



Relatório

Trabalho 2 - Decomposição de Sinais Áudio em Bandas (MP3)

Unidade Curricular:
Processamento Audiovisual (PA)

Licenciatura em Engenharia e Ciência de Dados

Realizado por:

Diogo Beltran Dória, 2020246139
Mariana Lopes Paulino, 2020190448

Ano Letivo 2022/2023

Introdução

O objetivo deste trabalho prático é a execução de experiências que envolvem a decomposição de um sinal de áudio em várias bandas e quantização dos sinais resultantes dessa decomposição recorrendo a scripts MATLAB e várias operações decorrentes do processamento de sinal durante a codificação do áudio em bandas com codecs nomeadamente MP1 e MP3.

Ao nosso dispor temos uma biblioteca com vários áudios que nos é possível utilizar. Após escolhermos um áudio temos de o processar, nomeadamente de o transformar em mono uma vez que este nos é apresentado em estéreo. Para fazer isto apenas consideramos uma das bandas do sinal que terá de ser cortado com o comprimento do primeiro áudio para efeitos de comparação.

Experiências Realizadas

Primeiramente foi-nos fornecido um sinal áudio “violin.wav” no qual corremos o áudio conseguindo ouvi-lo e processamos o mesmo e realizámos algumas operações de modo a obter o número de amostras que o áudio tem e o seu comprimento para que depois consigamos saber com que comprimento cortar o segundo áudio para melhor comparação.

Após esta análise inicial ao áudio do violino, carregamos o ficheiro de áudio “David Bowie - Blackstar.wav” e cortámos apenas o comprimento equivalente ao do áudio do violino, neste caso com 1,9998 segundos. Em seguimento ao corte do comprimento do áudio, tornamos esse corte do áudio em mono considerando apenas um canal do áudio.

As experiências realizadas ao áudio do violino resultaram nos seguintes gráficos:

3.2.3 32-channel pseudo-QMF filter bank

Construímos agora um banco de filtros PQMF de 32 canais, conforme implementado na norma MPEG-1 Layer-I, e verificamos a sua capacidade de reconstrução.

Primeiro construímos o banco de filtros a partir do filtro protótipo e examinamos as respostas de frequência de magnitude de todos os filtros (Fig. 1, à esquerda). Claramente, os filtros PQMF não são filtros passa-banda ideais com largura de banda = 689 Hz, mas a sua resposta se sobrepõe apenas com dois filtros vizinhos. A largura de banda total de cada filtro é de cerca de 2×689 Hz. Observamos que o ganho do filtro é de 15 dB, ou seja, $20 \cdot \log_{10}(32)/2$. Como resultado, passa duas vezes pelo filtro que produz um ganho de 32, que compensa a decimação em 32.

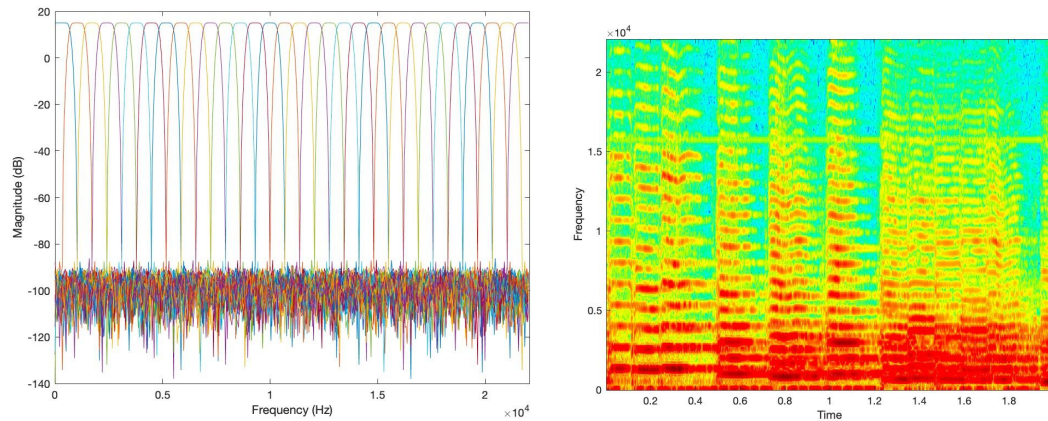


Fig. 1 *Esquerda:* Frequência de Magnitude da Resposta da Análise dos 32 filtros de um banco de filtros PQMF
Direita: Espectrograma de um sinal de áudio que contém algumas notas tocadas no violino

