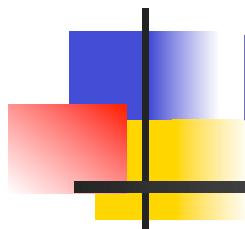


SCC0661 – Multimídia & Hipermídia



Prof.: Dr. Marcelo Manzato
(mmanzato@icmc.usp.br)

Aula 5 - Áudio

Instituto de Ciências Matemáticas e de Computação - ICMC
Sala 3-111

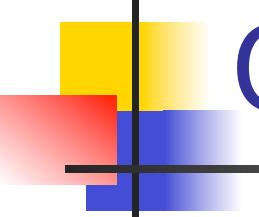


Sumário

- Características do Som.
- Digitalização.
- Compressão de Áudio.

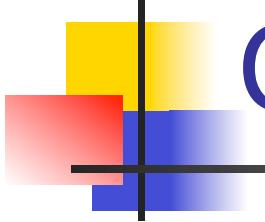


Características do Som



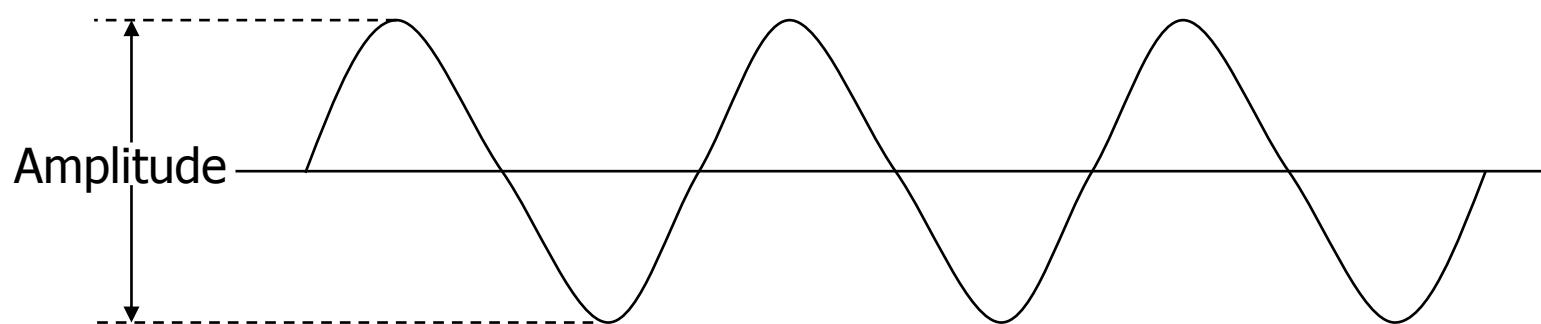
O quê é som?

- Som é um fenômeno físico produzido por variações (vibrações) na pressão do ar.
 - Cordas de violino, bater palmas, cordas vocais, ...
- Com as variações
 - as moléculas vizinhas vibram no ar criando um variação de pressão no ar à volta.
 - Essa alteração entre altas pressões e baixas pressões propaga-se no ar, em todas as direções, como uma onda (mecânica).

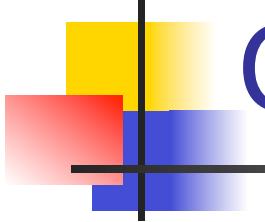


Características físicas do som.

- Som é uma onda mecânica.
 - Possui alguns aspectos, entre eles: amplitude e freqüência.



- Amplitude → Intensidade
 - Está relacionada ao volume do som. Quanto maior a amplitude, mais alto ouvimos o som.



Características físicas do som.

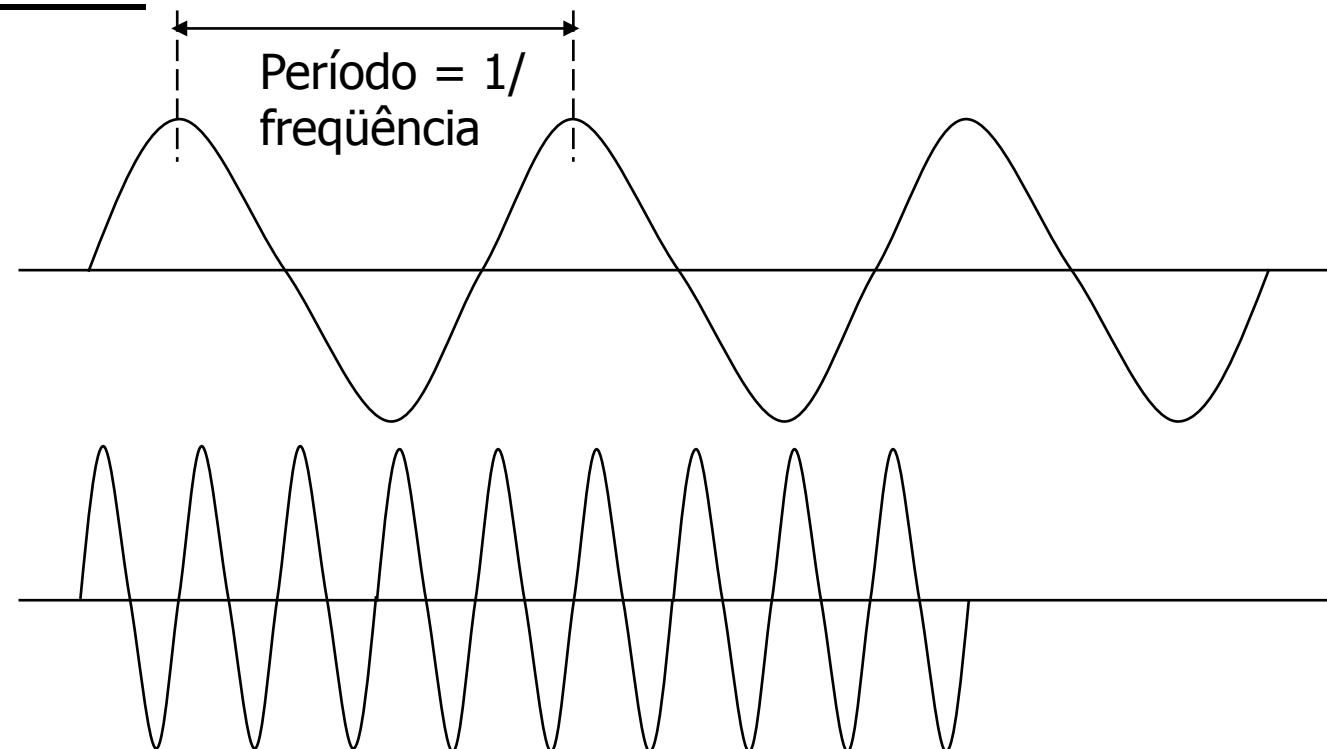
- Amplitude -> Intensidade
 - Medida em decibéis (dB).

Intensidade	Exemplos típicos
0dB	Límite da audição
25dB	Estúdio de gravação
50dB	Escritório
70dB	Conversação típica
90dB	<i>Home audio</i>
120dB	Límiar da dor
140dB	Show ao vivo, boate, etc.

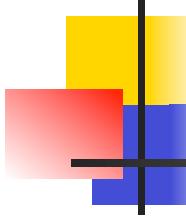


Características físicas do som.

