Um Método de Busca Rápida no Dicionário Adaptativo em Codificadores de Voz CELP

Ranniery da S. Maia, Cássio B. Ribeiro, Sergio L. Netto e Fernando Gil V. Resende Jr.

Programa de Engenharia Elétrica/COPPE, DEL/EE
Universidade Federal do Rio de Janeiro
CP 68504, Rio de Janeiro, RJ, 21945-970, BRASIL
Fone: (0xx21) 260 5010, Fax: (0xx21) 290 6626
{ranniery, cassio, sergioln, gil}@lps.ufrj.br

Resumo

Este trabalho apresenta um método de busca rápida no dicionário adaptativo em codificadores de voz por predição linear com excitação por códigos de dicionários (*code-excited linear prediction*, CELP). Resultados experimentais, que incluem testes com diferentes tipos de dicionários e tamanhos de segmentos, mostram que o sistema proposto é capaz de reduzir em torno de 30% a 50% o tempo total de codificação, quando comparado com o sistema sem busca rápida, introduzindo mínimas distorções no sinal.

1 Introdução

Atualmente, aplicações ligadas à Internet e até mesmo o grande desenvolvimento dos sistemas de telefonia móvel vêm requerendo codificadores de voz que obtenham um compromisso cada vez melhor entre qualidade e taxa de bits consumida. Uma das técnicas mais utilizadas atualmente para a codificação de voz a baixas taxas é a técnica CELP [1], cujo principal problema de implementação está na complexidade e tempo de processamento para a determinação da melhor excitação. Para amenizar estes problemas, algumas formas de acelerar esta técnica foram propostas na literatura [2], o que permitiu, em conjunto com o maior desenvolvimento dos processadores de sinal digital, a implementação da codificação CELP em tempo real. Os métodos empregados para busca rápida, na sua maioria, agilizam a busca no dicionário fixo, sendo a busca no dicionário adaptativo feita da forma convencional. Este trabalho propõe um método de busca rápida no dicionário adaptativo [3, 4], que é comparado ao mostrado em [5] para dicionários com e sem atrasos fracionários e diferentes tamanhos de segmentos de voz. A comparação é feita em termos de qualidade do sinal reconstruído e tempo de processamento.

O trabalho está organizado da seguinte forma: na Seção 2 apresentamos rapidamente a técnica de codificação CELP; na Seção 3 apresentamos o método de busca rápida proposto em [5]; na Seção 4 é explicado o método de busca rápida aqui proposto; na Seção 5 mostramos os resultados dos experimentos realizados comparando as performances dos métodos de busca rápida em termos de tempo de processamento e qualidade do sinal de voz decodificado; por fim, na Seção 6 incluímos as conclusões.

Este trabalho foi realizado com suporte financeiro de CNPq, CAPES, FAPERJ e FUJB/UFRJ.

2 A técnica CELP

A Figura 1 mostra o diagrama de blocos do codificador de voz CELP tido como padrão. A voz reconstruída $\hat{s}(n)$ é obtida ao passar o sinal de excitação $x(n)$, composto pelos dicionários adaptativo e fixo, pelo filtro de síntese $H(z) = 1/A(z)$. O filtro de ponderação do erro $W(z) = A(z)/A(z/\gamma)$ modifica o sinal de erro $e(n)$ durante o procedimento de análise-por-síntese para se obter as melhores componentes de excitação de cada dicionário.

A técnica CELP segmenta o sinal de voz em blocos, que são posteriormente divididos em sub-blocos. Os parâmetros do filtro de síntese $H(z)$ são determinados a cada bloco, enquanto que os ganhos e índices dos dicionários, que compõem a excitação $x(n)$, são determinados a cada sub-bloco. O sintetizador, ou decodificador, também faz parte do codificador e está selecionado pela linha pontilhada. Isto ocorre porque a técnica CELP utiliza o procedimento de análise-por-síntese, ou seja,

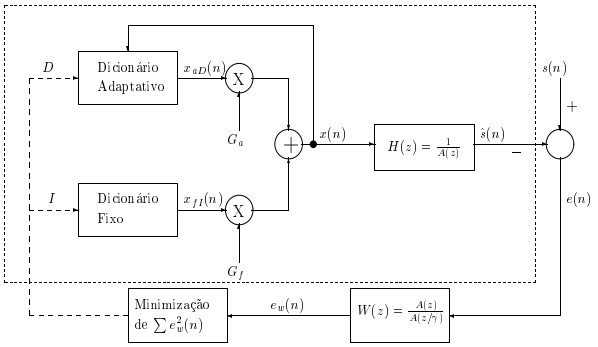

Figura 1: Estrutura do codificador CELP tida como padrão.