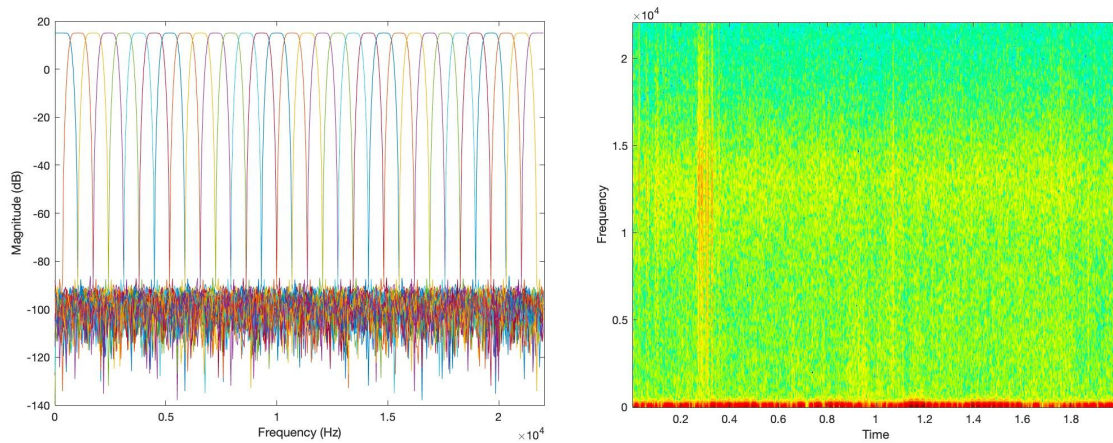


Fig. 2 *Esquerda:* Resposta Magnitude/Frequência da aplicação dos 32 filtros de análise de um banco de filtros PQMF; *Direita:* Espectrograma de um sinal de áudio com notas do áudio escolhido previamente (excerto da música “Blackstar”)

Vamos agora verificar a saída do banco de filtros PQMF quando alimentado com 2 segundos de sinal monofônico de violino amostrado em 44,100 Hz (Fig. 1, à direita). Conforme revelado pela reprodução da sub-banda 3, os sinais isolados da sub-banda são muito aliased, porque cada filtro PQMF não é ideal (Fig. 2, à esquerda).

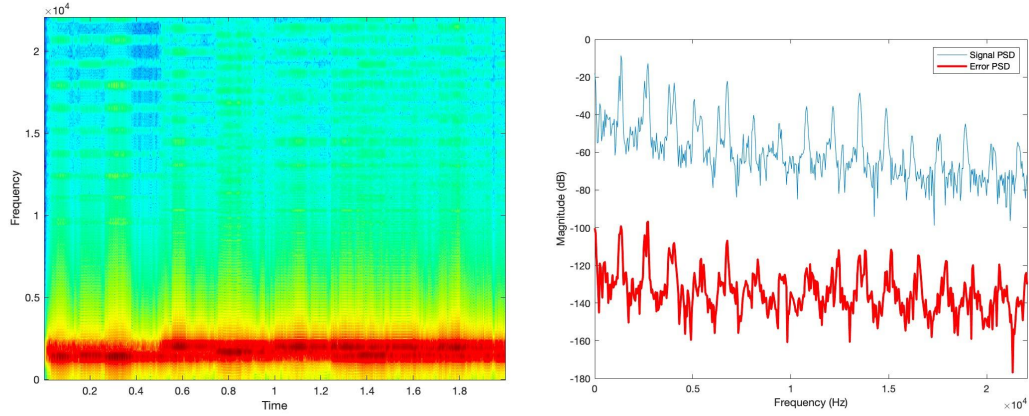


Fig. 3 *Esquerda*: Espectrograma da sub-banda #3; *Direita*: Periodograma do sinal original e de erro por volta da amostra 11000 no violino

$$\text{snr PQMF} = 84.2443$$

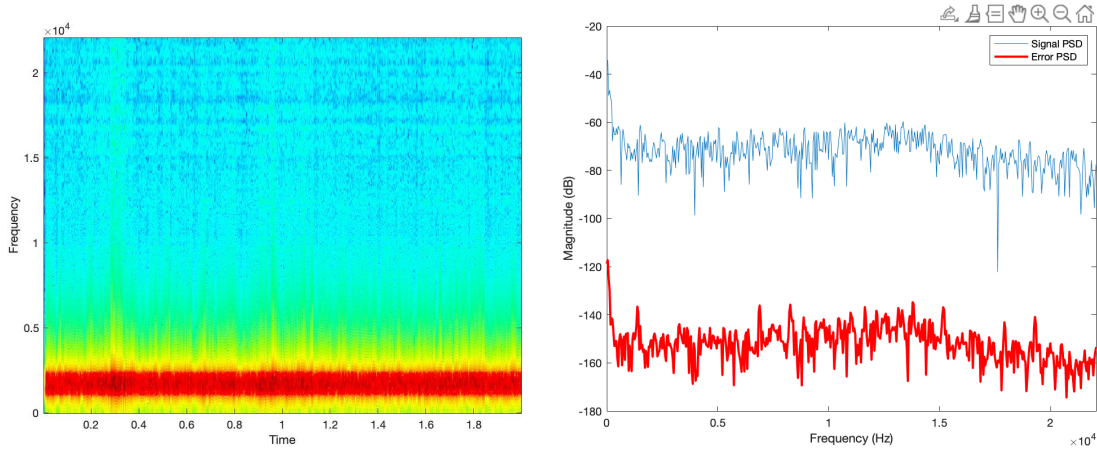


Fig. 4 *Esquerda*: Espectrograma da sub-banda #3; *Direita*: Periodograma do sinal original e de erro no sinal no áudio escolhido previamente (excerto da música “Blackstar”)

$$\text{snr PQMF} = 87.1114$$

O banco de filtros PQMF, no entanto, garante que o aliasing em bandas adjacentes seja cancelado quando sub-bandas são adicionadas. A potência do erro de reconstrução é cerca de 85 dB abaixo da potência do sinal (Fig. 2, à direita). Observamos que a saída está atrasada em 511 amostras (uma vez que os filtros de análise e síntese têm um atraso de 511/2 amostras).