Freqüência

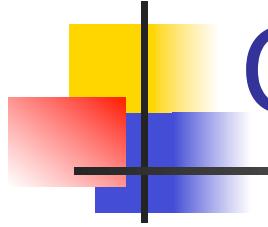


- Freqüência determina altura do som (altura \neq volume).
 - Freqüências altas = altura maior = sons agudos.
Freqüências baixas = altura menor = sons graves.



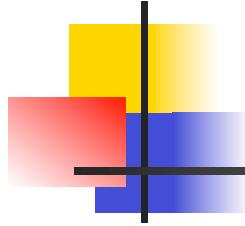
Características físicas do som.

Categoria	Intervalo de Freqüência
Infra-som	0 - 20 Hz
Som Audível	20 Hz - 20 KHz
Ultra-som	20 KHz - 1GHz
Hipersom	1 GHz - 10 GHz



Características físicas do som

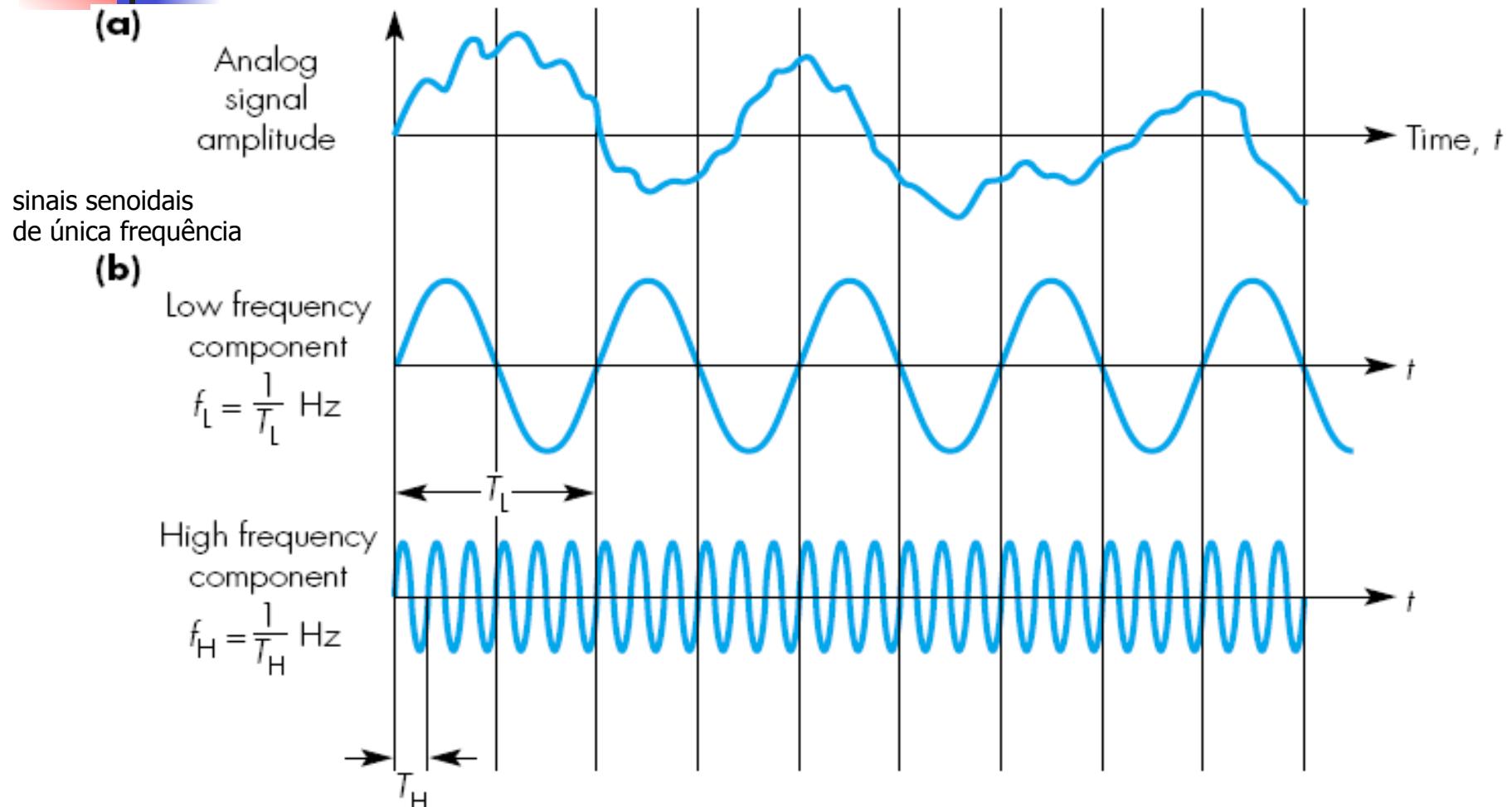
- Timbre
 - Característica sonora que permite distinguir sons de mesma frequência
 - Exemplo: nota LÁ (440 Hz) emitida de um violino e de um piano
 - O timbre é formado por duas características: forma de onda e envelope sonoro



Características físicas do som

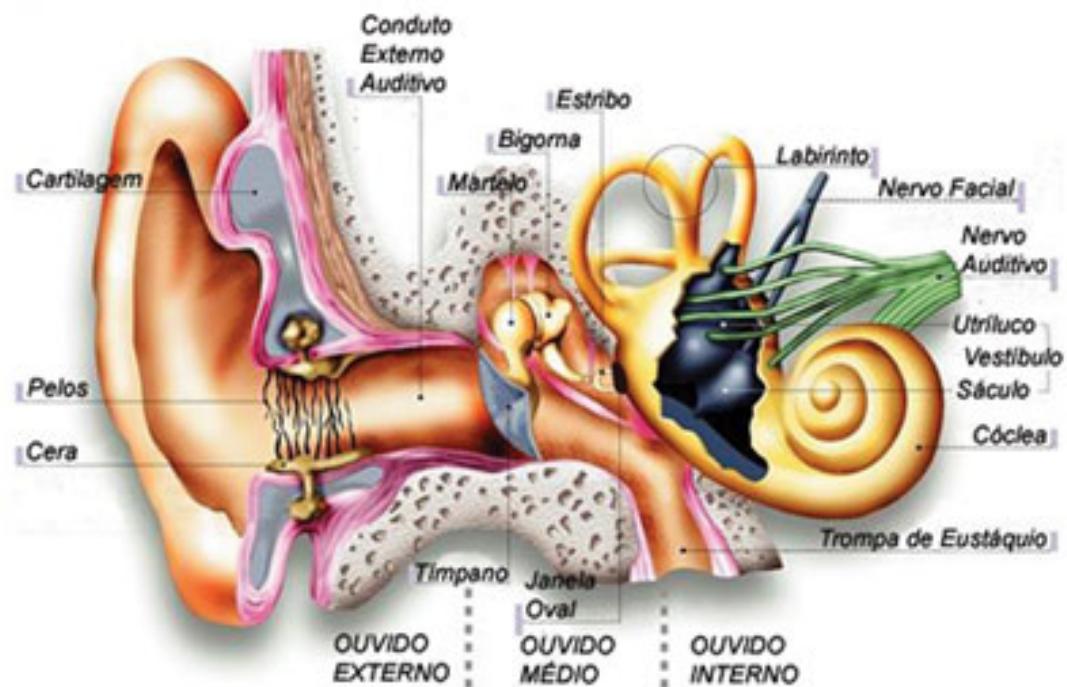
- Forma de onda
 - Som de um determinado instrumento produz ondas na frequência fundamental e várias frequências harmônicas com diferentes amplitudes
 - Combinação dessas frequências e amplitudes gera uma forma de onda irregular característica do instrumento
- Envelope sonoro
 - Ataque (Attack), Decaimento (Decay), Sustentação (sustain) e Relaxamento (release)

Características físicas do som



$T_{L/H}$ = time for one cycle = signal period

Como ouvimos sons?

