



Compressão de áudio

Marcio Dahia e Geber Ramalho



O que é compressão de dados?

- Processo de codificar uma mensagem a fim de reduzir o número de “bits” necessários para representá-la
- 2 tipos de compressão
 - Sem perdas
 - Não há eliminação de informação na mensagem.
 - Compressão → codificação eficiente
 - Com perdas
 - Informações são descartadas
 - Redundantes, pouco importantes ou irrelevantes sob algum critério
 - Não descarta a codificação eficiente
 - Grau de compressão x Distorção na mensagem



O que é compressão de dados?

- Por que uma codificação específica pra áudio?
 - Teoria da Informação: Quanto maior o conhecimento sobre a mensagem, emissor e receptor, melhor a compressão
 - Mp3 usa a modelagem psicoacústica para remover informações da irrelevantes
 - FLAC usa a correlação entre os canais (E/D) para aumentar a compressão



Tabela

| Formato | Tamanho (em MB) | TC(%)* |
|----------------------|-----------------|--------|
| WAV | 35,2 | - |
| RAR | 24,7 | 29,83 |
| FLAC | 22,6 | 35,80 |
| Mp3 192kbps (iTunes) | 4,92 | 86,02 |
| Mp3 192kbps (lame) | 4,79 | 86,39 |
| Ogg Vorbis (nível 5) | 3,64 | 89,66 |
| AAC 128kbps (iTunes) | 3,3 | 90,63 |
| Mp3 128kbps (iTunes) | 3,28 | 90,68 |
| Mp3 128kbps (lame) | 3,19 | 90,94 |

*Taxa de Compressão(TC) = tamanho da arq. comp./tamanho da arq. original *100



Codificação de áudio: Histórico

- 1988 : Criação Motion Pictures Expert Group (MPEG)
 - Padronização da codificação para transmissão e gravação de filmes
 - Imagem, seqüência de imagens (vídeo) e áudio
 - Surgimento de novos conceitos e vocabulário
 - Frame, Codec, Bitrate
 - **Padronização apenas do processo de decodificação**
 - Apenas sugestão de compressão



Codificação de áudio: Histórico

- 1992: MPEG-1.

- O padrão de codificação de áudio consistia em 3 modos de operação de acordo com o taxa de compressão (complexidade do processo de codificação)

| | TC |
|-----------------|--------------|
| Layer I | 1:4 |
| Layer II | 1:6 .. 1:8 |
| Layer III (MP3) | 1:10 .. 1:12 |

- As principais idéias foram sugeridas pela *Fraunhofer Institut für Integrierte Schaltungen* (IS 11172-3 e IS 13818-3).



Codificação de áudio: Histórico

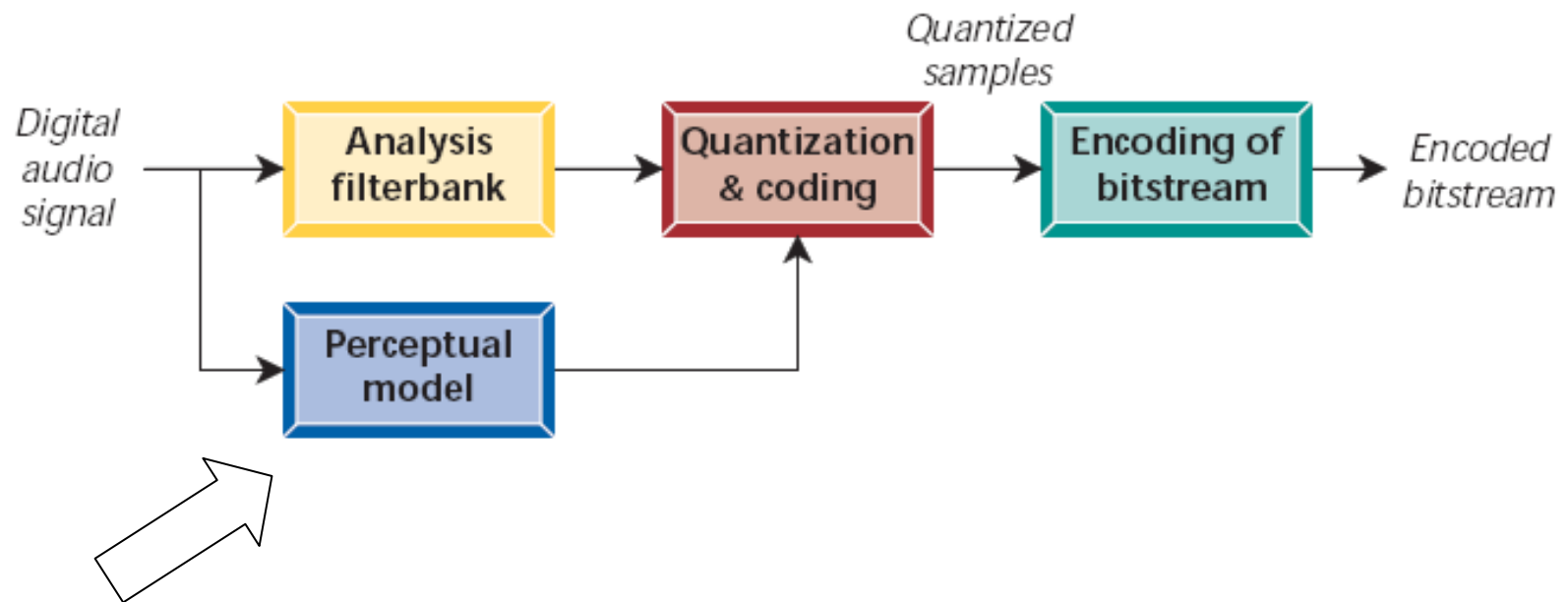
- 1994: Surge Xiph.org
 - Definiu padrão totalmente aberto para transmissão e armazenamento de áudio (vorbis e flac) e vídeo(theora).
- 1996: Surge o ID3
 - Padrão de fato para inserção de metadados no mp3
- 1997: MPEG-2
 - Advanced Audio Coding (AAC)
 - Estado da arte em compressão de áudio (melhorado em Mpeg-4)
 - 1:16
- 2003: Surge FLAC



