenta efetiva de comunicação.

De modo geral, são feitas três etapas para representar um sinal de maneira digital: amostragem, quantização e codificação. A amostragem é o processo de transformar um sinal analógico, isto é, contínuo na amplitude e contínuo no tempo, em um sinal discreto no tempo e contínuo na amplitude. Para discretizar a amplitude, é feita a quantização, processo que mapeia as infinitas possibilidades de amplitude de um sinal contínuo em um conjunto finito de valores pré determinados. Quanto maior a quantidade de níveis de quantização nas quais as amplitudes serão mapeadas, mais fielmente o sinal digital representará o sinal analógico. A codificação é o processo de associar uma sequência (em geral binária) a cada nível do processo de quantização. Para cada tipo de codificação existe certa quantidade de bits necessária para representar o sinal num determinado período de tempo, sendo esta quantidade chamada de taxa de codificação. Sendo assim, podemos comparar codificadores, sendo que aquele que tiver a menor taxa de codificação para uma mesma qualidade resultante

será mais eficiente.

Entre os codificadores de sinais de fala para telefonia que possuem baixa taxa de transmissão (codificação), os que mais se destacam atualmente são os codificadores baseados na técnica CELP (*Code Excited Linear Prediction*) [4]. O fato de apresentarem um bom compromisso entre taxa de transmissão e qualidade de codificação, faz com que estes codificadores sejam amplamente utilizados na área de Telecomunicações. Estes codificadores fazem uso de regressões lineares e dicionários de excitações (conceitos estes apresentados mais adiante) que serão utilizadas para a reprodução fiel de voz. Estes processamentos fazem parte da técnica conhecida por análise-por-síntese (*analysis-by-synthesis*, AbS), que demanda um enorme esforço computacional para ser executada. Devido a isto, existem diversos trabalhos que visam diminuir a complexidade computacional do esquema AbS. Alguns destes trabalhos são:

- *Adaptive encoding of fixed codebook in CELP coders* [5]: este artigo propõe uma implementação adaptativa para o dicionário fixo, partindo do pressuposto que a contribuição deste dicionário é periódica, assim como a do dicionário adaptativo (ou de *pitch*).

- *A fast search method of algebraic codebook by reordering search sequence* [6]: este artigo propõe uma aceleração na busca pela excitação do dicionário fixo do codificador G.729. Para tal, a sequência da busca no dicionário fixo é reordenada e a busca termina quando um determinado limiar de erro médio quadrático (*mean square error*, MSE) entre o sinal alvo e a possível excitação filtrada é atingido.

- *Computational improvement for G.729 standard* [7]: neste artigo uma aceleração na busca pela excitação do dicionário adaptativo é proposta. Para isto, a estimativa do atraso de *pitch*, utilizada na busca do dicionário adaptativo, só é feita caso um limiar do MSE, entre os coeficientes LSP de um segmento e dos coeficientes LSP do segmento anterior, é atingido. Do contrário, reutiliza-se o atraso do segmento anterior.

- *Joint position and amplitude search of algebraic multipulses* [8]: propõe-se, neste artigo, um novo método de obtenção da excitação do dicionário fixo

mais rápido que o dos codificadores G.729 e G.729A. Este método visa obter, de maneira conjunta, as posições e as amplitudes dos pulsos da excitação do dicionário fixo.

- *Iteration-free pulse replacement method for algebraic codebook search* [9]: um algoritmo rápido, sem iterações, para a busca pela excitação do dicionário fixo é proposto neste artigo. O método é composto por duas etapas. Um vetor inicial é determinado através de uma estimativa por verossimilhança. No segundo estágio, os pulsos do vetor inicial são substituídos pelos mais importantes de cada trilha.

Com exceção de [9], os trabalhos citados foram publicados anteriormente ao PESQ (*Perceptual Evaluation of Speech Quality*) [10]. Devido a isto, estes trabalhos não puderam se beneficiar da automação que um avaliador objetivo permite, diferentemente da presente dissertação, que pôde fazer ajustes finos que necessitam de avaliações sucessivas de qualidade e de custo tempo-financeiro proibitivo num cenário pré PESQ.

1.1 Proposta do trabalho

Esta dissertação tem por finalidade introduzir simplificações no codificador de fala baseado na técnica CS-ACELP (*Conjugate-Structure Algebraic Code Excited Linear Prediction*) definido pela Recomendação ITU-T G.729 [1], reduzindo a complexidade computacional do codificador. Estas simplificações são aplicadas à busca no dicionário adaptativo, mais especificamente na etapa de *open-loop*, responsável por encontrar uma primeira estimativa do atraso da excitação do dicionário adaptativo.

As simplificações propostas têm por objetivo diminuir a complexidade computacional do algoritmo (avaliada pelo tempo de codificação e pelo número de multiplicações) sem diminuir a qualidade percebida do sinal de fala reconstruído, medida através de técnicas objetivas e subjetivas.

1.2 Organização da dissertação

O Capítulo 2 descreve o funcionamento do codificador de fala ITU-T G.729, dando ênfase à busca no dicionário adaptativo, além de descrever as alterações efetuadas pelo ITU-T G.729 Anexo A [2]. Este capítulo contém ainda as descrições do método objetivo de avaliação de qualidade de fala PESQ e do método subjetivo de avaliação de qualidade de fala MOS (*Mean Opinion Score*) [3]. O banco de fala BD1 utilizado pelos testes objetivos também é descrito neste capítulo. Com isto, este capítulo provê o insumo técnico-teórico utilizado pelos demais capítulos que compõem esta dissertação.

O Capítulo 3 descreve quatro simplificações do codificador ITU-T G.729 e na sua versão acelerada ITU-T G.729 Anexo A (G.729A). O bloco do codificador em foco é a busca no dicionário adaptativo, mais especificamente o estágio de *open-loop*, em que o período de *pitch* de um determinado segmento do sinal de fala é estimado, sendo utilizado como primeira estimativa do atraso da excitação do dicionário adaptativo. Duas destas simplificações diminuem a complexidade computacional e a qualidade percebida do sinal de fala reconstruído, enquanto as outras duas aumentam estas características. Quando combinadas, estas técnicas diminuem o tempo de codificação dos codificadores G.729 e G.729A em 12% e 9%, respectivamente, sem alterar a qualidade percebida do sinal reconstruído.

O Capítulo 4 investiga a combinação das técnicas apresentadas no Capítulo 3 com dois outros esquemas de aceleração do estágio de *open-loop* na busca pela excitação do dicionário adaptativo apresentados. O primeiro esquema apresentado é proposto nesta dissertação e, quando combinado com as simplificações do Capítulo 3, resultado numa redução do tempo de codificação dos codificadores G.729 e G.729A em 12% e 11%, respectivamente, sem redução da qualidade percebida. O segundo esquema apresentado é proposto por [7] e a combinação deste esquema com as técnicas apresentadas no Capítulo 3 acarreta uma redução do número de multiplicações da etapa de *open-loop* para os codificadores G.729 e G.729A de 82% e 90%, respectivamente, mantendo-se a qualidade percebida. O resultado da combinação das quatro simplificações do Capítulo 3 com as duas técnicas do Capítulo 4 é uma redução de 87% e de 95% no número de multiplicações da etapa de *open-loop* para os codificadores G.729 e G.729A, respectivamente, sem que haja redução da qualidade do sinal

de fala reconstruído.

O Capítulo 5 resume toda a dissertação com comentários a respeito dos resultados obtidos e apresenta uma lista contendo propostas de trabalhos futuros.

Capítulo 2

Descrição do codificador G.729

2.1 Introdução

A recomendação ITU-T G.729 [1] contém a descrição de um algoritmo para codificação de sinais de fala a 8 kbit/s utilizando a técnica de *Conjugate-Structure Algebraic-Code-Excited Linear-Prediction* (CS-ACELP). Este codificador foi desenvolvido para operar com sinais de fala com banda telefônica (segundo a recomendação ITU-T G.712), freqüência de amostragem 8 kHz e codificação PCM linear com 16 bits por amostra. A saída do decodificador possui as mesmas especificações do sinal de entrada.

Em [11], podem ser encontrados dez Anexos (A-J) para o codificador G.729, sendo estes:

- Anexo A: contém modificações no algoritmo do G.729 original que visam diminuir sua complexidade computacional.
- Anexo B: inclui um esquema de compressão de silêncio no algoritmo do G.729 original.
- Anexo C: contém a descrição de algoritmos, em ponto flutuante, para os codificadores G.729 original e G.729 Anexo A. Existe uma versão chamada Anexo C+ que integra o G.729 original com os Anexos B, D e E em um algoritmo, em ponto flutuante, formando um codificador com compressão de silêncio e taxa de transmissão de 6,4, 8 ou 11,8 kbits/s.

- Anexo D: contém a descrição de um algoritmo para codificação de sinais de fala a 6,4 kbits/s baseado na técnica CS-ACELP.
- Anexo E: contém a descrição de um algoritmo para codificação de sinais de fala a 11,8 kbits/s baseado na técnica CS-ACELP.
- Anexo F: inclui a função de compressão de silêncio do Anexo B no algoritmo do codificador G.729 Anexo D.
- Anexo G: inclui a função de compressão de silêncio do Anexo B no algoritmo do codificador G.729 Anexo E.
- Anexo H: contém a descrição de um algoritmo que integra o G.729 original com os Anexos D e E, formando um codificador com taxa de transmissão de 6,4, 8 ou 11,8 kbits/s.
- Anexo I: contém a descrição de um algoritmo que integra o G.729 original com os Anexos B, D e E, formando um codificador com compressão de silêncio e taxa de transmissão de 6,4, 8 ou 11,8 kbits/s.
- Anexo J: contém a descrição de uma extensão do algoritmo G.729 com taxa de transmissão de 8 a 32 kbits. O G.729 Anexo J é também conhecido por G.729.1.

Este trabalho utiliza os algoritmos do G.729 original e do Anexo A [2], contidos no Anexo C [12].

Este capítulo tem por objetivo descrever o codificador G.729, com ênfase no estágio de busca no dicionário adaptativo, e as acelerações introduzidas pelo Anexo A. Na Seção 2.2, descreve-se, de maneira resumida, o codificador e o decodificador do G.729. Na Seção 2.3, explica-se como é feita a busca no dicionário adaptativo, dando-se ênfase ao estágio de *open-loop*. Na Seção 2.4, as alterações feitas pelo Anexo A são descritas, com ênfase nas modificações relativas à busca no dicionário adaptativo. Na Seção 2.5, explica-se o método de avaliação subjetiva MOS (*Mean Opinion Score*), o método de avaliação objetiva PESQ (*Perceptual Evaluation of Speech Quality*) [10] e a relação entre ambos. Na Seção 2.6, descreve-se o banco de voz utilizado nos testes objetivos (baseados no PESQ), assim como a avaliação PESQ-MOS dos codificadores G.729 e G.729 Anexo A.

2.2 Descrição do codificador G.729

2.2.1 Visão geral

Esta subseção visa descrever de maneira resumida o funcionamento do codificador CS-ACELP, que é baseado no modelo de codificação *Code-Excited Linear-Prediction* (CELP) [4] e é descrito na recomendação ITU-T G.729. As etapas do codificador serão explicadas mais detalhadamente na Subseção 2.2.2 e o decodificador será explicado com mais detalhes na Subseção 2.2.3. Segundo [13], segmentos com duração entre 10 e 30 ms de um dado sinal de fala são pseudo-estacionários e a sua geração pode ser modelada como por um filtro de predição linear, geralmente de ordem 10, como mostra a Figura 2.1. Em codificadores baseados no modelo CELP, a excitação deste modelo é determinada através de um processo denominado Análise por Síntese. Neste processo utiliza-se um dicionário adaptativo e um dicionário fixo. Baseado nisto, o codificador opera em segmentos de 10 ms, correspondentes a 80 amostras. Algumas etapas dividem estes segmentos em dois sub-segmentos de 5 ms. Os parâmetros do modelo CELP são:

- coeficientes $a_1, a_2, ..., a_{10}$ do filtro de predição linear.
- índice e ganho do dicionário adaptativo.
- posição das amostras não nulas, seus sinais e o ganho do dicionário fixo.

Os parâmetros do modelo CELP são quantizados, codificados e enviados a cada segmento. Eles são utilizados no decodificador para recuperar a excitação e os parâmetros do filtro de síntese, gerando-se ao final do processo de decodificação o sinal de fala reconstruído $\hat{s}(n)$, como mostra a Figura 2.2.

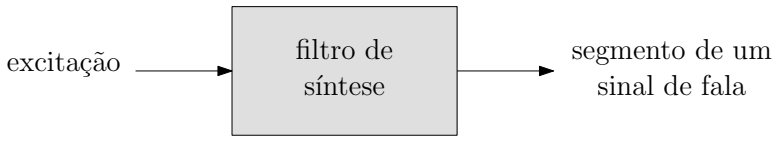


Figura 2.1: Modelo de geração do segmento de um sinal de fala.

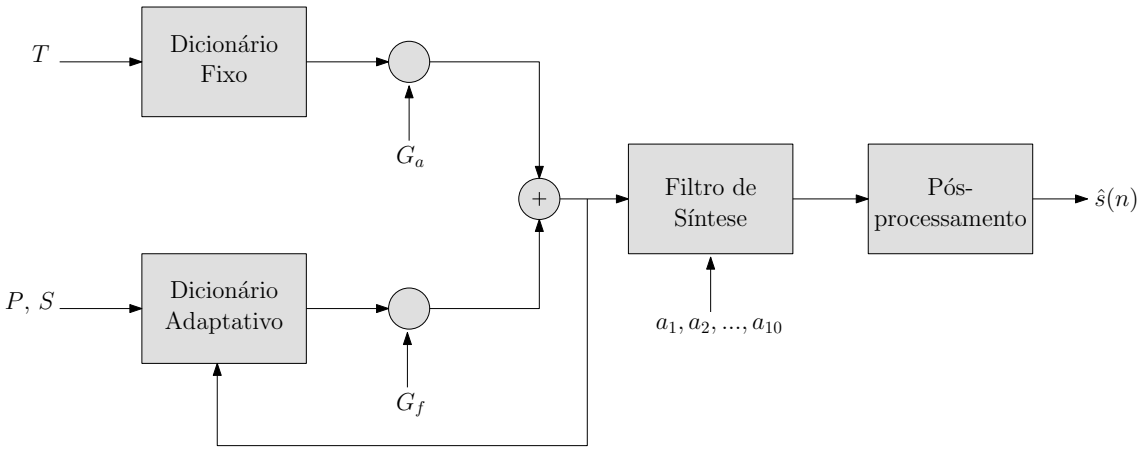


Figura 2.2: Princípio de funcionamento do decodificador do G.729.

2.2.2 Codificador

Na etapa de pré-processamento do codificador G.729, o sinal de entrada é escalado por um fator $\frac{1}{2}$ e processado por um filtro passa-altas. Ambos o escalamento e o filtro passa-altas são implementados pelo filtro da forma:

$$F_{pe}(z) = \frac{0,46363718 - 0,92724705z^{-1} + 0,46363718z^{-2}}{1 - 1,9059465z^{-1} + 0,9114024z^{-2}} \tag{2.1}$$

Esse processamento tem por objetivo eliminar componentes indesejáveis de baixa frequência e evitar o *overflow* na implementação em ponto fixo, respectivamente [1].

Em seguida, a cada segmento de 10 ms, é feita a análise de predição linear (*Linear Prediction*, LP) de ordem 10, que utiliza o filtro LP definido como:

$$\frac{1}{A(z)} = \frac{1}{1 + \sum_{i=1}^{10} a_i z^{-i}}. \tag{2.2}$$

A janela da análise LP é dividida em duas partes: a primeira parte é a metade de uma janela de Hamming e a segunda parte é um quarto de um ciclo de uma função seno. Esta janela é dada por:

$$w_{lp}(n) = \begin{cases} 0,54 - 0,46\cos\left(\frac{2\pi n}{399}\right), & n = 0, ..., 199 \\ \cos\left[\frac{2\pi(n-200)}{159}\right], & n = 200, ..., 239 \end{cases}. \tag{2.3}$$

Esta janela utiliza 120 amostras de segmentos anteriores, 80 amostras do segmento atual e 40 amostras do segmento futuro. O sinal janelado é utilizado no cálculo dos coeficientes do filtro LP, posteriormente convertidos para coeficientes

Line Spectrum Pairs (LSP) e quantizados através de uma quantização vetorial preditiva em dois estágios com 18 bits no total.

Para se determinar a excitação do modelo CELP, utiliza-se a abordagem chamada análise por síntese (*Analysis-by-Synthesis*, AbS), que consiste em escolher a excitação que gera a saída do filtro de síntese mais próxima do sinal-alvo, segundo o critério da máxima correlação. A excitação $u(n)$ do filtro de síntese é determinada a cada sub-segmento de 5 ms e é formada pela soma das excitações $u_a(n)$ e $u_f(n)$, ponderadas pelos seus respectivos ganhos:

$$u(n) = G_a u_a(n) + G_f u_f(n), \tag{2.4}$$

onde a excitação $u_a(n)$ é formada por uma versão atrasada da excitação $u(n)$ do filtro de síntese de segmentos anteriores e a excitação $u_f(n)$ é formada por um vetor que possui apenas quatro amostras não nulas, de amplitude unitária.

Uma vez a cada segmento de 10 ms estima-se o atraso de *pitch* pela análise em *open-loop*, que é utilizado para determinar o atraso do dicionário adaptativo. A etapa de *open-loop* será explicada em detalhes na Subseção 2.3.1. Em seguida, são feitas as seguintes etapas para cada sub-segmento:

- Obtém-se o sinal-alvo $x(n)$ ao processar o resíduo LP $r(n)$ pelo filtro de síntese $\frac{W(z)}{\hat{A}(z)}$, que são dados por:

$$r(n) = s(n) + \sum_{i=1}^{10} a_i s(n-i), \quad n = 0, ..., 39, \tag{2.5}$$

$$\frac{W(z)}{A(z)} = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \frac{1}{A(z)},$$

$$= \frac{\sum_{i=1}^{10} a_i \gamma_1^i z^{-i}}{\left(\sum_{i=1}^{10} a_i \gamma_2^i z^{-i}\right)\left(\sum_{i=1}^{10} a_i z^{-i}\right)}, \tag{2.6}$$

  onde γ_1 e γ_2 são funções do formato espectral do sinal de entrada.