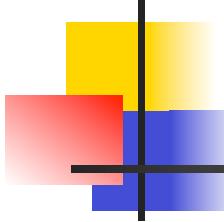


- As ondas sonoras atingem o tímpano.
- O tímpano faz os ossos do ouvido médio vibrarem.
- Essas vibrações são convertidas em impulsos nervosos que são transmitidos, via o nervo auditivo, para o cérebro.
- Quando esses impulsos chegam ao cérebro, “ouvimos” o som!



Como ouvimos sons?

- Assim, o ouvido funciona como um sensor ou transdutor que converte sons em estímulos nervosos que podem ser interpretados pelo cérebro.



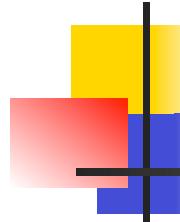
Som analógico



- Similarmente ao ouvido, o microfone é um transdutor.
 - Converte as variações de pressão do ar em sinais elétricos usáveis pelos equipamentos de áudio.
 - A saída de um microfone é uma voltagem elétrica analógica que varia no tempo do mesmo modo que as ondas mecânicas do som = **Sinal de Áudio**
- Alto-falantes: sinal elétrico → ondas mecânicas

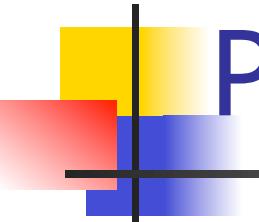


Digitalização

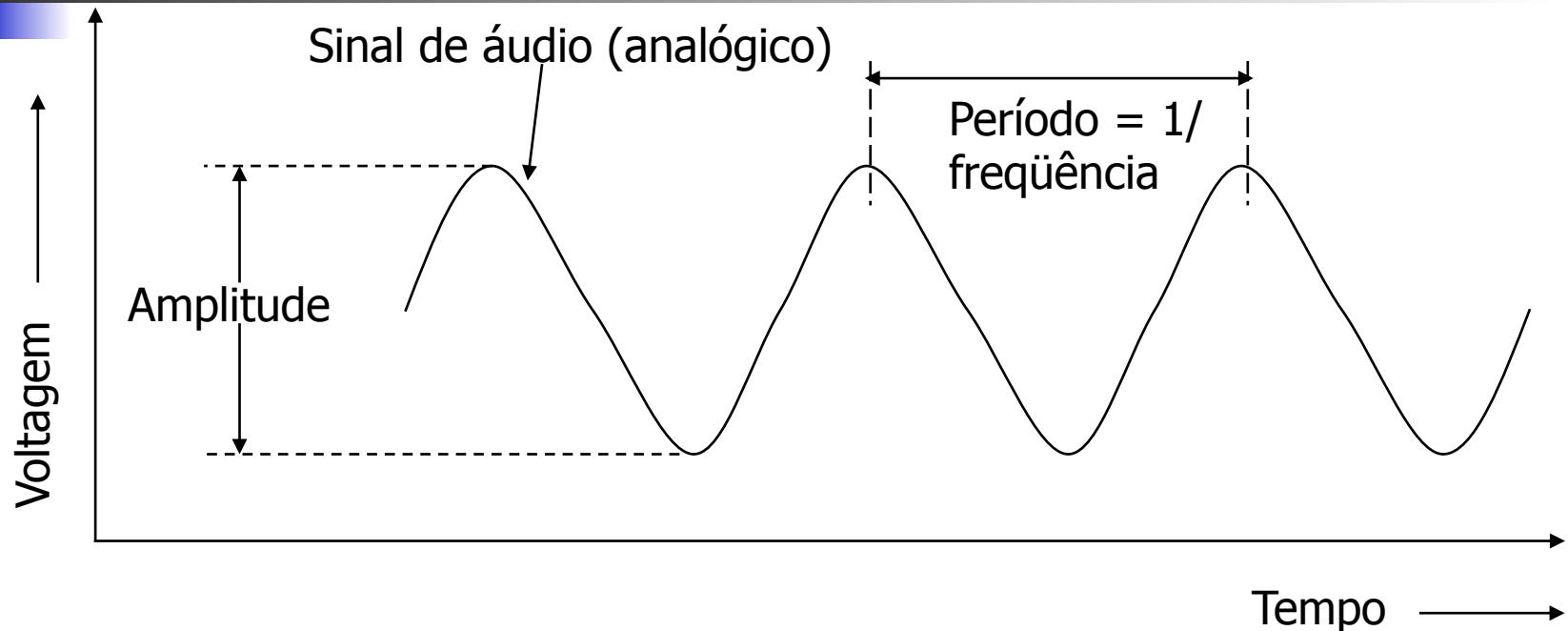


Princípios de Digitalização.

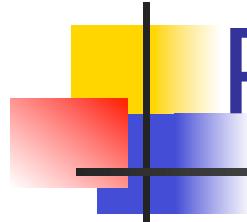
- Para poder ser utilizado em um computador, o som precisa de duas transformações:
 - **Eletrônica**: conversão de ondas mecânicas em sinais elétricos.
 - **Digital**: conversão de sinais elétricos em bits.



Princípios de Digitalização.



