

Áudio

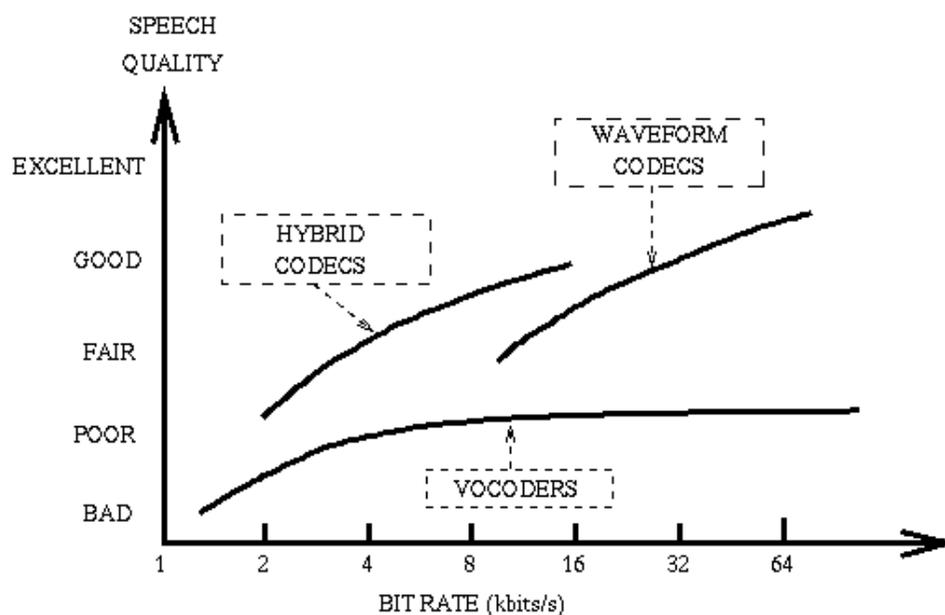
Profa. Débora Christina Muchaluat Saade
debora@midia.com.uff.br

Áudio

- ✓ **Sinal de Voz**
 - *50Hz a 10KHz*
- ✓ **Música**
 - *15Hz a 20KHz*
 - *Estéreo: 2 canais*
- ✓ **Codificadores de Sinal de Voz**
- ✓ **Codificadores de Áudio Genérico**

Codificação de Voz

- ✓ Os codificadores de voz podem ser classificados da seguinte forma:
 - *Baseados na forma do sinal (waveform codecs)*
 - *Baseados na fonte do sinal (source or voice codecs)*
 - *Híbridos (hybrid codecs)*

