

# UM SISTEMA CELP PARA A CODIFICAÇÃO DA FALA A 4,4 KBPS

RANNIERY DA S. MAIA, CÁSSIO B. RIBEIRO, SERGIO L. NETTO, FERNANDO GIL V. RESENDE JR.

*Programa de Engenharia Elétrica/COPPE, DEL/EE*  
*Universidade Federal do Rio de Janeiro*  
*C.P. 68504, 21945-970, Rio de Janeiro, RJ, Brasil*  
*Emails: {ranniery, cassio, sergioln, gil} @lps.ufrj.br*

**Resumo**— Atualmente é de grande interesse codificar a fala com boa qualidade e baixas taxas de bits. Dentre todas as técnicas, a predição linear com excitação por códigos de dicionários (*code-excited linear prediction*, CELP) surge como uma possível solução para atender a estes requisitos. Este trabalho tem o objetivo de apresentar a implementação de um algoritmo CELP que opera a 4,4 kbps com baixa complexidade computacional e um método de busca rápida no dicionário adaptativo.

**Abstract**— Speech coding at low bit rates with good quality is a research area of major interest. Among the existing techniques, the method of code-excited linear prediction (CELP) comes up as a possible solution for such demanding requests. This work describes a CELP codec at 4.4 kbps with low computing complexity, and a fast search method for the adaptive codebook.

**Key Words**— Speech coding; coders; coding schemes; CELP.

## 1 Introdução

O processamento digital da fala tem avançado muito nas últimas décadas. Isso tem sido motivado em grande parte pelo desenvolvimento da tecnologia digital, possibilitando a realização de tarefas antes consideradas de alto custo, e pela crescente necessidade dos sistemas de telefonia móvel de oferecer melhores produtos e serviços. Para o caso de codificação, a intenção é realizar a transmissão da fala com a menor taxa possível, mantendo uma boa qualidade em relação à naturalidade e à inteligibilidade do sinal reconstruído.

Dentre todas as técnicas de codificação da fala digital, a predição linear com excitação por códigos de dicionários (*code-excited linear prediction*, CELP) surgiu como uma possível alternativa capaz de atender a estes requisitos (Schroeder and Atal, 1985). Várias melhorias posteriores foram propostas tendo em vista diminuir a complexidade computacional (Kim, 1998; Woodard and Hanzo, 1995), ou melhorar a qualidade (Kleijn et al., 1988) desta técnica de codificação. Por exemplo, para operar a taxas abaixo de 8 kbps e obter boa qualidade, é necessário que o sistema seja cuidadosamente projetado usando algumas características como quantização eficiente dos coeficientes de predição linear (Viswanathan and Markhoul, 1976; Sugamura and Farvardin, 1988), e dicionários que melhorem a qualidade da fala reconstruída (da Silva and Alcaim, 1997).

O sistema CELP aqui proposto foi implementado com o objetivo de obter boa qualidade a uma taxa de 4,4 kbps e baixa complexidade computacional. Para isso, foi utilizado um dicionário fixo esparsos e uma quantização diferencial escalar eficiente dos parâmetros do filtro de síntese. Além disso, um método de busca rápida no dicionário

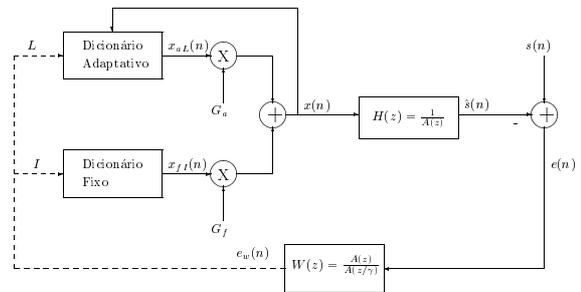


Figura 1. Diagrama de blocos do codificador CELP.

adaptativo também foi implementado.

Este trabalho está organizado da seguinte forma: a Seção 2 descreve as características gerais do sistema implementado; a Seção 3 mostra um procedimento para agilizar a busca no dicionário adaptativo; a Seção 4 mostra o desempenho do sistema com base em medidas objetivas de qualidade; por fim, a Seção 5 inclui as conclusões.