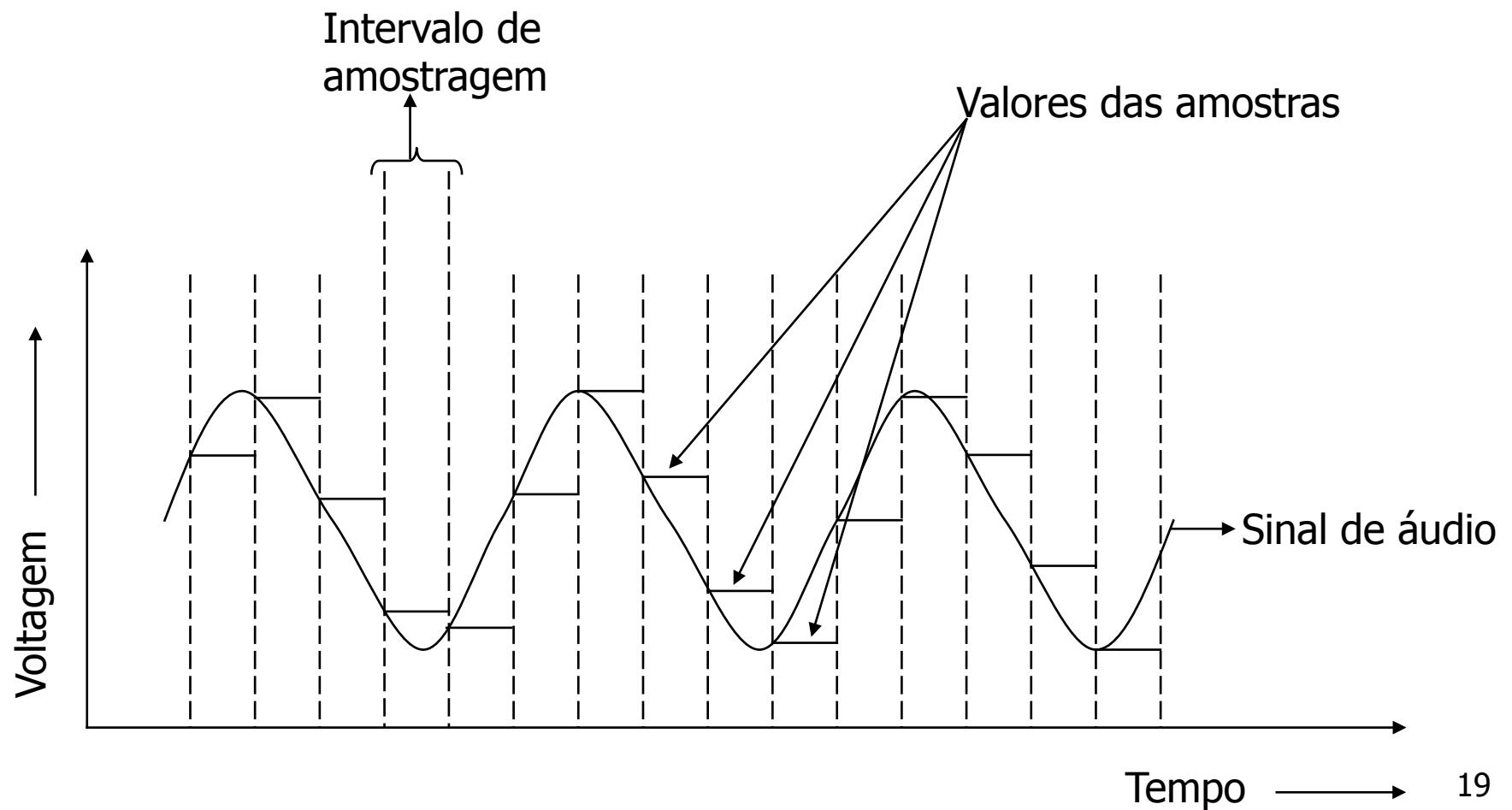
- **Freqüência:** taxa com que o sinal varia entre valores positivos e negativos. É medida em Hertz (Hz).
- **Amplitude:** diferença entre os máximos valores positivos e negativos do sinal de áudio. Pode ser expressa observando-se a voltagem (dependente do sistema). Normalmente expressa em decibéis (dB).

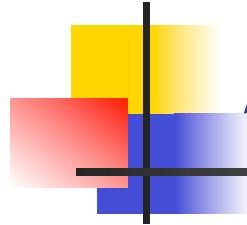


Princípios de Digitalização.

- Conversão analógico-digital.
 - Sinal de áudio possui duas dimensões: voltagem e tempo. Elas serão digitalizadas através de dois processos:
 - **Amostragem**: realiza leituras periódicas e instantâneas da voltagem em espaços de tempo uniformes.
 - **Quantização**: converte os valores analógicos amostrados em valores digitais.
 - Codificador:
 - Filtro digital + ADC (Analog to Digital Converter)

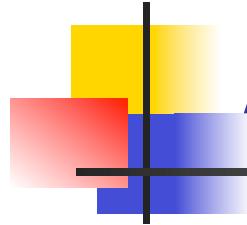
Amostragem.





Amostragem.

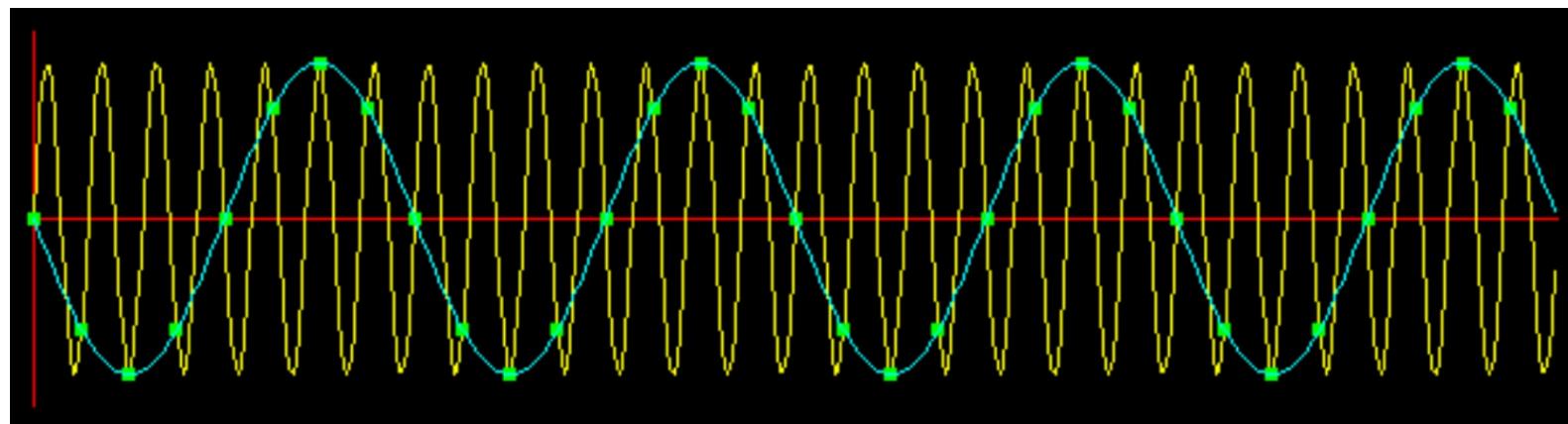
- O quanto deve ser amostrado?
 - Reconstruir exatamente o sinal → infinitas amostras.
 - Poucas amostras → sinal distorcido.



Amostragem.

- O quanto deve ser amostrado?
 - **Teorema de Nyquist:** “Para obter uma representação precisa de um sinal analógico, sua amplitude deve ser amostrada a uma **taxa** mínima igual ou superior ao dobro da componente de mais alta freqüência presente no sinal”. (taxa de Nyquist).
 - Ex. Se a freqüência mais alta do sinal é de 20KHz, para que a reconstrução seja precisa, a amostragem deve ser realizada a 40KHz, ou 40 Ksps.
 - sps = samples per second.

Amostragem.

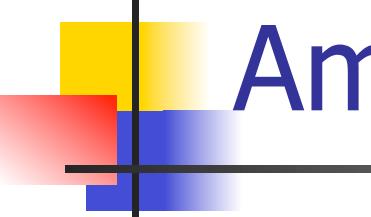


Sinal original (amarelo) → 17.500Hz

Digitalização à 20.000Hz (pontos verdes)

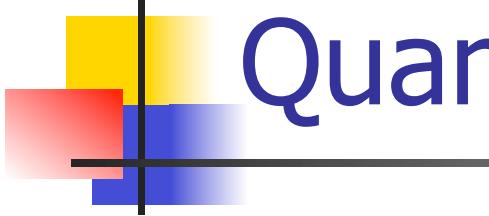
Onda resultante (azul) → 2.500Hz (**aliasing**)

$$f' = |f - amostragem|$$



Amostragem.