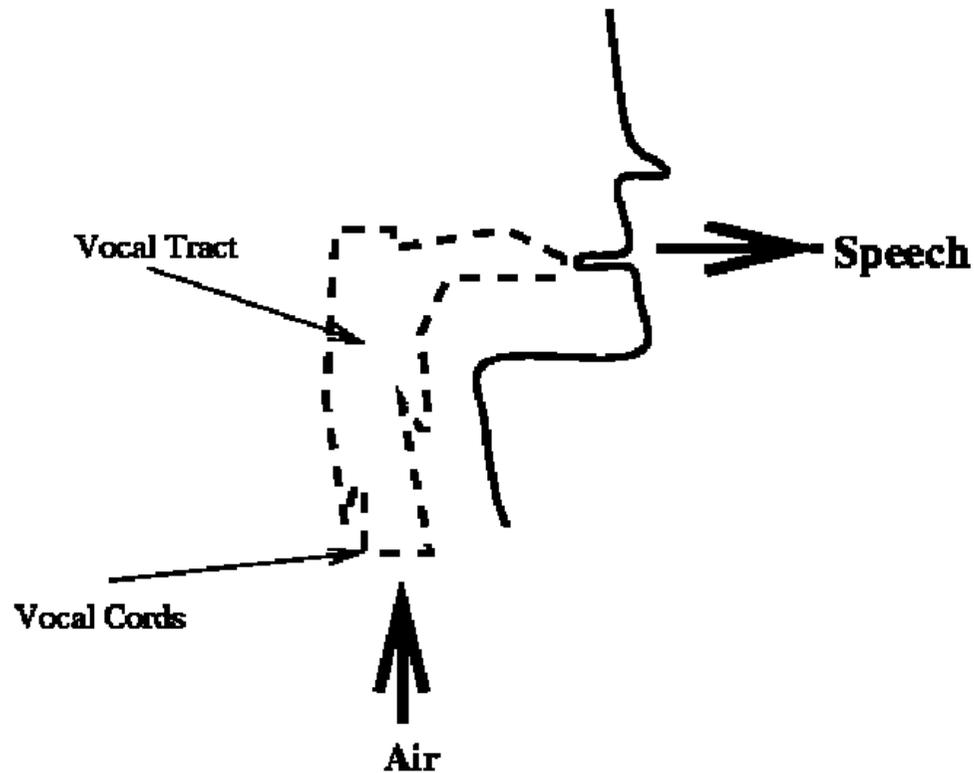


Codificação de Voz

- ✓ **Codificadores baseados na fonte do sinal - *source codecs, voice codecs, vocoders***
 - *O sinal é assumido como sendo unicamente voz e não qualquer forma de onda possível*