- A análise em *closed-loop* é então feita, a fim de se obter o índice e o ganho do dicionário adaptativo, através da busca em torno do atraso de *pitch* de *open-loop*. Um atraso de *pitch* fracionário com 1/3 de resolução é utilizado.

- O sinal-alvo $x(n)$ é atualizado através da subtração da contribuição (filtrada) do dicionário adaptativo, e este novo sinal-alvo $x'(n)$ é usado na busca do dicionário fixo para encontrar o ganho e a excitação ótima, que é composta por um vetor de 36 amostras nulas e 4 de amplitude unitária. Um dicionário algébrico com 17 bits com uma estrutura de laços aninhados [1] é utilizado para a obtenção das posições e sinais das 4 amostras não nulas.

- Os ganhos dos dicionários adaptativo e fixo são submetidos a uma quantização vetorial com 7 bits.

- Atualiza-se o dicionário adaptativo, incorporando-se a excitação $u(n)$ obtida pela equação (2.4).

A Figura 2.3 mostra o diagrama de blocos do codificador do G.729 e a Tabela 2.1 mostra a quantidade de bits destinada à quantização e à codificação dos parâmetros enviados pelo codificador a cada 10 ms, o que justifica a taxa de bits ser 8 kbits/s.

2.2.3 Decodificador

O princípio do decodificador é mostrado na Figura 2.2. O decodificador recebe as seguintes informações: coeficientes LSP; ganhos e índices do dicionário adaptativo; posições e sinais das amostras e ganhos do dicionário fixo. Para cada segmento de 10 ms, os coeficientes LSP são decodificados, interpolados e convertidos para os coeficientes do filtro LP. Em seguida, para cada sub-segmento de 5 ms os seguintes procedimentos são feitos:

- A excitação $u_a(n)$ do dicionário adaptativo é decodificada.

- A excitação $u_f(n)$ do dicionário fixo é decodificada.

- Os ganhos G_a e G_f dos dicionários adaptativo e fixo, respectivamente, são decodificados.

- A excitação do filtro de síntese $u(n)$ é construída pela adição das excitações dos dicionários adaptativo e fixo escalados pelos seus respectivos ganhos.

- Atualiza-se o dicionário adaptativo com a incorporação de $u(n)$.

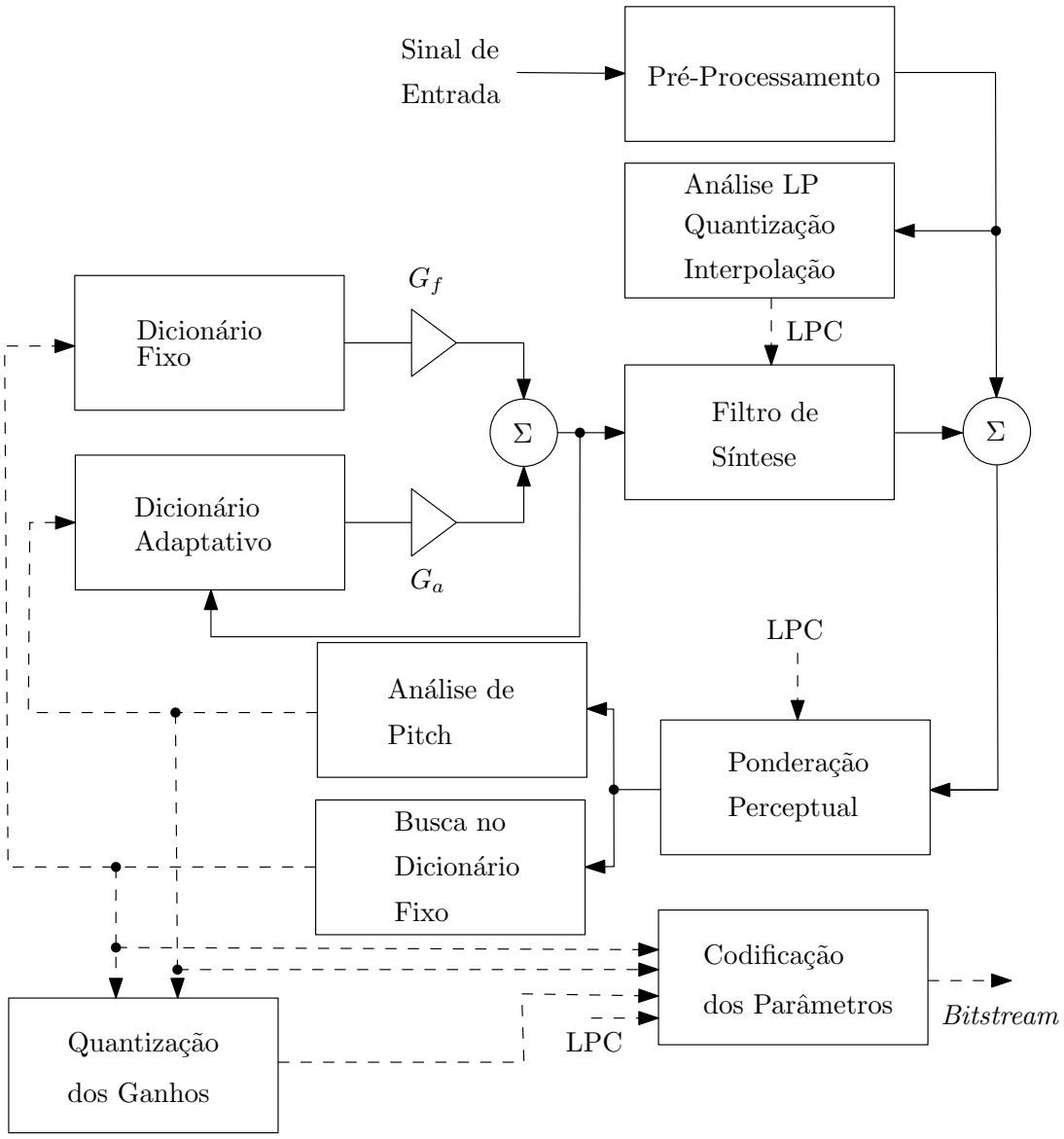


Figura 2.3: Diagrama de blocos do codificador do G.729.

- O sinal de fala é reconstruído processando-se a excitação pelo filtro de síntese.
- O sinal reconstruído de fala é submetido ao estágio de pós-processamento, que inclui um pós-filtro adaptativo, baseado em filtros de síntese de curto-termo e longo-termo, seguidos por um filtro passa-altas e uma operação de escalamento.

2.3 Busca no dicionário adaptativo

Dois parâmetros definem a excitação escolhida pela busca no dicionário adaptativo: o atraso e o ganho. A busca no dicionário adaptativo pode ser dividida em três

Tabela 2.1: Alocação de bits do algoritmo 8 kbit/s CS-ACELP (por segmento de 10 ms).

Paramêtro	Palavra de código	Sub-segmento 1	Sub-segmento 2	Total por segmento
Line spectrum pairs	$L0$, $L1$, $L2$, $L3$			18
Atraso do dicionário adaptativo	$P1$, $P2$	8	5	13
Paridade do atraso de *pitch*	$P0$	1		1
Índice do dicionário fixo	$C1$, $C2$	13	13	26
Sinal do dicionário fixo	$S1$, $S2$	4	4	8
Ganhos dos dicionários (estágio 1)	$GA1$, $GA2$	3	3	6
Ganhos dos dicionários (estágio 2)	$GB1$, $GB2$	4	4	8
Total				80

etapas: análise de *pitch* em *open-loop*, busca em *closed-loop* e cálculo do ganho. A Figura 2.4 mostra o príncipio de funcionamento da busca pelo atraso do dicionário adaptativo.

2.3.1 Análise de *pitch* em *open-loop*

Sejam T_1 e T_2 os atrasos do dicionário adaptativo do primeiro e segundo sub-segmentos de 5 ms. A fim de reduzir a complexidade da busca por estes atrasos, faz-se uma busca pelo atraso de *pitch* T_{op}, pois há uma alta probabilidade de T_1 e T_2 estarem em volta deste. Esta busca é chamada de análise de *pitch* em *open-loop*

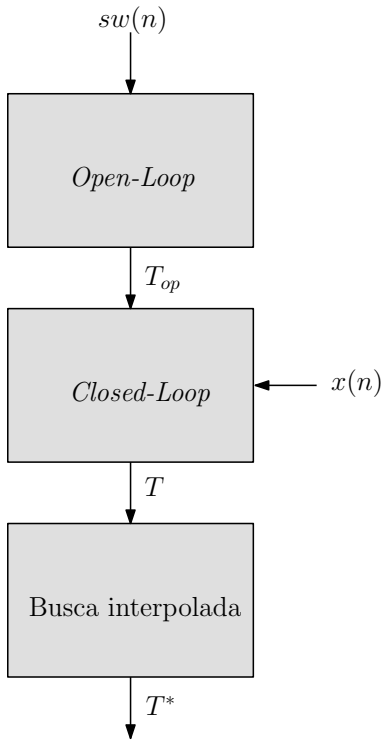


Figura 2.4: Princípio de funcionamento da busca pelo atraso do dicionário adaptativo do G.729.

e é feita uma vez a cada segmento, em que a estimativa de T_{op} utiliza o sinal de fala perceptualmente ponderado (*weighted speech signal*) $sw(n)$, definido por:

$$sw(n) = s(n) + \sum_{i=1}^{10} a_i \gamma_1^i s(n-i) - \sum_{i=1}^{10} a_i \gamma_2^i sw(n-i), \quad n = 0, ..., 39. \tag{2.7}$$

A análise em *open-loop* é feita da seguinte maneira no G.729:

1. Calculam-se três autocorrelações máximas $R(t_i)$, da forma:

$$R(k) = \sum_{n=0}^{79} sw(n) sw(n-k) \tag{2.8}$$

uma para cada intervalo intervalo:

$$\begin{aligned} i = 1 &: \quad k = 20, ..., 39 \\ i = 2 &: \quad k = 40, ..., 79 \\ i = 3 &: \quad k = 80, ..., 143. \end{aligned}$$

2. As autocorrelações máximas $R(t_i)$ são então normalizadas por:

$$R'(t_i) = \frac{R(t_i)}{\sqrt{\sum_n sw^2(n-t_i)}}. \tag{2.9}$$

3. A melhor autocorrelação máxima é escolhida dando-se preferência para os atrasos pertencentes aos intervalos de forma crescente, a fim de evitar a escolha de múltiplos do atraso de *pitch* . Isto é feito através da ponderação [1] das autocorrelações máximas normalizadas $R'(t_i)$.

Uma maneira de medir a complexidade computacional de um processo é baseada no número de operações algébricas que ela exige. O cálculo de T_{op} requer M multiplicações e A adições dadas por:

$$M = 124 \times 80 = 9920, \tag{2.10}$$

$$A = 124 \times 79 = 9796. \tag{2.11}$$

Neste trabalho, o número de multiplicações será utilizado como medida de complexidade computacional, uma vez que o número de adições é altamente correlacionado com o de multiplicações, como mostram as equações (2.10) e (2.11).

2.3.2 Busca em *closed-loop*

Após a obtenção da estimativa do período de *pitch* , é feita a busca pela excitação $u_a(n)$ do dicionário adaptativo a cada sub-segmento de 5 ms, através da maximização da correlação cruzada normalizada entre o sinal alvo $x(n)$ e $y_k(n)$, que é a excitação do filtro de síntese $u(n)$ passada, com um atraso k, convoluída com a resposta ao impulso $h(n)$ do filtro de síntese. A correlação cruzada normalizada é dada por:

$$R_{xy}(k) = \frac{\sum_{n=0}^{39} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{39} y_k(n)y_k(n)}} \tag{2.12}$$

Os valores de k para o cálculo de T_1 são limitados no intervalo $(T_{op} - 3) \leq k \leq (T_{op} + 3)$ e para o cálculo de T_2 são limitados no intervalo $(T_1 - 5) \leq k \leq (T_1 + 4)$. Após a maximização de $R_{xy}(k)$, é feita uma análise para valores fracionários, com resolução 1/3, ao redor dos atrasos T_1 e T_2 escolhidos, resultando em T_1^* e T_2^*.

2.4 G.729 Anexo A

O Anexo A do codificador G.729 introduz as seguintes alterações ao seu algoritmo original:

- O filtro perceptual $W(z)$ utiliza os parâmetros quantizados do filtro LP e é dado por:
$$W(z) = \frac{\hat{A}(z)}{\hat{A}(z/\gamma)}, \tag{2.13}$$
onde o coeficiente de ponderação possui valor fixo de $\gamma = 0,75$ [2].

- A ánalise de *pitch* em *open-loop* é simplificada ao se utilizar decimação no cálculo das autocorrelações do sinal de fala perceptualmente ponderado $sw(n)$. A aceleração na estimativa do período de *pitch* é feita ao se inserir um fator 2 de decimação no cálculo da autocorrelação $R(\tau)$, que passa a ser calculada por:
$$R(k) = \sum_{n=0}^{39} sw(2n)sw(2n-k) \tag{2.14}$$
No intervalo $80 \leq \tau \leq 143$, um fator 2 de decimação dos valores de τ também é inserido, e apenas valores pares são utilizados. Estas modificações alteram o cálculo das operações algébricas para:
$$M = 92 \times 80 = 3680, \tag{2.15}$$
$$A = 92 \times 79 = 3588. \tag{2.16}$$

- A computação da resposta ao impulso do filtro de síntese $\frac{W(z)}{\hat{A}(z)}$, a computação do sinal alvo $x(n)$ e atualização dos estados do filtro de síntese são simplificadas, uma vez que este é reduzido para $\frac{1}{\hat{A}(z)}$.

- O estágio de *closed-loop* é simplificado, pois a busca maximiza a correlação entre a excitação passada e o sinal alvo *backward filtered*. A energia da excitação passada filtrada não é considerada.

- A busca no dicionário fixo é simplificada. Ao invés de usar uma busca de laços aninhados, uma estrutura iterativa em árvore é utilizada [14].

- No decodificador, o pós-filtro de longo-termo é simplificado usando apenas atrasos inteiros.

2.5 Avaliação de qualidade de fala

Esta seção contém uma breve descrição dos métodos de avaliação MOS e PESQ, assim como a maneira com que eles se relacionam.

2.5.1 MOS (*Mean Opinion Score*)

O método de avaliação subjetiva MOS é descrito na recomendação ITU-T P.800 [3], sendo o método subjetivo mais utilizado. A técnica é feita da seguinte forma: reune-se um conjunto de pessoas em que cada sujeito deve escutar um número fixo de frases, avaliar a qualidade de cada sinal de fala e dar uma nota entre 1 e 5, segundo a Tabela 2.2, para cada um destes sinais. A média das notas para cada sinal de fala é o MOS daquele sinal. Este processo é bastante demorado e custoso, algo que fomentou a utilização de uma forma objetiva de avaliar as frases codificadas.

Tabela 2.2: Ecala MOS.

MOS	Qualidade do sinal de fala
5	Excelente
4	Bom
3	Regular
2	Ruim
1	Pobre

2.5.2 PESQ (*Perceptual Evaluation of Speech Quality*)

A recomendação ITU-T P.862 [10] descreve um método objetivo de avaliação, utilizado para estimar a nota MOS de um determinado sinal de fala com banda telefônica. Esta técnica é conhecida por PESQ (*Perceptual Evaluation of Speech Quality*), que estima a nota subjetiva MOS de forma aceitável (com coeficiente de correlação de aproximadamente 0,94 e erro absoluto de até 0,5 na escala MOS) quando o sinal de fala é afetado pelos seguintes processos ou degradações: filtragem; atraso variável; codificação com baixa taxa de bits; erros de canal.

Para calcular o valor PESQ, compara-se o sinal original com um sinal degradado, que é resultado do primeiro ser modificado por um sistema de comunicações. O

resultado obtido pelo cálculo PESQ, sendo este um resultado objetivo, pode ser mapeado na escala de avaliação subjetiva MOS através da seguinte equação [15]:

$$\text{MOS} = 0,999 + \frac{4}{1 + e^{-1,4945 \times \text{PESQ} + 4,6607}} \tag{2.17}$$

2.6 Banco de Voz

Nesta seção, descreve-se o banco de voz BD1 utilizado para todos os testes objetivos através do PESQ. Esse é composto por 40 sinais de fala, codificados com PCM 16-bit, com frequência de amostragem de 8 kHz.

Arquivos no formato WAV contendo diversos sinais de fala foram obtidos do *Open Speech Repository* (OSR) [16] e recortados manualmente por [17], dando origem a 596 arquivos no formato WAV contendo um sinal de fala cada. Destes, quarenta foram selecionados para compor a base de voz para este trabalho.

Os sinais de fala desta base de voz são distribuídos da seguinte maneira:

- 8 sinais de fala na língua chinesa (Mandarim);
- 8 sinais de fala na língua francesa;
- 8 sinais de fala na língua indiana;
- 8 sinais de fala na língua inglesa (Reino Unido);
- 8 sinais de fala na língua inglesa (Estados Unidos da América).

A Figura 2.5 mostra a distribuição dos sinais de fala contidos em BD1, em função da duração em segundos.

O desempenho dos codificadores G.729 e G.729A estimado pelo PESQ utilizando o banco de voz BD1 é de PESQ-MOS 3,85 e 3,74, respectivamente.

2.7 Conclusão

Neste capítulo, foi apresentada uma visão geral sobre o codificador ITU-T G.729, descrevendo o seu funcionamento, além de apresentar as modificações introduzidas pelo Anexo A.