Conceitos importantes

- Codec (COder-DECoder)
- Frames
 - Forma de dividir um arquivo de áudio
 - Uma unidade de codificação
 - **Mp3 -> 1152 amostras PCM**
- Bitrate (taxa de bits)
 - Numero de bits usados para codificar um frame
 - Medida em kbps
 - CBR (constant bitrate) - A mesma quantidade de bits para representar qualquer frame (WAV)
 - ABR (average bitrate) - Frames menos complexos deixam bits para os próximos (MP3)
 - VBR (variable bitrate) - O número de bits para cada frame deve garantir a qualidade definida no início da codificação (AAC, VORBIS)

Processo de Compressão MP3



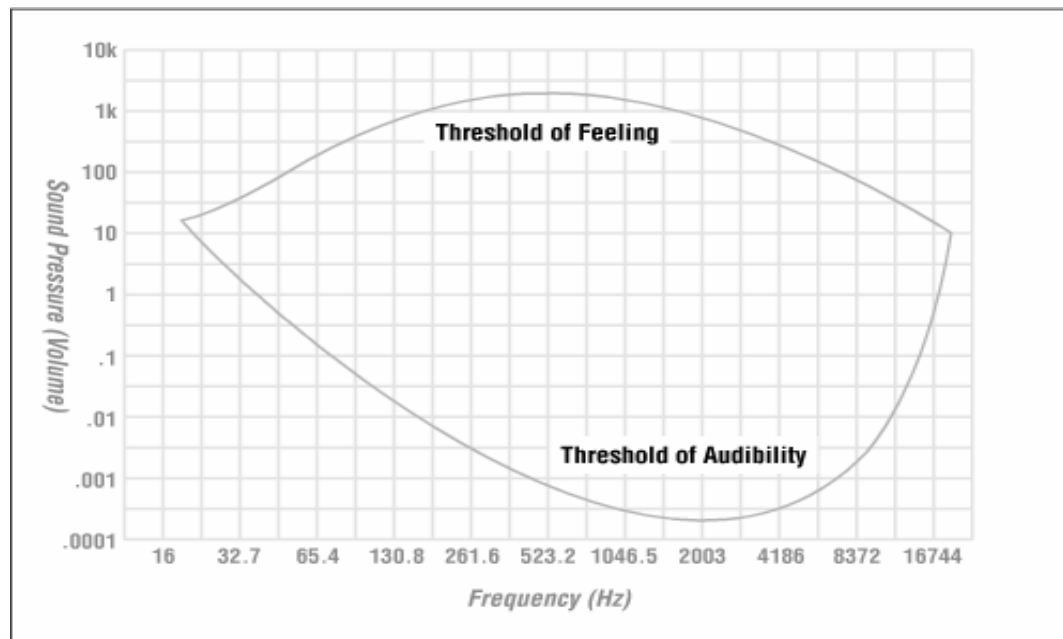


Modelo Perceptual

- Objetivo
 - Analisar o que pode ser considerado redundância
- Critérios
 - Limiar de audibilidade
 - Mascaramento de sons

Limiar de audibilidade

- Um codificador perceptual compara o sinal de entrada com o limiar de audição e descarta os sinais que estão abaixo ou acima

