Técnicas para Excitação em Codificadores de Voz do Tipo LPC

Abraham Alcaim*, José Roberto Boisson de Marca**
e Maurizio Copperi***

A última década foi palco de um excepcional desenvolvimento das técnicas de codificação de voz que evoluíram a partir do modelo tradicional LPC com a introdução de funções excitação mais elaboradas. Estas funções são obtidas através de um procedimento de análise por síntese. A uma taxa de 16 kbit/s esses métodos são capazes de fornecer uma qualidade de voz comparável à de um sistema log-PCM de 7 bits e também àquela do padrão CCITT para codificação em 32 kbit/s. Estes métodos também têm se mostrado úteis para taxas em torno de 8 kbit/s, sendo um elemento desta família o atual padrão (IS-54) para comunicações móveis nos Estados Unidos. Estas técnicas constituem ainda uma alternativa promissora para codificação em 4 kbit/s e sua complexidade é tal que permite uma implementação em apenas um pequeno número de processadores do tipo DSP. O presente trabalho apresenta uma visão unificada dos métodos hoje existentes para modelagem da excitação em codificadores de voz do tipo LPC, extração e representação digital dos parâmetros da excitação, bem como o desempenho obtido com essa classe de codificadores.

1. INTRODUÇÃO

Nos últimos vinte anos o modelo de predição linear (LPC) para produção da voz tem sido empregado com sucesso na codificação digital desses sinais [1]-[3]. Neste modelo, um filtro de síntese H(z), só com pólos e variante no tempo tem à sua entrada uma seqüência de excitação u(n), gerando então à sua saída o sinal sintetizado $\hat{s}(n)$. No modelo convencional mostrado na **Fig. 1**, a seqüência de excitação pode ter duas formas: na primeira, associada a sons sonoros, ela consiste de um trem periódico de pulsos espaçados por um período T_0 (o período fundamental ou tonal da voz – "pitch" em inglês); a segunda forma é a de uma seqüência assemelhando-se a um ruído branco e que é empregada na representação de sons surdos.

* CETUC-PUC/Rio, 22453 – Rio de Janeiro, RJ.

** Speech Research Department, AT&T, Bell Laboratories, Murray Hill, EUA, em licença sabática do CETUC-PUC/Rio.

*** R&D Department, SIP Headquarters, Via San Daimazzo 15, 10122 – Torino, Italia.

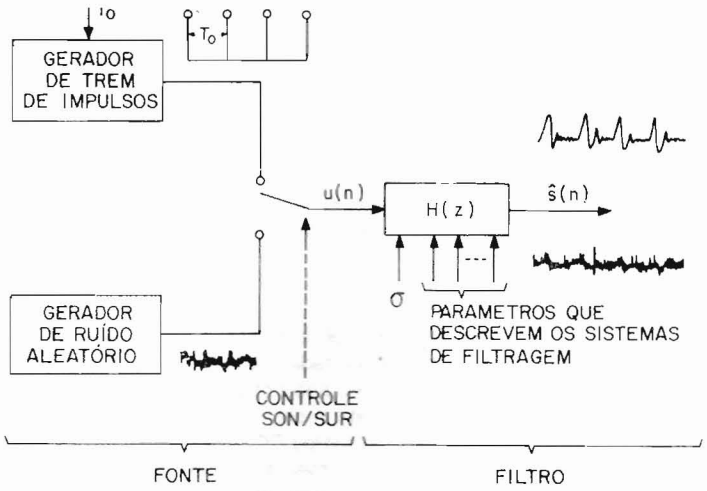

Figura 1. Modelo de excitação com um pulso por período fundamental para sons sonoros.

Embora o chamado vocoder LPC apresente uma qualidade de voz que pode ser considerada aceitável para taxas entre 2,4 e 4 kbit/s, não é possível melhorar o seu desempenho de forma perceptível aumentado-se a taxa de codificação. Esse comportamento é devido a limitações fundamentais do modelo de produção da **Fig. 1**, que são em parte provocadas pela maneira simplista e rígida de gerar a função excitação. Especialmente, é sabido que a classificação dos sons da fala em apenas duas categorias (exclusivamente sonoros e exclusivamente surdos) está longe de ser perfeita. Existem, por exemplo, sons tais como os fricativos sonoros (/v/, /z/, etc.) cuja formação envolve ambos os tipos de excitação. Em algumas situações chega a ser impossível afirmar que se um certo som é predominantemente sonoro ou surdo. É ainda verdade que a excitação orgânica que produz os sons sonoros não é exatamente periódica, apresentando pequenas (mas importantes) variações de um período para outro.

Portanto, as particularidades do mecanismo de produção de voz exigem que, para atingir um melhor desempenho na codificação, uma seqüência de excitação mais complexa seja utilizada. Na seção 2, é apresentado um procedimento para obtenção dos parâmetros representativos do sinal de excitação no qual não é necessário o conhecimento a priori do tipo de som que está sendo digitalizado. Em contraste com o vocoder tradicional LPC (**Fig. 1**), neste procedimento é permitido que a excitação contenha mais de um pulso por período fundamental. As posições e amplitudes dos pulsos que

compõem a seqüência de excitação são determinadas através de um método de análise por síntese que conduz a um problema de otimização que pode ser resolvido tanto no domínio do tempo quanto no da freqüência. Neste trabalho apenas o enfoque no domínio do tempo será abordado. Para concluir, na seção 2 são mencionados alguns métodos para representação eficiente dos parâmetros da excitação em forma binária.

Uma das primeiras aplicações encontradas para esta família de codificadores de voz foi na especificação de um sistema para utilização no serviço de comunicações móveis europeu. O codificador escolhido como padrão, descrito na seção 3, faz uso de uma excitação com múltiplos pulsos com a particularidade de que esses pulsos são igualmente espaçados, o que simplifica sobremaneira o projeto do codificador.

A seção 4 trata da sub-classe dos métodos de codificação aqui abordados que tem recebido maior atenção por parte dos pesquisadores, especificamente os codificadores com excitação por dicionário de códigos (CELP). Neste procedimento um conjunto (dicionário) de seqüências possíveis de serem usadas é pré-armazenado na memória do codificador. Para cada bloco de amostras de voz é escolhida uma (a melhor) seqüência dentre aquelas pertencentes ao dicionário para servir como excitação para aquele bloco.

Uma redução do número de pulsos necessários à formação da função excitação (ou do número de seqüências que compõem o dicionário no caso de sistema CELP) pode ser alcançada com a inclusão de um preditor com retardo longo. Equivalentemente a inclusão deste preditor, que será discutida na seção 5, permite uma melhora significativa no desempenho para uma dada taxa de bits, em especial para vozes femininas.

Da maneira como foram originalmente concebidos, os codificadores aqui analisados apresentam a desvantagem de um retardo de codificação muito longo para aplicações em rede. Esse problema pode ser praticamente eliminado através do uso de estruturas realimentadas, o que será visto na seção 6.

Valores típicos para o partilhamento dos bits entre os diferentes parâmetros a serem quantizados e dados sobre o desempenho dos sistemas aqui discutidos são encontrados nas seções 7 e 8, respectivamente. A seção 9 é dedicada aos comentários finais.

2. O MODELO DE EXCITAÇÃO MULTI-PULSO

Atal e Remde [4] desenvolveram em 1982 um novo procedimento de excitação onde não é feita suposição alguma à priori sobre a natureza da excitação, exceto que ela consiste de uma seqüência de pulsos. Este modelo é, portanto, válido para todas as classes de sons da fala. Além disso, não há tentativa de forçar que a excitação seja periódica ou não-periódica. Conseqüentemente, esta solução evita dois grandes problemas: a classificação sonoro-surdo e a detecção do período fundamental. Nesta abordagem, as incógnitas são as posições e amplitudes dos pulsos. As amostras de entrada são particionadas em blocos e, para cada bloco estes parâmetros (incógnitas do problema) são calculados. O método proposto para realizar este cálculo usa um procedimento de análise por síntese que é ilustrado na **Fig. 2**.

Para determinar a excitação em codificadores LPC com excitação multipulso (MP), a seqüência de pulsos u(n) é passada através do filtro de síntese LPC, cuja saída $\hat{s}(n)$ é o sinal de voz sintetizado. A diferença entre o sinal de voz original s(n) e $\hat{s}(n)$ é o erro de síntese (não ponderado) e(n).

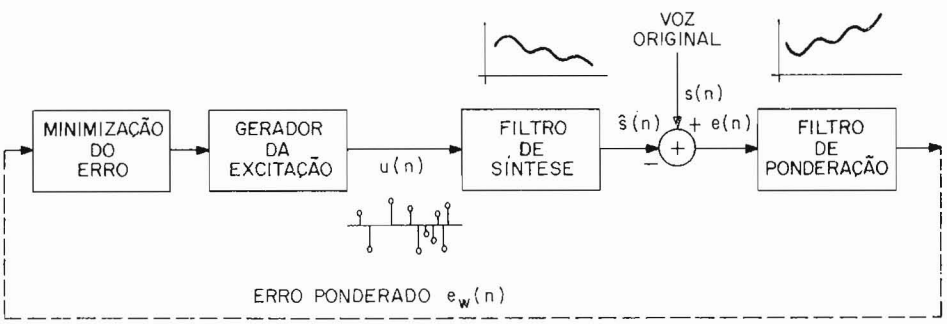

Figura 2. Procedimento para determinar a excitação em codificadores LPC com excitação multipulso [4].

Da teoria do mascaramento auditivo [5] sabe-se que nas regiões espectrais em torno das formantes a voz é menos sensível ao ruído. Ou seja, nessas regiões o sinal de voz consegue um melhor mascaramento do ruído. Por esse motivo, se o erro de síntese for ponderado de forma adequada ele será mais

significativo em termos da percepção humana. Isto é exatamente o que é feito no procedimento ilustrado na **Fig. 2**. O erro de síntese e(n) é filtrado por um filtro de ponderação que atenua o erro nas regiões de freqüência onde ele é melhor mascarado pela voz (as regiões das formantes) e acentua o erro nas outras regiões (vales do espectro). Isso faz com que, após o processo de minimização, a energia do erro seja relativamente maior nas regiões das formantes, onde o erro é mais tolerável.

Uma possível função de transferência para o filtro de ponderação W(z) é dada em [4]:

$$W(z) = \frac{H(\gamma^{-1}z)}{H(z)} = \frac{1 - \sum_{i=1}^{p} a_i z^{-i}}{1 - \sum_{i=1}^{p} a_i \gamma^i z^{-i}} \qquad (1)$$

onde H(z) é o filtro de síntese LPC de ordem p,

$$H(z) = \left[1 - \sum_{i=1}^{p} a_i z^{-i} \right]^{-1} \qquad (2)$$

e o parâmetro γ é um número real entre 0 e 1. Se γ=1 então W(z)=1, o que significa que não há ponderação. Por outro lado, para γ=0 o filtro de ponderação passa a ser igual ao inverso do filtro de síntese, isto é, $W(z)=H^{-1}(z)$. Testes de escuta indicaram que um bom valor de γ é 0,8 [4], [6].

