Codificação Perceptual de Áudio usando Quantização Retro-Adaptativa

João Manuel Rodrigues, Ana Maria Perfeito Tomé

Resumo – **This paper presents a new coding/compression algorithm for high-quality digital audio signals. The new coder comprehends a nonuniform filter bank, gain-adaptive logarithmic quantizers, arithmetic coding, and an explicit psychoacoustic model that adapts the quantizers according to perceptual criteria. The new system differs from other perceptual coders in that its quantizers are backward-adaptive: the adaptation is not based on the original signal but on the output signal. We discuss the advantages of backward-adaptation and its application to perceptual coding. Some results are presented.**

Resumo – **Neste artigo apresenta-se um novo algoritmo de codificação/compressão de sinais áudio digitais de alta qualidade. O novo codificador inclui um banco de filtros com bandas não uniformes, quantizadores logarítmicos adaptativos, codificação aritmética e um modelo psico-acústico explícito para adaptar a quantização de acordo com critérios perceptuais. O que distingue este sistema de outros codificadores perceptuais é que a quantização é retro-adaptativa: a adaptação baseia-se no sinal reconstruído e não no sinal original. Discutem-se as vantagens da retro-adaptação e a sua aplicabilidade à codificação perceptual. Apresentam-se alguns resultados obtidos.**

I. INTRODUÇÃO

Nos últimos anos tem havido uma grande actividade de investigação e desenvolvimento de métodos de codificação digital de sinais áudio de alta qualidade e sinais de voz de banda larga. Foi notado que os métodos tradicionais de codificação de fonte usados com sucesso em sinais de voz de banda estreita não resultam satisfatoriamente com áudio, e que para conseguir alta qualidade com débitos muito baixos é necessário explorar as limitações da percepção auditiva humana [1]. Isto levou ao desenvolvimento de sistemas de *codificação perceptual*: codificadores que aplicam conhecimentos de Psico-acústica de forma a eliminar ou reduzir o mais possível a audibilidade da distorção introduzida no processo de codificação. Actualmente a maioria dos codificadores perceptuais de áudio existentes seguem um esquema de codificação adaptativa no domínio da frequência: um banco de filtros ou uma transformada seguida de quantizadores adaptativos controlados por um algoritmo de adaptação perceptual dependente do próprio sinal a codificar. Alguns destes sistemas conseguem codificação "perceptualmente transparente" com cerca de 2 bits por amostra [2]. Apesar destes resultados, continua a busca de novos algoritmos que possam reduzir ainda mais o débito, o atraso de codificação e/ou a complexidade do sistema.

Este trabalho foi parcialmente financiado pela JNICT, bolsa de estudos número BM/2979/92-IA.

Neste artigo começamos por introduzir os princípios gerais de compressão de sinais digitais e da codificação perceptual de áudio em particular (Secção II). Na Secção III descrevemos um novo algoritmo de codificação perceptual de sinais áudio cuja maior diferença em relação aos codificadores existentes é a utilização de adaptação para trás, isto é: a adaptação depende do sinal já codificado e não do sinal a codificar. A aplicabilidade desta abordagem e as suas vantagens, entre as quais se contam uma menor complexidade de projecto e implementação e uma adaptação mais rápida e precisa, são discutidas na Secção IV. A Secção V descreve a metodologia de avaliação do codificador proposto e resume os resultados obtidos.

II. CODIFICAÇÃO E CRITÉRIOS PERCEPTUAIS

De uma forma muito geral pode dizer-se que a compressão de sinais digitais assenta sobre dois princípios fundamentais: a remoção de redundância e a redução de irrelevância. A remoção de redundância consiste na exploração das propriedades estatísticas dos sinais produzidos pela fonte, tais como a previsibilidade e a distribuição não uniforme de amplitudes das amostras. Um código PCM linear como o usado nos discos compactos não explora estas propriedades e em consequência é bastante redundante. Essa redundância pode, contudo, ser eliminada usando codificação preditiva e códigos de comprimento variável, por exemplo. Com um modelo da fonte completo e preciso é possível remover muita redundância e obter bons factores de compressão. O outro princípio de compressão—a redução de irrelevância—é muitas vezes deixado para segundo plano e, no entanto, ele pode ser muito importante como veremos adiante. A irrelevância está relacionada com a incapacidade de o receptor final detectar a distorção introduzida no processo de codificação. A filtragem das componentes de frequência superior a 20 kHz de um sinal áudio é um exemplo de redução de irrelevância: poupam-se bits não transmitindo componentes imperceptíveis para o ouvido humano. Naturalmente que para explorar eficazmente a irrelevância do sinal é necessário um bom modelo da percepção do receptor final.