3.2.4 Filter banks and lapped transforms

Agora mostramos como os filtros de análise e síntese podem ser implementados usando block transforms.

Quando o comprimento dos filtros usados nos filtros de sub-banda de canal M são exatamente M amostras, essas operações se reduzem a uma transformada linear de quadros de amostra M sucessivos não sobrepostos do sinal de entrada. Uma DFT de quatro amostras, por exemplo, pode implementar

um banco de filtros de quatro canais cujas respostas de impulso do filtro de sub-banda são o tempo invertido das linhas da matriz 4×4 da DFT.

Este tipo de banco de filtros não é muito seletivo em frequência, conforme mostrado na Fig. 5. A sobreposição de frequência entre bandas adjacentes é cerca de metade da passagem de banda do lóbulo principal (como na seção anterior sobre PQMF), mas os lóbulos laterais são muito altos. Isso não o torna um bom candidato para codificação de sub-banda.

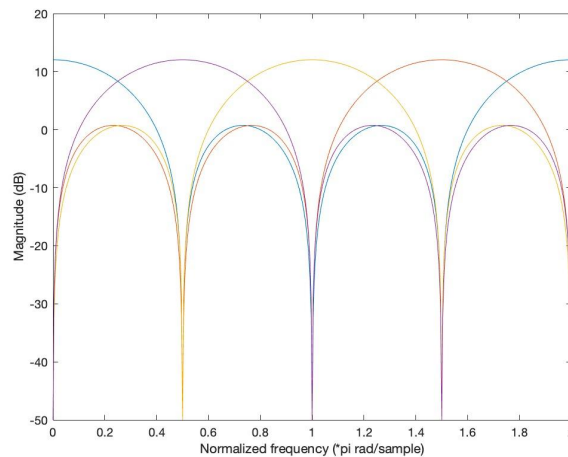


Fig. 5 Frequência de Resposta da magnitude dos quatro canais de um banco de filtros DFT. Estas respostas de frequência não têm dois lados no violino

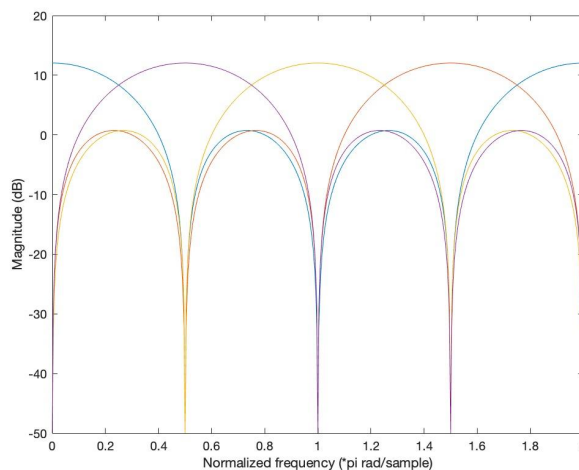


Fig. 6 Frequência de Resposta da magnitude dos quatro canais de um banco de filtros DFT. Estas respostas de frequência não têm dois lados no áudio escolhido previamente (excerto da música “Blackstar”)

Em geral, o comprimento das respostas ao impulso dos filtros de análise e síntese usados nos codificadores de sub-banda é maior que o número M de canais. As operações de filtragem, no entanto, ainda podem ser implementadas como a multiplicação de quadros L -amostra com matrizes $L \times M$ ou $M \times L$. Por exemplo, o banco de filtros PQMF de 32 canais introduzido na seção anterior, no qual o comprimento N da resposta ao impulso de cada filtro é de 512 amostras, pode ser eficientemente implementado da seguinte maneira (o que é muito semelhante à implementação de nosso exemplo

anterior Banco de filtros baseado em DFT, com adição de sobreposição). Obviamente, obtemos os mesmos resultados de antes (Fig. 7).

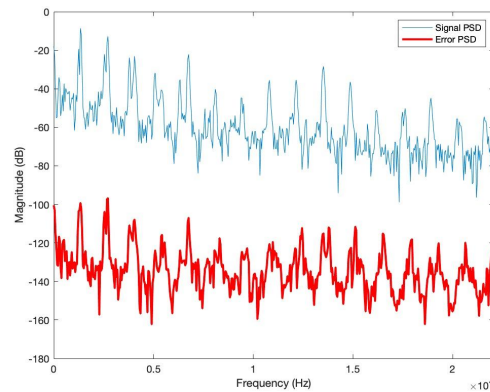


Fig. 7 Periodograma do sinal original e do de erro no violino por volta da amostra 11000