 - *Codificam apenas o suficiente para inteligibilidade e identificação do interlocutor*
 - *Codificadores de fonte para voz tentam replicar o processo físico da criação do som vocálico*

Codificadores Baseados na Fonte

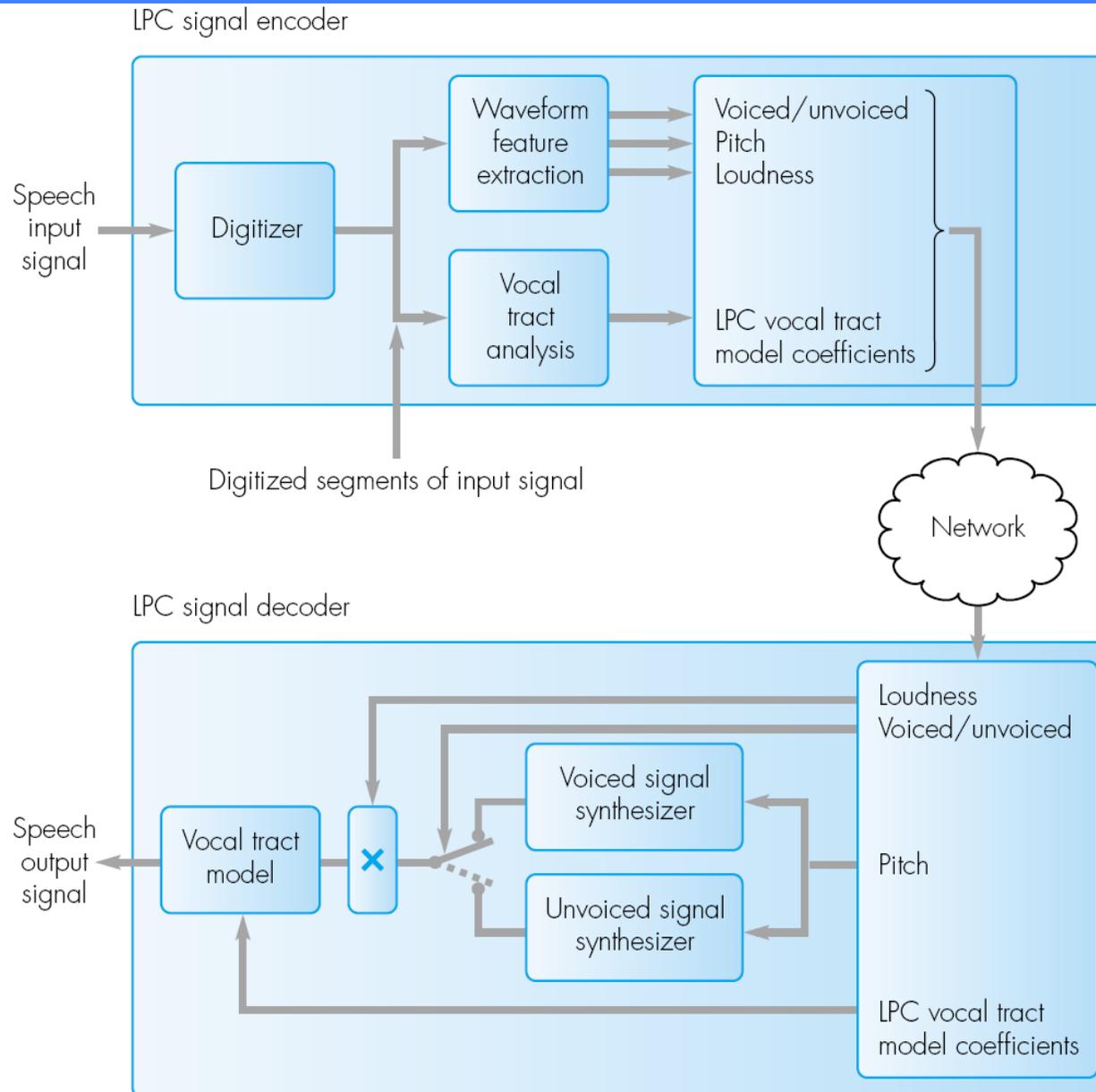
- ✓ Baseados no modelo do trato vocal humano