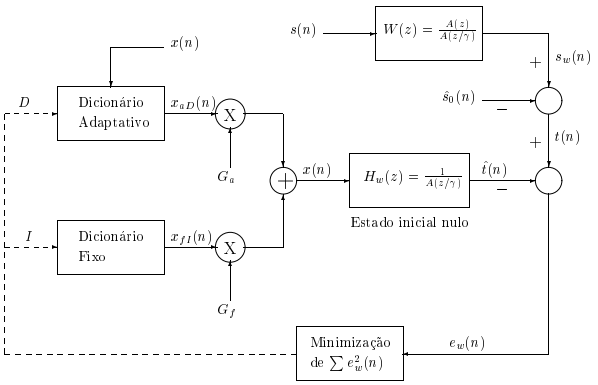

Figura 2: Estrutura do codificador CELP para implementação.

vários sub-blocos de voz são sintetizados para a escolha dos parâmetros (índices e ganhos) responsáveis pela reprodução do sub-bloco que ocasione a menor energia do erro ponderado $e_w(n)$.

O dicionário adaptativo, que substitui o filtro de *pitch* das primeiras versões da técnica CELP, é composto a partir de uma única seqüência formada com amostras das excitações em instantes passados, ou seja,

$$\{c_a(n)\} = \{x(-D_{max}), \ldots, x(-1)\}, \qquad (1)$$

onde D_{max} corresponde ao máximo atraso considerado no processo de busca da melhor excitação $x_{aD_{ot}}(n)$.

Na prática, o codificador CELP é implementado passando o filtro $W(z)$ para os dois ramos antes do somador que compõe o sinal de erro $e(n)$, e separando a resposta do filtro recursivo resultante $H_w(z) = 1/A(z/\gamma)$ em duas partes: a resposta com estado inicial zero, $\hat{t}(n)$, e a resposta à entrada zero, $\hat{s}_0(n)$. A Figura 2 mostra o diagrama do sistema levando-se em conta estas considerações.

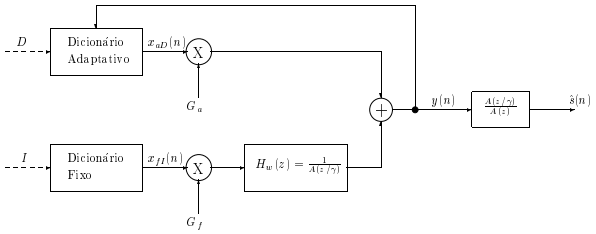

Figura 3: Novo modelo de síntese obtido a partir do modelo convencional, para o método de busca rápida no dicionário adaptativo de [5].

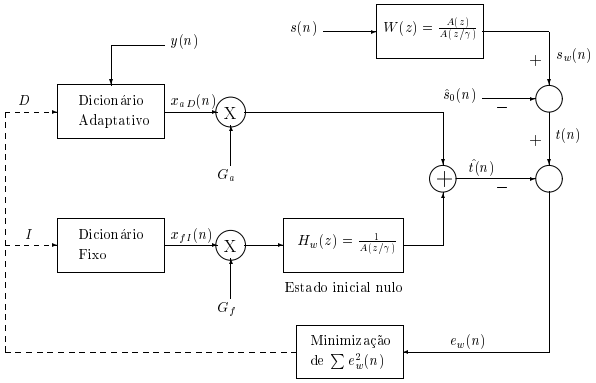

Figura 4: Estrutura prática do codificador CELP modificada para o método de busca rápida no dicionário adaptativo de [5].

3 O método de busca rápida de [5]

O método de busca no dicionário adaptativo apresentado em [5] modifica o sintetizador CELP mostrado na Figura 1 baseando-se no fato de que o ganho G_a, os parâmetros de $H(z)$ e o atraso D não variam muito em um bloco de voz sonoro. O novo modelo, que está mostrado na Figura 3, resulta de uma transformação do modelo convencional e uma posterior simplificação, conforme é mostrado em [5].