Os parâmetros da excitação (posições e amplitudes dos pulsos) são determinados de modo a minimizar a energia do erro ponderado e(n). Esta determinação é feita seqüencialmente, para cada bloco de amostras de entrada e, geralmente, após todas as posições terem sido encontradas é feita uma reotimização global das amplitudes dos pulsos. O sistema de codificação que opera de acordo com esses princípios é denominado *codificador LPC com excitação multipulso (MP-LPC).*

2.1. *Determinação das posições e amplitudes dos pulsos*

A excitação multipulso pode ser descrita matematicamente por

$$u(n) = \sum_{k=1}^{K} \alpha_k \, \delta(n - n_k) \qquad (3)$$

onde K é o número de pulsos por bloco de excitação, $\delta(n)$ é a seqüência impulso unitário, α_k é a amplitude do k-ésimo pulso e n_k sua posição.

É evidente que se fizermos K igual a N, o número de amostras por bloco de excitação, será possível reconstruir a voz original com uma qualidade muito alta. Para isso, entretanto, será necessário empregar uma também alta taxa de bits para codificar as posições e amplitudes de todos os N pulsos em vez de um número reduzido (muito menor que N). Assim, o valor de K deve ser determinado com base em um compromisso entre taxa de bits e desempenho. Felizmente, foi verificado que valores de K da ordem de N/10 são suficientes para a produção dos diferentes tipos de som da fala [4], [7]. Usando estes valores é possível projetar codificadores com taxa de bits total na faixa de 8 a 9,6 kbit/s. Para taxas da ordem de 16 kbit/s, K é tipicamente da ordem de 2N/10.

É possível desenvolver técnicas para extrair os parâmetros da excitação $\{\alpha_k\}$ e $\{n_k\}$ tanto no domínio do tempo como no domínio da freqüência. Como já mencionado aqui será visto apenas o caso mais usualmente adotado que é a formulação no domínio do tempo.

Idealmente, os parâmetros $\{\alpha_k\}$ e $\{n_k\}$ são obtidos de modo a minimizar a energia do erro ponderado $e_w(n)$, expressa por

$$\varepsilon = \sum_{n=1}^{N} e_w^2(n) \qquad (4)$$

O erro de síntese ponderado é definido por (vide **Fig. 2**)

$$e_W(n) = e(n) * w(n) \qquad (5)$$

onde w(n) é a transformada-z inversa de W(z),

$$e(N) = s(n) - \hat{s}(n) \qquad (6)$$

e * representa a operação de convolução discreta.

O sinal sintetizado (reconstruído) $\hat{s}(n)$ está relacionado à excitação u(n) através da expressão

$$\hat{s}(n) = u(n) * h(n) \qquad (7)$$

onde h(n) é a resposta impulsional do filtro de síntese. Substituindo (3), (5)-(7) em (4) resulta que

$$(\varepsilon = \sum_{n=1}^{N} \left[s_W(n) - \sum_{1=1}^{K} \alpha_1 h_W(n - n_1) \right]^2 \qquad 8)$$

onde

$$s^W(n) = s(n) * w(n)$$

e

$$h_W(n) = h(n) * w(n)$$

O problema de otimização que consiste em minimizar ε com relação a ambos os conjuntos $\{\alpha_k\}$ e $\{n_k\}$ é extremamente complexo. Um procedimento eficiente, embora sub-ótimo, consiste em determinar seqüencialmente as amplitudes e posições dos pulsos, um pulso de cada vez. Para isso, consideremos que as posições $n_1, n_2, \ldots, n_{j-1}$ e suas respectivas amplitudes já tenham sido determinadas. Denotando por $s_{j-1}(n)$ o sinal sintetizado quando a excitação é composta por estes (J-1) pulsos e $e_{W,j-1}(n)$ a diferença ponderada, entre s(n) e $s_{j-1}(n)$, tem-se que

$$e_{W,j-1}(n) = s_W(n) - \sum_{1=1}^{J-1} \alpha_1 \, h_W(n - n_1) \qquad (9)$$

A partir de (8) e (9) resulta a seguinte expressão para a energia do erro ponderado, após J pulsos terem sido alocados à excitação:

$$\varepsilon_J = \sum_{n=1}^{N} \left[e_{W,J-1}(n) - \alpha_J \, h_W(n - n_J) \right]^2 \qquad (10)$$

Derivando em relação a α_J e igualando a zero obtém-se a amplitude do J-ésimo pulso que minimiza ε_J:

$$\alpha_J^{op} = \frac{\sum_{n-1}^{N} e_{W,J-1}(n) \, . \, h_W(n - n_J)}{\sum_{n=1}^{N} H_W^2(n - n_J)} \qquad (11)$$

Substituindo agora (11) em (10), resulta que

$$\varepsilon_J = \sum_{n=1}^{N} e_{W,J-1}^2(n) - \frac{\left[\sum_{n-1}^{N} e_{W,J-1}(n) \, h_W(n - n_J) \right]^2}{\sum_{n=1}^{N} h_W^2(n - n_J)} \qquad (12)$$

Note que o segundo termo (que é não-negativo) do lado direito de (12) corresponde à redução do erro, devida à alocação do J-ésimo pulso adicional à excitação. Portanto, a posição ótima n_J^{op}, do J-ésimo pulso pode ser determinada como aquela que fornece o valor máximo do segundo termo de (12). Usando este valor de n_J em (11) obtém-se a amplitude do J-ésimo pulso.

O procedimento para determinar os parâmetros de excitação para um dado bloco de amostras (intervalo de análise) pode ser resumido como se segue. No início (J=1), sem qualquer pulso de excitação, gera-se o sinal sintetizado

com base na memória do filtro de síntese a partir de blocos (intervalos de análise) anteriores. Um sinal erro $e_{w,1}(n)$ é, então, calculado subtraindo-se esse sinal sintetizado do sinal de voz original e passando-se o resultado pelo filtro de ponderação. Determina-se, em seguida, a posição e amplitude do primeiro pulso que minimiza a energia de $e_{w,1}(n)$. As equações (11) e (12) são, então, usadas para determinar a posição e amplitude do segundo pulso, e assim por diante, até o K-ésimo pulso.

Uma forma alternativa para representar α_J^{op} e ε_J é em termos da seqüência de correlação cruzada [7], [8]

$$R_{s_w h_w}(k) = \sum_{n=1}^{N} s_w(n)\, h_w(n-k) \tag{13}$$

e da seqüência de autocorrelação

$$R_{h_w}(j,k) = \sum_{n=1}^{N} h_w(n-j)\, h_w(n-k) \tag{14}$$

Para obter essa representação, basta usar (8) com o limite superior K do somatório trocado por J e utilizar as definições (13) e (14). Derivando ε_J em relação a α_J e igualando a zero, obtém-se α_J^{op}:

$$\alpha_J^{op} = \frac{R_{s_w h_w} - \sum_{i=1}^{J-1} \alpha_i R_{h_w}(n_i,n_J)}{R_{h_w}(n_J,n_J)} \tag{15}$$

Substituindo esse valor de α_J^{op} em (10) resulta que

$$\varepsilon_J = \sum_{n=1}^{N} e^2_{w,J-1}(n) - \frac{\left[R_{s_w h_w}(n_J) - \sum_{i=1}^{J-1} \alpha_i R_{h_w}(n_i,n_J) \right]^2}{R_{h_w}(n_J,n_J)} \tag{16}$$

Esta forma de expressar α_J^{op} e ε_J permite, após algumas simplificações que serão feitas agora, reduzir a complexidade da busca de α_J^{op} e n_J^{op}. Para isso, observemos que, a partir de (1)

$$H_W(z) = H(Z)\, W(z)$$

$$= H(\gamma^{-1} z) \qquad (17)$$

de onde se obtém (utilizando propriedade de transformada-z) a resposta impulsional

$$h_W(n) = h(n)\gamma^n$$

Considerando que $\gamma<1$, a aproximação $h_W(n) \approx 0$ para n maior que um dado $n_g, n_g \leq N$, é válida na maioria das situações práticas. Por exemplo, um bloco de tamanho N=80 e valor de γ=0,8 resulta em $h_W(N) \approx 1{,}8 \times 10^{-8}\, h(0)$. É também verdade que a condição de causalidade exige que $h_w(n) = 0$ para $n< 0$. Estas duas propriedades de $h_w(n)$ permitem re-escrever (14) como

$$R_{h_W}(j,k) \approx \sum_{n=-\infty}^{\infty} h_W(n-j)\, h_W(n-k)$$

$$= \sum_{n=-\infty}^{\infty} h_W(n)\, h_W(n + |k-j|)$$

$$= \sum_{n=1}^{N-|k-j|} h_W(n) h_W(n + |k-j|) \qquad (18)$$

Portanto, $R_{h_W}(j,k) = R_{h_W}(l)$, $l = |k-j|$.

Usando (18), pode-se re-escrever (16) da seguinte forma:

$$\varepsilon_J = \sum_{n=1}^{N} e^2_{W,J-1}(n) - \frac{q^2(n_J)}{R_{h_W}(0)} \qquad (19)$$

onde

$$q(n_j) = R_{s_w h_w}(n_J) - \sum_{i=1}^{J-1} \alpha_i R_{h_w}(\,|\, n_j - n_i \,|\,) \qquad (20)$$

O numerador da fração do lado direito de (19) é sempre positivo e $R_{h_w}(0)$ não depende de n_J. Assim, a posição ótima n_J^{op} é aquela que maximiza $|\, q(n_J) \,|$. De (15), (18) e (20) resulta a correspondente amplitude ótima do J-ésimo pulso:

$$\alpha_J^{op} = \frac{q(n^{op})}{R_{h_w}(0)}$$

2.2. *Reotimização das amplitudes*

Como foi mencionado anteriormente, o procedimento de busca dos parâmetros da excitação que consiste em determinar um pulso de cada vez é sub-ótimo. Este procedimento pode ser melhorado se, a cada nova posição de pulso encontrada, for feita uma reotimização de suas amplitudes. Com isto leva-se em consideração a correlação entre as amplitudes dos pulsos. Resultados de simulação mostram, entretanto, que basta fazer a reotimização após todos os K pulsos terem sido encontrado, ou seja, uma única vez [9], [10]. Pouco ganho adicional se obtém ao fazer reotimização a cada novo pulso encontrado. Os valores finais de $\alpha_1,\ldots, \alpha_k$ são obtidos de forma a minimizar o erro total ε definido em (8). Derivando ε em relação a $\alpha_1,\ldots, \alpha_k$ e igualando a zero resulta no seguinte sistema de equações

$$\sum_{k=1}^{K} \alpha_k \, R_{h_w}(n_k^{op} , n_i^{op}) = R_{s_w h_w}(n_i^{op}) \quad , \quad i = 1,\ldots,k \qquad (21)$$

2.3. *Codificação das posições e amplitudes dos pulsos*

Existem diversas maneiras de codificar os parâmetros de excitação. Aqui serão ilustrados alguns procedimentos, começando-se com a codificação de amplitudes.

