

# Multimédia

## Compressão de Dados Multimédia: Áudio Digital

Prof. Dr. Rui Pedro Paiva

---

# Sumário



- ◆ Bibliografia
- ◆ Codecs de Áudio Digital
  - Sem compressão
    - ◆ WAV PCM
  - Destrutivos
    - ◆ mp3, AAC, outros
  - Não destrutivos
    - ◆ FLAC, ALAC, outros
- ◆ Formatos de Áudio Digital
  - MIDI, MusicXML, Outros



# Bibliografia

---

- ◆ Li et al (2014). “Fundamentals of Multimedia”, Springer.
- ◆ Peter Symes (2003). “Digital Video Compression”, McGraw-Hill.
  - Capítulo 20

# Codecs de Áudio Digital

- ◆ Vários codecs/formatos disponíveis, tanto para áudio digital como para som simbólico, destrutivos e não destrutivos
  - Áudio digital
    - ◆ Codecs/formatos “abertos”
      - mp3, aac, flac, ...
    - ◆ Codecs proprietários
      - alac (codec proprietário da Apple)
  - Som simbólico
    - ◆ Formatos “abertos”
      - midi, musicxml, ...
    - ◆ Formatos proprietários
      - E.g., sib (Sibelius), ...

# WAV PCM

- ◆ Definição breve
  - Waveform Audio File Format (WAV)
  - Desenvolvido pela IBM e Microsoft

# WAV PCM – Aspectos Gerais

## ◆ Compressão

- Tipicamente, **sem compressão**
- Representação de dados segundo o padrão **PCM** (Pulse-Code Modulation)
  - ◆ Amostragem e quantização
  - ◆ (ver Secção 1.2 Fundamentos de Compressão de Dados Multimédia: Áudio Digital)

# WAV PCM – Aspectos Gerais

## ◆ Dimensão e bit rate

- Depende da frequência de amostragem e nr. de bits de quantização (e também do número de canais)
- Qualidade de CD: 44.1 kHz, 16 bits, stereo → Ficheiros resultantes **muito grandes**
  - ◆ Bit rate = 1.41 Mbps
  - ◆ Muito espaço em disco
  - ◆ Inadequado para transmissão via Internet

## ◆ Qualidade do som

- Depende da frequência de amostragem e nr. de bits de quantização (e também do número de canais)
- Qualidade de CD: semelhante a som analógico para a maioria dos ouvintes

# WAV PCM – Aspectos Gerais

## ◆ Variantes

- (ver Secção 2.0 Compressão de Dados Multimédia: Fundamentos)
- WAV com compressão (não destrutiva) DPCM
  - ◆ Taxa de compressão de cerca de 2:1 (em comparação com PCM)
- WAV com compressão (destrutiva) ADPCM
  - ◆ Taxa de compressão na ordem de 4:1, mas com alguma perda de qualidade (sobretudo nas frequências altas)
- WAV com compressão (destrutiva)  $\mu$ -law PCM
  - ◆ Taxas de compressão na ordem de 2:1, mas com alguma perda de qualidade
- É um contentor → pode conter diferentes codecs (e.g., mp3, embora não seja comum)



# Codecs Destrutivos

## ◆ Metodologia geral

- Compressão destrutiva (**lossy**)
- Compressão perceptual, tirando partido das limitações da audição humano
  - ◆ Limitações nas componentes de alta frequência, mascaramento de sons, limitações na percepção de som estereofónico, etc.

## ◆ Taxas de compressão

- Tipicamente 10:1

# Codecs Destrutivos

## ◆ Aplicações principais

- Armazenamento/transferência com restrições de espaço/largura de banda
- Áudio digital de uso doméstico

## ◆ Exemplos

- mp3, AAC, ...

# Codificação Perceptual de Áudio

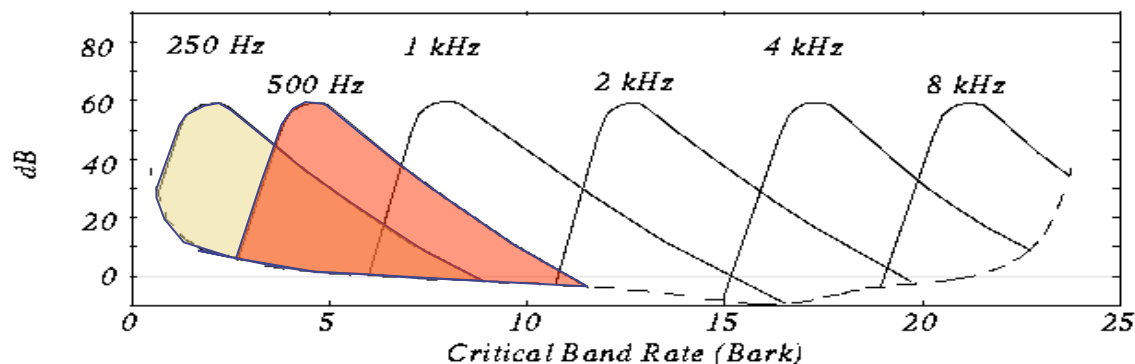
## ◆ Objectivo

- Base teórica dos codecs áudio destrutivos

# Codificação Perceptual de Áudio

## ◆ Banda crítica

- Ouvido humano comporta-se como um detector de frequências
  - ◆ Células diferentes respondem a frequência de formas diferentes (filtro passa-banda)
  - ◆ Cada célula tem uma gama de frequências à qual responde → **banda crítica**
  - ◆ Ouvido não distingue tão bem sons na mesma banda crítica → base da mascaramem de sons



# Codificação Perceptual de Áudio

## ◆ Banda crítica (cont.)

Band #	Lower bound (Hz)	Center (Hz)	Upper bound (Hz)	Bandwidth (Hz)
1	–	50	100	–
2	100	150	200	100
3	200	250	300	100
4	300	350	400	100
5	400	450	510	110
6	510	570	630	120
7	630	700	770	140
8	770	840	920	150
9	920	1000	1080	160
10	1080	1170	1270	190
11	1270	1370	1480	210
12	1480	1600	1720	240
13	1720	1850	2000	280
14	2000	2150	2320	320
15	2320	2500	2700	380
16	2700	2900	3150	450
17	3150	3400	3700	550
18	3700	4000	4400	700
19	4400	4800	5300	900
20	5300	5800	6400	1100
21	6400	7000	7700	1300
22	7700	8500	9500	1800
23	9500	10500	12000	2500
24	12000	13500	15500	3500
25	15500	18775	22050	6550

Largura das bandas críticas aumenta exponencialmente com a frequência

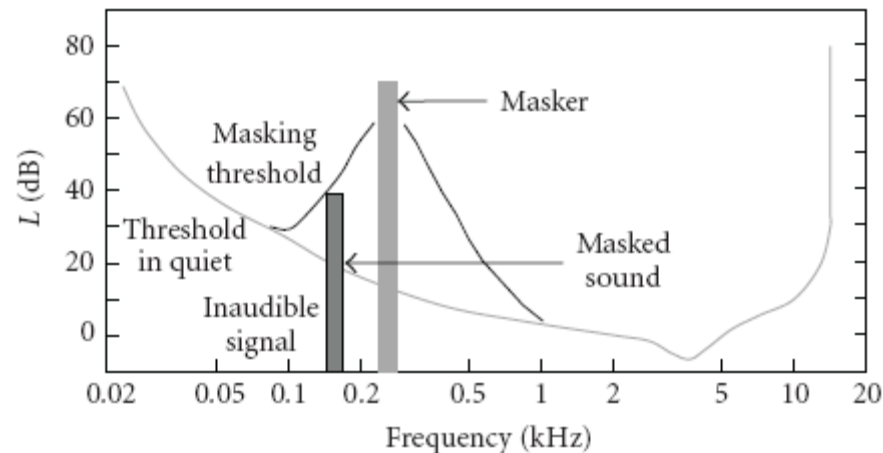
- até 500Hz: largura ~ 100Hz
- 7000 Hz: largura ~ 1300 Hz

