

Compressão de Áudio Digital

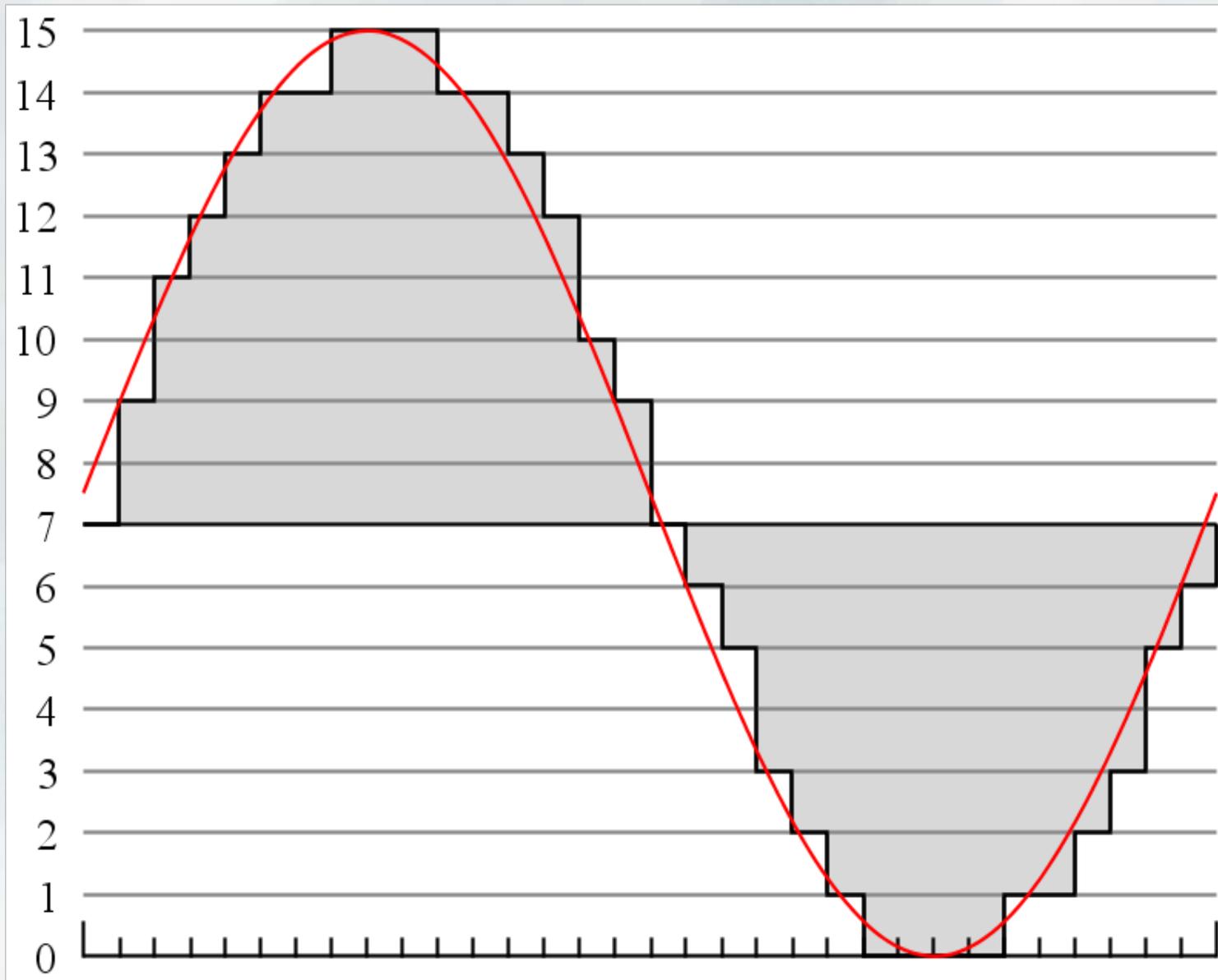
Aluno: Marcio Masaki Tomiyoshi
Supervisor: Marcelo Gomes de Queiroz