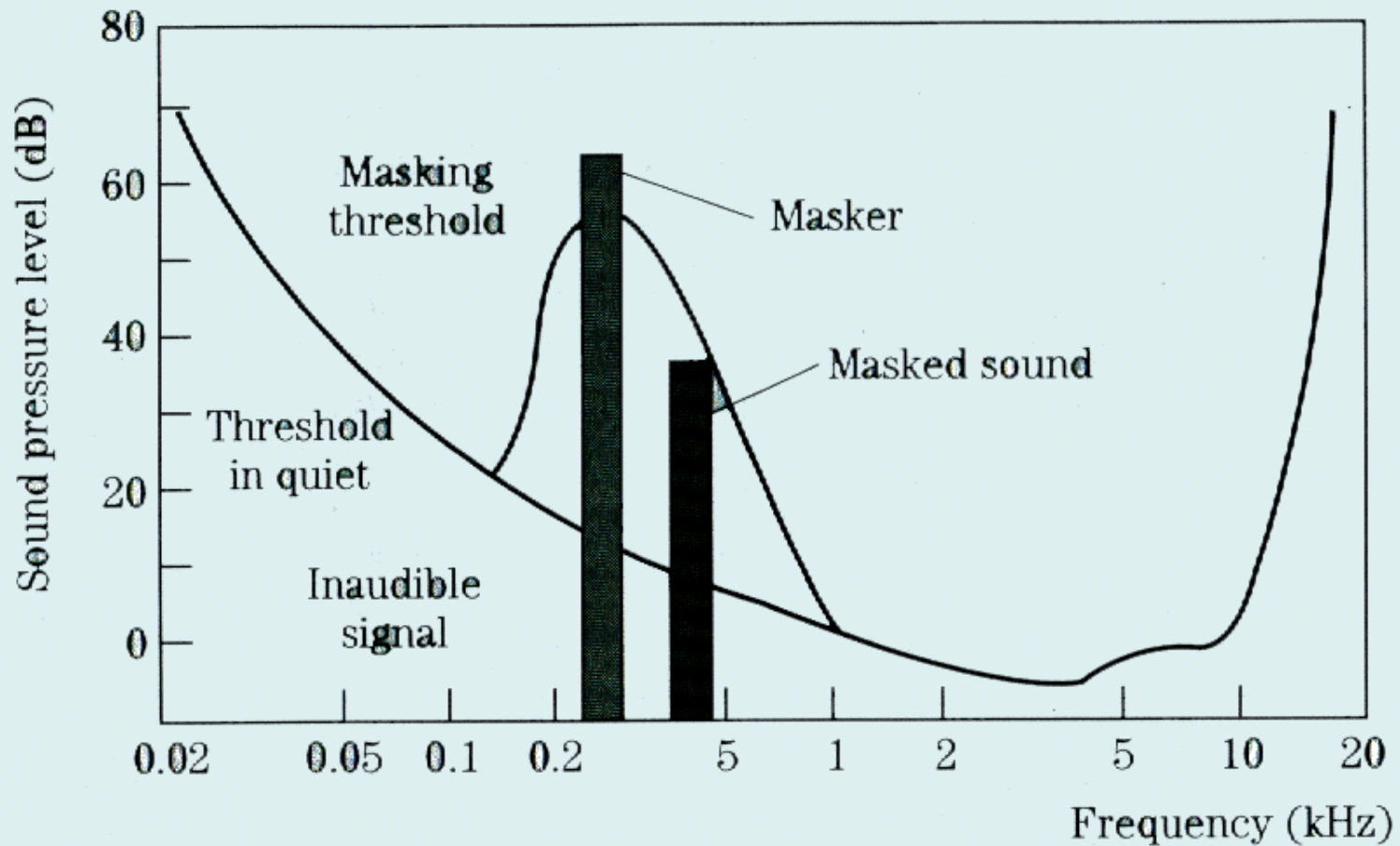




Mascaramento

- Dois tipos de mascaramento
 - Simultâneo
 - Temporal
- Mascaramento Simultâneo
 - Quando dois sons ocorrem simultaneamente, o de maior volume podem ocultar o de menor volume.
 - A curva de mascaramento é assimétrica
 - é mais fácil um tom mais baixo (na freqüência) mascarar um mais alto do que o contrário.