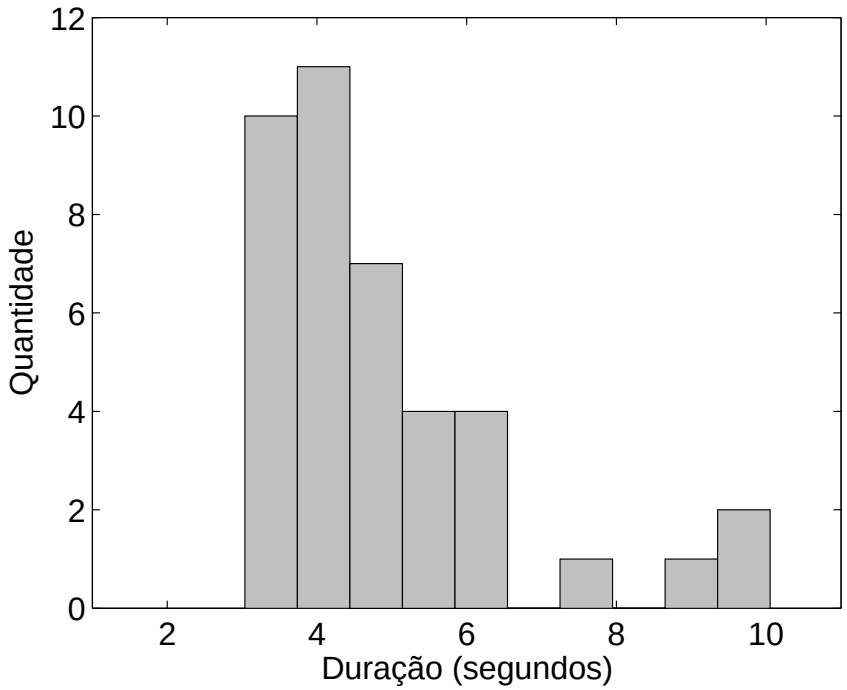


Figura 2.5: Distribuição do sinais de fala de BD1 em função da duração.

O estágio de *open-loop* da busca pela melhor excitação do dicionário adaptativo foi descrito de maneira bastante detalhada pois será amplamente abordado nos capítulos subsequentes.

O banco de voz BD1, utilizado em todos os testes objetivos nos próximos capítulos, foi descrito e o desempenho, para este banco, dos codificadores G.729 e G.729A foi medido através do PESQ.

Capítulo 3

Simplificações Principais

3.1 Introdução

Este capítulo descreve quatro simplificações no algoritmo do codificador ITU-T G.729 e na sua versão acelerada ITU-T G.729 Anexo A. A busca no dicionário adaptativo é o bloco em foco, mais especificamente o estágio de *open-loop*, em que o período de *pitch* de um determinado segmento do sinal de fala é estimado.

Diferentes estratégias são abordadas para as duas versões desse codificador, tendo como meta alcançar um excelente compromisso entre complexidade computacional e qualidade percebida do sinal de fala reconstruído. O resultado é um estágio de *open-loop* menos custoso computacionalmente para ambos os codificadores, mantendo-se a qualidade percebida de cada um, de acordo com as métricas objetiva ITU-T P.862 (PESQ - Perceptual Evaluation of Speech Quality) [10] e subjetiva ITU-T P.800 (CCR - Comparison Category Rating) [3].

A Seção 3.2 descreve de maneira detalhada o fator D_c de decimação no tempo, o fator D_t de decimação do atraso τ, a análise da vizinhança de T_{op} e a análise de estimativas de *pitch* distantes. Na Seção 3.3, descrevem-se os resultados da combinação das quatro modificações propostas tanto em testes objetivos usando a métrica PESQ, quanto em testes subjetivos usando testes CCR. Estas duas seções são apresentadas nas referências [18] e [19], em que a primeira referência faz a combinação no domínio do número M de multiplicações e a segunda referência a faz no domínio do tempo T de codificação de um segmento.

3.2 Modificações Propostas

Nesta seção, são descritas quatro modificações no algoritmo do estágio de *open-loop* dos codificadores G.729 e G.729A. Tais alterações quando combinadas de maneira correta, podem diminuir a complexidade computacional de maneira significativa para esse estágio.

3.2.1 Fator de decimação no tempo

A primeira e natural proposta de aceleração é considerar valores maiores para o fator de decimação $D_c = 2$ na equação (2.14). O cálculo da autocorrelação $R(\tau)$ passa a ser calculada como:

$$R(\tau) = \sum_{n=0}^{\lfloor \frac{79}{D_c} \rfloor} sw(D_c n) sw(D_c n - \tau) \tag{3.1}$$

Com a introdução deste parâmetro, o número de multiplicações necessário para se calcular a função de autocorrelação é igual a $\left(\left\lfloor \frac{79}{D_c} \right\rfloor + 1 \right)$.

A Figura 3.1 ilustra o efeito do uso de diferentes valores para o fator de decimação D_c nos codificadores G.729 e G.729A. É notável que a diminuição da complexidade computacional acompanha uma queda na qualidade percebida.

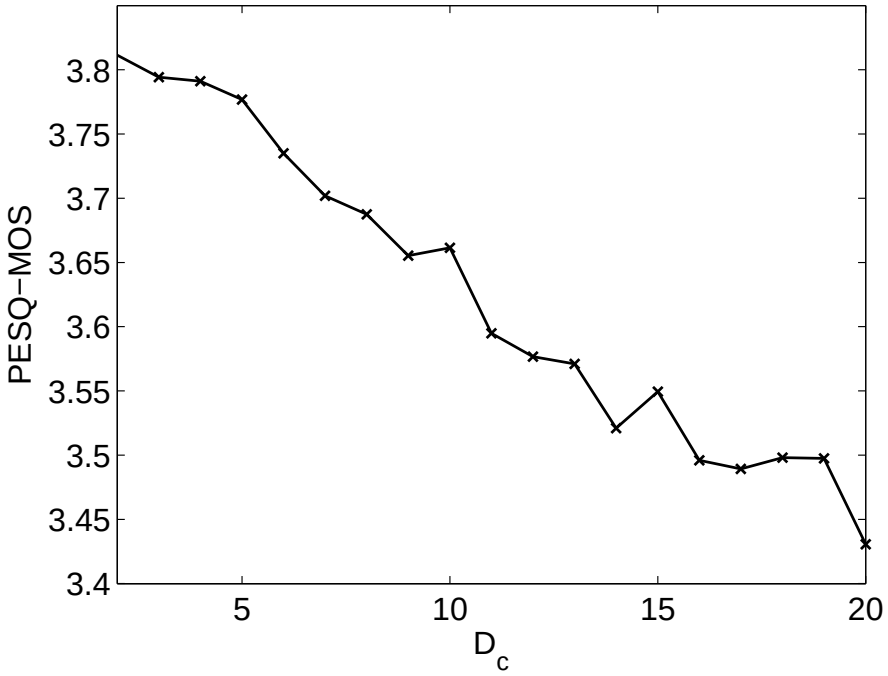
3.2.2 Fator de decimação do atraso τ

O codificador G.729 Anexo A considera a decimação por $D_t = 2$ do *lag* da autocorrelação dentro do intervalo $80 \leq \tau \leq 143$. Desta forma a função $R(\tau)$ só é calculada para valores pares dentro deste intervalo. Uma nova proposta de simplificação considera a decimação de τ por outros valores de D_t para todo o intervalo $20 \leq \tau \leq 143$. Sendo assim, a função de autocorrelação passa a ser calculada apenas para os valores $R(D_t \tau)$. Este esquema, em conjunto com o anterior, reduz o número de multiplicações e de adições para:

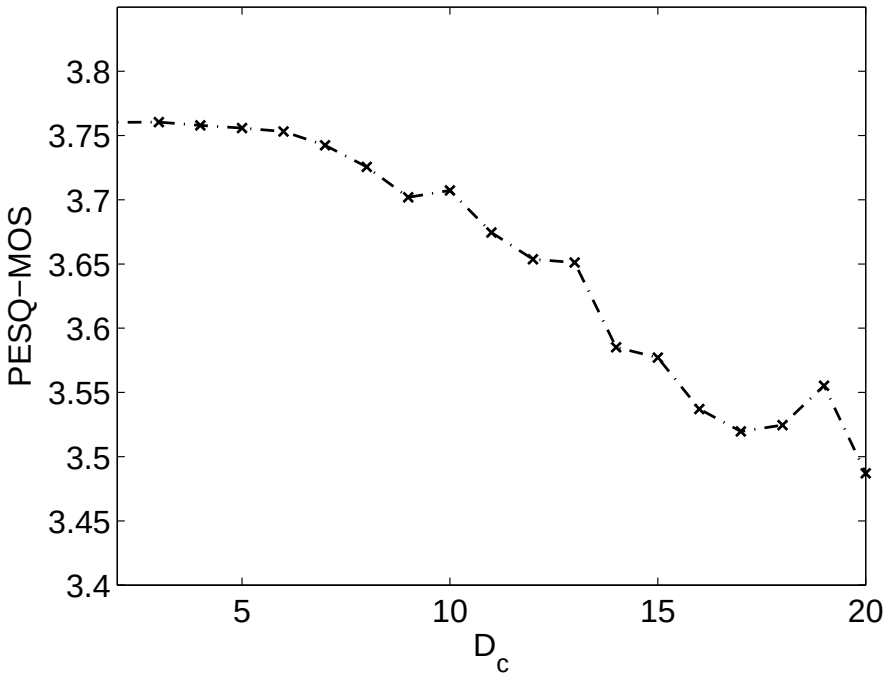
$$M = \left(\left\lfloor \frac{123}{D_t} \right\rfloor + 1 \right) \times \left(\left\lfloor \frac{79}{D_c} \right\rfloor + 1 \right) \tag{3.2}$$

$$A = \left(\left\lfloor \frac{123}{D_t} \right\rfloor + 1 \right) \times \left(\left\lfloor \frac{79}{D_c} \right\rfloor \right). \tag{3.3}$$

É natural que esta modificação também reduza a qualidade percebida conforme o fator D_t aumenta, como mostra a Figura 3.2 para diferentes valores de D_c.

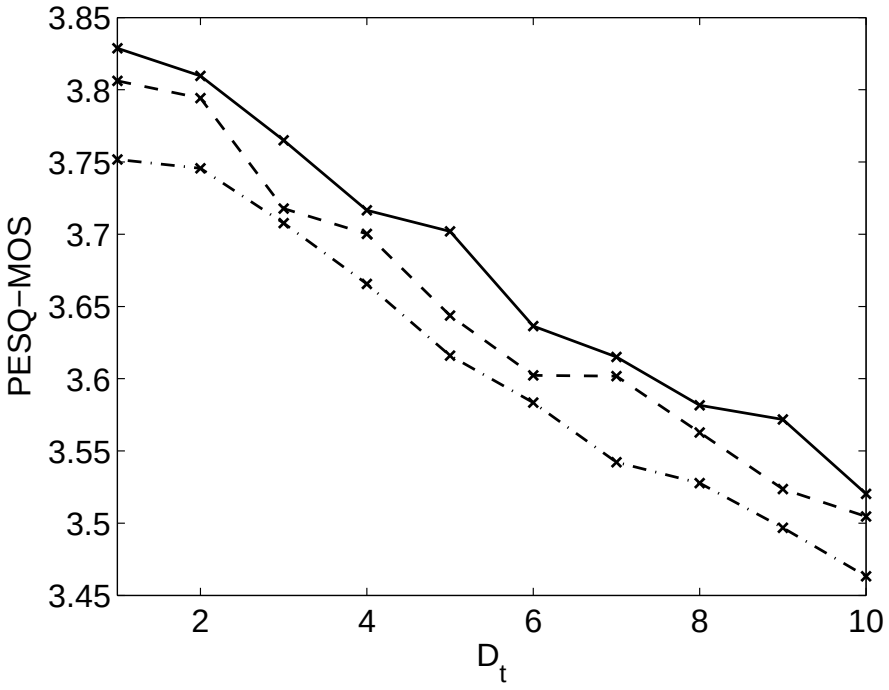


(a)

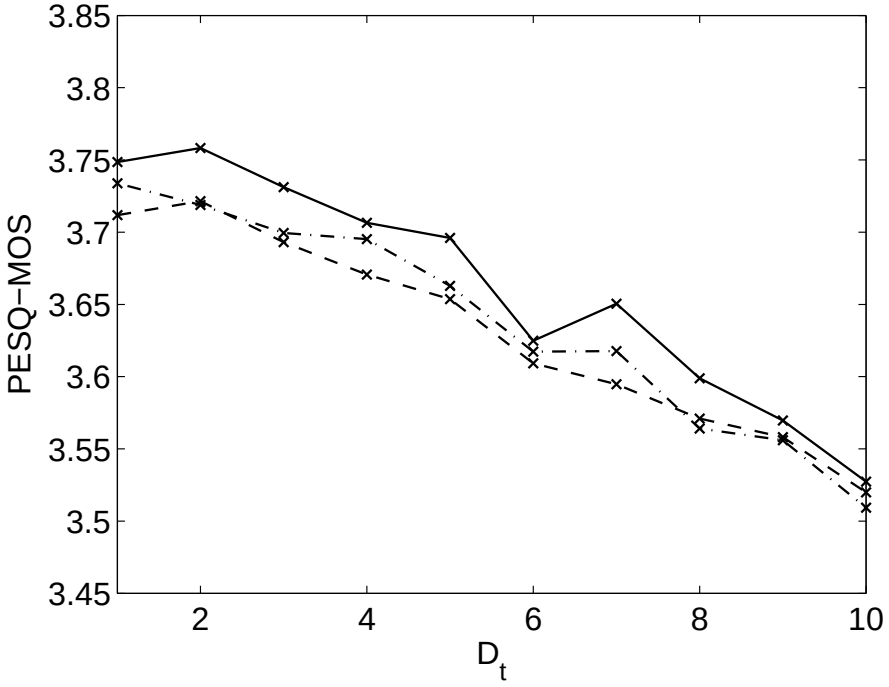


(b)

Figura 3.1: PESQ-MOS dos codecs (a) ITU-T G.729 e (b) ITU-T G.729A em função do fator de decimação D_c.



(a)



(b)

Figura 3.2: PESQ-MOS em função do fator de decimação D_t para: (a) ITU-T G.729 $D_c = 2$ (linha contínua), $D_c = 4$ (linha tracejada) e $D_c = 6$ (linha tracejada e pontilhada); (b) ITU-T G.729A $D_c = 2$ (linha contínua), $D_c = 8$ (linha tracejada e pontilhada) e $D_c = 10$ (linha tracejada).

3.2.3 Análise da vizinhança de T_{op}

Ao introduzir o parâmetro D_t no algoritmo do *open-loop*, alguns valores de τ não são levados em conta no cálculo da função de autocorrelação, pois são automaticamente desconsiderados na estimativa do melhor atraso do dicionário adaptativo (segundo o critério de maximização da função de autocorrelação). Isto causa a queda de qualidade ilustrada na Figura 3.2.

Com o intuito de atenuar este efeito, propõe-se calcular $R(\tau)$ em um intervalo de módulo Δ_t ao redor da estimativa inicial de T_{op}. Este processamento acrescenta $2\Delta_t \times (\left\lfloor \frac{79}{D_c} \right\rfloor + 1)$ multiplicações e $2\Delta_t \times (\left\lfloor \frac{79}{D_c} \right\rfloor)$ adições e gera um aumento na qualidade percebida do sinal reconstruído, como indica a Figura 3.3.

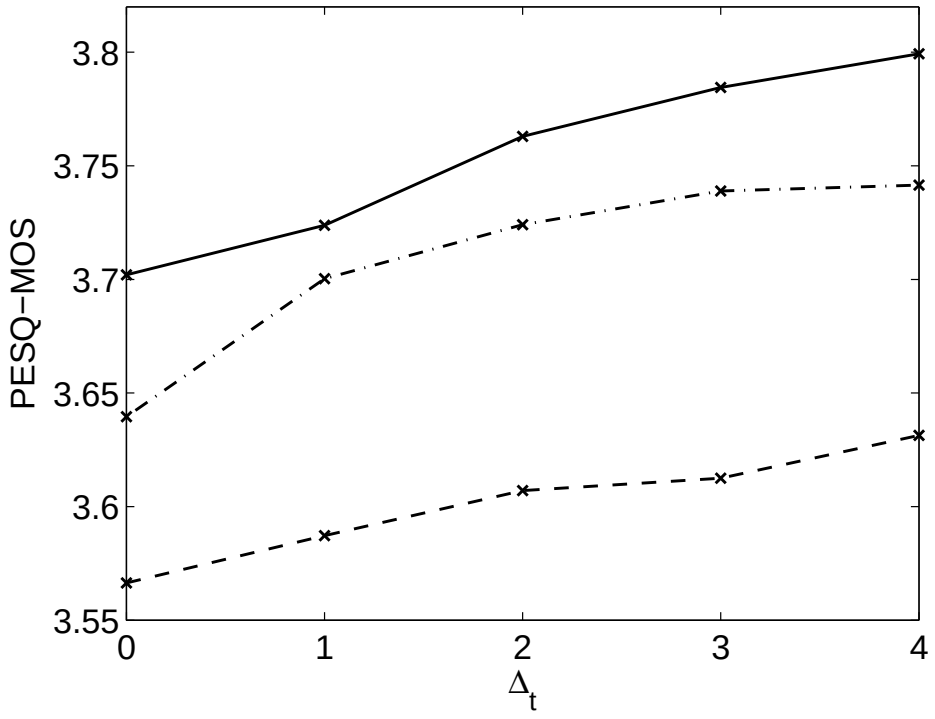
3.2.4 Análise de estimativas de *pitch* distantes

Observando a Figura 3.3 verifica-se que o parâmetro Δ_t não foi capaz de compensar totalmente a degradação inserida pelas decimações anteriores. Isto indica que o valor de T_{op} determinado pelos codificadores G.729 e G.729A originais está distante do valor estimado pelo algoritmo contendo as modificações anteriores, uma vez que caso estivesse próximo (na vizinhança de módulo Δ_t), tal valor seria detectado pela busca introduzida na Subseção 3.2.3.