Johnston e Brandenburg [1] observaram que as técnicas tradicionais de codificação de sinais, que exploram essencialmente a remoção de redundância, não permitem grandes desempenhos quando aplicadas aos sinais áudio. As razões disto são diversas. Por um lado, os sinais áudio são difíceis de modelar devido à sua vasta diversidade, grande gama dinâmica e largura de banda. Além disso, a expectativa de qualidade por parte do ouvinte é muito maior no caso do áudio do que noutros tipos de sinais. Por outro lado, o erro

médio quadrático, a relação sinal-ruído e outras medidas de distorção simples usadas na optimização dos codificadores tradicionais são modelos fracos da percepção auditiva humana, impossibilitando uma efectiva redução de irrelevância. Deste modo, sem um bom modelo da fonte de áudio, a solução passa necessariamente pela inclusão no codificador de conhecimento sobre fenómenos de percepção psico-acústica. Esta é a essência da *codificação perceptual de áudio*.

Os fenómenos psico-acústicos mais relevantes para a codificação perceptual são o limiar absoluto de audição, que traduz a diferente sensibilidade do ouvido a tons de diferentes frequências, e os efeitos de mascaramento no tempo (pré- e pós-mascaramento) e na frequência (mascaramento simultâneo). O mascaramento consiste na redução ou mesmo eliminação da audibilidade de um som (o mascarado) quando na presença de um som mais forte (o mascarante). O efeito de mascaramento é tanto maior quanto mais próximos forem os sons no tempo e na frequência. A presença do mascarante provoca uma elevação localizada do limiar de audibilidade do mascarado acima do limiar absoluto. A dependência com a frequência dos fenómenos de mascaramento simultâneo é mais facilmente descrita numa escala "natural" de frequência denominada escala Bark. Esta escala, aproximadamente logarítmica, foi revelada e confirmada por diversos estudos psico-acústicos independentes sobre as chamadas *bandas críticas*—uma banda crítica tem largura de 1 Bark qualquer que seja a sua frequência central—e mesmo por estudos fisiológicos do ouvido interno—um intervalo de 1 Bark corresponde a um comprimento fixo na membrana basilar da cóclea. Informações mais completas sobre estes fenómenos poderão ser encontradas em [3], [4] e outras referências aí citadas.

Os codificadores perceptuais procuram aproveitar estes efeitos tentando "esconder" o ruído de quantização sob o limiar de mascaramento global. Para tal, recorrem a técnicas de modelação dinâmica de ruído. Actualmente a técnica mais usada em codificadores perceptuais é a codificação adaptativa em sub-bandas ou por transformada, também chamada codificação no domínio da frequência [5]. Além de ser uma forma eficiente de explorar a previsibilidade do sinal, esta técnica proporciona um modo fácil de modelar o espectro do ruído através da variação da resolução dos quantizadores de cada banda.

O processo de codificação perceptual no domínio da frequência está representado genericamente na Figura 1. O sinal de entrada é decomposto por um banco de filtros multicadência ou uma transformada em diversas bandas de frequência devidamente subamostradas. O limiar de mascaramento é estimado a partir da representação espectral do sinal obtida pelo banco de filtros ou por outro analisador espectral em paralelo. As amostras de cada banda são quantizadas e codificadas com uma resolução e distribuição de bits condicionadas pelo limiar estimado. Por fim, as amostras codificadas são multiplexadas com a informação lateral necessária para a reconstrução do sinal no descodificador.

O descodificador, por seu lado, desmultiplexa os dados, recupera os valores quantizados das amostras de cada banda e recombina as várias bandas num banco de filtros de síntese que produz o sinal de saída.