## 2 Características do sistema proposto

A Figura 1 mostra o diagrama de blocos do codificador CELP. A fala reconstruída é obtida pela passagem do sinal de excitação  $x(n)$  pelo filtro de síntese

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} \quad (1)$$

Os coeficientes de predição linear (CPL),  $\{a_i : i = 1, \dots, p\}$ , são obtidos através da técnica de análise LPC (*linear predictive coding*) (Markhoul, 1975).

O sinal de excitação  $x(n)$  é formado por uma seqüência oriunda do dicionário adaptativo  $x_{aL}(n)$  e outra do dicionário fixo  $x_{fI}(n)$ , escaladas pelos respectivos ganhos  $G_a$  e  $G_f$ . O dicionário

adaptativo (Kleijn et al., 1988) é responsável pela periodicidade do sinal de fala, e substitui de maneira mais eficiente o filtro de *pitch* presente nos primeiros sistemas CELP (Schroeder and Atal, 1985). O dicionário fixo contém um conjunto de seqüências que representam sinais residuais sem correlações de curto e longo termo (Atal, 1982), que podem ser ruído branco gaussiano ou seqüências geradas a partir de um treinamento (da Silva and Alcaim, 1997).

A informação enviada ao decodificador inclui os coeficientes do filtro  $H(z)$ , os índices  $L$  e  $I$ , e os ganhos  $G_a$  e  $G_f$  dos dicionários adaptativo e fixo, respectivamente. Estes quatro últimos parâmetros são obtidos através de um procedimento de análise-por-síntese, ou seja, todas ou algumas combinações de seqüências candidatas são testadas, escolhendo-se aquela que minimiza a energia do erro ponderado  $e_w(n)$ , resultado da filtragem do sinal de erro  $e(n)$  (dado pela diferença entre os sinais original e reconstruído) pelo filtro de ponderação do erro  $W(z) = \frac{A(z)}{A(z/\gamma)}$ , onde para o sistema implementado  $\gamma = 0,8$ , como em (Kroon and Swaminathan, 1992; da Silva, 1996).

O codificador CELP implementado processa o sinal da fala em blocos de 30 ms, que é o intervalo de tempo entre duas análises LPC consecutivas. Cada bloco é dividido em 4 sub-blocos de 7,5 ms de duração, onde para cada um destes é obtida a melhor seqüência de excitação  $x(n)$ . Nas sub-seções seguintes são dados detalhes a respeito destes dois procedimentos principais do codificador CELP: a análise LPC e obtenção do sinal de excitação.

### 2.1 A análise LPC

No CELP proposto, a análise LPC é feita pelo método da autocorrelação de ordem 10 ( $p = 10$ ), usando uma janela de Hamming de 25 ms centrada no último sub-bloco de cada bloco, como mostrado na Figura 2. É usado ainda o procedimento de suavização espectral (Tohkura et al., 1978) que aplica uma janela binomial de 130 Hz à seqüência de autocorrelação, quando da execução do algoritmo de Levinson-Durbin para o cálculo dos CPL. Este procedimento evita picos de banda muito estreita no espectro de curto termo do sinal de fala, picos estes causados pela alta resolução em freqüência devida ao fato da janela de Hamming englobar muitos períodos de *pitch* em sons sonoros. Com o emprego da suavização espectral, as bandas dos formantes ficam mais largas, produzindo assim sons mais naturais (Tohkura et al., 1978).

Os CPL são transformados em freqüências do espectro de linha (*line spectral frequencies*, LSF) (Kabal and Ramachandran, 1986). Para produzir uma transição suave entre conjuntos adjacentes de coeficientes LSF, estes são interpolados. Para a interpolação, os sub-blocos de um

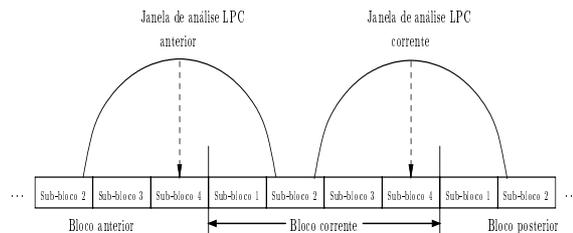


Figura 2. Posicionamento da janela de análise LPC.

determinado bloco são numerados de 1 a 4, conforme mostra a Figura 2. O  $i$ -ésimo coeficiente do  $n$ -ésimo sub-bloco é obtido por

$$w_i^n = (1 - q_n)w_i^a + q_n w_i^c, \quad (2)$$

onde  $w_i^c$  refere-se ao conjunto de coeficientes do bloco corrente,  $w_i^a$  ao do bloco anterior. O vetor de pesos usado foi  $\mathbf{q} = \{0, 25; 0, 5; 0, 75; 1\}$ . Deve-se reparar que  $w_i^4 = w_i^c, \forall i$ , já que a janela de Hamming tem seu centro coincidente com o sub-bloco 4.

