



C209 – Computação Gráfica e Multimídia
EC215 – Multimídia

Sistema de TV Digital

Parte 3: Processos de Compressão

Marcelo Vinícius Cysneiros Aragão

marcelovca90@inatel.br

Introdução

- O sinal de vídeo digitalizado constitui uma **taxa de bits elevada**, que pode superar 1Gbps, não podendo ser transmitido na largura de banda de 6 MHz reservada aos canais de TV.
- Os processos de modulação empregados nos sistemas ATSC, DVB-T e ISDB-T permitem transmitir um nível com taxa de bits de aproximadamente 20 Mbps nesta banda.
- Assim sendo, tornou-se necessária a redução da taxa de bits de 1 Gbps para 20 Mbps, a fim de que fosse possível transmitir esse sinal na banda de 6MHz.
- Esse processo de redução da taxa de bits é conhecido por “**compressão**”.

Introdução

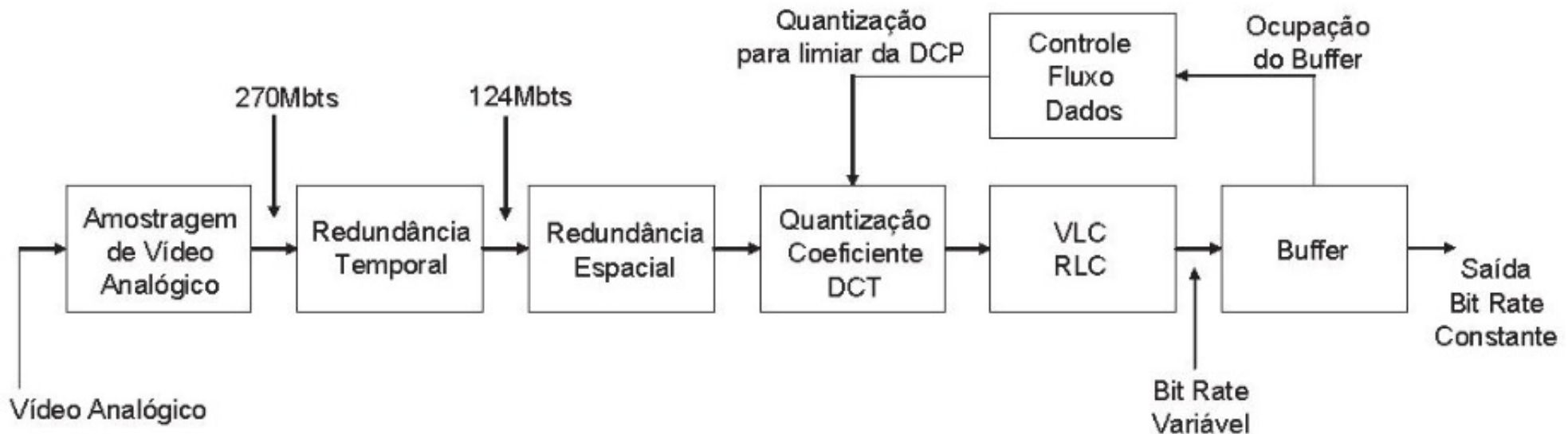
- Em 1986, um grupo de estudo denominado Joint Photograph Expert Group (**JPEG**), sob o apoio da International Standard Organization (ISO), foi criado visando desenvolver e padronizar um método de **compressão de sinal digitalizado de imagens estáticas**.
- Anos mais tarde decidiu-se criar um método de compressão de **imagens dinâmicas** baseado no princípio JPEG.
- Esse grupo foi denominado Moving Pictures Expert Group (**MPEG**) e deu origem ao padrão de compressão de vídeo conhecido como MPEG2, utilizado pelos três sistemas de TV digital ATSC, DVB-T e ISDB-T.

Introdução

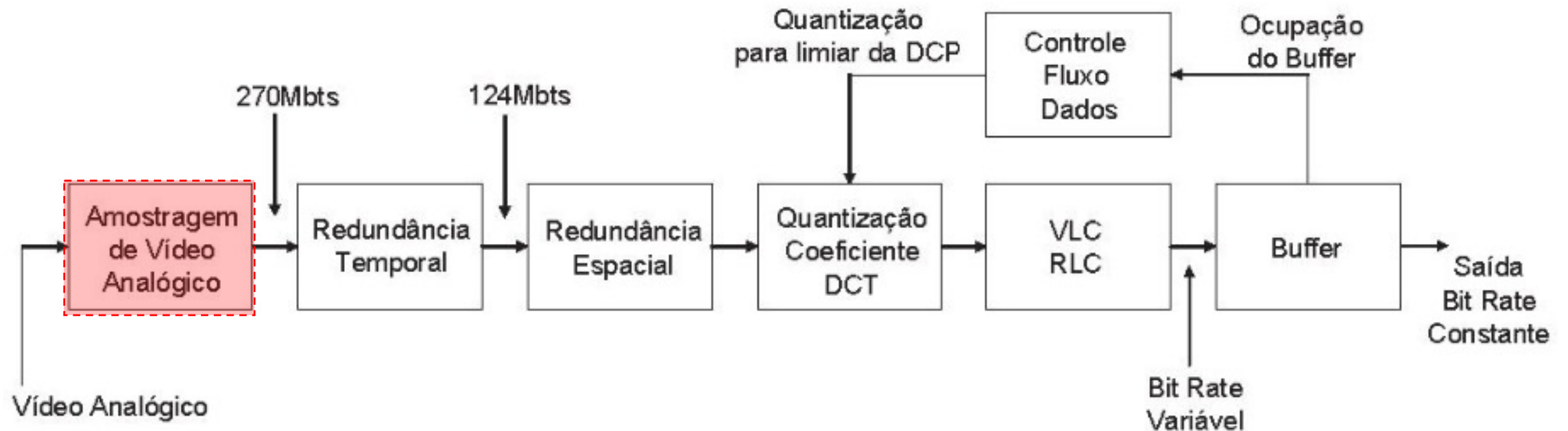
- Também existem muitos outros métodos de compressão de vídeo, como o MPEG-4/H264 e o WM9 (Windows Media 9).
- Contudo, por ser o **MPEG2 o método mais consagrado para uso em TV digital**, será este o foco do nosso estudo.
- O objetivo é formar uma noção dos princípios que norteiam o processo de compressão, sem entrar em detalhes profundos sobre o funcionamento dos blocos envolvidos.

MPEG2

- O princípio básico MPEG2 é utilizar técnicas que reduzem a qualidade da imagem de modo imperceptível, juntamente com técnicas que não afetam sua qualidade, usando principalmente o processo de eliminar redundâncias na informação.