LPC

- ✓ **Linear Predictive Coding => usado para sinal de voz**
- ✓ **Características do sinal são identificadas na codificação e usadas na decodificação, junto a um sintetizador de voz, para gerar o áudio final**
- ✓ **Principais características do sinal de voz:**
 - ***Parâmetros perceptuais:***
 - Pitch (relacionada a frequência de vibração das cordas vocais)
 - Período: duração do sinal
 - Loudness: altura da voz (quantidade de ar dos pulmões)
 - ***Parâmetros de excitação do trato vocal (modelo do trato vocal):***
 - Usados para identificar a origem do som
 - *Voiced sounds: sons gerados através das cordas vocais, relacionados às letras m, v e l*
 - *Unvoiced sounds: as cordas vocais ficam abertas com esses sons, relacionados às f e s*

LPC



LPC

- ✓ Som gerado é de voz metálica
- ✓ Taxas de 2.4 e 1.2 Kbps
- ✓ Codificadores LPC são usados em aplicações militares
 - *Banda limitada*
- ✓ Tecnologia desenvolvida durante a 2a. guerra mundial

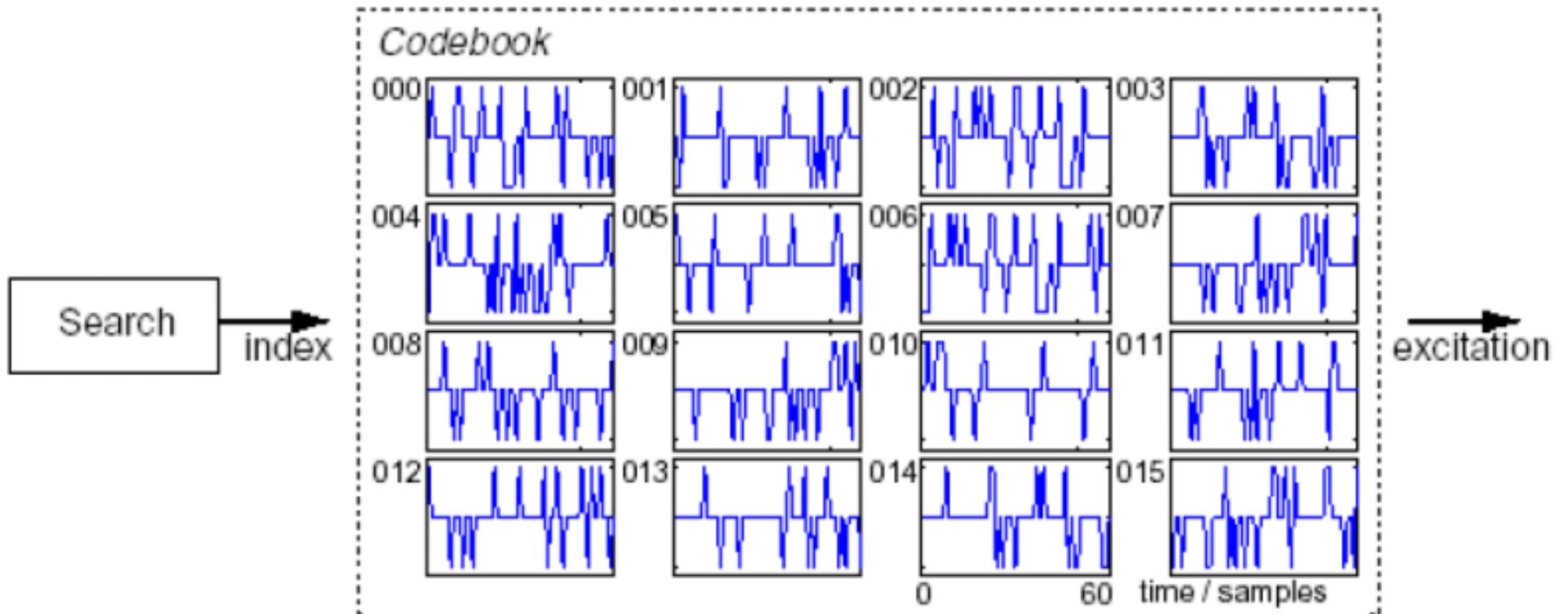
Codificação de Voz

- ✓ **Codificadores híbridos (*hybrid codecs*)**
 - *Utilizam uma combinação de análise da forma do sinal e modelagem da fonte*
- ✓ **Utilizam um modelo do trato vocal e também codificam o sinal de erro do processo de síntese em relação ao sinal original**
- ✓ **Três estratégias principais:**
 - *Multi-Pulse Excitation (MPE)*
 - *Regular Pulse Excitation (RPE)*
 - *Code-Excited Linear Prediction (CELP)*
- ✓ **Diferem na forma como o sinal de erro é codificado**

CELP

- ✓ **Code-excited Linear Prediction**
- ✓ **Gera os mesmos parâmetros LPC e ainda computa os erros entre a fala original e a fala gerada pelo modelo sintético**
- ✓ **Tanto os parâmetros LPC do modelo do trato vocal quanto uma representação comprimida dos erros são codificados**
 - *Template codebook*
- ✓ **O resultado do CELP tem qualidade boa a taxas baixas**

CELP



Padrões baseados em CELP

Padrão	Taxa de bits	Retardo na codificação	Exemplo de aplicação
G.728	16 Kbps	0.625 ms	Telefonia a taxa baixa
G.729	8 Kbps	25 ms	Telefonia em redes celulares
G.729(A)	8 Kbps	25 ms	Voz e dados digitais simultâneos
G.723.1	5.3/6.3 Kbps	67.5 ms	Videotelefonia e telefonia na Internet

Obs.: Retardo na codificação PCM padrão (taxa de amostragem a 8 KHz) => 0.125 ms

Codificação Perceptual

- ✓ **Áudio de propósito geral (voz ou música)**
- ✓ **Utiliza um Modelo Psicoacústico**
 - *Explora limitações do ouvido humano*
- ✓ **Áudio de origem é analisado e somente características perceptíveis pelo ouvido humano são codificadas/transmitidas**
- ✓ **Compressão perceptualmente sem perdas**