Introdução

- Motivações
- Com perdas (lossy)
 - MP3
- Sem perdas (lossless)
 - FLAC

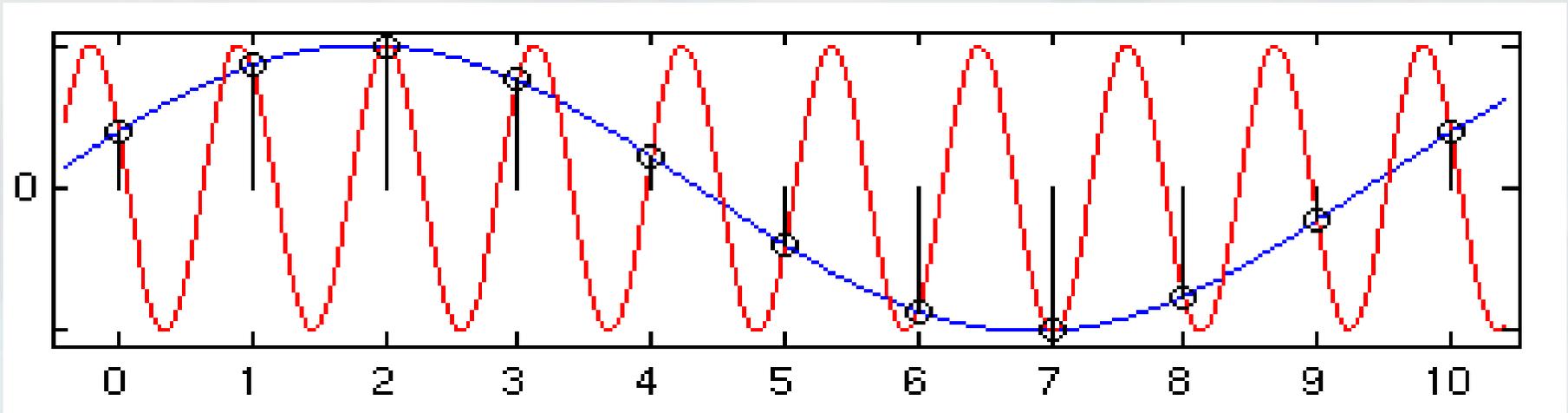
Representação Digital do Som

- Quantidade de canais
- Quantidade de bits por amostra
 - Número de intervalos para representar a amostra
- Taxa de amostragem
 - Indica quantas vezes por segundo a amplitude do sinal é armazenada



Representação de onda com 4 bits

Aliasing/Teorema de Nyquist



Para representarmos corretamente um sinal de frequência R Hz, precisamos ter ao menos uma taxa de amostragem $2R$ Hz.

Pulse Code Modulation (PCM)

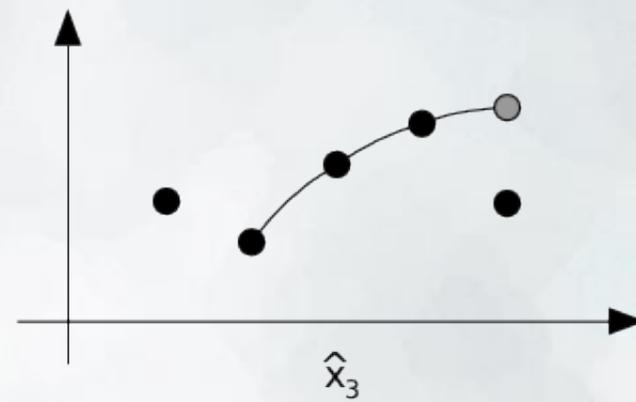
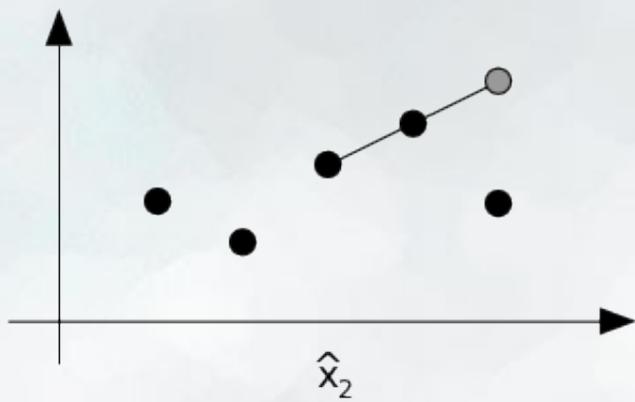
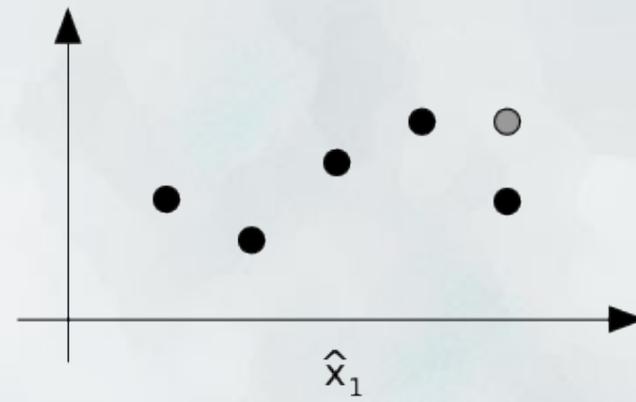
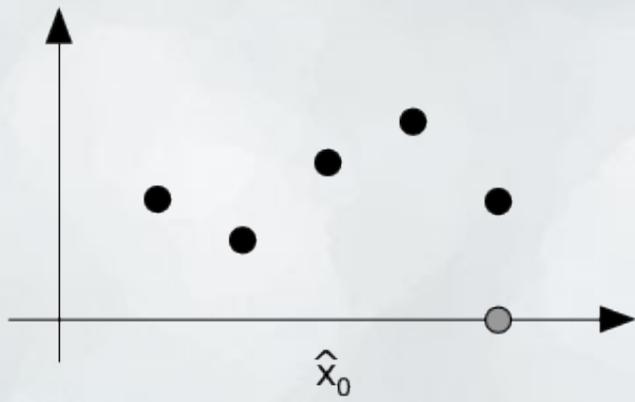
- Usado em CD, DAT, HiMD, DVD-Audio...
- Qualidade de CD
 - 44.1 KHz
 - 16 bits/amostra
 - 2 canais
- Taxa de bits do CD = 1411 kbps (~706 kbps por canal)