Mascaramento Simultâneo

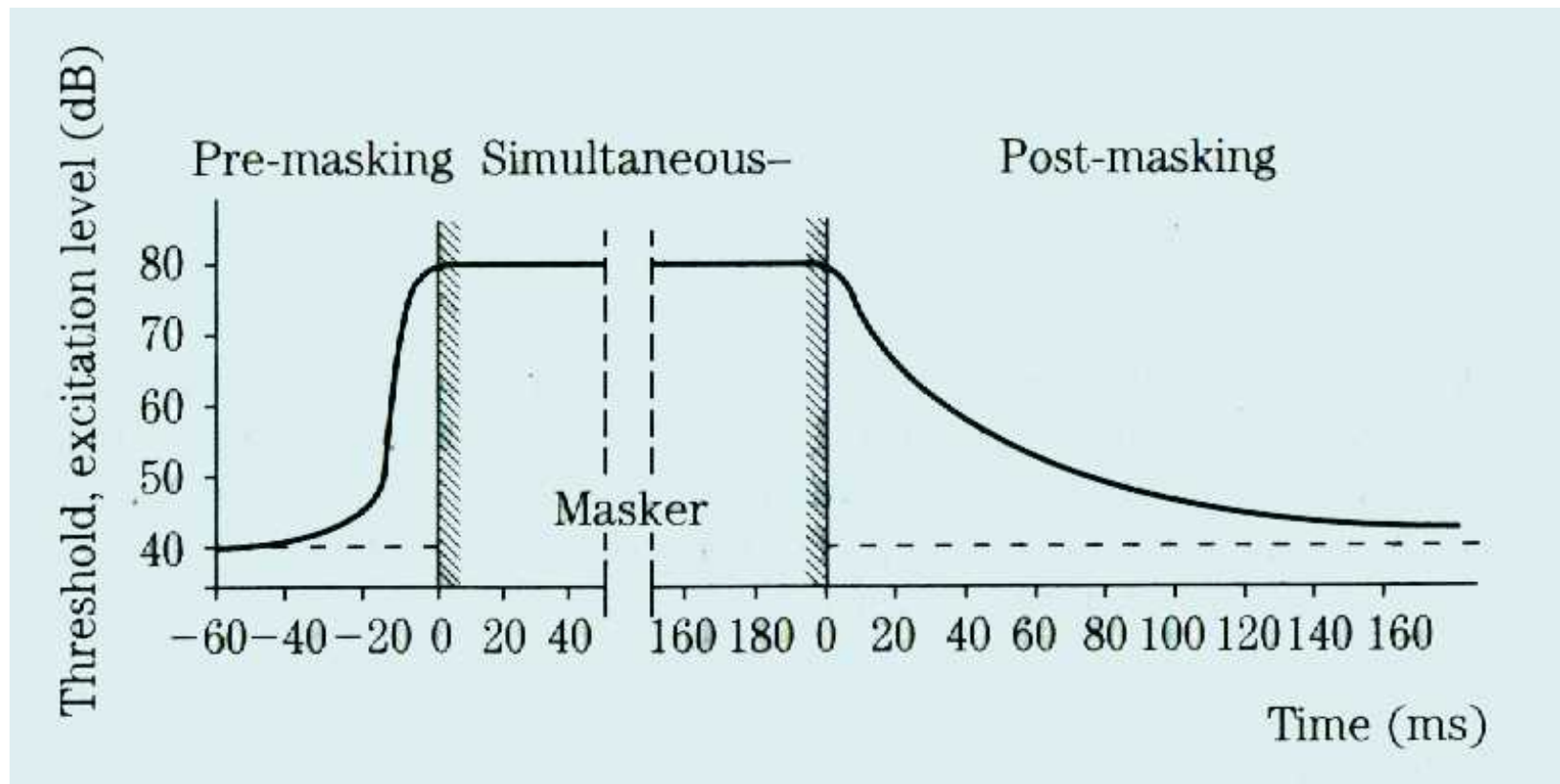




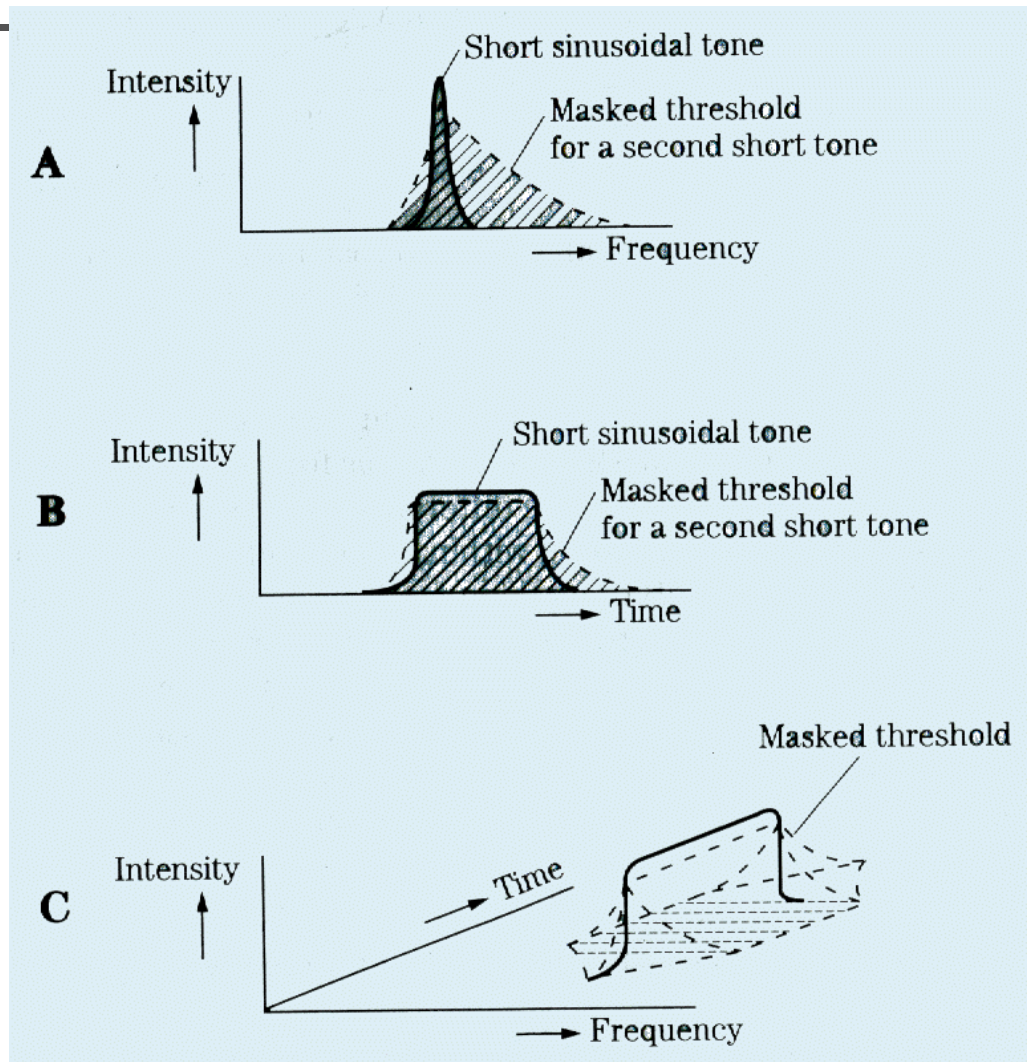
Mascaramento Temporal

- Ocorre quando sons são tocados próximos no tempo, mas não ao mesmo tempo.
- Um som de maior volume que ocorre antes ou depois de um som mais baixo pode mascará-lo

