

Universidade Federal do Rio de Janeiro

Escola Politécnica

Departamento de Eletrônica e de Computação

Aperfeiçoamento do Codificador de voz G.729 anexo A

Autor:

Lívia Cristina da Silva Santos

Orientador:

Prof. Sergio Lima Netto, Ph.D.

Coorientador:

Thiago de Moura Prego, M.Sc.

Examinador:

Prof. Eduardo Antônio Barros da Silva, Ph.D.

Examinador:

Prof. Amaro Azevedo de Lima, Ph.D.

DEL

Agosto de 2010

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

DEDICATÓRIA

A Lúcia Santos, minha mãe.

AGRADECIMENTO

Agradeço:

- Ao professor Sérgio Lima Netto e a Thiago de Moura Prego pela orientação acadêmica e paciência.
- Aos professores Eduardo Antônio Barros da Silva e Amaro Azevedo de Lima por aceitarem compor a banca examinadora.
- A Leonardo Santos, meu irmão, pela ajuda e apoio recebidos.

RESUMO

A grande utilização de dispositivos móveis ainda encontra limitações relacionadas à sua autonomia energética. Porém, uma das maneiras de aumentar o tempo de uso das baterias desses dispositivos é reduzir o esforço computacional exigido pelos códigos empregados em seus processos. Nesse nicho se encaixa a utilidade deste trabalho, que tem por objetivo acelerar o G.729 anexo A, um codificador de fala descrito e implementado pela *International Telecommunication Union* e baseado na estrutura CS-ACELP (*Conjugate-Structured Algebraic Code Excited Linear Prediction*).

Em uma etapa introdutória, é feita a descrição dos métodos utilizados para avaliar os resultados obtidos e da estrutura dos codificadores CS-ACELP. Essa descrição serve de subsídio para o entendimento deste trabalho.

Para a aceleração do algoritmo do G.729 anexo A, que será chamado G.729A ao longo deste trabalho, foi simplificada a busca de excitações no dicionário adaptativo nas etapas de *open loop* e *closed loop*. Isso foi feito através da limitação na quantidade de atrasos consideradas para a estimativa do *pitch*. Outro recurso utilizado para a aceleração do algoritmo é a implementação da decimação do cálculo de correlação e da decimação dos sinais no tempo. Na contramão da tendência de restrições e decimações citadas, são implementadas buscas proximais para a estimativa do *pitch*. Essas buscas aumentam a complexidade computacional do algoritmo, porém reduzem a degradação do sinal reproduzido.

A comparação do G.729A com o codificador resultante deste trabalho revela redução de 70,50% na complexidade computacional envolvida na busca do dicionário adaptativo e de 17% no tempo de codificação, sem que com isso seja comprometida a qualidade do sinal reproduzido.

Sumário

Lista de Figuras	ix
1 Introdução	1
1.1 Introdução	1
1.2 Proposta do trabalho	2
1.2.1 Justificativa	2
1.2.2 Objetivos	3
1.3 Organização do trabalho	3
2 O G.729A	5
2.1 Introdução	5
2.2 Métodos de qualificação de codificadores de voz	7
2.2.1 MOS	7
2.2.2 PESQ	7
2.3 Codificadores CS-ACELP	8
2.4 Codificador	9
2.4.1 Pré-processamento	9
2.4.2 Janelamento no sinal de voz	9
2.4.3 Análise LPC, quantização e interpolação	10
2.4.4 Filtro perceptivo	11
2.4.5 Análise por síntese	12
2.4.6 Busca no dicionário adaptativo	12
2.4.7 Busca no dicionário fixo	16
2.5 Decodificador	18
2.6 Conclusão	18

3	Trabalhos anteriores	19
3.1	Introdução	19
3.2	Banco de voz	19
3.2.1	Caracterização do banco de voz	20
3.2.2	Análise dos locutores do banco de voz	23
3.3	Modificações de trabalhos anteriores	24
3.3.1	Fixação de T_{op} para um dado intervalo	24
3.3.2	Decimação do sinal no tempo	25
3.3.3	Decimação do atraso k	25
3.3.4	Análise na vizinhança de T_{op}	25
3.3.5	Análise de estimativas de <i>pitch</i> distantes	26
3.3.6	Reutilização do atraso do dicionário adaptativo	26
3.4	Conclusão	28
4	Simplificações propostas	29
4.1	Introdução	29
4.2	Modificações no <i>open loop</i>	29
4.2.1	Fixação de T_{op} para todos os intervalos.	29
4.2.2	Análise da nota PESQMOS	32
4.2.3	Análise da frequência de ocorrência	34
4.2.4	Limitação da variação do intervalo de estimativa do <i>pitch</i>	36
4.2.5	Reprodução da decimação do sinal alvo no cálculo da auto- correlação	36
4.2.6	Reprodução da busca proximal de T_{op}	38
4.3	Resumo das simplificações no <i>open loop</i>	40
4.4	Modificações no <i>closed loop</i>	41
4.4.1	Busca proximal da estimativa do <i>lag</i>	41
4.4.2	Cálculo do resíduo do filtro LP	44
4.4.3	Decimação do sinal no tempo no <i>closed loop</i>	46
4.5	Resumo das modificações do <i>closed loop</i>	47
4.6	Conclusão	51

5 Conclusão	52
5.1 Contribuições do trabalho	52
5.2 Propostas para trabalhos futuros	53
Bibliografia	54

Lista de Figuras

2.1	Processo de geração de voz humana no codificador LPC.	9
2.2	Resposta em magnitude do filtro de pré-processamento F_{pe}	10
2.3	Janela de predição linear w_{lp}	10
2.4	Estrutura do dicionário adaptativo.	13
2.5	Diagrama de blocos do codificador G.729A	17
2.6	Princípio de funcionamento do decodificador do G.729.	18
3.1	Distribuição de T_a ao longo do intervalo (a) no G.729A original	21
3.2	Distribuição de T_b ao longo do intervalo (b) no G.729A original	21
3.3	Distribuição de T_c ao longo do intervalo (c) no G.729A original	22
3.4	Distribuição de T_{op} ao longo dos intervalos (a), (b) e (c) no G.729A original	22
3.5	Distribuição de frequência dos sinais de fala do banco BD1.	23
3.6	Quantidade de multiplicações necessárias na etapa de <i>open loop</i> para o cálculo da estimativa de T_{op} com a configuração que utiliza os parâmetros D_c, D_t, N_t, Δ_t	28
4.1	PESQMOS para τ_i ao longo dos intervalos $i = (a), (b)$ e (c).	31
4.2	Percentual de ocorrência da igualdade $T_{op} = \tau_i$ nos intervalos (a), (b) e (c)	32
4.3	Relação entre a nota PESQMOS e o parâmetro N_i , considerando T_i ordenado por nota PESQMOS.	34
4.4	Relação entre a nota PESQMOS e o parâmetro N_i , considerando T_i ordenado por frequência de ocorrência da igualdade $T_i = T_{op}$	35
4.5	Quantidade de multiplicações necessárias para o cálculo de T_{op} na configuração que utiliza os parâmetros N_a, N_b	37
4.6	Quantidade de multiplicações necessárias para o cálculo de T_{op} na configuração que utiliza os parâmetros N_a, N_b e D_c	39

4.7	Quantidade de multiplicações necessárias para o cálculo de T_{op} na configuração que utiliza os parâmetros N_a , N_b , D_c e Δ_t	41
4.8	Quantidade de multiplicações necessárias para o cálculo da estimativa do <i>lag</i> nas etapas de <i>open loop</i> e <i>closed loop</i> com a configuração que utiliza os parâmetros N_a , N_b , D_c , Δ_t e r_{cl}	43
4.9	Quantidade de multiplicações necessárias para o cálculo da estimativa do <i>lag</i> nas etapas de <i>open loop</i> e <i>closed loop</i> com a configuração que utiliza os parâmetros N_a , N_b , D_c , Δ_t e r_{cl} e o filtro de síntese para o cálculo do resíduo.	45
4.10	Quantidade de multiplicações necessárias para o cálculo da estimativa do <i>lag</i> nas etapas de <i>open loop</i> e <i>closed loop</i> com a configuração que utiliza os parâmetros N_a , N_b , D_c , Δ_t , r_{cl} e D_k e o filtro de síntese para o cálculo do resíduo.	47
4.11	Quantidade de multiplicações necessárias para o cálculo da estimativa do <i>lag</i> nas etapas de <i>open loop</i> e <i>closed loop</i> com a configuração que utiliza os parâmetros N_a , N_b , D_c , Δ_t e r_{cl} e o filtro de síntese para o cálculo do resíduo comparada com a curva C_o da Figura 3.6.	49
4.12	Distribuição de notas de teste comparativo subjetivo.	50

Capítulo 1

Introdução

1.1 Introdução

A área de Telecomunicações é a principal responsável pela transmissão de informação à distância, que marca a sociedade moderna. O desejo crescente do ser humano de se comunicar com qualquer ponto do globo e em qualquer instante impulsiona o desenvolvimento de dispositivos móveis, que devem atender a essa demanda da forma mais eficiente possível. Isso é obtido através do processamento digital de sinais, já que, com esse tipo de tecnologia, é possível utilizar algoritmos computacionais, reduzindo erros inerentes ao processamento analógico.

Para a realização do processamento digital de sinais são necessárias as etapas de amostragem, quantização e codificação. Na primeira delas, são tomadas amostras periódicas do sinal original, tornando-o discretizado no tempo.

Na quantização, a amplitude do sinal amostrado, grandeza que originalmente podia assumir qualquer valor, passa a ser restrita a níveis discretos. Essa quantidade de níveis discretos utilizados para representar um sinal é uma espécie de resolução do sinal e é função do número de bits empregados, sendo 2^n níveis quando são utilizados n bits.

A etapa estudada nesse trabalho é a de codificação do sinal, onde os níveis em que ele foi quantizado são representados por sequências de bits. A quantidade de bits empregados nesse processo diferencia os diversos tipos de codificadores no que

diz respeito à sua taxa. Os codificadores mais eficientes nesse quesito apresentam necessidade de uma quantidade menor de bits para transmitir uma informação e, assim, permitem que mais informação seja transmitida dentro de uma banda ou intervalo de tempo.

Porém, deve existir uma relação de compromisso entre a taxa de codificação e a qualidade do sinal transmitido. Nesse âmbito e no campo de processamento de voz, se destacam os codificadores baseados na técnica CS-ACELP, pois esses apresentam baixa taxa de codificação e boa qualidade do sinal de voz reconstituído, o que justifica seu uso em diversos padrões de telefonia.

Dentre os codificadores CS-ACELP está a recomendação G.729A da ITU (*International Telecommunication Union*), órgão de padronização na área de telecomunicações. A versão A do G.729 é especialmente focada na redução da complexidade computacional do codificador. O fato de ser desenvolvida por um órgão de padronização e de ter foco alinhado com o objetivo desse trabalho faz com que a implementação do G.729A elaborada pela ITU sirva de base para o desenvolvimento do estudo descrito ao longo deste trabalho.

1.2 Proposta do trabalho

O tema do trabalho é o processo de codificação utilizado no codificador de voz G.729A da ITU baseado na técnica CS-ACELP. Esse tema é explorado com o objetivo geral de implementar modificações que otimizem o sistema de codificação deste codificador e reduzam sua complexidade computacional sem alterar a qualidade perceptual do sinal decodificado. Para esse fim, parte-se das recomendações ITU-T G.729 [1] e ITU-T G.729A [2] e das alterações propostas em [3], [4] e [5].

1.2.1 Justificativa

Segundo [6], mais de 25% do esforço computacional de todo o processo de codificação do G.729 é gasto na etapa de *open loop* na busca da melhor excitação do dicionário adaptativo. Portanto, é notório que acelerações nesta etapa do processo de codificação podem levar a economias significativas ao final do processo. Além

disso, estudos recentes ([3], [4] e [5]) mostram que a qualidade do sinal de voz pode ser pouco afetada quando se reduz significativamente a complexidade computacional do codificador G.729A. Essas observações servem de motivação para que outras alterações sejam testadas nos algoritmos do G.729A.

1.2.2 Objetivos

Portanto, o objetivo geral previamente apresentado pode ser desdobrado nos seguintes objetivos específicos: caracterizar um conjunto de sinais de voz que serão codificados e avaliados; reproduzir e sistematizar simplificações propostas em [4] e simplificar o algoritmo de codificação, alterando as etapas de *open loop* e de *closed loop* na busca da melhor excitação do dicionário adaptativo.

1.3 Organização do trabalho

No capítulo 2 é descrita a codificação CS-ACELP, especificamente a empregada no codificador G.729A, assim como os métodos de qualificação de codificadores de voz utilizados na avaliação das modificações introduzidas ao longo deste trabalho. O capítulo segue com a explicação detalhada do funcionamento do codificador do G.729A, em especial da busca do dicionário adaptativo, onde são implementadas as modificações propostas neste trabalho. A finalização é feita com a descrição sucinta do decodificador.

No capítulo 3, o banco de sinais de voz utilizado é caracterizado em relação à sua composição, à sua distribuição de período de *pitch* e ao *pitch* médio dos seus locutores. Além disso, o capítulo 3 descreve as técnicas de decimação empregadas nos sinais de fala e nos cálculos de correlação e apresenta as limitações e a busca proximal aplicadas sobre a de estimativa do atraso do *pitch*. Portanto, esse capítulo reúne as modificações do *open loop* implementadas em trabalhos anteriores.

As modificações implementadas neste trabalho são apresentadas no capítulo 4. Essas modificações ocorrem em duas etapas do processo de codificação. A primeira delas é o *open loop*, onde há alteração nas técnicas utilizadas para se estimar do atraso do *pitch* e expansão das demais técnicas que se aplicam nessa etapa. Já

na etapa de *closed loop*, são descritas a busca proximal da estimativa do atraso do *pitch*, a nova implementação do cálculo do resíduo do filtro LP e a decimação do sinal. A combinação das técnicas implementadas nesse capítulo geram redução da complexidade computacional de até 70,50% em relação à busca do dicionário adaptativo do G.729A original sem alteração da qualidade do sinal reproduzido.

Finalmente, a conclusão resume a dissertação, introduzindo comentários e conclusões a respeito das modificações introduzidas, além de propostas para trabalhos futuros.

Capítulo 2

O G.729A

2.1 Introdução

O G.729 é um codificador de sinais de voz descrito pela recomendação G.729 [1] de 1996 da ITU (*International Telecommunication Union*). Este codificador baseado na estrutura de codificadores CS-ACELP (*Conjugate-Structure-Algebraic-Code-Excited Linear-Prediction*) apresenta alta qualidade e é destinado a operar sinais de voz com frequência na banda telefônica (amostrados a 8 kHz) e codificados com 16 bits por amostra usando codificação PCM.

O G.729 é um dos diversos codificadores de voz existentes baseados em CS-ACELP e a estrutura do próprio G.729 serve de base para diversos outros codificadores. Dentre eles está o G.729A [2], foco deste trabalho e cujas mudanças são implementadas no filtro perceptual, nas buscas nos dicionários adaptativo e fixo e no decodificador, itens descritos nas seções 2.4.4, 2.4.6, 2.4.7 e 2.5, respectivamente. Portanto, este capítulo trata dos aspectos envolvidos no processamento de voz implementado pela recomendação G.729A, tais como: métodos de qualificação de codificadores de voz, processo de codificação empregado no G.729A e seu processo de decodificação.

