

TREINAMENTO DA COMPENSAÇÃO DE EXCITAÇÕES DE VOZ ESPARSAS

Miguel Arjona Ramírez

Resumo - A excitação esparsa possibilita o emprego de algoritmos de busca eficientes para a obtenção de alta qualidade de reprodução da voz. Entretanto, conforme se diminui a taxa de transmissão, o espargimento começa a degradar a qualidade com o surgimento de efeitos auditivos que podem ser compensados pela fase. Descreve-se um procedimento para treinar a resposta impulsiva de um filtro para compensação de fase. Os filtros de compensação de dois codificadores ACELP de quatro pulsos à taxa de 5.3 kbit/s são treinados com os sinais de uma extensa base de dados. Um dos codificadores usa a busca focalizada (FOCS) e outro emprega a busca conjunta de posição e amplitude (JPAS), operacionalmente menos complexa. Efetuam-se testes subjetivos de audição envolvendo estes codificadores compensados e também versões sem compensação operando a 8 kbit/s. Os resultados indicam que a compensação treinada de espargimento eleva a qualidade do codificador de taxa baixa com JPAS a um nível próximo da qualidade dos codificadores de taxa mais alta. Finalmente, verifica-se que o truncamento da resposta impulsiva da compensação, além de reduzir a complexidade operacional da busca fixa, melhora ligeiramente a qualidade da voz quando aplicado antes do treinamento.

Abstract - Sparse excitation from fixed codebooks provides for efficient search algorithms and high-quality speech. However, as the bit rate decreases, sparsity causes hearing effects that may be phase-compensated. A procedure is described for training the impulse response of a phase compensation filter over a large database. Two four-pulse ACELP coders operating at 5.3 kbit/s have their compensation filters trained. One of the coders searches the fixed codebook with the focused search (FOCS) algorithm and the other one uses the joint position and amplitude search (JPAS) algorithm. Listening tests were performed involving both compensated and uncompensated versions of these coders at 5.3 kbit/s and the uncompensated version at 8 kbit/s. The lower rate coders provided with trained sparsity compensation are perceived to deliver speech quality equivalent to that achieved with 8 kbit/s coders. Additionally, the beneficial hearing effect is more pronounced with the computationally efficient JPAS. Further, truncation of the dispersion response prior to its training is shown to decrease the operational complexity of the search algorithm while enhancing the speech quality by a small amount.

Palavras-chave: Codificação de voz, CELP, ACELP, multipulsos algébricos, excitação esparsa, compensação impulsiva, dispersão impulsiva.

O autor é do Departamento de Engenharia de Sistemas Eletrônicos, Escola Politécnica, Universidade de São Paulo, São Paulo, SP.
E-mail: miguel@lps.usp.br

1. INTRODUÇÃO

A maior parte das amostras das excitações esparsas é nula e as posições das amostras não-nulas são atribuídas de forma estocástica ou determinística. O uso de dicionários fixos esparsos permitiu sua aplicação generalizada devido à existência de algoritmos de busca eficientes. Além disso, um efeito benéfico que também contribui para essa popularidade é a melhora da qualidade perceptual que a excitação esparsa proporciona [1]. Em particular, dentre os codificadores de voz com algoritmos ACELP e dicionários multipulso algébricos padronizados mais recentemente encontra-se o codificador G.729 da ITU-T [2], que opera à taxa de 8 kbit/s. Este tipo de excitação determinística esparsa é usada também em codificadores de voz com taxas mais baixas de operação como o G.723.1 da ITU-T [3] para a taxa de 5.3 kbit/s. Porém, este último codificador apresenta distorção mais audível em decorrência do espargimento mais acentuado da sua excitação. Por esta razão, foi escolhido para a aplicação de dois métodos de compensação do espargimento enquanto o primeiro codificador foi tomado como referência.

Como já foi mencionado, a motivação principal para o uso de excitações esparsas é a aplicação de algoritmos computacionalmente mais eficientes para a busca da excitação fixa. Assim, enquanto os codificadores G.729 e G.723.1 de referência usam o algoritmo de busca focalizada (“focused search”) da excitação fixa, desenvolveu-se o método de busca em árvore “depth first tree search”, DFTS, especialmente para um codificador multimídia para voz digital e dados simultâneos (“digital simultaneous voice and data” - DSVD), que veio a ser incorporado na recomendação G.729A da ITU-T [4].