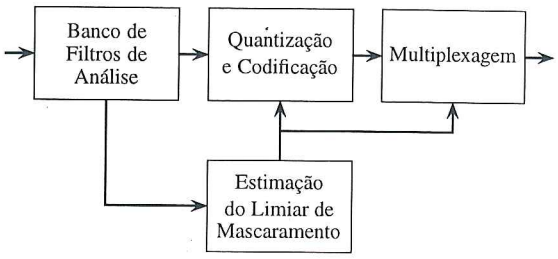

Figura 1 - Um codificador perceptual no domínio da frequência típico. (Adaptado de [1].)

III. ESTRUTURA DO CODIFICADOR

A Figura 2 representa o sistema de codificação tratado neste artigo. O sinal de entrada, amostrado a uma frequência de 44100 amostras por segundo e quantizado linearmente com 16 bits por amostra, é fornecido a um banco de filtros de análise (T) que o decompõem em 62 bandas de diferentes larguras. Um conjunto de quantizadores logarítmicos de ganho adaptativo (Q) discretiza as amostras de cada banda que de seguida são codificadas com um código de comprimento variável para transmissão ao receptor. O algoritmo de adaptação aplica um modelo psico-acústico para estimar continuamente o limiar de mascaramento a partir das amostras previamente quantizadas e estabelece os ganhos dos quantizadores de forma que o ruído de quantização resultante não ultrapasse o limiar calculado ou se mantenha dentro de uma certa vizinhança deste.

No receptor, o código é interpretado e as amostras quantizadas são recuperadas. O banco de filtros de síntese recombina as várias bandas para formar uma réplica do sinal original. Os ganhos dos desquantizadores são adaptados pelo mesmo algoritmo que no transmissor de forma que não é necessário transmitir informação adicional para manter circunstâncias iguais num e noutro extremo do canal de comunicação.

Nas próximas secções descrevemos com mais pormenor os vários blocos que compõem este sistema de codificação.

A. O Banco de Filtros

O banco de análise decompõe o sinal em bandas de largura não uniforme usando uma estrutura de dois andares. O primeiro andar é formado por uma ELT [6] que divide o sinal em 256 bandas uniformes. No segundo andar, as primeiras 32 bandas passam inalteradas para a saída, enquanto as restantes são agrupadas em grupos de duas, quatro, oito ou dezasseis, e recombinadas com ELTs inversas para produzir bandas mais largas mas com melhor resolução temporal. Note-se que este tipo de estrutura difere da estrutura em árvore mais usual que se quis evitar por razões discutidas em [3]. Os filtros protótipos das ELTs foram projectados para uma sobreposição de 75% (factor de sobreposição $K = 2$) e foram optimizados segundo o critério de minimização da energia na banda de corte, definida como $[\pi/M, \pi]$ onde M é o número de bandas. Disto resultaram protótipos com respostas semelhantes o que, como foi observado por Cox [7], é uma condição desejável para conseguir cancela-

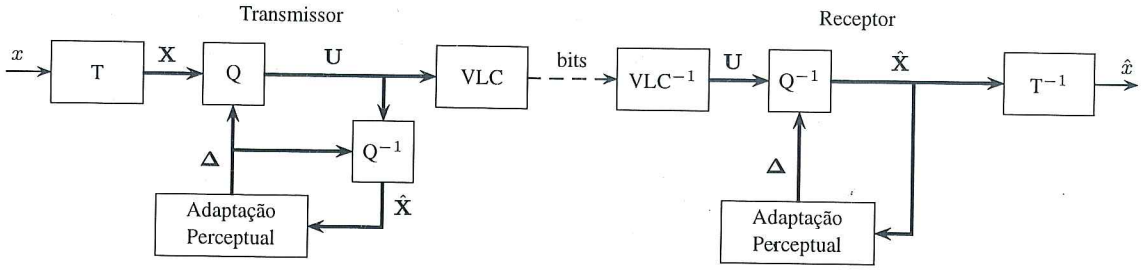

Figura 2 - Sistema de codificação perceptual com adaptação para trás.

mento parcial do *aliasing* entre os dois andares da estrutura. Todas as saídas passam por linhas de atraso que normalizam o atraso global para permitir a reconstrução perfeita no banco de síntese que tem uma estrutura dual.