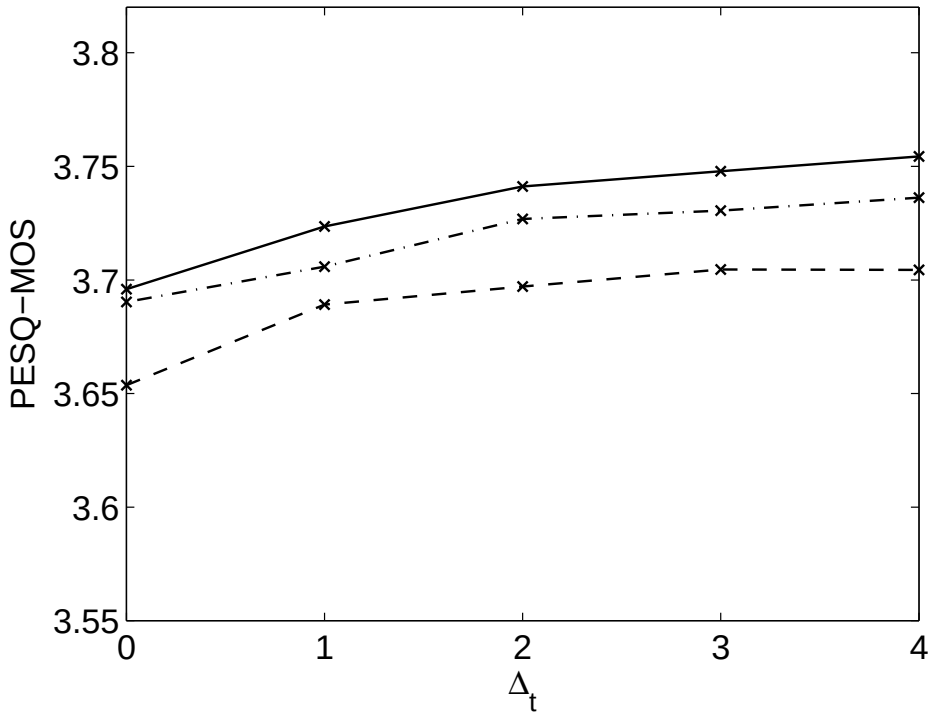
Para solucionar este problema, uma nova modificação seleciona as N_t melhores estimativas de T_{op} e aplica a análise de vizinhança de módulo Δ_t para cada uma delas. O método, então, escolhe a melhor autocorrelação dentre as $N_t \times (2\Delta_t + 1)$ candidatas de cada sub-intervalo, o que gera um aumento da complexidade computacional e da qualidade estimada pelo PESQ, como indica a Figura 3.4.

3.3 Combinando as quatro técnicas

Esta seção analisa o desempenho da combinação dos quatro esquemas apresentados nas subseções 3.2.1–3.2.4. Ao introduzirmos os parâmetros D_c, D_t, Δ_t e N_t ao algoritmo dos codificadores G.729 e do G.729A, o nível de redução da complexidade computacional passa a depender da combinação destes quatro parâmetros. Com isto, o número M de multiplicações e o número A de adições passam a ser dados

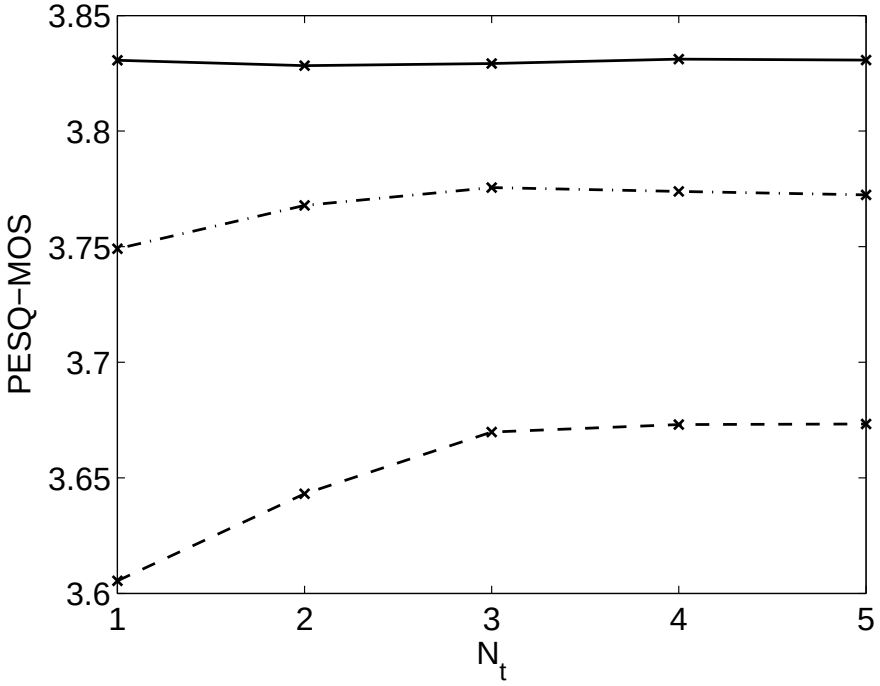


(a)

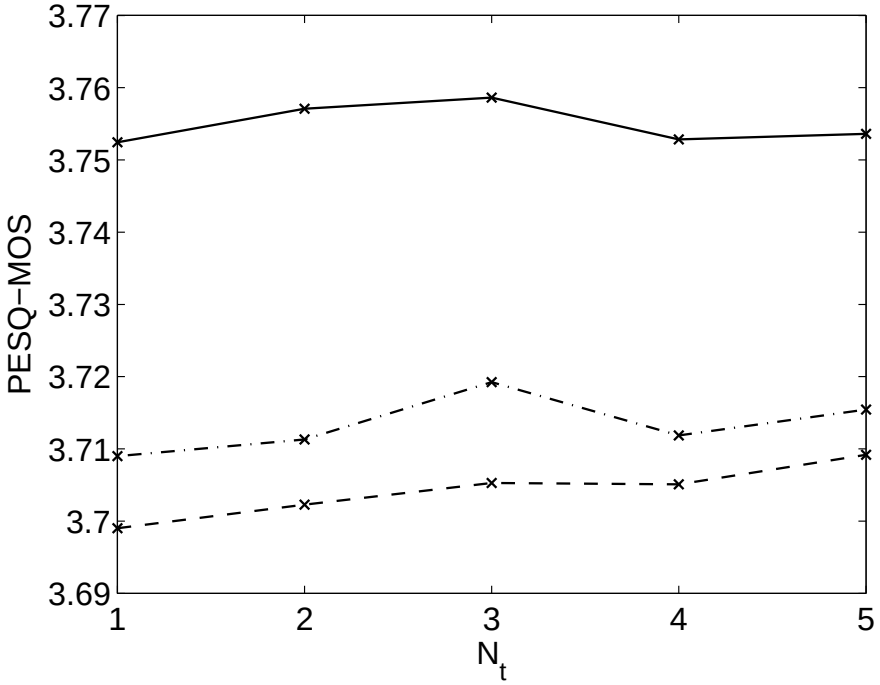


(b)

Figura 3.3: PESQ-MOS em função do fator de decimação Δ_t para: (a) ITU-T G.729 e (b) ITU-T G.729A, onde $D_t = 5$ e $D_c = 2$ (linha contínua), $D_c = 5$ (linha tracejada e pontilhada) e $D_c = 10$ (linha tracejada).



(a)



(b)

Figura 3.4: PESQ-MOS em função do parâmetro N_t para: (a) ITU-T G.729 e (b) ITU-T G.729A, onde $D_c = 2$, $D_t = 2$ e $\Delta_t = 1$ (linha contínua), $D_c = 8$, $D_t = 7$ e $\Delta_t = 5$ (linha tracejada e pontilhada) e $D_c = 10$, $D_t = 8$ e $\Delta_t = 7$ (linha tracejada).

por:

$$M = (\left\lfloor \frac{123}{D_t} \right\rfloor + 1 + 6\Delta_t N_t) \times (\left\lfloor \frac{79}{D_c} \right\rfloor + 1), \tag{3.4}$$

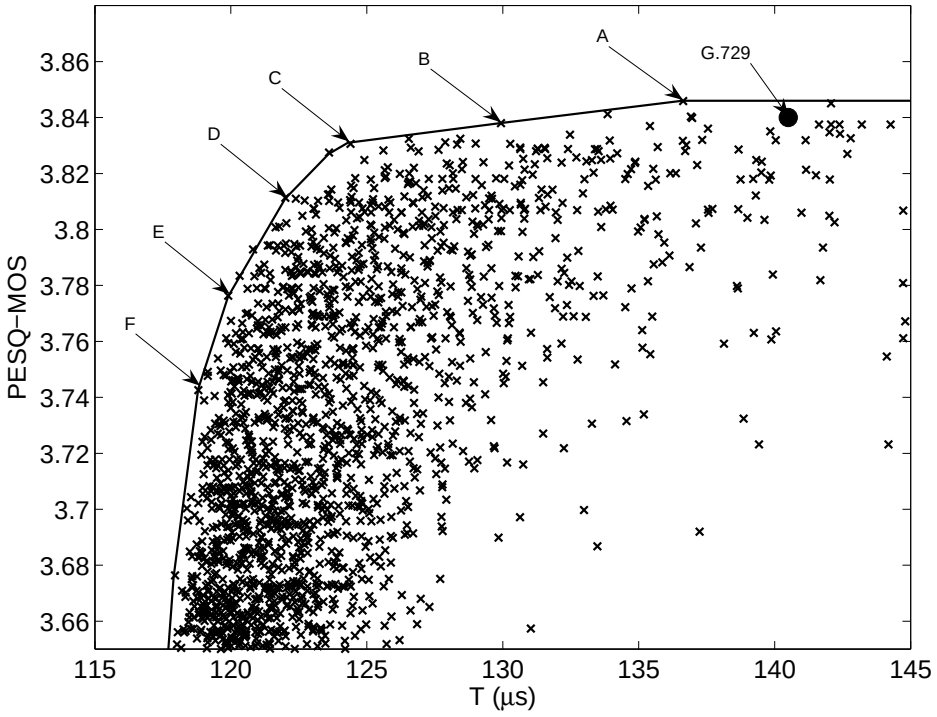
$$A = (\left\lfloor \frac{123}{D_t} \right\rfloor + 1 + 6\Delta_t N_t) \times (\left\lfloor \frac{79}{D_c} \right\rfloor). \tag{3.5}$$

Os quatro esquemas apresentados anteriormente foram incorporados simultaneamente nos dois codificadores utilizando os seguintes intervalos:

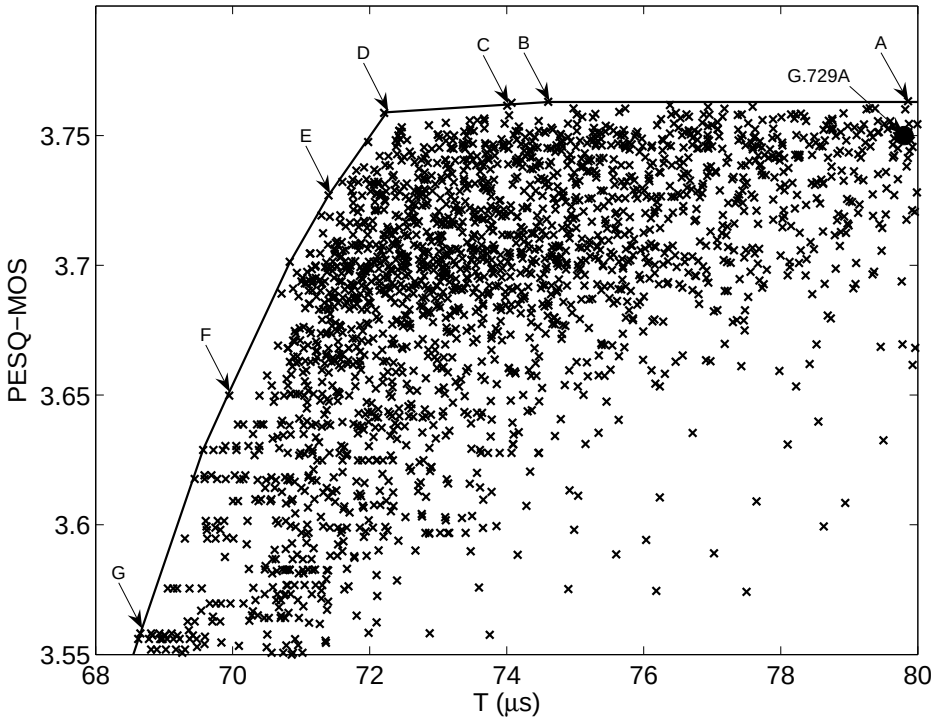
$$\begin{cases} 1 \leq D_c \leq 10 \\ 1 \leq D_t \leq 10 \\ 0 \leq \Delta_t \leq D_t - 1 \\ 1 \leq N_t \leq 5 \end{cases}. \tag{3.6}$$

Neste teste, o tempo T de codificação de um segmento foi utilizado para determinar a complexidade computacional de uma dada configuração $[D_c, D_t, \Delta_t, N_t]$. Os processos de codificação e de decodificação foram executados em uma máquina AMD Turion MK-36, 2.01 GHz com 480 MB RAM. O algoritmo dos codificadores G.729 e G.729A disponibilizado pela ITU-T em [11] foi implementado em C ANSI, onde o codificador (decodificador) pode ser dividido em três etapas: leitura de um arquivo armazenado no disco rígido, codificação (decodificação) e escrita em um arquivo no disco rígido. O tempo T se refere somente à etapa de codificação. O tempo T de codificação é calculado como a quantidade n_{clk} de *clocks* necessários para a execução da etapa de codificação multiplicada pelo tempo T_{clk} de cada *clock*. Para medir a quantidade n_{clk} de *clocks*, utilizou-se o programa *AMD CodeAnalyst Performance Analyzer* [20], disponibilizado gratuitamente para ser baixado.

As notas PESQ-MOS resultantes, para o banco BD1, em função do tempo T de codificação de um segmento, para os codificadores G.729 e G.729A são exibidas na Figura 3.5. Alguns pontos do fecho côncavo (melhor compromisso qualidade/complexidade computacional) são destacados e detalhados nas Tabelas 3.1 e 3.2, que também incluem o número de multiplicações M. Estes resultados indicam que as modificações propostas podem coletivamente reduzir o tempo total de codificação em cerca de 12% (configuração C na Tabela 3.1) ou 9% (configuração D na Tabela 3.2) se comparado com as implementações originais dos codecs, mantendo o nível de qualidade do sinal de fala reconstruído.



(a)



(b)

Figura 3.5: PESQ-MOS em função do tempo T de codificação de um segmento para diferentes configurações dos codificadores modificados. O círculo grande representa o codificador original: (a) ITU-T G.729 e (b) ITU-T G.729A.

A qualidade dos sinais de fala obtidos pelos codificadores modificados (G.729 com a configuração C e G.729A com a configuração D) também foi aferida subjetivamente através de um teste do tipo *comparison category rating* (CCR), como descrito na recomendação ITU-T P.800 [3]. Neste teste, 32 sinais diferentes de fala na língua portuguesa do Brasil foram codificados/decodificados pelas versões originais (sinal original) e modificadas (sinal modificado) dos codificadores G.729 e G.729A. Então, 25 ouvintes foram instruídos a comparar diretamente os sinais modificados com seus sinais originais equivalentes seguindo uma ordem aleatória. Em cada comparação,

Tabela 3.1: Características das configurações do G.729 modificado com o melhor compromisso PESQ-MOS $\times$ T selecionadas na Figura 3.5(a).

Ponto	D_c	D_t	Δ_t	N_t	MOS	M	$T(\mu s)$
G.729	1	1	0	1	3,84	9920	140,5
A	2	10	9	3	3,85	7000	136,6
B	2	6	5	3	3,84	4440	129,9
C	2	2	1	1	3,83	2720	124,3
D	2	4	3	1	3,81	1960	122,0
E	4	4	3	1	3,78	980	119,9
F	4	5	2	1	3,74	740	118,8

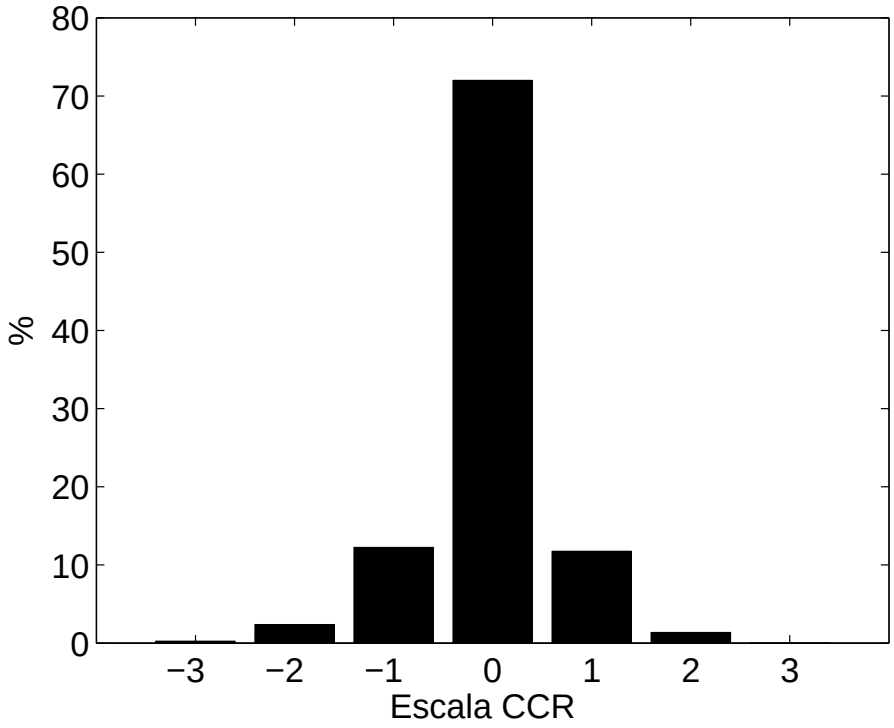
Tabela 3.2: Características das configurações do G.729A modificado com o melhor compromisso PESQ-MOS $\times$ T selecionadas na Figura 3.5(b).