© Ze-Nian Li 2014, p. 463

# Codificação Perceptual de Áudio

## ◆ Mascaramento

- Mascaramento de componente de frequência depende da sua frequência e intensidade
- **Mascaramento simultânea** (frequência)
  - ◆ Um som de amplitude elevada tende a mascarar um som de menor intensidade na mesma gama de frequências (banda crítica)

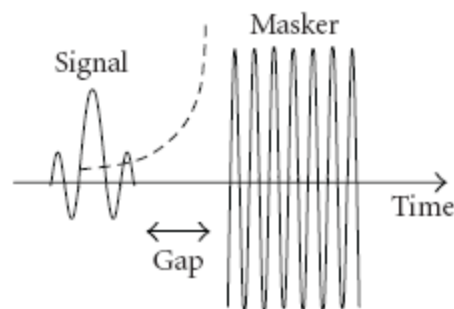


# Codificação Perceptual de Áudio

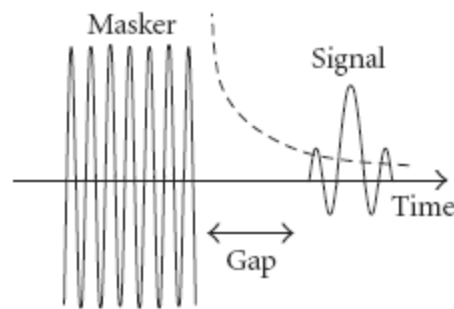
## ◆ Mascaramento (cont.)

### ■ Mascaramento temporal

- ◆ Um som de amplitude elevada tende a mascarar durante um determinado período temporal (antes, mas sobretudo depois) sons com amplitude inferior numa região vizinha de frequências



Backward masking



Forward masking



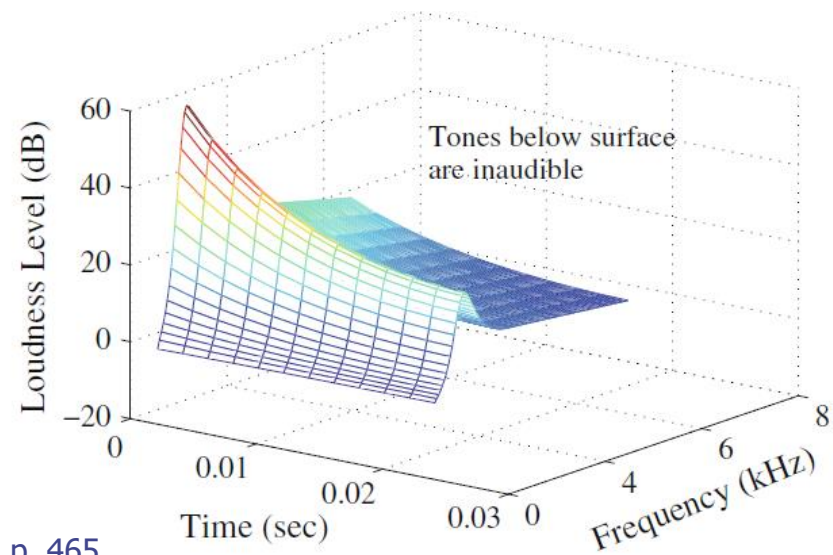
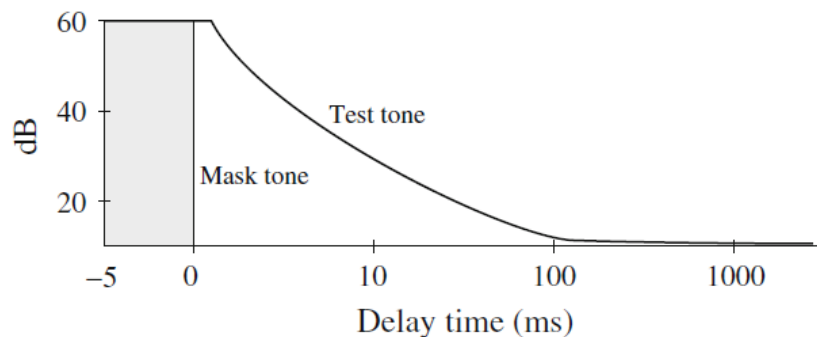
# Codificação Perceptual de Áudio

## ◆ Mascaramento (cont.)

### ■ Mascaramento temporal

- ◆ **Forward masking (post-masking): células ciliadas ficam saturadas** → necessitam de tempo para recuperar

- Tempo de recuperação depende da duração do tom mascarador



© Ze-Nian Li 2014, p. 465



# Codificação Perceptual de Áudio

- ◆ Mascaramento (cont.)
  - Mascaramento temporal
    - ◆ **Backward masking (pre-masking)**: causas menos conhecidas
      - Efeito temporal substancialmente menor (2-5 mseg)

# MP3

## ◆ Definição breve

- MPEG-1/2 Audio Layer 3 (mp3)
- Padrão para compressão de áudio criado pelo Movie Pictures Expert Group em 1992

# MP3 – Aspectos Gerais

## ◆ Princípios gerais

- O sistema auditivo humano **não consegue ouvir detalhe extremamente fino**
  - ◆ **Menor sensibilidade nas altas frequência**
    - Ouvido humano apenas consegue distinguir sons muito intensos
  - ◆ **Mascaragem de som**
    - Sons de baixa intensidade “**mascarados**” por sons de maior intensidade
  - ◆ **Insensibilidade a stereo**
    - Em algumas situações, pelas características do som, o ouvido humano não consegue aperceber-se da direcção de onde provém
- Compressão perceptual: reduzir **redundância perceptual**
  - ◆ **Compressão destrutiva**

# MP3 – Aspectos Gerais

## ◆ Método de Compressão

- Algoritmos de **compressão perceptual** (destrutiva) e **entrópica** (não destrutiva)
- Compressão perceptual
  - ◆ Tira partido das limitações da audição humana, descartando informação não perceptível (pela maioria dos utilizadores)
    - **Modelos psicoacústicos** para identificação **de sons menos relevantes perceptualmente**
    - Menor detalhe na representação de
      - Componentes de **alta frequência** ( $> 16\text{Hz}$ , tipicamente)
      - Sons “**mascarados**” por sons de maior intensidade
    - **Conversão para sinal mono** em situações de insensibilidade a stereo

# MP3 – Aspectos Gerais

## ◆ Método de Compressão

### ■ Transform Coding

- ◆ Teoria da Informação → Mais eficiente codificar vectores do que escalares (correlação temporal, i.e., entre valores próximos)
  - → Transformar blocos, e.g., blocos de um áudio original,  $A_O$ , num áudio transformada,  $A_T$ , usando uma transformada  $T$

$$A_T = T(A_O)$$

- ◆ Modified DCT (**MDCT**) → descorrelação em frequência
  - Evidenciar energia em diferentes frequências do som
  - **Quantização adaptativa**
    - Mais forte nos componentes de frequência menos importantes (altas frequências, sons mascarados)
      - → Reduzir o número de bits para as representar
  - **Principal operação destrutiva de todo o processo**

# MP3 – Aspectos Gerais

## ◆ Parâmetros fundamentais

### ■ **Frequência de amostragem**

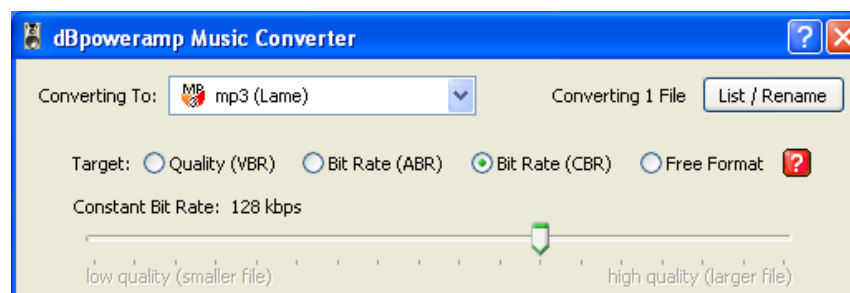
- ◆ 16 a 48 kHz

### ■ **Nr. de canais**

- ◆ 2 canais na versão MPEG-1 e 5.1 na MPEG-2

### ■ **Bit rate**

- ◆ Alta (320 Kbps) → pouca compressão destrutiva → qualidade elevada
- ◆ Média (128 Kbps) → pouca/média compressão destrutiva → qualidade elevada/média
- ◆ Baixa (64 Kbps) → muita compressão destrutiva → qualidade baixa