Uma aplicação de larga escala de codificadores esparsos eficientes ocorre no sistema D-AMPS TDMA de telefonia celular IS-136, que incorpora o codec IS-641-A como versão melhorada para a taxa plena, denominado “enhanced full rate (EFR) codec” [5].

Os algoritmos de busca DFTS e EFR ganham eficiência sobre a busca focalizada de referência na determinação das posições dos quatro pulsos porque realizam buscas preliminares sobre subconjuntos de pares de posições [5, 6].

Um algoritmo de busca eficiente de excitação esparsa, denominado “busca conjunta de posição e amplitude” (“joint position and amplitude search” - JPAS) [7], preserva mais a informação de amplitude dos pulsos e tem mais precisão na determinação da posição do pulso dominante, apresentando um ganho de eficiência decorrente da redução do número de combinações de posições pesquisadas para os demais pulsos.

Os algoritmos acima foram anteriormente comparados em complexidade e em qualidade de reprodução [8] para a taxa de operação de 8 kbit/s. Para a medida de complexidade

operacional de implementações com aritmética de ponto fixo usou-se a unidade WMOPS (“weighted million operations per second”), contabilizada no pior caso, e tomou-se a complexidade da busca focalizada, 6,31 WMOPS, como referência. Assim, a complexidade medida para a busca JPAS foi abaixo de 25% enquanto as complexidades medidas para as buscas DFTS e EFR foram de 30% e 54%, respectivamente. As qualidades de reprodução da voz pelos codificadores com os três algoritmos de busca eficiente excederam a qualidade percebida do codificador de referência com a busca focalizada.

Assim, para avaliar o impacto de um método de busca eficiente no processo de compensação de espargimento, implementou-se a busca JPAS num codificador com taxa de operação de 5,3 kbit/s.

2. COMPENSAÇÃO DE FASE

O espargimento da excitação fixa é mais notado auditivamente quando a excitação ideal do filtro de síntese apresenta uma distribuição mais uniforme de energia no tempo. Isto ocorre quando o segmento de voz em questão é surdo, solicitando uma contribuição relativamente maior do dicionário fixo na composição da excitação total, que não é periódica. Por outro lado, quando o segmento de voz é sonoro, o dicionário adaptativo fornece a maior contribuição relativa e a componente complementar do dicionário fixo é apropriada para a modelagem da excitação ideal, que é de natureza periódica impulsiva e, portanto, esparsa neste caso. De fato, a melhora perceptual ocasionada pela excitação esparsa, comentada na Seção 1, deve decorrer deste casamento de características entre o dicionário esparsa e a excitação necessitada pelos segmentos quase-periódicos.

As observações acima permitem compreender que os efeitos causados pela excitação esparsa sobre o sinal de voz reconstruído ocorram predominantemente para os sons surdos, quando podem chegar a ser percebidos como uma componente quase-periódica estranha à natureza do sinal [9]. Sabe-se também que tais efeitos podem ser significativamente atenuados pela adição de uma componente aleatória ao espectro de fase nas altas frequências [9]. Embora este procedimento reduza a periodicidade do sinal, ele não afeta significativamente a sensação de tom porque o “pitch” é percebido preponderantemente a partir dos harmônicos mais baixos da frequência fundamental.

Essa alteração do espectro de fase é efetuada no nível da excitação porque neste ponto se pode aproveitar o filtro de síntese na redução de efeitos transitórios causados pelo tratamento do sinal em blocos. Esta localização da compensação de espargimento permite que ela seja implementada através de um filtro $F(z)$ aplicado à excitação fixa como está representado na Figura 2. Isto ocorre porque os espectros de fase da excitação e do filtro se somam nesta disposição.

Assim, o filtro de compensação de espargimento $F(z)$ é um passa-tudo não-recorrente (FIR) projetado para ter um espectro de fase nulo até 1,5 kHz e uniformemente distribuído entre 0 e $\pi/2$ na faixa entre 1.5 e 4 kHz. A resposta impulsiva resultante está apresentada na Figura 1 depois do truncamen-

to dentro dos limites do sub-bloco $0 \leq n < 60$.