Com este novo modelo, o codificador CELP prático passa a ser aquele mostrado na Figura 4, onde o dicionário adaptativo é atualizado com a seqüência $y(n)$ obtida segundo mostra a Figura 3. Pode-se perceber claramente que neste novo modelo as seqüências candidatas $x_{aD}(n)$ não são filtradas em hipótese alguma durante todo o processo de codificação, reduzindo assim o tempo de processamento para a determinação da melhor excitação $x(n)$.

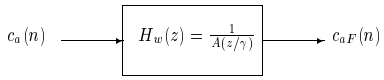

Figura 5: Geração do dicionário adaptativo filtrado usado no método proposto de busca rápida.

4 O método de busca rápida proposto

O método aqui proposto para a busca rápida no dicionário adaptativo também evita a filtragem de cada seqüência candidata $x_{aD}(n)$ por $H_w(z)$. Neste método, a seqüência $c_a(n)$, da qual é composto o dicionário adaptativo segundo mostra (1), é inteiramente filtrada por $H_w(z)$, gerando assim uma seqüência $c_{aF}(n)$ que formará o dicionário adaptativo filtrado, conforme mostra a Figura 5. A busca da melhor excitação neste novo dicionário passa então a ser feita sem que seja necessário filtrar cada seqüência candidata individualmente, ou seja, utiliza-se do modelo mostrado na Figura 4 com o dicionário filtrado no lugar do original.

Sendo determinado o melhor atraso D_{ot} no dicionário adaptativo filtrado, a leitura da melhor seqüência $x_{aD_{ot}}(n)$ é feita em $c_a(n)$ (dicionário original), que depois é atualizado com $x(n)$, ou seja, a partir do instante em que é determinado o melhor atraso D_{ot}, o procedimento de busca passa a ser exatamente igual ao método utilizado pelo sistema sem busca rápida, e que é caracterizado pelo modelo mostrado na Figura 2

Portanto, neste método existem somente duas filtragens relativas ao dicionário adaptativo para cada subbloco: a de $c_a(n)$ por $H_w(z)$ no início do processo de busca, e a da melhor seqüência $x_{aD_{ot}}(n)$, lida no dicionário adaptativo original, também por $H_w(z)$ para a determinação do ganho G_a. Este procedimento garante uma aceleração do processo de busca de $x(n)$.

5 Experimentos

Os dois métodos de busca rápida foram implementados no sistema CELP descrito em [3] sem a quantização dos ganhos G_a e G_f. As medidas objetivas utilizadas corresponderam à razão sinal-ruído segmentada perceptual (RSRSP) [3], à distância cepestral (DC) [6] e à distância de Itakura (DI) [7]. As duas últimas quantificam as diferenças no domínio da freqüência, enquanto que a RSRSP trata mais as diferenças no domínio do tempo.

As sentenças escolhidas para o experimento corresponderam a quatro sinais de voz compondo frases típicas da língua portuguesa falada no Brasil, sendo dois sinais gerados por locutores do sexo masculino (M1 e M2) e dois por locutores do sexo feminino (F1 e F2). Os sinais foram digitalizados a 8 kHz com 16 bits por amostra. Após a digitalização, foi realizada uma filtragem passa-altas para a remoção de ruídos de baixa freqüência.

Daqui em diante o método de busca rápida proposto em [5] será referenciado como Método I, enquanto que o método proposto neste trabalho será referenciado como Método II.

5.1 Experimento 1

Neste experimento, verificamos as performances do sistema CELP sem e com os métodos I e II para busca rápida no dicionário adaptativo. Os testes foram feitos em termos de medidas objetivas de qualidade e tempo de processamento em uma estação de trabalho Sun Ultra60. Foram considerados blocos de 20 ms com sub-blocos de 5 ms, e foi usado um dicionário adaptativo com atrasos fracionários cujas resoluções foram distribuídas da seguinte forma: oitavas de 20 a 55, quartas de 55 a 101 e unitárias de 101 a 146. A Tabela 1 sintetiza os resultados aqui obtidos na codificação das 4 frases consideradas.