# MP3 – Aspectos Gerais

## ◆ Parâmetros fundamentais (cont.)

### ■ Tipo de bit rate

- ◆ Constant bit rate (CBR),
  - Simples e mais rápida
  - Menos flexível
- ◆ Variable bit rate (VBR)
  - Tira partido do dinamismo do som → bit rates mais altas em regiões mais complexas e mais baixas em regiões menos complexas (silêncio, poucos instrumentos)
  - Qualidade global aumenta
- ◆ Average bit rate (AVR)
  - Anterior, com especificação da bit rate média



# MP3 – Aspectos Gerais

## ◆ Dimensão e bit rate

- Depende da **bit rate** especificada
- Ficheiros resultantes **significativamente mais pequenos** que WAVE PCM com qualidade de CD (tipicamente de **1:8 a 1:12; 128 kbps ~ 1:11**)
  - ◆ Estas taxas permitem obter relações da ordem de 1 Mbyte por minuto de música. Bit rates típicas: 96 kbps, 128 kbps, 192 kbps



# MP3 – Aspectos Gerais

## ◆ Qualidade do som

- Depende da **bit rate** especificada
- Em termos **perceptuais** será, idealmente, qualidade de CD (mesmo sem 16 bits por amostra, stereo)
  - ◆ Utiliza métodos de compressão pensados para música (gama de frequências ampla, áudio polifónico, ...)
- Tipicamente, **128 kbps** aceitável para a maior parte dos utilizadores
  - ◆ Actualmente, 192 kbps mais comum devido ao aumento de espaço em disco
  - ◆ Máximo: 320 kbps (a partir daí, zona do lossless)
- Distorções por vezes captadas
  - ◆ Falta de “brilho”: componentes de alta frequência eliminados
  - ◆ Sons sibilantes: ruído de quantização
  - ◆ Pré-eco: clips com ataques abruptos (percussão) e algo aleatórios

# MP3 – Aspectos Gerais

## ◆ Qualidade do som (cont.)

- Avaliação perceptual da qualidade (testes de audição)
  - ◆ 5.0 = “Transparente”: diferenças indetectáveis face ao original; equivalente a qualidade de CD com 14 a 16 bits de quantização
  - ◆ 4.0 = Diferença perceptível mas não incomodativa
  - ◆ 3.0 = Ligeiramente incomodativo
  - ◆ 2.0 = Incomodativo
  - ◆ 1.0 = Muito incomodativo
- Exemplo: bit rate = 64Kbps → qualidade mp3 de 3.6 a 3.8
  - ◆ mp2: 2.1 a 2.6

# MP3 – Aspectos Gerais

## ◆ Encoders

- **Liberdade de implementação**, desde que em conformidade com a especificação: resultado deve ser um bitstream interpretável por qualquer decodificador
  - ◆ → Velocidades e qualidades diferentes
- Alguns encoders otimizados para bit rates elevadas (e.g., lame), outros para baixas
  - ◆ Exemplos de codecs mp3: lame, blade, Xing, ...
- Qualidade dos encoders tem vindo a melhorar
- Em 1998:
  - ◆ MP2 192 kbps ~ mp3 128 kbps ~ AAC 96 kbps
- Comparação
  - ◆ Listening tests

# MP3 – Aspectos Gerais

## ◆ Popularidade

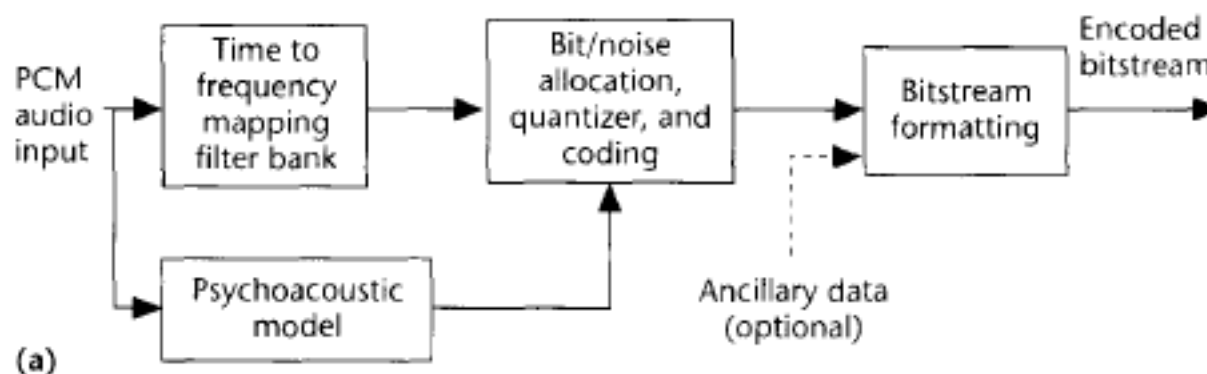
- 2ª metade dos anos 1990: sistemas peer-to-peer de partilha de ficheiros (e.g., Napster 1999)
- Pouco espaço em disco: ~ 500 MB
- Audio players em tempo real, e.g., Winamp (1997)

## ◆ DRM: Não

- Violações generalizadas de direitos de autor
  - ◆ → Processos contra Napster (acabou por ser fechado e reaberto posteriormente noutros moldes) e utilizadores individuais
- DRM:
  - ◆ Dados encriptados
  - ◆ Escuta apenas possível nos computadores autorizados
- Mas...
  - ◆ Utilizadores não podem escutar a música comprada em quaisquer dispositivos ou computadores...

# MP3 – Algoritmo de Compressão

## ◆ Codificação



## ◆ Decodificação

- inverter a ordem das setas

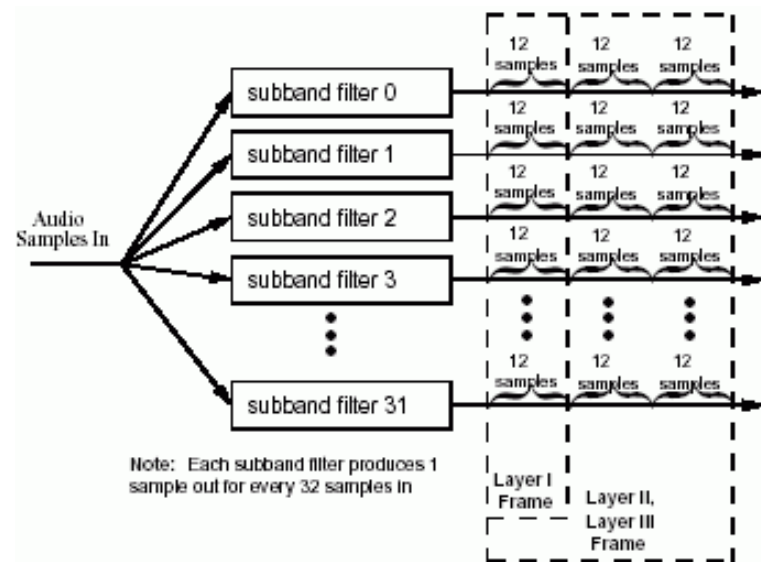
# MP3 – Algoritmo de Compressão

- ◆ **1) Divisão do sinal em frames (janelas) de curta duração**
  - Estacionaridade pode ser assumida
  - Tipicamente, 1152 amostras por canal (stereo, 5.1)

# MP3 – Algoritmo de Compressão

## ◆ 2) Filtragem passa-banda

- Divisão do sinal de entrada em múltiplas bandas de frequência
  - ◆ Espectro de frequências audíveis
- Aumenta a probabilidade de remoção de sons redundantes (simula bandas críticas do ouvido humano)
- Separação do sinal em **32 bandas**
  - ◆ Aproxima bandas críticas
  - ◆ Sub-amostragem em cada banda 32x
    - das 1152 amostras, 36 amostras para cada banda
- Inversão destrutiva...