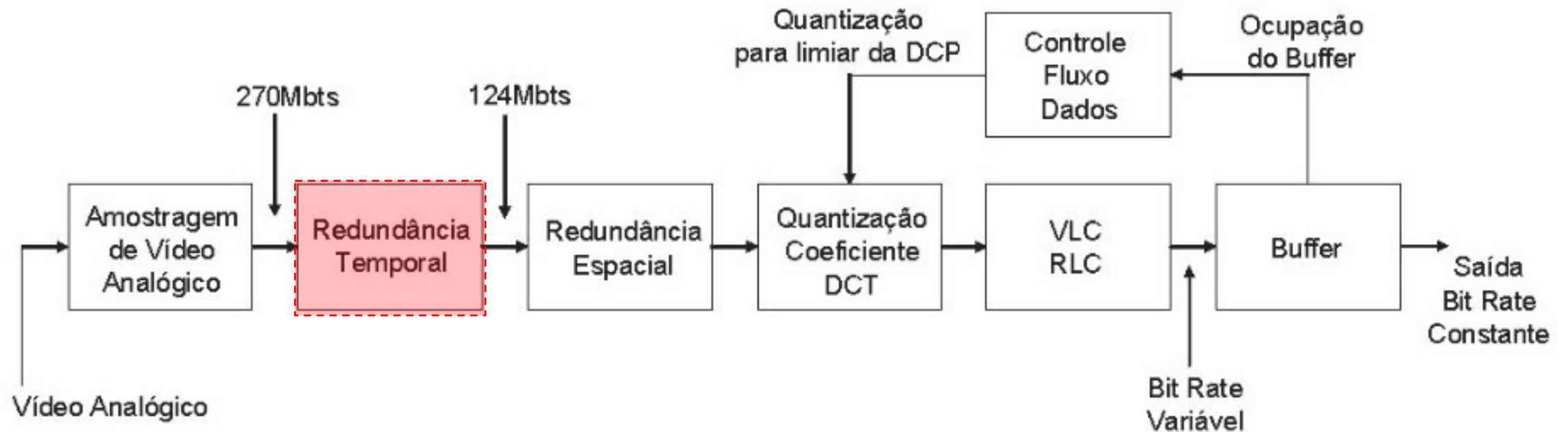
Amostragem e Digitalização de Vídeo



Amostragem e Digitalização de Vídeo

- O primeiro passo do processo é fazer a amostragem e a digitalização dos sinais de vídeo, como visto nas aulas anteriores.
- Como foi visto, a digitalização dos sinais Y , P_B e P_R de **TV analógica tradicional** com formato 4:2:2 ocorreu para a taxa de bits de **270 Mbps**.
- Já a digitalização dos sinais Y , P_B e P_R de **HDTV** com formato 4:2:2 e 10 bits por amostra pode levar a uma taxa de bits de **1,24 Gbps (1080i) ou 1,1 Gbps (720p)**.

Redundância Temporal



Redundância Temporal

- Consiste no aproveitamento da **similaridade** existente entre os quadros sucessivos que formam uma imagem dinâmica.
- Um exemplo prático seria a imagem de um jogador de golfe sacando uma bola. Após o saque, a diferença entre os quadros sucessivos seria somente a posição da bola a cada instante.
- Dessa maneira, o algoritmo enviaria a informação de um quadro completo mais o **vetor de deslocamento** da imagem efetivamente dinâmica.

Redundância Temporal



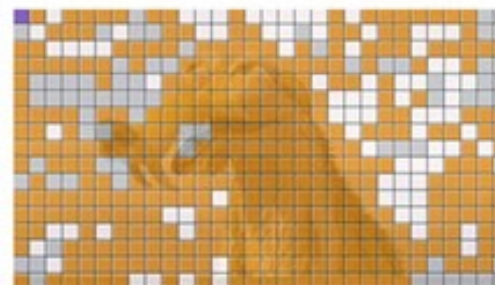
A sequence of video frames



Frame $n-1$



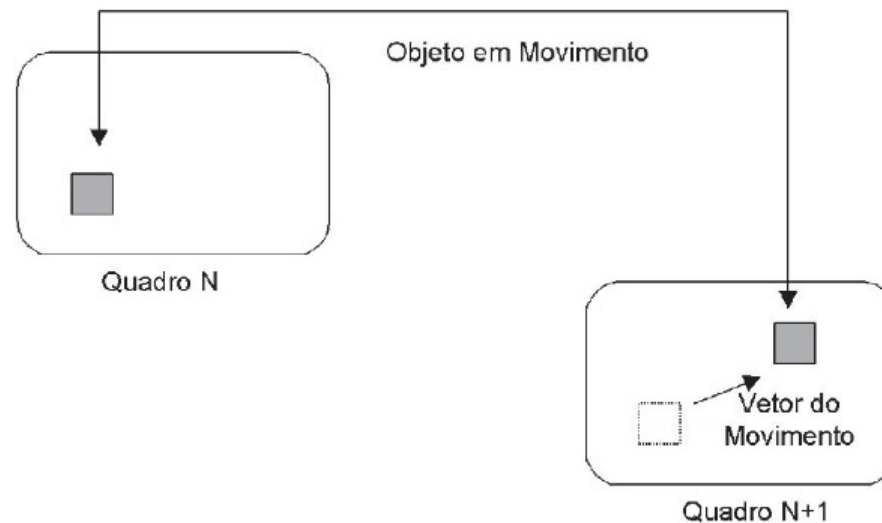
Frame n



- Similar information (partially redundant)
- Independent information
- No change in information (fully redundant)

Redundância Temporal

- A diferença entre o quadro N e o quadro N+1 é o deslocamento do bloco.
- Então, em vez de enviar os dados dos dois quadros completos, envia-se a informação do quadro N completo mais a informação do vetor movimento do quadro N+1.



Redundância Temporal

- Há um **jogo de compromisso** nesse processo, pois caso haja um erro na transmissão do quadro N, por exemplo, o quadro N+1 não poderá ser formado corretamente.
- Para minimizar esse problema, durante a transmissão, são enviados alguns quadros com a informação completa dentro da sequência de quadros (*Group of Pictures*).
- A **compensação de movimentos** se baseia nas semelhanças entre as imagens subsequentes e permite a transmissão apenas das diferenças. Esse método pode ser entendido como a **subtração** de uma imagem pela sua predecessora.