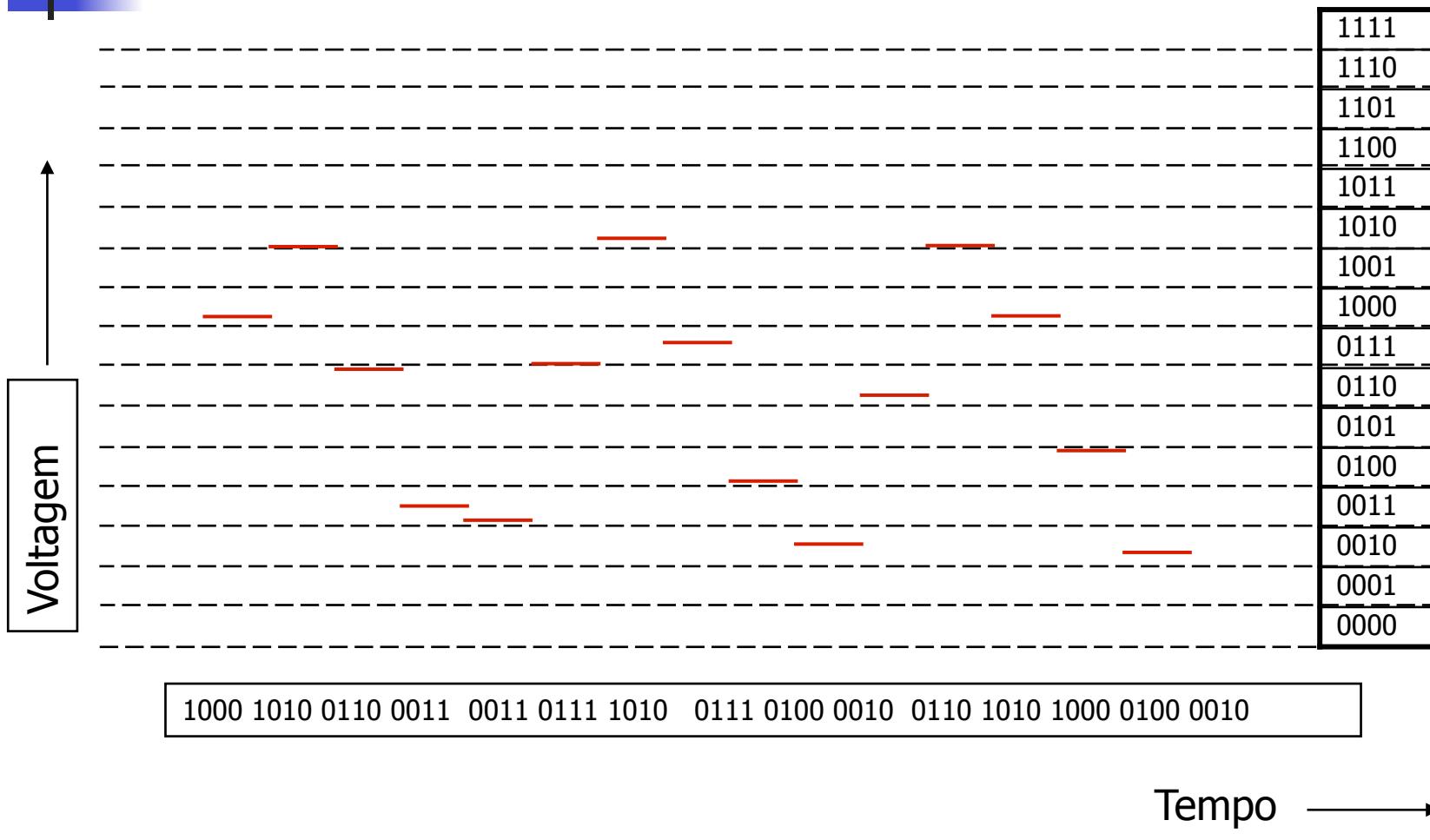
- Filtros anti-aliasing.
 - Removem componentes de frequência muito altas
 - Direcionados a:
 - Faixa audível do ouvido humano
 - Até 20 KHz → filtro remove frequências acima, sendo o sinal posteriormente amostrado a 40 KHz
 - Largura de banda do canal de comunicação
 - Canal suporta a transferência até 12 KHz → filtro remove frequências acima, sendo o sinal posteriormente amostrado a 24 KHz



Quantização.

- Processo pelo qual os valores analógicos das amostras tomadas da amplitude do sinal são convertidos em valores digitais.
- Para reconstruir exatamente o sinal:
 - Necessidade de um número infinito de bits.
- Usando um número finito de bits:
 - Representa-se cada amostra através de um número correspondente de níveis discretos.

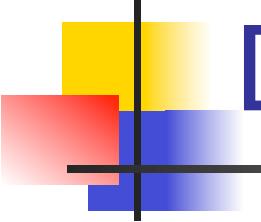
Quantização.





Quantização.

- Amostragem e Quantização
 - Número de amostras x número de níveis.
 - Compromisso.
 - Quantização resulta em distorções.

