

Uma Comparação de Quantizadores Vetoriais Preditivos para Codificação dos Parâmetros LSF da Voz

Rodrigo C. de Lamare e Abraham Alcaim

CETUC - PUC-RIO

22453-900 Rio de Janeiro - Brasil

e-mails: delamare@infolink.com.br

alcaim@cetuc.puc-rio.br

Resumo

A quantização vetorial preditiva chaveada (QVPC) mostrou-se mais eficiente do que esquemas como a quantização vetorial preditiva (QVP) convencional e a quantização vetorial sem memória (QV) na redução da taxa de bits dos parâmetros LSF em aplicações de codificação de voz. Neste artigo é realizada uma comparação entre diferentes esquemas de QVPC, QVP e QV, tanto em canal ideal como em canal ruidoso.

1 Introdução

Grande parte dos algoritmos modernos de codificação de voz é baseada na técnica de análise de predição linear (ou análise LPC - “Linear Predictive Coding”), onde um sinal de excitação é aplicado a um filtro só de pólos que representa a informação da envoltória espectral do sinal de voz. Este artigo trata do problema de transmissão eficiente da informação da envoltória espectral da voz usando quantização vetorial (QV). As frequências espectrais em linha ou parâmetros LSF foram escolhidos para representar os coeficientes LPC, uma vez que estes se mostraram adequados para procedimentos de quantização e interpolação. Os parâmetros LSF apresentam, em geral, uma correlação significativa entre quadros de voz sucessivos, especialmente para segmentos sonoros. Na quantização vetorial sem memória (QV), cada vetor de LSFs é quanti-

zado de maneira independente de qualquer outro conjunto de LSFs. Contudo, esta não é a melhor forma de codificar os parâmetros LSF, já que consideráveis ganhos podem ser conseguidos ao explorar a inerente correlação entre conjuntos de LSFs adjacentes. Alguns esquemas de quantização vetorial preditiva (QVP), que exploram a correlação entre quadros, foram propostos nos últimos anos. Neste trabalho, foi conduzida uma comparação entre algumas das mais populares e bem sucedidas técnicas de QV, onde inclui-se a quantização vetorial sem memória, simplesmente denotada por QV, a quantização vetorial preditiva (QVP) e uma abordagem mais recente denominada quantização vetorial preditiva chaveada (QVPC). Foram excluídos deste estudo métodos que requerem maiores retardos como QV matricial e QVP bidirecional. Este artigo é organizado da seguinte maneira. As Seções 2, 3 e 4 descrevem as estruturas de quantização vetorial dos parâmetros LSF. A Seção 5 apresenta e discute os resultados de simulações e a Seção 6 as conclusões.

2 Quantização Vetorial Sem Memória

A quantização vetorial sem memória (QV) é uma técnica que codifica cada vetor de LSFs de maneira independente de vetores adjacentes [1]. Paliwal e Atal [2] demonstraram que um esquema de QV multiestágios com estrutura em árvore é capaz de codificar de forma eficiente os parâmetros LSF com

24 bits por quadro.

3 Quantização Vetorial Preditiva

Uma maneira simples de obter vantagens na exploração da memória de uma fonte é usar quantização vetorial preditiva (QVP) [1]. Um preditor linear vetorial forma uma estimativa dos vetores de entrada como uma combinação linear de observações passadas, e o vetor residual de predição é quantizado vetorialmente como expresso por:

$$\delta_{j+1} = \mathbf{f}_{j+1} - \hat{\mathbf{f}}_j \cdot \rho^t \quad (1)$$

onde ρ é o vetor com os coeficientes de correlação das LSFs e $\hat{\mathbf{f}}_j$ é a versão quantizada de \mathbf{f}_j , o vetor de LSFs no instante j . Neste trabalho, restringiremos os experimentos a preditores de primeira ordem, como expresso em (1), que demonstraram atingir a maior parte dos ganhos de codificação.