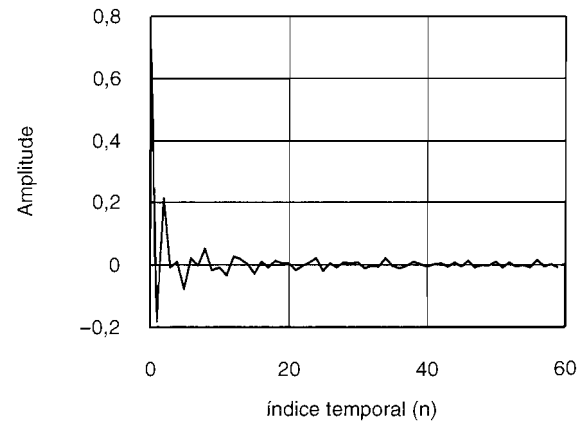


Figura 1. Resposta impulsiva projetada para compensar a fase da excitação esparsa.

3. TREINAMENTO DA COMPENSAÇÃO DE ESPARGIMENTO

A compensação de fase é um componente muito valioso num dispositivo destinado à compensação das conseqüências do espargimento da excitação. Entretanto, considerou-se que a eficácia da compensação poderia ser aumentada após um processo de treinamento [10]. Portanto, elaborou-se um processo de treinamento que vai moldando gradativamente a resposta impulsiva do filtro de compensação, inicialmente identificada com a resposta impulsiva do filtro de compensação de fase projetado.

A excitação composta injetada no filtro de síntese do decodificador resulta da combinação

$$e(n) = \eta_f c_f(n) * f(n) + \eta_a c_a(n), \quad (1)$$

onde $c_f(n)$ e $c_a(n)$ representam os vetores-código fixo e adaptativo, respectivamente, cujos ganhos são η_f e η_a , e $f(n)$ é a resposta impulsiva do filtro de compensação.

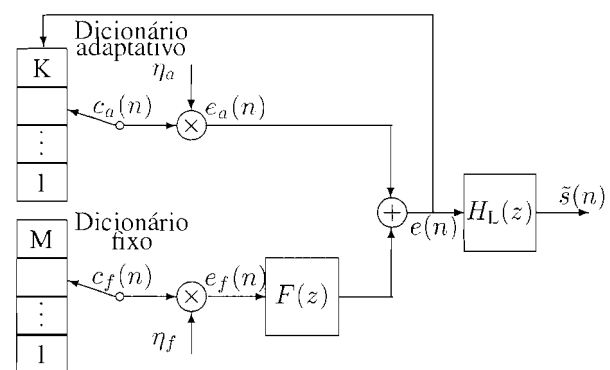


Figura 2. Síntese CELP com excitação compensada contra espargimento.

No codificador, durante a fase de busca da excitação fixa, o vetor-alvo $u_f(n)$ é reconstruído como

$$\tilde{u}_f(n) = \eta_f c_f(n) * f(n) * h(n), \quad (2)$$

onde $h(n)$ é a resposta impulsiva do filtro de síntese ponderado.

Permutando-se a ordem das duas convoluções na Equação (2), obtém-se

$$\tilde{u}_f(n) = \eta_f c_f(n) * h(n) * f(n), \quad (3)$$

que é mais conveniente de se usar na fase de treinamento, em que a resposta impulsiva $f(n)$ da compensação é incógnita. Além disso, notando-se que

$$q_u(n) = \eta_f c_f(n) * h(n) \quad (4)$$

é a reconstrução sem compensação do vetor-alvo, sua versão compensada na Equação (3) pode se expressar como

$$\tilde{u}_f(n) = q_u(n) * f(n) \quad (5)$$

ou, com notação matricial, ela pode ser escrita como

$$\tilde{\mathbf{u}}_f = \mathbf{Q}\mathbf{f}. \quad (6)$$

A matriz \mathbf{Q} na Equação (6) representa a operação de convolução com a reconstrução sem compensação q_u . Portanto, ela tem uma estrutura Toeplitz triangular inferior com o vetor \mathbf{q}_u na primeira coluna.

