

Efeitos de Canais com Erros na Transmissão dos Parâmetros LSF da Voz em Redes de Pacotes

Fabian D. Backx, Rodrigo C. de Lamare e Abraham Alcaim

Resumo—Os efeitos de canais com erros na transmissão dos parâmetros LSF quantizados da voz são examinados em redes de pacotes. A codificação dos parâmetros LSF do codificador MELP, onde é empregado um quantizador multiestágios sem memória com busca em árvore, é comparada à codificação das LSFs de um quantizador multiestágios com predição chaveada e busca em árvore, em canais sujeitos a erros de bits e perdas de quadros.

Palavras-Chave—Quantização das LSFs, quantizadores vetoriais multiestágio, erros nos canais, redes de pacotes.

Abstract—The effects of channel errors on the transmission of LSF parameters are examined in a packet network scenario. The LSF quantisation scheme used in the MELP coder, where a tree-structured memoryless multistage vector quantiser is employed, is compared to the LSF quantisation achieved by a switched-predictive tree-structured multistage vector quantiser, in channels subject to bit errors and frame losses.

Keywords—LSF quantisation, multistage vector quantisers, channel errors, packet networks.

I. 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 da transmissão da informação da envoltória espectral da voz, usando esquemas de quantização vetorial (QV), em presença de erros em redes de pacotes. As frequências em linhas espectrais ou parâmetros LSF foram escolhidos para representar os coeficientes LPC, uma vez que são mais adequadas para procedimentos de quantização e interpolação [1].

Os parâmetros LSF apresentam, em geral, uma correlação significativa entre quadros de voz sucessivos, especialmente para segmentos sonoros. Na quantização vetorial sem memória (QVSM) [2], [3], usada no codificador MELP [4], cada vetor de LSFs é quantizado 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. Um esquema eficiente de codificação das LSFs, onde emprega-se quantização vetorial preditiva chaveada (QVPC), foi apresentado em [5], [6].

Neste artigo, o QVSM usado no MELP é comparado ao esquema QVPC em canais sujeitos a erros de bits e perdas

de quadros. Essas condições adversas são usualmente encontradas em redes de pacotes. Aplicações importantes incluem a Internet e a telefonia móvel celular. A Seção 2 deste artigo descreve o sistema QVSM usado no MELP e a Seção 3 o esquema QVPC [5], [6]. A Seção 4 apresenta os problemas envolvidos com a transmissão de LSFs em redes de pacotes, onde as perturbações do canal de comunicações são modeladas por inserções de erros de bits e perdas de quadros. A Seção 5 apresenta e discute os resultados de simulações e a Seção 6 resume as principais conclusões do trabalho.

II. QUANTIZAÇÃO VETORIAL SEM MEMÓRIA NO MELP

A quantização vetorial sem memória (QVSM) codifica cada vetor de parâmetros LSF da voz de modo independente dos vetores de LSFs adjacentes [2]. Paliwal e Atal [3] demonstraram que um esquema QVSM multiestágios com estrutura em árvore é capaz de codificar de forma eficiente os parâmetros LSF com 24 bits por quadro de 20 ms. A estrutura do QVSM usado no MELP consiste em 4 estágios cujos dicionários têm 7, 6, 6 e 6 bits, respectivamente, totalizando 25 bits para quantizar cada vetor de 10 LSFs. Como critério de distorção foi escolhido o erro-quadrático ponderado [7]. Este critério propicia uma melhor quantização das formantes, melhorando as qualidades objetiva e subjetiva da voz codificada e é uma medida de baixa complexidade. Esta medida é dada por:

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

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

$$\alpha(\mathbf{f}) = \left\{ \begin{array}{ll} P(\mathbf{f}_i)^{0,3}, & 1 \leq i \leq 8 \\ 0, 64P(\mathbf{f}_i)^{0,3}, & i = 9 \\ 0, 16P(\mathbf{f}_i)^{0,3}, & i = 10 \end{array} \right\} \quad (2)$$

onde \mathbf{f}_i é o i -ésimo componente do vetor não quantizado das LSFs e $P(\mathbf{f}_i)$ é a potência do filtro de predição inverso na frequência \mathbf{f}_i . O procedimento de busca em árvore usa uma aproximação dos $M = 8$ melhores vetores-código, que são guardados a cada estágio do procedimento de busca.

III. QUANTIZAÇÃO VETORIAL PREDITIVA CHAVEADA

Um esquema eficiente de quantização vetorial preditiva chaveada (QVPC), desenvolvido recentemente para quantização das LSFs [5], [6], é descrito nesta seção. Esse esquema combina QVSM e quantização vetorial preditiva (QVP) para codificar quadros com alta correlação temporal de forma separada de quadros com baixa correlação temporal.