Em geral, as amplitudes são normalizadas de modo a reduzir o número de bits necessário para quantizar cada uma delas. Bons candidatos para o parâmetro de normalização são

$$\alpha_{max} = \max\{ |\alpha_1| , |\alpha_2| , ..., |\alpha_K| \}$$

e

$$\alpha_{rms} = \sqrt{\frac{1}{K} \sum_{i=1}^{K} \alpha_i^2}$$

Esse parâmetro é normalmente digitalizado com um log-PCM de 6 ou 7 bits e as amplitudes são codificadas usando um PCM uniforme tipicamente de 3, 4 e 5 bits para taxas totais de 8, 9, 6 e 16 kbit/s [8], [11].

Para codificação das posições dos pulsos pode-se considerar uma solução direta, que consiste em quantizar a primeira posição (n_1^{pp}) com $\lceil \log_2 N \rceil$ bits[1]. As outras posições podem ser codificadas com um esquema diferencial simples que quantiza as diferenças entre duas posições consecutivas com 5 bits. Esse método tem a desvantagem de limitar a distância entre dois pulsos consecutivos.

Uma técnica mais elaborada de codificação das posições é baseada no fato de que existem C_N^K maneiras de se distribuir K pulsos em N posições. Associando-se um índice $I \in \{0,1,...,C_N^K - 1\}$ a cada uma das possíveis seqüências resultantes da distribuição de K pulsos em N posições, observa-se que serão suficientes $\lceil \log_2 C_N^K \rceil$ bits para representar I. Por exemplo, em um sistema a 9,6 kbit/s com N=80 e K=8 são necessários 35 bits para codificar as 8 posições. Quando comparado com o método que codifica as diferenças com 5 bits, isto representa uma economia de 700 bit/s. Para um sistema de 16 kbit/s com N=160, essa economia alcançaria 3,8 kbit/s.

Para evitar que uma grande quantidade de memória seja gasta no armazenamento das C_N^K seqüências de pulsos o seguinte algoritmo foi proposto em [12]:

[1] [A] = menor inteiro maior que A.

Passo 1: Associar à seqüência de excitação um vetor N-dimensional $c = (c_1, c_2, ..., c_N)$ que contém o número 1 nas posições n_1^{op} , n_2^{op} , ... , n_K^{op} e zero nas outras posições.

Passo 2: Fazer I = 0, i = N+1, j = K.

Passo 3: i = i-1

Passo 4: Se $c_i = 1$, fazer $I = I + C_{i-1}^{J}$.

Se $c_i = 0$, retornar ao passo 3.

Passo 5: j = j-1

Se $j \leq 0$, parar.

Caso contrário, retornar ao passo 3.

3. EXCITAÇÃO COM PULSOS REGULARMENTE ESPAÇADOS

O conceito de pulsos regularmente (ou uniformemente) espaçados foi introduzido por Kroon et alii. [13], [14] com o objetivo de reduzir a complexidade do codificador.

É claro que se os K pulsos que compõem a excitação multipulso forem igualmente espaçados, as posições de K-1 desses pulsos podem ser determinadas a partir da posição de um único pulso. Com isso, o tempo de processamento pode ser economizado, pois deixa de haver a necessidade de se fazer a busca de cada oposição individualmente. Como existem K pulsos de excitação para cada bloco de N amostras, o espaçamento entre os pulsos é $\Delta = N/K$. Isso significa que existem Δ possíveis seqüências de posição da excitação candidatas. A **Fig. 3** ilustra as 4 possíveis seqüências para o caso típico de N=40 e K=10.

Um método para se escolher a melhor seqüência de posição da excitação, dentre as possíveis candidatas, pode ser resumido através dos seguintes passos [14]:

1) Para cada uma das Δ seqüências candidatas calcular as amplitudes ótimas a partir de (21);

2) Calcular, para cada seqüência de posições candidata, o erro $\varepsilon^{(\iota)}$,$\iota = 1,..., \Delta$, expresso por

Figura 3. Possíveis seqüências de posição da excitação de pulsos regularmente espaçados, para o caso N=40 e K=10.

$$\varepsilon^{(\iota)} = \sum_{n=1}^{N} \left[s_W(n) - \sum_{i=1}^{K} \alpha_i^{(\iota)} h_W(n - n_i^{(\iota)}) \right]$$

onde $n_i^{(\iota)}$ e$\alpha_i^{(\iota)}$ são, respectivamente, a posição e a amplitude do i-ésimo pulso da ι-ésima seqüência de excitação candidata;

(3) Determinar o valor de ι que minimiza $\varepsilon^{(\iota)}$.

É importante notar que o método descrito pelos passos acima necessita da solução de Δ sistemas de equações lineares.

Uma vantagem adicional da utilização de pulsos regularmente espaçados é a redução do número de bits necessários para codificar a informação relativa às posições dos pulsos. Esta redução permite que seja feita uma quantização mais fina dos outros parâmetros do codificador ou que mais pulsos possam ser incluídos na excitação.

Uma configuração especial do codificador LPC com excitação com pulsos regularmente espaçados foi escolhida pelo CEPT Groupe Speciale Mobile

como padrão para o sistema rádio móvel digital Pan-Europeu [15]. Essa configuração será brevemente descrita a seguir.

3.1. *Padrão para o sistema móvel pan-europeu*

Após pré-processamento, o sinal de voz é dividido em segmentos de 20ms não superpostos. Para cada segmento são calculados 8 coeficientes LAR ("log-area ratio"), os quais são quantizados uniformemente com 6 bits para os coeficientes 1 e 2, 5 bits para os coeficientes 3 e 4, 4 bits para os coeficientes 5 e 6, e 3 bits para os coeficientes 7 e 8.

Neste sistema é utilizado um preditor com retardo longo de 1ª ordem cujos parâmetros são calculados a cada 5ms. O ganho é quantizado com 2 bits e o retardo com 7 bits.

A excitação regularmente espaçada é obtida, a cada 5ms, a partir do resíduo final, o qual resulta da filtragem do sinal de voz através do inverso do filtro H(z) e do inverso do filtro que realiza a predição com retardo longo (veja Seção 5). Esse resíduo é, então, filtrado por um filtro passa-baixa, otimizado a cada quadro no sentido de minimizar o valor médio quadrático do erro de síntese ponderado. O sinal na saída desse filtro é composto por 40 amostras denotadas por x(k), k=0,..., 39. Define-se agora quatro seqüências de excitação candidatas de comprimento 13:

$\underline{X}_0 = \{ x(0), x(3), x(6), x(9), ..., x(36) \}$

$\underline{X}_1 = \{ x(1), x(4), x(7), x(10), ..., x(37) \}$

$\underline{X}_2 = \{ x(2), x(5), x(8), x(11), ..., x(38) \}$

$\underline{X}_3 = \{ x(30), x(6), x(9), x(12), ..., x(39) \}$

É possível mostrar [14] que a seqüência ótima é aquela que tem maior energia. É importante ressaltar [14] que essa solução, utilizando o filtro passa-baixa otimizado, é equivalente à solução do sistema de equações (21).

Finalmente, a seqüência escolhida é quantizada com um quantizador adaptativo de 3 bits que utiliza como parâmetro de ajuste do quantizador o valor

máximo das magnitudes das amostras da seqüência. Este parâmetro é digitalizado logaritmicamente com 6 bits e o índice da seqüência é quantizado com 2 bits. Resultam portanto 260 bits a cada 20ms, o que significa uma taxa de 13kbit/s. A qualidade da voz é bastante superior à que se obtém com os sistemas rádio móveis analógicos [15].

4. EXCITAÇÃO POR DICIONÁRIO DE CÓDIGOS EM CODIFICADORES LPC

Com o objetivo de reduzir ainda mais a taxa de bits de codificadores LPC, o sinal excitação pode ser obtido de um dicionário que contém um conjunto de excitações candidatas ou seqüências de inovação [16]. Essa técnica é então chamada de *excitação por dicionário de códigos* e o codificador é denotado pela sigla CELP ("Code-excited linear prediction"). Observe-se que, nessa técnica, o sinal excitação é representado por uma dentre um conjunto de formas de onda pré-armazenadas. Já a excitação multipulso busca parametrizar o sinal excitação. Ambas as técnicas, porém, utilizam um procedimento de análise por síntese para determinar o sinal excitação.

4.1. *Escolha da seqüência de inovação ótima*

O procedimento usado por codificadores CELP para escolher a seqüência de inovação ótima $c_k(n)$ segue o modelo da **Fig. 4**. Esta seqüência é selecionada de um dicionário contendo $M=2^m$ seqüências candidatas de comprimento N. Assim, um índice representado por uma palavra-código de m bits é suficiente para especificar a seqüência de inovação. Valores típicos para m e N são 10 e 40 respectivamente [16]-[18].

Cada seqüência $c_k(n)$, $k=1, ..., M$ do dicionário é multiplicada por um fator de ganho σ_k, que efetivamente aumenta o tamanho do dicionário. Em seguida, a seqüência resultante

$$u_k(n) = \sigma_k c_k(n)$$

é filtrada por um filtro recursivo, que usa um preditor com retardo longo $P_L(z)$, que será discutido em detalhe na próxima seção, de modo a introduzir periodicidade. A saída $v(n)$ deste filtro é a excitação ao filtro de síntese LPC, $H(z)$.

O fator de ganho σ_k é determinado para cada seqüência $c_k(n)$ através da minimização do valor médio quadrático do erro de síntese ponderado

$$\varepsilon_k = \sum_{n=1}^{N} [\, s_W(n) - \hat{s}_W(n) \,]^2 \qquad (22)$$

onde

$$\hat{s}_W(n) = \sum_{i=1}^{N} \sigma_k \, c_k(i) \, g(n-i)$$

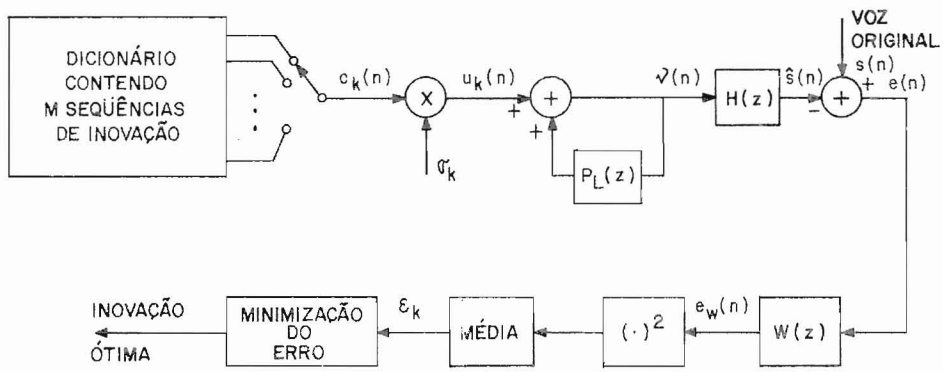

Figura 4. Procedimento básico para escolha da inovação ótima em codificadores CELP

e g(n) é a resposta impulsional do filtro cuja função de transferência é

$$G(z) = \frac{1}{1 - P_L(z)} \, . \, H(z) \, . \, W(z)$$

Fazendo

$$\frac{\partial \varepsilon_k}{\partial \sigma_k} = 0$$

obtém-se facilmente a seguinte expressão para o fator de ganho ótimo:

$$\sigma_k = \frac{\sum_{n=1}^{N} s_W(n) \sum_{i=1}^{N} c_k(i) g(n-i)}{\sum_{n=1}^{N} \left[\sum_{i=1}^{N} c_k(i) g(n-i) \right]^2}$$

Usando-se este valor de σ_k em (22) resulta que:

$$\varepsilon_k = \sum_{n=1}^{N} s_W^2(n) - \frac{\left[\sum_{n=1}^{N} s_W(n) \sum_{i=1}^{N} c_k(i) g(n-i) \right]^2}{\sum_{n=1}^{N} \left[\sum_{i=1}^{N} c_k(i) g(n-i) \right]^2} \qquad (23)$$

A escolha da seqüência de inovação ótima é feita através da otimização de ε_k ou, equivalente, da maximização do segundo termo do lado direito de (23). A alta complexidade computacional da busca da inovação ótima pode ser reduzida de diversas formas como será visto em seguida.