O G.729A é apenas um dos diversos anexos da recomendação G.729, cujas implementações são resumidamente descritas abaixo:

- Anexo A: este anexo contém modificações no algoritmo do G.729 original que visam reduzir sua complexidade computacional.

- Anexo B: adiciona um esquema de compressão de silêncio ao algoritmo do G.729 original.
- Anexo C: implementação em ponto flutuante para o G.729 original e do anexo A. Abrange a recomendação C+, que implementa em ponto flutuante o G.729 anexos B, D e E.
- Anexo D: implementa um codificador de sinais de fala baseado na técnica CS-ACELP com taxa de 6.4 kbit/s.
- Anexo E: implementa um codificador de sinais de fala baseado na técnica CS-ACELP com taxa de 11.8 kbit/s destinado a uma maior gama de sinais, incluindo sinais com ruído de fundo e música.
- Anexo F: implementa a função compressão de silêncio do G.729 anexo B no anexo D.
- Anexo G: implementa a função compressão de silêncio do G.729 anexo B no anexo E.
- Anexo H: implementação combinada do G.729 original e dos anexos D e E, gerando um codificador com taxa de 6,4, 8 ou 11,8 kbit/s.
- Anexo I: implementação combinada do G.729 original e dos anexos B, D e E, gerando um codificador com supressão de silêncio e taxa de 6,4, 8 ou 11,8 kbit/s.
- Anexo J: implementação do G.729 original com taxa que varia entre 8 e 32 kbit/s.

Os métodos de qualificação de codificadores de voz utilizados ao longo deste texto são o PESQ e o MOS, cujas descrições são apresentadas na seção 2.2. A descrição geral do funcionamento de codificadores CS-ACELP, é feita na seção 2.3, enquanto as seções 2.4 e 2.5 apresentam a estrutura e funcionamento do codificador e do decodificador do G.729A, respectivamente. No processo de codificação merecem atenção especial a análise por síntese e a busca no dicionário adaptativo, cujo entendimento é fundamental para compreensão deste trabalho.

2.2 Métodos de qualificação de codificadores de VOZ

As diferenças entre as estruturas do G.729 e seus anexos têm resultados diretos na qualidade do sinal reproduzido por eles. Com o intuito de quantificar estas diferenças, são utilizados os métodos de qualificação de codificadores de voz descritos nesta seção.

2.2.1 MOS

O *Mean Opinion Score* (MOS) é um avaliador de qualidade de sinais de voz descrito na recomendação P.800.1 da ITU [7]. Esse método de qualificação de codificador consiste na reunião de um grupo de pessoas, que compara a qualidade do sinal codificado com o sinal original, permitindo a quantificação da avaliação subjetiva de sinais de voz através da escala MOS. Essa escala é exibida na Tabela 2.1 e relaciona a percepção do avaliador com uma nota. Ao final do processo, a nota MOS de um determinado sinal é a média das notas subjetivas dadas por cada pessoa ao sinal em questão.

Tabela 2.1: Escala MOS

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

Porém, realizar testes com seres humanos pode ser um processo dispendioso. Este fato impulsiona a utilização de outros métodos de qualificação de sinais de voz.

2.2.2 PESQ

O *Perceptual Evaluation of Speech Quality* (PESQ) consta na Recomendação P862 da ITU [8] e tem como objetivo estimar a nota MOS de um sinal de voz. O PESQ

oferece precisão aceitável em seus resultados em relação à clareza da voz quando afetados pelos seguintes parâmetros ou processos:

- codificadores de forma de onda;
- codificadores paramétricos e híbridos a partir de 4 kbps;
- erros no canal de transmissão;
- perdas de pacotes.

Para calcular a nota PESQ é feita a comparação do sinal original com o sinal codificado. O cálculo da estimativa da nota MOS é dado por 2.1[8]:

$$PESQMOS = 0,999 + \frac{4}{1 + e^{-1,4945 * PESQ + 4,6607}}. \quad (2.1)$$

Este método tem como vantagem o fato de permitir que sejam testadas diversas configurações de codificação e que apenas um número reduzido destas configurações tenha de ser submetida à avaliação de seres humanos, que é mais dispendiosa.

2.3 Codificadores CS-ACELP

Codificadores CS-ACELP partem do princípio de que a voz humana, apesar de não ser estritamente estacionária, é estacionária por partes se forem considerados segmentos de 10 a 30 ms [9]. Por esta ótica, a voz humana pode ser interpretada como a resposta de um filtro a uma excitação, conforme esquematizado na Figura 2.1. Assim, a tarefa desenvolvida pelos codificadores CS-ACELP é usar mecanismos de predição linear para determinar os parâmetros do filtro LP (*Linear Prediction*) a partir da análise do sinal e determinar as excitações de entrada. Estas excitações são oriundas de dicionários de excitações, cujas buscas serão descritas nas seções 2.4.6 e 2.4.7. Depois da codificação, as informações relativas ao filtro de síntese e as excitações dos dicionários são transmitidas ao decodificador para que este reconstrua o sinal estimado no codificador.



Figura 2.1: Processo de geração de voz humana no codificador LPC.

2.4 Codificador

No G.729A, cada segmento de sinal de voz de 10 ms passa por um pré-processamento onde é escalado e tem suas baixas frequências filtradas. No codificador, também são feitos a análise LP, que calcula os coeficientes do filtro LP; o mapeamento desses coeficientes em LSP (*Line Spectral Pairs*); sua interpolação e a escolha das excitações dos dicionários fixo e adaptativo, uma para cada sub-segmento de 5 ms. Todas essas etapas são detalhadas nas subseções a seguir.

2.4.1 Pré-processamento

Na etapa de pré-processamento do codificador G.729A, o sinal de entrada é escalado por um fator $\frac{1}{2}$ e processado por um filtro passa-altas, cujos objetivos são respectivamente evitar o *overflow* na implementação em ponto fixo e eliminar componentes de baixa frequência. A implementação do escalamento e do filtro passa-altas de frequência de corte 140 Hz é feita por um filtro da forma [1]:

$$F_{pe}(z) = \frac{0,46363718 - 0,927247052z^{-1} + 0,463637181z^{-2}}{1 - 1,9059465z^{-1} + 0,9114024z^{-2}}, \quad (2.2)$$

cuja resposta em magnitude é apresentada na Figura 2.2.

2.4.2 Janelamento no sinal de voz

A janela $\omega_{lp}(n)$ da análise LP, com formato exibido na Figura 2.3, 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 senoidal [1]. Essa janela atua em 240 amostras, sendo 120 amostras de segmentos anteriores, 80 amostras do segmento atual e 40 amostras do segmento futuro.

$$\omega_{lp}(n) = \begin{cases} 0,54 - 0,46 \cos \frac{2\pi n}{399}, & n = 0, \dots, 199 \\ \cos \frac{2\pi(n-200)}{159}, & n = 200, \dots, 239 \end{cases} \quad (2.3)$$

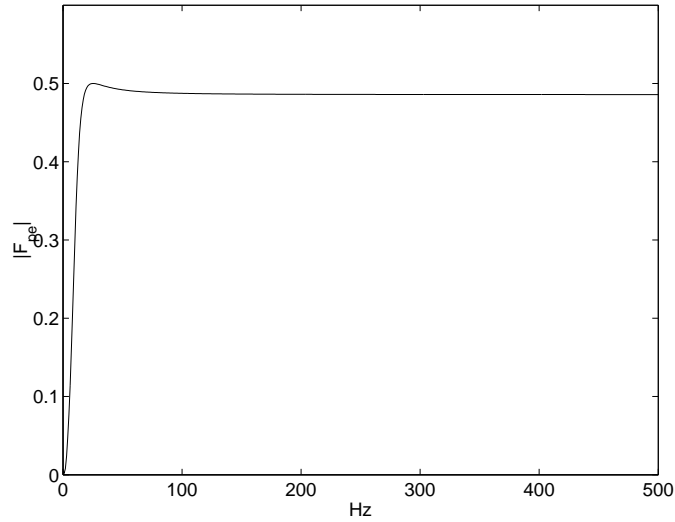


Figura 2.2: Resposta em magnitude do filtro de pré-processamento F_{pe} .

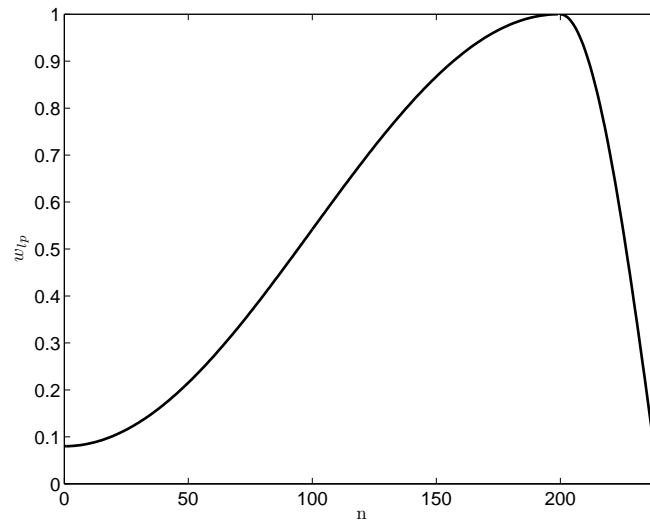


Figura 2.3: Janela de predição linear w_{lp} .

O sinal janelado é utilizado na análise LPC (*Linear Prediction Coding*) descrita a seguir.

2.4.3 Análise LPC, quantização e interpolação

A codificação de sinal consiste em representá-lo por parâmetros capazes de reproduzir suas características. Para este fim, no G.729A, a cada segmento de 10 ms é feita a análise LPC, onde a autocorrelação do sinal é usada para o cálculo dos

coeficientes LPC a_i do filtro LP de ordem $M = 10$ dado por [1]:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{i=1}^M a_i z^{-i}}. \quad (2.4)$$

Por questões de sensibilidade a erros de quantização, os coeficientes LPC são posteriormente mapeados em coeficientes LSP, cujos valores são quantizados com quantização vetorial de 2 estágios com total de 18 bits e interpolados com coeficientes LSP de sub-segmentos adjacentes. Esta interpolação é feita apenas no primeiro sub-segmento e é realizada recalculando-se os valores dos 10 coeficientes LSP quantizados no sub-segmento atual $\hat{q}_i^{(1)}$, que passam a ser a média entre seus valores atuais $\hat{q}_i^{(atual)}$ e os valores destes coeficientes no sub-segmento anterior $\hat{q}_i^{(anterior)}$, isto é [1]:

$$\hat{q}_i^{(1)} = 0,5\hat{q}_i^{(atual)} + 0,5\hat{q}_i^{(anterior)} \quad i = 1, \dots, 10. \quad (2.5)$$

2.4.4 Filtro perceptivo

O ouvido humano é mais sensível a erros e ruídos em componentes de menor amplitude na frequência do que erros e ruídos em componentes de maior amplitude [9]. Sendo assim, as componentes de menor amplitude na frequência são consideradas mais importantes no cálculo do erro do processo de síntese. Esta variação no grau de importância das componentes em função de sua frequência é implementada no G.729A pelo filtro perceptivo $W(z)$, que utiliza os parâmetros quantizados do filtro LP e é dado por [2]:

$$W(z) = \frac{A(z)}{A(z/\gamma)}, \quad (2.6)$$

com coeficiente de ponderação $\gamma = 0,75$ [2].

O filtro de síntese é resultado da atuação conjunta dos filtros LP e perceptual, portanto leva em consideração tanto informações extraídas do sinal (contribuição do filtro LP) quanto informações sobre a percepção humana sobre este sinal (contribuição do filtro perceptual). Matematicamente, a atuação conjunta destes dois filtros é representada por:

$$H_w(z) = H(z)W(z) = \frac{1}{A(z)} \frac{A(z)}{A(\frac{z}{\gamma})} = \frac{1}{A(\frac{z}{\gamma})}. \quad (2.7)$$

2.4.5 Análise por síntese

A análise por síntese (*Analysis-by-Synthesis*, AbS) é o processo utilizado para determinar a melhor estimativa do sinal de fala a ser codificado $s(n)$. Para isso, o codificador extrai parâmetros do sinal $s(n)$ e utiliza esses parâmetros para sintetizar um conjunto de excitações, gerando estimativas do sinal de fala. Essas estimativas são avaliadas segundo um critério específico e a excitação ótima é aquela que atende melhor a esse critério. No G.729A, a análise por síntese é feita em cada sub-segmento do sinal e é caracterizada pelas etapas seguintes.

1. As excitações dos dicionários são sintetizadas pelo filtro de síntese;
2. a excitação cuja saída do filtro de síntese possui maior correlação com o sinal alvo é a excitação ótima, que é usada para reconstituir o sinal.

Observa-se que, para a realização da análise por síntese como ela foi descrita, seria necessário sintetizar todas as combinações de todas as excitações dos dicionários fixo e adaptativo, o que acarretaria em um esforço computacional impraticável. Para evitar o emprego desse esforço, são desenvolvidas técnicas simplificadoras para a busca de um número reduzido de excitações cujas combinações serão sintetizadas e analisadas. Esses mecanismos de busca são descritos nas seções 2.4.6 e 2.4.7

2.4.6 Busca no dicionário adaptativo

As variações em segmentos consecutivos de sinal não são bruscas, por isto se faz uso do dicionário adaptativo, que é um dicionário em forma de vetor composto por 256 amostras. Nele cada excitação ν é composta por 40 amostras passadas deslocadas, conforme mostra a Figura 2.4, onde ν_1 e ν_2 são excitações com *lag* 1 e 2, respectivamente. A busca neste dicionário consiste em encontrar a melhor excitação para a recomposição do sinal e pode ser dividida em três etapas consecutivas: o *open loop*, onde se estima o atraso de *pitch* T_{op} da excitação; o *closed loop*, onde se utiliza a estimativa do *pitch* para buscar a excitação ótima do dicionário adaptativo; e busca interpolada, onde se faz uma busca considerando valores fracionários para o atraso de *pitch*.

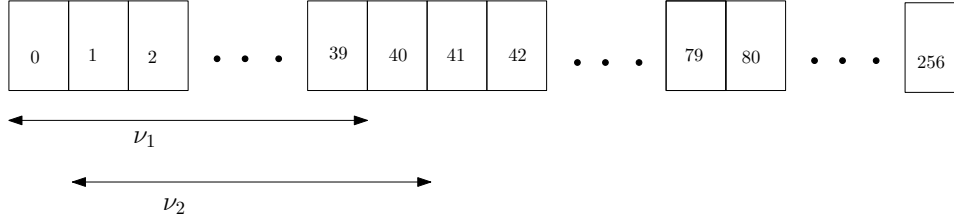


Figura 2.4: Estrutura do dicionário adaptativo.

O objetivo principal da etapa de *open loop* é fornecer uma estimativa do atraso do *pitch*.

Para encontrar a estimativa do atraso T_{op} é calculada a autocorrelação normalizada $R'(k)$ do sinal de fala ponderado (*weighted speech signal*) $s_w(n)$ decimado por 2, como definido nas equações a seguir [2]:

$$s_w(n) = r(n) - \sum_{i=1}^M a'_i s(n-i), \quad n = 0, \dots, 39; \quad (2.8)$$

$$r(n) = s(n) + \sum_{i=1}^M a'_i s(n-i), \quad n = 0, \dots, 39; \quad (2.9)$$

$$R'(k) = \frac{\sum_{n=0}^{39} s_w(2n)s_w(2n-k)}{\sqrt{\sum_n s_w^2(2n-k)}}. \quad (2.10)$$

onde $r(n)$ é o resíduo do filtro LP e a'_i são os coeficientes do filtro perceptual modificado pelo filtro passa-baixas $G(z)$ da forma [2]:

$$G(z) = \frac{1}{1 - 0,7z^{-1}}. \quad (2.11)$$

A autocorrelação normalizada é feita para os intervalos (a), (b) e (c), caracterizados por:

- (a): $20 \leq k \leq 39$;
- (b): $40 \leq k \leq 79$;
- (c): $80 \leq k \leq 142$ (k é par).