Uma busca de diferentes QVs é realizada para cada quadro e o vetor de LSFs com melhor desempenho, no que se refere a um critério de distorção, é codificado e transmitido. O sistema QVPC examinado neste trabalho é mostrado na Figura 1. Ele opera com 21 bits por quadro de 20 ms para quantização de $p = 10$ LSFs e usa 4 QVs multiestágios com busca em árvore, sendo 3 QVPs e 1 QVSM, onde cada QV possui 3 estágios. Para o QVPC, o procedimento de busca em árvore usa uma aproximação dos $M = 12$ melhores vetores-código. Com relação à alocação de bits, cada QV utiliza 7 bits no primeiro estágio, 6 no segundo e 6 no último, além de 2 bits para chavear entre 4 QVs, totalizando 21 bits por vetor de 10 LSFs. O erro quadrático-ponderado usado no QVPC, tanto para projeto dos dicionários como para realização da busca dentro de cada QV, emprega a função de ponderação [7] 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 $f_{p+1} = 0, 5$.

A seleção do melhor QV, dentre os 4 usados neste sistema, é feita com base na distorção espectral definida 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. A base de dados para treinamento dos quantizadores foi dividida em 4 subconjuntos de acordo com a DE entre vetores LSF consecutivos. Os preditores usados nos esquemas QVP foram projetados com base nos coeficientes de correlação obtidos para cada subconjunto de vetores LSF de treinamento.

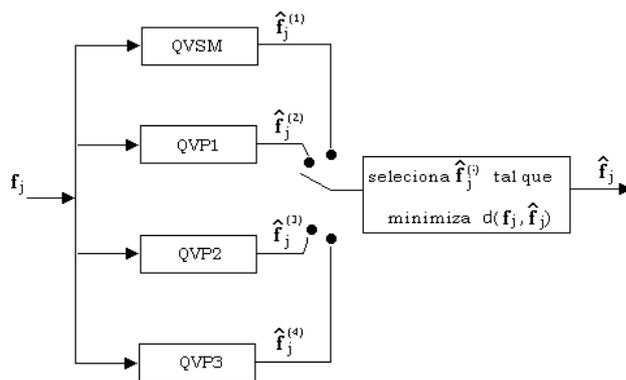


Fig. 1. Sistema QVPC com 3 QVPs e 1 QVSM.

IV. TRANSMISSÃO DAS LSFs COM ERROS EM REDES DE PACOTES

Ultimamente, o crescente interesse da aplicação de Voz sobre IP (VoIP) tem estimulado a utilização de sistemas de codificação a baixas taxas em terminais ligados a redes de pacotes (como ATM, por exemplo [8]), sobre as quais pode trafegar voz digitalizada. Ao longo desse artigo, as PDUs

(“Packet Data Units”) das camadas 2 e 3 do modelo OSI (“Open System Interconnection”) serão referenciadas pela palavra *pacote*, ficando reservado o uso da palavra *quadro* para a seqüência de bits na qual é mapeado um conjunto de parâmetros LSF pelo quantizador.

Uma rede de pacotes pode ser modelada por um remetente e um destinatário ligados por um canal de comunicações. No remetente, um conjunto de parâmetros LSF é quantizado e codificado a cada quadro. No caso do MELP, cada quadro contém uma palavra-código de 7 bits seguida de três palavras-código de 6 bits. Já para o quantizador descrito na Seção 3, cada conjunto de parâmetros LSF é mapeado em um quadro contendo uma palavra-código de 2 bits, uma de 7 bits e duas de 6 bits. Cada seqüência de conjuntos de parâmetros LSF codificados está associada a uma seqüência de quadros, que por sua vez é caracterizada por uma seqüência de bits. Essa última é encapsulada em pacotes, de acordo com o mecanismo de transmissão da rede, e enviada pelo canal, que pode introduzir erros. No destinatário, o enquadramento é desfeito pelo decodificador e os parâmetros LSF quantizados são recebidos.

Numa abordagem inicial, o canal de comunicações foi modelado por um típico canal binário simétrico (*Binary Symmetric Channel-BSC*) onde um símbolo (0 ou 1) da seqüência de bits tem probabilidade p_e de ser recebido com erro na saída do canal. O próximo passo foi levar em consideração a possibilidade de congestionamento na rede de pacotes.