Tabela 1: Experimento 1 - Medidas objetivas de qualidade em dB e tempo de processamento (TP) com dicionário adaptativo com atrasos fracionários e blocos de 20 ms com sub-blocos de 5 ms.

Sistema CELP sem busca rápida				
Locutor	RSRSP	DC	DI	TP (s)
M1	17,75	2,87	1,02	39,774
M2	18,43	2,93	1,04	51,068
F1	19,23	2,98	1,10	37,534
F2	17,01	3,14	1,21	45,100
Sistema CELP com o Método I				
Locutor	RSRSP	DC	DI	TP (s)
M1	16,24	3,20	1,25	26,806
M2	16,79	3,19	1,24	34,364
F1	18,05	3,01	1,10	25,547
F2	15,97	3,16	1,23	30,573
Sistema CELP com o Método II				
Locutor	RSRSP	DC	DI	TP (s)
M1	17,60	2,93	1,05	27,250
M2	18,43	2,93	1,04	35,030
F1	18,89	2,89	1,02	25,969
F2	16,77	3,07	1,17	31,329

Dos resultados apresentados, pode-se perceber que ambos os métodos de aceleração possuem tempos totais de

processamento bastante próximos, com uma ligeira vantagem do Método I, como esperado. Em termos de medidas objetivas de qualidade, o Método II introduz uma distorção sensivelmente menor que a do outro método. Aliás, para as sentenças F1 e F2 a qualidade melhorou em relação ao sistema sem busca rápida, segundo a DC e a DI. Isso pode ter ocorrido pelo fato destas medidas avaliarem a qualidade do sinal basicamente no domínio da freqüência. Por fim, pode-se também perceber que os dois métodos reduzem em cerca de 32% o tempo total de codificação em relação ao sistema sem busca rápida.

5.2 Experimento 2

Neste caso, realizamos testes semelhantes aos do Experimento 1, considerando aqui, porém, blocos de 30 ms com sub-blocos de 7,5 ms. Os resultados encontrados, qualitativamente análogos aos do caso anterior, são relacionados na Tabela 2.

Tabela 2: Experimento 2 - Medidas objetivas de qualidade em dB e tempo de processamento (TP) com dicionário adaptativo com atrasos fracionários e blocos de 30 ms com sub-blocos de 7,5 ms.

Sistema CELP sem busca rápida				
Locutor	RSRSP	DC	DI	TP (s)
M1	16,26	3,12	1,20	37,708
M2	17,03	3,08	1,16	48,466
F1	17,74	3,14	1,21	35,961
F2	15,63	3,33	1,37	43,304
Sistema CELP com o Método I				
Locutor	RSRSP	DC	DI	TP (s)
M1	15,45	3,40	1,42	25,425
M2	16,10	3,27	1,30	32,612
F1	17,04	3,34	1,35	24,299
F2	15,07	3,38	1,40	29,137
Sistema CELP com o Método II				
Locutor	RSRSP	DC	DI	TP (s)
M1	16,16	3,22	1,26	25,618
M2	16,86	3,09	1,17	32,958
F1	17,56	3,17	1,24	24,312
F2	15,45	3,29	1,33	29,257

5.3 Experimento 3

Por fim, foi utilizado o dicionário adaptativo sem atrasos fracionários, com faixa de 20 a 146, no lugar daquele com atrasos fracionários utilizado anteriormente. Os resultados obtidos são mostrados na Tabela 3, usando-se blocos de 20 ms com sub-blocos de 5 ms, como no Experimento 1 acima.

Tabela 3: Experimento 3 - Medidas objetivas de qualidade em dB e tempo de processamento (TP) com dicionário adaptativo sem atrasos fracionários e blocos de 20 ms com sub-blocos de 5 ms.