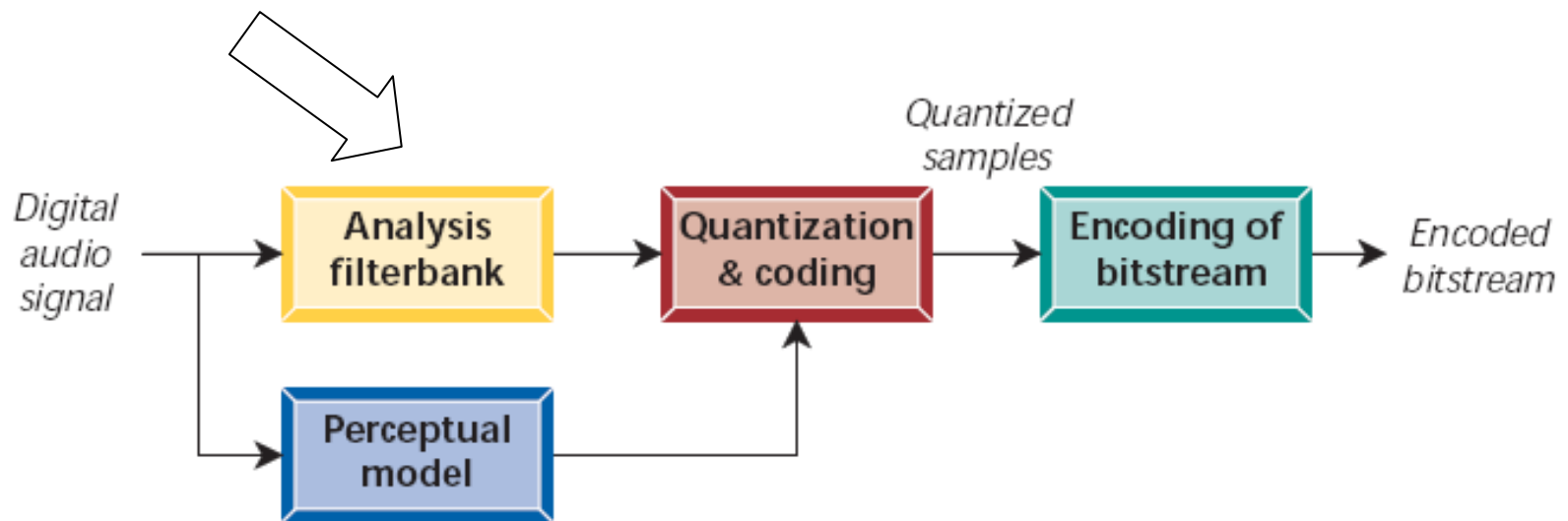
Mascaramento Temporal (cont.)



Mascaramento Temporal e Simultaneo



Processo de Compressão MP3



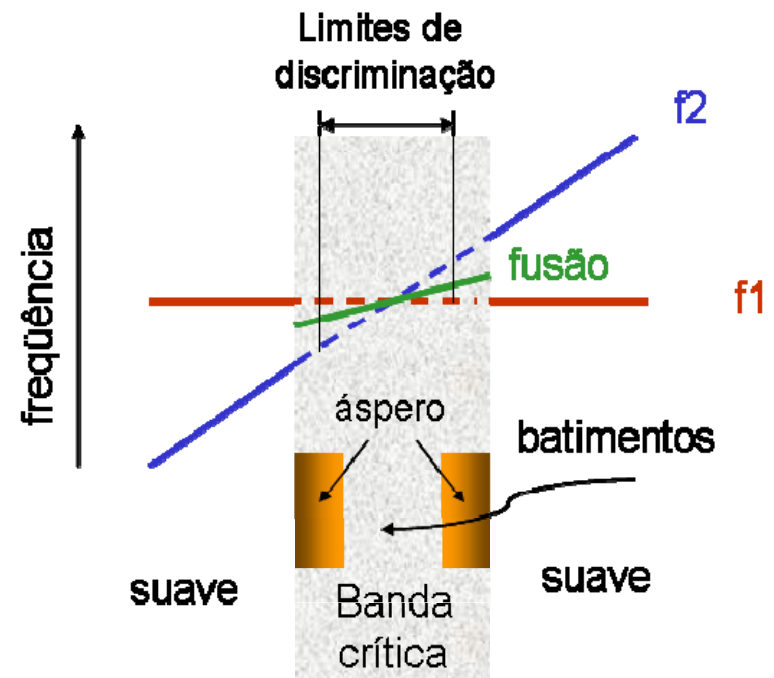


Analisis Filterbank

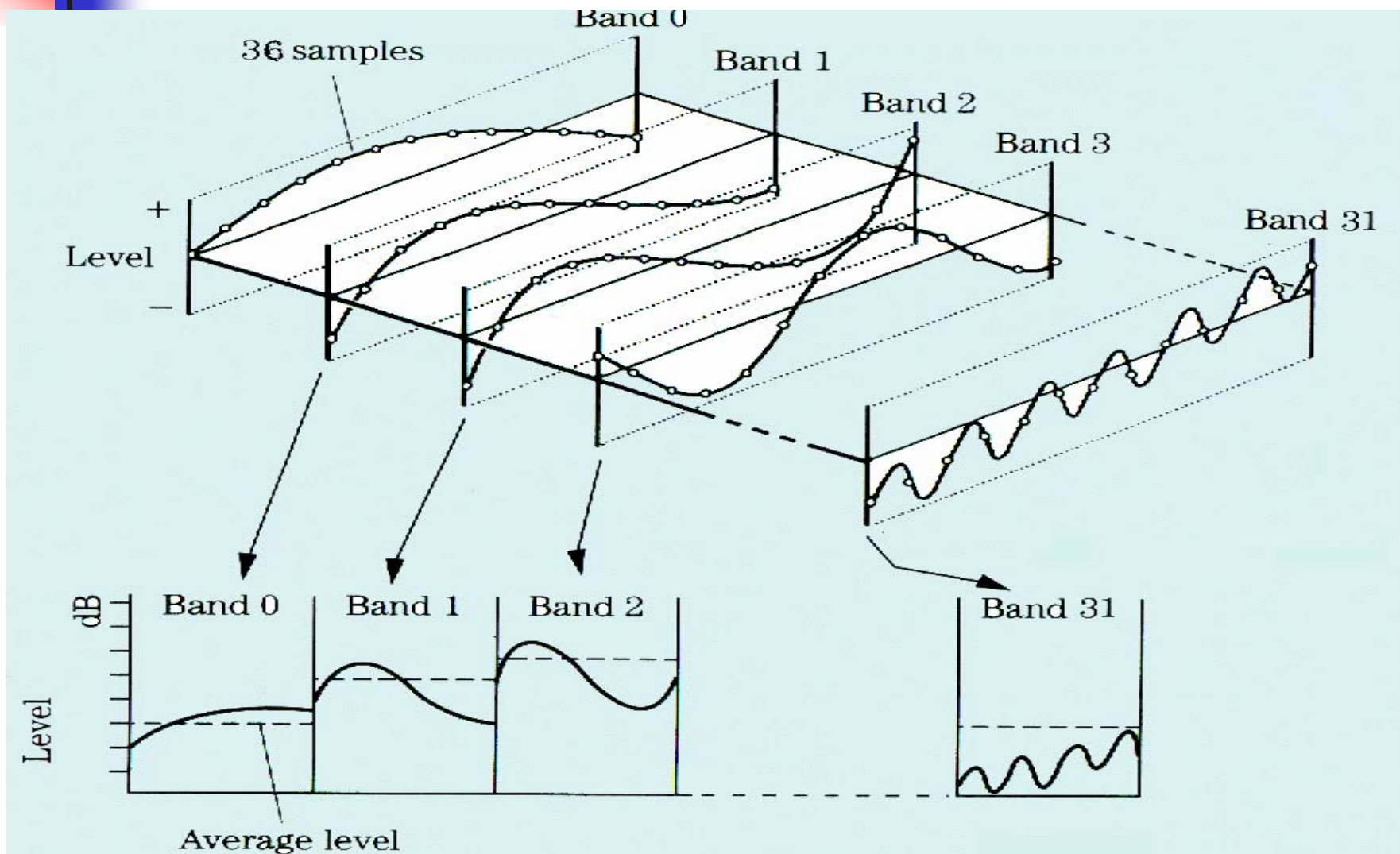
- Objetivo: transformar o sinal para o domínio da frequência para evidenciar as redundâncias e preparar o sinal para retirada
 - Composto por 2 fases
 - Filtragem em sub-bandas
 - Transformada Cosseno Discreta Modificada (MDCT)