4.2 *Estruturas do dicionário e técnicas de redução de complexidade*

Existem várias maneiras de construir o dicionário a ser usado em codificadores CELP. A maneira mais simples é gerar as amostras que compõem as seqüências de inovação a partir de um processo Gaussiano com variância unitária [16]. A justificativa para isso é que a seqüência de excitação busca modelar a seqüência erro de predição, a qual, por sua vez, pode ser aproximada por amostras independentes com distribuição Gaussiana. Quando um dicionário desse tipo é empregado, os codificadores CELP também são chamados de codificadores LPC com *excitação estocástica*.

Uma maneira de reduzir a complexidade dos codificadores CELP consiste em limitar o número de componentes não nulas das seqüências que compõem o dicionário (vetores-código), forçando com isso que os membros do dicionário sejam vetores esparsos. As operações de filtragem podem então

ser feitas a um custo computacional mais baixo, utilizando-se algoritmos de multiplicação de vetores esparsos por matrizes [19]. Um dicionário Gaussiano que utiliza vetores esparsos e que, além disso, proporciona uma pequena melhoria de desempenho, é obtido através de uma ceifagem central de seqüências Gaussianas com variância unitária. Níveis de ceifagem usualmente empregados são ± 1,2 e ± 1,3 [18], [20].

Usando vetores esparsos obtidos a partir de deslocamentos circulares de um vetor Gaussiano com ceifagem central é possível obter relações recursivas eficientes para a saída da filtragem de $c_k(n)$ por g(n) e para sua energia [21]. Note-se que ambas são usadas na busca da seqüência ótima. A obtenção de um dicionário desse tipo, com L vetores, é feita tomando-se inicialmente um vetor Gaussiano com ceifagem central e de comprimento L, cujos elementos são

$$\tau(1), \tau(2), ..., \tau(L)$$

Define-se em seguida $\tau(j+L) = \tau(j)$. O k-ésimo vetor é obtido tomando-se como seu primeiro elemento $\tau(k)$. Seus elementos são, então,

$$\tau(k), \tau(k + 1), ..., \tau(k + N - 1)$$

Utilizando esse tipo de dicionário e definindo

$$y_k(n) = \sum_{i=1}^{N} c_k(i)g(n - i)$$

É possível mostrar as relações recursivas mencionadas acima. Elas são expressas por [21]

$$y_{k+1}(n) = \begin{cases} y_k(n + 1) - \tau(k)g(n) & , \ 1 \le n \le N - 1 \ \text{se} \ \tau(k) \neq 0 \\ y_k(n + 1) & , \ 1 \le n \le N - 1 \ \text{se} \ \tau(k) = 0 \\ \sum_{i=1}^{N} \tau(i + k)g(N - i) & , \ n = N \end{cases}$$

$$\sum_{n=1}^{N} y_{k+1}^2(n) = \begin{cases} \sum_{n=1}^{N-1} y_{k+1}^2(n) + y_{k+1}^2(N) & , \text{ se } \tau(k) \neq 0 \\ \sum_{n=2}^{N} y_k^2(n) + y_{k+1}^2(N) & , \text{ se } \tau(k) = 0 \end{cases}$$

Essas relações são eficientes porque grande parte dos vetores de $\tau(j)$ são nulos. A percentagem desses valores que são nulos pode ser obtida da probabilidade de $| \tau(j) |$ ser inferior ao limiar de ceifagem. Como $\tau(j)$ é Gaussiana essa probabilidade é facilmente calculada com o auxílio de tabelas.

Uma outra maneira de gerar o dicionário em codificadores CELP é através do emprego de algoritmos de quantização vetorial que utilizam como seqüência de treinamento os resíduos ou erros de predição obtidos do sinal de voz [22], [23]. É também possível utilizar mais de um dicionário, cada um associado a uma característica espectral diferente [19], [23]. A escolha da classe espectral mais apropriada para um segmento de voz particular é baseada nos valores dos parâmetros LPC daquele segmento. O emprego de classificação espectral permite uma significativa redução de complexidade, uma vez que o tamanho do dicionário associado a cada padrão é usualmente muito menor que o do dicionário único.

A redução de complexidade pode ainda ser obtida por uma seleção inicial de um sub-grupo de N_c seqüências mais prováveis. Nesse procedimento inicial não são usadas algumas operações de filtragem, como por exemplo a ponderação [22], [24] ou o preditor com retardo longo [24]. Em seguida, as N_c inovações escolhidas passam pelo sistema completo, sendo então efetuada uma busca exaustiva sobre este sub-grupo de inovações.

A tentativa de diminuir o tempo necessário para realizar a busca da seqüência de inovação ótima levou à investigação de dicionários mais estruturados. Em [25], por exemplo, foi proposto o uso de seqüências com apenas dois elementos não-nulos, a saber +1 e 1. para usar um dicionário desse tipo o sinal de voz é pré-processado de modo a reduzir o número de componentes relevantes do bloco de amostras a ser representado. Dicionários possuindo uma estrutura algébrica também são capazes de aumentar a velocidade computacional dos codificadores CELP [26]-[28].

Uma outra técnica eficiente para caracterização da excitação é através de um modelo completo [21], em que a excitação u(n) é expressa por

$$u(n) = \tau_1 u_1(n) + \tau_2 u_2(n) \qquad (24)$$

onde $u_1(n)$ e $u_2(n)$ são as componentes "pulsada" e "ruidosa" da excitação e τ_1 e τ_2 são os ganhos respectivos. O dicionário para representar $u_1(n)$ é obtido de amostras Gaussianas com ceifagem central e infinita, resultando em amplitude +1, 0 e 1, utilizando o método de deslocamentos circulares descrito anteriormente. O dicionário para $u_2(n)$, por outro lado, é obtido a partir de amostras Gaussianas com ceifagem infinita, resultando em amplitudes +1 e -1. Um procedimento de busca sub-ótimo, porém útil, consiste em escolher inicialmente a componente "pulsada" $u_1(n)$. Em seguida, essa componente é combinada com as componentes "ruidosas" e τ_1, τ_2 e $u_2(n)$ são determinados de modo a minimizar o erro médio quadrático de síntese ponderado. Esse modelo parece ser útil principalmente em segmentos sonoros onde o filtro de predição com retardo longo não consegue reproduzir a componente pulsada do resíduo LPC [21].

O padrão adotado, em final de 1989, pela Telecommunications Industry Association (TIA) dos Estados Unidos, para sistemas móveis celulares digitais, emprega um método semelhante à técnica de Lin mas que permite uma maior simplificação do método de busca da excitação ótima [29]. O codificador escolhido, denominado de VSELP, faz uso também de dois dicionários sendo que as seqüências que os compõem são formadas através da combinação linear, com coeficientes binários, de m vetores (seqüências) base. A k-ésima seqüência do i-ésimo dicionário (i=1 ou 2) é dada por:

$$u_k^{(i)}(n) = \sum_{j=1}^{m} \theta_j v_{ij}(n)$$

com $\theta_j = \pm 1$. Portanto, cada dicionário é composto de 2^m seqüências. A função excitação total é então obtida conforme a eq.(24). Note-se que, devido à estrutura das m seqüências, para cada bloco é necessário realizar filtragem apenas nas seqüências de base. É ainda possível mostrar [29] que o cálculo dos termos necessários à busca da melhor função excitação pode ser feito de maneira recursiva, para cada dicionário, a partir de qualquer um de seus vetores. Por fim como o negativo de cada seqüência é também uma seqüência pertencente ao dicionário e os termos a serem avaliados durante

o processo de busca são função do valor absoluto (vide eq. 23) das amostras das seqüências, faz-se necessário realizar as operações apenas para metade das seqüências e após a escolha da melhor decidir entre ela e o seu simétrico. Os dicionários para o VSELP foram obtidos através de treinamento com sinais de voz a partir de dicionários iniciais com amostras gaussianas.

Uma anáiise comparativa dos diferentes métodos para redução de complexidade em algoritmos CELP é apresentada em [30].

5. PREDIÇÃO COM RETARDO LONGO

Tanto no caso do sistema multi-pulso quanto no da técnica CELP é possível melhorar o desempenho do codificador com a utilização de um preditor com retardo longo. A função deste tipo de preditor é explorar a alta correlação entre amostras separadas de um período fundamental, o que permite uma redução substancial do número de pulsos da excitação em um sistema multipulso ou, semelhantemente, uma redução do tamanho do dicionário em um sistema CELP. Essa redução se justifica já que com o emprego do preditor de retardo longo a função excitação deverá aproximar apenas a parte descorrelatada do sinal e que portanto não pode ser estimada [9], [31], [32].

A **Fig. 5** mostra o filtro com preditor com retardo longo utilizado. O preditor com retardo longo $P_L(z)$ é tipicamente de 1ª ordem, embora algumas vezes seja empregada uma ordem superior (geralmente igual a 3). O preditor de 1ª ordem é especificado por um coeficiente B e um retardo L. Nesse caso, a excitação ao filtro de síntese H(z) é dada por

$$v(n) = B\, v(n - L) + u(n)$$

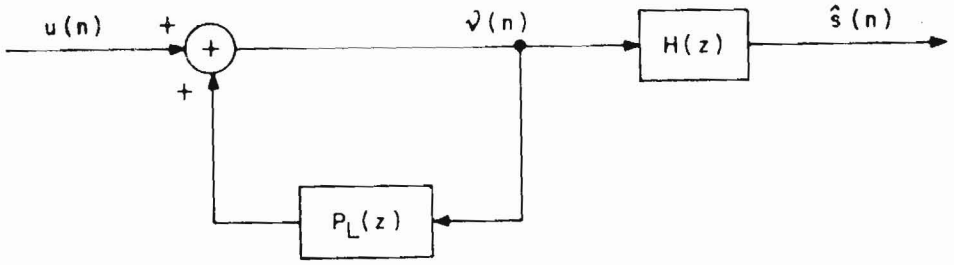

Figura 5. O uso de predição com retardo longo sobre a excitação.