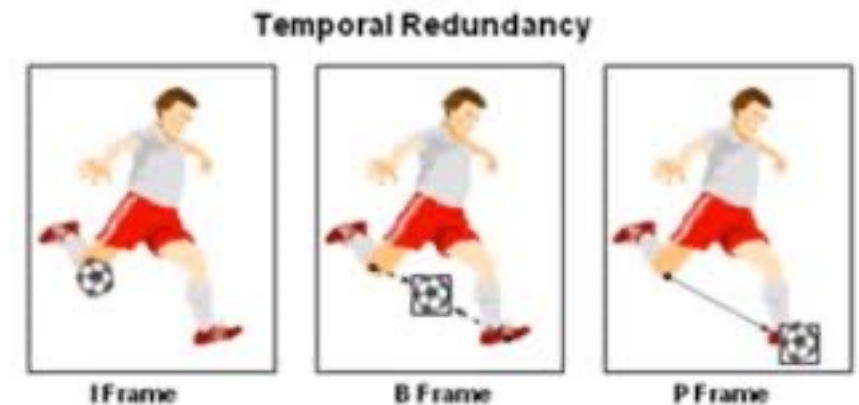
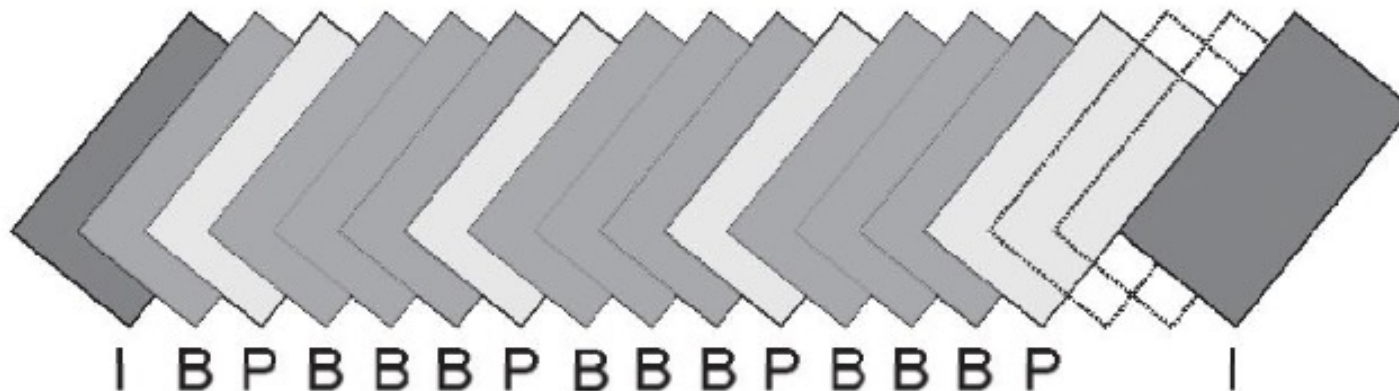
Redundância Temporal

- Para garantir que esse processo não gere perdas, o caminho de realimentação do codificador contém um decodificador funcional que irá subtrair exatamente o sinal que um codificador terá à disposição para somar ao sinal recebido.
- A figura do próximo slide mostra a sequência de quadros com os seguintes símbolos:
 - I – *Intra Frames*: **quadros completos**, usados no início da sequência;
 - B – *Bidirectional Frames*: possuem a informação das **diferenças** entre o quadro atual e os quadros anterior e posterior.
 - P – *Predicted Frames*: são **baseados no anterior**, possuem uma alta compressão e trazem mudança em relação ao último quadro, podendo ainda ser usados como referência para o próximo quadro.

Redundância Temporal

- Se for transmitida uma imagem com o fator $N = 15$ (número total de quadros) e $M = 3$ (número de quadros tipo B que fica entre dois quadros tipo P), tem-se a transmissão mostrada na figura abaixo com a seguinte sequência de quadros.