Filtragem de sub-banda

- Baseado no conceito de banda-critica
 - regiões da membrana basilar que estabelecem limites na percepção de frequências
 - 32 subbandas de mesma largura
 - Pouca justificativa psicoacustiva
 - Motivo: eficiência
- 1 subbanda é selecionada para cada 32 amostras
 - 36 amostras em cada subbanda



Filtragem em Sub-bandas



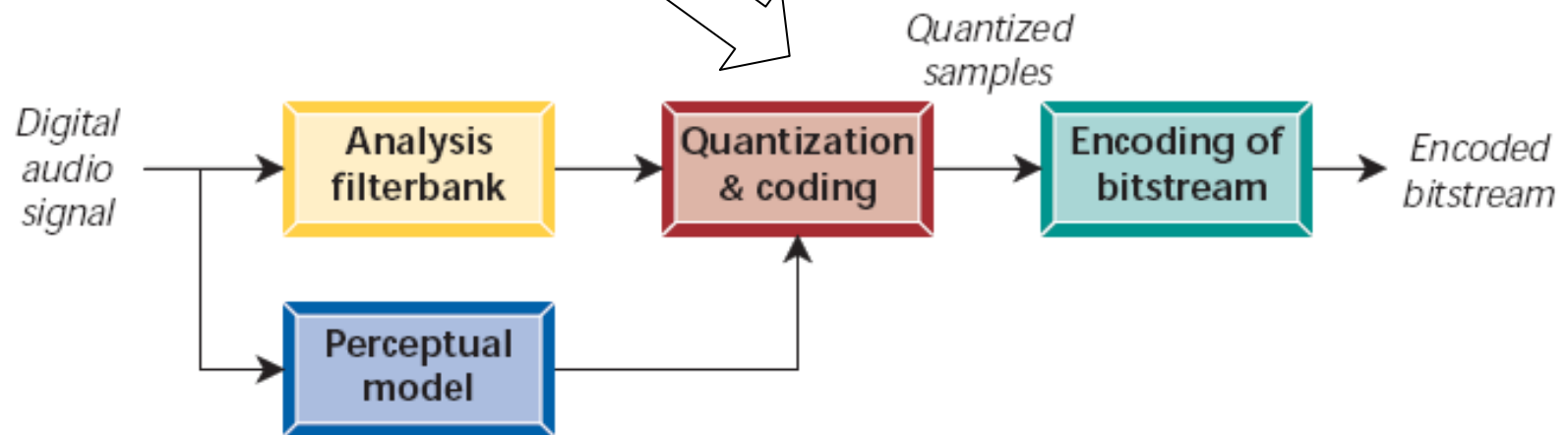


Transformada Cosseno Discreta Modificada (MDCT)

- Cada banda passa é transformada do domínio do tempo pra domínio da frequência (50% overlap)

$$X(m) = \sum_{k=0}^{n-1} f(k)x(k) \cos\left[\frac{\pi}{2n}\left(2k + 1 + \frac{n}{2}\right)(2m + 1)\right], \quad m = 0 \dots \frac{n}{2} - 1$$

Processo de Compressão MP3





Quantização e codificação

- Quantização
 - As informações irrelevantes (detectadas na fase de modelagem perceptual) são efetivamente retirada
 - O número de bits para cada banda é determinado
- Codificação
 - Dissociação inter-canais (caso estéreo)
 - Os Canais E e D são transformados em média $(E+D/2)$ e *side* $(E-D/2)$
 - Codificação de Huffman