Assim, o vetor de erro ponderado no sub-bloco m é dado por

$$\begin{aligned} \mathbf{e}_m &= \mathbf{u}_{f,m} - \tilde{\mathbf{u}}_{f,m} \\ &= \mathbf{u}_{f,m} - \mathbf{Q}_m \mathbf{f} \end{aligned} \quad (7)$$

de forma tal que o erro quadrático total acumulado ao longo de toda a base de dados pode ser representado como

$$\begin{aligned} \sum_{m=1}^{N_S} \mathbf{e}_m^T \mathbf{e}_m &= \sum_{m=1}^{N_S} (\mathbf{u}_{f,m}^T - \mathbf{f}^T \mathbf{Q}_m^T) (\mathbf{u}_{f,m} - \mathbf{Q}_m \mathbf{f}) \\ &= \sum_{m=1}^{N_S} (\mathbf{u}_{f,m}^T \mathbf{u}_{f,m} - 2\mathbf{f}^T \mathbf{Q}_m^T \mathbf{u}_{f,m} + \mathbf{f}^T \mathbf{Q}_m^T \mathbf{Q}_m \mathbf{f}) \end{aligned} \quad (8)$$

onde N_S é o número total de sub-blocos contidos na base de dados.

A minimização do erro quadrático ocorre quando seu gradiente em relação à resposta impulsiva de compensação é o vetor nulo, isto é,

$$\begin{aligned} \mathbf{0} &= \nabla \sum_{m=1}^{N_S} (\mathbf{e}_m^T \mathbf{e}_m) = \sum_{m=1}^{N_S} \frac{\partial}{\partial \mathbf{f}} (\mathbf{e}_m^T \mathbf{e}_m) \\ &= -2 \sum_{m=1}^{N_S} \mathbf{Q}_m^T \mathbf{u}_{f,m} + 2 \sum_{m=1}^{N_S} \mathbf{Q}_m^T \mathbf{Q}_m \mathbf{f}. \end{aligned} \quad (9)$$

Após uma reordenação da Equação (9), obtém-se o sistema de equações

$$\mathbf{A}\mathbf{f} = \mathbf{b}, \quad (10)$$

onde a matriz do sistema é $\mathbf{A} = \sum_{m=1}^{N_S} \mathbf{Q}_m^T \mathbf{Q}_m$ e o vetor constante é $\mathbf{b} = \sum_{m=1}^{N_S} \mathbf{Q}_m^T \mathbf{u}_{f,m}$. Este sistema acumulativo é resolvido na última etapa de cada iteração, fornecendo a nova resposta impulsiva de compensação. Há outros métodos que usam parte desta solução do sistema linear como incremento sobre a solução da iteração anterior [11].

Se a resposta \mathbf{f} da compensação for truncada além disso a um comprimento $L_T < L$, o gradiente na Equação (9) será composto apenas pelas primeiras L_T derivadas parciais $\frac{\partial}{\partial f(0)}, \frac{\partial}{\partial f(1)}, \dots, \frac{\partial}{\partial f(L_T-1)}$ de tal forma que o sistema linear de equações (10) torna-se $L_T \times L_T$ neste caso, sendo dado por

$$\overline{\mathbf{A}} \overline{\mathbf{f}} = \overline{\mathbf{b}}, \quad (11)$$

onde $\overline{\mathbf{A}}$ representa a submatriz $L_T \times L_T$ superior esquerda da matriz de autocorrelação \mathbf{A} enquanto $\overline{\mathbf{f}}$ e $\overline{\mathbf{b}}$ representam os subvetores compostos pelos primeiros L_T elementos de \mathbf{f} e \mathbf{b} , respectivamente.

Usou-se o codificador ACELP G.723.1 operando a 5,3 kbit/s [3] para treinar as respostas impulsivas de filtros de compensação de espargimento, tomando-se a resposta impulsiva de fase projetada que está representada na Figura 1 como a primeira estimativa. Efetuaram-se dois treinamentos completos, usando-se num deles o algoritmo de busca focalizada [2] e no outro o algoritmo de busca conjunta de posição e amplitude [12, 8] para buscar a excitação no dicionário fixo ACELP. Todos os sinais contidos na partição de teste da base de dados TIMIT foram usados nos treinamentos, totalizando uma duração de 5187 s ou 691,6 mil sub-blocos. Os treinamentos foram executados com o critério de maximização da relação sinal-ruído segmentada (WSNRSEG) no nível do vetor-alvo, que é o procedimento seguido pelo codificador durante a busca da excitação. As medidas obtidas nas situações inicial e final estão na Tabela 1.