A Tabela I mostra a resolução no tempo e na frequência das 62 bandas resultantes. Todas as bandas têm largura inferior a 1 Bark, o que permite um aproveitamento eficiente do fenómeno de mascaramento simultâneo. Simultaneamente, a resolução temporal é bastante boa, especialmente nas frequências mais altas, para evitar a violação das condições de mascaramento no tempo. Um banco de filtros com decomposição uniforme não poderia satisfazer ambas as condições. O uso de estruturas compostas de ELTs garante ainda as propriedades de decimação máxima e reconstrução perfeita e resulta numa complexidade computacional de apenas 21 adições e 12 multiplicações por amostras. A combinação análise-síntese introduz um atraso de 40.6 ms.

Tabela I
PERÍODO DE AMOSTRAGEM (Δt) E LARGURA (Δf E Δz) DE CADA BANDA DO BANCO DE FILTROS.

Bandas	Δt (ms)	Δf (Hz)	Δz (Bark)
1–32	5.80	86	0.85–0.19
33–40	2.90	172	0.35–0.22
41–48	1.45	345	0.42–0.26
49–54	0.73	689	0.47–0.31
55–62	0.36	1378	0.54–0.23

B. *Quantização*

Cada banda do sinal é quantizada por um quantizador independente. Para simplificar a implementação, todos os quantizadores têm 127 níveis distribuídos simetricamente segundo uma curva característica de forma fixa. Apenas o ganho ou passo de quantização Δ é variado de banda para banda e de amostra para amostra. Os quantizadores são *mid-tread*, isto é, o zero é um dos níveis de saída. A curva característica tem uma forma aproximadamente logarítmica semelhante às da lei-A ou lei-μ usadas em telefonia digital. Para amostras de amplitude moderada (em relação a Δ), o quantizador tem um comportamento semelhante ao de um quantizador uniforme de passo Δ. Para amostras de amplitude mais elevada, o passo de quantização aumenta proporcionalmente à amplitude. Isto permite uma grande gama dinâmica e proporciona também alguma modelação de ruído visto que, mesmo quando o sinal aumenta bruscamente, a relação sinal-ruído não ultrapassa os 31 dB, o que é suficiente em termos perceptuais.

C. *Codificação de Comprimento Variável*

Os quantizadores debitam cada um dos 127 níveis possíveis com diferente frequência. Para eliminar a redundância associada a esta distribuição não uniforme, os valores quantizados são codificados com um codificador aritmético adaptativo [8].

Um codificador aritmético é controlado por um modelo estatístico da sequência a codificar. Uma vez que cada banda apresenta uma distribuição de probabilidades diferente, o nosso sistema mantém 62 tabelas de probabilidades distintas e comuta entre elas à medida que processa os símbolos de cada banda. Assim, o codificador aritmético integra simultaneamente a função de multiplexagem dos dados das várias bandas. A adaptação do modelo é feita após a codificação de cada amostra, actualizando os contadores de ocorrências na tabela adequada.

D. *Algoritmo de Adaptação Perceptual*

O algoritmo de adaptação baseia-se num modelo psico-acústico explícito que estima continuamente o valor do limiar de mascaramento. O modelo é semelhante ao usado em [9] e compreende os seguintes passos:

1. Espalhamento da energia no tempo. As amostras de cada banda passam por um rectificador de lei quadrática e por um filtro passa-baixo recursivo de primeira ordem. As constantes de tempo dos filtros variam de banda para banda, modelando a dependência com a frequência do pós-mascaramento. Este passo foi inspirado pelo modelo descrito em [10].
2. Espraiamento na frequência. Faz-se uma convolução com uma função de espraiamento na frequência para modelar o mascaramento simultâneo. Este passo é implementado por uma multiplicação por uma matriz de dimensão 256×256 felizmente bastante esparsa.
3. Dedução do índice de mascaramento. De cada banda subtrai-se, numa escala logarítmica, o índice de mascaramento de ruído por tons (TMN) para obter a estimativa do limiar de mascaramento. O índice de mascaramento de tons por ruído (NMT) não foi considerado nesta versão do codificador.
4. Correcção para o limiar absoluto. O limiar de mascaramento calculado no passo anterior é comparado com o limiar absoluto de audição e é corrigido nos pontos em que lhe for inferior.