Codificação de Huffman

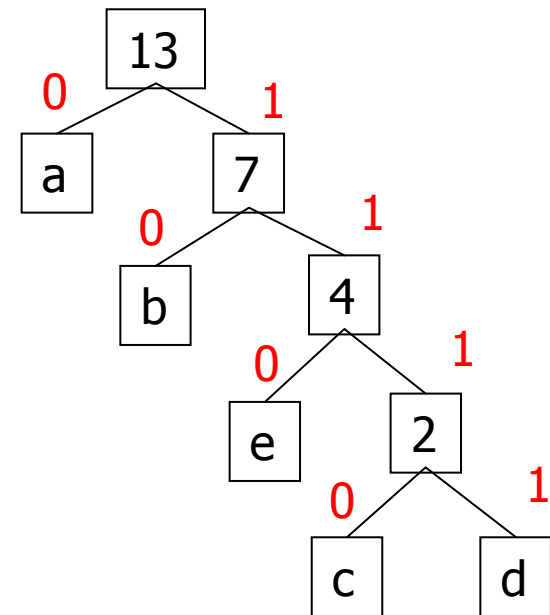
- Mensagem

- aaaaaabbbbcdee
- 104 bits

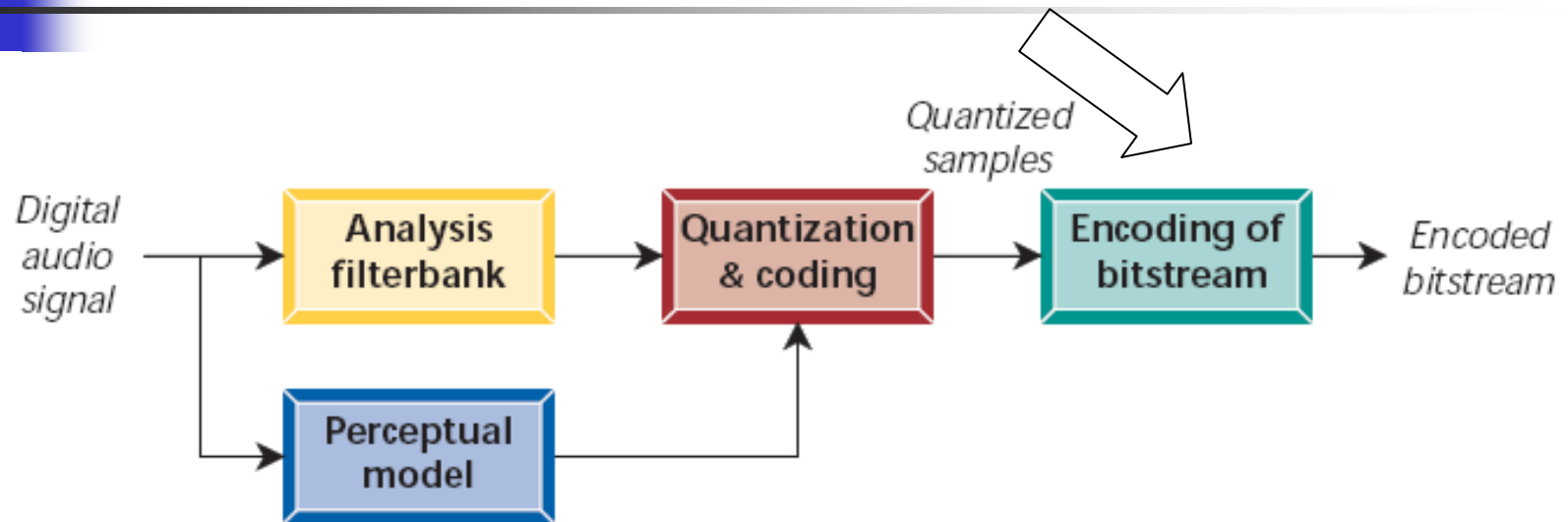
| símbolo | ocorrência | código |
|---------|------------|--------|
| a | 6 | 0 |
| b | 3 | 10 |
| c | 1 | 1110 |
| d | 1 | 1111 |
| e | 2 | 110 |

- Mensagem comprimida

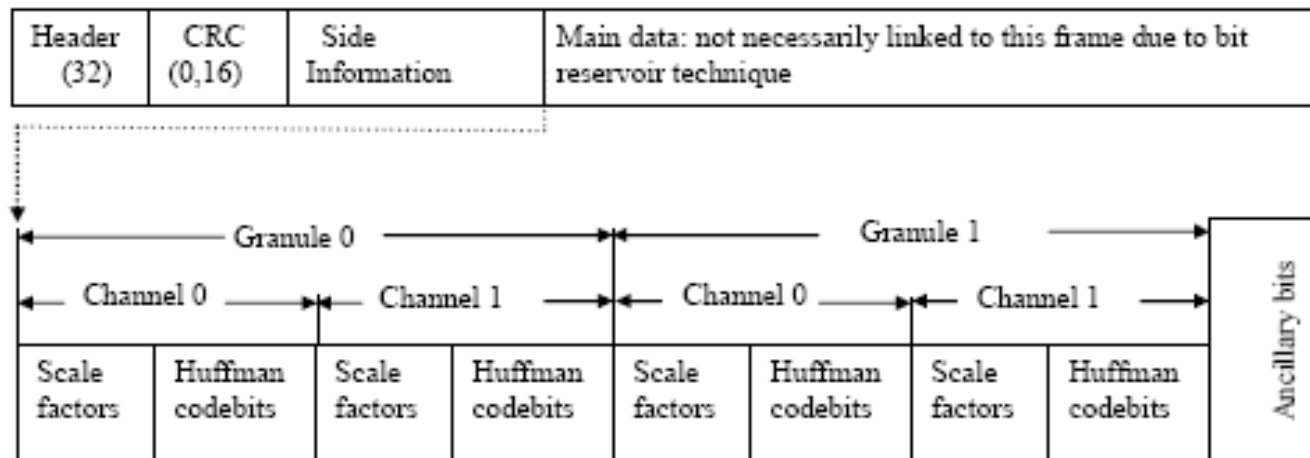
- 00000010101011101111110110
- 26 bits. TC = 75%