Desta expressão observa-se que $v(n)$ é o resultado da composição de dois termos. Um deles, $B\ v(n-L)$, explora a estrutura quase periódica dos sons sonoros e representa uma predição de $v(n)$ com um retardo L que é igual a uma estimativa do período fundamental. O outro termo, u(n), necessita apenas representar, como se mencionou anteriormente, a parte descorrelatada da excitação. Formas de onda típicas de $v(n)$ e u(n) em um codificador LPC com excitação multipulso, que utiliza preditor com retardo longo, podem ser visualizados na **Fig. 6**, na qual é mostrada ainda a forma de onda do sinal sintetizado.

Os parâmetros do preditor (B e L, no caso de predição de 1ª ordem) podem ser determinados utilizando tanto um procedimento em malha aberta como um procedimento em malha fechada. Ambos serão descritos em seguida.

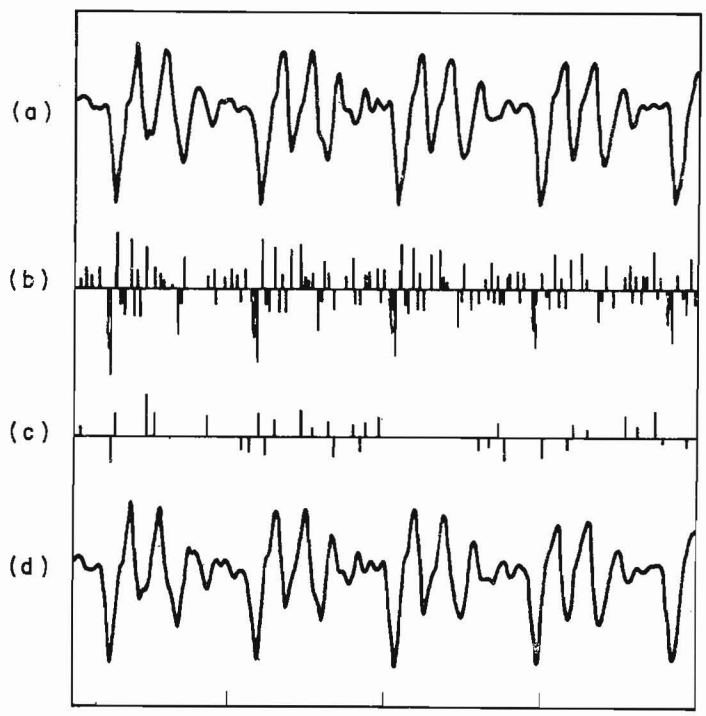

Figura 6. Formas de onda em diversos pontos de um codificador LPC com excitação multipulso que utiliza predição com retardo longo: a) voz original, b) excitação $v(n)$ na entrada do filtro de síntese, c) excitação descorrelatada u(n), d) voz sintetizada.

5.1. *Análise em malha aberta do preditor com retardo longo*

No procedimento em malha aberta os parâmetros B e L são calculados a partir do sinal de voz original, s(n), ou do resíduo $e_p(n)$. Um método usual consiste em igualar L ao retardo para o qual o coeficiente de autocorrelação normalizado do sinal (s ou e_p) é máximo. O ganho B é simplesmente o valor desse coeficiente [33]. Para um intervalo de análise desses parâmetros com duração de M amostras, o coeficiente de autocorrelação normalizado ρ(k) é definido para um retardo k, por

$$\rho(k) = \frac{\sum_{n=1}^{M-k} v(n)\, v(n+k)}{\left\{ \sum_{n=1}^{M-k} v^2(n) \sum_{n=k+1}^{M} v^2(n) \right\}^{1/2}}$$

onde v(n) é normalmente o resíduo $e_p(n)$, de modo a reduzir a influência das formantes, embora seja possível também usar o sinal de voz s(n). Assim L é o valor de k, $L_1 \leq k \leq L_2$,para o qual ρ(k) é máximo e $B = \rho(L)$. L_1 e L_2 são, respectivamente, os valores mínimo e máximo permissíveis ao período fundamental. Se B for menor que um certo valor de limiar, determinado experimentalmente, o som é considerado ser de natureza não sonora e, evidentemente, B é fixado em zero. Alguns testes adicionais também podem ser incluídos de modo a diminuir os erros na estimativa do período fundamental [34].

5.2. *Análise em Malha Fechada do Preditor com Retardo Longo*

Antes de apresentar a análise em malha fechada é importante considerar o diagrama em blocos do procedimento de análise por síntese mais detalhado mostrado na **Fig. 7**.

Neste diagrama, o filtro de ponderação W(z) foi combinado com o filtro de síntese H(z) resultando em $1/A(\gamma^{-1} z)$ onde A(z) = 1/H(z) é o filtro inverso. Além disso, observa-se que apenas uma parcela da saída do sistema

$$G(z) = \frac{1}{1 - P_L(Z)} \cdot \frac{1}{A(\gamma^{-1} z)}$$

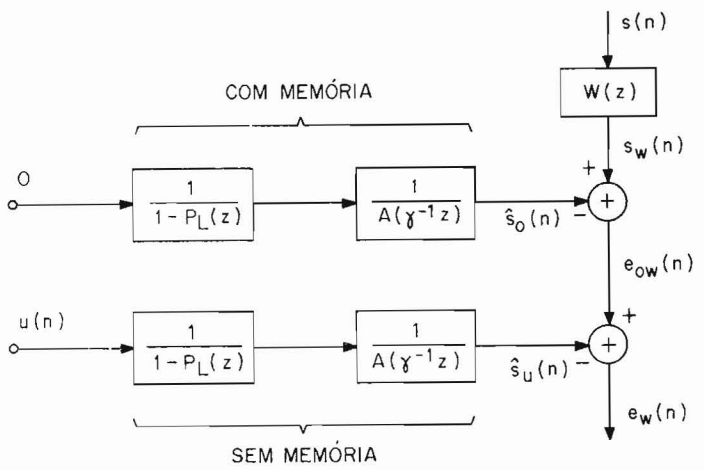

Figura 7. Diagrama em blocos detalhado do procedimento de análise por síntese.

é determinada (e portanto dependente) da excitação u(n). Uma outra parcela da saída de G(z) é obtida a partir da memória do filtro para a saída atual considerando-se entrada nula. Desse modo, é possível separar as duas parcelas através dos dois ramos mostrados na **Fig. 7**. No ramo superior $s_0(n)$ representada a saída de G(z) para uma entrada nula, levando-se em conta apenas a memória. Já no ramo inferior, $\hat{s}_u(n)$ representa a saída do mesmo filtro para uma entrada u(n) e com estado inicial zero [17], [35].

A rigor, na análise em malha fechada do preditor com retardo longo, L e B deveriam ser calculados juntamente com o sinal excitação u(n) de modo a minimizar a energia do erro ponderado $e_w(n)$. Entretanto, tal procedimento levaria a uma complexidade excessiva. Uma possibilidade sub-ótima é fazer a análise em duas etapas: na primeira os parâmetros L e B são calculados minimizando-se a energia de

$$e_{ow}(n) = s_w(n) - \hat{s}_0(n)$$

e, em seguida, a excitação u(n) é determinada de modo a minimizar a energia de $e_w(n)$.

Consideremos agora a determinação dos valores de B e L que minimizam a energia de $e_{ow}(n)$. Para isso, é de interesse dividir o ramo superior da **Fig. 7** em dois outros. Um deles leva em conta a saída $\hat{s}_{os}(n)$ do filtro $1/A(\gamma^{-1} z)$, para uma entrada nula (considerando-se apenas o efeito da memória). O outro leva em conta apenas a entrada $v_o(n)$, que representa o sinal de saída de $1/(1-P_L(z))$, para uma entrada nula, resultante apenas da contribuição da memória do preditor de retardo longo. O diagrama em blocos que representa essa divisão do ramo superior da **Fig. 7** é mostrado na **Fig. 8**.

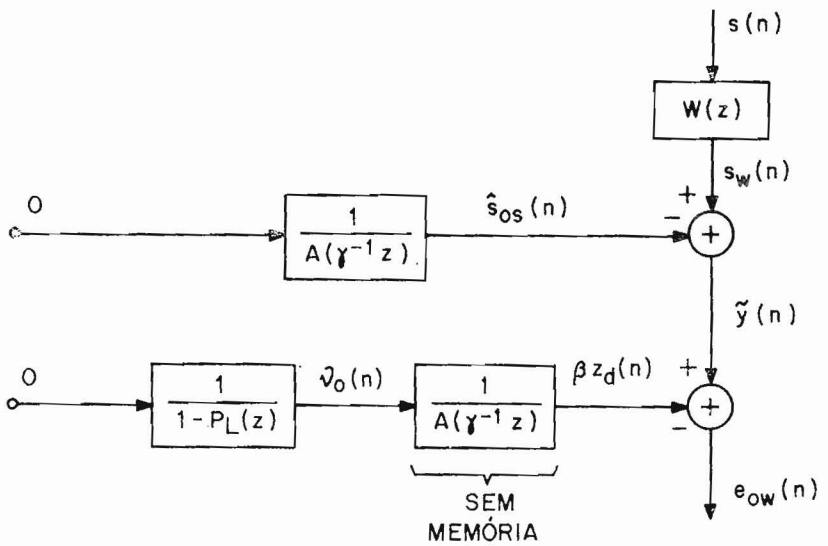

Figura 8. Diagrama em blocos utilizado para análise em malha fechada do preditor com retardo longo.

Deseja-se, então, minimizar

$$\varepsilon_{ow} = \sum_{n=1}^{N} e_{ow}^2(n) \tag{25}$$

onde

$$e_{ow}(n) = \tilde{y}(n) - \beta z_d(n)$$

$$\tilde{y}(n) = s_w(n) - \hat{s}_{os}(n)$$

onde $\beta z_d(n)$ é a saída de $1/A(\gamma^{-1}z)$ para a entrada $v_0(n)$. Derivando (25) em relação a β e igualando a zero, resulta

$$\beta = \frac{\sum_{n=1}^{N} \tilde{y}(n)\, z_d(n)}{\sum_{n=1}^{N} z_d^2(n)} \qquad (26)$$

Substituindo (26) em (25) resulta que o valor ótimo de L é o que maximiza

$$\gamma_L = \frac{\left[\sum_{n=1}^{N} \tilde{y}(n)\, z_d(n)\right]^2}{\sum_{n=1}^{N} z_d^2(n)}$$

É importante observar que $v_0(n) = B\, v_0(n-L)$ está bem definido para $n \leq L$ e representa a saída de $1/(1-P_L(z))$ quando a entrada é a excitação ótima de blocos anteriores. Para $n > L$ pode-se fazer a aproximação de $v_0(n)$ pelo resíduo LPC (saída de $A(\gamma^{-1}z)$ quando a entrada é s(n)).