Os coeficientes LSF de cada sub-bloco são quantizados escalarmente de forma diferencial com um total de 32 bits, distribuídos como indicado na Tabela 1, onde  $\hat{w}_i$  representa a versão quantizada de  $w_i$ . Neste esquema foram atribuídos 4 bits para os dois primeiros elementos, porque são eles os que mais afetam a representação do primeiro formante dos sons sonoros. A quantização da diferença entre um coeficiente  $w_i$  e o coeficiente anterior já quantizado evita que o erro de quantização se propague aos demais coeficientes (Sugamura and Farvardin, 1988). Para cada elemento, foi projetado um quantizador não-uniforme através do treinamento de uma base de dados pelo algoritmo de Lloyd I (Gersho and Gray, 1992). A base de dados para o treinamento consistiu de 15 sinais de fala, englobando 9 frases em português (Alcaim et al., 1992) de dois locutores do sexo masculino e um do sexo feminino (cada um pronunciou 3 frases), e 6 frases em inglês pronunciadas por 3 locutores do sexo masculino e 3 do sexo feminino.

Os quantizadores foram avaliados com uma outra base de dados formada por 20 frases de 10 locutores diferentes, onde cada um foi responsável pela pronúncia de 2 frases. A medida de distorção usada foi a distância espectral ( $DE$ ) dada por (Gray and Markel, 1976) entre os coeficientes obtidos do sinal de fala para cada sub-bloco e os quantizados e interpolados, conforme o procedimento descrito em (da Silva, 1996). A Tabela 2 mostra os resultados para blocos de 20 ms com sub-blocos de 5 ms (20/5), blocos de 30 ms com sub-blocos de 6 ms (30/6), e blocos de 30 ms com sub-blocos de 7,5 ms (30/7,5).

Tabela 1. Alocação de bits para quantização escalar diferencial das LSF.

| Elemento          | No. bits | Elemento             | No. bits |
|-------------------|----------|----------------------|----------|
| $w_1$             | 4        | $w_6 - \hat{w}_5$    | 3        |
| $w_2 - \hat{w}_1$ | 4        | $w_7 - \hat{w}_6$    | 3        |
| $w_3 - \hat{w}_2$ | 3        | $w_8 - \hat{w}_7$    | 3        |
| $w_4 - \hat{w}_3$ | 3        | $w_9 - \hat{w}_8$    | 3        |
| $w_5 - \hat{w}_4$ | 3        | $w_{10} - \hat{w}_9$ | 3        |
| Total             | 32 bits  |                      |          |

Tabela 2. Desempenho da quantização escalar diferencial das LSF para os vários tamanhos de blocos e sub-blocos.

|        | DE   | %DE $\in [2; 4]$ | %DE $> 4$ |
|--------|------|------------------|-----------|
| 20/5   | 1,52 | 18,73            | 0,78      |
| 30/6   | 1,82 | 31,06            | 2,70      |
| 30/7,5 | 1,81 | 30,90            | 2,91      |

## 2.2 Os parâmetros da excitação

O dicionário adaptativo usado varre a faixa de atrasos de 20 a 146, correspondendo a um total de 127 atrasos possíveis. Durante o processo de codificação, a saída de  $H(z)$  para cada seqüência candidata  $\{x_{aL}(n)\}$ , que será designada por  $\{y_{aL}(n)\}$ , é correlacionada com o sinal que se deseja reproduzir, e que será designado por  $\{t(n)\}$ . Esta correlação é obtida através de

$$C = \sum_{n=0}^{N-1} y_{aL}(n)t(n), \quad (3)$$

onde  $N$  é o número de amostras das seqüências (tamanho de um sub-bloco em amostras). Somente seqüências candidatas que ocasionam  $C > 0$  são levadas em conta. Caso  $C \leq 0$  para todas as seqüências candidatas de um determinado sub-bloco, o índice  $L$  é ajustado para 0 para indicar que naquele sub-bloco a excitação será formada apenas pelo dicionário fixo. Como descrito em (Kroon and Swaminathan, 1992), este procedimento melhora a qualidade da fala reproduzida, porque é obtida uma estimativa mais consistente do atraso. Outra vantagem é a de que o ganho  $G_a$  assume somente valores positivos, o que reduz a faixa dinâmica e ocasiona uma quantização mais eficiente.