4 Quantização Vetorial Preditiva Chaveada

A correlação entre quadros de voz pode ser explorada através de métodos de QV com memória como a QVP. Contudo, existem situações de mudanças rápidas na envoltória espectral da voz e, portanto, baixas correlações entre conjuntos de LSFs adjacentes. De fato, esta observação motivou a combinação de técnicas de QVP com QV sem memória para codificar quadros com baixa correlação de maneira separada de quadros com alta correlação. Uma busca de ambos esquemas é realizada para cada quadro e o melhor candidato, no que diz respeito a um critério de distorção, é codificado e transmitido [3].

Entre os esquemas de QVPC, destacamos aquele denominado *safety-net*, proposto por Eriksson et al. [4], mostrado na Figura 1. Este sistema chaveia entre uma estrutura QV particionada e uma estrutura QVP particionada, conseguindo uma quantização eficiente dos parâmetros LSF com apenas 20 bits por quadro. Outra técnica de QVPC, introduzida por McCree e De Martin [5], usa 2 estruturas multiestágios de QVP com busca em árvore, onde cada preditor é projetado para um banco de

dados de treinamento específico e opera com 21 bits por quadro. Neste trabalho, investigaremos o desempenho destes algoritmos, descritos na Tabela 1, juntamente com os métodos QV e QVP.

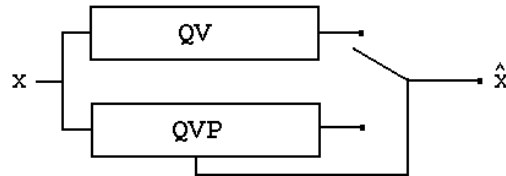


Figura 1: O princípio do “safety net”.

Tabela 1: Esquemas de QVPC.

Sigla	Estrutura dos Quantizadores
QVPC	1 esquema de QVP e 1 de QV.
QVPCP	2 esquemas de QVP.

5 Resultados Experimentais

O banco de dados de gravações para treinamento usado neste trabalho consiste em 2400 segundos de voz, extraídos de listas de frases foneticamente balanceadas para o Português falado no Rio de Janeiro [6], pronunciados por 40 locutores, 20 masculinos e 20 femininos. Um outro conjunto de 30 segundos de voz gravado por 6 diferentes locutores, 3 masculinos e 3 femininos, foi usado para avaliação. A voz gravada foi filtrada por um filtro passa-baixas em 3,4 kHz, por um filtro passa-altas em 120 Hz e amostrada em 8 kHz. Em seguida, foi realizada uma análise LPC de 10ª ordem usando o método da autocorrelação a cada 20 ms, com janela de análise de Hamming com 24 ms. Uma expansão de largura de banda de 15 Hz foi aplicada a cada pólo dos vetores de coeficientes LPC a fim de evitar picos espectrais agudos no espectro LPC, que podem resultar em voz metálica. Como critério de distorção foi escolhido o erro-quadrático ponderado [7], uma vez que este é capaz de propiciar uma melhor quantização dos formantes, que são mais importantes para o sistema auditivo, melhorando as qualidades objetiva e subjetiva da voz codificada e também porque é uma medida de baixa

complexidade para implementação. Esta medida é dada por:

$$d(\mathbf{f}, \hat{\mathbf{f}}) = \sum_{i=1}^p \alpha_i(\mathbf{f}) (f_i - \hat{f}_i)^2 \quad (2)$$

onde $\alpha(\mathbf{f}) = (\alpha_1(\mathbf{f})\alpha_2(\mathbf{f}) \dots \alpha_p(\mathbf{f}))$ é a função de ponderação definida por

$$\alpha_i(\mathbf{f}) = \frac{1}{f_i - f_{i-1}} + \frac{1}{f_{i+1} - f_i} \quad (3)$$

onde $i = 1, \dots, p$ e $f_0 = 0$ e $f_{p+1} = 0,5$.