# MP3 – Algoritmo de Compressão

## ◆ 3) Aplicação do modelo psicoacústico

- Questão: como representar o sinal de entrada com o número de bits disponíveis?
  - ◆ Ideia: sob certas condições de mascaramento, o sistema auditivo humano não escuta o ruído de quantização
  - ◆ Determinar limiares de mascaramento
  - ◆ Representar amostras com menos (ou zero) bits, em função do **SMR** (Signal to Mask Ratio) para cada banda
    - $SMR = \text{rácio entre a energia do sinal e o limiar de mascaramento}$
- O sinal de entrada passa simultaneamente por um **modelo psicoacústico**, o qual determina o SMR



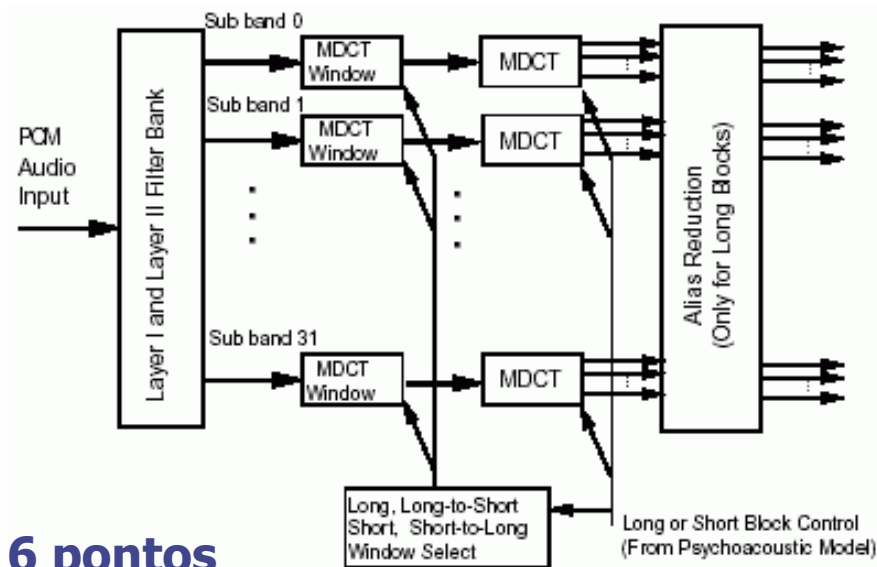
# MP3 – Algoritmo de Compressão

## ◆ 3) Aplicação do modelo psicoacústico (cont.)

### ■ a. Converter frame (1152 amostras) para o domínio de frequência

- ◆ Necessária melhor resolução em frequência para cálculo dos limiares
- ◆ Modified DCT

- Blocos longos → maior resolução em frequência: MDCT com **18 pontos**
- Transientes: blocos curtos → maior resolução temporal: MDCT com **6 pontos**

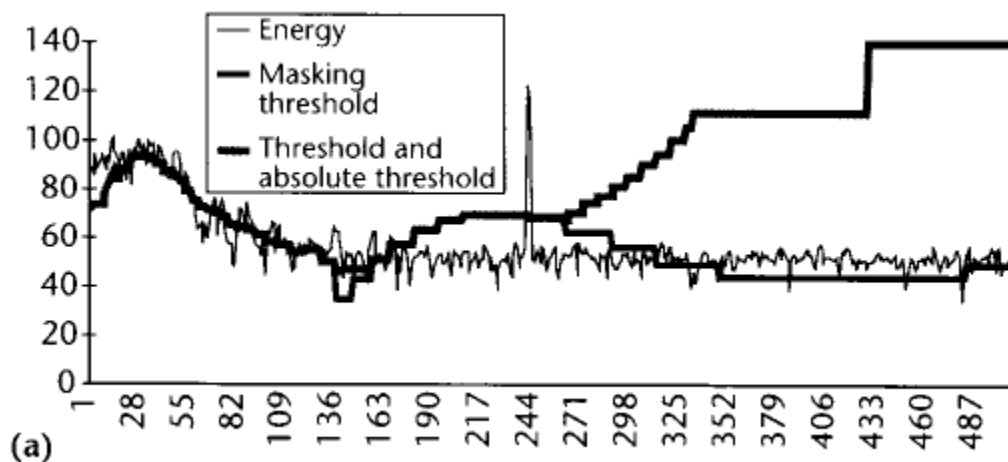


# MP3 – Algoritmo de Compressão

## ◆ 3) Aplicação do modelo psicoacústico (cont.)

### ■ b. Determinar o limiar de mascaramento global

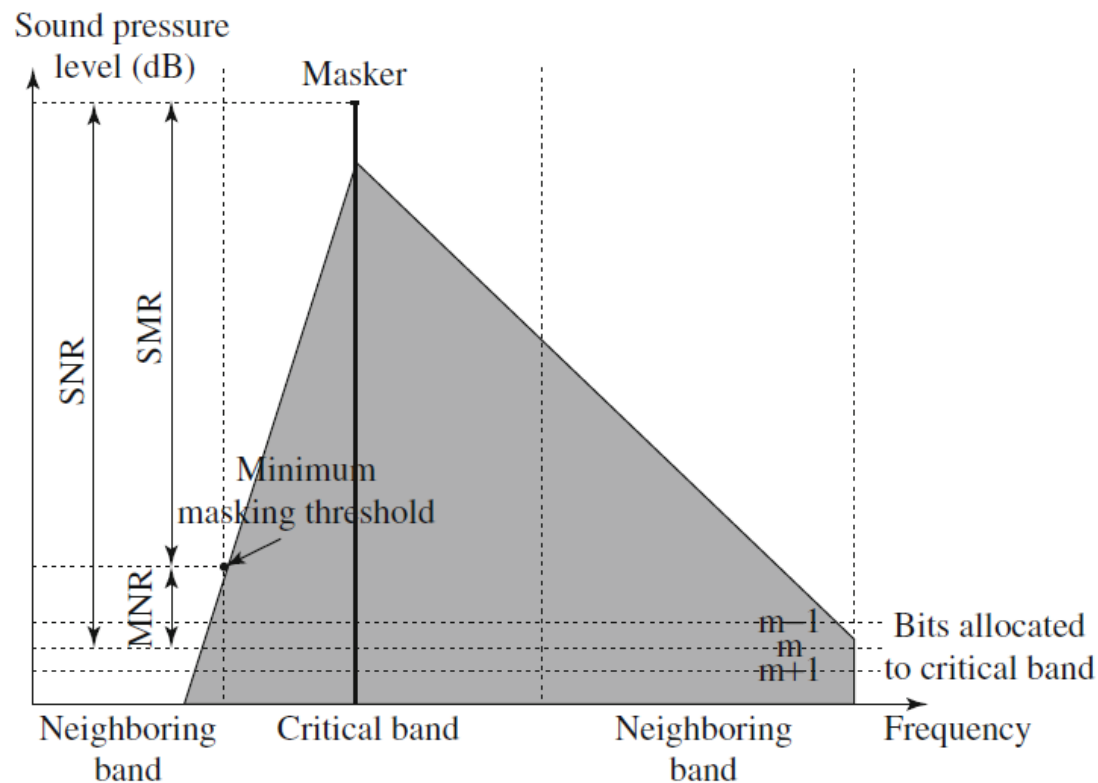
- ◆ Limiar de mascaramento em função da frequência
- ◆ Determinado pela combinação do efeito de mascaramento de todos os coeficientes da MDCT
  - → **Coração do codec**
  - Função de dispersão ao longo de cada banda crítica determinada empiricamente para os componentes do sinal