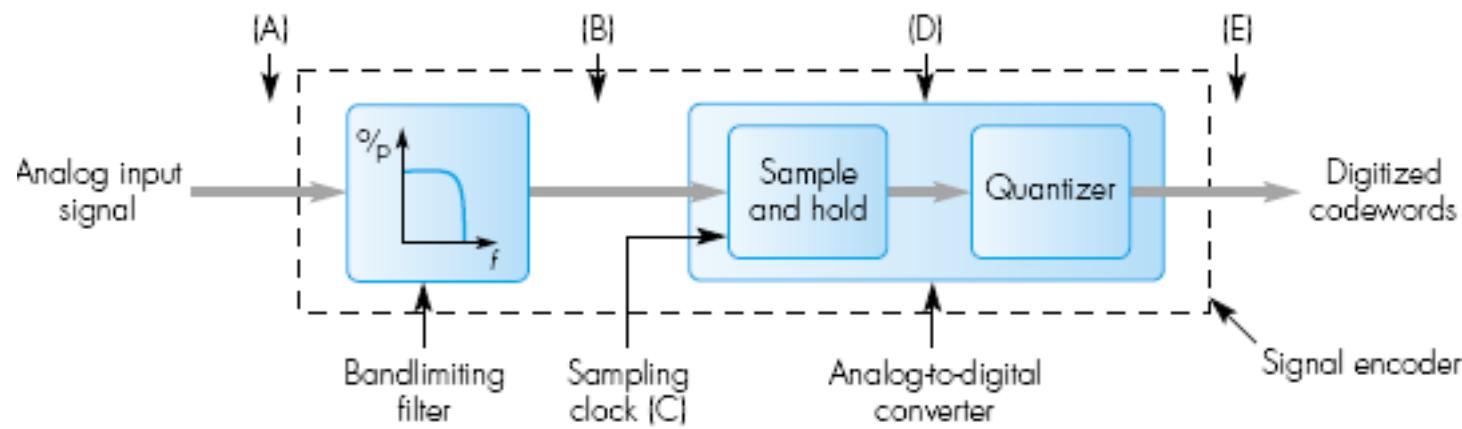


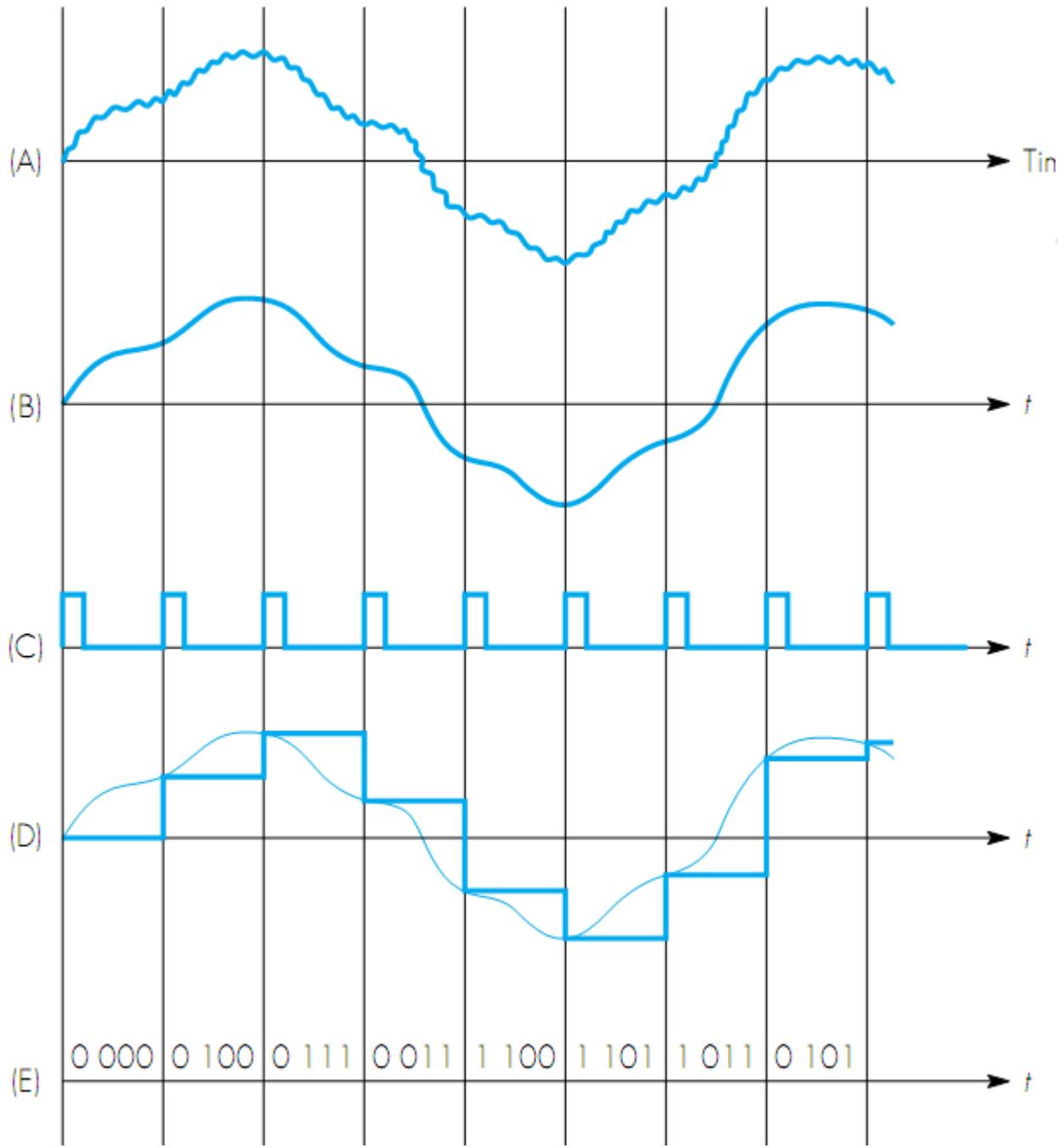
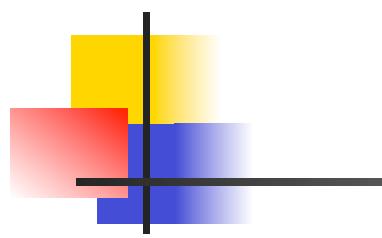
Digitalização.

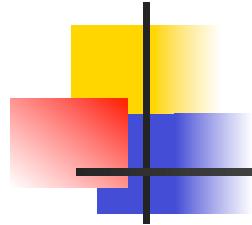
- Taxas comuns de amostragem:
 - 8.000Hz, 11.025Hz, 22.050Hz e 44.100Hz (CD).
- Números comuns de bits por amostra:
 - 4, 8, 16 e 24.
- Canais de som:
 - 1 (mono), 2 (stereo), 3, 5, 7, ...
- Qualidade de CD:
 - Amostras a 44.100Hz (44,1 KHz), 16 bits por amostra e 2 canais de som (stereo).

Digitalização.

- Circuito que realiza amostragem e quantização:
 - Conversor analógico-digital (*analog to digital converter* – ADC).
 - Caminho inverso: DAC. Usado na reprodução de áudio digital.







Digitalização.

- Após a captura
 - os dados amostrados e quantizados devem ser “guardados” em algum formato – mídia de representação.
 - WAV e MP3, por exemplo.



Digitalização.

- Aspectos quantitativos.
 - Quantos bytes serão necessários para armazenar 1 segundo de áudio, capturado com qualidade de CD?



Digitalização.

- Aspectos quantitativos.
 - Quantos bytes serão necessários para armazenar 1 segundo de áudio, capturado com qualidade de CD?
 - $1(\text{segundo}) * 44.100 \text{ (taxa de amostragem)} * 2 \text{ (16 bits por amostra)} * 2 \text{ (som estéreo)} = 176.400 \text{ bytes.}$
 - Necessidade para transmissão: 1,41Mbps!

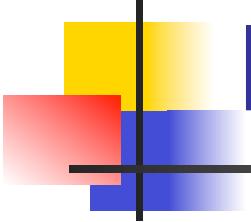


Compressão de Áudio



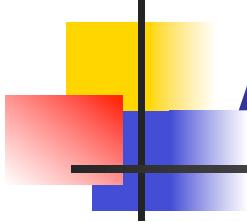
PCM Linear

- Pulse Code Modulation
 - Digitalização por meio de amostragem e quantização
 - Amostragem em intervalos regulares
 - Limitações:
 - Erro de quantização
 - Aliasing caso teorema de Nyquist não seja satisfeito



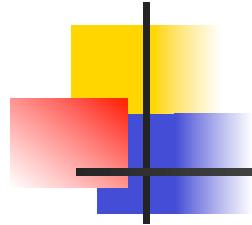
Differential PCM (DPCM)

- Amostras adjacentes de áudio são parecidas. DPCM faz previsão da amostra seguinte e codifica apenas a diferença.
 - Mudanças bruscas entre amostras adjacentes causam distorções.
- Menor quantidade de bits necessários para codificar os valores
 - Normalmente, salva-se 1 bit para cada amostra
 - Sinal de voz: redução de 64 kbps (PCM) para 56 kbps (DPCM)



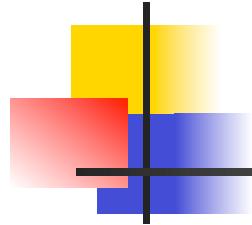
Adaptive DPCM (ADPCM)

- Utilização de número variável de bits para codificação das diferenças
 - Menos bits → menores diferenças
 - Mais bits → maiores diferenças
- ITU-T Recommendations:
 - G.721, G.722, G.726



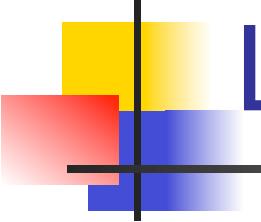
PCM Logarítmico

- Explora quantização não-linear.
 - Amplitudes maiores → maiores intervalos de quantização.
 - Amplitudes menores implicam em maior percepção de ruído de amostragem.
- Desempenho:
 - Utilizando 8 bits equivale à quantização linear com 12 bits.