A análise em malha fechada para cálculo dos parâmetros do preditor com retardo longo permite que o ganho do preditor assim projetado seja equivalente ao ganho de um preditor de terceira ordem projetado em malha aberta. Ainda assim o ideal seria utilizar um preditor de ordem superior com parâmetros extraídos em malha fechada. A restrição que se coloca é o número de bits necessários para quantização dos coeficientes adicionais. Por outro lado, é sabido que o ganho do preditor do retardo longo é extremamente sensível à precisão na estimativa do retardo L. Portanto, um maior ganho poderia ser obtido com um preditor de 1ª ordem se a resolução na determinação de L fosse melhorada. Esta linha de raciocínio foi explorada em [36] com a proposição de um preditor com retardo fracionário, ou seja, a forma genérica para L passa a ser:

$$L = I + \frac{\iota}{D}\,, \quad \iota = 0,1\,,\ldots,\, D-1 \qquad (27)$$

Para a determinação dos parâmetros l e ι é necessário aumentar, através de um procedimento de interpolação, por um fator D a freqüência de amostragem do sinal do qual está sendo feita a extração dos parâmetros do preditor. De (27) pode ser verificado que após a interpolação o número de amostras correspondentes ao retardo L passa a ser um número inteiro, lD + ι. Foi verificado [36] que um preditor de 1ª ordem com D=4, e que portanto requer dois bits adicionais para representação de ι, apresenta um desempenho comparável a um preditor de 3ª ordem, que necessita de 4 a 6 bits para representação dos coeficientes adicionais.

Vale ser salientado que o preditor com retardo longo pode ser implementado, no caso dos sistemas CELP, na forma de um dicionário adicional [29], [36], [48] em que cada seqüência seria uma versão retardada por um dado valor de l e ι da seqüência de excitação escolhida pelo codificador em blocos anteriores. Com esse tipo de implementação os períodos mais curtos, em geral correspondentes à voz feminima, podem ser privilegiados com uma melhor resolução. Por exemplo, considerando um dicionário com 256 seqüências (8 bits), os períodos entre 20 e 41 amostras podem ser representados com D=4, enquanto que, para os períodos entre 42 e 102 amostras pode ser usada uma resolução com D=2 e finalmente para os períodos maiores (freqüências mais baixas) a representação seria feita apenas com L inteiro (D=1). Estratégias como estas melhoram substancialmente o comportamento de sistemas CELP no seu ponto mais fraco que é na digitalização da voz feminina.

Um último ponto a ressaltar é que o uso de melhor preditor com retardo longo permite reduzir a importância dos demais dicionários que compõem a excitação. Em verdade esta observação torna possível melhorar ainda mais o sinal reconstruído, já que as seqüências-excitação que procuram modelar a parte descorrelatada da voz tendem a dar uma característica ruidosa àquele sinal. Esse problema pode ser minorado reduzindo o valor do ganho para a seqüência excitação sempre que o preditor de retardo longo estiver funcionando bem. Um algoritmo para esse fim é apresentado em [37].

6. ESTRUTURAS DE ANÁLISE REALIMENTADAS EM CODIFICADORES CELP

Os desenvolvimentos obtidos com os codificadores CELP vistos até aqui consistem de esquemas que operam com um retardo de pelo menos duas

vezes a duração do intervalo de análise, devido principalmente à utilização de uma estrutura de análise dos parâmetros do filtro não realimentada. Os retardos envolvidos são tipicamente da ordem de 40 a 60ms, sendo que com retardos desta ordem é possível obter boa qualidade do sinal de voz reconstruído e taxas de bits abaixo de 8 kbit/s. Em um desses desenvolvimentos, por exemplo [38], foi possível obter taxas tão baixas quanto 3,6 kbit/s com um retardo de codificação de 45ms. Para isso, foi empregada uma estrutura em que segmentos do sinal de voz são classificados em cinco categorias, sendo cada uma das quais codificada com um esquema diferente tanto em termos dos parâmetros dos filtros, como do tipo e dimensão da excitação. Vê-se, então, que embora seja possível obter um bom desempenho a baixas taxas, esses codificadores apresentam um sério inconveniente para muitas aplicações, que é o alto retardo envolvido.

Para a taxa de 16 kbit/s foram propostos esquemas de codificação CELP que operam com retardos da ordem de 2ms, além de apresentar desempenho equivalente ao da recomendação G.721 do CCITT para 32 kbit/s e boa robustez (pouca sensibilidade) a erros no canal [39]-[41]. Retardos desta ordem, entretanto, só puderam ser obtidos através da utilização de estruturas realimentadas para análise dos parâmetros dos filtros de síntese. Nessas estruturas os parâmetros são determinados ou a partir do sinal sintetizado ou, recursivamente, a partir do vetor excitação, sendo necessária a transmissão apenas de um código relativo ao vetor excitação (em geral de comprimento muito pequeno). Neste caso sobram, portanto, todos os bits para excitação, de forma que as deficiências que possam advir da estrutura realimentada e do pequeno comprimento do vetor excitação são de certa forma compensadas pela taxa de bits disponível de 16 kbit/s, que é relativamente alta. Um problema importante que poderá surgir refere-se à tentativa de utilizar métodos de codificação deste tipo a baixas taxas (abaixo de 8 kbit/s), uma vez que o número de bits disponíveis pode não ser suficiente para compensar as deficiências mencionadas.

As estruturas de análise realimentadas podem ser de dois tipos: em bloco ou recursivas.

Nas estruturas em bloco os parâmetros dos filtros são determinados a partir de um bloco de amostras de sinal sintetizado. Esses parâmetros são então usados durante o próximo segmento de voz a ser codificado. Em uma implementação desse tipo [39], o bloco de amostras do sinal sintetizado, a partir do qual é feita a análise, tem duração de 20ms, enquanto que o

segmento de voz que utiliza os parâmetros resultantes dessa análise tem duração de 5ms. Nessa implementação o preditor com retardo longo não foi utilizado. Entretanto, para evitar a deterioração que isso acarretaria na qualidade da voz feminina, a ordem do preditor com retardo curto foi aumentada de 10 para 50. Isso tem como objetivo explorar a redundância associada com a periodicidade dos sons sonoros para a voz feminina. A ordem 50 foi obtida experimentalmente através da determinação do ganho de predição em função da ordem do preditor. Este é atualmente o codificador mais bem posicionado para ser recomendado pelo CCITT como padrão para a taxa de 16 kbit/s.

Nas estruturas recursivas a determinação dos parâmetros é mais flexível e é feita amostra por amostra. Sua atualização é baseada em considerações de complexidade [40], [42]. Uma possibilidade consiste em adaptar os coeficientes do preditor com retardo curto usando um algoritmo de gradiente similar ao empregado na recomendação G.721 do CCITT para 32 kbit/s. Nesse caso, são utilizados 6 zeros e 2 pólos. Já na determinação dos parâmetros do preditor com retardo longo é empregada uma estrutura realimentada híbrida que é composta de uma estrutura em bloco e uma estrutura recursiva [43].

7. QUANTIZAÇÃO DA EXCITAÇÃO E DOS PARÂMETROS DOS FILTROS EM CODIFICADORES CELP

Em codificadores CELP que não empregam as estruturas realimentadas vistas na seção anterior, os parâmetros a serem quantizados são os ganhos associados à excitação e ao preditor com retardo longo, os coeficientes do filtro de retardo curto, também conhecidos por coeficientes LPC, o índice k associado à excitação escolhida e o retardo L do preditor com retardo longo.

Os parâmetros mencionados são usualmente discretizados através de quantizadores escalares e não-uniformes. A forma de representação dos coeficientes LPC varia, podendo ser utilizados o log da razão de área (LAR), coeficientes de reflexão (e sua versão arcseno) (RC) e pares de linhas espectrais (LSP). Cada uma delas apresenta vantagens e desvantagens discutidas em diversos trabalhos na literatura. O uso de quantizadores vetoriais para discretização dos coeficientes do preditor com retardo curto poderão trazer grande economia em termos de número de bits necessários para uma dada qualidade. Resultados recentes [44] indicam que 24 bits seriam suficientes para representar um filtro de 10ª ordem. No entanto, a

pouca robustez de quantizadores vetoriais a erros no canal tem restringido o seu uso na prática.

A Tabela 1 [45] fornece a distribuição dos bits para codificadores já adotados como padrão ou em vias de padronização. O sistema FS1016 foi escolhido como padrão pelo Departamento de Defesa dos Estados Unidos para a taxa de 4,8 kbit/s.

Um outro aspecto que vale ressaltar, relacionado à codificação dos parâmetros da excitação e dos parâmetros dos filtros em codificadores CELP é o benefício que poderia ser obtido de uma alocação de bits adaptativa [47]. Embora pequena, a melhoria de desempenho que se obtém com essa técnica é significativa. Essa melhoria resulta do fato de que os processos de produção e percepção da voz são altamente não estacionários, de forma que é importante se ter flexibilidade na alocação da taxa de bits para os parâmetros da envoltória espectral e da excitação.

Parâmetro	Codificador			
	FS1016 [46]	IS54 [29]	GSM [15]	LD-CELP [40]
Taxa de Bits (bits/s)	4.600	7.950	13.000	16.000
Preditor de Curta Duração				
Ordem:	10	10	8	50
Representação:	LSP	RC	LAR	LPC
Número de bits:	34	38	36	–
Atualização de parâmetro: (ms)	30	20	20	2.5
Preditor de Retardo Longo				
Número de bits (ganho):	5	4	2	–
Número de bits (retardo):	8/6/8/6	7	7	–
Atualização dos parâmetros: (ms)	7,5	5	5	–
Excitação				
Número de bits ganho:	4	5,25	45*	2
Índice e sinal:	10	14	2	8
Tamanho do bloco (ms):	7,5	5	5	0,625

(*) inclui amplitudes dos pulsos.

Tabela 1. Comparação dos parâmetros e distribuição de bits para alguns codificadores de voz [45]

8. DESEMPENHO

A razão sinal-ruído segmentada média (RSS) de um codificador CELP de *referência* (sem quantização dos parâmetros) usado em [18] é de cerca de 14dB para uma frase composta predominantemente por vogais e semi-vogais (sons sonoros). Entretanto, seus valores locais apresentam uma grande variação, podendo ir desde 5dB até 25dB. Para codificadores a uma taxa de 8 kbit/s o valor médio de desempenho é de aproximadamente 13dB, apresentando também amplas variações locais como mostra a **Fig. 9** [23]. A **Fig. 10** ilustra uma comparação entre os sinais de entrada e saída para um codificador a 8 kbit/s.

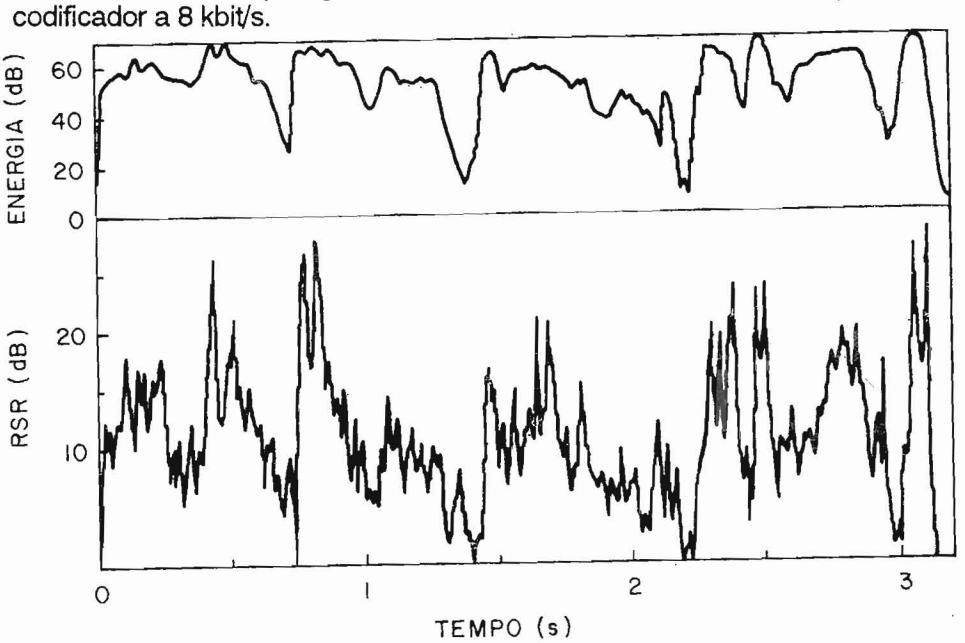

Figura 9. Energia local do sinal de voz e razão sinal-ruído segmentada para um codificador CELP a 8 kbit/s [23].