Iteração	Focalizada		Conjunta	
	SNRSEG (dB)	WSNRSEG (dB)	SNRSEG (dB)	WSNRSEG (dB)
Inicial	9,20	4,65	8,94	4,53
Final	9,26	4,73	8,99	4,59

Tabela 1. Desempenho objetivo de codificadores com compensação de fase projetada e treinada a 5,3 kbit/s com as buscas focalizada e conjunta da excitação fixa ACELP.

As respostas impulsivas das compensações de espargimento obtidas com o treinamento para a busca conjunta e para a busca focalizada estão representadas nas Figuras 3 e 4, sendo que esta última é apresentada como um incremento sobre a primeira para ressaltar a diferença que existe entre ambas.

4. COMPLEXIDADE E DESEMPENHO RESULTANTES

Executou-se um teste de avaliação auditiva subjetiva do codificador G.723.1 a 5,3 kbit/s com cada algoritmo de busca, focalizada e conjunta, usando-se cada um deles tanto na versão sem compensação quanto em duas versões compensadas, além de dois codificadores G.729 a 8 kbit/s usando

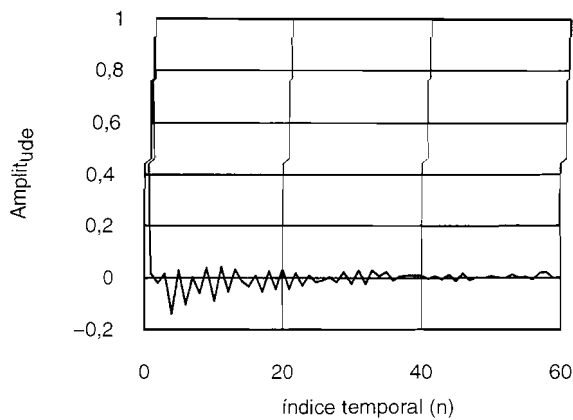


Figura 3. Resposta impulsiva da compensação treinada com a busca conjunta no codificador G.723.1 à taxa de 5,3 kbit/s.

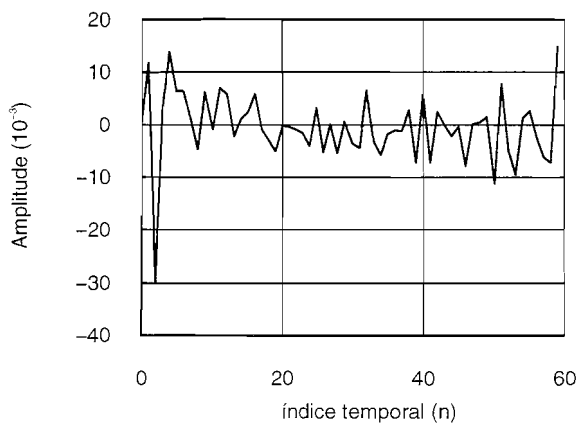


Figura 4. Resposta impulsiva da compensação treinada com a busca focalizada no codificador G.723.1 à taxa de 5,3 kbit/s, representada como um incremento sobre a resposta treinada com a busca conjunta.

cada um dos dois algoritmos de busca sem compensação. Ao todo, nove condições foram avaliadas por 12 ouvintes femininos e 12 masculinos. Usaram-se quatro pares de frases, que foram reproduzidas através de fones de ouvido durante as audições. Cada par de frases é composto por duas sentenças curtas proferidas em português por dois locutores femininos e dois masculinos, com duração aproximada de 3 s e com um intervalo entre elas de 0,5 s de silêncio. O teste subjetivo foi do tipo de categorização absoluta (“absolute category rating” - ACR) com uma escala de 5 pontos onde 5 é excelente e 1 é ruim [13]. Os índices médios de opinião (MOS) resultantes são mostrados na Tabela 2, evidenciando que o codificador com busca conjunta compensada a 5,3 kbit/s tem um desempenho muito próximo do codificador com busca focalizada a 8 kbit/s. Ao contrário, o codificador com busca focalizada compensada a 5,3 kbit/s não consegue atingir o nível de qualidade dos codificadores que operam a 8 kbit/s.