Ponto	D_c	D_t	Δ_t	N_t	MOS	M	$T(\mu s)$
G.729A	2	1	0	1	3,75	3680	79,8
A	2	7	4	3	3,76	3600	79,9
B	4	4	2	3	3,76	1340	74,6
C	3	3	2	1	3,76	1458	74,0
D	5	4	2	1	3,76	688	72,3
E	7	5	3	1	3,73	516	71,4
F	3	7	0	1	3,65	486	69,9
G	10	9	0	1	3,56	112	68,7

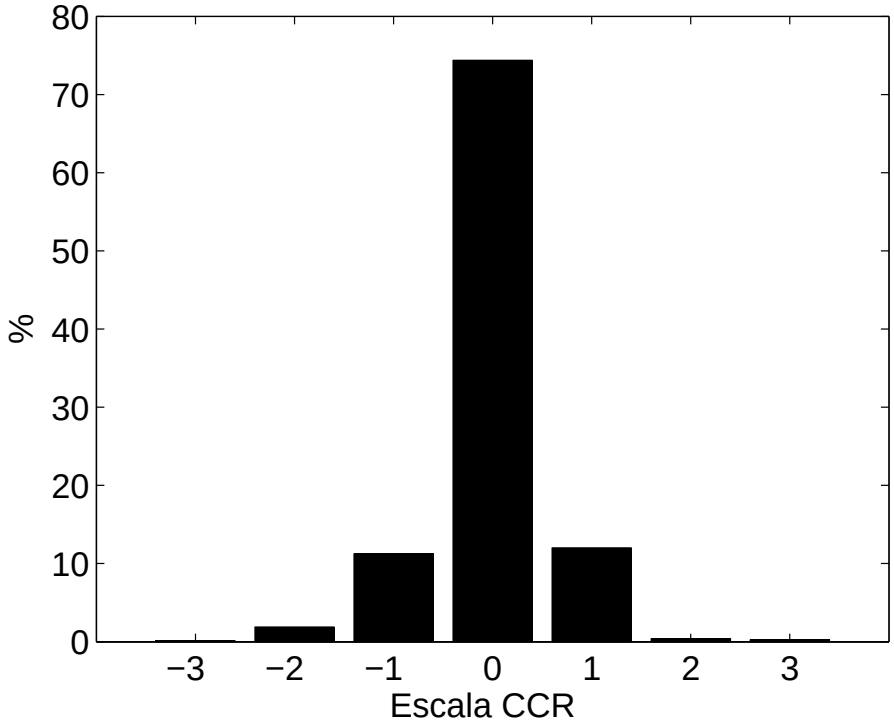
deu-se uma nota em resposta as perguntas “Qual sinal possui melhor qualidade? O quão melhor?” segundo a Tabela 3.3 [3], onde os sinais + e − favorecem os sinais modificado e original, respectivamente. A Figura 3.6 mostra o resultado do teste CCR e indica, para fins práticos, equivalência completa entre as versões modificadas e originais para ambos os codificadores, em relação à qualidade percebida do sinal de fala reconstruído.

Tabela 3.3: Descrição da escala do teste CCR [3]: Notas positivas favorecem o codificador modificado.

Escala CCR	Significado
+3	Muito Melhor
+2	Melhor
+1	Pouco Melhor
0	Igual
−1	Pouco Pior
−2	Pior
−3	Muito Pior



(a)



(b)

Figura 3.6: Histogramas do teste CCR: (a) G.729 (configuração C na Tabela 3.1) original × modificado; (b) G.729A (configuração D na Tabela 3.2) original × modificado.

3.4 Conclusão

Quatro modificações foram propostas com o intuito de acelerar o estágio de *open-loop* na busca pela excitação $u_a(n)$ do dicionário adaptativo dos codificadores G.729 e G.729A, mantendo total compatibilidade com os decodificadores originais. Quando postas juntas, as simplificações propostas podem reduzir o tempo T de codificação em aproximadamente 12% e 9%, respectivamente, sem afetar a qualidade do sinal de fala reconstruído, como confirmado pelas métricas objetiva e subjetiva.

Capítulo 4

Simplificações Secundárias

4.1 Introdução

Este capítulo investiga a combinação das técnicas apresentadas no Capítulo 3 com dois esquemas de aceleração da busca pela excitação do dicionário adaptativo apresentados pela Seção 4.2 e por [7].

A Seção 4.2 apresenta uma nova proposta de simplificação do estágio de *open-loop* através da fixação do atraso τ no intervalo $80 \leq \tau \leq 143$. Investiga-se a combinação desta técnica com a inserção dos parâmetros D_c, D_t, Δ_t e N_t. O resultado desta combinação é semelhante ao observado no Capítulo 3 para o codificador G.729. Para o codificador G.729A, o resultado é uma redução de aproximadamente 11% do tempo T de codificação mantendo-se a qualidade percebida, contra 9% observado no Capítulo 3.

A Seção 4.3 apresenta a proposta feita por [7], em que o valor de T_{op} é recalculado se um pré-requisito for preenchido, do contrário reaproveita-se a estimativa do segmento anterior. Estuda-se a combinação desta técnica com a inserção dos parâmetros D_c, D_t, Δ_t e N_t, que resulta em uma redução de 11% no tempo T de codificação de um segmento para o codificador G.729 e de 11% para o codificador G.729A.

Na Seção 4.4, é apresentado o resultado da combinação das 6 técnicas apresentadas nesta dissertação. Consegue-se uma redução do tempo T de codificação de um segmento de 14% e 12% para os codificadores G.729 e G.729A, respectivamente, sem que ocorra redução da qualidade do sinal de fala reconstruído, confirmado por

testes subjetivos informais.

4.2 Estimativa fixa de pitch para um dado intervalo

Nesta seção, apresentamos uma nova técnica simplificada de busca no dicionário adaptativo dos codificadores G.729 e G.729A. Mais adiante, esta nova proposta será combinada às propostas vistas no Capítulo 3 e o efeito disto na complexidade computacional e na qualidade resultantes será investigado.

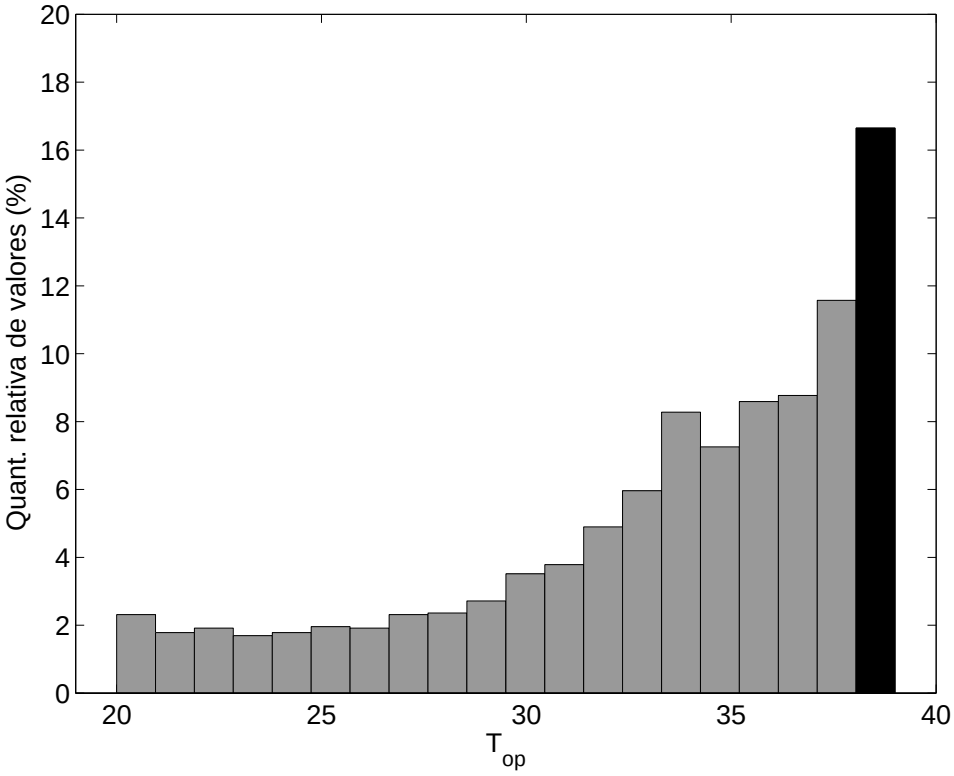
Em um experimento inicial, foi verificada a importância de cada intervalo de τ no cálculo da função de autocorrelação $R(\tau)$, onde: (a) $20 \leq \tau \leq 39$, (b) $40 \leq \tau \leq 79$, (c) $80 \leq \tau \leq 143$. Neste estudo, usou-se $T_i = \bar{\tau}_i$ em um dado intervalo $i = a, b, c$, e a busca por T_{op} foi feita de forma padrão nos outros dois intervalos. Assim foram definidas três configurações para o G.729 e o G729A modificados:

- 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.

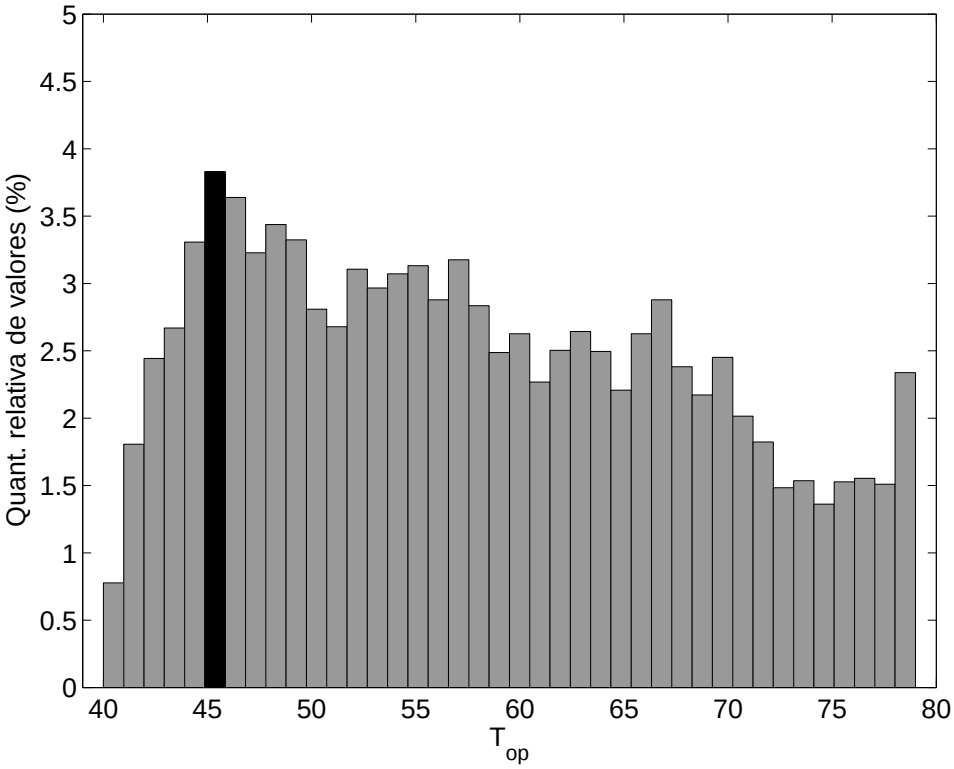
A importância de cada intervalo foi avaliada pela perda de qualidade do sinal de voz quando T_i foi pré-determinado no intervalo i: quanto maior a perda de qualidade do sinal, maior a importância do intervalo.

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. Os histogramas de cada T_{op} para cada intervalo são mostrados nas Figuras 4.1 e 4.2 para os codecs G.729 e G.729A, respectivamente, de onde se determina que:

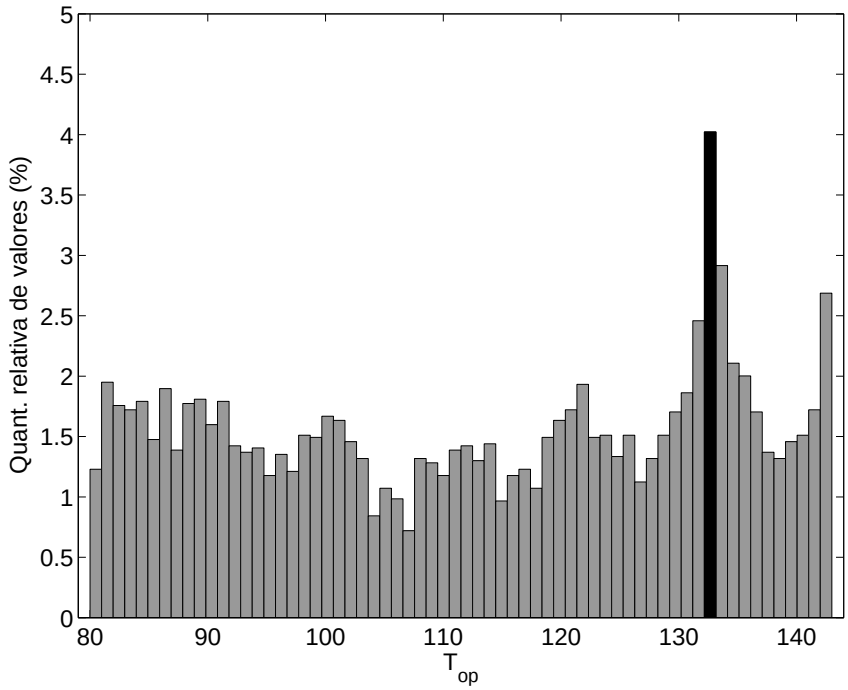
$$\text{G.729} \begin{cases} \bar{\tau}_a = 39 \\ \bar{\tau}_b = 45 \\ \bar{\tau}_c = 133 \end{cases} \tag{4.1}$$


Quant. relativa de valores (%)
0
2
4
6
8
10
12
14
16
18
20
20
25
30
35
40
T op

(a)

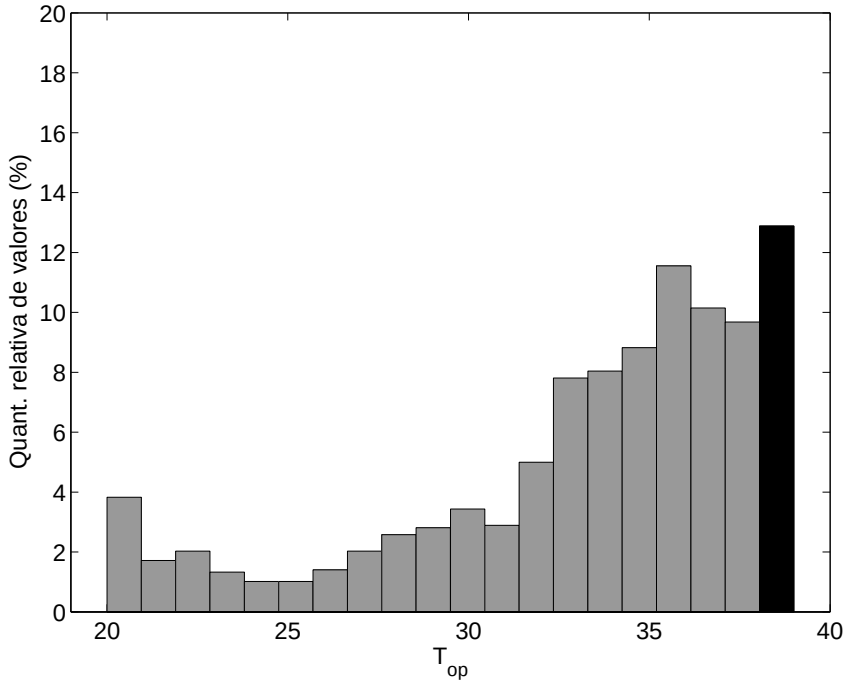

Quant. relativa de valores (%)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
40
45
50
55
60
65
70
75
80
T op

(b)

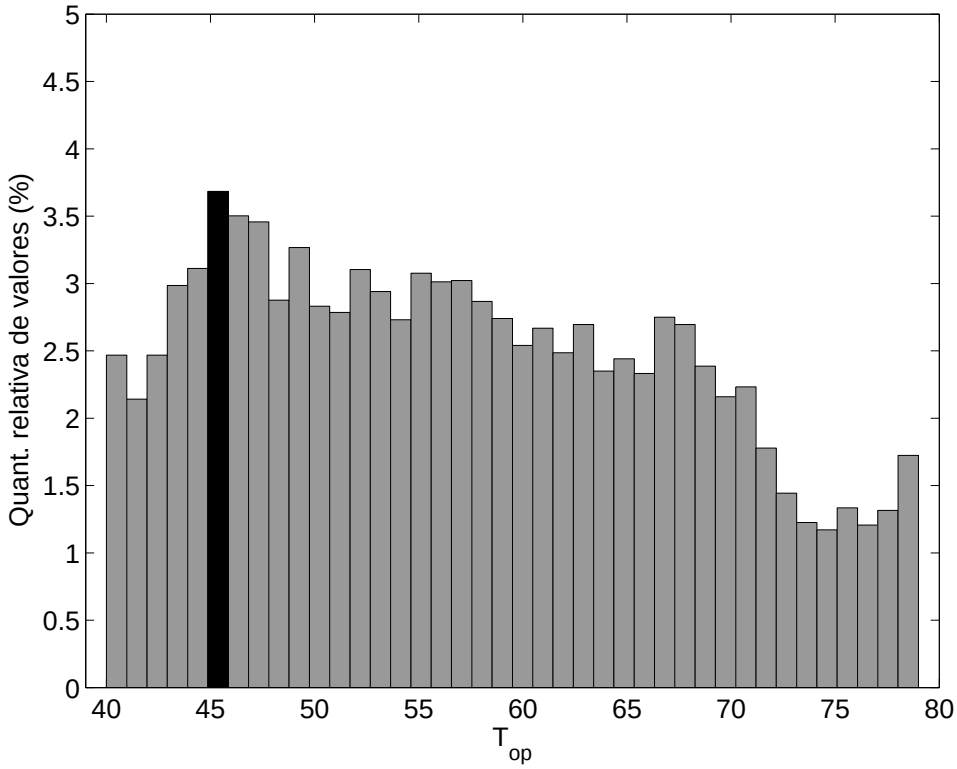


(c)