Os passos de quantização são então determinados de forma que a potência de ruído que vão introduzir não ultrapasse o limiar calculado. (A potência de ruído de quantização é estimada por $N = \Delta^2/12$, que se verificou ser uma aproximação razoável em operação normal.) Os passos são ainda multiplicados por um parâmetro global ϕ—chamado *nível de qualidade*—que permite controlar o compromisso qualidade/compressão: $\phi > 1$ aumenta os passos de quantização, degradando a qualidade mas conseguindo maior compressão; $\phi < 1$ diminui os passos, garantindo uma "margem de segurança" abaixo do limiar à custa de um débito mais elevado. Na versão actual, o parâmetro ϕ é transmitido apenas no início de um sinal mas no futuro poderá vir a ser transmitido mais regularmente de forma a possibilitar dinamicamente a satisfação de requisitos de qualidade ou débito.

Devido à estrutura multiresolução da saída do banco de filtros, o algoritmo de adaptação é executado incrementalmente, em alternância com a quantização, em 16 fases sincronizadas pelas amostras das bandas mais altas que têm o período de amostragem menor. Com o banco de filtros utilizado e este algoritmo evita-se a análise espectral paralela e as múltiplas conversões entre diferentes partições do plano tempo-frequência que são necessárias noutros codificadores, como os recomendados pelo MPEG [11]. O algoritmo de adaptação completo pode ser implementado com cerca de 70 operações aritméticas por amostra.

IV. Adaptação para Trás num Codificador Perceptual de Áudio

Uma característica particular do sistema proposto é que a estimação do limiar de mascaramento e logo a adaptação dos passos de quantização se faz a partir das amostras já quantizadas. Trata-se portanto de quantização retro-adaptativa que, embora seja usada em codificadores tradicionais como o G.722 do CCITT [12], não é comum em codificadores perceptuais de áudio. Nesta secção discutimos as vantagens e inconvenientes do uso da técnica de adaptação para trás em codificação de sinais e verificamos a sua aplicabilidade à codificação perceptual de áudio.

A. Vantagens e Inconvenientes

A consequência imediata do uso de adaptação para trás é a eliminação da necessidade de transmissão de informação lateral para adaptação, visto que os passos de quantização são gerados localmente tanto no transmissor como no receptor. Num codificador com adaptação para a frente, pelo contrário, os parâmetros de adaptação têm que ser transmitidos ao receptor, ocupando uma fracção considerável da capacidade disponível do canal. Não tendo que quantizar, codificar e multiplexar esta informação lateral, um codificador retro-adaptativo tem um algoritmo mais simples e é mais fácil de projectar. Outra consequência é que os passos de quantização podem ser adaptados amostra-a-amostra enquanto que num sistema adaptado para a frente a adaptação só é feita entre blocos de várias amostras de forma a minimizar a quantidade de informação lateral gerada.

A maior desvantagem de um sistema de codificação retro-adaptativo é o seu comportamento na presença de erros de transmissão. Após a ocorrência de um erro, o receptor perde o sincronismo com o transmissor e o efeito do erro propaga-se por tempo indeterminado. Existem formas de minimizar este problema, tais como a reinicialização periódica dos parâmetros adaptáveis ou o uso dos chamados *factores de fuga* (*leakage factors*) [13]. Contudo, no sistema que desenvolvemos não adoptámos qualquer destas estratégias, pelo que, nesta fase, o sistema não é robusto na presença de erros de transmissão. Outra potencial desvantagem dos sistemas adaptados para trás é a maior complexidade computacional do receptor devido à necessidade de inclusão do algoritmo de adaptação. Em aplicações tipo difusão, onde há um grande número de receptores, esta característica pode ser economicamente inconveniente se o algoritmo de adaptação, e particularmente o modelo psico-acústico, for excessivamente complexo. Por outro lado, em aplicações de comunicação bidireccional *half-duplex* como num leitor/gravador digital de áudio ou na compressão/descompressão de ficheiros de áudio digital, não se levanta qualquer problema visto que a operação de descodificação pode aproveitar integralmente os recursos computacionais já disponíveis para a operação de codificação.