Todos os codificadores utilizados foram implementados com aritmética de ponto fixo e suas complexidades operacionais foram medidas em milhões de operações ponderadas

Condição	Ouvintes		
	Fem.	Masc.	Todos
Original	3,65	4,13	3,89
FOCS	3,52	3,58	3,55
FOCF	3,50	3,75	3,63
FOCT	3,52	3,63	3,57
JPAS	3,60	3,48	3,54
JPAF	3,54	3,60	3,57
JPAT	3,60	3,73	3,67
FOC8	3,60	3,79	3,70
JPA8	3,77	3,83	3,80

FOCS: Busca focalizada sem compensação a 5,3 kbit/s

FOCF: Busca focalizada compensada em fase

FOCT: Busca focalizada com compensação treinada

JPAS: Busca conjunta sem compensação a 5,3 kbit/s

JPAF: Busca conjunta compensada em fase

JPAT: Busca conjunta compensada por filtro treinado

FOC8: Busca focalizada sem compensação a 8 kbit/s

JPA8: Busca conjunta sem compensação a 8 kbit/s

Tabela 2. Qualidade subjetiva de codificadores com buscas ACELP em aritmética de ponto fixo nas variedades desprovidas e providas de compensação de espargimento à taxa de 5,3 kbit/s e apenas nas variedades desprovidas de compensação à taxa de 8 kbit/s em testes subjetivos de categorização absoluta a partir de gravações feitas com ruído ambiente de escritório.

Busca Fixa	Função	Complexidade (WMOPS)	
		Normal	Compensada
Focalizada	Codif.	3,54	3,94
Conjunta	Codif.	1,24	1,59
Focalizada	Decod.	0,59	0,62
Conjunta	Decod.	0,55	0,59

Tabela 3. Medidas de complexidade operacional no pior caso de implementações em ponto fixo de dois algoritmos de busca ACELP nas versões sem compensação (normal) e compensada operando à taxa de 5,3 kbit/s com os sinais da partição de teste da base de dados TIMIT.

por segundo (WMOPS), tomadas nos piores casos. Conforme aparece na Tabela 3, a complexidade do codificador pode aumentar mais do que um quarto para acomodar a compensação de espargimento. Mas há condições de redução da complexidade porque as amplitudes das respostas impulsivas de compensação decaem rapidamente com o afastamento a partir da origem (Figuras 3 e 4), podendo muito bem ser truncadas em analogia com as respostas mais breves usadas no filtro de conformação de pulsos em codificadores que operam por volta de 4 kbit/s [14], como será visto em seguida.

5. TRUNCAMENTO DAS RESPOSTAS DE COMPENSAÇÃO

Os comprimentos das respostas impulsivas dos filtros de compensação de espargimento têm sido considerados idênticos ao comprimento do sub-bloco até agora. Porém,

isto acarreta uma maior complexidade operacional em virtude das convoluções mais longas que têm que ser executadas. Portanto, é desejável o truncamento das respostas das compensações a comprimentos inferiores ao do sub-bloco.

Observa-se, por outro lado, que as respostas de compensação obtidas até este momento decaem rapidamente ao afastar-se da origem. De fato, este comportamento pode ser constatado nas respostas de compensação projetadas como a da Figura 1 bem como nas treinadas como a que está representada na Figura 3.

Dessa forma, foram obtidas duas respostas de compensação truncadas em 20 e 10 amostras simplesmente pelo truncamento da resposta projetada, que está representada na Figura 1. Deve-se recordar que o truncamento no domínio do tempo é a aproximação espectral de menor erro quadrático médio em decorrência do teorema de Parseval.

Ao contrário, as respostas treinadas não foram simplesmente truncadas, mas foram obtidas por processos de treinamento iniciados com as respectivas respostas projetadas truncadas. Os resultados obtidos estão dispostos nas Tabelas 4, demonstrando que não houve perda de desempenho em decorrência do truncamento. Na realidade, nota-se um desempenho ligeiramente melhor das versões mais truncadas comparando-se as duas Tabelas com a Tabela 1. As Figuras 5 e 6 mostram as respostas impulsivas após o treinamento das compensações de espargimento truncadas em 10 amostras.