# MP3 – Algoritmo de Compressão

## ◆ 3) Aplicação do modelo psicoacústico (cont.)

### ■ c. Determinar o mask-to-noise ratio



© Ze-Nian Li 2014, p. 472

# MP3 – Algoritmo de Compressão

## ◆ 3) Aplicação do modelo psicoacústico (cont.)

### ■ c. Determinar o signal-to-mask ratio e noise-to-mask ratio

- ◆ SMR = Distância entre a intensidade do mascarador e o limiar de mascaramento (em DB)
- ◆  $SNR(m)$  = Signal-to-Noise Ratio com  $m$  bits = Distância entre a intensidade do mascarador e o ruído de quantização com  $m$  bits de quantização
- ◆  $MNR$  = Mask-to-Mask Ratio =  $SMR - SNR(m)$  = Distância entre o limiar de mascaramento e o ruído de quantização
  - **Ruído de quantização não será audível se  $MNR < 0$**
  - **$MNR$  negativo com valor alto  $\rightarrow$  alocar menos bits**

# MP3 – Algoritmo de Compressão

## ◆ 3) Aplicação do modelo psicoacústico (cont.)

### ■ d. Joint Stereo

- ◆ 2 modos de operação
  - Codificação **middle/side** (mid-side), se diferenças baixas entre os 2 canais; **L/R** caso contrário
    - Mid:  $X = (L + R) / 2$
    - Side:  $Y = L - R$  (valores baixos)
  - Não destrutiva
  - **Intensity Stereo** (IS): consiste em “juntar” (join) os dois canais num só, em regiões onde o sistema auditivo humano não necessite dos 2 canais (tipicamente acima dos 2 KhZ)
    - Som do “centro” → os dois canais são muito semelhantes →  $IS = L + R$
    - Destrutiva

# MP3 – Algoritmo de Compressão

## ◆ 3) Aplicação do modelo psicoacústico (cont.)

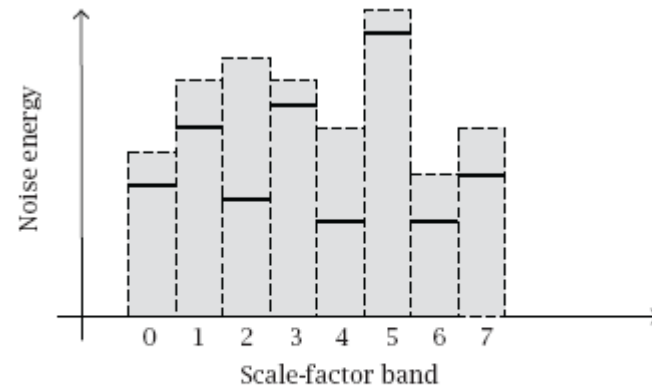
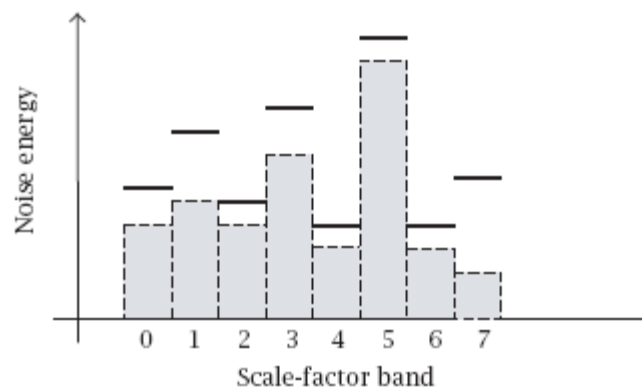
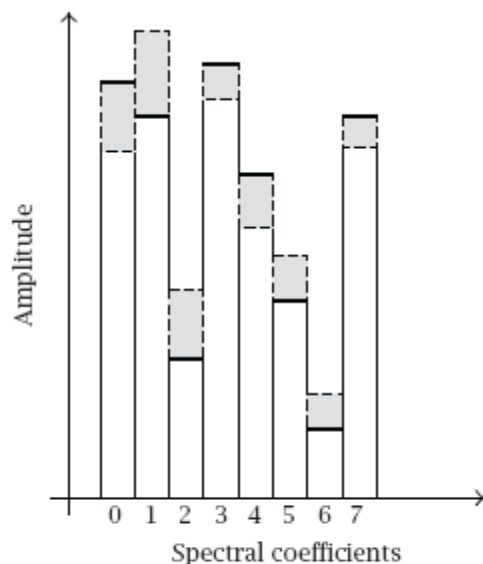
### ■ d. Joint Stereo (cont.)

- ◆ Comutação frame a frame
- ◆ Aplicação: todas as bandas ou nenhuma...
  - E se a localização do som for diferente em função da gama de frequências?

# MP3 – Algoritmo de Compressão

## ◆ 4) Bit allocation

- Usa os SMRs para decidir sobre a atribuição do número total de bits para **quantização** dos componentes de frequência em cada banda → nr. de bits necessários para que o **ruído de quantização** esteja **abaixo do limiar de audição**



— Original value  
--- Quantized value  
■ Quantization noise

— Masking threshold  
■ Quantization noise

— Masking threshold  
■ Quantization noise



# MP3 – Algoritmo de Compressão

## ◆ 4) Bit allocation (cont.)

- **a. Quantização dos coeficientes da MDCT**
  - ◆ Em função do NMR
    - Mais bits se NMR baixo
  - ◆ Menos bits para informação menos relevante
  - ◆ Número de bits com base na bit rate definida
- **b. Codificação Huffman** dos coeficientes quantizados



# MP3 – Algoritmo de Compressão

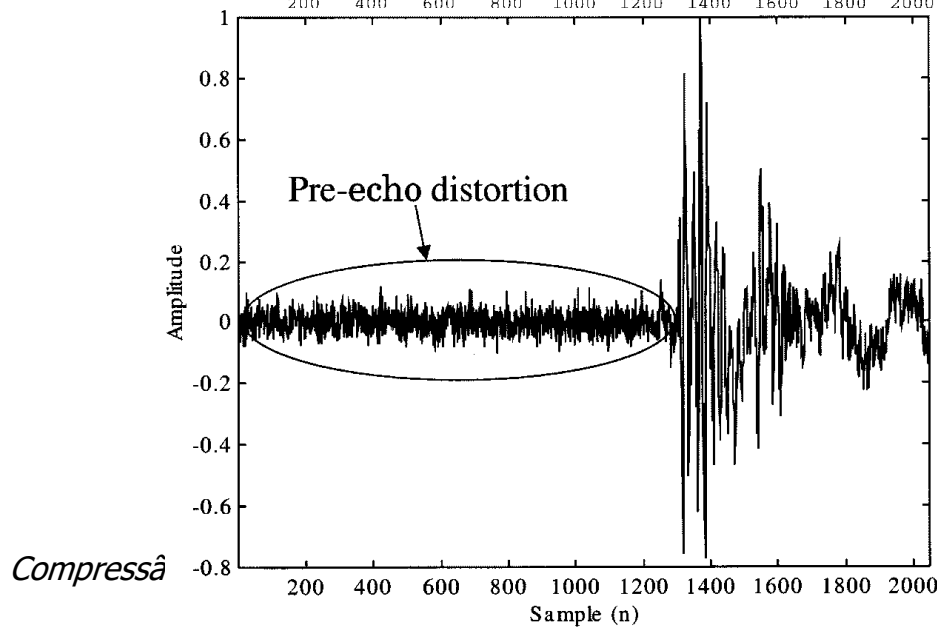
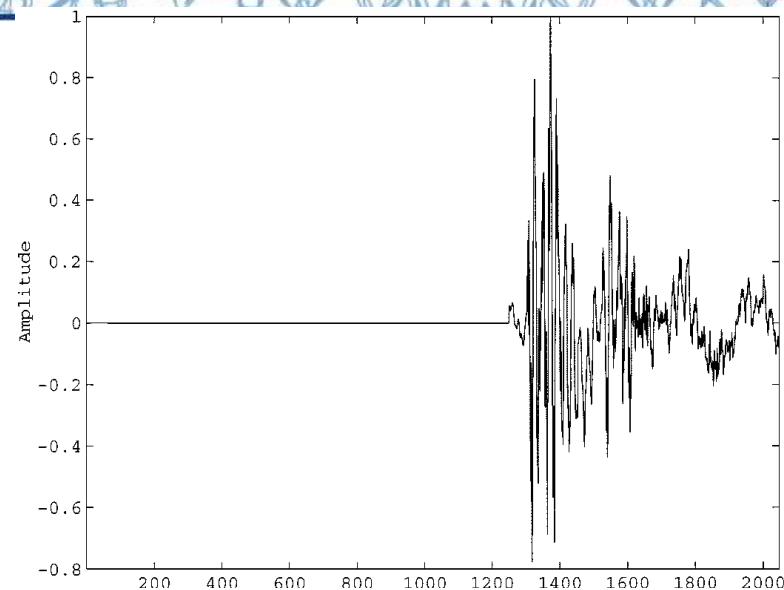
## ◆ Outros mecanismos