B. Aplicabilidade

Como se referiu atrás, no codificador perceptual de áudio apresentado neste artigo a estimação do limiar de mascaramento baseia-se em amostras previamente quantizadas. Consequentemente, por muito exacto que seja o modelo psico-acústico usado, a estimativa do limiar vem sempre afectada pelos erros de quantização dessas amostras. Além disso, como essas amostras foram por sua vez quantizadas com base num limiar estimado anteriormente, o novo limiar vem afectado indirectamente pelos erros de estimação do limiar prévio e de todos os anteriores.

Inevitavelmente, levanta-se a questão da possibilidade de esta realimentação dos erros de quantização comprometer irremediavelmente todo o processo, invalidando por completo o uso de retro-adaptação num codificador perceptual. Uma breve reflexão sobre o comportamento do sistema tanto em condições normais como em condições extremas fornece-nos bons argumentos para refutar essa hipótese:

1. O limiar calculado em cada ponto no tempo e na frequência depende da energia do sinal em muitos outros pontos no passado e em bandas adjacentes. Logo, é pouco provável que os erros de quantização se conjuguem para afectar o limiar num mesmo sentido. Por outras palavras: as operações de espalhamento temporal e espectral *filtram* os erros de quantização.
2. Se o sinal numa banda diminui muito abaixo do respectivo passo de quantização, passa a ser quantizado com o nível zero, o que provoca uma tendência de decaimento rápido no limiar e consequentemente, no passo de quantização. Esse decaimento respeita as condições de pós-mascaramento e só é interrompido quando o passo diminui suficientemente abaixo do nível do sinal ou quando o limiar "encosta" ao limiar imposto pelas outras bandas ou ao limiar absoluto, como é desejável. A característica *mid-tread* dos quan-

tizadores é um factor determinante para este comportamento. Quantizadores *mid-rise* resultariam num decaimento mais lento ou mesmo num crescimento instável dos passos de quantização.

3. Se o sinal crescer acima do ponto de saturação do quantizador, a sua energia é subavaliada, obrigando o limiar a aumentar mais lentamente que o desejável. Isto conduz a uma sobrecodificação desnecessária do sinal. A grande gama dinâmica e a modelação intrínseca de ruído proporcionada pelos quantizadores logarítmicos permitem minorar muito este problema.

Assim, temos razões para concluir que a estimação do limiar de mascaramento não é grandemente afectada pelo erro de quantização, especialmente se houver o cuidado de usar quantizadores *mid-tread* com boa gama dinâmica. Para o confirmar, realizámos uma experiência muito simples: usando um sinal representativo como entrada, registámos o limiar "exacto" Ψ_0 calculado a partir das amostras antes da quantização, e o limiar "perturbado" Ψ_ϕ calculado a partir do sinal quantizado com nível de qualidade ϕ. Os limiares foram então comparados fazendo o histograma da razão Ψ_ϕ/Ψ_0. A Figura 3 mostra o resultado obtido para $\phi = 4$. Mesmo a este nível, correspondente a uma quantização muito grosseira, quase 60% dos valores estimados para o limiar "perturbado" não se desviam mais que 0.25 dB em relação ao limiar "exacto". Para $\phi = 2$ e $\phi = 1$, essa fracção cresce para 76% e 93%, respectivamente.

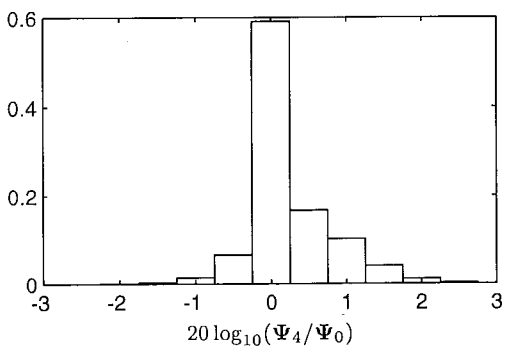

Figura 3 - Histograma das diferenças (em dB) entre o limiar estimado a partir de amostras quantizadas grosseiramente ($\phi = 4$) e o limiar estimado a partir de amostras não quantizadas.