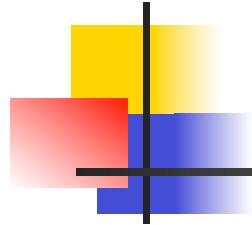
Lossless vs. lossy

- Remoção de redundâncias
 - Estatística
 - Temporal
 - Psico-acústica



Lossless vs. lossy

- Lossless coding
 - Capaz de reconstruir perfeitamente as amostras originais
 - Razão de compressão de aprox. 2:1
 - Elimina apenas redundâncias estatística e temporal
- Lossy coding
 - Não é capaz de reconstruir perfeitamente as amostras originais
 - Razão de compressão de aprox. 10:1 ou 20:1
 - Baseado no modelo psico-acústico

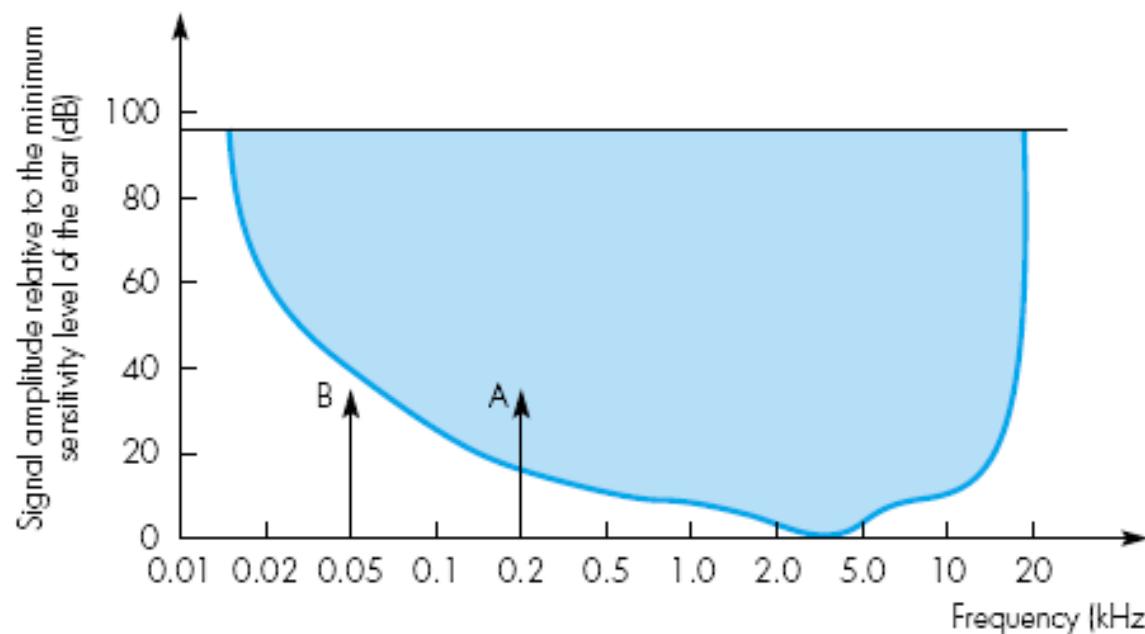


Perceptual Coding

- Modelo psico-acústico
 - Sensibilidade do ouvido.
 - Mascaramento de freqüência.
 - Mascaramento temporal.