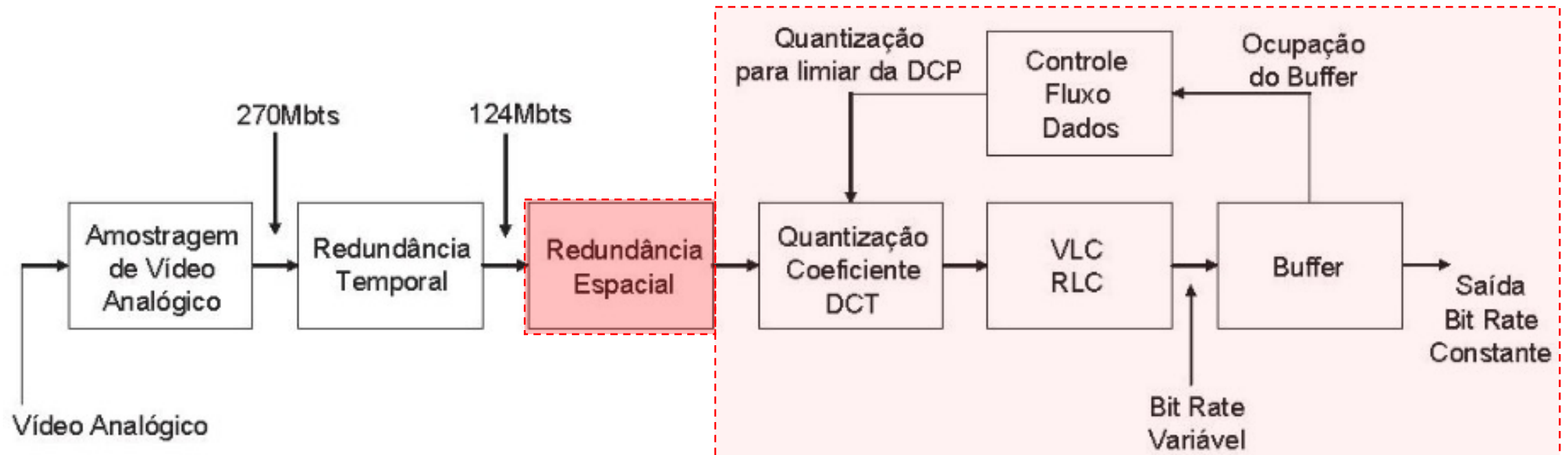
I B P B B B P B B B P B B B P



Redundância Temporal

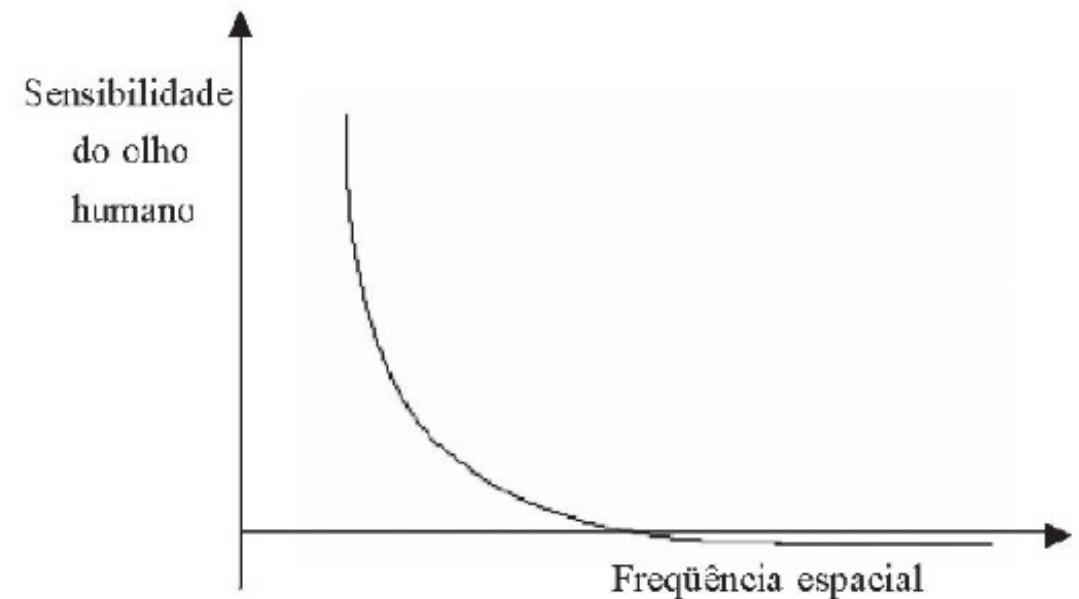
- Além de introduzir o conceito de estimação de movimentos, o MPEG2 introduz um método para regular a taxa de transmissão através do controle do processo de quantificação.
- Um buffer na saída do codificador irá permitir uma **taxa de transmissão constante**.
- Caso o buffer se encontre próximo de exceder sua capacidade de armazenamento, ele irá ordenar que o processo de quantificação **diminua a resolução do vídeo**, diminuindo assim o montante de dados gerados.

Redundância Espacial



Redundância Espacial

- A redundância espacial ou frequência espacial consiste na **semelhança dos pixels adjacentes** de uma imagem.
- Um exemplo é um avião passando no céu sem nuvens, em que a informação relevante a ser transmitida é o avião, e o fundo é a parte da imagem azul cujo conteúdo de imagem é praticamente uniforme.
- O gráfico da figura ao lado mostra que o **olho humano possui baixa sensibilidade para frequências espaciais altas**.



Redundância Espacial

- Baseando-se nesse conceito físico de que **altas frequências podem ser eliminadas sem causar degradações perceptíveis na imagem**, a ideia do MPEG2 é converter a amplitude espacial da imagem em frequência espacial por meio da DCT (transformada discreta do cosseno), conforme já vimos anteriormente na disciplina.
- É importante notar que a eliminação destas componentes caracteriza um processo “**com perdas**”, pois parte da informação foi completamente descartada e não pode mais ser recuperada.

Considerações Gerais

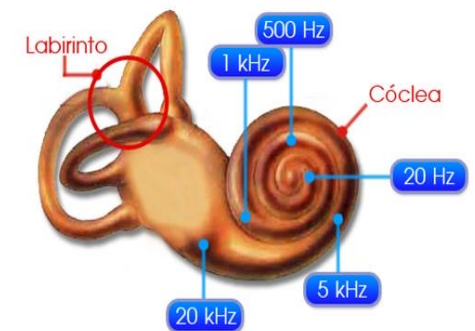
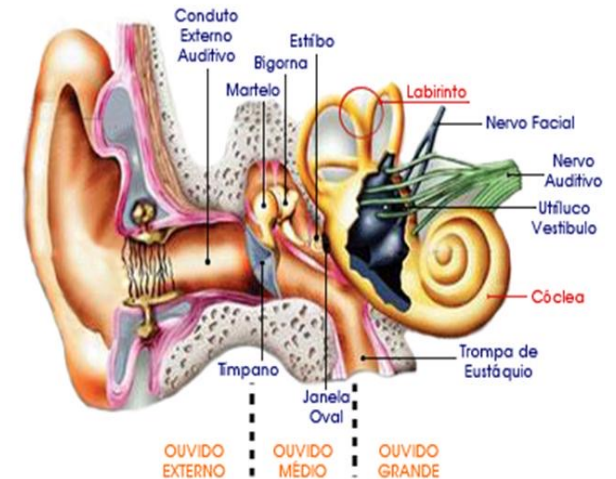
- Após todas essas etapas, a taxa de compressão obtida pode ser superior a 50:1, o que significa que um sinal de vídeo de HDTV digitalizada que possui uma taxa superior a **1 Gbps** passa a ter uma taxa de aproximadamente **20 Mbps** depois de comprimido.
- Como mencionado, há outros métodos de compressão (como o MPEG 4, WM9 e o H264). Entretanto, como não são tradicionalmente empregados para transmissão de TV terrestre, não serão discutidos.

Compressão de Áudio

- A gama de frequências audíveis pelo ouvido humano vai de 20 Hz a 20 kHz, e o limite superior varia de pessoa para pessoa.
- Para a qualidade de CD, amostra-se o sinal de áudio a uma frequência de 44,1 kHz.
 - Obedecendo ao critério de Nyquist, idealmente 40 kHz seria suficiente. Entretanto, um filtro passa-faixa ideal é impraticável, então é introduzida uma banda de transição de 2,05 kHz.
- Utilizando-se o código de 16 bits, tem-se 706 Kbps; se a transmissão for para música estéreo (ou seja, com dois canais), a taxa será de 1,4 Mbps.
- Surge então a necessidade de um algoritmo de compressão que reduza essa taxa, para possibilitar uma transmissão de maneira eficiente.

Compressão de Áudio

- O codificador de áudio MPEG consegue reduzir a taxa de dados de um CD em aproximadamente 12 vezes, sem perda significativa da qualidade.
- O principal fator do qual o processo se vale para essa redução consiste no uso da **característica do ouvido humano de não ser um aparelho perfeito para captar sons.**
- O MPEG aproveita tais imperfeições para comprimir o áudio.



Compressão de Áudio

- Ao se ouvir uma orquestra tocando simultaneamente violinos, trombetas e os instrumentos de percussão, **os ouvintes não são capazes de captar todos os sons simultaneamente** porque alguns dos sons estão “camuflados” pelos sons dos instrumentos mais fortes, uma vez que o ouvido humano é mais sensível aos sons mais graves do que aos mais agudos.