O dicionário fixo é formado por 512 seqüências, cada uma com cerca de 90% de amostras nulas (dicionário esparsa). Ele foi gerado anulando-se todas as amostras de seqüências de ruído branco gaussiano com média 0 e variância 1 que estivessem abaixo de um determinado limiar, no caso 1,645 obtido heurísticamente, para possibilitar o grau de esparsamento em que 90 a 95% das amostras das seqüências do dicionário são nulas (Kroon and Swaminathan, 1992; Gersho, 1994). Além de uma boa qualidade, este tipo de dicionário permite agilizar a busca pela melhor excitação. O dicionário fixo utilizado no algorit-

mo de codificação proposto difere do apresentado em (Kroon and Swaminathan, 1992) porque as amostras com valores acima do limiar de ceifagem não são subtraídas deste valor.

Os ganhos  $G_a$  e  $G_f$  são quantizados de maneira escalar com 4 e 5 bits, respectivamente, após serem otimizados ao final do procedimento de análise-por-síntese para cada sub-bloco. Os quantizadores não-uniformes para cada coeficiente foram obtidos através de um treinamento pelo algoritmo de Lloyd I de um conjunto de treinamento, obtido ao codificar 17 sinais de fala pronunciados por locutores do sexo masculino e do sexo feminino nas línguas portuguesa e inglesa, pelo codificador CELP implementado sem quantização dos ganhos.

A Tabela 3 mostra as características de quantização dos parâmetros da excitação.

Tabela 3. Alocação de bits para quantização dos parâmetros da excitação.

| Parâmetro    | Faixa        | No. de bits |
|--------------|--------------|-------------|
| Ganho do DF  | -0,05 a 0,05 | 5           |
| Índice do DF | 0 a 511      | 9           |
| Ganho do DA  | 0 a 2        | 4           |
| Índice do DA | 0 a 127      | 7           |
| Total        | 25 bits      |             |

## Obtenção da melhor excitação

A Figura 3 mostra o diagrama de blocos detalhado do codificador CELP implementado. Este diagrama é obtido a partir do mostrado na Figura 1 após duas modificações: (1) o filtro perceptual  $W(z)$  é deslocado para os dois ramos à esquerda do somador, transformando o filtro  $H(z)$  no filtro de síntese ponderado  $H_w(z) = 1/A(z/\gamma)$ ; (2) a resposta do filtro  $H_w(z)$  no intervalo  $0 \leq n < N$ , onde  $N$  é o número de amostras de cada sub-bloco, é decomposta em duas partes: a resposta com estado inicial zero  $\hat{t}(n)$ , e a resposta à entrada zero  $\hat{s}_0(n)$ . Esta última seqüência não depende da seqüência de excitação candidata  $x(n)$  e, portanto, pode ser subtraída do sub-bloco de fala ponderado  $s_w(n)$  (resultado da filtragem de  $s(n)$  por  $W(z)$ ) antes de se iniciar o procedimento de análise-por-síntese. O sinal resultante

$$t(n) = s_w(n) - \hat{s}_0(n), \quad (4)$$

é geralmente denominado *signal alvo*, sendo portanto o sinal que se deseja reproduzir com o processo de análise-por-síntese.

Como pode ser visto na Figura 3, para testar cada seqüência candidata  $x(n)$ , é necessário apenas passá-la pelo filtro  $H_w(z)$  com estado inicial nulo. Esta estrutura propicia uma redução significativa no esforço computacional do processo de codificação.