A taxas em torno de 4 kbit/s, a RSS que se obtém com codificadores CELP é de cerca de 11dB. Porém, com o estado atual da tecnologia, a qualidade da voz não é mais suficientemente natural, particularmente durante transições rápidas do sinal de voz e em regiões de alta freqüência.

Tanto o codificador CELP como o MP-LPC apresentam uma boa qualidade subjetiva para a taxa de 8 kbit/s. Para taxas mais baixas, porém, o MP-LPC

tem seu desempenho consideravelmente degradado. Por outro lado, o codificador CELP tem se mostrado o mais indicado para operação na faixa entre 4,0 e 9,6 kbit/s.

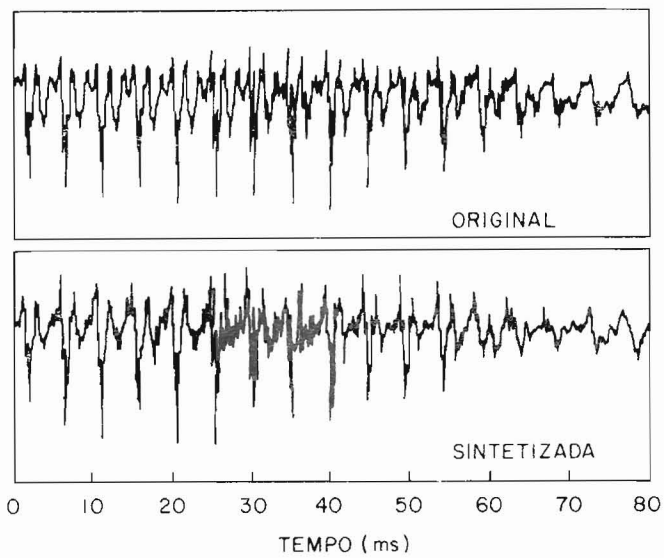

Figura 10. Comparação de formas de onda para um codificador CELP a 8 kbit/s [23].

A Tabela 2 apresenta valores de MOS (Indicador de Média de Opiniões), para os mesmos quatro sistemas mencionados na Tabela 1, tendo como referência que o valor de MOS para um sistema μ –log PCM a 64 kbit/s é de 4,2.

Codificador	FS1016	IS54	GSM	LD-CELP
Desempenho (MOS)	3,5	4,0	3,9	4,1

Tabela 2. Desempenho em termos de MOS para os codificadores da Tabela 1 [45]

O LD-CELP fornece um desempenho praticamente idêntico ao codificador ADPCM padrão para 32 kbit/s (G.721).

9. CONCLUSÕES

Neste artigo foi apresentada uma visão unificada da teoria e aspectos mais importantes relativos a três diferentes métodos para geração da função excitação de um filtro de síntese LPC, quais sejam: multi-pulso, pulsos regularmente espaçados e excitação vetorial (dicionário de códigos). Todos esses métodos têm por base para determinação da seqüência de excitação (quase) ótima um critério de distorção ponderado que depende da envoltória espectral do bloco de amostras de voz a ser quantizado. Especificamente, é permitido que o espectro do ruído de quantização tenha um nível de energia maior na região das formantes.

Ainda que esses modelos possam todos ser enquadrados na classe de procedimentos de análise-por-síntese, a complexidade na implementação e a taxa para qual eles são mais indicados variam consideravelmente.

A complexidade computacional exigida por uma busca exaustiva da excitação ótima em sistemas multipulso e CELP conduziu à proliferação de estratégias para redução dessa complexidade, fazendo com que uma implementação em tempo real fosse possível. Em verdade, com a tecnologia atual todos os codificadores mencionados na Tabela 1 podem hoje ser implementadas empregando uma pastilha DSP para o codificador (onde a complexidade está concentrada) e outra para o decodificador.

Afortunadamente, essas técnicas recentes para redução do custo computacional podem também melhorar o desempenho. Esse é o caso, por exemplo, do uso de vetores esparsos em sistemas CELP, ao invés de seqüências estocásticas densamente preenchidas.

O emprego do preditor com retardo longo, que realiza uma busca de uma componente da seqüência excitação a partir de excitações utilizadas para blocos anteriores, garante um ganho adicional na qualidade subjetiva do sinal sintetizado.

Com relação a desenvolvimentos relativamente recentes para taxas próximas a 2 kbit/s, os codificadores CELP se mostraram uma alternativa promissora para superar a qualidade sintética da voz nessas taxas.

Embora avanços notáveis tenham sido conseguidos para essa classe promissora de codificadores, com aplicações diversos como comunicações móveis, armazenamento de voz e transmissão digital de voz em canais faixa estreita, vários problemas ainda não foram resolvidos. Eles se referem à qualidade de voz percebida pelo usuário e dependem de muitos fatores, inclusive de retardo de codificação.

Inicialmente, é necessário um esforço para superar as deficiências do modelo da produção da voz, explorando também os limites da audição humana definindo, assim, critérios de erro mais significativos de ponto de vista da percepção auditiva. Desta forma talvez seja possível obter uma reprodução transparente da voz em taxas baixas.

Apesar do continuado avanço tecnológico dos processadores específicos DSP, haverá dificuldades de implementação de um codificador CELP que usa dicionários muito grandes e quantização vetorial dos coeficientes LPC com um alto número de bits. Portanto algoritmos mais eficientes deverão ser desenvolvidos para que as implementações apresentem desempenho comparável aos chamados codificadores de referência que são executados "off-line".

Ainda com relação à qualidade de voz sintetizada parece ser possível uma melhoria através de um maior refinamento do preditor de retardo longo, conduzindo a um contorno do período fundamental com variação suave ao longo dos diversos segmentos de voz.

Uma quarta área de pesquisa é a análise de codificadores com baixo retardo e alta qualidade para taxas inferiores a 10 kbit/s. Como mencionado na seção 6 isto já foi atingido com sucesso para a taxa de 16 kbit/s. No entanto, para taxas menores o desafio continua, já que tamanhos de blocos mais longos propiciam o emprego de um número menor de bits e uma melhor qualidade.

Finalmente, os codificadores CELP e multipulso tendem a ser vulneráveis a erros no canal, em especial aqueles que operam em taxas baixas e portanto devem fazer uso de quantizadores vetoriais e preditor de retardo longo. Para que os mesmos possam ser utilizados em canais reais eles devem ser tornados intrinsicamente mais robustos, bem como técnicas apropriadas para correção e recuperação contra erros devem ser desenvolvidas.

REFERÊNCIAS

[1] ATAL, B.S. e HANAUER, S.L., "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", *J. Acoust. Soc. Amer.*, vol. 50, pp. 637-655, 1971.

[2] SCHROEDER, M.R., "Linear Predictive Coding of Speech: Review and Current Directions", *IEEE Communications Magazine*, vol. 23, nº 8, pp. 54-61, Agosto 1985.

[3] MARKEL, J.D. e GRAY, Jr. A.H., "Linear Prediction of Speech", *Springer-Velag*, Berlin, 1976.

[4] ATAL, B.S. e REMDE, J.R., "A New Model of LPC Excitation for Producing Natural – Souding Speech at Low Bit Rates", Proc. *IEEE International Conf. on Acoustics. Speech and Signal Processing*, pp. 614-617, Paris, 1982.

[5] ATAL, B.S. e SCHROEDER, M.R., "Predictive Coding of Speech Signals and Subjective Error Criteria", *IEEE Trans. on Acoustics. Speech and Signal Processing*, vol. ASSP-27, nº 3, pp. 247-254, Junho 1979.

[6] HAWKINS, H.A. et alii, "Perceptual Weightings and Optimal Pulse Positioning in Multi-Pulse LPC Speech Coding", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, pp. 477-480, Tampa, 1985.

[7] WAKE, Y. et alii, "A Multi-Pulse LPC Speech Codec Using Digital Signal Processors", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing*, pp. 1429-1432, Tampa, 1985.

[8] OZAWA, K., ONO, S. e ARASEKI, T., "A Study on Pulse Search Algorithms for Multipulse Excited Speech Coder Realization", *IEEE Journal on Selected Areas in Communications*, vol. SAC-4, nº 1, pp. 133-141, Janeiro 1986.

[9] KROON, P. e DEPRETTERE, E.F., "Experimental Evaluation of Different Approaches to the Multi-Pulse Coder", *Proc. IEEE Intl. Conf. on Acoustics. Speech and Signal Processing*, pp. 10.4.1-10.4.4, San Diego, 1984.

[10] LEFEVRE, J.P. e PASSIEN, O., "Efficient Algorithms for Obtaining Multipulse Excitation for LPC Coders", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing*, pp. 957-960, Tampa, 1985.

[11] MONTAGNA, R. e OMOLOGO, M., "Some Results on Multipulse Linear Predictive Coding", *Proc. IEEE Intl. Conf. on Acoustics. Speech and Signal Processing*, pp. 802-806, Houston, 1986.

[12] BEROUTI, M. et alii, "Efficient Computation and Encoding of the Multipulse Excitation for LPC", *Proc. IEEE Intl. Conference on Acoustics Speech and Signal Processing*, pp. 10.1.1-10.1.4, San Diego, 1984.

[13] KROON, P. SLUYTER, R.J. e DEPRETTERE, E.F., "A Low Complexity Regular Pulse Coding Scheme with a Reduced Transmission Delay", *Proc. IEEE International Conf. on Acoustics. Speech and Signal Processing*, Tokyo, pp. 3083-3086, 1986(a).

[14] KROON, P. DEPRETTERE, E.F. e SLUYTER, R.J., "Regular-Pulse Excitation – A Novel Approach to Effective and Efficient Multipulse Coding of Speech", *IEEE Trans. on Acoustics. Speech and Signal Processing*, vol. ASSP-34, pp. 1054-1063, Outubro 1986(b).

[15] VARY, P. et al., "Speech Codec for the European Mobile Radio System", *Proc. IEEE Intl. Conf. on Acoustics. Speech and Signal Processing*, pp. 227-230, New York, N.Y., Abril 1988.

[16] SCHROEDER, M.R. e ATAL, B.S., "Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing*, pp. 937-940, Tampa, 1985.