Grandes economias em complexidade e armazenamento podem ser conseguidas através do uso de procedimentos sub-ótimos, como QV multiestágios e de técnicas de busca em árvore, onde uma aproximação dos M melhores vetores-código é guardada a cada estágio do procedimento de busca. De fato, em todas as estruturas deste trabalho foram usados esquemas de QV multiestágios com busca em árvore com parâmetro de busca $M = 12$ e 4 estágios.

O desempenho da QV foi avaliado através da distância espectral (DE) expressa por:

$$DE = \left[\sum_{f=0}^{4000} \frac{1}{4000} \left(10 \log_{10} \left| \frac{S(f)}{\hat{S}(f)} \right| \right)^2 \right]^{1/2} \quad (dB) \quad (4)$$

onde $S(f)$ e $\hat{S}(f)$ representam a envoltória espectral original e quantizada, respectivamente. O percentual de *outliers* (vetores de LSFs com DE entre 2 dB e 4 dB e com DE superior a 4 dB) é também examinado para todas as estruturas analisadas.

No projeto dos quantizadores vetoriais preditivos chaveados, o banco de dados de treinamento foi dividido em 2 subconjuntos de acordo com a DE entre vetores de LSFs consecutivos. Para o esquema QVPC os subconjuntos foram divididos de tal forma que 1/3 das LSFs do treinamento com maior DE foi usado para o projeto da estrutura QV e o restante para o da estrutura QVP. Já para a técnica QVPCP, que utiliza 2 estruturas QVP, a proporção de LSFs de treinamento foi de 50% com maior DE para uma estrutura QVP e de 50% com menor DE para a outra. Estas partições para treinamento foram as melhores dentre as várias testadas. Os preditores usados em esquemas de QVP dos diferentes sistemas de codificação foram projetados com base nos coeficientes de correlação calcu-

lados a partir de cada subconjunto de treinamento de parâmetros LSF.

5.1 Desempenho Em Canal Sem Ruído

Na Figura 3, as DEs médias das técnicas de quantização são mostradas como função do número de bits por quadro de voz. De acordo com esta figura, está claro que os esquemas QVPC e QVPCP alcançam o melhor desempenho entre os métodos descritos neste trabalho. O desempenho em termos da DE e o percentual de *outliers* entre 2 e 4 dB são mostrados nas Tabelas de 2 a 6, para as taxas de bits de 20 a 24 bits por quadro, respectivamente. É importante notar que nenhum dos quantizadores examinados possui DE maior que 4 dB. Os resultados em termos da DE média mostram que a melhor estrutura é a QVPCP, enquanto que no que se refere aos *outliers* o melhor esquema é o QVPC. Os métodos de QV preditiva chaveada mostraram-se capazes de economizar 4 bits em comparação com o QV sem memória.

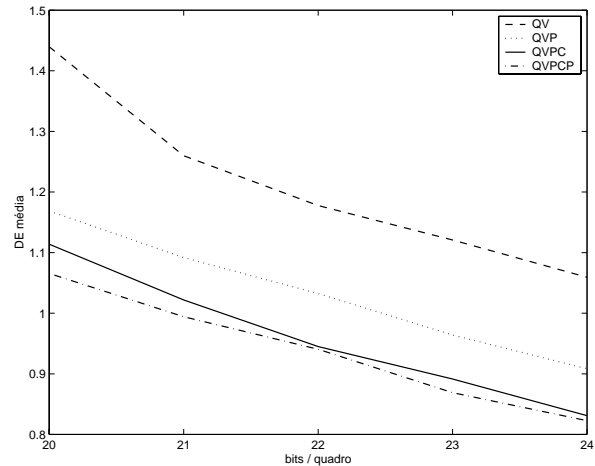


Figura 2: Desempenho dos quantizadores em termos da DE média.

5.2 Desempenho em Canal Ruidoso

Na seção anterior, verificamos que esquemas de QV com memória apresentam desempenho superior ao das estruturas de QV sem memória em canais sem ruído. Contudo, para ser útil em aplicações

Tabela 2: O desempenho dos esquemas de QV com 20 bits/quadro.