### ■ Reserva de bits

- ◆ Algumas frames são muito simples → não é necessário “gastar” todos os bits disponíveis → reserva para frames mais complexas
- ◆ Frames complexas: podem usar a reserva de bits
- ◆ Reserva nula → degradação mais notória da qualidade em zonas complexas

# MP3 – Limitações Principais

- ◆ Dimensão das frames pouco flexível → pré-eco
  - Erro na MDCT propaga-se por toda a janela temporal
  - Ruído “antes” do próprio sinal
  - Mais notório antes do que depois devido às características do *backward temporal masking*



# MP3 – Limitações Principais

- ◆ Joint stereo em todas as bandas
- ◆ Esquema híbrido de transformação
  - Filtragem passa-banda + MDCT: compatibilidade com mp1 e mp2
    - ◆ Filtragem inversa com reconstrução imperfeita
  - Codecs modernos utilizam MDCT directamente

# AAC

## ◆ Definição breve

- Advanced Audio Coding (AAC)
- Sucessor do mp3
  - ◆ Especificado como Part 7 do standard MPEG-2 (1997) e como Parte 3 do standard MPEG-4 (1999)
  - ◆ Referido como MPEG-4 AAC ou simplesmente AAC
  - ◆ → Codec destrutivo baseado também em compressão perceptual

# AAC – Aspectos Gerais

## ◆ Melhorias face ao mp3

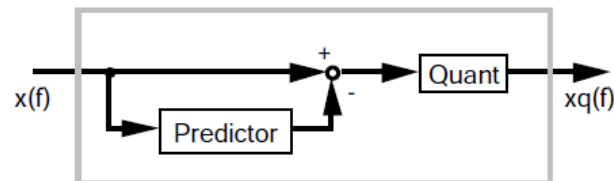
- **Melhor qualidade média** que mp3 para a mesma bit rate
  - ◆ Especialmente em **bit rates < 128 kbps**
    - Em geral, 96 kbps AAC ~ 128 kbps mp3
- **MDCT pura**, sem banco de filtros → inversão perfeita
- Bit rates arbitrárias e **janelas de duração variável**
  - ◆ Mais pequenas em zonas transitórias
  - ◆ Maiores em zonas estacionárias
- Mais frequências de amostragem (de 8 a 96 kHz)
- Até 48 canais (mp3: 2 canais na versão MPEG-1 e 5.1 na MPEG-2)
- Melhor manipulação de **frequências acima de 16 kHz**
- **Joint stereo mais flexível** (por bandas de frequência)

# AAC – Aspectos Gerais

## ◆ Melhorias face ao mp3 (cont.)

### ■ Temporal Noise Shaping (TNS)

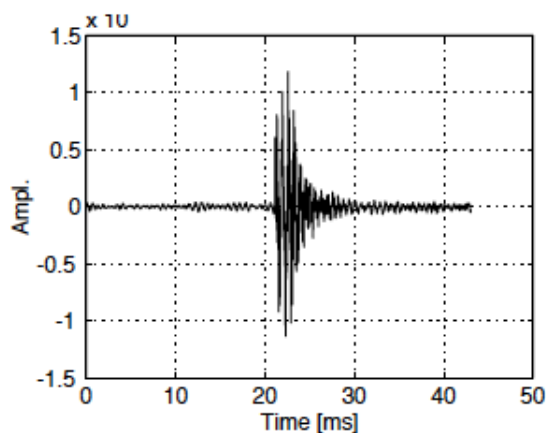
- ◆ Objectivo: Melhorar tratamento de transitórios muito rápidos → controlar pré-eco nos ataques de sinais
- ◆ Modelo de previsão forward ( $D^*PCM$ ) dos coeficientes da MDCT + quantização
  - → na desquantização, a forma do sinal de ruído no domínio temporal será semelhante à forma do sinal original (e não aproximadamente constante em amplitude)



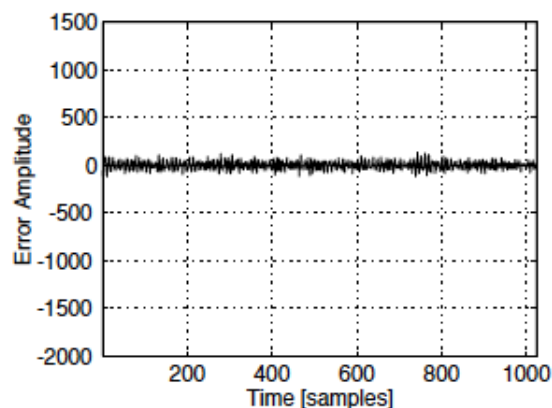
Open-loop predictive coding  
© Jürgen Herre

# AAC – Aspectos Gerais

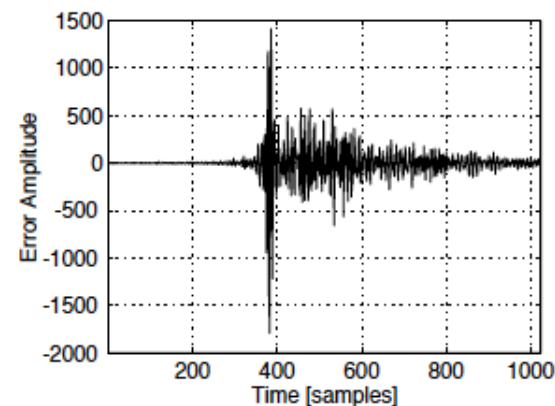
- ◆ Melhorias face ao mp3 (cont.)
  - Temporal Noise Shaping (TNS)