[17] TRANCOSO, I.M. e ATAL, B.S., "Efficient Procedures for Finding the Optimum Innovation in Stochastic Coders", *Proc. IEEE International Conf. on Acoustics. Speech and Signal Processings*, Tokyo, pp. 2375-2378, 1986.

[18] KROON, P. e ATAL, B.S., "Strategies for Improving the Performance of CELP Coders at Low Bit Rates", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, New York, Abril 1988.

[19] DAVIDSON, G. e GERSHO, A., "Complexity Reduction Methods for Vector Excitation Coding", *Proc. IEEE Intl. Conf. on Acoustics. Speech and Signal Processing*, Tokyo, pp. 3055-3058, 1986.

[20] KROON, P. e DEPRETTERE, E.F., "A Class of Analysis-by-Synthesis Predictive Coders for High Quality Speech Coding at Rates Between 4.8 and 16 kbit/s", *IEEE J. on Selected Areas in Commun.*, vol. 6, nº 2, pp. 353-363, Fevereiro 1988.

[21] LIN, D., "Vector Excitation Coding Using a Composite Source Model", *Proc. European Signal Processing Conference*, 1988.

[22] COPPERI, M. e SERENO, D., " Vector Quantization and Perceptual Criteria for Low-Rate Coding of Speech", *Proc. IEEE Intl. Conf. on Acoustics. Speech and Signal Processing*, pp. 7.6.1-7.6.4, Tampa, 1985.

[23] COPPERI, M. e SERENO, D., "CELP Coding for High Quality Speech at 8 kbit/s", *Proc. IEEE Intl. Conf. on Acoustics Speech and Signal Processing*, pp. 1655-1688, Tokyo, 1986.

[24] HERNANDEZ-GOMES, L.A., et alii., "On the Behavior of Reduced Complexity Code-Excited Linear Prediction (CELP)", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, pp. 469-472, Tokyo, 1986.

[25] SCHROEDER , M.R. e SLOANE, N.J.A., "New Permutation Codes Using Hadamard Unscrambling", *IEEE Trans. on Information Theory* vol. IT-88, pp. 144-146, Janeiro 1987.

[26] ADOUL, J.P. et alii, "Fast CELP Coding Based on Algebraic Coders", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Dallas, 1987.

[27] ADOUL, J.P. e LAMBLIN, C.,, "A Comparison of Some Algebraic Structures for CELP Coding Speech", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Dallas, 1987.

[28] LAFLAMME, C. et alii,"On Reducing Computation Complexity of Codebook Search in CELP Coder Through the Use of Algebraic Codes", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing,* Albuquerque, USA, Abril 1990.

[29] GERSON, I.A. e JASIUK , M.A., "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kbit/s", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, pp. 461-464, Albuquerque, USA, Abril 1990.

[30] KLEIJN, W.B., KRASINSKI, D.J. e KETCHUM, R.H., "Fast Methods for the CELP Speech Coding Algorithms", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, nº 8, pp. 1330-1342, Agosto 1990.

[31] SINGHAL, S. e ATAL, B.S., "Improving Performance of Multipulse LPC Coders at Low Bit Rates", *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing*, pp. 1.3.1-1.3.4, San Diego, 1984.

[32] OZAWA, K. e ARASEKI, T., "High Quality Multipulse Speech Coder with Pitch Prediction", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, pp. 1689-1692, Tokyo, 1986.

[33] ATAL, B.S. e SCHROEDER, M.R., "Adaptive Predictive Coding of Speech Signals", *Bell Syst. Tech. J.*, vol. 49, pp. 1973-1986, Outubro 1970.

[34] ALCAIM, A., BOISSON DE MARCA, J.R. e JACCOUD, C.F.B., "Análise de Desempenho do Codificador APC-AB a 16 kbit/s", *Anais do 7º Simpósio Brasileiro de Telecomunicações*, pp. 198-202, Florianópolis, SC, Setembro 1989.

[35] OMOLOGO, M. e SERENO, D., "A Comparison Between Two Speech Coders at 8 kbit/s for Mobile Communications", *CSELT Technical Reports*, vol. XVI, nº 5, pp. 449-453, Agosto 1988.

[36] KROON, P. e ATAL, B.S. "Pitch Predictors with High Temporal Resolution", *Proc. IEEE 1990 Intl. Conf. on Acoustics, Speech and Signal Processing*, Albuquerque, U.S.A., Abril 1990.

[37] SHOHAM, Y., "Constrainel-Stochastic Excitation Coding of Speech at 4.8 kbit/s, em *Advances in Speech Coding*, eds. B.S. Atal, V. Cuperman e A. Gersho, Boston: Kluwer Academic Publishers, 1990.

[38] WANG, S. e GERSHO, A., "Phoneticallly-Based Vector Excitation Coding of Speech at 3.6 kbps", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Glasgow, 1989.

[39] CHEN, J.H., "A Robust Low-Delay CELP Speech Coder at 16 kbit/s", *Proc. IEEE Global Telecommunications Conference*, pp. 1237-1241, Dallas, USA, Novembro 1989.

[40] CHEN, J. H. e COX, R.V., "LD-CELP: A High Quality 16 kbit/s Speech Coder with Low Delay", *Proc. IEEE Global Telecommunications Conf.*, pp. 528-532, San Diego, U.S.A., Dezembro 1990.

[41] CUPERMAN, V. et alii, "Backward Adaptation for Low Delay Vector Excitation Coding of Speech at 16 kbit/s", *Proc. IEEE Global Telecommunications Conference*, pp. 34.2.1-5, Dallas, U.S.A., Novembro 1989.

[42] PENG, R. e CUPERMAN, V., "Low Delay Analysis-by-Synthesis Speech Coding using Lattice Predictors", *Proc. IEEE Global Telecommunications Conf.*, pp. 951-954, San Diego, U.S.A., Dezembro 1990.

[43] PETTIGREW, R. e CUPERMAN, V., "Backward Pitch Prediction for Low-Delay Speech Coding", *Proc. IEEE Global Telecommun. Conference*, pp. 34.3.1-6, Dallas, U.S.A., Novembro 1989.

[44] PALIWAL, K. e ATAL, B.S., "Efficient Vector Quantization of LPC Parameters at 24 bit/frame", *J. Acoust. Soc. Am.*, Suppl. 1, vol. 87, p. S39, 1990 (Spring).

[45] KROON, P. e ATAL, B.S., "Predictive Coding of Speech Using Analysis-by-Synthesis Techniques", em *Advances in Speech Signal Processing*, eds. S. Furui e M.M. Sondhi, New York: Marcel Dekker Inc., 1991.

[46] CAMPBELL, J.P., TREMAIN, T.E. e WELCH, V.C., "The Proposed Federal Standard 1016 4.800 bps Voice Coder: CELP", *Speech Technology*, pp. 58-64, Maio 1990.

[47] JAYANT, N.S. e CHEN, J.H., "Speech Coding with Time-Varyng Bit Allocations to Excitation and LPC Parameters", *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Glasgow, Escócia, Maio 1989.

[48] ROSE, R.C. e BARNWELL, III, T.P., "Design and Performance of an Analysis-by-Synthesis Class of Predictive Speech Coders", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, nº 9, pp. 1489-1503, Setembro 1990.

MAURIZIO COPPERI nasceu em 1950 em Torino. Recebeu o diploma em Eletrônica do Instituto Técnico G. Peano, em 1969 e o grau de Dr. "cum laude" em Matemática, em 1977, da universidade de Torino. O Dr. Copperi trabalha no CSELT (Centro de Pesquisas da Companhia Telefônica Italiana SIP) desde 1969, nas áreas de eletroacústica, codificação de voz e avaliação de qualidade subjetiva. Em 1981 trabalhou por um curto período na Universidade de Notre Dame (South Bend, Indiana) com o professor James Melsa no projeto de um codificador de voz a 9,6 Kbit/s. Foi representante do Grupo Especialista de Codificação de Voz para o algoritmo ADPCM do CCITT a 32 Kbit/s e dirigiu um projeto de pesquisa em compressão de voz a 4,8/9,6 Kbit/s para comunicações para satélites representando a Agência Espacial Européia. Em 1987 foi indicado Gerente de Programa de Aplicações de Processamento de Voz e Rádio Celular na Companhia Telefônica Italiana SIP. Ele é autor de mais de 70 artigos em revistas e conferências internacionais, é também inventor de 5 patentes internacionais relevantes para o processamento de voz. Apresentou diversos seminários sobre novas técnicas de compreensão de voz no Politéonico de Torino e na CCCNN, a escola de pós-graduação da Holding STET.

ABRAHAM ALCAIM formou-se em Engenharia Elétrica pela PUC/Rio em 1975, obteve o título de Mestre em Ciências em Engenharia Elétrica pela mesma Universidade 1977, e os títulos de D.I.C. e PhD em Engenharia Elétrica pelo Imperial College of Science and Tecnology, University of London, em 1981. Tem mais de 15 anos de experiência nas áreas de codificação digital e transmissão de formas de onda e processamento digital de sinais de voz. É autor de diversos artigos nessas áreas, publicados em revistas e conferências nacionais e internacionais. Em 1984 desempenhou atividades na área de processamento de voz, como Pesquisador Visitante no Centre Nacional d'Etudes de Télécommunications (CNET), em Lannion, França. É Professor Associado do Centro de Estudos em Telecomunicações da PUC/Rio, onde atua, desde 1976, no grupo de Sistemas de Telecomunicações. Foi presidente da Comissão de Programa do SBT/IEEE International Telecommunications Symposium (ITS'90), realizado no Rio de Janeiro em setembro de 1990 e é o presidente da Comissão de Programa do ITS'94, a ser realizado também no Rio de Janeiro, em agosto de 1994. Desde de dezembro de 1991 se encontra em licença sabática da PUC/Rio, como Pesquisador Assessor do Centro Científico Rio da IBM Brasil, onde vem desenvolvendo trabalhos na área de codificação digital de imagens.

JOSÉ ROBERTO BOISSON DE MARCA formou-se em Engenheiro Eletricista (Telecomunicações) pela PUC/Rio em 1972, tendo posteriormente recebido os graus de M.Sc. (1975) e PhD (1977) em Engenharia Elétrica, ambos pela University of Southern California, Los Angeles. Foi engenheiro de telecomunicações da EMBRATEL, professor da UNICAMP e desde 1978 é professor do Centro de Estudos de telecomunicações da PUC/Rio. Desempenhou também as atividades de Consultor científico do AT&T Bell Laboratories, Murray Hill (1986), Pesquisador Visitante da Universidade de Toronto (1981) e por duas vezes a função de Professor Visitante no Politécnico de Turim (1984 e 1989). O Prof. Boisson atuou ainda como presidente da Sociedade Brasileira de Telecomunicações (1984-1987), coordenador do Comitê Accoccoroc (CCCA) do CNPq (1987-1989), o Coordenador Central de Projetos Patrocinados pela PUC/Rio (1984-1986). É membro senior do IEEE e presidente do Comitê para a América Latina da IEEE Communications Society. Em 1990, exerceu o cargo de Diretor de Desenvolvimento Científico e Tecnológico do CNPq e foi o Coordenador Geral do SBT/IEEE International Telecommunications Symposium (ITS'90).