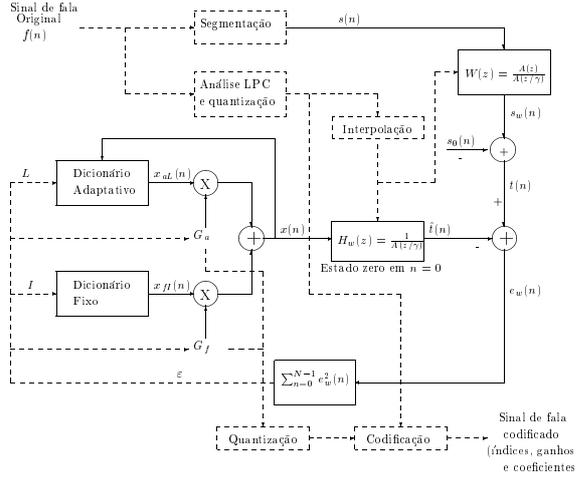


Figura 3. Diagrama de blocos detalhado do codificador CELP implementado.

Usando notação vetorial, da Figura 3 tem-se que o vetor reconstruído  $\hat{\mathbf{t}}$  gerado pelos vetores  $\mathbf{x}_{aL}$  e  $\mathbf{x}_{fI}$ , e pelos ganhos  $G_a$  e  $G_f$ , é dado por

$$\hat{\mathbf{t}} = G_a \mathbf{H}_w \mathbf{x}_{aL} + G_f \mathbf{H}_w \mathbf{x}_{fI}, \quad (5)$$

onde  $\mathbf{H}_w$  é uma matriz Toeplitz triangular inferior cujos elementos são amostras da resposta ao impulso do filtro  $H_w(z)$ , ou seja,

$$\mathbf{H}_w = \begin{pmatrix} h_w(0) & 0 & \cdots & 0 \\ h_w(1) & h_w(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_w(N-1) & h_w(N-2) & \cdots & h_w(0) \end{pmatrix}. \quad (6)$$

Definindo a resposta de  $H_w(z)$  à entrada  $x_{aL}(n)$  como  $y_{aL}(n)$ , e à entrada  $x_{fI}(n)$  como  $y_{fI}(n)$ , então

$$\mathbf{y}_{aL} = \mathbf{H}_w \mathbf{x}_{aL} \quad \text{e} \quad \mathbf{y}_{fI} = \mathbf{H}_w \mathbf{x}_{fI}. \quad (7)$$

A busca nos dicionários é feita de forma seqüencial: primeiro determina-se o melhor vetor no dicionário adaptativo considerando-o como único dicionário; depois é feita a procura pelo vetor no dicionário fixo considerando o vetor ótimo  $\mathbf{x}_{aL_{ot}}$  e o ganho  $G_a$  determinados. Neste caso o vetor alvo é  $\mathbf{t}$ , e deseja-se determinar o vetor  $\mathbf{x}_{aL}$  que minimiza o quadrado da norma do vetor erro ponderado

$$\varepsilon = \|\mathbf{e}_w\|^2 = \|\mathbf{t} - G_a \mathbf{y}_{aL}\|^2. \quad (8)$$

O ganho  $G_a$  que minimiza  $\varepsilon$  é obtido fazendo  $\partial\varepsilon/\partial G_a = 0$ , isto é

$$G_a = \frac{\mathbf{t}^T \mathbf{y}_{aL}}{\mathbf{y}_{aL}^T \mathbf{y}_{aL}}. \quad (9)$$

Substituindo (9) em (8), tem-se

$$\varepsilon = \mathbf{t}^T \mathbf{t} - \frac{(\mathbf{t}^T \mathbf{y}_{aL})^2}{\mathbf{y}_{aL}^T \mathbf{y}_{aL}}. \quad (10)$$

O erro  $\varepsilon$  dado por (10) é avaliado para cada vetor candidato  $\mathbf{x}_{aL}$ , sendo escolhido aquele que ocasionar o menor de todos os valores. Tendo feito isso, o ganho é calculado através de (9) com  $\mathbf{y}_{aL_{ot}}$  sendo a resposta do filtro  $\mathbf{H}_w$  ao melhor vetor escolhido no dicionário adaptativo  $\mathbf{x}_{aL_{ot}}$ .