O cálculo de $R'(k)$ nestes intervalos gera T_a , T_b e T_c^* , que são os valores de k que maximizam a função $R'(k)$ nos intervalos (a) , (b) e (c) respectivamente. Neste ponto, é empregado um recurso de redução de complexidade característico do G.729A: a decimação por 2 no cálculo da autocorrelação normalizada no intervalo (c) , que faz com que $R'(k)$ seja calculada para 32 valores de k e não para os 64 necessários caso este recurso não fosse utilizado. Em seguida, é feita uma busca proximal em torno de T_c^* , calculando-se o valor de $R'(k)$ para os valores $(T_c^* - 1)$ e $(T_c^* + 1)$ e fazendo-se T_c assumir o valor dentre $(T_c^* - 1)$, T_c^* e $(T_c^* + 1)$ que maximiza a autocorrelação normalizada.

Os valores T_i (com $i = a, b, c$) são candidatos a T_{op} . A escolha do valor definitivo de T_{op} é feita através da comparação dos valores de $R'(T_i)$ depois de serem ponderados por pesos que privilegiam o intervalo (a) em detrimento do (b) e o intervalo (b) em detrimento do (c) . Esta ponderação tem por finalidade evitar que os valores de T_i sejam múltiplos harmônicos do verdadeiro período de *pitch* e é dada pelo algoritmo:

- se $|2T_b - T_c| < 5$ ou $|3T_b - T_c| < 7$
 $max_b = R'(T_b) + 0,25R'(T_c)$
- se $|2T_a - T_b| < 5$ ou $|3T_a - T_b| < 7$
 $max_a = R'(T_a) + 0,20R'(T_b)$,

Os valores max_a , max_b e $R'(T_c)$, associados a T_a , T_b ou T_c , respectivamente, são comparados e T_{op} assume o valor do T_i referente ao maior deles.

No *closed loop*, o resultado da análise do pitch é usado para buscar o sinal alvo do dicionário adaptativo $x(n)$, dado por [2]:

$$x(n) = H_w(z)\{r(n)\}. \quad (2.12)$$

onde $r(n)$ é o resíduo do filtro LP.

O sinal $x(n)$ é empregado na busca proximal feita em torno de T_{op} para a estimativa do atraso do *pitch*. Para isto, realiza-se o cálculo da correlação cruzada $R_{xy}(k)$ entre o sinal $x(n)$ e $y_k(n)$, onde $y_k(n)$ é resultado da convolução da excitação $u(n)$,

com atraso k , do dicionário adaptativo passada pelo filtro de síntese com a resposta ao impulso desse mesmo filtro $h(n)$. A correlação cruzada é dada por [2]:

$$R_{xy}(k) = \frac{\sum_{n=0}^{39} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{39} y_k^2(n)}}, \quad (2.13)$$

onde os valores de k para o cálculo de T_1 (estimativa do período de *pitch* no primeiro sub-segmento) são limitados no intervalo $(T_{op} - 3) \leq k \leq (T_{op} + 3)$ e para o cálculo de T_2 (estimativa do período de *pitch* no segundo sub-segmento) são limitados no intervalo $(T_1 - 5) \leq k \leq (T_1 + 4)$. Ao final do *closed loop*, T_1 e T_2 são os valores de k que maximizam o numerador da equação (2.13), para os primeiro e segundo sub-segmentos e que serão utilizados da etapa seguinte. Com o intuito de se reduzir a complexidade computacional, maximiza-se o numerador de $R_{xy}(k)$ dado por [2]:

$$R_{xy}(k) = \sum_{n=0}^{39} x(n)y_k(n) = \sum_{n=0}^{39} x_b(n)u(n-k), \quad (2.14)$$

onde $x_b(n)$ é o sinal alvo filtrado pelo filtro de síntese (*backward filtered*), é .

Na última etapa da busca pela excitação $u_a(n)$ do dicionário adaptativo, é feita a busca interpolada, que é uma análise para valores fracionários com resolução 1/3 ao redor dos atrasos T_1 e T_2 . Disto resultam T_1^* e T_2^* , que determinam unicamente as excitações do dicionário adaptativo para o primeiro e segundo sub-segmentos.

Para o cálculo do ganho do dicionário adaptativo G_a normaliza-se a correlação cruzada $R_{xy}(n)$ através da relação [1]:

$$G_a = \frac{\sum_{n=0}^{39} x(n)y_k(n)}{\sum_{n=0}^{39} y_k^2(n)}. \quad (2.15)$$

A busca no dicionário adaptativo no G.729A é um dos itens que o difere do G.729, pois neste se reduz o número de operações algébricas feitas, o que é uma das maneiras de reduzir a complexidade computacional de um processo. No caso do cálculo de

T_{op} no G.729A são necessárias M multiplicações e A adições dadas por [5]:

$$M = (124 - 32) \times 40 = 3680 \quad (2.16)$$

$$A = (124 - 32) \times 39 = 3588. \quad (2.17)$$

Já para a etapa do *closed loop* as a quantidade de operações algébricas empregadas no cálculo do sinal $y_k(n)$ e na maximização da correlação cruzada $R_{xy}(n)$:

$$M = 2 \times 20 \times (40 + 1) + (7 + 10) \times 40 = 2320 \quad (2.18)$$

$$A = 2 \times 20 \times (39) + (6 + 9) \times 39 = 2145. \quad (2.19)$$

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 (2.16) a (2.19).

2.4.7 Busca no dicionário fixo

Com base no sinal sinal alvo $x'(n)$ do dicionário fixo, dado por:

$$x'(n) = s(n) - G_a \times H_w\{x(n)\}, \quad (2.20)$$

e utilizando o critério da máxima correlação, é feita a busca da melhor excitação do dicionário fixo, que é um dicionário algébrico composto por vetores de 36 amostras nulas e 4 pulsos de amplitude unitária $\delta(n)$ de acordo com [1]:

$$c(n) = \sum_{i=0}^3 s_i \delta(n - m_i), \quad n = 0, \dots, 39, \quad (2.21)$$

onde m_i são valores inteiros que variam de 0 a 39 e indicam as posições em que os pulsos se encontram e s_i são os sinais desses pulsos. Por isto, a busca é essencialmente o processo de obtenção das posições e sinais s_i das 4 amostras não nulas e é feita utilizando-se a busca em profundidade em uma estrutura de árvore.

O ganho do dicionário fixo G_f é resultado da comparação entre a energia da contribuição atual do dicionário fixo, que é estimada utilizando-se mecanismos de predição linear, e a energia da contribuição anterior deste mesmo dicionário. O cálculo de G_f não será detalhado por fugir ao escopo deste trabalho.

O sinal $u(n)$ é resultante da soma das excitações $u_a(n)$ e $u_f(n)$ ponderadas pelos seus respectivos ganhos:

$$u(n) = G_a u_a(n) + G_f u_f(n), \quad (2.22)$$

onde a excitação do dicionário adaptativo $u_a(n)$ é formada por uma versão atrasada da excitação $u(n)$ do filtro de síntese de segmentos anteriores e a excitação do dicionário fixo $u_f(n)$ é formada por um vetor com apenas quatro amostras não nulas de amplitude unitária. Então, é feita uma retroalimentação do sinal $u(n)$ no dicionário adaptativo. Os ganhos dos dicionários são quantizados, codificados e transmitidos juntamente com os demais parâmetros do codificador. Porém, como este trabalho é focado na aceleração da busca do dicionário adaptativo, a busca do dicionário fixo não será descrita além do necessário.

O diagrama de blocos do processo de codificação como um todo é apresentado na Figura 2.5.

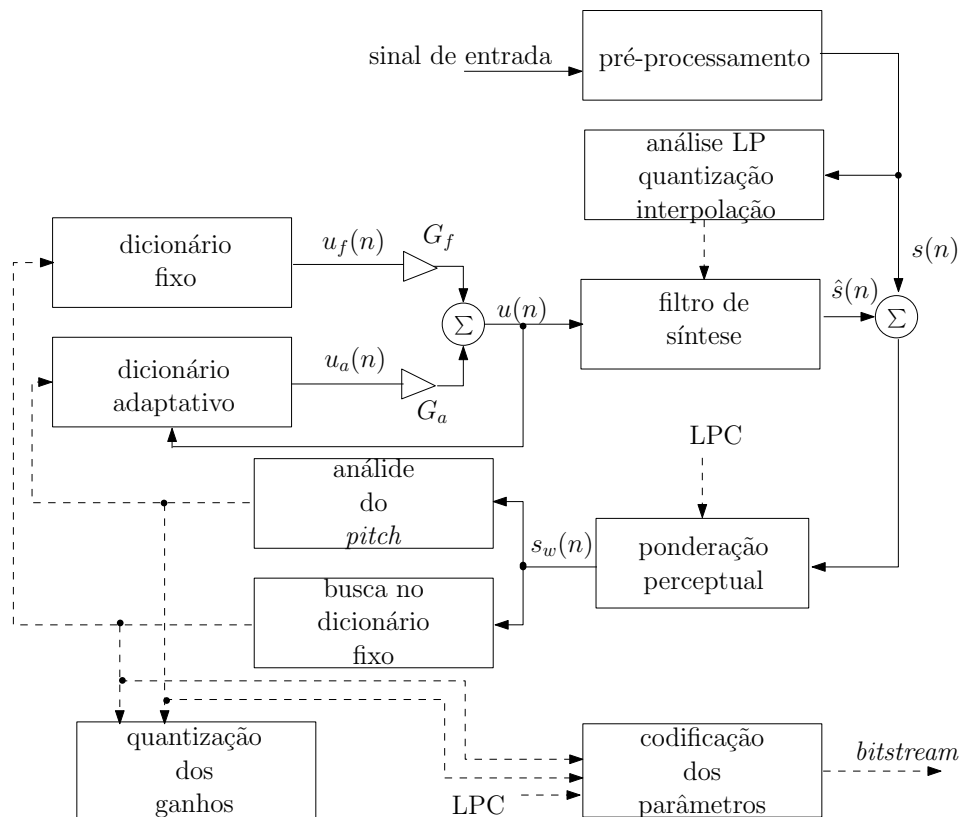


Figura 2.5: Diagrama de blocos do codificador G.729A

2.5 Decodificador

Através dos ganhos dos dicionários (G_a e G_f), dos coeficientes LPC, do atraso do dicionário adaptativo (T) e dos índices e sinais das excitações do dicionário fixo (C e S respectivamente), o decodificador recupera o filtro de síntese e as excitações dos dicionários fixo e adaptativo. As excitações são combinadas, filtradas e passam por um pós-processamento, onde são aplicados mecanismos de compensação de erros, controle de ganho, escalamento e um filtro passa-altas, que não serão detalhados neste trabalho. O resultado desse processo é a estimativa do sinal de voz $\hat{s}(n)$, conforme pode ser visualizado na Figura 2.6.

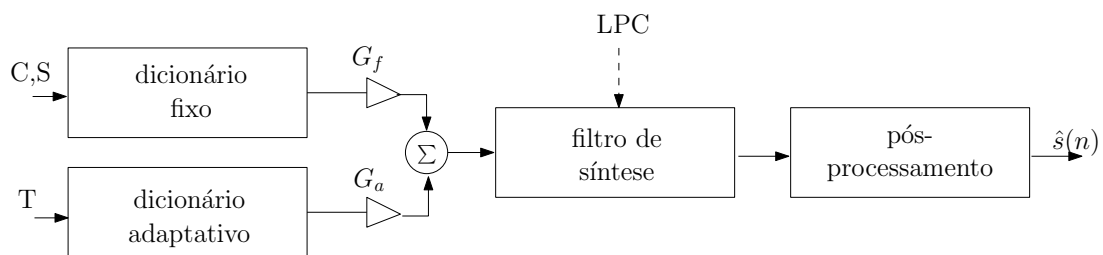


Figura 2.6: Princípio de funcionamento do decodificador do G.729.

2.6 Conclusão

Este capítulo descreveu os métodos de avaliação de codificadores de voz MOS e PESQ, que serão utilizados ao longo deste trabalho. Além disto, foram apresentados tanto o funcionamento geral do codificador de voz G.729A quanto o processamento e análise do sinal de voz realizados por este codificador.

Dentre as etapas apresentadas neste capítulo destacam-se a análise por síntese e a busca no dicionário adaptativo. A primeira fornece a descrição da estrutura do processo de codificação e é fundamental para o entendimento do codificador utilizado neste trabalho. Já a busca no dicionário adaptativo destaca-se por ser a área em que são realizadas as alterações propostas neste trabalho e que serão descritas nos capítulos que se seguem.

Capítulo 3

Trabalhos anteriores

3.1 Introdução

O principal objetivo deste capítulo é apresentar um resumo das mudanças no G.729A anteriores à elaboração deste trabalho e que, portanto, serviram de referência e motivação. Estas modificações serão descritas na seção 3.3. Porém, antes de dar início à descrição destas modificações, a seção 3.2 descreverá o conjunto de sinais de voz sobre o qual as alterações serão testadas.

3.2 Banco de voz

Para que as modificações implementadas em um codificador sejam analisadas é necessário um conjunto de sinais de voz onde os efeitos da codificação possam ser avaliados. Para este fim, será utilizado ao longo deste trabalho um conjunto de sinais chamado banco de voz BD1, formado por sinas de fala retirados do *Open Speech Repository* (OSR) e montado por Thiago de Moura Prego [5].

Nesta seção será feito um estudo de caracterização do banco BD1 e serão descritas sua formação e suas características estatísticas.

O banco BD1 é formado por 40 sinais de fala distribuídos da seguinte forma:

- 8 sinais em mandarim;
- 8 sinais em francês;

- 8 sinais em indiano;
- 8 sinais em inglês britânico;
- 8 sinais em inglês americano;

Seis dos sinais em inglês americano empregam locutores femininos, todos os demais sinais de fala usam locutores masculinos.

3.2.1 Caracterização do banco de voz

O desempenho do G.729A no banco BD1 é avaliado através da estimativa PESQ-MOS, que nesse banco tem valor médio de 3,75. Foram feitas duas caracterizações do comportamento estatístico do G.729A aplicado ao banco BD1. Na primeira, observa-se a frequência de ocorrência da igualdade $k = T_i$, com $20 \leq k \leq 39$ se $i = (a)$, $40 \leq k \leq 79$ se $i = (b)$ e $80 \leq k \leq 143$ se $i = (c)$, conforme descrito no capítulo anterior. Ou seja, essa análise descreve o quanto cada valor de k é frequente como aquele que maximiza a correlação no intervalo i . Como resultado, a distribuição dos valores de T_a , T_b e T_c dentro dos intervalos (a) , (b) e (c) segundo sua frequência é apresentada nas Figuras 3.1, 3.2 e 3.3, onde verificam-se as modas $T_a = 39$, $T_b = 45$ e $T_c = 143$ de cada intervalo.

A segunda forma de analisar estatisticamente o comportamento do G.729A no banco BD1 enfatiza a quantificação da participação efetiva de cada valor de T_i na estimativa do atraso do período de *pitch*. Para isto é analisado o percentual de vezes que cada valor de T_i aparece como T_{op} em cada intervalo i . A Figura 3.4 representa esse resultado para os intervalos a , b e c .