Sistema CELP sem busca rápida				
Locutor	RSRSP	DC	DI	TP (s)
M1	17,26	3,01	1,12	7,336
M2	18,37	2,87	1,00	9,489
F1	18,61	2,90	1,04	7,024
F2	16,77	2,84	1,00	8,439
Sistema CELP com o Método I				
Locutor	RSRSP	DC	DI	TP (s)
M1	16,12	3,27	1,31	3,576
M2	16,75	3,20	1,25	4,614
F1	17,68	3,51	1,51	3,416
F2	15,72	3,18	1,24	4,105
Sistema CELP com o Método II				
Locutor	RSRSP	DC	DI	TP (s)
M1	17,23	3,02	1,12	3,702
M2	18,13	2,86	1,00	4,747
F1	18,55	2,94	1,06	3,559
F2	16,56	2,96	1,07	4,246

Cabe notar aqui a redução drástica do tempo total de processamento devido ao uso de um dicionário adaptativo sem atrasos fracionários, causada pela diminuição das seqüências candidatas e pelo fato de não haver necessidade de se utilizar o filtro interpolador para a obtenção dos atrasos com frações de amostras. Neste caso o tempo total de codificação foi reduzido em torno de 50% em relação ao sistema sem busca rápida. Em termos qualitativos, porém, temos que o Método II introduz significativamente menos distorção também neste caso. Pode-se perceber também que o uso do dicionário adaptativo sem atrasos fracionários melhora as sentenças M2, F1 e F2 para o caso sem busca rápida, e as sentenças M2 e F2 para os caso do Método II; quando consideradas a DC e a DI (tabelas 1 e 3). Isto pode ter ocorrido pelo fato destas medidas analisarem mais as diferenças entre os sinais no domínio da freqüência, apesar de se esperar sempre que a introdução de um dicionário com atrasos fracionários melhore a qualidade.

5.4 Experimento 4

Foi realizado um teste subjetivo informal, que consistiu em apresentar sinais de voz decodificados pelo sistema CELP sem busca rápida e com o Método II para um

total de 23 ouvintes, a fim de que os mesmos dessem suas respectivas opiniões. A Tabela 4 mostra os percentuais de ouvintes que acharam melhor o sistema sem busca rápida, o sistema com o Método II, ou qualidade indistinguível, para cada sentença. Pode-se perceber, dos resultados apresentados, que a qualidade do sistema com o Método II foi julgada superior para a sentença F1 e inferior para a sentença M1, enquanto que para o caso das sentenças M2 e F2, as qualidades foram consideradas indistinguíveis.

Tabela 4: Experimento 4 - Avaliação subjetiva informal: percentuais de ouvintes que julgaram ser melhor o sistema sem busca rápida (SR), o sistema com o Método II, ou qualidade idistinguível, para blocos de 20 ms com sub-blocos de 5 ms e dicionário adaptativo sem atrasos fracionários.

Sentença	Melhor SR (%)	Melhor II (%)	Iguais (%)
M1	60,87	8,70	30,43
M2	17,39	39,13	43,48
F1	17,39	56,52	26,09
F2	26,09	30,43	43,48

6 Conclusão

Este trabalho apresentou um método de busca rápida no dicionário adaptativo em codificadores CELP e a posterior comparação com outro método existente na literatura. Os testes foram feitos em termos de medidas objetivas de qualidade e tempo de processamento, para dicionários com e sem atrasos fracionários e diferentes tamanhos de blocos de voz. Os resultados mostraram que o método proposto produz qualidade superior, enquanto reduz o tempo de processamento na mesma proporção (em torno de 30% a 50% do tempo total de codificação) que o método de referência. Um teste subjetivo informal indicou que o método proposto obtém qualidade comparável ao método convencional de busca.

Referências

[1] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 937–940, 1985.

[2] W. B. Kleijn, D. J. Krasinski, and R. H. Ketchum, "Fast methods for the CELP speech coding," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1330–1342, Aug. 1990.

[3] R. S. Maia, "Codificação CELP e análise espectral de voz," Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, Mar. 2000.

[4] R. S. Maia, C. B. Ribeiro, F. G. V. Resende Jr. e S. L. Netto, "Um sistema CELP para a codificação da fala a 4,4 kb/s," *XIII Congresso Brasileiro de Automática*, Florianópolis, Brasil, Set. 2000.

[5] L. M. da Silva and A. Alcaim, "A modified CELP model with computacionally efficient adaptive codebook search," *IEEE Signal Processing Letters*, vol. 2, pp. 44–45, Mar. 1995.

[6] N. Kitawaki, H. Nagabuchi, and K. Itoh, "Objective quality evaluation for low-bit-rate speech coding systems," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 242–248, Feb. 1988.

[7] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*. New York:NY, Macmillan, 1993.