snr lapped 84.2436

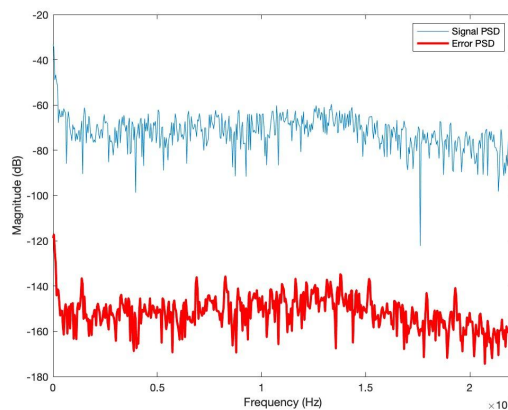


Fig. 8 Periodograma do sinal original e do de erro no áudio escolhido previamente (excerto da música “Blackstar”)

snr lapped = 87.1139

3.2.5 Perceptual audio coding

O processo de filtragem de sub-banda desenvolvido nas seções anteriores transforma o fluxo original de amostras na frequência de amostragem F_s em 32 sub-bandas paralelas amostradas em $F_s/32$. Este ainda precisa de uma etapa de quantização e codificação para produzir a compressão geral da taxa de bits. A quantização uniforme de amostras de sub-banda 11 não permite muita transparência em taxas de bits baixas, conforme mostrado no teste de amostra de quatro bits por sub-banda abaixo (fator de compressão = 4). O sinal de saída resultante é degradado; seu espectrograma é dado na Fig. 9.

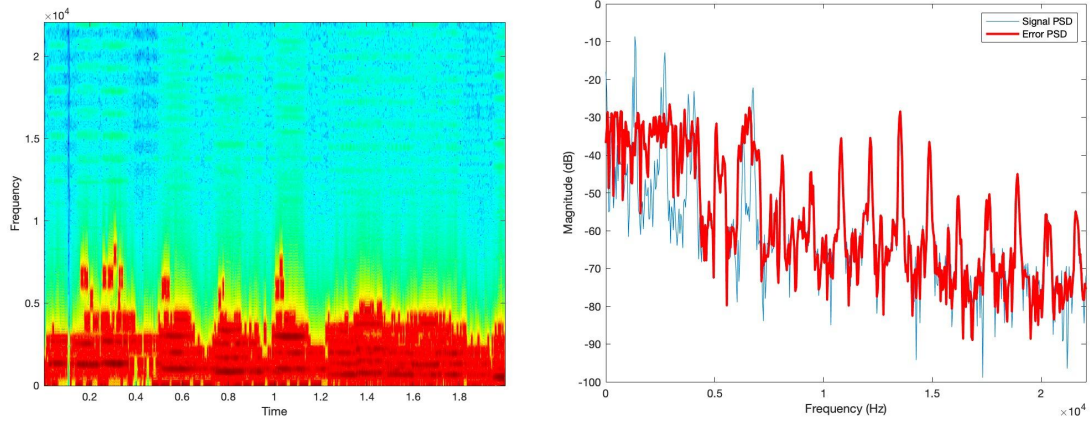


Fig. 9 Codificação da sub-banda utilizando uma quantização uniforme fixa de quatro bits *Esquerda*: Espectrograma; *Direita*: Periodograma do sinal original e do erro computada por volta da amostra 11000 onde são quase idênticas tirando frequências baixas no áudio do violino

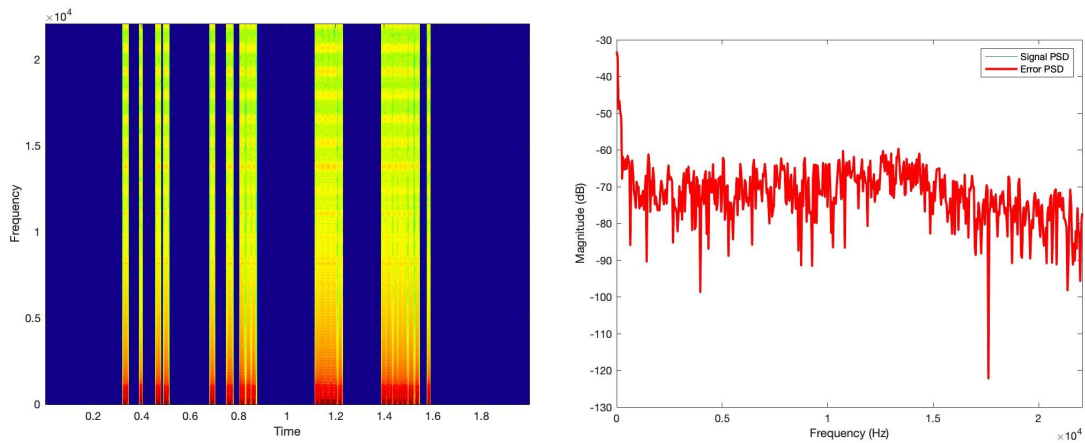


Fig. 10 Codificação da sub-banda utilizando uma quantização uniforme fixa de quatro bits *Esquerda*: Espectrograma; *Direita*: Periodograma do sinal original e do erro onde são quase idênticas tirando frequências baixas no áudio escolhido previamente (excerto da música “Blackstar”)

O sinal de saída está fortemente distorcido. O SNR cai para 10,3 dB. O mesmo acontece no nosso excerto de áudio, onde o SNR cai para 2,2 dB.