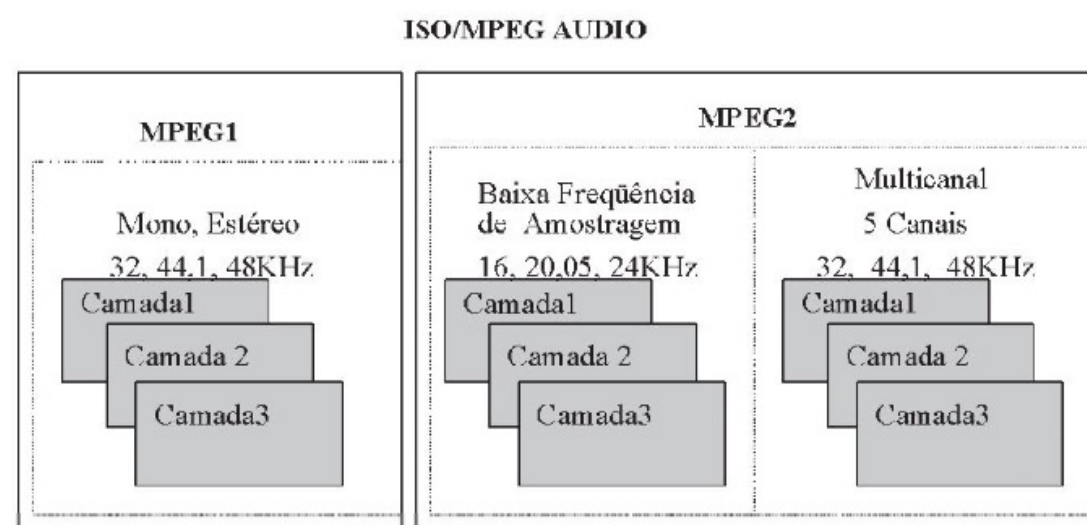
Compressão de Áudio

- Uma gravação em CD dessa orquestra contém todos os sons dos instrumentos da orquestra. Mas, se **as partes que as pessoas não são capazes de ouvir forem removidas**, consegue-se ouvir a mesma orquestra com a mesma qualidade do CD.
- Esse é o funcionamento geral do sistema de compressão do MPEG Áudio.



Compressão de Áudio

- O MPEG é o primeiro padrão internacional no domínio de **compressão de áudio de alta fidelidade** (*High Fidelity*, ou simplesmente Hi-Fi).
- Em particular, o MPEG define a sintaxe da sequência de bits de áudio codificado a baixas taxas de bits. O algoritmo de codificação não é definido pela norma MPEG.



MPEG1 Áudio

- O **MPEG1 Áudio** providencia o esquema de compressão para fontes de áudio estéreo e mono com qualidade semelhante a um CD de áudio.
- A parte do algoritmo correspondente é dividida em três camadas:
 - **MPEG1 Áudio Camada 1**: oferece **menor taxa de compressão**; é usado em sistemas de áudio de consumo; a sua vantagem é o **menor custo de implementação**;
 - **MPEG1 Áudio Camada 2**: oferece **maior taxa de compressão**; é encontrado em aplicações profissionais ou em aplicações de **consumo mais sofisticado**;
 - **MPEG1 Áudio Camada 3**: oferece **maior taxa de compressão** e é encontrado em Rede Digital de Serviços Integrados (RDSI); **seu custo é maior e apresenta maior complexidade de codificação e decodificação**.

MPEG1 Áudio

- A ideia principal é que uma aplicação do MPEG1 viabiliza usar a camada que oferece o **melhor compromisso entre a necessidade de computação e o desempenho de compressão**.
- O áudio pode ser codificado em qualquer uma das camadas.
- O MPEG1 Áudio destina-se a transformar sinais de áudio amostrados a taxas de 32 kHz, 44,1 kHz ou 48 kHz e codifica-los para uma taxa de 32 Kbps a 192 Kbps.

MPEG2 Áudio

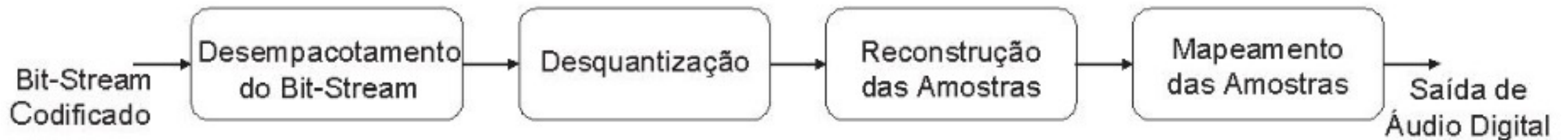
- O **MPEG2 Áudio** foi desenvolvido com base no MPEG1, fazendo com que os seus benefícios maior qualidade e maior eficiência, propiciando um padrão de qualidade para satisfazer os requisitos necessários para as **redes digitais**.
- Além dos canais mono e estéreo, foi adicionado suporte para múltiplos canais.
- Assim, o MPEG2 tem a possibilidade de suportar até **cinco canais de som** (esquerdo, direito, centro, dois canais de *surround* e mais um canal de graves), ou, em extensão, sete canais mais um canal de graves.
- Isso possibilita ter em casa o som com o mesmo desempenho de algumas salas de cinema, permitindo a reprodução do som estereofônico mais realista.

MPEG2 Áudio

Codificador MPEG Áudio



Decodificador MPEG Áudio



MPEG2 Áudio

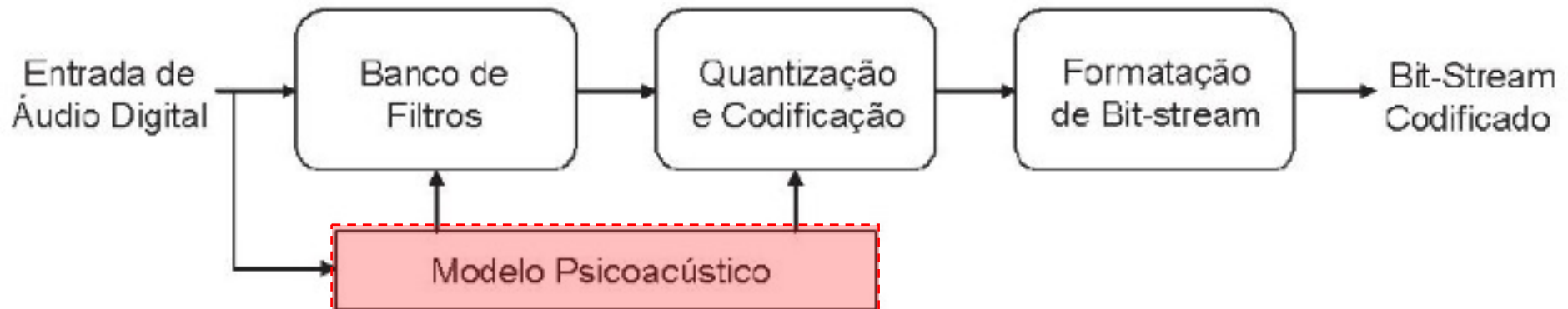
- O MPEG2 possui outras funções, como a característica “**multilíngue**”, permitindo, por exemplo, que um filme possa ser ouvido em diferentes idiomas, de acordo com a escolha do telespectador.
- Para efetuar a compressão de áudio, existem duas alternativas:
 - reduzir a **frequência de amostragem**;
 - reduzir a resolução da amostra (ou seja, diminuir o número de **níveis de quantização**).
- Para se obter uma boa qualidade, não se pode fazer grandes alterações na frequência de amostragem, pois ela é limitada a no mínimo 44,1 kHz (recapitulando: devido à audição humana de até 20 kHz e ao teorema de Nyquist).