Tendo escolhido o melhor vetor do dicionário adaptativo  $\mathbf{x}_{aL_{ot}}$  e determinado o ganho correspondente  $G_a$ , a busca do dicionário fixo é iniciada pela definição do novo sinal alvo  $\mathbf{t}_1$  dado por

$$\mathbf{t}_1 = \mathbf{t} - G_a \mathbf{y}_{aL_{ot}}. \quad (11)$$

A partir daí, a busca é realizada da mesma maneira que no dicionário adaptativo, ou seja, determina-se o vetor  $\mathbf{x}_{fI}$  que minimize o erro

$$\varepsilon = \mathbf{t}_1^T \mathbf{t}_1 - \frac{(\mathbf{t}_1^T \mathbf{y}_{fI})^2}{\mathbf{y}_{fI}^T \mathbf{y}_{fI}}, \quad (12)$$

e depois determina-se o ganho  $G_f$

$$G_f = \frac{\mathbf{t}_1^T \mathbf{y}_{fI_{ot}}}{\mathbf{y}_{fI_{ot}}^T \mathbf{y}_{fI_{ot}}}, \quad (13)$$

onde  $\mathbf{y}_{fI_{ot}}$  corresponde à resposta do filtro  $\mathbf{H}_w$  ao melhor vetor do dicionário fixo  $\mathbf{x}_{fI_{ot}}$  encontrado.

No final, os ganhos  $G_a$  e  $G_f$  são recalculados através de (14) e (15), para serem posteriormente quantizados.

$$G_a = \frac{\mathbf{t}^T \mathbf{y}_{aL_{ot}} - \mathbf{y}_{aL_{ot}}^T \mathbf{y}_{fI_{ot}} \mathbf{t}_1^T \mathbf{y}_{fI_{ot}}}{\mathbf{y}_{aL_{ot}}^T \mathbf{y}_{aL_{ot}}}, \quad (14)$$

$$G_f = \frac{\mathbf{t}_1^T \mathbf{y}_{fI_{ot}} - \mathbf{y}_{aL_{ot}}^T \mathbf{y}_{fI_{ot}} \mathbf{t}^T \mathbf{y}_{aL_{ot}}}{\mathbf{y}_{fI_{ot}}^T \mathbf{y}_{fI_{ot}}}. \quad (15)$$

Este método de busca da melhor excitação é sub-ótimo, pois não testa todas as combinações possíveis de excitações  $x_{aL}(n)$  e  $x_{fI}(n)$  para determinar o sinal mais próximo de  $t(n)$ . No entanto, o esforço computacional para isso é muito grande. Em (da Silva, 1996), são mostrados outros métodos de busca sub-ótima, sendo o utilizado aqui uns dos mais simples.

### 2.3 Taxa de bits do algoritmo proposto

Como observado nas sub-seções anteriores, 25 bits são utilizados para quantizar a excitação, e 32 bits para os parâmetros do filtro de síntese. Portanto a taxa de bits  $T$  pode ser determinada da seguinte forma:

$$T = \frac{25 \text{ bits}}{7,5 \text{ ms}} + \frac{32 \text{ bits}}{30 \text{ ms}} = 4,4 \text{ kbps}. \quad (16)$$

### 3 Busca rápida no dicionário adaptativo

O dicionário adaptativo corresponde a uma seqüência  $c_a(n)$  que armazena amostras da melhor excitação em instantes passados, ou seja,

$$\{c_a(n)\} = \{x(n) : n = -L_{max}, \dots, -1\}, \quad (17)$$

onde  $L_{max}$  é o comprimento do dicionário adaptativo. As seqüências candidatas  $x_{aL}(n)$  são obtidas a partir deste vetor de acordo com cada atraso  $L$ .

A busca pela melhor excitação exige grande esforço computacional, devido à filtragem de todas as seqüências candidatas por  $H_w(z)$ . Em (da Silva and Alcaim, 1995) é proposto um método de busca rápida no dicionário adaptativo através da modificação do modelo de síntese do codificador CELP, evitando que cada seqüência  $x_{aL}(n)$  seja filtrada por  $H_w(z)$ . O método proposto aqui também atinge este objetivo, só que de uma maneira diferente. Trata-se de filtrar o conteúdo do dicionário adaptativo  $c_a(n)$  por  $H_w(z)$  antes da busca pela melhor excitação, obtendo a seqüência correspondente ao dicionário adaptativo filtrado  $c_{aF}(n)$ . Em notação vetorial tem-se