Quantizador	DE (dB)	>2 dB (%)
QV-20	1,4397	14,7508
QVP-20	1,1687	6,3787
QVPC-20	1,1139	3,3887
QVPCP-20	1,0657	3,9894

Tabela 3: O desempenho dos esquemas de QV com 21 bits/quadro.

Quantizador	DE (dB)	>2 dB (%)
QV-21	1,2598	6,8439
QVP-21	1,0918	5,5150
QVPC-21	1,0219	2,5931
QVPCP-21	0,9943	2,7926

Tabela 4: O desempenho dos esquemas de QV com 22 bits/quadro.

Quantizador	DE (dB)	>2 dB (%)
QV-22	1,1777	4,9834
QVP-22	1,0325	4,6512
QVPC-22	0,9450	2,1277
QVPCP-22	0,9405	2,6596

Tabela 5: O desempenho dos esquemas de QV com 23 bits/quadro.

Quantizador	DE (dB)	>2 dB (%)
QV-23	1,1209	3,7209
QVP-23	0,9641	3,9867
QVPC-23	0,8915	1,3298
QVPCP-23	0,8690	1,7952

Tabela 6: O desempenho dos esquemas de QV com 24 bits/quadro.

Quantizador	DE (dB)	>2 dB (%)
QV-24	1,0594	3,0565
QVP-24	0,9083	2,5914
QVPC-24	0,8310	0,7979
QVPCP-24	0,8227	1,9282

práticas, é fundamental que o esquema de codificação dos parâmetros LSF também se comporte bem em canais ruidosos. Então, desenvolvemos um estudo do comportamento destas estruturas de QV em condições ruidosas. Assumimos um canal binário simétrico sem memória com probabilidade de erro de bit p_e [8]. O desempenho para todos os métodos estudados em condições ruidosas iguais, para 21 bits por quadro, em termos da DE média

está mostrado nas Tabelas 7 a 9 e na Figura 4.

Tabela 7: Desempenho em Canal Ruidoso: comparação da DE entre as estruturas de QV para diferentes taxas de erro de bits.

p_e %	QVPC	QVPCP	QVP	QV
0	1,02	0,99	1,09	1,26
0,01	1,03	1,01	1,13	1,26
0,02	1,04	1,02	1,14	1,27
0,05	1,06	1,04	1,21	1,28
0,1	1,07	1,11	1,28	1,29
0,5	1,42	1,57	1,99	1,52
1	1,87	2,04	2,83	1,81
2	2,46	2,98	3,75	2,28
5	3,89	4,55	5,90	3,53

Tabela 8: Desempenho em Canal Ruidoso: comparação dos *outliers* entre 2 e 4 dB das estruturas de QV para diferentes taxas de erro de bits.

p_e %	QVPC	QVPCP	QVP	QV
0	2,59	2,79	5,51	6,84
0,01	3,18	2,92	6,98	6,97
0,02	3,32	3,32	7,58	7,01
0,05	4,18	4,25	9,17	7,17
0,1	4,32	5,91	9,90	7,24
0,5	12,56	18,54	29,34	11,69
1	22,20	28,90	43,65	15,88
2	30,10	49,17	51,89	22,52
5	44,32	39,73	20,27	32,36

Tabela 9: Desempenho em Canal Ruidoso: comparação dos *outliers* maiores que 4 dB das estruturas de QV para diferentes taxas de erro de bits.

p_e %	QVPC	QVPCP	QVP	QV
0	0	0	0	0
0,01	0	0,06	0,13	0
0,02	0	0,13	0,13	0,13
0,05	0	0,13	0,33	0,13
0,1	0	0,60	1,33	0,53
0,5	3,00	3,12	6,84	2,72
1	6,58	7,30	19,07	6,18
2	14,62	18,87	34,95	11,69
5	32,22	55,81	79,13	28,90