Resta ainda lembrar que num sistema adaptado para a frente também nunca podemos aplicar o limiar "exacto" visto que, antes de serem transmitidos ao receptor, os passos de quantização têm de ser quantizados e decimados para limitar a quantidade de informação lateral. O MPEG *Layer* I—exemplo típico de um codificador adaptado para a frente—só transmite um factor de escala (proporcional ao passo de quantização) por cada 12 amostras, e o seu valor é quantizado com uma resolução de 2 dB. Nestas condições, é perfeitamente admissível que um sistema retro-adaptativo consiga efectivamente produzir melhores estimativas do limiar de mascaramento que um sistema com adaptação para a frente.

V. AVALIAÇÃO DO CODIFICADOR

O sistema de codificação perceptual de áudio com adaptação para trás (BAPAC) foi implementado em *software* com um programa escrito em MATLAB e C. Para avaliar o seu desempenho, codificaram-se sete trechos musicais—extraídos na sua maioria do disco compacto EBU SQAM [14]—em três níveis de qualidade decrescente: $\phi = 1$, $\phi = 2$ e $\phi = 3$. Cada trecho foi ainda codificado com uma implementação disponível em *shareware* do MPEG Layer III a 64 kbit/s. As quatro versões codificadas de cada trecho foram avaliadas em termos do débito produzido e da qualidade subjectiva medida em testes de audição.

Os testes de audição inspiraram-se na metodologia de teste de *estímulo triplo* com *referência escondida* que foi usada com bons resultados nos testes realizados no âmbito do MPEG e do CCIR [2]. Em cada teste eram apresentados três sinais ao ouvinte: R, X e Y. O sinal R era sempre o trecho original para ser usado como sinal de referência. Um de X e Y, escolhido aleatoriamente pelo computador, era uma das quatro versões codificadas enquanto o outro era uma cópia da referência R. O ouvinte podia escutar os sinais repetidamente e pela ordem que entendesse. A sua tarefa consistia em classificar a degradação percebida de cada um dos sinais X e Y em relação à referência R, atribuindo uma pontuação tirada da escala de degradação de 5 pontos do CCIR. Participaram dez pessoas nos testes de audição. Cada ouvinte completou, por uma ordem aleatória, dois testes de cada uma das quatro versões codificadas dos sete trechos.

A Tabela II mostra a pontuação média obtida por cada versão codificada dos vários trechos. Também se apresenta a média das pontuações ou *Mean Opinion Score* (MOS) e o débito médio obtido por cada codificador.

Tabela II

RESULTADOS DOS TESTES DE AVALIAÇÃO: PONTUAÇÕES MÉDIAS DE CADA VERSÃO CODIFICADA, MOS E DÉBITO MÉDIO (EM BITS POR AMOSTRA).

	BAPAC $\phi = 1$	BAPAC $\phi = 2$	BAPAC $\phi = 3$	Layer III 64 kb/s
Castanholas	4.20	3.85	3.70	4.20
Cravo	4.30	3.45	2.55	4.30
Sarasate	4.60	3.75	2.40	4.75
Sting	4.75	4.65	4.30	4.70
Stravinsky	4.85	4.50	3.90	4.40
Suzanne	3.00	1.85	1.40	3.25
Violino	3.00	1.65	1.20	3.40
MOS	4.10	3.39	2.78	4.14
Débito	2.35	1.78	1.46	1.42

A 2.35 bits por amostra, o algoritmo proposto permite uma codificação de alta qualidade. No entanto, para um débito comparável ao do codificador de Layer III, apresenta uma qualidade bastante inferior. Isto pode dever-se, em parte, a uma adaptação demasiado lenta do codificador aritmético que é inicializado com tabelas optimizadas para a situação $\phi = 1$. Alguns trechos obtiveram consistentemente pontuações baixas em todas as versões, o que indicia deficiências no modelo psico-acústico.

VI. CONCLUSÕES

Foi apresentado um novo algoritmo de codificação de áudio que aplica quantização retro-adaptativa e respeita critérios perceptuais. O sistema é muito simples e pode ser implementado com cerca de 110 a 120 operações aritméticas por amostra. Os testes realizados revelam alta qualidade de codificação com débitos médios abaixo dos 100 kbit/s.