$$\mathbf{c}_{aF} = \mathbf{H}_w \mathbf{c}_a, \quad (18)$$

onde  $\mathbf{H}_w$  é a matriz mostrada em (6), mas com dimensão  $L_{max}$ . A busca pela melhor excitação passa então a ser feita em  $c_{aF}(n)$ , sem que haja a necessidade de filtrar as seqüências candidatas por  $H_w(z)$ . Sendo determinado o melhor atraso  $L$ , o restante do procedimento é feito como mostrado na Seção 2.2, ou seja, faz-se a leitura da seqüência ótima  $x_{aL_{ot}}(n)$  em  $c_a(n)$  (dicionário não filtrado) e depois faz-se a filtragem por  $H_w(z)$ , para obter  $y_{aL_{ot}}(n)$ . De posse deste último, determina-se o ganho  $G_a$  através de (9), e o novo sinal alvo  $t_1(n)$  através de (11) para ser realizada a busca no dicionário fixo.

### 4 Desempenho

Foram usadas as seguintes medidas de qualidade objetivas para avaliar o desempenho do algoritmo CELP implementado: razão sinal-ruído segmentada perceptual (RSR-SP) (da Silva, 1996), distância cepstral (DC) (Gray and Markel, 1976), e distância de Itakura (DI) (Deller et al., 1993). Os testes foram realizados em 4 sentenças, cada uma delas composta de duas frases retiradas de (Alcaim et al., 1992) e pronunciadas por um locutor diferente, onde 2 foram do sexo masculino (M1 e M2), e 2 do sexo feminino (F1 e F2). As frases e os locutores foram diferentes dos utilizados nos projetos dos quantizadores dos CPL e ganhos dos dicionários. Os sinais foram digitalizados com 16 bits por amostra a 8 kHz, com posterior filtragem passa-altas para a remoção de ruídos de baixa freqüência.

A Tabela 4 mostra medidas de qualidade para o algoritmo proposto. O desempenho quanto à qualidade é comparado com DoD-CELP (*Department of Defense CELP*, CELP do Departamento de Defesa) que foi padronizado pelo Governo dos Estados Unidos para comunicação segura (*Federal Standard 1016*, FS-1016) (Gersho, 1994), e opera a uma taxa semelhante de 4,8 kbps. A Tabela 5 mostra os resultados para este sistema. Pode-se observar que o codificador proposto obtém melhor performance em todas as medidas utilizadas para todas as sentenças.

Tabela 4. Medidas objetivas de qualidade para o algoritmo CELP proposto.

| Locutor | RSR-SP (dB) | DC (dB) | DI (dB) |
|---------|-------------|---------|---------|
| M1      | 15,75       | 3,16    | 1,23    |
| M2      | 16,59       | 3,07    | 1,16    |
| F1      | 17,11       | 3,22    | 1,27    |
| F2      | 15,07       | 3,19    | 1,25    |

Tabela 5. Medidas objetivas de qualidade para o algoritmo DoD-CELP 4,8 kbps.

| Locutor | RSR-SP (dB) | DC (dB) | DI (dB) |
|---------|-------------|---------|---------|
| M1      | 9,43        | 3,54    | 1,53    |
| M2      | 9,92        | 3,51    | 1,50    |
| F1      | 13,89       | 3,40    | 1,40    |
| F2      | 11,45       | 3,51    | 1,49    |

A Tabela 6 mostra os resultados do algoritmo proposto com o procedimento de busca rápida no dicionário adaptativo (algoritmo acelerado). Observando os resultados, pode-se perceber que os resultados não diferem de forma significativa do algoritmo sem busca rápida.

Tabela 6. Medidas objetivas de qualidade para o algoritmo proposto acelerado.

| Locutor | RSR-SP (dB) | DC (dB) | DI (dB) |
|---------|-------------|---------|---------|
| M1      | 15,64       | 3,13    | 1,21    |
| M2      | 16,38       | 3,04    | 1,15    |
| F1      | 16,94       | 3,12    | 1,19    |
| F2      | 14,97       | 3,20    | 1,27    |

Os algoritmos com e sem o método de busca rápida no dicionário adaptativo foram executados em uma mesma máquina (uma estação de trabalho Sun Ultra-60) a fim de verificar a velocidade de processamento de cada um, no que diz respeito ao tempo gasto durante o procedimento de análise-por-síntese para a determinação dos melhores ganhos e índices, e também na análise LPC. A Tabela 7 mostra os resultados para cada sentença. Pode-se observar que o método de busca rápida reduziu o tempo de processamento entre 25% e 37%.