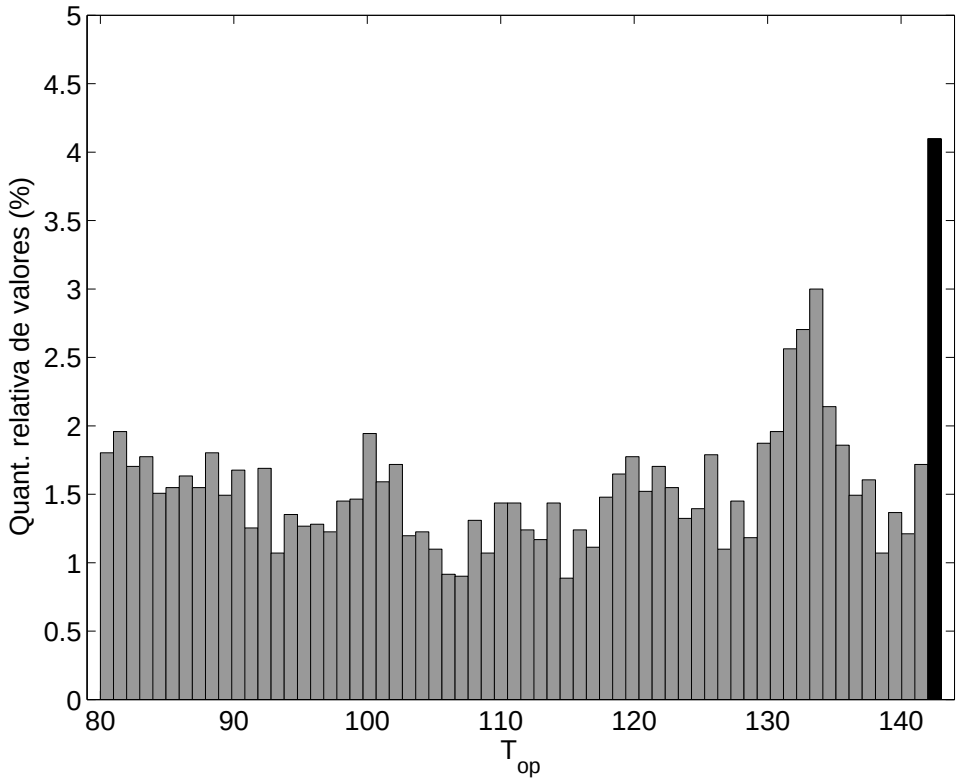
Figura 4.1: Histograma (com destaque para a moda) da variável T_{op} para o codificador G.729 nos intervalos: (a) $20 \leq T_{op} \leq 39$; (b) $80 \leq T_{op} \leq 79$; (c) $80 \leq T_{op} \leq 143$.



(a)



(b)



(c)

Figura 4.2: 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) $80 \leq T_{op} \leq 143$.

$$\text{G.729A} \begin{cases} \bar{\tau}_a = 39 \\ \bar{\tau}_b = 45 \\ \bar{\tau}_c = 143 \end{cases} \tag{4.2}$$

As Tabelas 4.1 e 4.2 mostram a distribuição de T_{op} em cada intervalo e a avaliação PESQ média para todo o banco BD1 para cada configuração dos codecs G.729 e G.729A, respectivamente. As colunas (a)%, (b)% e (c)% representam o percentual de T_{op} pertencentes aos intervalos (a), (b) e (c), respectivamente, para todos os sinais do banco BD1. Com base nestas tabelas, 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 PESQ dos codificadores G.729 e G729A padrões, enquanto a complexidade computacional requerida pelo cálculo de $R(\tau)$ é decrescida de 54% e 35%, respectivamente. Nestes casos, o intervalo (b) parece suprir toda a deficiência provocada ao se fazer T_c fixo.

Tabela 4.1: Distribuição de T_{op} e avaliação PESQ para o banco de sinais de voz BD1 para cada configuração do codificador G729 modificado.

Configuração	(a)%	(b)%	(c)%	PESQ
G.729	11,58	59,07	29,35	3,84
I	3,27	64,93	31,80	3,78
II	14,93	7,53	77,54	3,65
III	17,98	76,52	5,50	3,83

Tabela 4.2: 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

4.2.1 Combinação com as técnicas do Capítulo 3

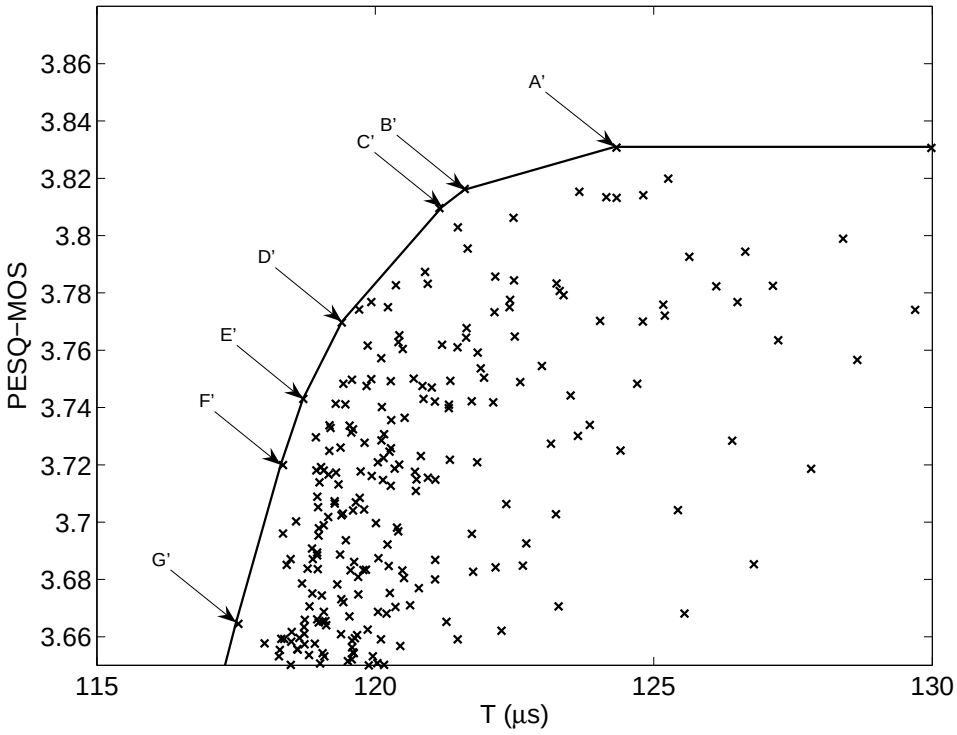
Nesta subseção, investiga-se a combinação da técnica fixar o valor de T_c com os parâmetros D_c, D_t, Δ_t e N_t. Verifica-se a maneira com que estes métodos todos trabalham em conjunto para reduzir a complexidade computacional dos codificadores G.729 e G.729A e como eles afetam a qualidade percebida do sinal resultante de fala. Para isto, o algoritmo PESQ foi utilizado para estimar a qualidade percebida do codificadores e o tempo T de codificação foi utilizado para representar a complexidade computacional.

As notas PESQ-MOS resultantes, para o banco BD1, em função do tempo T de codificação de um segmento, para os codificadores G.729 e G.729A são exibidas nas Figuras 4.3(a) e 4.4(a). Alguns pontos do fecho côncavo (melhor compromisso qualidade/complexidade computacional) são destacados e detalhados nas Tabelas 4.3 e 4.4, que também incluem o número de multiplicações M. Estes resultados indicam que as modificações propostas podem coletivamente reduzir o tempo total de codificação em cerca de 12% (configuração A' na Tabela 4.3) ou 11% (configuração D' na Tabela 4.4) se comparado com as implementações originais dos codecs, mantendo o nível de qualidade do sinal de fala reconstruído.

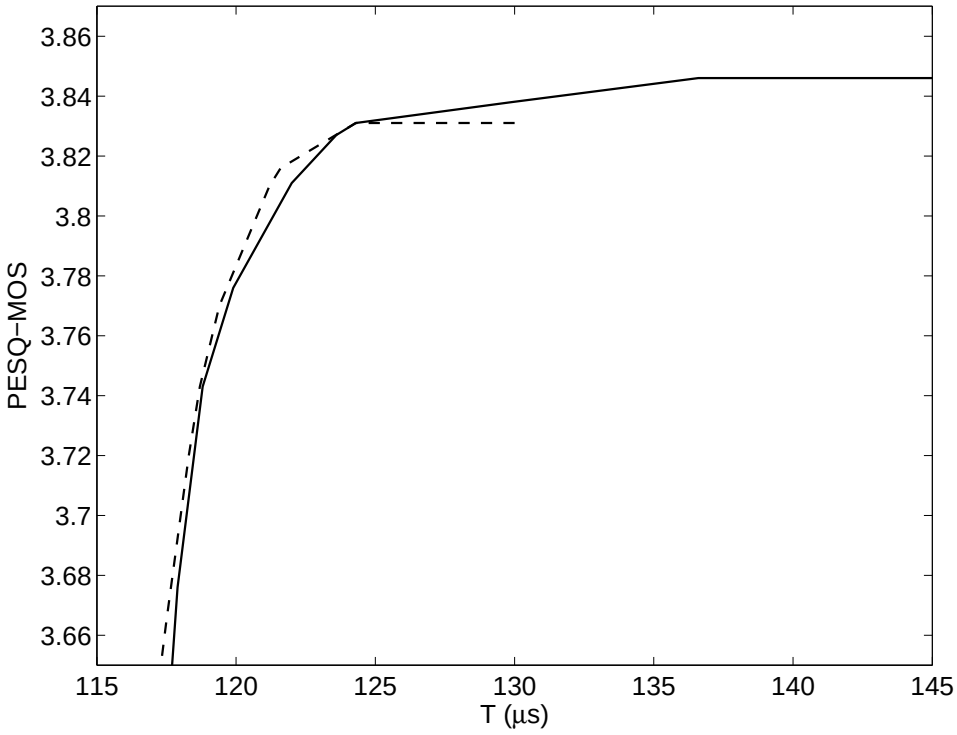
4.3 Reutilização do atraso do dicionário adaptativo

Segundo [13], um determinado segmento de fala pode ser classificado como um dos seguintes tipos:

- Sonoro: basicamento um trem de impulsos (ou pulsos glotais).
- Surdo: pode ser considerado ruído branco.
- Misto: contém componentes sonoros e surdos.
- Silêncio: é, na verdade, a ausência de excitação.
- Plosivo: silêncio por um curto período de tempo, seguido de excitação sonora ou surda (fecha-se o trato vocal, aumentando a pressão do ar e soltando-o em seguida de uma só vez).

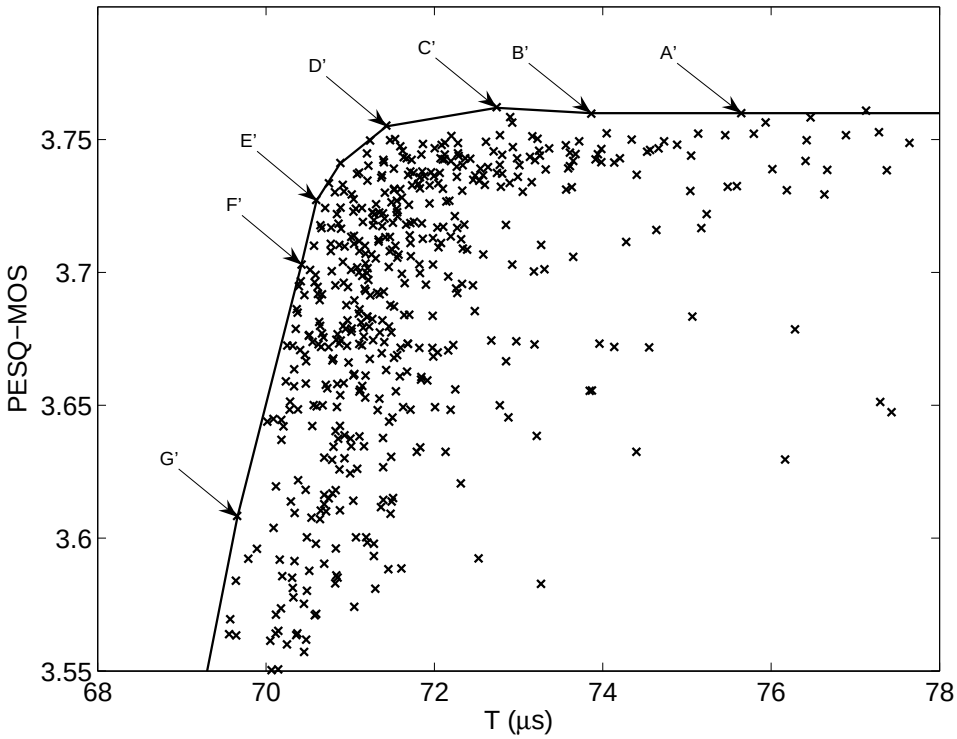


(a)

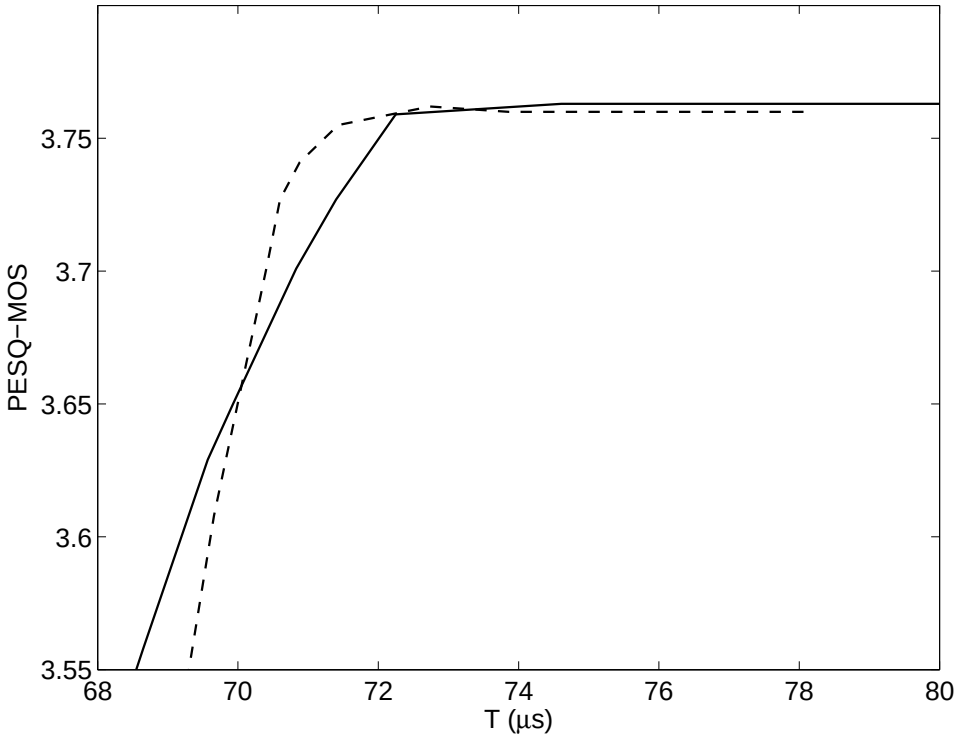


(b)

Figura 4.3: PESQ-MOS em função do tempo T de codificação para: (a) diferentes vetores $[D_c\ D_t\ \Delta_t\ N_t]$ no G.729 com T_c fixo; (b) fecho côncavo da Figura 3.5(a) (linha cheia) e fecho côncavo da Figura 4.3(a) (linha tracejada).



(a)



(b)

Figura 4.4: PESQ-MOS em função do tempo T de codificação para: (a) diferentes vetores $[D_c\ D_t\ \Delta_t\ N_t]$ no G.729A com T_c fixo; (b) fecho côncavo da Figura 3.5(b) (linha cheia) e fecho côncavo da Figura 4.4(a) (linha tracejada).

- Sussurro: formado por componentes de baixa amplitude, essencialmente ruidosos, até mesmo para os segmentos originalmente sonoros.
- Ejetivo: apenas sai ar pela cavidade oral.
- Clique/Implosivo: o ar é trazido para dentro do trato vocal.

A técnica proposta por [7] utiliza como base o fato de o cálculo da estimativa T_{op} do atraso de *pitch*, obtida no estágio de *open-loop*, descrito no Capítulo 2, possuir três propriedades importantes, como ilustrado na Figura 4.5:

- O contorno de *pitch* inicia na fronteira entre trechos de fala dos tipos surdo e sonoro (UV-V) ou dos tipos silêncio e sonoro (S-V).
- O contorno de *pitch* é uma curva suave.
- O contorno de *pitch* é definido apenas em trechos de fala do tipo sonoro.

Com base nas Figuras 4.5(b) e (c), pode-se concluir que existe uma grande probabilidade de segmentos próximos de fala possuírem períodos de *pitch* próximos e, consequentemente, valores semelhantes para T_{op}.

Tabela 4.3: Características das configurações do G.729 modificado com T_c fixo com o melhor compromisso PESQ-MOS $\times$ T selecionadas na Figura 4.3(a).

Ponto	D_c	D_t	Δ_t	N_t	MOS	M	$T(\mu s)$
G.729	1	1	0	1	3,84	9920	140,5
A’	2	1	0	1	3,83	2400	124,3
B’	2	2	1	1	3,82	1440	121,6
C’	2	2	0	1	3,81	1200	121,2
D’	2	3	0	1	3,77	800	119,4
E’	2	5	2	1	3,74	960	118,7
F’	4	3	0	1	3,72	400	118,3
G’	5	5	1	1	3,67	288	117,5

A técnica de reutilização do atraso do dicionário adaptativo utiliza a função WD-LSP (*weighted delta linear spectral pairs*) para determinar as fronteiras UV-V e S-V.