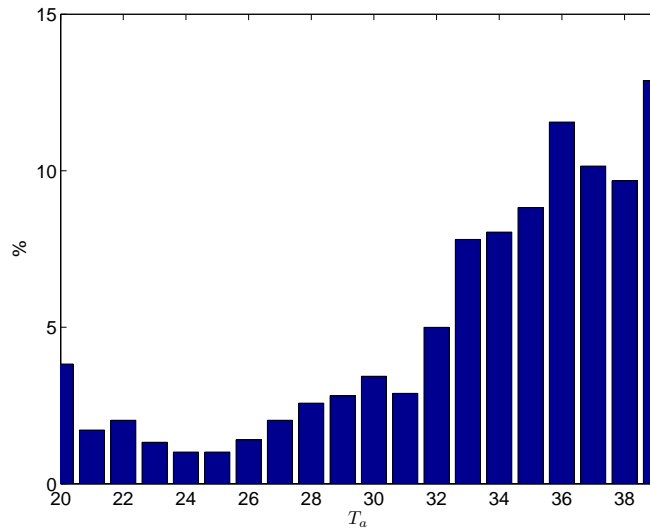


Figura 3.1: Distribuição de T_a ao longo do intervalo (a) no G.729A original

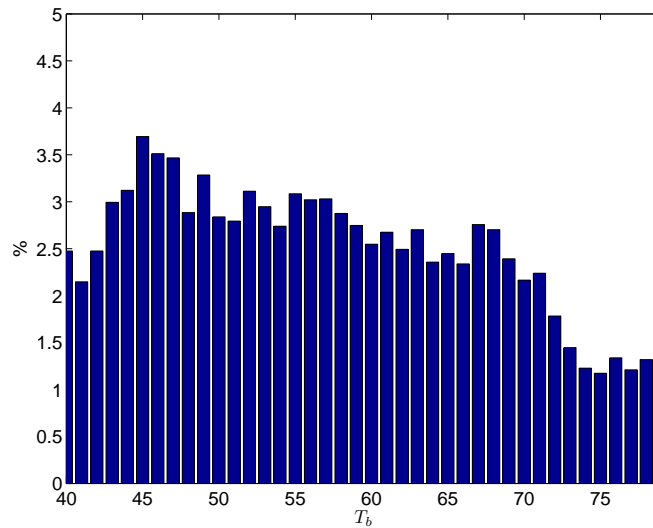


Figura 3.2: Distribuição de T_b ao longo do intervalo (b) no G.729A original

A Tabela 3.1 apresenta o percentual de T_{op} que se encontra em cada um dos intervalos e reforça algo que é possível notar na Figura 3.4: a participação do intervalo (b) é significativamente maior que a dos demais intervalos para este banco de sinais.

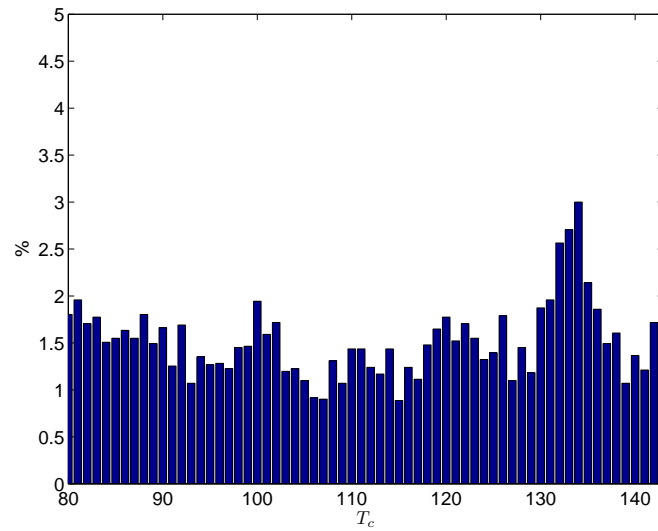


Figura 3.3: Distribuição de T_c ao longo do intervalo (c) no G.729A original

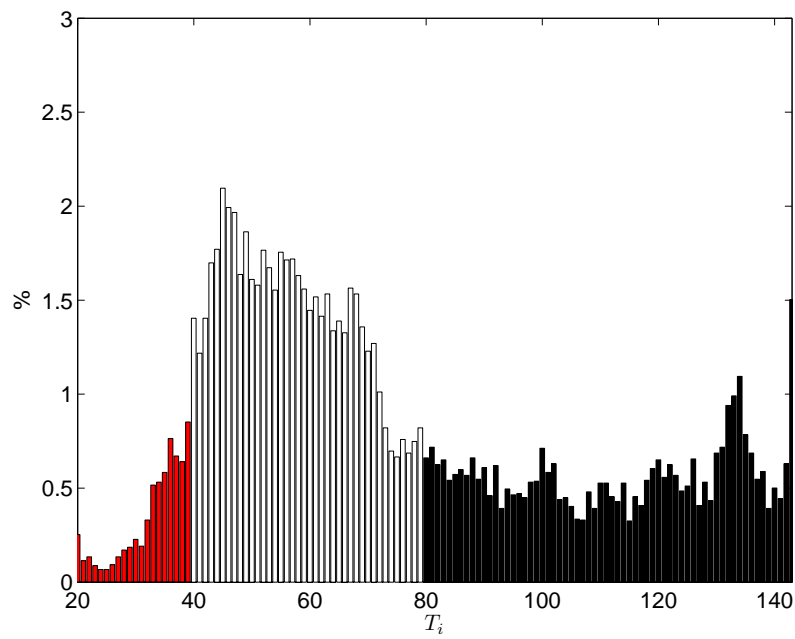


Figura 3.4: Distribuição de T_{op} ao longo dos intervalos (a), (b) e (c) no G.729A original

Tabela 3.1: Distribuição de T_{op} ao longo dos intervalos (a), (b) e (c) no G.729A original.

Intervalo	Percentual
(a)	6,6126
(b)	56,7365
(c)	36,6508

3.2.2 Análise dos locutores do banco de voz

Outra característica importante do banco BD1 diz respeito às frequências de período de *pitch* dos segmentos sonoros do sinais de fala presentes nele. Essa grandeza é chamada F_{pitch} e é dada por:

$$F_{pitch} = \frac{F_s}{k}, \quad (3.1)$$

onde F_s é a frequência de amostragem e $20 \leq k \leq 143$. A distribuição de F_{pitch} é apresentada no histograma da Figura 3.5

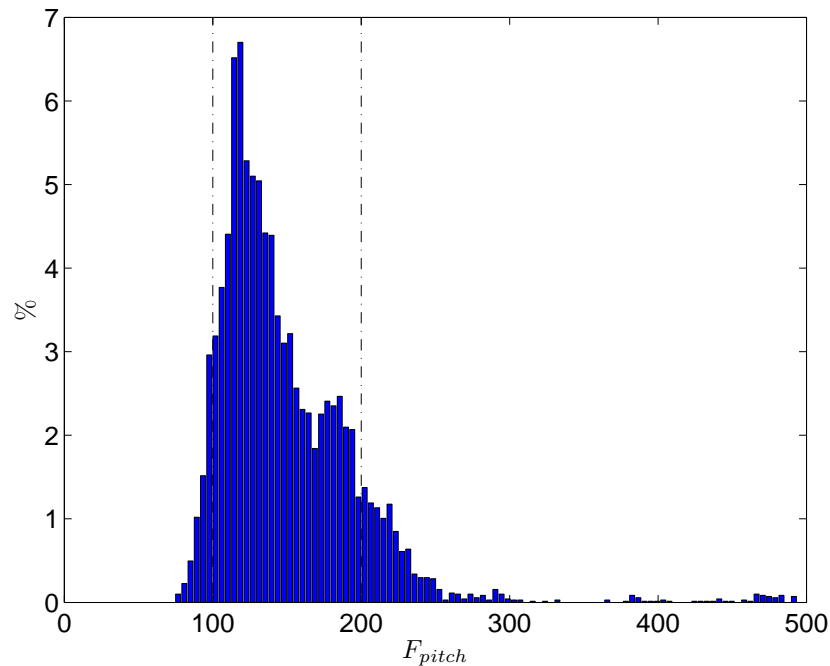


Figura 3.5: Distribuição de frequência dos sinais de fala do banco BD1.

Assim, o intervalo (a) compreende frequências de 205,13 a 400 Hz; o intervalo (b), as frequências de 101,267 a 200 Hz e o intervalo (c), as frequências de 55,94 a 100

Hz. Observando-se esses intervalos de frequência na Figura 3.5, nota-se que cerca de 11,22% estão no intervalo (a), 82,46% no intervalo (b) e 6,32% no intervalo (c).

Deste modo, está caracterizado o banco de voz herdado de trabalhos anteriores e que servirá de insumo para avaliações do impacto das modificações sobre a qualidade do codificador de fala G.729A.

3.3 Modificações de trabalhos anteriores

3.3.1 Fixação de T_{op} para um dado intervalo

Em [3], foi testado um método de redução de complexidade computacional baseado na observação da distribuição de $T_a = \bar{\tau}_a$, $T_b = \bar{\tau}_b$ e $T_c = \bar{\tau}_c$ ao longo de seus respectivos intervalos (ver Figuras 3.1, 3.2 e 3.3). Deste modo, o artigo [3] sugere que T_i seja fixado em $\bar{\tau}_i$ em um intervalo i , enquanto o cálculo da estimativa do atraso do período de *pitch* T_{op} nos demais intervalos se mantém inalterado, gerando as configurações I, II e III conforme a seguinte identificação:

- I: T_a é mantido fixo em sua moda $\bar{\tau}_a$, enquanto as buscas nos intervalos (b) e (c) são mantidas como no G.729A padrão.
- II: T_b é mantido fixo em sua moda $\bar{\tau}_b$, enquanto as buscas nos intervalos (a) e (c) são mantidas como no G.729A padrão.
- III: T_c é fixado em sua moda $\bar{\tau}_c$, enquanto as buscas nos intervalos (a) e (b) são mantidas como no G.729A padrão.

Para o banco de voz BD1 e utilizando o codificador G.729A, os resultados apresentados em [3] e reproduzidos na Tabela 3.2 mostram que na configuração III não há alteração na nota PESQMOS. Portanto, esta referência mostra ser possível reduzir o número de multiplicações necessárias para o cálculo da autocorrelação do sinal em um segmento para o valor $M = (124 - 63) \times 40 = 2240$ sem comprometer a qualidade do sinal.

Tabela 3.2: Fixação de T_i

Configuração	PESQMOS
<i>I</i>	3,70
<i>II</i>	3,49
<i>III</i>	3,75

3.3.2 Decimação do sinal no tempo

Conforme evidenciado na equação (2.10), o G.729A calcula a autocorrelação do sinal de fala ponderado usando fator de decimação 2, o que tem como objetivo reduzir a complexidade computacional do G.729. Essa idéia é estendida em [4] através da implementação de um fator de decimação no tempo D_c variável, com $2 \leq D_c \leq 10$. Considerando-se os efeitos da decimação do sinal no tempo isoladamente, o número de multiplicações necessárias para o cálculo da autocorrelação passa a ser dado por:

$$M = 124 \times \left\lfloor \frac{79}{D_c} + 1 \right\rfloor. \quad (3.2)$$

A redução da complexidade computacional do cálculo da autocorrelação é acompanhada pela redução da qualidade do sinal reproduzido.

3.3.3 Decimação do atraso k

Outro recurso utilizado no G.729A que é estendido em [4] diz respeito à decimação do *lag* no cálculo da autocorrelação no intervalo (*c*), onde o G.729A padrão usa fator de decimação 2, já que k é par. Essa característica sugere a extrapolação da decimação para os intervalos (*a*) e (*b*) e a variação do parâmetro de decimação do *lag* D_t , fazendo $1 \leq D_t \leq 10$. O número de multiplicações necessárias para o cálculo da autocorrelação, que passa a ser calculada apenas para os valores $k\Delta_t$, é dado por:

$$M = \left\lfloor \frac{123}{D_t} + 1 \right\rfloor \times \left\lfloor \frac{79}{D_c} + 1 \right\rfloor, \quad (3.3)$$

e sua redução é acompanhada da queda de qualidade do sinal segundo [4].

3.3.4 Análise na vizinhança de T_{op}

A combinação das três alterações apresentadas anteriormente reduz a complexidade computacional do processo de codificação, porém este benefício traz o ônus da

redução da qualidade do sinal reproduzido. Visando minimizar este efeito a custo de um pequeno aumento na complexidade computacional é introduzido o parâmetro Δ_t , que delimita uma vizinhança de T_{op} em que a autocorrelação é calculada sem decimação do *lag*[4]. Assim, no trecho $(T_{op} - \Delta_t) \leq k \leq (T_{op} + \Delta_t)$ recalcula-se $R(k)$, o que eleva o número de multiplicações necessárias em um segmento para [4]:

$$M = \left\lfloor \frac{123}{D_t} + 1 \right\rfloor \times \left\lfloor \frac{79}{D_c} + 1 \right\rfloor + 2 \times \Delta_t \times \left\lfloor \frac{79}{D_c} + 1 \right\rfloor, \quad (3.4)$$

tendo como resultado a elevação da qualidade de codificação.

3.3.5 Análise de estimativas de *pitch* distantes

Mesmo com a implementação da análise na vizinhança de T_{op} , a qualidade do sinal continua bastante afetada pelos parâmetros D_c e D_t . A segunda proposta de [4] para a recuperação da qualidade do sinal é a análise da vizinhança não só de T_{op} , mas também nas N_t melhores estimativas de T_{op} , elevando-se o número de multiplicações em $N_t \times (2\Delta_t + 1)$ para cada segmento em que se calcula a autocorrelação.

3.3.6 Reutilização do atraso do dicionário adaptativo

O cálculo da estimativa T_{op} do atraso do *pitch* obtida no *open loop* utiliza uma característica importante a respeito do contorno do *pitch*. Esse contorno se inicia na fronteira entre trechos de fala do tipo surdo e sonoro (UV-V) ou dos tipos silêncio e sonoro (S-V), caracterizados da seguinte maneira:

- sonoro - um trem de impulsos (ou pulsos glotais);
- surdo - pode ser considerado ruído branco;
- silêncio - ocorre ausência de excitação.

Baseando-se no fato de que existe uma grande probabilidade de segmentos próximos de fala possuírem períodos de *pitch* próximos e, conseqüentemente, valores semelhantes para T_{op} , a reutilização do atraso do dicionário adaptativo é apresentada [5] como recurso de redução da complexidade computacional do codificador G.729A.

A técnica de reutilização do atraso do dicionário adaptativo utiliza a função WDLSP (*weighted delta linear spectral pairs*) para determinar as fronteiras UV-V e S-V. Essa função WD-LSP para o i -ésimo segmento é dada por:

$$F_i = \sum_{k=1}^{10} w_k \times [LSP_i(k) - LSP_{i-1}(k)]^2, \quad (3.5)$$

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

Assim, se o valor desta função for superior a um determinado limiar η , o atraso T_{op} é reestimado. Caso contrário, pode-se assumir que não houve alteração significativa no valor do *pitch* e portanto, reutiliza-se o valor da estimativa do atraso do segundo sub-segmento T_2 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.