Processo de Compressão MP3



Encoding



http://www.mp3-tech.org/programmer/frame_header.html



Metadados ID3

- Padrão não prevê metadados
 - Solução: Usar 128 bits antes do primeiro frame (v1.0)

| Tamanho | Descrição |
|---------|------------|
| 3 | TAG |
| 30 | Título |
| 30 | Artista |
| 30 | Álbum |
| 4 | Ano |
| 30 | Comentário |
| 1 | Gênero |



Curiosidade: AAC

- Evolução do mp3
 - Fim da retro-compatibilidade com Layer I e II
 - Codificação para até 48 canais
 - De 576 para 1024 amostras por canal por frame
 - VBR
 - Uso de codebooks
 - Ao invés de codificar huffman para cada subbanda separadamente, codificar seqüências de subbandas
 - Dissociação inter-frames
 - Predição

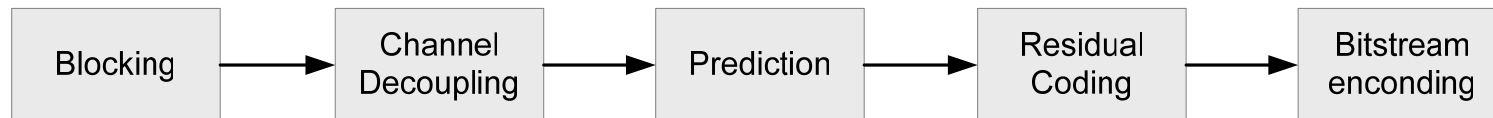


Curiosidade: AAC

- Diversos modos com complexidade variável
 - MPEG-2 AAC LC / Low Complexity
 - MPEG-2 AAC Main
 - MPEG-2 AAC SSR / Scalable Sampling Rate
 - MPEG-4 AAC LC / Low Complexity
 - MPEG-4 AAC Main
 - MPEG-4 AAC SSR / Scalable Sampling Rate
 - MPEG-4 AAC LTP / Long Term Prediction
 - MPEG-4 AAC HE / High Efficiency
 - MPEG-4 AAC LD / Low Delay

Compressão sem perdas

- Os compressores em geral possuem os seguintes passos



- Exemplo prático FLAC



Compressão sem perdas (FLAC)

- *Blocking*
 - Separação do conteúdo em uma série de blocos contíguos
 - Tamanho do bloco pode variar de 16 a 64k amostras
 - Na versão atual, o tamanho do bloco é fixo,
 - Depende da frequência de amostragem e número de canais
 - Blocos possuem Sub-bloco
 - Informações sobre um canal específico
 - Blocos são codificados em Frames
 - Cabeçalho
 - taxa de amostragem, resolução, CRC do frame etc.
 - Os frames são independente entre si.
 - Sub-frame: contém o dado codificado e um header que indica a predição usada na compressão



Compressão sem perdas (FLAC)

- Dissociação inter-canais
 - Remoção redundância de informação causada pela correlação entre os canais (estéreo)
 - Executa 4 transformações, usa a melhor delas.
 - **Canais independentes**: codificados independentemente
 - **Média-Lado (*mid-side*)**: Os Canais são transformados em média (média dos 2 canais) e side (canal esquerdo menos direito)
 - **Esquerda-Lado/Direita-Lado**: Codifica um canal (esquerda ou direita) e a diferença entre este e o canal não codificado. Geralmente, dão melhores resultado.

Compressão sem perdas (FLAC)

■ Predição

- Função matemática para tentar descrever o sinal
- 4 funções existentes
 - **Zero**: prediz sempre zero. Resíduo = Entrada
 - **Constante**: Usado quando todas as amostras de uma canal possuem o mesmo valor, e.g., silêncio.
 - **“Preditor” Linear Fixo**: função linear com coeficientes fixos que prediz no máximo 4 amostras subseqüentes
 - **Predição FIR Linear**: Utiliza um filtro de até 32 estágios.
 - Os coeficientes do filtro são achados a partir a da recursão de Levinson-Durbin.
 - Acha a solução para o sistema de equações referentes a uma matriz constante diagonal
 - Algoritmo $O(n^2)$

$$\begin{bmatrix} a & b & c & d & k \\ f & a & b & c & d \\ g & f & a & b & c \\ h & g & f & a & b \\ j & h & g & f & a \end{bmatrix}$$

Compressão sem perdas (FLAC)

- Codificação Residual
 - Codificação da diferença (resíduo) entre o que foi predito e a entrada
 - FLAC utiliza o código Rice para codificar os resíduos de predição (passo anterior)
- Dado um número N (de k bits) a ser codificado e m um número de bits que divide N , a codificação Rice (R) é representada por
 - $\langle \text{Unário}(N \text{ div } M) \rangle \langle \text{Binário}(N \text{ mod } M) \rangle$
 - Ex.: $N=23$ (10111) , $m=32 \rightarrow R= 11111110110$

Compressão sem perdas (FLAC)



- Por que Rice?
 - É ótimo para distribuições geométricas
 - $\Pr[x] = (1 - p)^{x-1}p$
 - Huffman é ótimo nessa distribuição também
 - Mas não tem uma forma simples de decodificar (encontrar onde começa e termina cada código)
- Para a escolher m , FLAC utiliza 2 métodos
 - **Método 1:** m é escolhido baseado na variância do sinal residual. O bloco inteiro é codificado com esse m .
 - **Método 2:** um bloco é particionada em varias partes de tamanhos distintos. m é escolhido pra cada partição e baseado na média do resíduo daquela partição especifica