Em uma rede de pacotes, havendo congestionamento, poderá ocorrer a situação de estouro do buffer (*buffer overflow*) nos *switches* ou roteadores levando ao descarte ou perda de pacotes. Nesse caso, não se tem mais um canal BSC, mas sim um canal com perda, ou apagamento, de quadro (*Frame Erasure-FE*). Supõe-se, sem perda de generalidade, que exatamente um quadro do quantizador é encapsulado em um pacote. Para observar o desempenho dos quantizadores neste tipo de canal foi utilizado o modelo descrito a seguir: cada pacote (e por conseguinte, cada quadro) tem probabilidade p_q de ser perdido, admitindo-se que a perda ou não de um quadro independe da situação (perda ou não) dos quadros anteriores. Todavia, o congestionamento da rede não é a única causa para descarte de pacotes. De fato, existem esquemas de recuperação de erro baseados somente na distinção entre quadros *bons* e quadros *apagados*. Se for detectado que ocorreu erro no quadro recebido, este deve ser tratado como se tivesse sido descartado pela rede. Para codificadores padronizados, já foram identificados quais os bits mais sensíveis e, portanto, passíveis de algum esquema de proteção [9].

Para combater os efeitos de perda de quadro, existem algumas contramedidas. A primeira e mais simples delas é silenciar (*muting*) a saída de áudio enquanto não se estiver recebendo as LSFs perdidas por causa do descarte de quadros. Entretanto, pode-se obter um melhor desempenho se ao invés do *muting* for realizada a repetição (*repeating*) do último quadro de parâmetros LSF recebido. Uma solução um pouco mais atraente e adotada neste artigo foi a interpolação dos parâmetros LSF. Devido às suas propriedades, as LSFs podem ser interpoladas linearmente para formar um conjunto válido de parâmetros LSF. Seja \mathbf{f}_i o conjunto de parâmetros LSF

quantizados recebidos no quadro i . Considere o caso em que as L seguintes LSFs são perdidas e f_{i+L+1} é recebida. Então, as LSFs interpoladas $f_{i+1} \dots f_{i+L}$ são dadas por:

$$f_{i+x} = \frac{L+1-x}{L+1} \times f_i + \frac{x}{L+1} \times f_{i+L+1} \quad (5)$$

para $x = 1 \dots L$. Assim sendo, recebe-se, por exemplo, um conjunto de parâmetros LSF#1 e, por causa das imperfeições do canal, deixa-se de receber as LSF#2 e recebe-se as LSF#3. A interpolação permite que se obtenha uma aproximação das LSF#2 às custas de um pequeno retardo adicional. Nota-se que é possível realizar a interpolação de mais de um conjunto de parâmetros LSFs às custas de um retardo cada vez maior [10], [11]. De acordo com [11], se n quadros consecutivos de duração t cada um, são perdidos, então o atraso devido à interpolação D_i é dado por

$$D_i = n \times t + RTT/2 \quad (6)$$

onde *Round Trip Time* (RTT) é o tempo que um pacote leva para ir e voltar de um destinatário (valores típicos para RTT são de 10-700ms). Ainda segundo [11], atrasos aceitáveis para aplicações de VoIP devem ser menores que 800ms.

V. RESULTADOS DE SIMULAÇÕES

Os desempenhos foram aferidos utilizando-se uma coleção de 1505 conjuntos de parâmetros LSF, correspondendo a seqüências de $1505 \times 25 = 37625$ bits no caso do MELP e $1505 \times 21 = 31605$ bits no caso do QVPC. Foram usados como critérios de avaliação objetiva a distorção espectral média explicitada em (4) e a distribuição dos *outliers*. São definidas duas classes de *outliers*: porcentagem de LSFs quantizadas com distorção espectral entre 2 e 4 dB e porcentagem de LSFs quantizadas com distorção espectral superior a 4 dB.

A primeira simulação visa contemplar a situação em que o canal de comunicações é modelado como um BSC. O desempenho dos dois quantizadores foi simulado para probabilidades de erro de bit (PEB) iguais a: 0%, 0.25%, 0.50%, 1%, 2%, 5%, 10% e 15%. Os gráficos de distorção espectral versus probabilidade de erro de bit e porcentagem de outliers versus probabilidade de erro de bit são mostrados nas Figuras 2 e 3 respectivamente. Nota-se na Figura 2 que para pequenas probabilidades de erro de bit (menores do que 0.50%) o quantizador vetorial da Seção 3 (aqui referido como QVPC) apresenta um desempenho um pouco superior ao do MELP. Para uma probabilidade de erro de bit de 0.50%, o desempenho dos dois quantizadores é comparável. Para todo o resto da faixa de probabilidade de erro de bit considerada, o MELP tem desempenho superior. Por outro lado, a observação da Figura 3 mostra o melhor desempenho do MELP em termos de porcentagem de outliers acima de 4 dB para probabilidades de erro de bit (PEB) acima de 2%. Esse sistema, entretanto, é inferior ao QVPC em termos de porcentagem de outliers entre 2 e 4 dB nessa faixa de PEB.