AudioPaK

- Armazena a diferença entre a amostra atual e um valor estimado a partir das anteriores
- Estimadores utilizados:

$$\left\{ \begin{array}{l} \hat{x}_0[n] = 0 \\ \hat{x}_1[n] = x[n-1] \\ \hat{x}_2[n] = 2x[n-1] - x[n-2] \\ \hat{x}_3[n] = 3x[n-1] - 3x[n-2] + x[n-3] \end{array} \right.$$

- Vantagem: fácil de calcular



Estimadores do AudioPaK

AudioPaK

- Codificação de Golomb/Rice
 - Divide o número a ser armazenado em:
 - Bit de sinal
 - m bits menos significativos
 - Codificação unária dos bits mais significativos restantes
 - Parâmetro m é estimado através da fórmula:
 - $m = \lceil \log_2 (E(|e_i[n]|)) \rceil$

AudioPaK

Exemplo para $m = 2$:

13_{10} ($= 1101_2$)

- Bit de sinal: 0
- m bits menos significativos: 01
- Codificação unária dos bits restantes ($11_2 = 3_{10}$): 0001

Código de Rice para 13 é 0010001

Assim, gastamos 7 bits ao invés de 16 na representação.

μ -law

- Usa uma escala logarítmica para quantização
- Representa sinais de baixa amplitude com maior precisão e sinais de grande amplitude com menor precisão
 - Sinais de grande amplitude tendem a mascarar o ruído introduzido
- Transformação realizada ($m = 255$):

$$\begin{cases} 255 - \frac{127}{\ln(1-m)} \times \ln(1+m|x|) & \text{para } x \geq 0 \\ 127 - \frac{127}{\ln(1-m)} \times \ln(1+m|x|) & \text{para } x < 0 \end{cases}$$

- 8 bits/sample
 - Comprime um CD para metade do tamanho

IMA ADPCM

- Armazena a diferença entre a amostra atual e o valor decodificado da amostra anterior
- Quanto maior a diferença, maior o erro introduzido
- Algoritmo vai se adaptando ao sinal de entrada, adaptando a escala de representação do erro ao longo do tempo
- 4 bits/sample
 - Comprime um CD para 25% do tamanho

MPEG-1

- Três *layers*
- Divisão do espectro de frequências em subbandas
- Explora limitações do ouvido humano
 - Regiões críticas
 - Mascaramento temporal/simultâneo
- Bitrate varia entre 32 e 384 kbps
 - Com cerca de 192 kbps no modo *layer III* é considerado comparável à qualidade de um CD, ficando em torno de 15% do tamanho original

Resultados

Codificador	Tamanho (em kB)	Taxa de compressão
PCM WAV	121723	100%
ZIP	115044	94,5%
AudioPaK*	90260	74,1%
RAR	89327	73,4%
FLAC	85135	69,9%
μ -law*	60862	50%
IMA ADPCM*	30535	25%
MP3 (LAME -V2)	17984	14,8%

Arquivo codificado em diversos formatos.

Codificadores implementados estão marcados com (*)

Conclusão

- Explorando as correlações entre as amostras em arquivos de áudio é possível comprimir sem perdas com grande eficiência.
- Modelos de compressão com perdas que se aproveitam de técnicas de psicoacústica conseguem uma maior taxa de compressão sem que haja uma queda muito grande na qualidade.

Referências

1. Imagens extraídas de <http://www.wikipedia.org>
2. M. Hans and RW Schafer. Lossless compression of digital audio. Signal Processing Magazine, IEEE, 18(4):21–32, 2001.
3. F.R. Moore. Elements of computer music. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1990.
4. D.Y. Pan. Digital Audio Compression. Digital Technical Journal, 5(2):28–40, 1993.

Dúvidas?