Tabela 4.4: Características das configurações do G.729A modificado com T_c fixo com o melhor compromisso PESQ-MOS $\times$ T selecionadas na Figura 4.4(a).

Ponto	D_c	D_t	Δ_t	N_t	MOS	M	$T(\mu s)$
G.729A	2	1	0	1	3,75	3680	79,8
A'	1	6	3	1	3,76	2240	75,6
B'	2	2	1	1	3,76	1440	73,9
C'	3	2	1	1	3,76	972	72,7
D'	4	7	5	1	3,76	780	71,4
E'	6	4	0	1	3,73	210	70,6
F'	3	7	0	1	3,70	243	70,4
G'	6	7	0	1	3,61	126	69,7

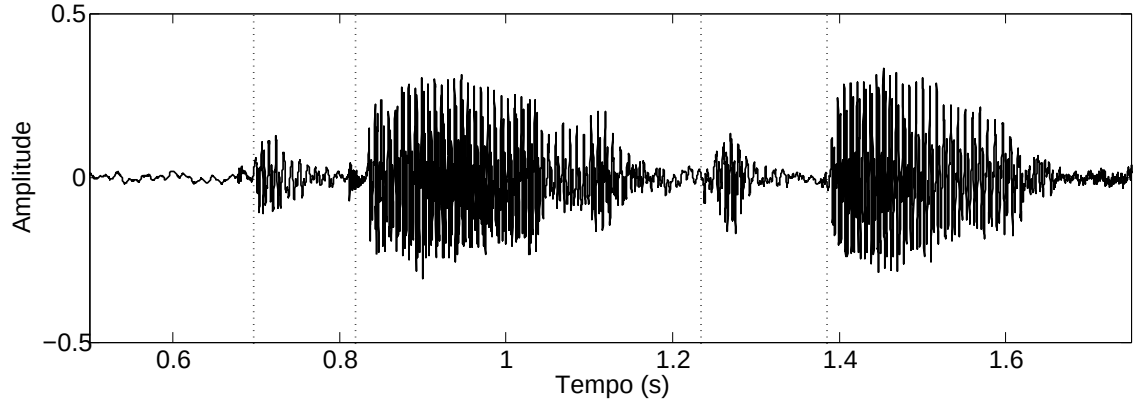
A função WD-LSP para o i-ésimo segmento é dada por:

$$F_i = \sum_{k=1}^{10} w_k \times [LSP_i(k) - LSP_{i-1}(k)]^2, \tag{4.3}$$

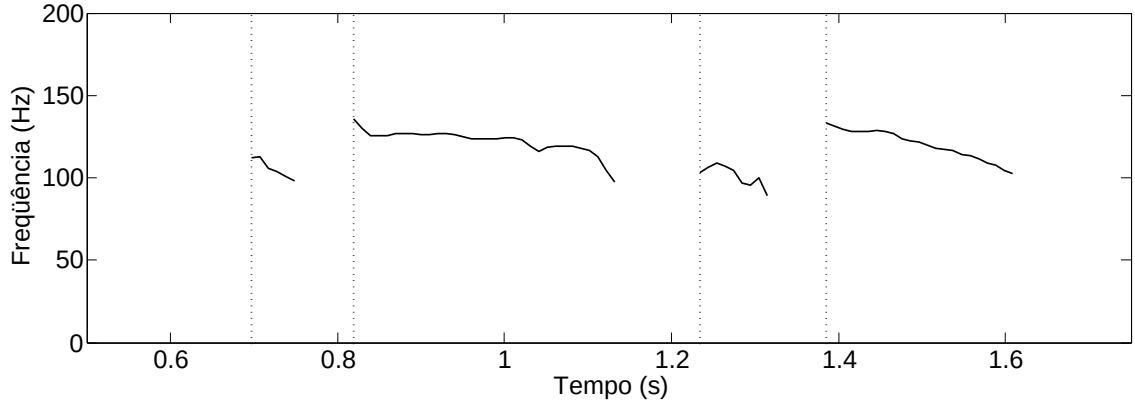
onde $LSP_i(k)$, $k = 1, ..., 10$ são os coeficientes LSP do i-ésimo segmento e w_k é um coeficiente fixo de ponderação associado ao k-ésimo coeficiente LSP.

Se o valor desta função for superior a um dado limiar η, o valor de T_{op} é re-estimado. Caso contrário, T_{op} é atualizado pelo atraso T_2 do dicionário adaptativo referente ao segundo sub-segmento do segmento anterior. Naturalmente, quanto maior for o limiar η, maior será a redução na complexidade computacional, ao custo de uma pequena redução na qualidade de sinal de fala reconstruído, como mostra [7]. Neste trabalho, os pesos w_k utilizados tanto para o G.729, quanto para o G.729A são:

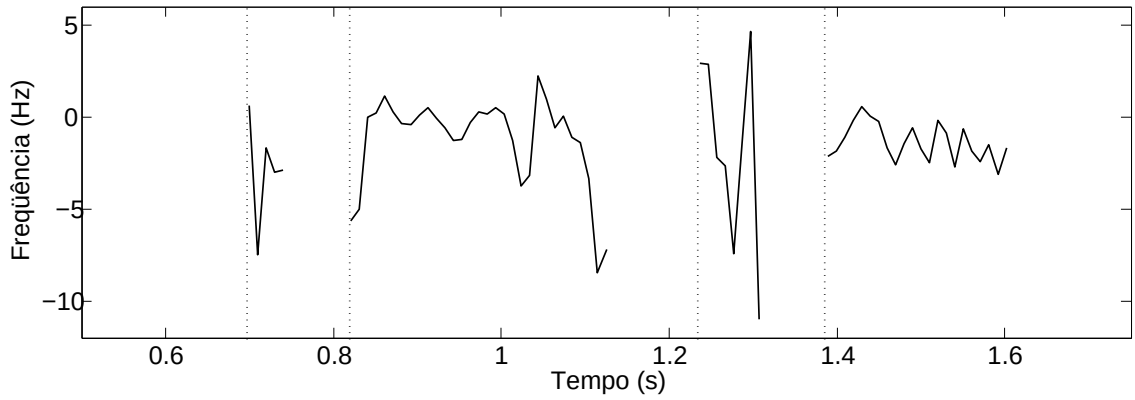
$$\begin{cases} w_1 &= 0{,}57 \\ w_2 &= 2{,}45 \\ w_3 &= 5{,}78 \\ w_4 &= 7{,}32 \\ w_5 &= 7{,}30 \\ w_6 &= 8{,}20 \end{cases}.$$



(a)



(b)



(c)

Figura 4.5: Propriedades do contorno de *pitch* de um sinal de fala, onde a linha pontilhada representa fronteiras UV-V ou S-V. (a) Trecho do sinal de fala uk100.wav; (b) Contorno de *pitch* de (a); (c) Variação do *pitch* de (a).

$$\begin{cases} w_7 &= 7{,}01 \\ w_8 &= 5{,}94 \\ w_9 &= 4{,}17 \\ w_{10} &= 2{,}17 \end{cases}.$$

Estes pesos foram obtidos através da metodologia descrita em [7], porém, neste caso, considerando segmentos de fala das 5 classes definidas por [7] (fronteira UV-V, fronteira S-V, sonoro, surdo e silêncio). A relação entre o percentual $R\%$ de reutilização do atraso do dicionário adaptativo e o limiar η para este conjunto de coeficientes de ponderação é caracerizada na Tabela 4.5, onde as colunas G.729 e G.729A representam a nota PESQ-MOS para os respectivos codificadores. Como esperado, valores maiores de η acarretam maiores taxas de reutilização do atraso do dicionário adaptativo.

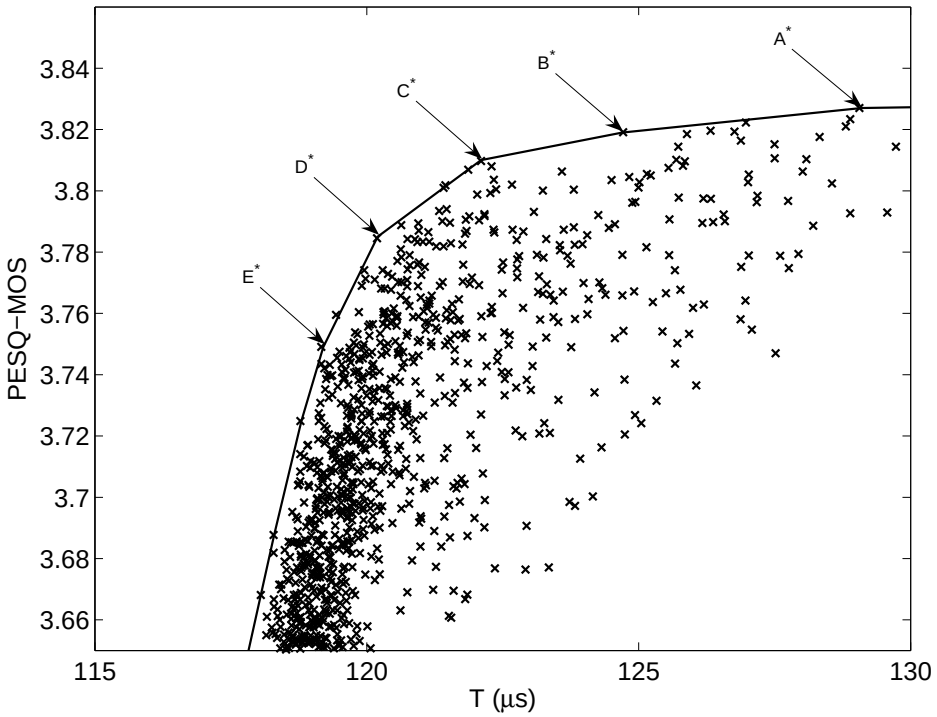
Tabela 4.5: Percentual $R\%$ de reutilização em função do limiar η.

η	$R\%$	G.729	G.729A
$0,0043$	1%	3,83	3,75
$0,0110$	5%	3,82	3,74
$0,0170$	10%	3,82	3,75
$0,0220$	15%	3,81	3,74
$0,0270$	20%	3,81	3,72
$0,0315$	25%	3,81	3,70

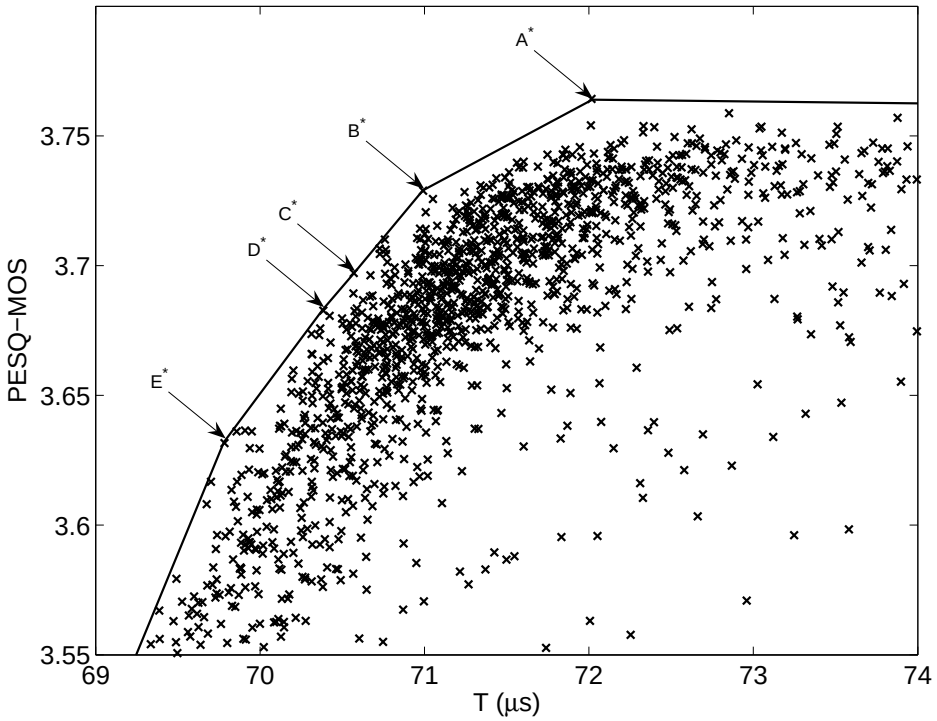
4.3.1 Combinação com as técnicas do Capítulo 3

Nesta subseção, investiga-se a combinação da técnica proposta por [7] com os parâmetros D_c, D_t, Δ_t e N_t. Verifica-se a maneira com que estes métodos todos trabalham em conjunto para reduzir a complexidade computacional dos codificadores G.729 e 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 codificadores e o número médio $\bar{M}$ de multiplicações foi utilizado para representar a complexidade computacional.

As notas PESQ-MOS resultantes, para o banco BD1, em função do tempo T de codificação de um segmento, para os codificadores G.729 e G.729A são exibidas nas Figuras 4.6(a) e 4.6(b). Alguns pontos do fecho côncavo (melhor compromisso qualidade/complexidade computacional) são destacados e detalhados nas Tabelas 4.6 e 4.7. Estes resultados indicam que as modificações propostas podem



(a)



(b)

Figura 4.6: PESQ-MOS em função do tempo T de codificação para diferentes configurações dos codificadores modificados: (a) ITU-T G.729 e (b) ITU-T G.729A. A linha contínua representa os fechos côncavos caracterizados pelas Tabelas 4.6 e 4.7.

Tabela 4.6: Características das configurações do G.729 modificado com o melhor compromisso PESQ-MOS $\times$ T selecionadas na Figura 4.6(a).

Ponto	D_c	D_t	Δ_t	N_t	η	MOS	$\bar{M}$	$T(\mu s)$
G.729	1	1	0	1	-	3,84	9920	140,5
A*	1	2	1	1	0.011	3,83	5168	129,1
B*	2	2	1	1	0.011	3,82	2584	124,7
C*	2	4	3	1	0.011	3,81	1862	122,1
D*	4	3	2	1	0.011	3,79	1026	120,2
E*	4	4	2	1	0.022	3,75	918	119,2

Tabela 4.7: Características das configurações do G.729A modificado com o melhor compromisso PESQ-MOS $\times$ T selecionadas na Figura 4.6(b).

Ponto	D_c	D_t	Δ_t	N_t	η	MOS	$\bar{M}$	$T(\mu s)$
G.729A	2	1	0	1	-	3,75	3680	79,8
A*	6	3	2	1	0.011	3.76	718,2	72,0
B*	8	4	1	1	0.011	3.73	351,5	71,0
C*	8	5	1	1	0.011	3.70	294,5	70,6
D*	7	5	0	1	0.011	3.68	285,0	70,38
E*	6	7	0	1	0.022	3.63	214,2	69,8

coletivamente reduzir o tempo T de codificação de um segmento em aproximadamente 11% (configuração B* na Tabela 4.6) ou 11% (configuração B* na Tabela 4.7) se comparado com as implementações originais dos codecs, mantendo o nível de qualidade do sinal de fala reconstruído.

4.4 Combinação de todas as técnicas

A fim de se obter o melhor compromisso entre complexidade computacional e qualidade de codificação, fez-se a combinação de todas as técnicas apresentadas nesta dissertação. A Figura 4.7 mostra os vários pontos resultantes dessa combinação de técnicas, com destaque para o fecho côncavo, que representa o melhor compromisso

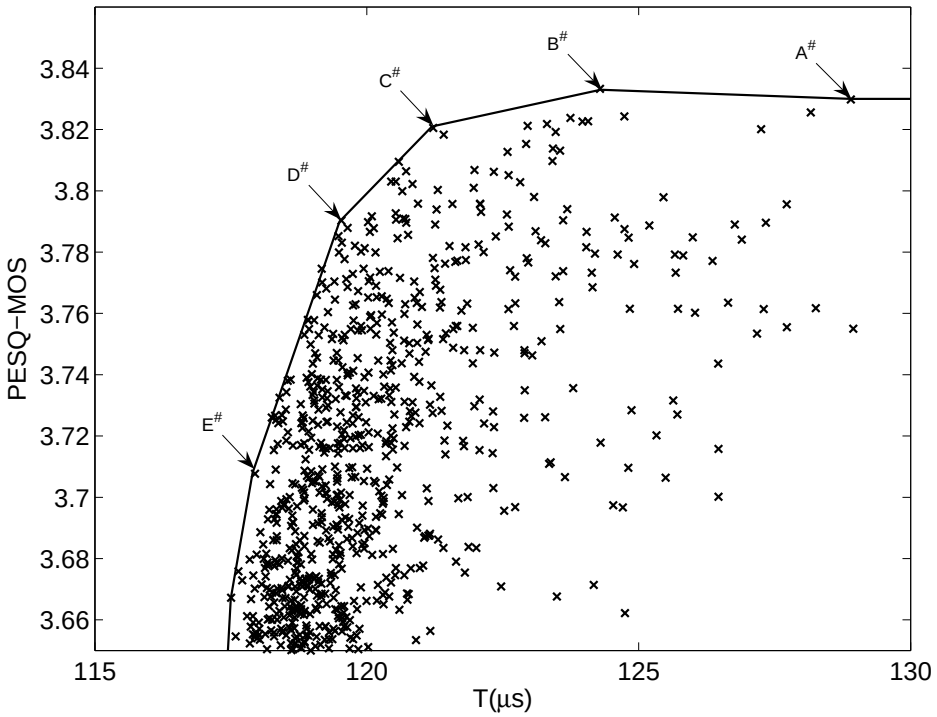
entre o tempo T de codificação de um segmento e a nota PESQ-MOS. Alguns pontos do fecho côncavo da Figura 4.7(a) estão detalhados na Tabela 4.8 e alguns pontos do fecho côncavo da Figura 4.7(b) estão detalhados na Tabela 4.9. A coluna $(U; V)$ representa o intervalo de confiança de 95%. Observando estas duas tabelas, pode-se concluir que a combinação das 6 técnicas apresentadas é capaz de reduzir em 14% (ponto C$^{\#}$ da Tabela 4.8) e 12% (ponto D$^{\#}$ da Tabela 4.9) o tempo T de codificação de um segmento para os codificadores G.729 e G.729A, respectivamente, sem que haja redução da qualidade de codificação.