A seguir, foi simulado o modelo de perda ou descarte de quadros, abordado na Seção 4, para os seguintes valores de taxa de perda de quadro (TPQ): 0%, 5%, 10%, 15%, 20% e 25%. Os mesmos critérios objetivos foram mantidos. A apresentação dos resultados é feita nas figuras 4 e 5 que

contêm, respectivamente, os gráficos de distorção espectral versus taxa de perda de quadro e porcentagem de *outliers* versus taxa de perda de quadro. Das figuras 4 e 5 observa-se que o desempenho do quantizador QVPC é superior ao do MELP para toda a faixa de taxa de perda de quadros considerada. O ganho de qualidade (segundo o critério de distorção espectral média) é de aproximadamente 0.25 dB em toda a faixa de taxa de perda de quadro considerada. O melhor desempenho do quantizador proposto em [5] é confirmado pela Figura 5: para toda a faixa de interesse, a quantidade de *outliers* com distorção espectral maior do que 4 dB do MELP é 4 pontos percentuais maior do que para os *outliers* de mesma classe para o outro quantizador. A quantidade de *outliers* com distorção espectral entre 2 e 4 dB é comparável para os dois quantizadores.

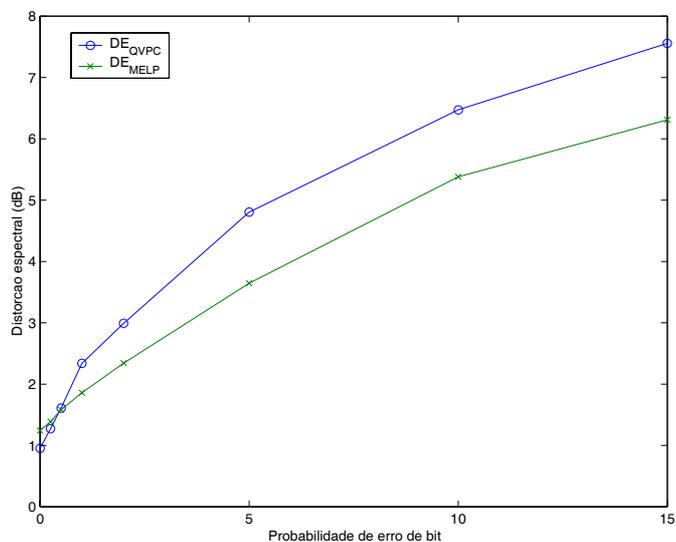


Fig. 2. Desempenho em termos de DE para PEB.

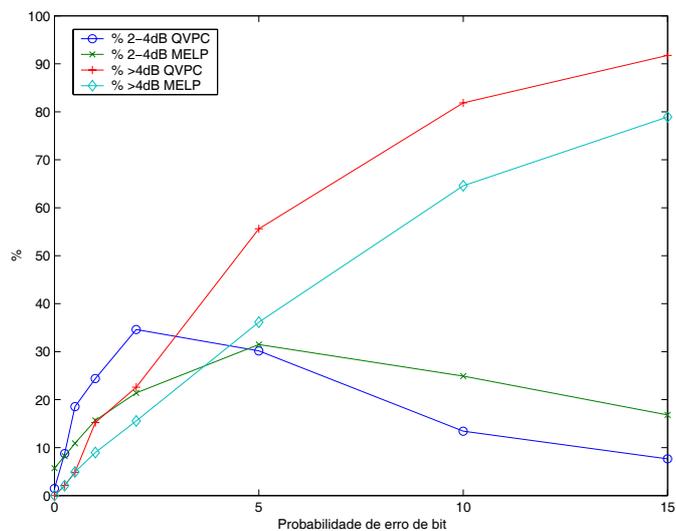


Fig. 3. Desempenho em termos de outliers para PEB.

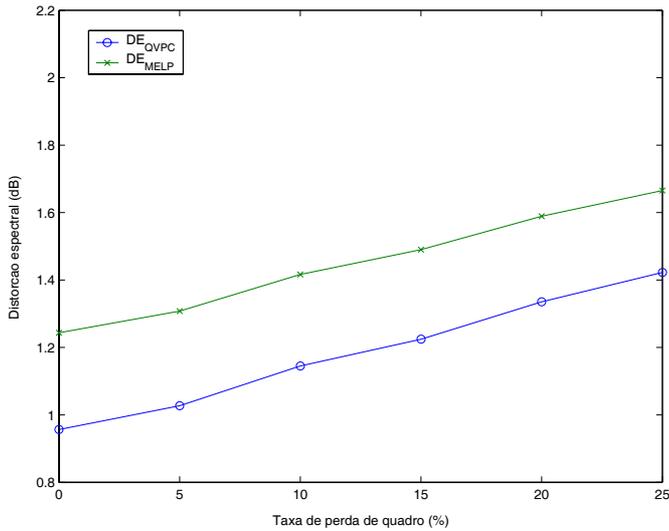


Fig. 4. Desempenho em termos de DE para TPQ.