Castanholas  
© Jürgen Herre



Ruído de quantização sem TNS  
© Jürgen Herre



Ruído de quantização com TNS  
© Jürgen Herre

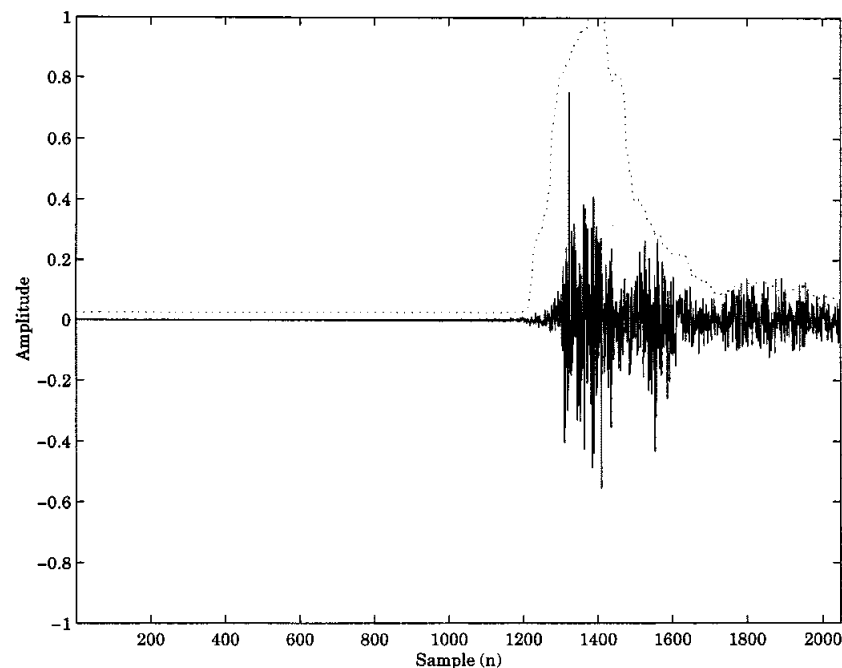


# AAC – Aspectos Gerais

## ◆ Melhorias face ao mp3 (cont.)

### ■ **Perceptual Noise Substitution (PNS)**

- ◆ Bandas de frequência em que o sinal seja tipo ruído → usar modelo de ruído
  - Percepção de ruído: contorno importante, mas sinal exacto não tão importante
- ◆ Vantagem: codificados apenas os (poucos) parâmetros do modelo em vez de todos os coeficientes da MDCT



# AAC – Aspectos Gerais

## ◆ Melhorias face ao mp3 (cont.)

### ■ Long-Term Prediction (LTP)

- ◆ Modelos de previsão → maior eficiência de codificação, principalmente para **sinais estacionários**
- ◆ Reduz redundância em frames sucessivas
- ◆ Armazenam-se coeficientes do modelo de previsão em vez do sinal

# AAC – Aspectos Gerais

## ◆ Outras funcionalidades

- Suporte de DRM (Digital Rights Management)

## ◆ Utilização mais comum

- iPod e iTunes da Apple (extensão .m4a ou .m4p)
  - ◆ AAC 128 kbps
  - ◆ iTunes: DRM para “proteger” ficheiros → .m4p
- Apple iPhone
- Sony Playstation 3
- Nintendo Wii
- MPEG-4 video
- Suportado por diversos dispositivos móveis

## ◆ Definição breve

- Windows Media Audio (WMA)
- Codec perceptual, desenvolvido pelo Microsoft, com a intenção de competir com o mp3
  - ◆ No entanto, não se tornou muito popular
    - Alguns sites de venda de música online utilizam-no em vez de mp3 devido às possibilidades de DRM (e.g., Sapo Música)
  - ◆ WMA 10 (anteriormente wmaPRO) posiciona-se como competidor do AAC

# WMA – Aspectos Gerais

## ◆ Possibilidades

- CBR e VBR
- Compressão não destrutiva: 470 a 940 kbps
- DRM

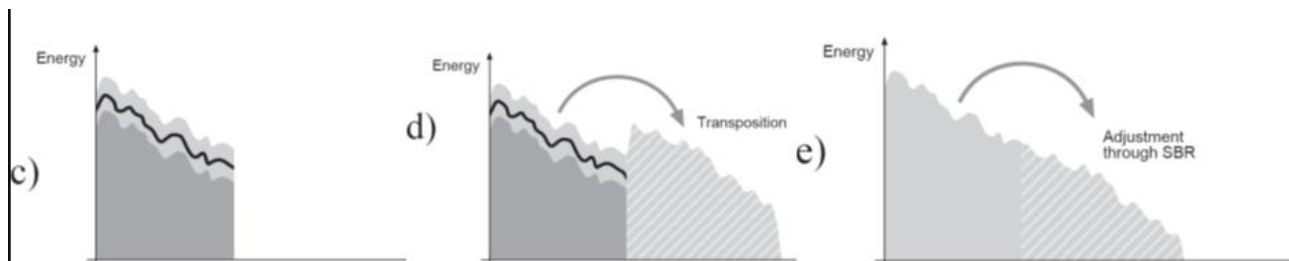
## ◆ Qualidade

- Semelhante a mp3 lame
  - ◆ Mas melhor em bit rates até 64 kbps
- Inferior a AAC

# Outros codecs perceptuais similares

## ◆ HE-AAC

- High-Efficiency AAC
- AAC+ SBR (Spectral Band Replication)



## ◆ Codecs livres

- Musepack
- **Ogg Vorbis**
  - ◆ Popularidade crescente: melhor qualidade que mp3 para as mesmas bit rates

# Codecs Não-Destrutivos

## ◆ Metodologia geral

- Compressão não-destrutiva (**lossless**)
- Utilização de **modelos de previsão**, e.g., DPCM, e **métodos de compressão entrópica**

## ◆ Taxas de compressão

- Tipicamente 2:1



# Codecs Não-Destrutivos

## ◆ Aplicações principais

- Arquivamento
- Áudio profissional de alta qualidade

## ◆ Exemplos

- FLAC, ALAC, Monkey's Audio, ...

## ◆ Definição breve

- Free Lossless Audio Codec (FLAC)
- Codec popular de compressão áudio open-source, criado em 2001, baseado no codec Shorten (já descontinuado)
- Compressão **não destrutiva**

# FLAC – Aspectos Gerais

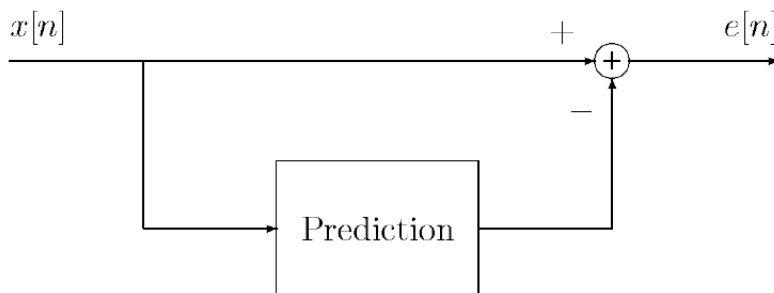
## ◆ Método de Compressão

### ■ Codificação **middle / side** (Mid-Side)

- ◆ Mid:  $X = (L + R) / 2$
- ◆ Side:  $Y = L - R$  (valores baixos)

### ■ Codificação **DPCM**

- ◆ **Modelos de previsão linear**: prever o valor de uma amostra com base em valores anteriores
  - Modelos: previsão linear fixa (até à 4ª ordem), previsão linear FIR (até à ordem 32)



$$e[n] = x[n] - Q \left\{ \sum_{k=1}^M \hat{a}_k x[n-k] - \sum_{k=1}^N \hat{b}_k e[n-k] \right\}$$

# FLAC – Aspectos Gerais

## ◆ Método de Compressão (cont.)

- **Compressão entrópica não destrutiva** dos resíduos
  - ◆ Resíduos codificados pelo algoritmo entrópico de **Golomb-Rice**
    - Alfabeto de resíduos segue uma distribuição quase-geométrica (ideal para Golomb-Rice)
      - Resíduos baixos mais frequentes que altos → menos bits
  - ◆ **RLE** utilizado em blocos com valores idênticos, e.g., silêncio

# FLAC – Aspectos Gerais

## ◆ Dimensão e bit rate

- Ficheiros resultantes (música) comprimidos com dimensão típica ~ **50%** do tamanho original
  - ◆ Bit rate ~ qualidade de CD / 2 = 705.6 kbps
  - ◆ Maior compressão para fala

## ◆ Qualidade do som

- Compressão não destrutiva → qualidade permitida pela codificação PCM subjacente

# FLAC – Aspectos Gerais

- ◆ Digital Rights Management (DRM)
  - Não
- ◆ Possibilidades
  - Streaming
  - Tagging
  - Imagens associadas (capas, etc.)
  - Codec livre e open-source → suportado por muitas aplicações de software
- ◆ Limitações
  - Suporte para dispositivos móveis limitado