Tabela 7. Tempos de processamento do algoritmo proposto sem (ASBR) e com (ACBR) método de busca rápida no dicionário adaptativo para as sentenças consideradas.

| Sentença | Duração (s) | ASBR (s) | ACBR (s) |
|----------|-------------|----------|----------|
| M1       | 5,25        | 6,8      | 4,26     |
| M2       | 6,75        | 8,74     | 5,54     |
| F1       | 5           | 6,46     | 4,81     |
| F2       | 6           | 7,84     | 4,92     |

## 5 Conclusões

Foi mostrado neste trabalho um algoritmo de codificação CELP que opera a 4,4 kbps com baixa complexidade computacional, além de um procedimento de busca rápida no dicionário adaptativo que agiliza o processo de codificação enquanto mantém a qualidade em relação ao sistema convencional. O algoritmo proposto obteve qualidade superior ao padrão FS-1016 (DoD-CELP 4,8 kbps), o que foi verificado através de três medidas objetivas de qualidade.

## Referências

- Alcaim, A., Solemicz, J. A. and de Moraes, J. A. (1992). Frequência de ocorrência dos fonos e listas de frases foneticamente balanceadas no português falado no rio de janeiro, *Revista da Sociedade Brasileira de Telecomunicações* **7**(1).
- Atal, B. S. (1982). Predictive coding of speech at low bit rates, *IEEE Transactions on Communications* **30**(4): 600–614.
- da Silva, L. M. (1996). *Contribuições para a melhoria da codificação CELP a baixas taxas de bits*, PhD thesis, PUC-RJ.
- da Silva, L. M. and Alcaim, A. (1995). A modified celp model with computationally efficient adaptive codebook search, *IEEE Signal Processing Letters* **2**(3): 44–45.
- da Silva, L. M. and Alcaim, A. (1997). Multipulse stochastic codebook for celp speech coding, *Proceedings of ICSP '97*, Seoul, Korea.
- Deller, J., Proakis, J. and Hansen, J. (1993). *Discrete-time Processing of Speech Signals*, Macmillan, New York.
- Gersho, A. (1994). Advances in speech and audio compression, *Proceedings of the IEEE* **82**(6): 900–918.
- Gersho, A. and Gray, R. M. (1992). *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston.
- Gray, A. H. and Markel, J. D. (1976). Distance measures for speech processing, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **24**: 380–391.
- Kabal, P. and Ramachandran, R. P. (1986). The computation of line spectral frequencies using chebyshev polynomials, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **34**(6): 1419–1426.
- Kim, H. K. (1998). Adaptive encoding of fixed codebook in celp coders, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 149–152.
- Kleijn, W. B., Krasinski, D. J. and Ketchum, R. H. (1988). Improved speech quality and efficient vector quantization in selp, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 155–158.
- Kroon, P. and Swaminathan, K. (1992). A high-quality multirate real-time celp coder, *IEEE Journal on Selected Areas in Communications* **10**(5).
- Markhoul, J. (1975). Linear prediction: a tutorial review, *Proceedings of the IEEE* **63**(4): 561–579.
- Schroeder, M. R. and Atal, B. S. (1985). Code-excited linear prediction (celp): high-quality speech at very low bit rates, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tampa, pp. 937–940.
- Sugamura, N. and Farvardin, N. (1988). Quantizer design in lsp analysis-synthesis, *IEEE Journal on Selected Areas in Communications* **6**(2): 432–440.
- Tohkura, Y., Itakura, F. and Hashimoto, S. (1978). Spectral smoothing technique in parcor speech analysis-synthesis, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26**(6).
- Viswanathan, R. and Markhoul, J. (1976). Quantization properties of transmission parameters in linear predictive systems, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **23**: 309–321.
- Woodard, J. P. and Hanzo, L. (1995). Improvements to the analysis-by-synthesis loop in celp codecs, *Proceedings of the Radio Receivers and Associated Systems Conference*, Bath, UK, pp. 114–118.