Pode-se ver que esta faixa fixa do quantizador $[-1, +1]$ não considera adequadamente a variação do nível do sinal da sub-banda através das sub-bandas, bem como no tempo (Fig. 11, à esquerda).

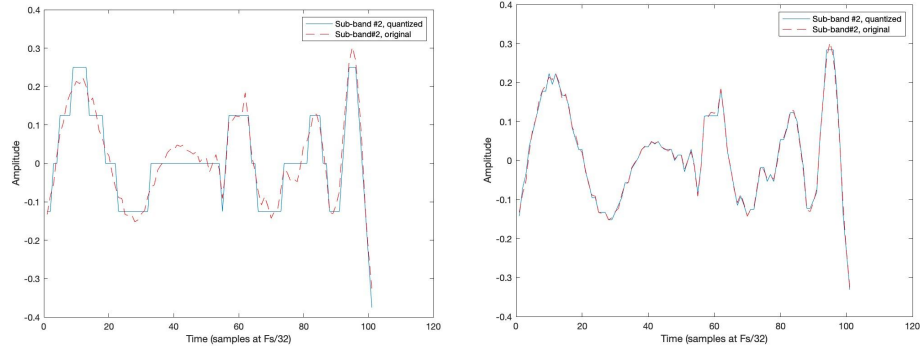


Fig. 11 Excerto da sub-banda #2 antes e depois da quantização. *Esquerda*: Utilizando uma quantização fixa uniforme de quatro bits; *Direita*: Utilizando uma quantização de nível adaptativo de quatro bits uniforme no sinal do violino.

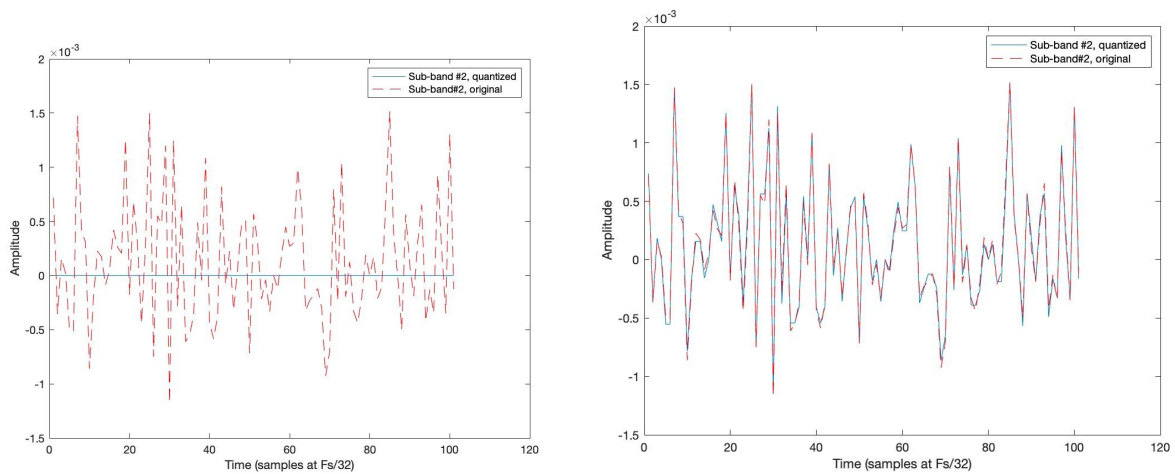


Fig. 12 Excerto da sub banda #2 antes e depois da quantização. *Esquerda*: Utilizando uma quantização fixa uniforme de quatro bits; *Direita*: Utilizando uma quantização de nível adaptativo de quatro bits uniforme no áudio escolhido previamente (excerto da música “Blackstar”)

Um meio óbvio de aumentar sua qualidade é, portanto, aplicar um fator de escala a cada quantizador de sub-banda. Como no codificador MPEG-1 Layer-I, calculamos um novo fator de escala a cada amostra de 12 sub-bandas (ou seja, a cada $32 \times 12 = 384$ amostras na taxa de amostragem original). Os erros de quantização resultantes são muito reduzidos (veja a Fig. 11, à direita).

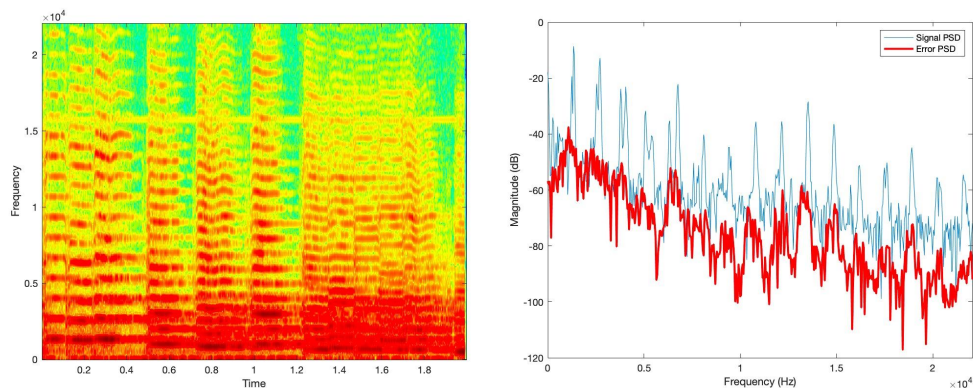


Fig. 13 Codificação da sub banda utilizando uma quantização uniforme de quatro bits adaptativa
Esquerda: Espectrograma; *Direita:* Periodograma do sinal original e do de erro por volta da amostra 11000 do sinal do violino