# ALAC

## ◆ Definição breve

- Apple Lossless Audio Codec (ALAC)
- Codec proprietário da Apple, criado em 2004, também conhecido como Apple Lossless ou ALE (Apple Lossless Encoder)
  - ◆ Extensão .m4a



# ALAC – Aspectos Gerais

## ◆ Método de Compressão

- Não destrutiva
- Pouca informação disponível, embora tenha sido projectado para ser **baseado no FLAC**, com melhor desempenho na descodificação

## ◆ Dimensão e bit rate

- Ficheiros resultantes (música) comprimidos com dimensão típica ~ **50%** do tamanho original
- Bit rate ~705.6 kbps

## ◆ Qualidade

- Compressão não destrutiva → qualidade permitida pela codificação PCM subjacente

# ALAC – Aspectos Gerais

## ◆ DRM

- Não suporta (embora seja possível devido ao contentor)

## ◆ Armazenamento

- Num contentor MPEG 4 (ver adiante) com extensão .m4a
  - ◆ Mas não é nenhuma variante do AAC

# ALAC – Aspectos Gerais

## ◆ Vantagens

- Decodificação rápida
  - ◆ Argumenta-se que o ALAC é mais leve que o FLAC
- Utilizável em iPods
- Codec *crackado* (reverse engineering, David Hammerton e Cody Brocious)
  - ◆ Código open-source do decodificador
- Tags ID

# Outros codecs não-destrutivos similares

## ◆ Monkey's Audio (.ape, 2000)

- Codec proprietário, desenvolvido por Matthew T. Ashland
- Método de Compressão
  - ◆ Não destrutiva
  - ◆ Codificação **middle / side**
  - ◆ Modelo de **previsão linear**
  - ◆ Codificação do erro com **Golomb-Rice**
- Dimensão e bit rate
  - ◆ Ficheiros resultantes comprimidos com rácios típicos de **2:1**
  - ◆ Bit rate ~ 705.6 kbps
- DRM: Não

# Outros codecs não-destrutivos similares

## ◆ Monkey's Audio (cont.)

### ■ Limitações

- ◆ Software proprietário (mas código disponibilizado em 2008)
- ◆ Decodificação mais lenta que FLAC (lenta em dispositivos móveis)
- ◆ Suporte limitado fora de ambiente Windows

# Outros codecs não-destrutivos similares

## ◆ MPEG-4 Audio Lossless Coding (ALS, 2005)

- Codec lossless desenvolvimento no âmbito do standard MPEG-4
- Compressão
  - ◆ Semelhante a FLAC
- Outras possibilidades
  - ◆ Até 32 bits de quantização
  - ◆ Frequências de amostragem arbitrárias
  - ◆ Streaming, ...

## ◆ E ainda outros...

- Shorten (um dos primeiros, descontinuado), TTA, WavPack, Real-Audio Lossless Format (RALF), WMA lossless, MPEG-4 Scalable to Lossless (SLS), ...

# Formatos de Áudio Digital

## ◆ WAVE (Waveform Audio File Format)

- Ver slides iniciais desta secção
- 1ª parte do ficheiro: informação de formatação
  - ◆ Identificação do codec
  - ◆ Número de canais
  - ◆ Frequência de amostragem
  - ◆ Número médio de bytes/s
  - ◆ Tamanho dos blocos
- 2ª parte: informação dependente do codec.
  - ◆ No caso do PCM é incluído apenas o número de bits de quantização.
  - ◆ Pode também ser incluído um conjunto de marcas, a definição de uma ordem de execução à custa dessas marcas e ainda uma conjunto de informação relativo a cada marca
  - ◆ Finalmente aparecem os valores de amostragem



# Formatos de Áudio Digital

## ◆ Outros

- ASF (Advanced Streaming Format)
  - ◆ Encapsula dados codificados com WMA possibilitando DRM e streaming

# Formatos de Som Simbólico

## ◆ MIDI

- (ver Secção 1.1 Fundamentos de Multimédia: Áudio Digital)
- Formato aberto

# Formatos de Som Simbólico

## ◆ MusicXML

- Formato aberto, baseado em **XML**
- Exemplo:
  - ◆ Representação de dó central (nota C4)

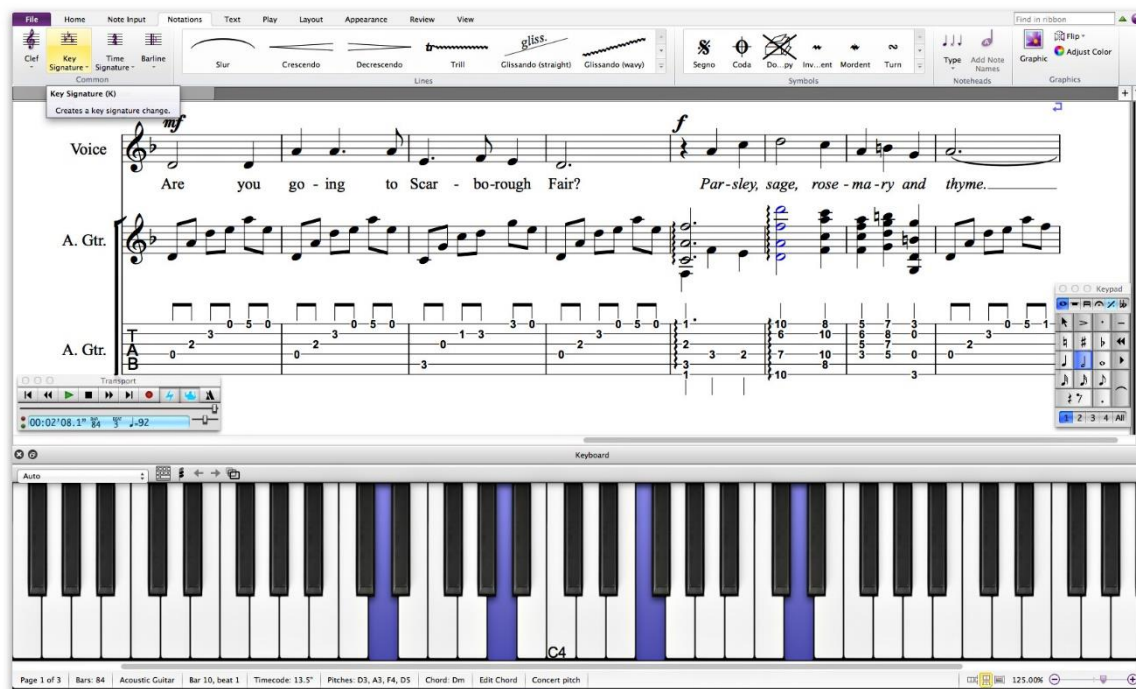
© <https://en.wikipedia.org/wiki/MusicXML>

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE score-partwise PUBLIC
    "-//Recordare//DTD MusicXML 4.0 Partwise//EN"
    "http://www.musicxml.org/dtds/partwise.dtd">
<score-partwise version="4.0">
  <part-list>
    <score-part id="P1">
      <part-name>Music</part-name>
    </score-part>
  </part-list>
  <part id="P1">
    <measure number="1">
      <attributes>
        <divisions>1</divisions>
        <key>
          <fifths>0</fifths>
        </key>
        <time>
          <beats>4</beats>
          <beat-type>4</beat-type>
        </time>
        <clef>
          <sign>G</sign>
          <line>2</line>
        </clef>
      </attributes>
      <note>
        <pitch>
          <step>C</step>
          <octave>4</octave>
        </pitch>
        <duration>4</duration>
        <type>whole</type>
      </note>
    </measure>
  </part>
</score-partwise>
```

# Formatos de Som Simbólico

## ◆ Outros

- sib (Sibelius)
  - ◆ Formato proprietário



©  
[https://www.sibelius.com/products/sibelius\\_first/images/cw\\_700x700\\_startcomposing\\_rightaway3.jpg](https://www.sibelius.com/products/sibelius_first/images/cw_700x700_startcomposing_rightaway3.jpg)