A Figura 3.6 e a Tabela 3.3 mostram os pontos dos experimentos descritos em [5] cuja combinação dos parâmetros D_c , D_t , Δ_t , N_t e η gera resultados de maior razão qualidade/complexidade computacional para os sinais do banco BD1. Observa-se que a combinação adequada desses parâmetros consegue caracterizar um codificador G.729A modificado com a mesma qualidade do G.729A original, porém com necessidade de $M = 189$ multiplicações. Esse ponto é indicado pelo símbolo D*, com PESQMOS = 3,74. Portanto, dispõe-se de uma redução de 94,86% da complexidade computacional sem redução significativa da nota PESQMOS.

Tabela 3.3: Detalhamento dos dados da curva C_o da Figura 3.6 [5].

Símbolo	D_c	D_t	Δ_t	N_t	N_t	M	PESQMOS
G.729A	2	1	0	1	-	3680	3,75
A	2	6	5	1	0,017	1140,0	3,76
B	3	2		1	0,017	874,8	3,75
C	5	2	0	1	0,011	456,0	3,75
D	6	4	0	1	0,017	189,0	3,74
E	10	5	0	1	0,022	81,6	3,66

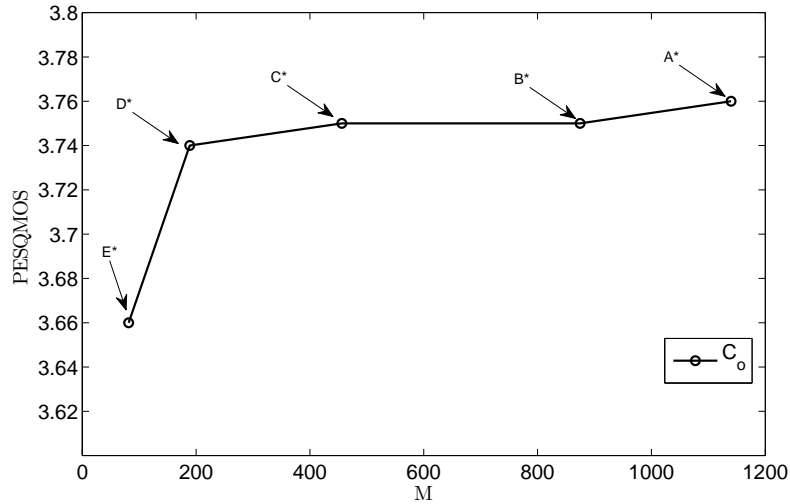


Figura 3.6: Quantidade de multiplicações necessárias na etapa de *open loop* para o cálculo da estimativa de T_{op} com a configuração que utiliza os parâmetros D_c , D_t , N_t , Δ_t .

3.4 Conclusão

Neste capítulo foi caracterizado o banco de sinais de voz BD1, que será utilizado ao longo deste trabalho. Também foi descrito o G.729A modificado, cujas idéias e configurações servirão de base para as modificações propostas no próximo capítulo. Para entendimento dessas modificações, é importante ter em mente os mecanismos de fixação da estimativa do *pitch* em um dado intervalo, da decimação do sinal no tempo, de decimação do atraso no *lag*, e de análise na vizinhança de T_{op} , assim como o objetivo da análise da estimativa de *pitch* distantes e a reutilização do atraso do dicionário adaptativo.

Capítulo 4

Simplificações propostas

4.1 Introdução

Neste capítulo serão apresentadas simplificações de trabalhos anteriores implementadas no G.729A e descritas novas modificações implementadas neste trabalho. Todas essas alterações têm o intuito de reduzir a complexidade computacional empregada nos cálculos da estimativa do atraso do *pitch* no codificador G.729A sem reduzir a qualidade do sinal reproduzido.

Com essa finalidade, modificações descritas em [4] serão reproduzidas e expandidas, fazendo com que a relação entre complexidade computacional e qualidade seja função dos parâmetros introduzidos por essas modificações, que atuam na etapa de *open loop*.

Encerradas as alterações implementadas no *open loop*, serão descritos e apresentados os resultados que as alterações da etapa de *closed loop* provocam na qualidade do sinal.

4.2 Modificações no *open loop*

4.2.1 Fixação de T_{op} para todos os intervalos.

O resultado da seção 3.3.1 do capítulo anterior sugere que a percepção sobre o sinal reproduzido, para o banco utilizado, é menos sensível a alterações no intervalo

(c) que nos demais intervalos. Porém, essa observação é resultante de experimentos que fixam somente a moda de cada intervalo.

A fim de tornar possível uma observação mais genérica, são fixados outros valores de T_a , T_b e T_c e não somente a moda destes. Assim, em um dado intervalo i , T_i é fixado em cada um dos seus possíveis valores e, portanto, o *open loop* realiza os cálculos necessários para a maximização da autocorrelação do sinal apenas em dois outros intervalos, pois no terceiro o valor de T_i já é previamente fixado.

Os resultados dessa configuração apresentam dois aspectos. O primeiro deles é de caráter qualitativo, pois diz respeito à degradação introduzida no sinal reproduzido; já o segundo aspecto tem caráter estatístico, pois diz respeito ao percentual de vezes em que o valor fixado de T_i é escolhido como melhor valor para T_{op} . As figuras que se seguem detalham esses dois aspectos decorrentes da fixação de T_i .

Na Figura 4.1, cada barra indica o valor médio da nota PESQMOS para os sinais do banco BD1 quando $T_i = \tau_i$ e a linha tracejada no valor PESQMOS 3,75 indica o valor PESQMOS atribuído ao G.729A original. Desse modo, pode ser feita a comparação da qualidade do sinal reproduzido em função do valor τ_i em que T_i foi fixado. Nesta figura, torna-se notório que a fixação de qualquer valor de T_{op} no intervalo (c) mantém a percepção sobre o sinal reproduzido praticamente inalterada. Já alterações no intervalo (a) e (b) penalizam a qualidade do sinal reproduzido em valores que variam de 5 a 10 centésimos na nota PESQMOS.

Por outro lado, na Figura 4.2, cada barra expressa em termos percentuais¹ a quantidade de sub-segmentos em que $T_{op} = \tau_i$ considerando o universo de todos os sub-segmentos do banco BD1. Portanto, esse percentual P indica o quanto foi alterada neste experimento a participação de cada valor de T_i na distribuição de T_{op} .

Conclui-se que as técnicas implementadas resultam na alteração da distribuição de T_{op} nos três intervalos. Porém, quando se considera a nota PESQMOS, o comportamento apresentado pelos três intervalos é diferente. Nesse caso, no intervalo

¹Aqui o somatório não equivale a 100%, pois este não representa um histograma com variável τ_i .

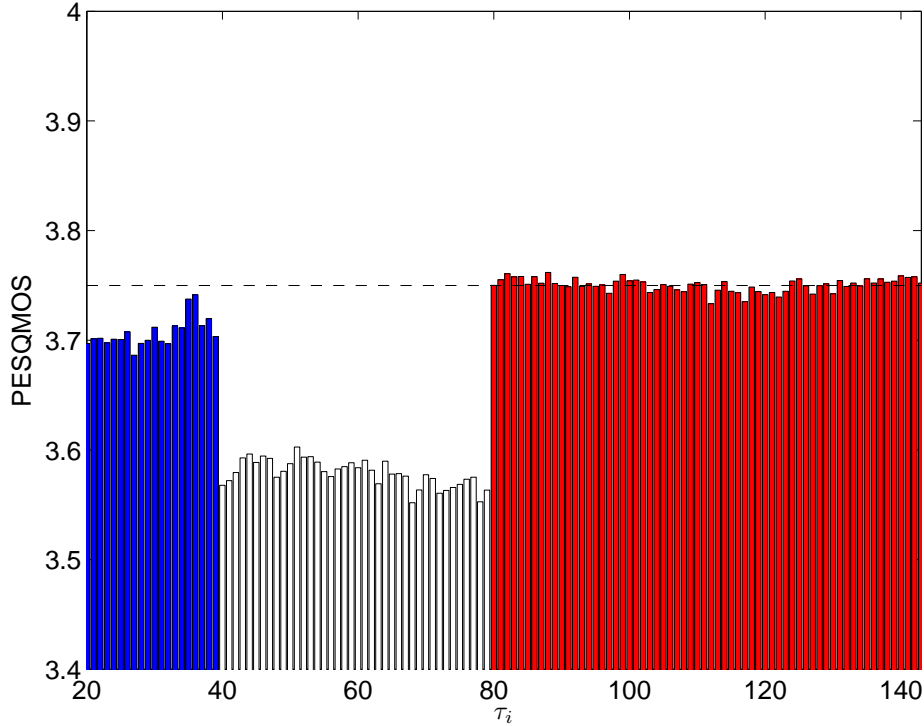


Figura 4.1: PESQMOS para τ_i ao longo dos intervalos $i = (a)$, (b) e (c) .

(c) é atingido o objetivo de redução da complexidade computacional sem reduzir a nota PESQMOS e esse efeito independe do valor de τ_i em que T_i é fixado. Já nos intervalos (a) e (b), a redução da complexidade computacional vem acompanhada da degradação do sinal reproduzido.

O objetivo deste trabalho é implementar simplificações capazes de reduzir a complexidade computacional sem introduzir degeneração no sinal. Portanto, se mostra necessário um estudo que indique qual fator deve ser privilegiado nas etapas posteriores. Pois, se por um lado a Figura 4.1 mostra a degeneração introduzida pela fixação de T_i , por outro, a Figura 4.2 mostra a quantidade de vezes que a decisão de se fixar T_i em τ_i desprezando os demais valores se mostra acertada, poupando esforço computacional do processo de codificação.

Sendo assim, os resultados observados nessa seção levam ao questionamento sobre que parâmetro deve ser utilizado como critério de escolha dos valores em que T_i deve ser fixado de modo que sejam minimizadas a complexidade computacional e a degeneração oriunda da codificação. Responder a essa questão é o objetivo das

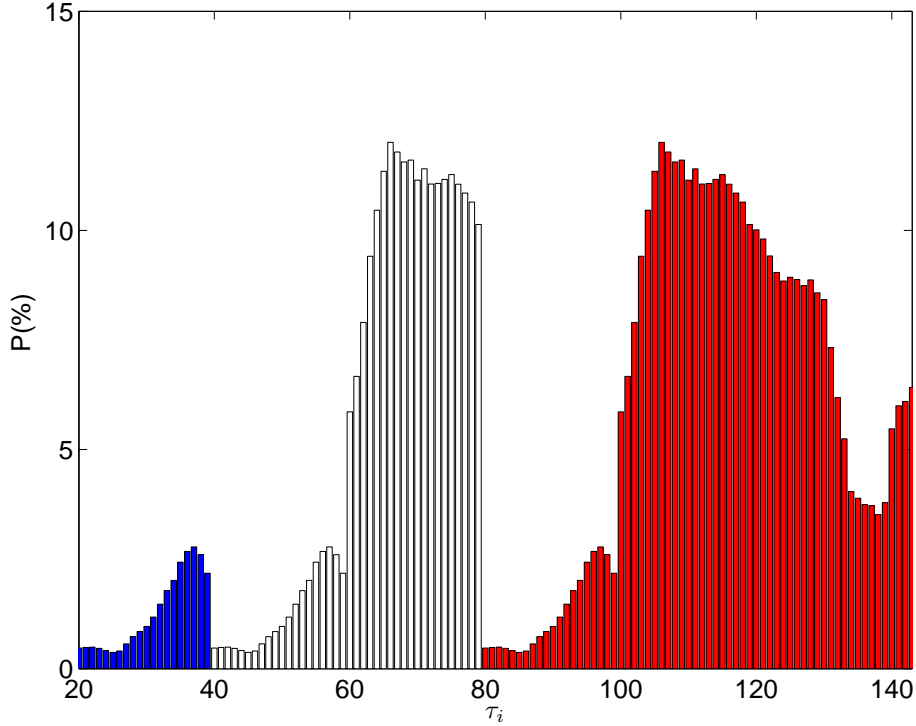


Figura 4.2: Percentual de ocorrência da igualdade $T_{op} = \tau_i$ nos intervalos (a), (b) e (c)

seções 4.2.3 e 4.2.2.

4.2.2 Análise da nota PESQMOS

Almeja-se entender que parâmetro relacionado a T_i tem mais influência sobre a qualidade do sinal reproduzido: as notas PESQMOS associadas a T_i quando este é fixado em τ_i ou a frequência de vezes que $T_{op} = \tau_i$. Para atingir esse objetivo, são empregadas duas técnicas que consistem em limitar em N_i o número de valores que T_i pode assumir em cada um dos intervalos i , distribuídos da seguinte forma:

- $1 \leq N_a \leq 20$, pois $20 \leq T_a \leq 39$;
- $1 \leq N_b \leq 40$, pois $40 \leq T_b \leq 79$;
- $1 \leq N_c \leq 64$, pois $80 \leq T_c \leq 143$.

Esta seção descreve a primeira técnica empregada para verificar que parâmetro T_i tem mais influência sobre a qualidade do sinal reproduzido. Essa técnica privilegia a integridade do sinal reproduzido, ou seja, emprega o critério de maior nota

PESQMOS quando se fixa $T_i = \tau_i$ como critério de escolha para os valores de T_i , que são considerados para análise do atraso de *pitch*.

Assim, tomando como exemplo o intervalo (a), ao invés de se buscar o valor de T_a dentro do intervalo $20 \leq T_a \leq 39$, como no G.729A original, a busca é feita somente no conjunto de N_a elementos. Este mecanismo reduz a quantidade de operações matemáticas necessárias para o cálculo de T_a , enquanto as buscas de T_b e T_c permanecem inalteradas. Desse modo, tem-se um conjunto de N_a elementos formado pelos valores de τ_a que maximizam a nota PESQMOS quando se fixa $T_i = \tau_a$. Processo semelhante é implementado para os intervalos (b) e (c).

A título de exemplificação, considere-se o detalhamento dos dados apresentados nas Figuras 4.1 e 4.2 para alguns pontos selecionados. Para esses pontos, os valores de τ_i e os valores correspondentes do percentual P e da nota PESQMOS apresentados na Tabela 4.1.

Tabela 4.1: Detalhamento de informações em para alguns pontos selecionados das Figuras 4.1 e 4.2.

τ_a	P(%)	PESQMOS
24	0,41	3,70
27	0,57	3,69
36	2,67	3,74
38	2,60	3,72

Utilizando a nota PESQMOS como critério de ordenação e $N_a = 4$, teria-se como resultado a seguinte sequência de valores considerados para T_a : {36, 38, 24, 27}.

A Figura 4.3 mostra a qualidade do sinal reproduzido em função do número N_i de valores possíveis de T_i no intervalo i , considerando T_i ordenado por notas PESQMOS quando $T_i = \tau_i$. É possível notar que com o aumento do parâmetro N_i , a nota PESQMOS aumenta quando $i = (a)$ e quando $i = (b)$. Porém, no intervalo (c), o aumento das possibilidades de valores de T_c não altera significativamente a nota PESQMOS, que vale 3,77. E mesmo essa pequena alteração na nota PESQMOS não

representa melhora na qualidade do sinal, já que a precisão do PESQ não permite diferenciar 3,75 de 3,77.