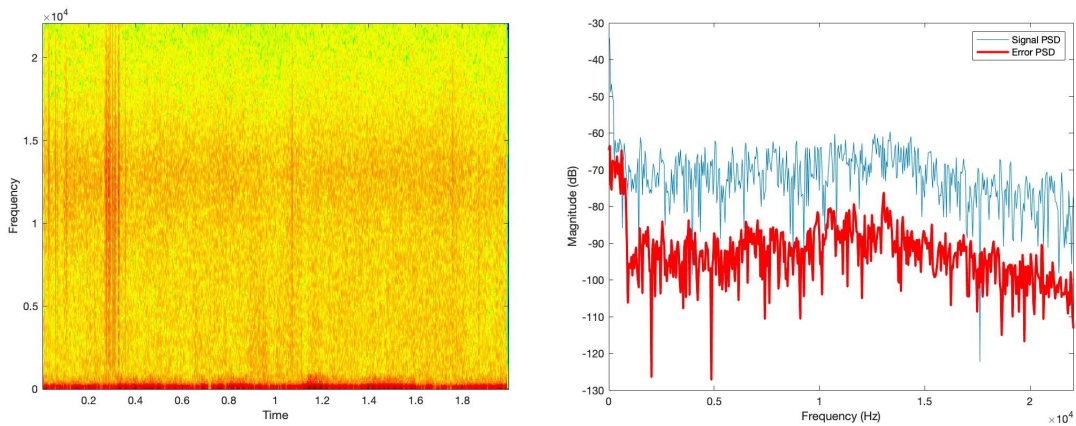


Fig. 14 *Esquerda:* Espectrograma; *Direita:* Periodograma do sinal original e do de erro no áudio escolhido previamente (excerto da música “Blackstar”)

O sinal resultante agora tem uma qualidade muito maior, conforme mostrado na Fig. 13. O SNR geral aumentou para 25 dB. Para a nossa experiência no nosso excerto aconteceu o mesmo onde o SNR passou a 27 dB aproximadamente.

O refinamento final, de longe o mais eficaz sendo resultado de anos de pesquisa de áudio, consiste em aceitar mais ruído de quantização (alocando menos bits e, portanto, aceitando um SNR mais baixo) em faixas de frequência onde não será ouvido e usando esses bits para bandas mais perceptivamente proeminentes. A informação perceptiva necessária é fornecida por um modelo psicoacústico.

Para qualquer quadro de 512 amostras retirado do sinal de entrada, o modelo psicoacústico MPEG-1 Layer-I calcula um limiar de mascaramento global, obtido pela primeira detecção de mascaradores tonais e de ruído proeminentes separadamente e combinando seus limiares individuais. O máximo deste limiar global e o limiar auditivo absoluto é então tomado como o limiar final (Fig. 15, à esquerda).

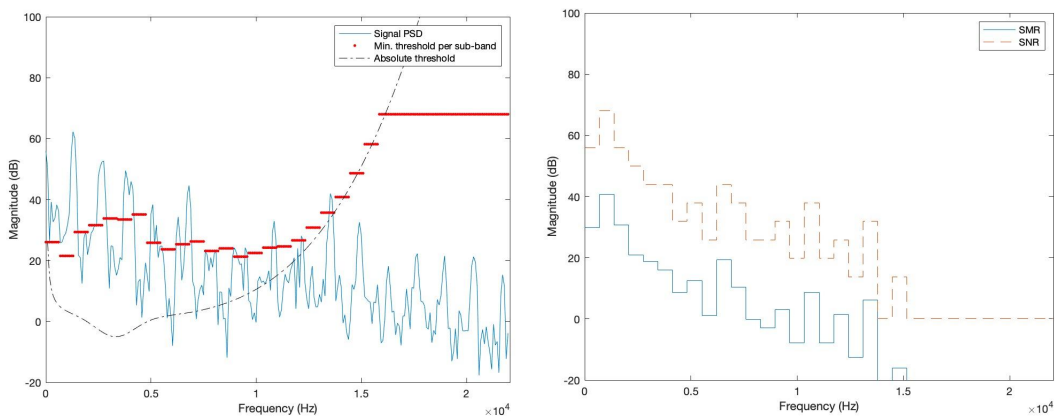


Fig. 15 MPEG-1 utilizando o modelo psicoacústico. *Esquerda*: Masking threshold por sub-banda; *Direita*: Rácios Sinal-Máscara (SMR) e o máximo SNR obtido após a alocação de bits no sinal de áudio do violino.

A SNR está sistematicamente acima do SMR de modo a que o ruído da quantização seja inaudível.