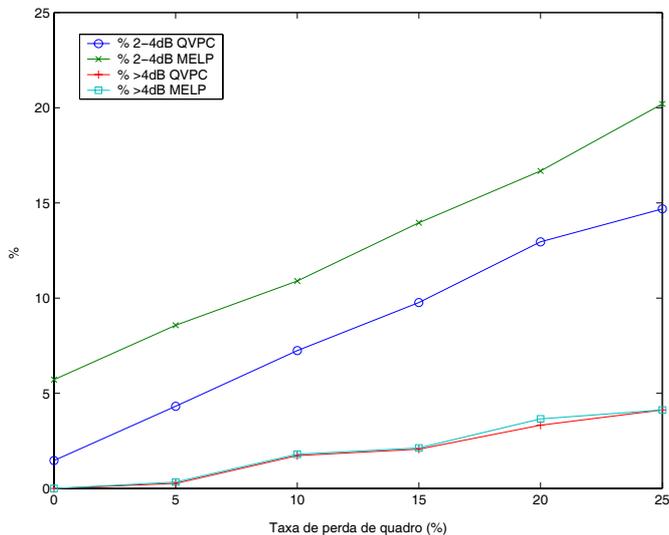


Fig. 5. Desempenho em termos de outliers para TPQ.

VI. CONCLUSÕES

Neste trabalho foram examinados os efeitos de erros de bits e perdas de quadros na transmissão dos parâmetros LSF em redes de pacotes. Duas estruturas distintas de quantização vetorial foram consideradas. Uma que emprega 25 bits/vetor, adotada no MELP, e outra que usa 21 bits/vetor, proposta em [5] e denominada QVPC. Como o QVPC utiliza esquemas preditivos, ele apresenta resultados de desempenho inferiores ao MELP em canais com probabilidade de erro de bit (PEB) elevadas. Já quando se considera perdas de quadros, onde os quadros perdidos (ou descartados) são interpolados na recepção, o QVPC é sempre superior ao MELP, independentemente da taxa de perdas de quadros considerada. Esses resultados são devidos à superioridade do QVPC de 21 bits, quando comparado ao MELP de 25 bits em canais sem erro, e ao emprego da estratégia de interpolação entre quadros recebidos.

REFERÊNCIAS

- [1] F. K. Soong e B.H. Juang, "Line spectrum pair (LSP) and speech data compression," Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 1984.
- [2] A. Gersho e R. M. Gray, Vector quantization and signal compression, Kluwer Academic Publishers, 1992.
- [3] 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.
- [4] L. M. Supplee, R. P. Cohn, J. S. Collura e A. V. McCree, "MELP: the new federal standard at 2400 bps", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 1591-1594, 1997.
- [5] R. C. de Lamare e A. Alcaim, "Analysis of LSF switched-predictive vector quantisers," Proc. International Symposium on Signal Processing and its Applications, Kuala-Lumpur, Malaysia, 2001.
- [6] R. C. de Lamare e A. Alcaim, "Noisy channel performance of LSF switched-predictive vector quantisers", *Proc. IEEE International Conference on Information, Communications and Signal Processing*, Singapore, 2001.
- [7] R. Laroia, N. Phamdo e N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 641-644, 1991.
- [8] D. Rahikka, J. Collura, T. Fuja, D. Sridhara e T. Fazel, "Error coding strategies for MELP vocoder in wireless and ATM environments," Proc. IEE Seminar on Speech Coding for Algorithms for Radio Channels, 2000.
- [9] N. R. Sollenberger, N. Seshadri, e R. Cox, "The evolution of IS-136 TDMA for third-generation wireless services", *IEEE Personal Communications*, vol. 6, pp. 8-18, Junho 1999.
- [10] E. Daniel e K. Teague, "Federal standard 2.4 kbps MELP over IP", *Proc. 43rd IEEE Midwest Symp. on Circuits and Systems*, 2000.
- [11] J. Wang e J. Gibson, "Parameter interpolation to enhance the frame erasure robustness of CELP coders in packet networks", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2001.