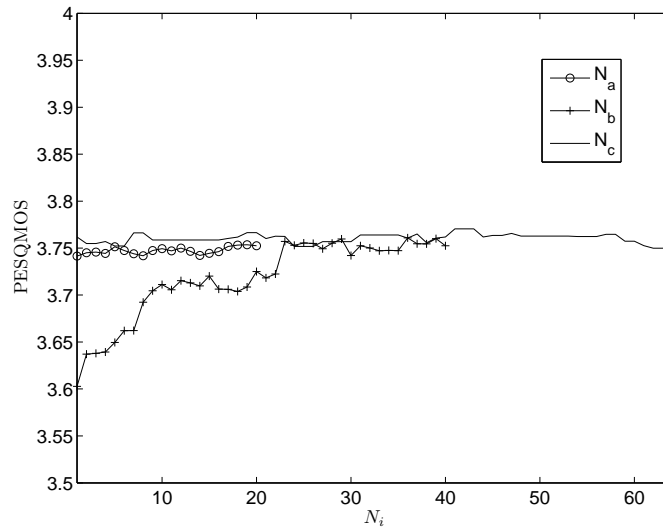


Figura 4.3: Relação entre a nota PESQMOS e o parâmetro N_i , considerando T_i ordenado por nota PESQMOS.

4.2.3 Análise da frequência de ocorrência

Nesta seção é descrita a segunda técnica, que privilegia a redução da complexidade computacional. Ou seja, aqui emprega-se a frequência de ocorrência da igualdade $T_{op} = T_i$ quando se fixa $T_i = \tau_i$, que é a frequência de relativa de $T_i = \tau_i$, como critério de escolha para os valores de T_i que são considerados para análise do atraso de *pitch*.

Assim, tomando como exemplo o intervalo (a), ao invés de se buscar o valor de T_a dentro do intervalo $20 \leq T_a \leq 39$, como no G.729A original, a busca é feita somente no conjunto de N_a elementos. Este mecanismo também reduz a quantidade de operações matemáticas necessárias para o cálculo de T_a , enquanto as buscas de T_b e T_c permanecem inalteradas. Desse modo, tem-se um conjunto de N_a elementos formado pelos valores de τ_a que maximizam a frequência relativa de $T_a = \tau_a$. Processo semelhante é implementado para os intervalos (b) e (c).

Considerando-se novamente o conjunto de dados da Tabela 4.1, porém, desta vez, utilizando a análise da frequência relativa e $N_a = 4$, teria-se como resultado os seguintes valores possíveis para T_a : $\{36, 38, 27, 24\}$.

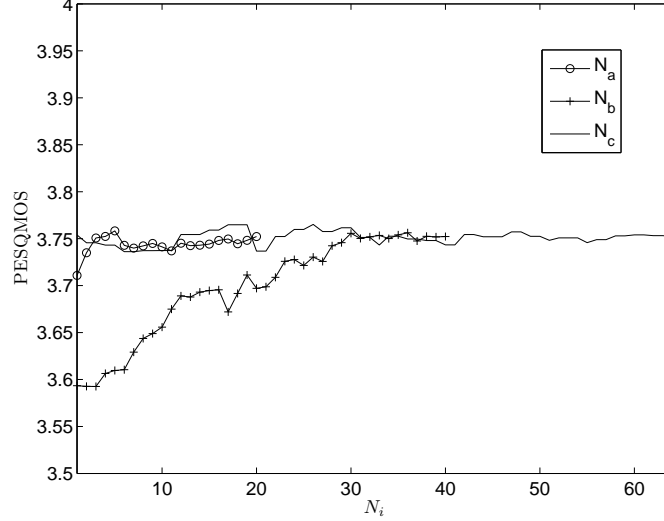


Figura 4.4: Relação entre a nota PESQMOS e o parâmetro N_i , considerando T_i ordenado por frequência de ocorrência da igualdade $T_i=T_{op}$.

A Figura 4.4 mostra a qualidade do sinal reproduzido em função do parâmetro N_i de valores possíveis de T_i no intervalo i , considerando T_i ordenado por frequência de ocorrência da igualdade $T_i=T_{op}$. Nota-se que, com o aumento do valor do parâmetro N_i , a nota PESQMOS aumenta quando $i = (a)$ e quando $i = (b)$. Porém, no intervalo (c) , o aumento das possibilidades de valores de T_c não altera significativamente a nota PESQMOS, que vale 3,75 quando $N_c = 1$.

A comparação dos resultados desta técnica com a técnica descrita na seção anterior mostra que a nota PESQMOS e a maximização da função autocorrelação $R'(k)$ definida na (2.10) são correlacionadas. Comparando-se as Figuras 4.3 e 4.4, observa-se que a influência sobre a qualidade do sinal reproduzido é a mesma tanto quando N_i equivale ao número de valores de T_i em que $T_i=\tau_i$ quanto quando N_i equivale ao número de valores de T_i que garante maior nota PESQMOS. Portanto, a partir deste ponto, escolhe-se que serão utilizados os valores de T_i ordenados por notas PESQMOS, ou seja, N_i é a quantidade de valores T_i assumidos como aqueles que

garantem maior nota PESQMOS quando $T_i = \tau_i$. Além disto, a partir deste ponto será assumida a igualdade $N_c = 1$, que corresponde ao valor $T_c = 88$.

4.2.4 Limitação da variação do intervalo de estimativa do *pitch*

Diante dos resultados observados nas seções 4.2.3 e 4.2.2, a análise do impacto causado pela combinação das alterações nos intervalos (*a*) e (*b*) conjuntamente se apresenta como extensão natural. Para este fim, é utilizada uma configuração em que são implementadas as mudanças de limitação do tamanho do intervalo T_a e T_b simultaneamente através da fixação dos valores de N_a e N_b . Assim, para cada um dos valores de N_a , N_b é variado dentro do intervalo $1 \leq N_b \leq 40$. Com isso, o número de multiplicações M necessárias em um segmento na etapa de *open loop* é dado por:

$$M = (N_a + N_b + 1) \times \left\lfloor \frac{79}{2} + 1 \right\rfloor. \quad (4.1)$$

A avaliação PESQMOS em função do número de multiplicações M empregadas no *open loop* para esta configuração é apresentado na Figura 4.5. O fecho destacado é formado pelos pontos com maior relação qualidade/complexidade computacional. Além disto, os valores dos parâmetros N_a e N_b , do número de multiplicações M e da nota PESQMOS associados a configuração representada por cada ponto deste eixo são apresentados na Tabela 4.2.

Através dessa tabela, é possível notar que há queda na qualidade do sinal reconstituído à medida que a quantidade de multiplicações empregadas na etapa *open loop* do processo de codificação diminui. Contudo, obtém-se a nota 3,75 para a configuração C da Tabela 4.2 com apenas 41,30% do número de multiplicações do G.729A original.

4.2.5 Reprodução da decimação do sinal alvo no cálculo da autocorrelação

Com o objetivo de reduzir a complexidade computacional de seu algoritmos, o G.729A original usa um fator de decimação $D_c = 2$ no cálculo da autocorrelação do

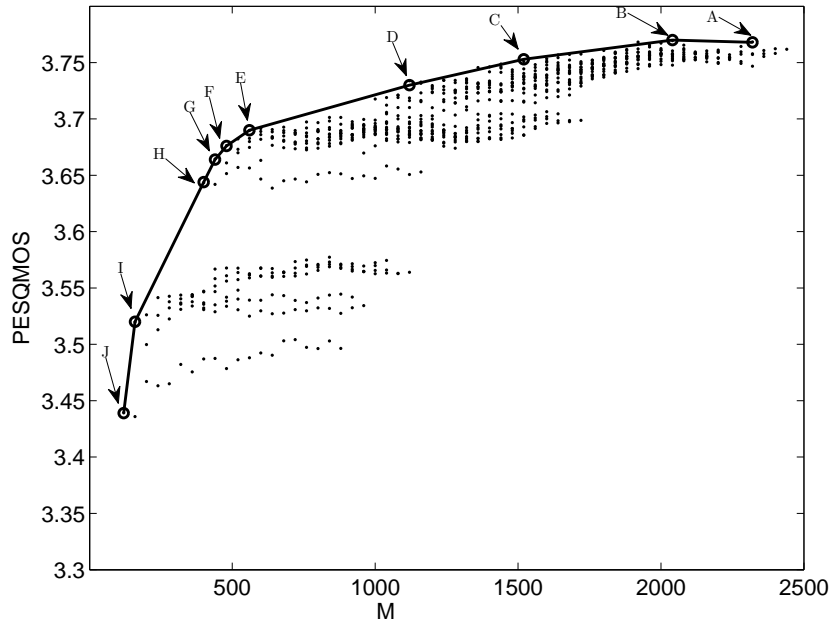


Figura 4.5: Quantidade de multiplicações necessárias para o cálculo de T_{op} na configuração que utiliza os parâmetros N_a , N_b .

signal. Esse fator serve de motivação para que outros valores de fator de decimação sejam experimentados, o que é objeto de estudo nesta seção. Deste modo, além dos parâmetros N_a e N_b , é introduzido o parâmetro D_c (ver seção 3.3.2), fazendo com que as configurações sejam combinações desses três parâmetros, que variam segundo a forma:

- $1 \leq N_a \leq 20$;
- $1 \leq N_b \leq 40$;
- $2 \leq D_c \leq 10$.

Com isso, o número de multiplicações M necessárias em um segmento na etapa de *open loop* é dado por:

$$M = (N_a + N_b + 1) \times \left\lceil \frac{79}{D_c} + 1 \right\rceil. \quad (4.2)$$

A Figura 4.6 apresenta os dois fechos: o primeiro, representado pela curva C_1 , mostra os dados da Tabela 4.2 e o segundo, representado pela curva C_2 , mostra os

Tabela 4.2: Identificação dos pontos do fecho da Figura 4.5.

Símbolo	N_a	N_b	M	PESQMOS
G.729A	20	40	3680	3,75
A	20	37	2320	3,77
B	20	30	2040	3,77
C	20	17	1520	3,75
D	20	7	1120	3,73
E	12	1	560	3,69
F	10	1	480	3,68
G	9	1	440	3,66
H	8	1	400	3,64
I	2	1	160	3,52
J	1	1	120	3,44

resultados da Tabela 4.3, com a introdução do parâmetro D_c . Como pode ser observado, a introdução do parâmetro D_c gera deslocamento da curva para a esquerda, o que demonstra a redução mais efetiva da complexidade alcançada. Comparando-se os pontos C e D das curvas C_1 e C_2 , respectivamente, verifica-se que a variação do valor da decimação do sinal possibilita para a configuração D da Tabela 4.3 a redução de até 87% do total de multiplicações empregadas na etapa de *open loop* para a configuração correspondente à nota PESQMOS 3,75. Porém, de maneira geral, à medida que o fator de decimação da autocorrelação do sinal aumenta, a qualidade desse sinal diminui.

4.2.6 Reprodução da busca proximal de T_{op}

Além da degradação gerada pela decimação do sinal no cálculo da autocorrelação, a introdução dos parâmetros N_a , N_b e N_c faz com que alguns valores de T_a e T_b sejam desprezados. Com o objetivo de compensar essas perdas, é introduzida uma busca proximal em torno do valores T_i (ver seção 3.3.4).

Para implementação da busca proximal é introduzido o parâmetro Δ_t , com $1 \leq \Delta_t \leq 5$. Com a variação do valor desse parâmetro é possível fazer a busca se estender

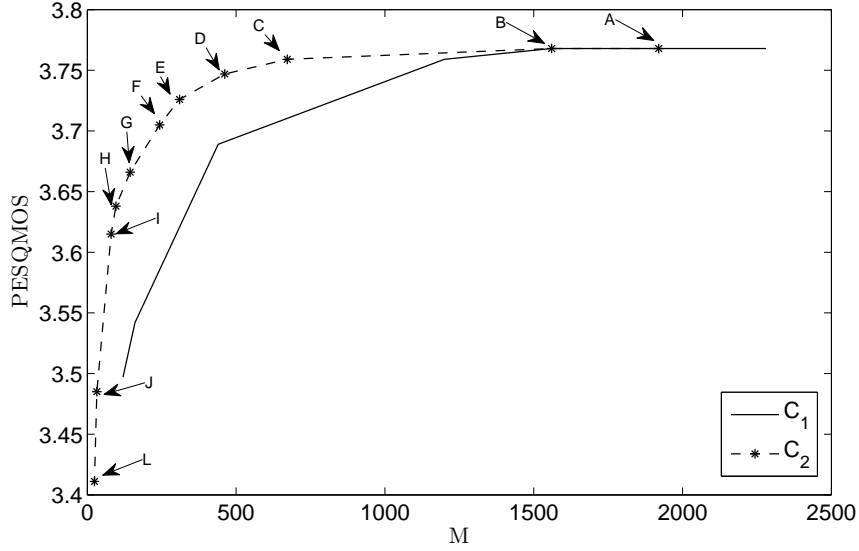


Figura 4.6: Quantidade de multiplicações necessárias para o cálculo de T_{op} na configuração que utiliza os parâmetros N_a , N_b e D_c .

a um número maior ou menor de elementos da vizinhança de T_i , de tal modo que, depois de escolhido T_i , a busca proximal é feita em $T_i - \Delta_t < T_i < T_i + \Delta_t$.

Com essa configuração o número de multiplicações necessárias na etapa de *open loop* é dado por:

$$M = (N_a + N_b + 1 + 2\Delta_t) \times \left\lceil \frac{79}{D_c} + 1 \right\rceil \quad (4.3)$$

O efeito da introdução da busca proximal é apresentado na Figura 4.7, que contém os fechos C_1 , C_2 e C_3 . As curvas C_1 e C_2 apresentam os resultados das Tabelas 4.2 os parâmetros N_a e N_b e 4.3 com o parâmetro D_c , como na Figura 4.6. Já a curva C_3 mostra os resultados apresentados na Tabela 4.4, onde é adicionado o parâmetro Δ_t .

Novamente a nota PESQMOS de valor 3,75 associada ao G.729A original é obtida com complexidade computacional inferior àquela apresentada nas curvas C_1 e C_2 . Conclui-se que a busca proximal atende à expectativa de funcionar como ajuste fino na estimativa do valor de T_{op} , permitindo aumento da qualidade do sinal reproduzido, mesmo para valores elevados do parâmetro D_c .

Tabela 4.3: Dados da curva C_2 da Figura 4.6.

Símbolo	N_a	N_b	D_c	M	PESQMOS
G.729A	20	40	2	3680	3,75
A	10	37	2	1920	3,77
B	1	37	2	1560	3,77
C	9	38	6	672	3,76
D	5	27	6	462	3,75
E	3	27	8	310	3,73
F	3	23	9	243	3,70
G	1	10	7	144	3,67
H	1	10	10	96	3,64
I	1	8	10	80	3,62
J	1	2	10	32	3,48
L	1	1	10	24	3,41

4.3 Resumo das simplificações no *open loop*

Empregando-se as quatro técnicas descritas nas seções anteriores, a redução da complexidade computacional empregada é função de N_a , N_b , D_c e Δt . Os parâmetros N_a e N_b realizam a limitação da gama de valores que T_a e T_b podem assumir e o parâmetro D_c realiza a decimação do sinal no cálculo da autocorrelação. Esses parâmetros aliados à fixação de T_c reduzem a complexidade computacional empregada na etapa de *open loop* do processo de codificação, porém também reduzem a qualidade do sinal reproduzido. Para minimizar esse efeito, é utilizado o parâmetro Δt , que reduz a degradação do sinal decodificado com custo de um pequeno aumento na complexidade computacional.