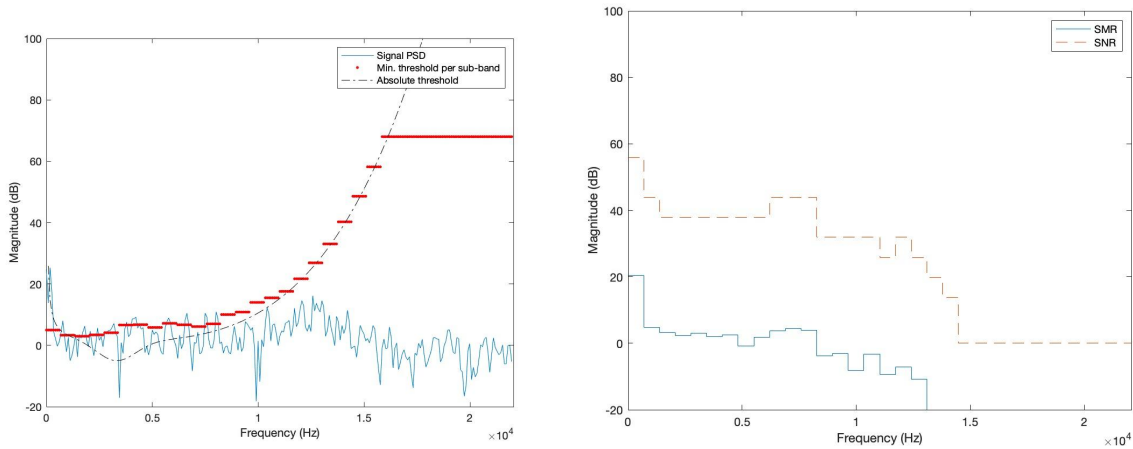


Fig. 16 MPEG-1 utilizando o modelo psicoacústico. *Esquerda*: Masking threshold por sub-banda; *Direita*: Rácios Sinal-Máscara (SMR) e o máximo SNR obtido após a alocação de bits no sinal de áudio escolhido previamente (excerto da música “Blackstar”)

Aqui mais uma vez a SNR está sistematicamente acima do SMR de modo a que o ruído da quantização seja inaudível.

As Signal-to-mask ratios (SMR) são calculadas para cada banda de uma forma muito conservadora, como a razão entre o máximo do sinal PSD e o mínimo do limite de mascaramento em cada banda. A alocação de bits é realizada por um algoritmo iterativo, que dá prioridade às sub-bandas com maior SMR. O SNR resultante em cada sub-banda deve ser maior ou igual ao SMR para empurrar o nível de ruído abaixo do limite de mascaramento (Fig. 15, à direita).

Vamos finalmente testar isso no sinal completo, adicionando alocação de bits perceptuais à quantização uniforme adaptativa. Observe que simplificamos o esquema de quantização aqui, comparado ao MPEG-1, considerando a quantização com qualquer número de bits no intervalo [0...16]. O ruído de quantização agora é muito pequeno em algumas sub-bandas proeminentes como a sub-banda #2 (Fig. 17, à esquerda). Também é mais importante em outras sub-bandas como a sub-banda #20 (Fig. 17, à direita).