Perceptual Coding – Modelo psico-acústico

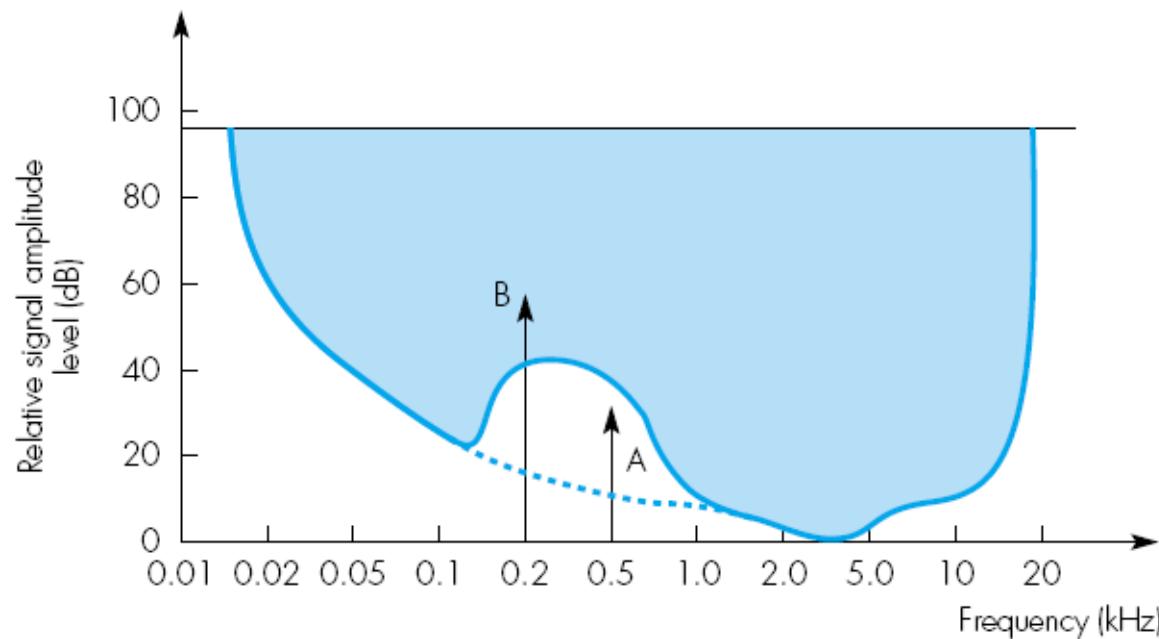
- Sensibilidade da audição



= Hearing sensitivity of the human ear

Perceptual Coding – Modelo psico-acústico

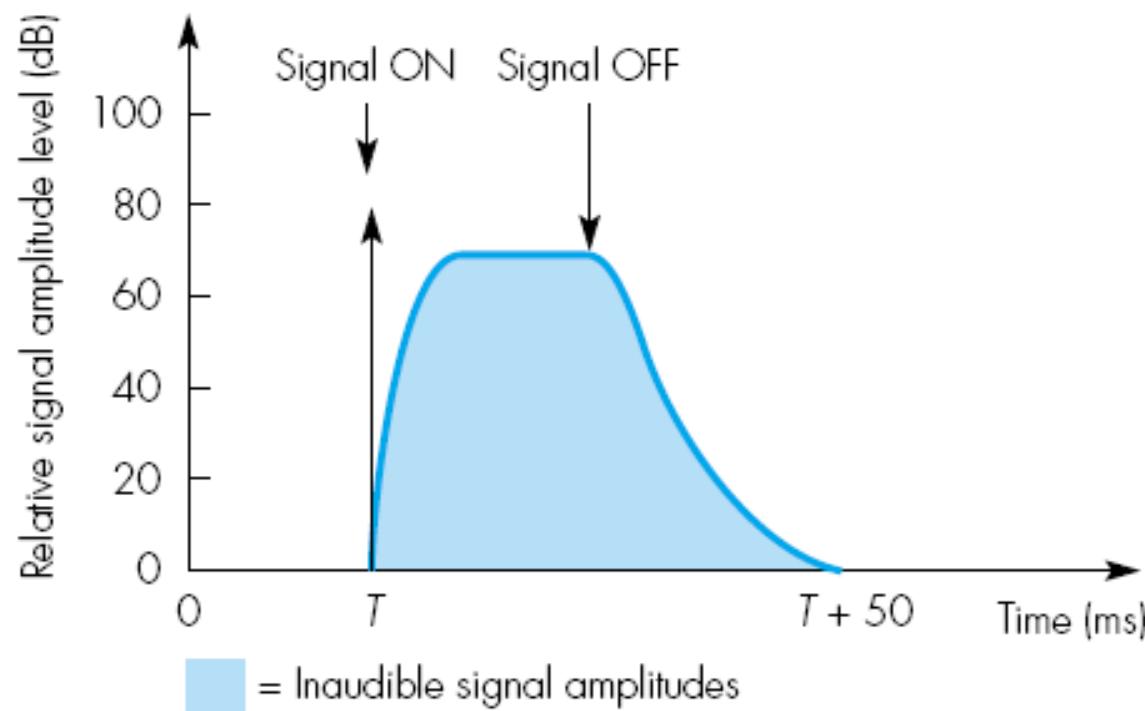
- Mascaramento de frequência

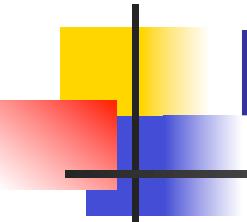


= Hearing sensitivity of the human ear

Perceptual Coding – Modelo psico-acústico

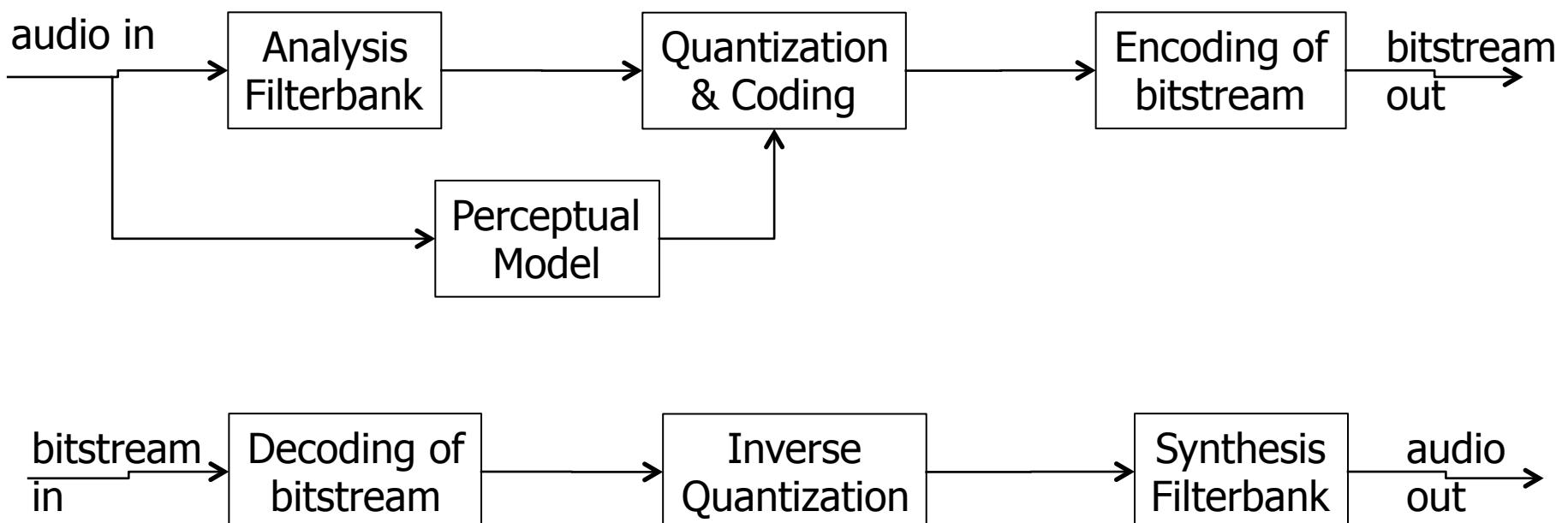
- Mascaramento temporal

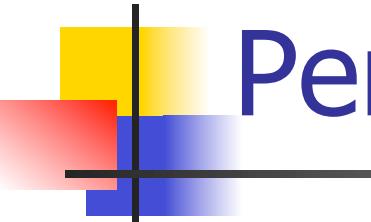




Perceptual Coding

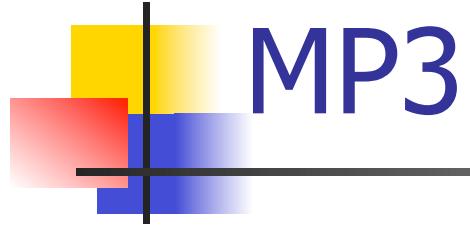
- Diagrama de blocos do encoder/decoder:





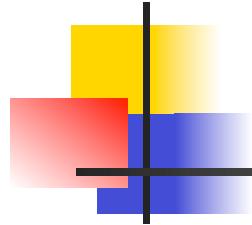
Perceptual Coding

- Filter bank
 - Usado para decompor o sinal de entrada em sub-bandas ou componentes espectrais (domínio de frequências)
- Perceptual model
 - Analisa o sinal de entrada e computa limiares de mascaramento dependentes do sinal
- Quantization and coding
 - Componentes espectrais são quantizados e codificados
 - Objetivo é manter ruído de quantização abaixo do limiar de mascaramento
- Empacotamento
 - Construção do fluxo de bits, consistindo basicamente dos dados codificados e metadados



MP3

- MPEG-1 Audio Layer 3.
 - Utiliza um modelo psico-acústico complexo.
 - Sensibilidade do ouvido.
 - Mascaramento temporal.
 - Mascaramento de freqüência.
 - Elimina as freqüências que o ouvido humano não consegue captar.
 - Compressão com perda. As perdas não são perceptíveis.
 - Som de alta qualidade e alta razão de compressão (aprox. 12:1)



Para Saber Mais

- Luther, A. C. *Using Digital Video*. AP Professional, 1995.
- Halsall, F. *Multimedia Communications: Applications, Networks, Protocols, and Standards*, Addison-Wesley Publishing, 2001. ISBN: 0201398184. Capítulos 2 e 4.