MPEG2 Áudio

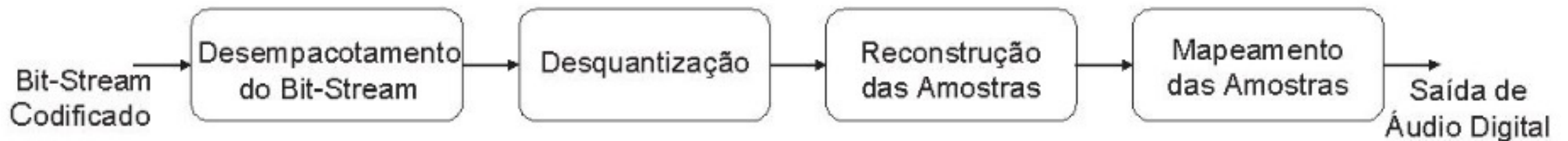
- A razão de escolher código de 16 bits deve-se ao desejo de obter boa **relação sinal-ruído (SNR)**, de quantificação proveniente do processo de digitalização do sinal.
- Para cada bit que se adiciona ao código há uma melhoria de 6 dB nesta relação.
- Ao se utilizar código de 8 bits por amostra, pode-se ainda perceber **ruído de fundo** nos momentos de silêncio de música. Mas, para esse nível de codificação, se o nível de som é elevado, não se ouve nenhum ruído. Esse efeito é chamado de **Efeito Máscara**, e é a chave para o método de codificação utilizado pelo MPEG2.
- Com o sistema de CD Áudio consegue-se uma SNR de 90 dB, o que está dentro da gama dinâmica do ouvido humano, pois nesta faixa não se ouve nenhum ruído.

Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

Codificador MPEG Áudio

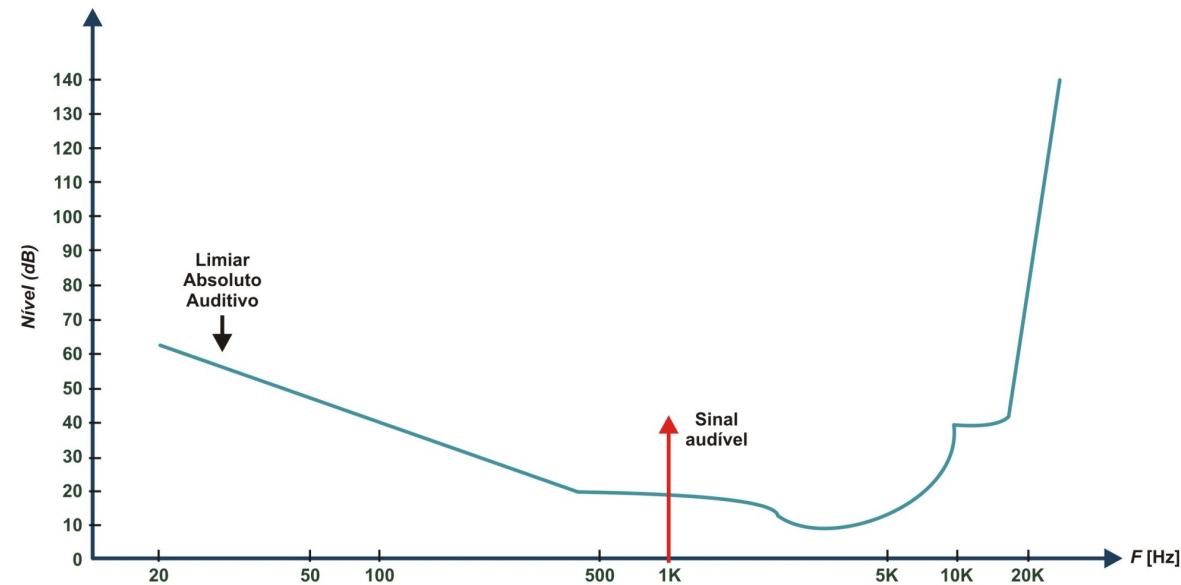


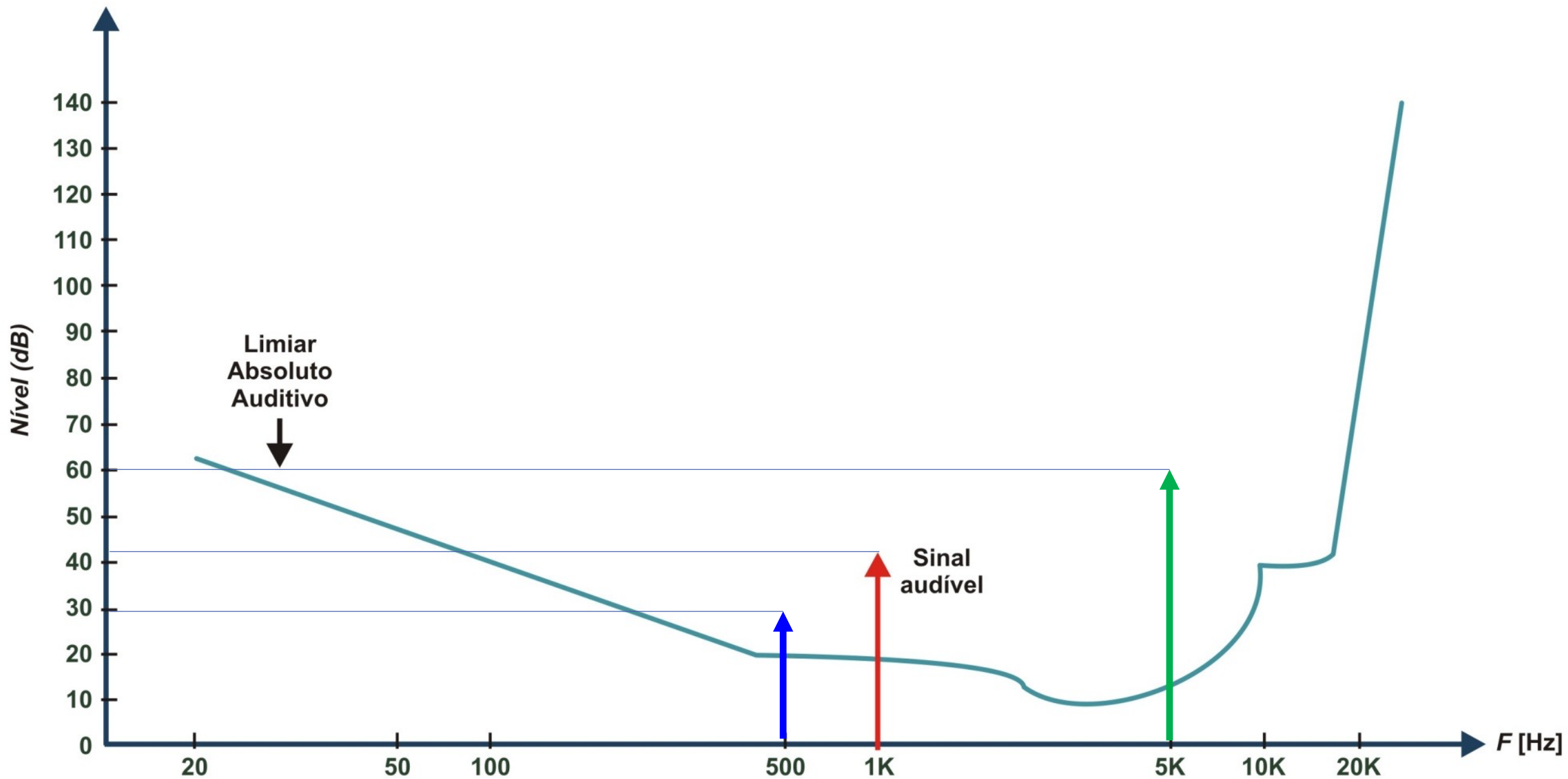
Decodificador MPEG Áudio



Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

- O **modelo psicoacústico** baseia-se no efeito da sensibilidade do ouvido humano, que tem uma gama dinâmica de 96 dB em níveis de som, variando em frequência.
- Entretanto, a sensibilidade não é uniforme para toda a gama de frequência, sendo mais sensível na faixa de 1 kHz a 3 kHz.





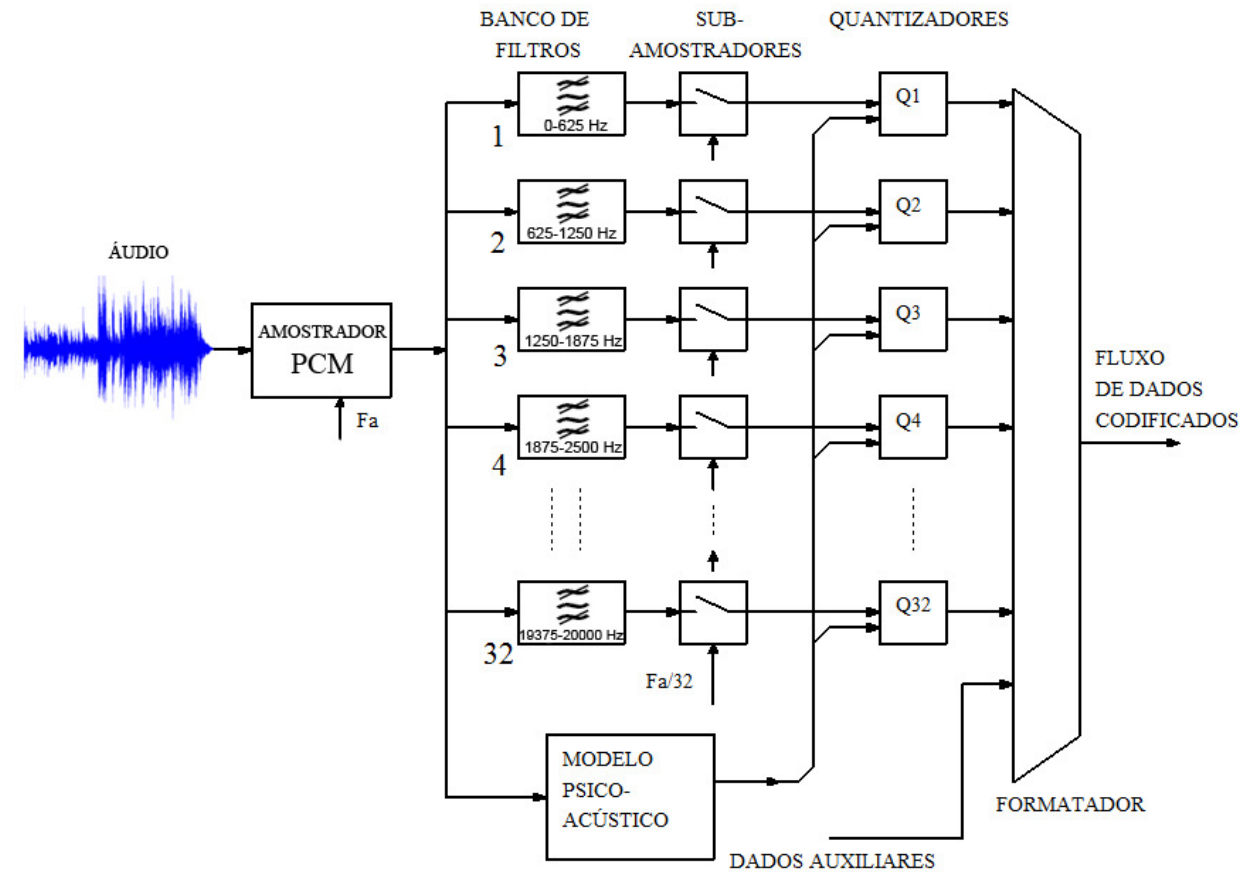
Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

- Quando dois tons de frequências próximas estão simultaneamente presentes, o ouvido humano percebe melhor o som com nível mais elevado.
- Quando os dois sons tem frequências mais afastadas uma da outra, o ouvido humano percebe mais nitidamente a presença simultânea dos dois.
- Tal efeito recebe o nome de Efeito Máscara – efeito que permite aumentar o nível de ruído de fundo à volta do som forte, uma vez que o ruído já será mascarado.
- A largura de banda das curvas dos sinais que causam o Efeito Máscara não é proporcional à frequência.
- Essa largura, chamada de **Dimensão de Banda Crítica**, é de cerca de 100 Hz para baixas frequências, aumentando muito para frequências mais elevadas.

Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

- **MPEG2 Áudio Camada 1**

- O mapeamento tempo-frequência é constituído por um **banco de filtros** polifásicos com 32 sub-bandas, com baixa complexidade computacional e opções de implementação flexíveis;
- Essas sub-bandas são igualmente espaçadas em frequência.



Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

- O modelo psicoacústico usa uma **Fast Fourier Transform** (FFT), de modo a obter uma informação espectral detalhada do sinal.
 - A saída da FFT é usada para determinar as máscaras do sinal (ruído).
- Cada máscara produz um **limiar de mascaramento** dependente da sua tonalidade, intensidade e frequência.
 - Para cada sub-banda, os limiares de mascaramento são combinados de modo a formar um limiar global de mascaramento.
 - O limiar é comparado com o nível do sinal para cada sub-banda, produzindo uma relação sinal-máscara (Signal Mask Rate - SMR), que é a entrada do quantizador.

Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

- O quantizador/codificador examina as amostras de cada sub-banda, procura o valor máximo absoluto e quantiza a 6 bits.
 - Esse valor é chamado de fator de escala da sub-banda.
- É possível mesmo que em bandas muito mascaradas seja obtido o resultado “0 bit”, **de modo que nenhuma amostra seja codificada.**
- Finalmente, as amostras da sub-banda são linearmente quantizadas.

Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

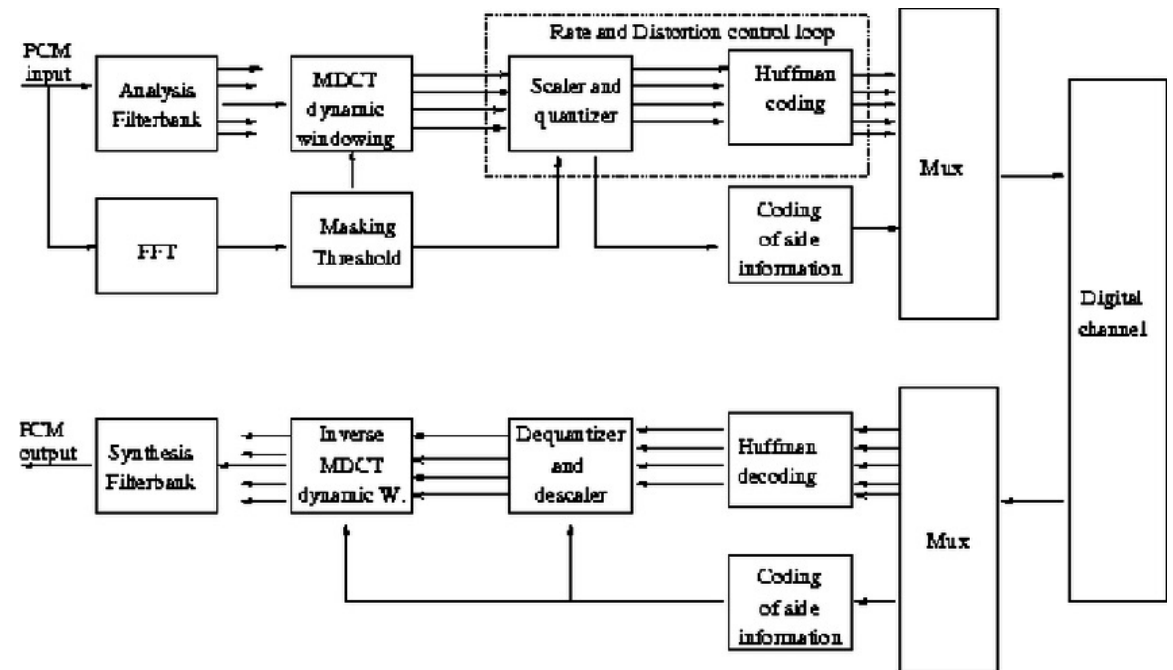
- **MPEG2 Áudio Camada 2**

- O mapeamento tempo-frequência é igual ao da Camada 1, ou seja, também usa um banco de filtros polifásicos de 32 sub-bandas.
- O modelo psicoacústico é semelhante ao da Camada 1, porém, **usa uma FFT com maior resolução de frequência.**
- Usa também o mesmo procedimento que a Camada 1 para produzir as relações sinal-máscara, para cada uma das 32 sub-bandas.
- O quantizador/codificador é também semelhante ao da Camada 1.
- **A Camada 2 permite a cada banda uma sequência de três fatores de escala sucessivos, capturando muito mais redundância do sinal que a Camada 1.**

Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

- **MPEG2 Áudio Camada 3**

- O mapeamento tempo-frequência usa um banco de filtros polifásicos de 32 bandas principais, seguido de uma filtragem usando a Transformada Discreta do Cosseno Modificada (MDCT) para subdivisão em 18 sub-bandas de cada uma das 32 sub-bandas principais, resultando em 576 sub-bandas, provendo um **cálculo mais preciso dos limiares de mascaramento**.



Modelo Psicoacústico, Efeito Máscara e Bandas Críticas

- O quantizador/codificador é mais sofisticado, com taxa de bits variável.
- O sistema de codificação exige grande poder computacional, mas permite resultados de alta qualidade a taxas de transmissão muito baixas, como 64 Kbps.
- Esta camada pode ser usada em Redes Digitais de Serviços Integrados (RDSI), que pode transportar áudio em tempo real com qualidade de CD.
- O nível de compressão e de qualidade alcançados por este sistema devem-se principalmente à utilização do processo de codificação por sub-banda, e baseado num modelo psicoacústico do ouvido humano que permite mascarar a percepção de qualquer ruído introduzido no sinal de áudio após o processo de compressão.

Referências

- YAMADA, F.; SUKYS, F.; BEDICKS JR., G.; AKAMINE, C.; RAUNHEITTE, L. T. M.; DANTAS, C. E.
Parte I - SISTEMAS DE TV DIGITAL. **Revista Mackenzie de Engenharia e Computação**, v. 5, n. 5, 17 mar. 2010.
- MPEG. **The Moving Picture Experts Group website**. <<https://mpeg.chiariglione.org/>>
- MPEG-1: coded storage of sampled sound waves. **MIT Media Lab**.
<<https://sound.media.mit.edu/resources/mpeg4/audio/faq/mpeg1.html>>
- MPEG-2: coded transmission/storage of sampled sound waves. **MIT Media Lab**.
<<https://sound.media.mit.edu/resources/mpeg4/audio/faq/mpeg2.html>>