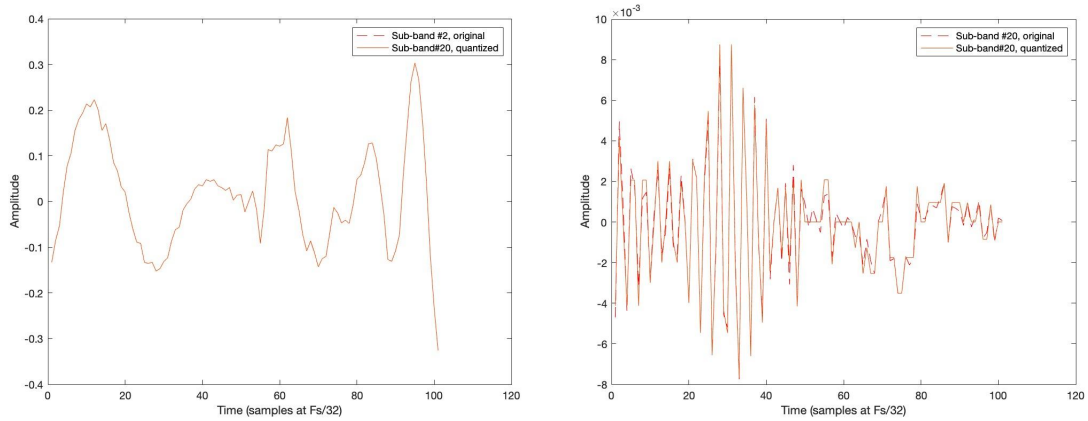


Fig. 17 Excertos das sub-bandas #2 e #20 do sinal de áudio do violino

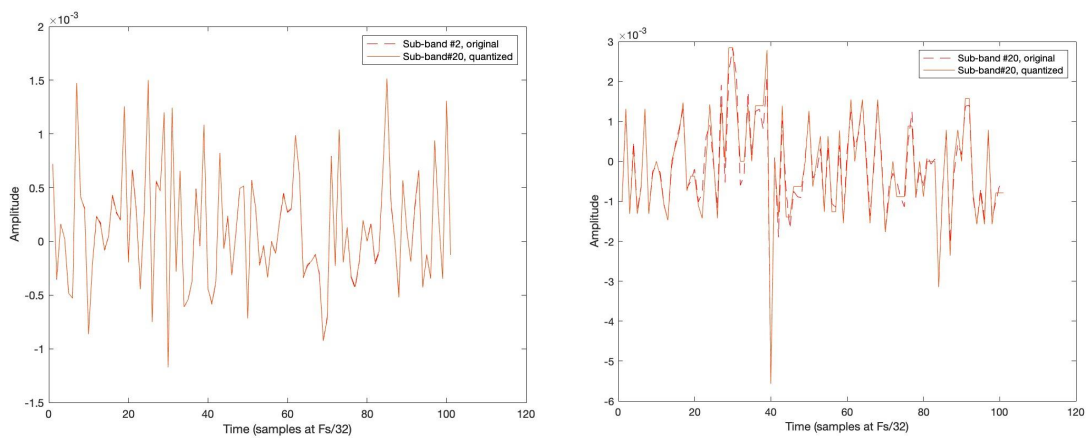


Fig. 18 Excertos das sub-bandas #2 e #20 do áudio escolhido previamente (excerto da música “Blackstar”)

O SNR geral aumentou para 38.6457 dB. O SNR perceptual é realmente muito maior, já que a maior parte do ruído não pode ser ouvida. O sinal resultante é um áudio de alta qualidade (Fig. 19).

Já na nossa segunda experiência o SNR apenas subiu para 28.3550 dB que já não transmite uma diferença muito grande no nosso áudio, onde mantemos a qualidade no entanto não é tão alta como na do violino.

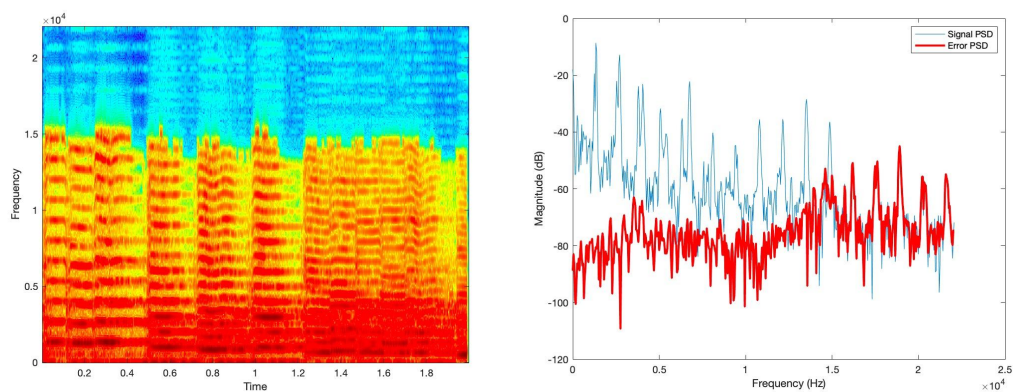


Fig. 19 Codificação da sub-banda usando quantização adaptativa perceptual uniforme. *Esquerda:* Espectrograma; *Direita:* Periodograma do sinal original e do sinal de erro do sinal do violino

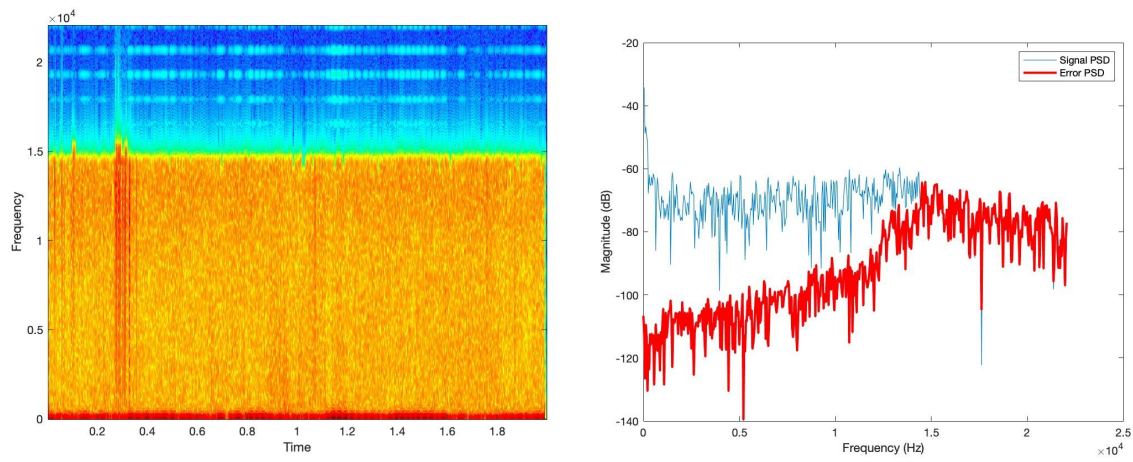


Fig. 20 Codificação da sub-banda usando quantização uniforme perceptual adaptativa *Esquerda:* Espectrograma; *Direita:* Periodograma do sinal original e do sinal de erro no áudio escolhido previamente (excerto da música “Blackstar”)