Comprimento da resposta de compensação: 20 amostras

Iteração	Busca focalizada		Busca conjunta	
	SNRSEG (dB)	WSNRSEG (dB)	SNRSEG (dB)	WSNRSEG (dB)
Inicial	9,21	4,67	8,96	4,55
Final	9,29	4,74	9,01	4,60

Comprimento da resposta de compensação: 10 amostras

Iteração	Busca focalizada		Busca conjunta	
	SNRSEG (dB)	WSNRSEG (dB)	SNRSEG (dB)	WSNRSEG (dB)
Inicial	9,24	4,71	8,99	4,59
Final	9,33	4,78	9,04	4,64

Tabela 4. Desempenhos objetivos das respostas de compensação de espargimento truncadas treinadas em codificadores operando a 5,3 kbit/s com as buscas focalizada e conjunta do dicionário fixo ACELP.

Busca Fixa	Comprimento da compensação	Complexidade (WMOPS)
Focalizada	10	3,79
	20	3,88
Conjunta	10	1,42
	20	1,48

Tabela 5. Medidas de complexidade operacional no pior caso de implementações em ponto fixo de dois algoritmos de busca ACELP nas versões com compensação treinada truncada operando à taxa de 5,3 kbit/s com os sinais da partição de teste da base de dados TIMIT.

Por outro lado, a consequência imediata do truncamento da resposta impulsiva da compensação de espargimento é a redução da complexidade em até 10%, como pode ser verificado comparando a Tabela 5 com a Tabela 3 apresentada anteriormente.

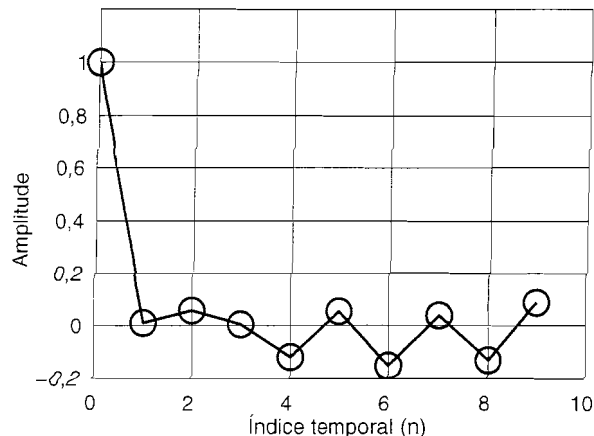


Figura 5. Resposta impulsiva da compensação truncada em 10 amostras treinada com a busca conjunta no codificador G.723.1 à taxa de 5,3 kbit/s.

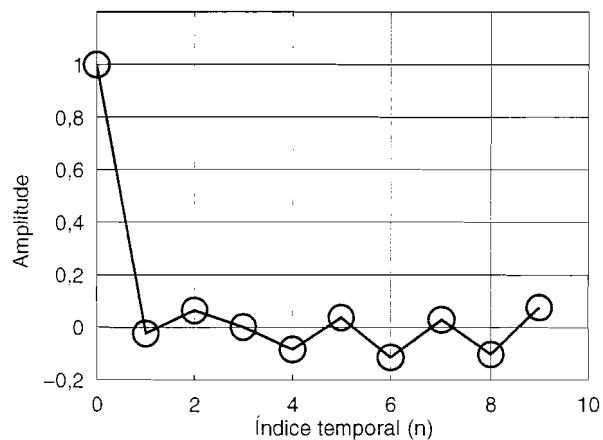


Figura 6. Resposta impulsiva da compensação truncada em 10 amostras treinada com a busca focalizada no codificador G.723.1 à taxa de 5,3 kbit/s.

6. CONCLUSÃO

Testaram-se vários codificadores de voz com dicionários fixos de quatro pulsos operando nas taxas de 5,3 e 8 kbit/s. Os codificadores à taxa mais baixa foram melhorados com dois tipos de filtros de compensação de espargimento: um projetado para tornar a fase mais aleatória nas altas frequências e outro cuja resposta impulsiva foi refinada num processo de treinamento a partir da resposta impulsiva do primeiro. Ademais, dois algoritmos de busca da excitação fixa foram usados para cada condição: a busca focalizada de referência e a busca conjunta de posição e amplitude. Em testes subjetivos de audição, a compensação de espargimento treinada foi capaz de elevar a qualidade da busca conjunta à taxa de 5,3 kbit/s a um nível muito próximo ao da busca focalizada a 8 kbit/s

ao passo que não pôde promover uma melhora significativa de qualidade da busca focalizada. Uma possível causa deste comportamento é o maior cuidado exercido pela busca conjunta na determinação dos sinais dos pulsos. Verificou-se ainda que o truncamento da resposta da compensação a comprimentos menores que um sub-bloco, além de reduzir a complexidade do codificador, não prejudica seu desempenho.