Mostrou-se que o erro de quantização introduzido no sinal não perturba significativamente a estimativa do limar de mascaramento. Este resultado sugere que um sistema retro-adaptado amostra-a-amostra pode permitir um melhor acompanhamento do limiar de mascaramento real do que um sistema adaptado para a frente bloco-a-bloco.

As características discutidas na Secção IV mostram que a retro-adaptação é uma técnica a ter em conta pelo menos em determinadas aplicações de codificação perceptual de áudio. Estamos convictos que os defeitos apontados, particularmente a falta de robustez a erros de transmissão, poderão ser superados sem grandes custos adicionais. É plausível que uma solução híbrida, com alguma informação lateral para corrigir os parâmetros retro-adaptados, possa trazer grandes benefícios e alargar o leque de aplicações possíveis.

Estão a ser estudadas diversas modificações ao sistema proposto no sentido de melhorar o seu desempenho. Uma dessas modificações consiste na utilização de um banco de filtros simplificado, com divisão uniforme.[1] Apesar de se perder resolução temporal nas bandas mais altas, a adaptação amostra-a-amostra deve permitir evitar os problemas que surgem noutros codificadores existentes que nem têm resolução espectral tão boa. Outras modificações em estudo são um algoritmo de adaptação mais rápido para o codificador aritmético e modelos psico-acústicos mais elaborados.

REFERÊNCIAS

[1] James D. Johnston e Karlheinz Brandenburg, "Wideband coding—perceptual considerations for speech and music", em *Advances in Speech Signal Processing*, Sadaoki Furui e M. Mohan Sondhi, Eds., capítulo 4. Marcel Dekker, Inc., New York, 1991.

[2] Peter Noll, "Wideband speech and audio coding", *IEEE Communications Magazine*, pp. 34–44, Nov. de 1993.

[3] João Manuel de Oliveira e Silva Rodrigues, "Compressão digital de sinais Áudio aplicando critérios perceptuais e adaptação para trás", Master's thesis, Departamento de Electrónica e Telecomunicações da Universidade de Aveiro, Aveiro, Nov. de 1995.

[4] Aníbal João Sousa Ferreira, "Codificação perceptual de Áudio digital estereofónico", Master's thesis, Faculdade de Engenharia da Universidade do Porto, Porto, Jan. de 1992.

[5] José M. Tribolet e Ronald E. Crochiere, "Frequency domain coding of speech", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, pp. 512–530, Out. de 1979.

[6] Henrique S. Malvar, *Signal Processing with Lapped Transforms*, Artech House, Norwood, MA, 1992.

[7] Richard V. Cox, "The design of uniformly and nonuniformly spaced pseudoquadrature mirror filters", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, pp. 1090–1096, Out. de 1986.

[8] Ian H. Witten, Radford M. Neal, e John G. Cleary, "Arithmetic coding for data compression", *Communications of the Association for Computing Machinery*, vol. 30, no. 6, pp. 520–540, Jun. de 1987.

[9] James D. Johnston, "Transform coding of audio signals using perceptual noise criteria", *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314–323, Fev. de 1988.

[10] John G. Beerends e Jan A. Stemerdink, "A perceptual audio quality measure based on a psychoacoustic sound representation", *Journal of the Audio Engineering Society*, vol. 40, no. 12, pp. 963–978, Dez. de 1992.

[11] ISO/IEC JTC1/SC29/WG11 (MPEG), *ISO/IEC CD 11172: Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s*, 1992.

[12] Xavier Maitre, "7 kHz audio coding within 64 kbit/s", *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 283–298, Fev. de 1988.

[13] Nuggehally S. Jayant e Peter Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice-Hall, Englewood Cliffs, N.J., 1984.

[14] European Broadcasting Union, Brussels, *Sound Quality Assessment Material: Recordings for Subjective Tests*, Abr. de 1988.

[1] Um sistema com esta alteração está a ser implementado num processador digital de sinal por alunos finalistas da cadeira de projecto do curso de Engenharia Electrónica e Telecomunicações no ano lectivo de 1995/96.