A tendência de deslocamento dos fechos que contêm os pontos de melhor relação qualidade/complexidade computacional para esquerda nos gráficos de PESQMOS em função de M à medida que estes parâmetros são introduzidos indica a redução da complexidade alcançada. Para o ponto cuja nota PESQMOS corresponde à nota do G.729A original, existe um conjunto de valores de parâmetros para a configuração

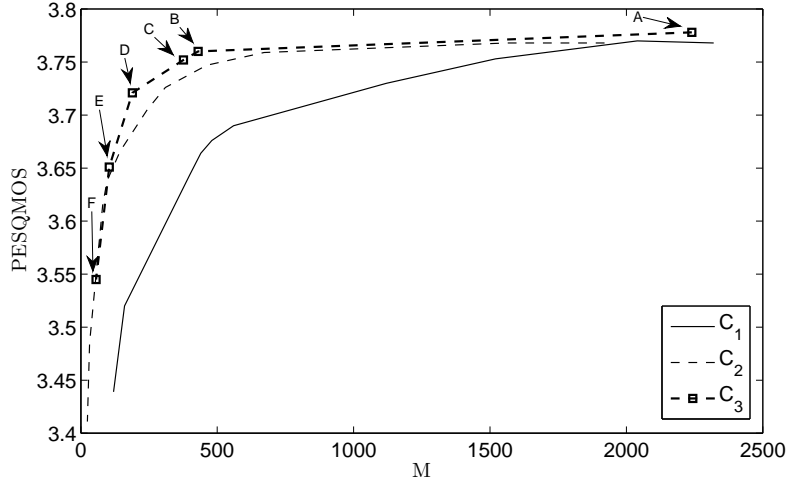


Figura 4.7: Quantidade de multiplicações necessárias para o cálculo de T_{op} na configuração que utiliza os parâmetros N_a , N_b , D_c e Δ_t .

C da Tabela 4.4 que produz redução de até 89,78% da complexidade computacional envolvida na etapa de *open loop*, quando esta é comparada com a do próprio G.729A original.

4.4 Modificações no *closed loop*

4.4.1 Busca proximal da estimativa do *lag*

Depois de explorada a redução da complexidade do *open loop*, surge como extensão dessa tendência a tentativa de redução da complexidade computacional na busca proximal feita no *closed loop*.

Conforme descrito na seção 2.4.6, imediatamente após ser encerrada a etapa de *open loop*, o G.729A original emprega uma etapa de *closed loop*, onde é feita a busca proximal em torno de T_{op} para o primeiro sub-segmento, com $(T_{op} - 3) \leq T_1 \leq (T_{op} + 3)$, enquanto a busca de T_2 é feita em $(T_{op} - 5) \leq T_2 \leq (T_{op} + 4)$, o que consome 2320 multiplicações. Porém, nesse trabalho já foi introduzido o parâmetro Δ_t , que também desempenha a tarefa de realizar busca proximal, com a diferença de que isso é feito na etapa de *open loop*. Assim, há necessidade de se analisar o impacto que o intervalo de busca do *closed loop* tem na configuração atual do G.729A, que

Tabela 4.4: Dados da curva C_3 da Figura 4.7.

Símbolo	N_a	N_b	D_c	Δ_t	M	PESQMOS
G.729A	20	40	2	0	3680	3,75
A	11	36	2	4	2240	3,78
B	11	27	8	2	430	3,76
C	1	35	10	5	376	3,75
D	1	9	9	5	189	3,72
E	1	1	10	5	104	3,65
F	1	1	10	2	56	3,54

conta com os parâmetros N_a , N_b , D_c e Δ_t . Esse é o objetivo desta seção.

Para realizar a análise da influência do intervalo do *closed loop*, introduz-se o parâmetro r_{cl} , que define os limites do intervalo do primeiro sub-segmento com $1 \leq r_{cl} \leq 5$. Assim, durante o *closed loop* o intervalo de busca de T_1 é $(T_{op} - r_{cl}) \leq T_1 \leq (T_{op} + r_{cl})$. Assim, as configurações passam a ser combinações de cinco parâmetros, que variam segundo a forma:

- $1 \leq N_a \leq 20$;
- $1 \leq N_b \leq 40$;
- $2 \leq D_c \leq 10$;
- $1 \leq \Delta_t \leq 5$;
- $1 \leq r_{cl} \leq 5$.

A Figura 4.8 apresenta os fechos C_1 , C_2 , C_3 e C_4 , onde C_4 apresenta os pontos de melhor compromisso qualidade/complexidade computacional quando são empregados os parâmetros N_a , N_b , D_c , Δ_t e r_{cl} e cujos valores são detalhados na Tabela 4.5. As curvas C_1 , C_2 e C_3 são as mesmas anteriormente apresentadas na Figura 4.7 somadas com 2320, que é o número do *closed loop* do G.729A original e estão aqui para efeito de comparação. Porém, aqui está representado também o número de multiplicações empregadas no *closed loop* e não somente as do *open loop* como

contabilizado na seção 4.2. Esse número de multiplicações de cada segmento passa a ser dado por:

$$M = \left\lceil \frac{79}{D_c} + 1 \right\rceil (N_a + N_b + 1 + 2\Delta_t) + 40 \times (2 r_{cl} + 1 + 10) + 1640, \quad (4.4)$$

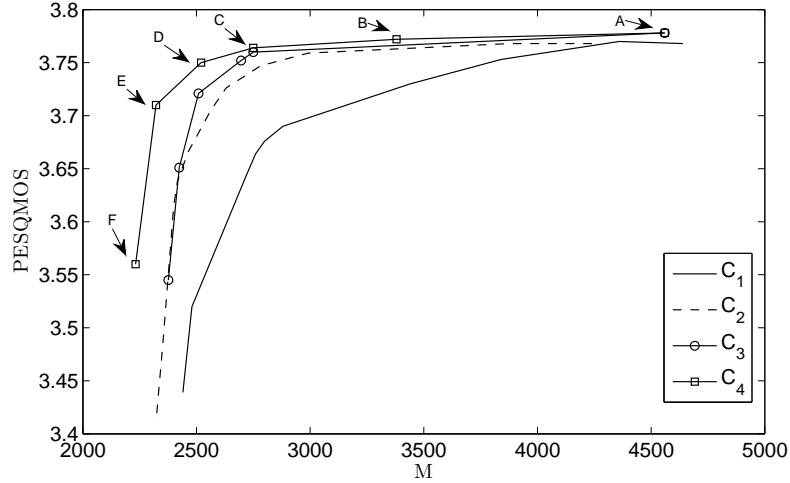


Figura 4.8: Quantidade de multiplicações necessárias para o cálculo da estimativa do *lag* nas etapas de *open loop* e *closed loop* com a configuração que utiliza os parâmetros N_a , N_b , D_c , Δ_t e r_{cl} .

Tabela 4.5: Dados da curva C_4 da Figura 4.8.

Símbolo	N_a	N_b	D_c	Δ_t	r_{cl}	M	PESQMOS
G.729A	20	40	2	0	0	6000	3,75
A	11	36	2	4	3	4400	3,78
B	9	29	4	3	5	3140	3,77
C	11	27	8	2	3	2590	3,76
D	11	27	10	3	1	2440	3,75
E	2	9	9	3	1	2236	3,71
F	1	1	10	3	1	2152	3,56

É notória a redução do número de multiplicações necessárias ao cálculo de T . Por exemplo, para a nota PESQMOS 3,75 para a configuração D da Tabela 4.5 há redução de 59% em relação às etapas de *open loop* e *closed loop* do G.729A original.

4.4.2 Cálculo do resíduo do filtro LP

Conforme descrito na seção anterior, a determinação da vizinhança de T_{op} onde é feita a análise do atraso do período de *pitch* foi alterado em relação ao G.729A original, passando a ser feita através do parâmetro r_{cl} . Porém o método empregado nessa análise permanece inalterado, despertando o interesse da avaliação da possibilidade de redução da complexidade computacional empregada nesse processo. Esse é o objetivo das alterações apresentadas nesta seção.

Conforme descrito na seção 2.4.6, no G.729A original, a escolha do atraso do período de *closed loop* T se dá através da comparação dos sinais de excitações do dicionário adaptativo com o sinal $r(n)$, que é resultado da filtragem do sinal alvo desse dicionário $x(n)$ pelo inverso do filtro de síntese ponderado. Essa filtragem é dada pela equação:

$$r(n) = \sum_{j=n}^{39} x(j)h_{inv}(j-n), \quad 0 \leq n \leq 39, \quad (4.5)$$

onde $h_{inv}(n)$ é o inverso da resposta ao filtro de síntese ponderado. A maximização da correlação cruzada do sinal $r(n)$ com as excitações do dicionário adaptativo determina o valor do atraso do *pitch* do *closed loop*.

Através da equação (4.5) é possível verificar que esta implementação para o cálculo de $r(n)$ tem um custo de 1640 multiplicações por segmento. Porém, esta não é a única abordagem possível para o cálculo de $r(n)$. Como abordagem alternativa, propõe-se a utilização da função de transferência do filtro de síntese ao invés de sua resposta ao impulso, já que esse método pode reduzir a complexidade computacional desta etapa de codificação. Visando este objetivo, o cálculo de $r(n)$ passa a ser implementado da seguinte maneira:

$$r(n) = \sum_{i=0}^{10} a_i x(n-i), \quad 0 \leq n \leq 39, \quad (4.6)$$

onde a_i são os coeficientes do filtro de síntese.

Com essa alteração, o número de multiplicações empregadas no cálculo do sinal $r(n)$ é reduzido para 770 por segmento, enquanto o número de multiplicações

envolvidas nas etapas *open loop* e *closed loop* passa a ser dado por:

$$M = \left\lceil \frac{79}{D_c} + 1 \right\rceil (N_a + N_b + 2\Delta_t) + 40 \times (2 r_{cl} + 11) + 770, \quad (4.7)$$

Essa nova abordagem para o cálculo do sinal $r(n)$ aplicada ao banco de sinais de voz BD1 em conjunto com os parâmetros N_a , N_b , D_c , Δ_t e r_{cl} conforme haviam sido implementados anteriormente reduz a quantidade de multiplicações empregadas no *closed loop*. Essa redução é notória quando se analisa a Figura 4.9, onde a curva C_3 apresenta os pontos de melhor compromisso qualidade/complexidade computacional quando são considerados apenas os parâmetros atuantes na etapa de *open loop*, a curva C_4 apresenta os pontos de melhor compromisso qualidade/complexidade computacional considerando-se o parâmetro r_{cl} e a curva C_5 , onde além desses parâmetros é implementada a nova abordagem para o cálculo de $r(n)$. O deslocamento entre elas é quantificado na Tabela 4.6, que identifica do valores dos parâmetros dos pontos dessa curva.

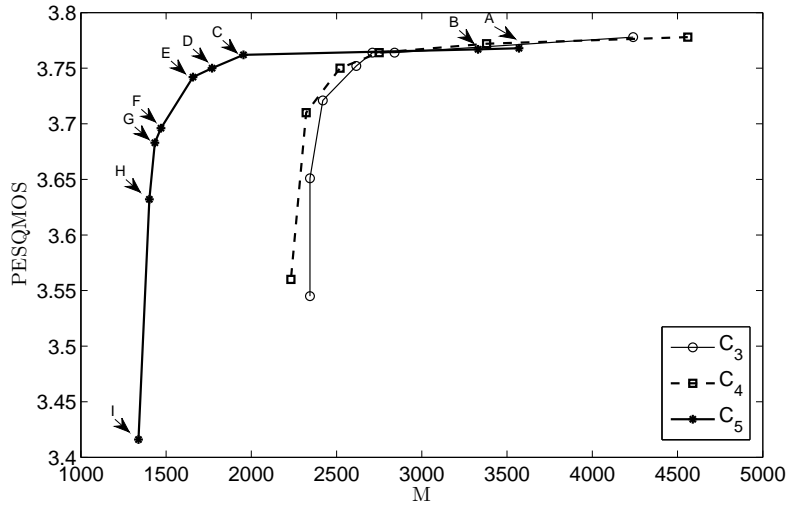


Figura 4.9: Quantidade de multiplicações necessárias para o cálculo da estimativa do *lag* nas etapas de *open loop* e *closed loop* com a configuração que utiliza os parâmetros N_a , N_b , D_c , Δ_t e r_{cl} e o filtro de síntese para o cálculo do resíduo.

Observa-se que a nota PESQ MOS 3,75, característica da G.729A para o banco BD1 é obtido com o emprego de $M = 1770$ multiplicações, o que significa que há redução de 70, 50% da complexidade computacional utilizada nas etapas de *open loop* e *closed loop* em relação ao G.729A original.

Tabela 4.6: Dados da curva C_5 da Figura 4.9.

Símbolo	N_a	N_b	D_c	Δ_t	r_{cl}	M	PESQMOS
G.729A	20	40	2	0	0	6000	3,75
A	9	35	2	4	3	3570	3,77
B	8	36	2	1	3	3330	3,77
C	1	28	6	3	3	1954	3,76
D	6	27	7	3	1	1770	3,75
E	3	28	10	2	2	1658	3,74
F	2	9	9	4	1	1470	3,70
G	2	9	10	3	1	1434	3,68
H	1	8	10	2	1	1402	3,63
I	1	2	10	1	1	1338	3,41

4.4.3 Decimação do sinal no tempo no *closed loop*

A decimação do sinal alvo na cálculo da autocorrelação no *open loop*, descrita na seção 4.2.5 leva à redução do número de multiplicações nessa etapa. Tendo essa redução da complexidade computacional como motivação, esta seção aborda a extensão do conceito de decimação do sinal na etapa de *closed loop*.

Para esse fim, é utilizado o fator de decimação do sinal D_k , o que faz com o cálculo de $R_{xy}(k)$ passe a ser dado por:

$$R_{xy}(k) = \sum_{n=0}^{\lfloor \frac{39}{D_k} \rfloor} x(nD_k) y_k(nD_k), \quad (4.8)$$

e que a quantidade de multiplicações envolvidas nas etapas de *open loop* e *closed loop* seja dada por:

$$M = \left\lfloor \frac{79}{D_c} + 1 \right\rfloor (N_a + N_b + 1 + 2\Delta_t) + \left\lfloor \frac{39}{D_k} + 1 \right\rfloor (2 r_{cl} + 11) + 770; \quad (4.9)$$

A nova técnica é aplicada aos pontos da curva C_4 da Figura 4.8, portanto os parâmetros N_a , N_b , D_c , Δ_t e r_{cl} têm seus valores determinados por esses pontos enquanto D_k varia da seguinte forma: $1 \leq D_k \leq 5$.

O efeito da introdução da decimação do sinal no tempo é apresentado na Figura 4.10, que contém os fechos C_3 , C_4 , C_5 e C_6 . A curva C_3 corresponde aos pontos gerados pelo G.729A com modificações apenas do *open loop*. Enquanto as curvas C_4 e C_5 correspondem às versões do G.729A implementadas nas seções 4.4.1 e 4.4.2. Já a curva C_6 , mostra os resultados da implementação descrita nessa seção e tem seus dados detalhados na Tabela 4.7.