REFERÊNCIAS

- [1] W. B. Kleijn, D. J. Krasinski, and R. H. Ketchum. "Fast methods for the CELP speech coding algorithm." *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 8, pp. 1330–1342, Aug. 1990.
- [2] R. Salami, C. Laflamme, J.-P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham. "Design and description of CS-ACELP, a toll quality 8 kb/s speech coder." *IEEE Trans. Speech Audio Processing*, vol. 6, no. 2, pp. 116–130, Mar. 1998.
- [3] "Dual rate speech coder for multimedia applications transmitting at 5.3 and 6.3 kbit/s," ITU-T Recommend. G.723.1, Mar. 1996.
- [4] R. Salami, C. Laflamme, B. Bessette, and J.-P. Adoul. "Description of ITU-T recommendation G.729 annex A: Reduced complexity 8 kbit/s CS-ACELP codec." in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Munich, 1997, vol. 2, pp. 775–778.
- [5] T. Honkanen, J. Vainio, K. Järvinen, and P. Haavisto. "Enhanced full rate codec for IS-136 digital cellular system." in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Munich, 1997, vol. 2, pp. 731–734.
- [6] R. Salami, C. Laflamme, B. Bessette, and J.-P. Adoul. "ITU-T G.729 Annex A: Reduced complexity 8 kb/s CS-ACELP codec for digital simultaneous voice and data." *IEEE Commun. Mag.*, vol. 35, no. 9, pp. 56–63, Sept. 1997.
- [7] M. Arjona Ramírez. "A low-complexity search algorithm for speech coders with sparse excitation." *Revista da Sociedade Brasileira de Telecomunicações*, a ser publicado.
- [8] M. Arjona Ramírez and M. Gerken. "Joint position and amplitude search of algebraic multipulses." *IEEE Trans. Speech Audio Processing*, vol. 8, no. 5, pp. 633–637, Sept. 2000.
- [9] R. Hagen, E. Ekudden, B. Johansson, and W. B. Kleijn. "Removal of sparse-excitation artifacts in CELP," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Seattle, 1998, vol. 1, pp. 145–148.
- [10] M. Arjona Ramírez. "Compensação de espargimento em codificadores de voz," in *Anais do Simpósio Brasileiro de Telecomunicações*. Gramado, 2000, CD-ROM.
- [11] K. Yasunaga, H. Ehara, K. Yoshida, and T. Morii. "Dispersed-pulse codebook and its application to a 4 kb/s speech coder," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Istanbul, 2000, vol. 3, pp. 1503–1506.
- [12] M. Arjona Ramírez and M. Gerken. "A multistage search of algebraic CELP codebooks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, 1999, vol. 1, pp. 17–20.
- [13] P. Kroon. "Evaluation of speech coders." in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds., pp. 467–494. Elsevier Science, Amsterdam, 1995.
- [14] T. Amada, K. Miseki, and M. Akamine. "CELP speech coding based on an adaptive pulse position codebook," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, 1999, vol. 1, pp. 13–16.

Miguel Arjona Ramírez graduou-se em Engenharia Eletrônica em 1980 pelo Instituto Tecnológico de Aeronáutica (ITA) e recebeu os títulos de Mestre e Doutor em Engenharia Elétrica pela Escola Politécnica da Universidade de São Paulo (EPUSP) em 1992 e 1997, respectivamente. Tem atuado na Escola Politécnica desde 1991, sendo atualmente Professor Doutor do Departamento de Engenharia de Sistemas Eletrônicos. O autor tem trabalhado com processamento digital de sinais de voz e sinalização telefônica desde o desenvolvimento de algoritmos para circuitos integrados dedicados de codificação de voz (Philips International Institute, 1981), passando pelo desenvolvimento de Unidades de Resposta Auditiva (URAs) para aplicações de consulta telefônica a bancos de dados (Itautec, 1982 a 1990) até a pesquisa de algoritmos de codificação do sinal de voz (EPUSP, desde 1991). Suas áreas de interesse em pesquisa são codificação de voz, compressão de sinais, reconhecimento de voz e processamento digital de sinais.