Os resultados, tanto em termos da DE média como de % de *outliers*, mostram que a estrutura QVPC é superior ao esquema QV sem memória

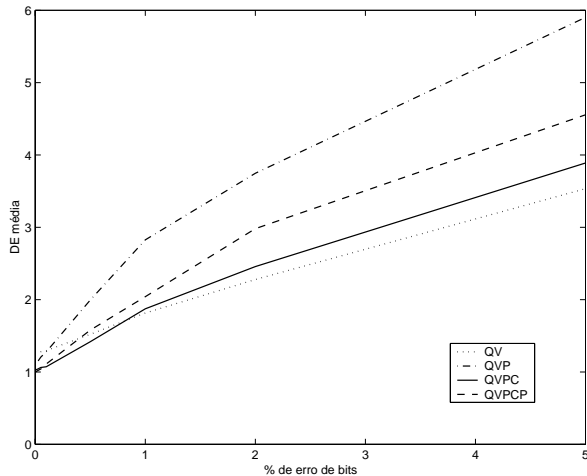


Figura 3: Desempenho das estruturas de QV para 21 bits/quadro em termos da DE média como função da taxa de erro de bits.

para $p_e \leq 0,1\%$, enquanto que para taxas de erro de bits maiores o método QV sem memória apresenta melhor desempenho. Já os esquemas que usam apenas sistemas de predição (QVP e QVPCP) são os que apresentam os piores resultados devido à propagação de erros.

6 Conclusões

Neste artigo apresentamos uma análise comparativa de quantizadores vetoriais preditivos para os parâmetros LSF da voz. O desempenho das diferentes estruturas de QV foram avaliados em canais ruidosos e em condições sem ruído. Os esquemas denominados QVPC e QVPCP apresentaram os melhores desempenhos, sendo capazes de economizar 4 bits em comparação com o QV sem memória. A avaliação dos métodos de QV em canal ruidoso revelou que o esquema QVPCP é menos robusto a erros de canal, enquanto o QVPC é capaz de propiciar um desempenho semelhante ao QV sem memória, até mesmo superando-o para $p_e \leq 0,1\%$. Note-se que o sistema QVPC é capaz de chavear para o modo QV, evitando a propagação de erros nestas situações. Entretanto, em geral, os esquemas de QV com memória perdem rapidamente a vantagem sobre os esquemas sem memória, devido à propagação de erros quando a taxa de er-

ros de bits cresce.

Referências

- [1] A. Gersho e R. M. Gray, *Vector quantization and signal compression*, Kluwer Academic Publishers, 1992.
- [2] K. K. Paliwal e B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", *IEEE Trans. Speech and Audio Processing*, vol. 1, no. 1, pp. 3-14, 1993.
- [3] M. Yong, G. Davidsson e A. Gersho, "Encoding of LPC spectral parameters using switched-adaptive interframe vector prediction", *Proc. ICASSP-88*, vol. 1, pp. 402-405, 1988.
- [4] T. Eriksson, J. Lindén e J. Skoglund, "Exploiting Interframe Correlation in Spectral Quantization: A Study of Different Memory VQ Schemes", *Proc. ICASSP-96*, 1996.
- [5] A. McCree e J. C. De Martin, "A 1.7 KB/S Melp Coder with Improved Analysis and Quantization", *Proc. ICASSP-98*, 1998.
- [6] A. Alcaim, J. A. Solewicz, e J. A. Moraes, "Frequência de ocorrência dos fonos e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro", *Revista da Sociedade Brasileira de Telecomunicações*, vol. 7, pp. 23-41, 1992.
- [7] R. Laroia, N. Phamdo e N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers", *Proc. ICASSP-91*, pp. 641-644, 1991.
- [8] S. L. Q. Dall'Agnol, A. Alcaim e J. R. B. de Marca, "Performance of LSF Vector Quantizers for VSELP Coders in Noisy Channels", *European Transactions on Telecommunications*, vol.5, No. 5, pp. 19-29, 1994.