A qualidade dos sinais de fala obtidos pelos codificadores modificados (G.729 com a configuração C$^{\#}$ e G.729A com a configuração D$^{\#}$) também foi aferida subjetivamente através de um teste do tipo *absolute category rating* (ACR), como descrito na recomendação ITU-T P.800 [3]. Neste teste, 32 sinais diferentes de fala na língua portuguesa do Brasil foram codificados/decodificados pelas codificadores originais (G.729 e G.729A) e modificados (G.729$^{\#}$ e G.729A$^{\#}$). Então, 15 ouvintes foram instruídos a dar uma nota para cada sinal, segundo a Tabela 2.2, em resposta à pergunta "Como você avalia a qualidade de reprodução?". A Tabela 4.10 mostra o resultado do teste ACR e indica, para fins práticos, equivalência entre as versões modificadas e originais para ambos os codificadores, em relação à qualidade percebida do sinal de fala reconstruído.

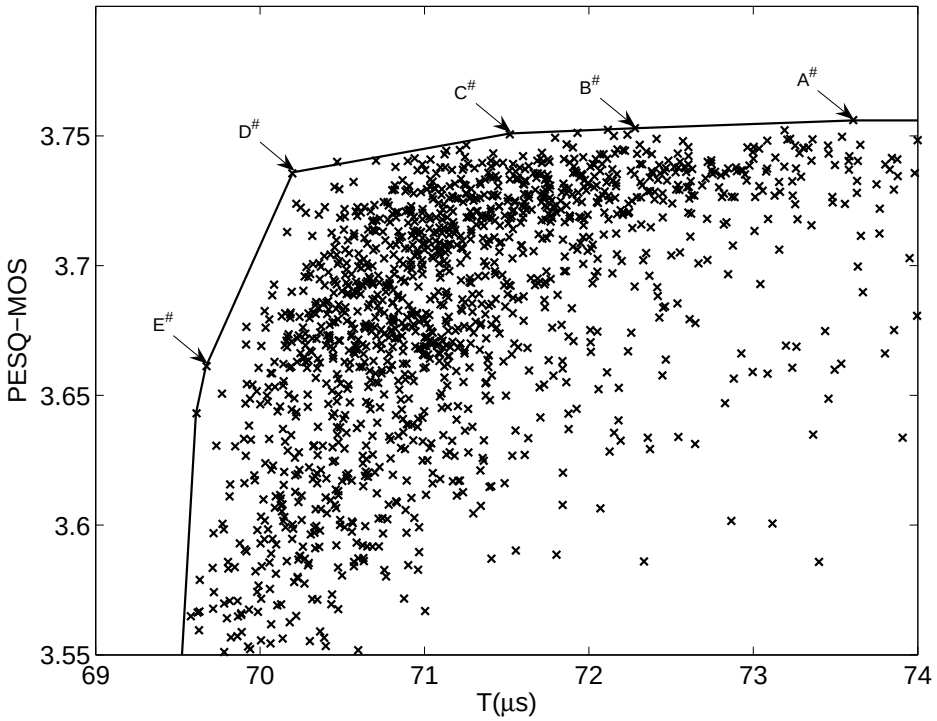
4.5 Conclusão

Dois esquemas de simplificação do estágio de *open-loop* na busca pela excitação do dicionário adaptativo são apresentados: fixação de τ no intervalo $80 \leq \tau \leq 143$ e a reutilização do atraso do dicionário adaptativo [7]. Ao combinar-se o primeiro esquema com as simplificações apresentadas no Capítulo 3, é possível reduzir o tempo T de codificação de um segmento para os codificadores G.729 e G.729A em 12% e 11%, respectivamente, sem que haja diminuição da qualidade estimada pelo PESQ. Como resultado da combinação do segundo esquema com os parâmetros D_c, D_t, Δ_t e N_t, o tempo T de codificação de um segmento pode ser reduzido em 11% e 11% para os codecs G.729 e G.729, respectivamente, sem que a qualidade percebida do sinal de fala reconstruído seja afetada.

Ao combinar-se simultaneamente ambos os esquemas apresentados neste capítulo com os parâmetros D_c, D_t, Δ_t e N_t, obtém-se uma redução de 14% e 12% do tempo T de codificação de um segmento para os codificadores G.729 e G.729A, respectivamente, sem diminuir a qualidade de codificação original destes codificadores.



(a)



(b)

Figura 4.7: PESQ-MOS em função do tempo T de codificação para os codificadores modificados: (a) G.729 e (b) G.729A. A linha contínua representa os fechos detalhados nas Tabelas 4.8 e 4.9.

Tabela 4.8: Características das configurações do G.729 modificado em destaque na Figura 4.7(a).

Ponto	D_c	D_t	Δ_t	N_t	η	MOS	$(U;V)$	$\bar{M}$	$T(\mu s)$
G.729	1	1	0	1	-	3,84	$(3,797;3,879)$	9920	140,5
A$^{\#}$	1	1	0	1	0.011	3,83	$(3,790;3,870)$	4560	128,9
B$^{\#}$	1	2	1	1	0.017	3,83	$(3,789;3,877)$	2592	124,3
C$^{\#}$	2	2	1	1	0.017	3,82	$(3,779;3,862)$	1296	121,2
D$^{\#}$	4	2	1	1	0.017	3,79	$(3,744;3,836)$	648	119,5
E$^{\#}$	6	3	0	1	0.022	3,71	$(3.658;3.757)$	238	117,9

Tabela 4.9: Características das configurações do G.729A modificado em destaque na Figura 4.7(b).

Ponto	D_c	D_t	Δ_t	N_t	η	MOS	$(U;V)$	$\bar{M}$	$T(\mu s)$
G.729A	2	1	0	1	-	3,75	$(3,705;3,800)$	3680	79,8
A$^{\#}$	2	6	5	1	0.017	3,76	$(3,714;3,798)$	1140	73,6
B$^{\#}$	3	2	1	1	0.017	3,75	$(3,710;3,796)$	874,8	72,3
C$^{\#}$	5	2	0	1	0.011	3,75	$(3,711;3,790)$	456,0	71,5
D$^{\#}$	6	4	0	1	0.017	3,74	$(3,689;3,782)$	189,0	70,2
E$^{\#}$	10	5	0	1	0.022	3,66	$(3,610;3,712)$	81,6	69,7

Tabela 4.10: Resultado do teste ACR com 15 ouvintes.

Codificador	MOS	$(U;V)$
G.729	4,1	$(4,06;4,23)$
G.729$^{\#}$	4,1	$(4,06;4,24)$
G.729A	4,0	$(3,95;4,13)$
G.729A$^{\#}$	3,9	$(3,77;3,96)$

Capítulo 5

Conclusão

5.1 Contribuições do trabalho

Este trabalho apresentou o funcionamento do codificador definido pela Recomendação ITU-T G.729 [1], bastante utilizado no cenário de Telecomunições. Foram apresentados também seis diferentes esquemas que simplificam o bloco de busca no dicionário adaptativo.

O Capítulo 2 descreve detalhadamente o codificador ITU-T G.729 [1] e as alterações feitas pelo Anexo A [2], enfatizando o bloco de busca pela excitação do dicionário adaptativo. Os métodos objetivo PESQ [10] e subjetivo MOS [3] de avaliação de qualidade de fala também foram descritos neste capítulo.

O Capítulo 3 fez a introdução de quatro modificações no bloco da busca pela excitação do dicionário adaptativo dos codificadores G.729 e G.729A: D_c, D_t, Δ_t e N_t. As duas primeiras capazes de diminuir a complexidade computacional, porém também diminuindo a qualidade de codificação, enquanto as outras duas aumentam a complexidade computacional e também a qualidade de codificação. A combinação das quatro modificações tem a capacidade de manter a qualidade de codificação diminuindo o tempo de codificação de um segmento em 12% e 9% para os codecs G.729 e G.729A, respectivamente, se comparados com os valores originais.

O Capítulo 4 descreveu duas outras técnicas de simplificação do bloco da busca pela excitação do dicionário adaptativo e a combinação destas técnicas com as simplificações do Capítulo 3: utilização de T_c fixo e reaproveitamento do atraso do dicionário adaptativo. A combinação da primeira técnica com os parâmetros D_c,

D_t, Δ_t e N_t resultou em uma diminuição de 12% e 11% do tempo de codificação de um segmento para os codificadores G.729 e G.729A, respectivamente, mantendo a qualidade do sinal de fala reconstruído. O resultado da combinação da técnica de reutilização do atraso do dicionário adaptativo com as quatro simplificações do Capítulo 3 foi a redução de 11% e 11% do tempo de codificação de um segmento, respectivamente, mantendo-se a qualidade do sinal de fala reconstruído. Combinar simultaneamente as duas técnicas apresentadas neste capítulo com as quatro simplificações do Capítulo 3, resultou em uma redução no tempo T de codificação de um segmento para os codificadores G.729 e G.729 de 14% e 12%, respectivamente, sem acarretar uma redução da qualidade de codificação original, o que foi confirmado por testes subjetivos informais.

5.2 Propostas para trabalhos futuros

A seguir estão algumas sugestões de continuação deste trabalho:

- A Seção 4.2 descreve o processo de fixar T_c no valor da moda de seu histograma. Esta não é, necessariamente, a melhor estratégia e estudar outras estatísticas, como a média, por exemplo, é um possível trabalho futuro.

- O filtro de ponderação é dado por $W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)}$ e por $W(z) = \frac{A(z)}{A(z/\gamma)}$ nos codificadores G.729 e G.729A, respectivamente. No codificador G.729 os coeficientes de ponderação γ_1 e γ_2 são adaptativos, enquanto $\gamma = 0,75$ é fixo no codificador G.729A. Um possível trabalho futuro é estudar valores fixos para γ_1, γ_2 e γ diferentes de $0,75$, utilizando o avaliador objetivo de qualidade PESQ.

- Para se obter a excitação do dicionário fixo e a excitação do dicionário adaptativo, obtém-se de maneira independente os segmentos das respectivas excitações. Isto pode gerar descontinuidades na junção de dois segmentos consecutivos. Um possível trabalho futuro é estudar formas de reduzir ou eliminar estes transitórios, possivelmente aumentando a qualidade do sinal reconstruído de fala.

- A busca pela excitação do dicionário fixo consome cerca de 40% do tempo

de codificação de um segmento, segundo [7]. Um possível trabalho futuro é estudar uma maneira de reduzir a complexidade computacional deste procedimento através da interrupção do algoritmo de busca caso o coeficiente de correlação entre o sinal alvo $x(n)$ atualizado e a possível excitação filtrada do dicionário fixo $c(n) * h(n)$ seja superior a um dado limiar.

Referências Bibliográficas

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

[2] ITU-T, "Rec. G.729 Annex A, *Reduced complexity 8 kbit/s CS-ACELP speech codec*", 1996.

[3] ITU-T, "Rec. P.800, *Methods for subjective determination of transmission quality*", 1996.

[4] SCHROEDER, M. R., ATAL, B. S., "*Code-excited linear prediction* (CELP): *High quality speech at very low bit rates*", *Proc. IEEE Int. Conf. Acoust.*, v. 2, pp. 437–440, 1985.

[5] KIM, H. K., "*Adaptive encoding of fixed codebook in CELP coders*". In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, v. 1, pp. 149–152, 12–15 May 1998.

[6] HA, N. K., "*A fast search method of algebraic codebook by reordering search sequence*". In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '99*, v. 1, pp. 21–24, 15–19 March 1999.

[7] HWANG, S.-H., "*Computational improvement for G.729 standard*", *IEE Electronic Letters*, v. 36, n. 13, pp. 1163–1164, 2000.

[8] RAMIREZ, M. A., GERKEN, M., "*Joint position and amplitude search of algebraic multipulses*", v. 8, n. 5, pp. 633–637, Sept. 2000.

[9] LEE, E. D., YUN, S. H., LEE, S. I., et al., "*Iteration-free pulse replacement method for algebraic codebook search*", *Electronics Letters*, v. 43, n. 1, pp. 59–60, Jan. 4 2007.

[10] 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.

[11] ITU-T, "http://www.itu.int/rec/T-REC-G.729/", G.729 web page.

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

[13] J. R. DELLER, J. G. PROAKIS, J. H. L. H., *Discrete-Time Processing of Speech Signals*. Macmillan Coll Div, 1995.

[14] SALAMI, R., LAFLAMME, C., BESSETTE, B., et al., "*ITU-T G.729 Annex A: reduced complexity 8 kb/s CS-ACELP codec for digital simultaneous voice and data*", v. 35, n. 9, pp. 56–63, Sept. 1997.

[15] KONDOZ, A. M., *Digital Speech: Coding for Low Bit Rate Communication Systems*. Wiley, 1999.

[16] OSR, "http://www.voiptroubleshooter.com/open_speech/index.html", Open Speech Repository.

[17] DE M. PREGO, T., "Aperfeiçoamento do codificador de voz CELP", UFRJ/Poli, 2007.

[18] DE M. PREGO, T., NETTO, S., "Algoritmo de Busca Eficiente no Dicionário Adaptativo para o Codec ITU-T G.729", *SBrT - Simpósio Brasileiro de Telecomunicações*, 2008.

[19] DE M. PREGO, T., NETTO, S. L., "*Efficient Search in the Adaptive Codebook for ITU-T G.729 Codec*", v. 16, n. 10, pp. 881–884, Oct. 2009.

[20] AMD, "http://developer.amd.com/cpu/CodeAnalyst/Pages/default.aspx", AMD CodeAnalyst Performance Analyzer.

Apêndice A

Medida do tempo de codificação

Neste apêndice será detalhada a maneira a qual foi feita a medida do tempo T de codificação de um segmento, utilizada como medida de complexidade computacional nesta dissertação.

A.1 AMD CodeAnalyst Performance Analyzer

O programa AMD CodeAnalyst Performance Analyzer [20] foi utilizado para medir o tempo T de codificação de um segmento e pode ser baixado gratuitamente na página *http://developer.amd.com/cpu/CodeAnalyst/Pages/default.aspx*. Este programa possui um conjunto de ferramentas que verificam quais processos estão sendo executados em um determinado instante e fornece o somatório final da quantidade de vezes que cada processo estava sendo executado. A este método de verificação se dá o nome de *profiling*. O AMD CodeAnalyst subdivide os processos em funções, permitindo uma análise detalhada de cada processo.

Estas verificações podem ocorrer de diversas maneiras e a utilizada nesta dissertação é chamada de *Event-Based Profiling* (EBP). O EBP consiste em verificar quais processos estão sendo executados em determinados instantes a cada n ocorrências de um determinado evento, no caso deste trabalho, a cada $n = 10^4$ clocks.

A.2 Rotina para medir o tempo de codificação

Para medir o tempo T de codificação de um segmento, utilizou-se o AMD CodeAnalyst via linha de comando, através de 3 comandos chave: *caprofile*, *cadataanalyze* e *careport*.

O primeiro comando é responsável por fazer o processo de *profiling*, armazenando o resultado em um arquivo codificado do tipo .prd. O segundo comando é responsável por transformar um arquivo do tipo .prd em arquivos do tipo .ebp (caso seja feito um *profiling* do tipo EBP) que pode ser importado para o modo gráfico do AMD CodeAnalyst ou pode ser interpretado pelo comando *careport*. Este último comando interpreta os arquivos do tipo .ebp, gerando informação passível de interpretação por seres humanos.

A.2.1 Exemplo

- Para fazer a codificação utilizando o G.729 contendo todas as modificações propostas utilizamos o comando: *caprofile /e 0x07601:10000 /o teste /l C:\UFRJ\g729_win\Release\coder.exe ..\..\base_voz\teste_menor\ch1.wav ch1.celp 1 2 0 3 0.022 100* , onde $D_c = 1$, $D_t = 2$, $\Delta_t = 0$, $N_t = 3$ e $\eta = 0.022$ são passados na linha de comando, enquanto o parâmetro $T_c = 133$ é fixo e não é passado pela linha de comando. O último argumento da linha de comando *100* significa que a codificação é feita 100 vezes, enquanto a leitura e escrita de arquivo acontece apenas 1 vez. Isto foi feito para diminuir a influência da leitura e da escrita de arquivo na medida do tempo de codificação. Os argumentos */e 0x07601:10000* determinam que é um EBP, o evento é número de clocks e é contado a cada 10000 eventos. Os argumentos */o teste* determinam que o resultado do EBP será gravado no arquivo *teste.prd*.

- Para transformar o arquivo teste.prd em teste.ebp utilizamos o comando: *cadataanalyze /i teste.prd /o teste*. São criados o diretório com o nome *teste.ebp.dir* e o arquivo *teste.ebp.dir\teste.ebp*.

- Para verificarmos o número N de clocks que o processo *coder.exe* exigiu para ser concluído utilizamos o comando: *careport /i teste.ebp.dir\teste.ebp > teste.txt*.

- Do arquivo *teste.txt* obtemos a linha: *985831.00 21.03 C:\UFRJ\g729_win\Release\coder.exe*. O número *985831.00* representa o número N de clocks que o processo *C:\UFRJ\g729_win\Release\coder.exe* necessitou para ser executado. *21.03* é o percentual de clocks do processo *coder.exe* em relação ao total de clocks aferidos.

- O tempo T de codificação de um segmento é obtido por $\frac{985831 \times 10000}{382 \times 100 \times 2.01 \times 10^9} = 128.39\ \mu$s, uma vez que o arquivo *ch1.wav* possui 382 segmentos e o processador do computador no qual o processo *coder.exe* foi executado possui frequência de clock igual a 2,01 GHz.