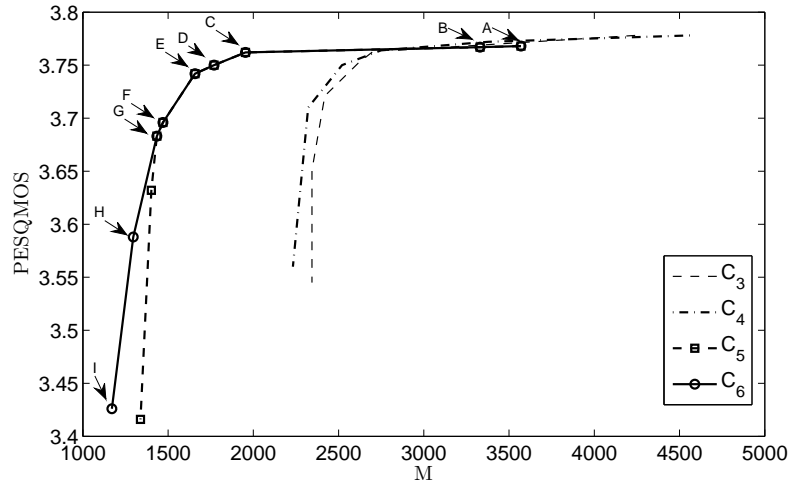


Figura 4.10: Quantidade de multiplicações necessárias para o cálculo da estimativa do *lag* nas etapas de *open loop* e *closed loop* com a configuração que utiliza os parâmetros N_a , N_b , D_c , Δ_t , r_{cl} e D_k e o filtro de síntese para o cálculo do resíduo.

Há redução da complexidade computacional em relação a implementação da seção anterior. Porém, essa redução de complexidade é acompanhada pela degradação do sinal reproduzido. Para a nota PESQ MOS 3,75 associada ao G.729A padrão, o número de multiplicações se mantém como na seção anterior.

4.5 Resumo das modificações do *closed loop*

As técnicas apresentadas nas seções anteriores geram redução de complexidade computacional tanto pelos parâmetros r_{cl} e D_k quanto pela alteração do método do cálculo do sinal r . O parâmetro r_{cl} permite variação na vizinhança de T_{op} em que o *lag* do primeiro sub-segmento seja procurado. A alteração do método de cálculo do sinal r reduz a complexidade computacional sem necessidade de decimação, o que é

Tabela 4.7: Dados da curva C_6 da Figura 4.10.

Símbolo	N_a	N_b	D_c	Δ_t	r_{cl}	D_k	M	PESQMOS
G.729A	20	40	2	0	0	1	6000	3,75
A	9	35	2	4	3	1	3570	3,77
B	8	36	2	1	3	1	3330	3,77
C	1	28	6	3	3	1	1954	3,76
D	6	27	7	3	1	1	1770	3,75
E	3	28	10	2	2	1	1658	3,74
F	2	9	9	4	1	1	1470	3,70
G	2	9	10	3	1	1	1434	3,68
H	1	31	10	5	1	3	1296	3,59
I	3	27	10	2	2	5	1170	3,43

feito com o parâmetro D_k . Porém a decimação introduzida por esse parâmetro não atende os objetivos desse trabalho. Pois de maneira diferente das demais simplificações implementadas anteriormente, o parâmetro D_k reduz a complexidade computacional para configurações cuja nota PESQMOS é inferior a do G.729A padrão.

O comportamento de redução de complexidade resultante dessas modificações é comprovado pela tendência de deslocamento para a esquerda dos fechos que contêm os pontos de melhor relação qualidade/complexidade computacional dos gráficos de PESQMOS em função de M à medida que essas modificações são implementadas. Para o ponto D da Tabela 4.7 cuja nota corresponde a nota do G.729A original, existe um conjunto de valores de parâmetros que produz redução de 70,50% da complexidade computacional utilizada nas etapas de *open loop* e *closed loop* quando esta é comparada com o próprio G.729A original.

A Figura 4.11 mostra a curva C_6 e a curva C_o dada pelos pontos da Tabela 3.3 apresentada na seção 3.3.5 somada com 2320, que é o número de multiplicações empregadas na etapa de *closed loop*. A comparação entre as duas curvas permite observar que, para os pontos de melhor compromisso qualidade/custo computacional (E de C_6 e D* de C_o), há redução de 33,91% da complexidade computacional da

versão do G.729A obtida nesse trabalho em relação à melhor curva obtida em [5].

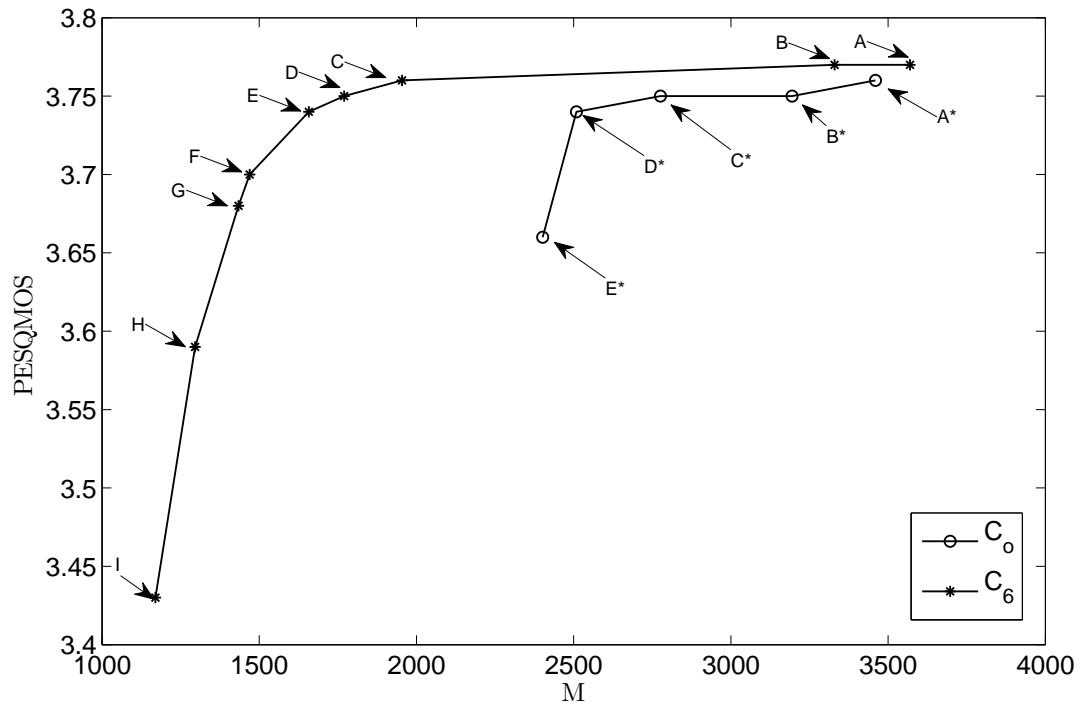


Figura 4.11: Quantidade de multiplicações necessárias para o cálculo da estimativa do *lag* nas etapas de *open loop* e *closed loop* com a configuração que utiliza os parâmetros N_a , N_b , D_c , Δ_t e r_d e o filtro de síntese para o cálculo do resíduo comparada com a curva C_o da Figura 3.6.

A redução da complexidade computacional obtida nesse trabalho se confirma na Tabela 4.8, apresentada a seguir. Nela, além dos parâmetros, do número de multiplicações e da nota PESQMOS associada aos pontos D e E da Tabela 4.7 e do G.729A original, é apresentada a informação do tempo T gasto no processo de codificação. Esses dados revelam redução de 17% no tempo total.

A qualidade dos sinais de fala resultantes da configuração do ponto E da Tabela 4.7 e o G.729A original foi avaliada através de testes comparativos. Nestes testes foram utilizados 32 sinais diferentes de fala na língua portuguesa do Brasil. Os 33 avaliadores participantes dos testes, ouviram os sinais de fala resultantes da codificação/decodificação feita pelo G.729A original (sinal A) e posteriormente pela versão modificada (sinal B). Os ouvintes foram instruídos a dar uma nota para cada

Tabela 4.8: Introdução do tempo codificação para as versão original do G.729A e das versões caracterizadas pelos parâmetros dos pontos D e E da curva C_6 da Figura 4.10.

Símbolo	N_a	N_b	D_c	Δ_t	r_{cl}	D_k	M	PESQMOS	T(μs)
G.729A	20	40	2	0	0	1	6000	3,75	369,18
D	6	27	7	3	1	1	1770	3,75	306,59
E	3	28	10	2	2	1	1658	3,74	304,86

sinal, em resposta à pergunta “Como você avalia a degradação presente no sinal B comparado ao sinal A?”. Notas negativas expressam percepção de que o sinal B é mais degradado que o sinal A, notas positivas expressam o oposto e a nota zero indica que não foi percebida degradação entre os sinais.

A Figura 4.12 mostra o histograma do resultado do teste comparativo e indica, para fins práticos, equivalência entre a versão modificada e original, em relação à qualidade percebida do sinal de fala reconstruído. Portanto, esses resultados indicam que o objetivo desse trabalho foi alcançado com sucesso.

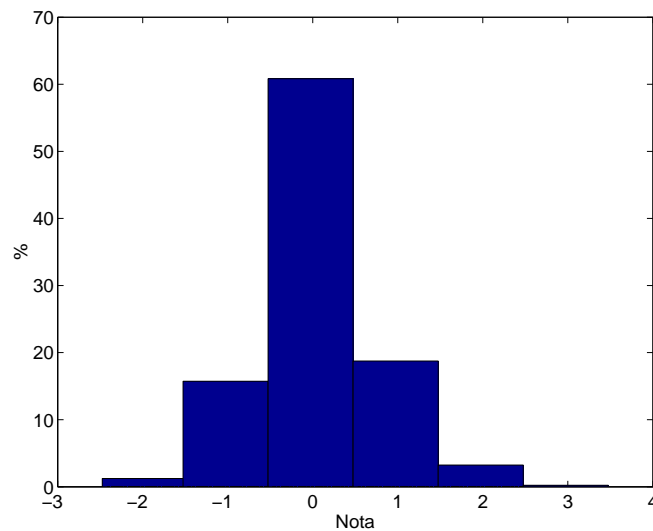


Figura 4.12: Distribuição de notas de teste comparativo subjetivo.

4.6 Conclusão

Neste capítulo foram implementadas modificações nas etapas de *open loop* e *closed loop* do G.729A propostas em trabalhos anteriores, além de propostas e implementadas novas modificações nessas duas etapas de busca do dicionário adaptativo. Todas essas modificações têm o intuito de reduzir a complexidade computacional empregada no cálculo da estimativa do *lag* sem reduzir a qualidade do sinal reproduzido.

Na etapa de *open loop*, a limitação do conjunto de valores para fixação de T_{op} , a decimação do sinal no cálculo da autocorrelação e a busca proximal de T_{op} proporcionaram uma redução da complexidade computacional de até 89,78% sem alteração da qualidade do sinal reproduzido.

Para a nota PESQMOS 3,75, quando considerada a flexibilização da busca proximal no *closed loop*, introduzida através do parâmetro r_{cl} e a alteração do cálculo do sinal $r(n)$, houve redução de 70,50% do número de multiplicações empregadas nas etapas de *open loop* e *closed loop* em relação ao G.729A original e de 33,91% em relação à versão do G.729A apresentada em [5]. Além disso, há redução de 17% no tempo total empregado na codificação do G.729A original.

Portanto, embora a decimação do sinal no cálculo da correlação cruzada no *closed loop* tenha provocado redução da complexidade computacional fora da faixa de interesse, as demais técnicas descritas ao longo deste capítulo possibilitaram o alcance do objetivo de redução de complexidade computacional do processo de codificação enquanto testes subjetivos informais confirmam que não houve comprometimento da qualidade do sinal reproduzido, que é o objetivo deste trabalho.

Capítulo 5

Conclusão

5.1 Contribuições do trabalho

No capítulo 2 é descrita a codificação empregada no G.729 anexo A, baseado na estrutura CS-ACELP. No texto é dada atenção especial ao mecanismo de busca do dicionário adaptativo do codificador do G.729 anexo A, onde são implementadas as modificações propostas neste trabalho. Também são descritos os métodos de qualificação de codificadores de voz subjetivo MOS [7] e objetivo PESQ [8] utilizados na avaliação das modificações descritas nos capítulos que se seguem.

No capítulo 3, as modificações do *open loop* implementadas em trabalhos anteriores tem descritas suas técnicas de decimação de sinais no tempo, decimação de cálculos de correlação, da busca proximal da estimativa do atraso do *pitch*, assim como da estimativa de *pitch* distantes. Essas técnicas são introduzidas pelos parâmetros D_c , D_t , Δ_t e N_t , respectivamente, e a combinação adequada desses fatores reduz em 94,86% o esforço computacional empregado na etapa de *open loop* sem alterar a qualidade da codificação. Além disso, esse capítulo ainda traz a caracterização do banco de sinais de voz utilizado em relação à sua composição, à sua distribuição de atraso de *pitch* (*lag*) e à frequência de voz dos seus locutores.

As modificações implementadas nesse trabalho são apresentadas no capítulo 4. Em relação ao *open loop*, são descritas a alteração na técnica de limitação impostas sobre a estimativa do atraso do *pitch* e a expansão das demais técnicas dessa etapa. Assim, a limitação do conjunto de valores para fixação de T_{op} , a decimação

do sinal no cálculo da autocorrelação e à busca proximal do T_{op} , feitas através dos parâmetros N_a , N_b , D_c e Δ_t , proporcionaram uma redução da complexidade computacional de até 89,78% em relação ao G.729A original sem alteração da qualidade do sinal reproduzido. Já na etapa de *closed loop* são descritas a expansão da busca proximal da estimativa do atraso do *pitch*, usando o parâmetro r_{cl} , a alteração na implementação do cálculo do resíduo do filtro LP e a introdução da decimação do sinal no tempo através do parâmetro D_k . Quando consideradas todas as alterações do capítulo 4, há redução de 70,50% da complexidade computacional empregada nas etapas de *open loop* e *closed loop* em relação ao G.729A original também sem alteração da qualidade do sinal reproduzido.

A redução obtida com o codificador G.729A oriunda da combinação de todas as técnicas apresentadas no capítulo 4 leva a uma redução de 17% no tempo T de codificação de um segmento para o codificador G.729A, sem que isso ocasione a redução da qualidade de codificação original, o que foi confirmado por testes subjetivos.

5.2 Propostas para trabalhos futuros

A seguir são listadas sugestões para trabalhos futuros:

- estudar uma maneira de reduzir a complexidade computacional da busca do dicionário fixo através da decimação da correlação cruzada sinal alvo $x(n)$ atualizado e a possível excitação filtrada do dicionário fixo $c(n) * h(n)$ na busca no dicionário;
- combinar as técnicas aqui apresentadas com outras vistas na literatura;
- aplicar as propostas em outros codificadores que utilizam a filosofia CELP, como o MELP (*Mixed Excited Linear Prediction*), por exemplo, o que já vem sendo feito pelo aluno de mestrado Marcelo M. Ventura.

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 CSACELP Speech Codec*, 1996.
- [3] DE M. PREGO, T. L. NETTO, S., “Nova Aceleração da Busca no Dicionário Adaptativo do Codificador ITU-T G.729 Anexo A”. *SBrT - Simpósio Brasileiro de Telecomunicações*, 2009.
- [4] DE M. PREGO, T. L. NETTO, S., “Efficient Search in the Adaptive Codebook for ITU-T G.729 Codec”, *IEE Electronic Letters*, v. 16 , n. 10, 2009.
- [5] PREGO, T. D. M., *Aceleração dos codificadores de fala G.729 E G.729A.* , Tese de mestrado COPPE/UFRJ, 2009.
- [6] HWANG, S. H., “Computational improvement for G.729 standard”, *IEE Electronic Letters*, v. 36, n. 13, 2000.
- [7] ITU-T, *Rec. P.800, Methods for subjective determination of transmission quality*, 1996.
- [8] 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.
- [9] BISPO, B. C., *Otimização do codificador de voz CELP.* , Projeto final DEL-EE/UFRJ, 2005.