Codificação Perceptual

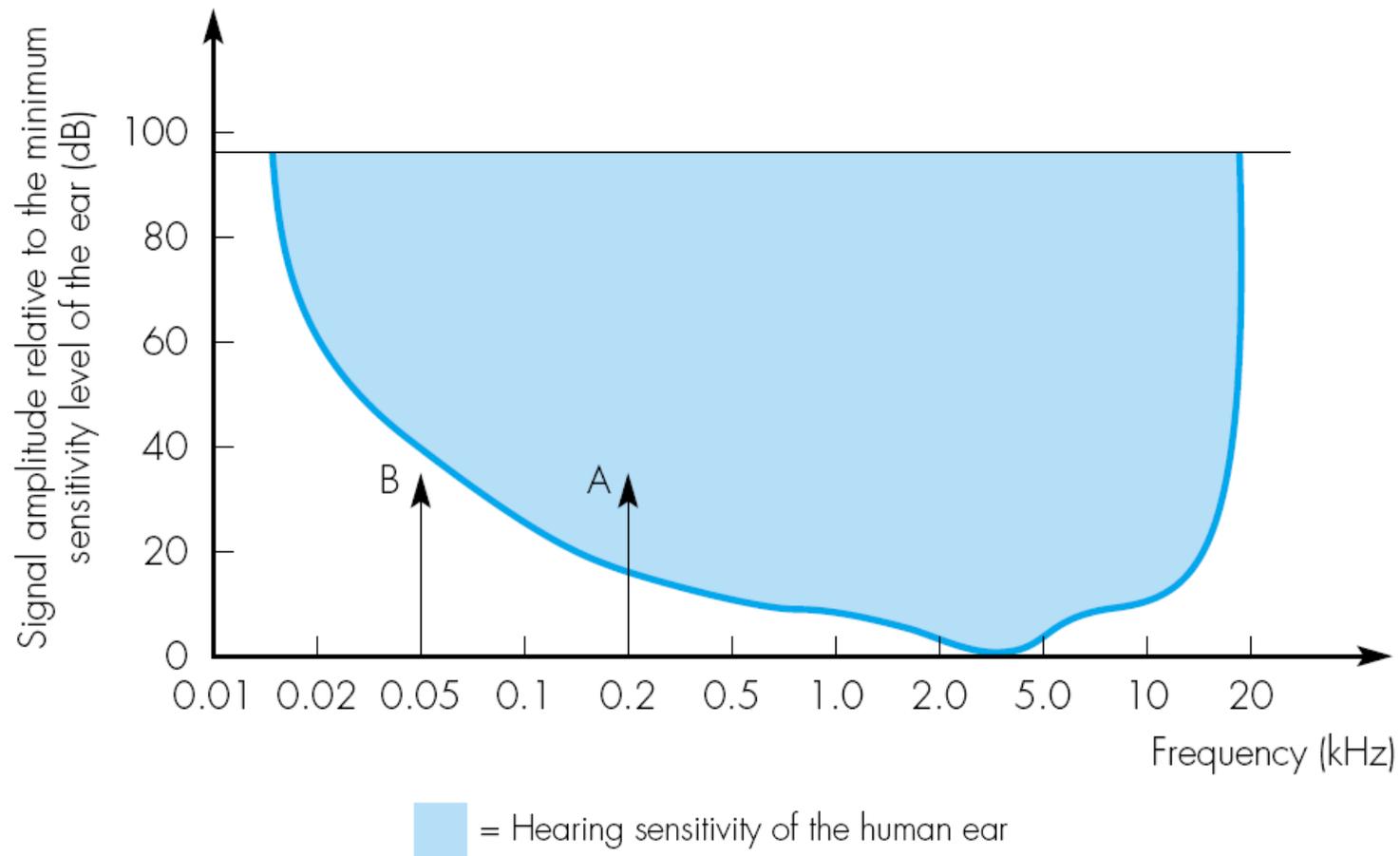
- ✓ **Ouvido humano é sensível a frequências na faixa de [15Hz, 20kHz]**
- ✓ **Nível de sensibilidade é não-linear**
 - *Ouvido é mais sensível a algumas sub-faixas que outras*
- ✓ **Quando vários sinais estão presentes:**
 - *um sinal mais forte pode reduzir a sensibilidade do ouvido a outros sinais próximos na frequência*
 - Mascaramento de frequências
 - *Quando o ouvido escuta um som alto, ele leva um tempo curto para conseguir escutar um som mais baixo*
 - Mascaramento temporal

Sensibilidade do Ouvido

- ✓ **Domínio dinâmico de um sinal (dynamic range)**
 - *Razão entre a amplitude máxima e a mínima medida em decibéis (dB)*
 - $D = 20 \log_{10} (V_{\max}/V_{\min}) \text{ dB}$
- ✓ **Domínio dinâmico do ouvido (som mais alto e mais baixo) = 96 dB**
- ✓ **Sensibilidade do ouvido varia com a frequência do sinal**
 - *Faixa de maior sensibilidade => [2, 5] KHz*

Sensibilidade do Ouvido

- ✓ **Sinal A seria ouvido, sinal B não seria**

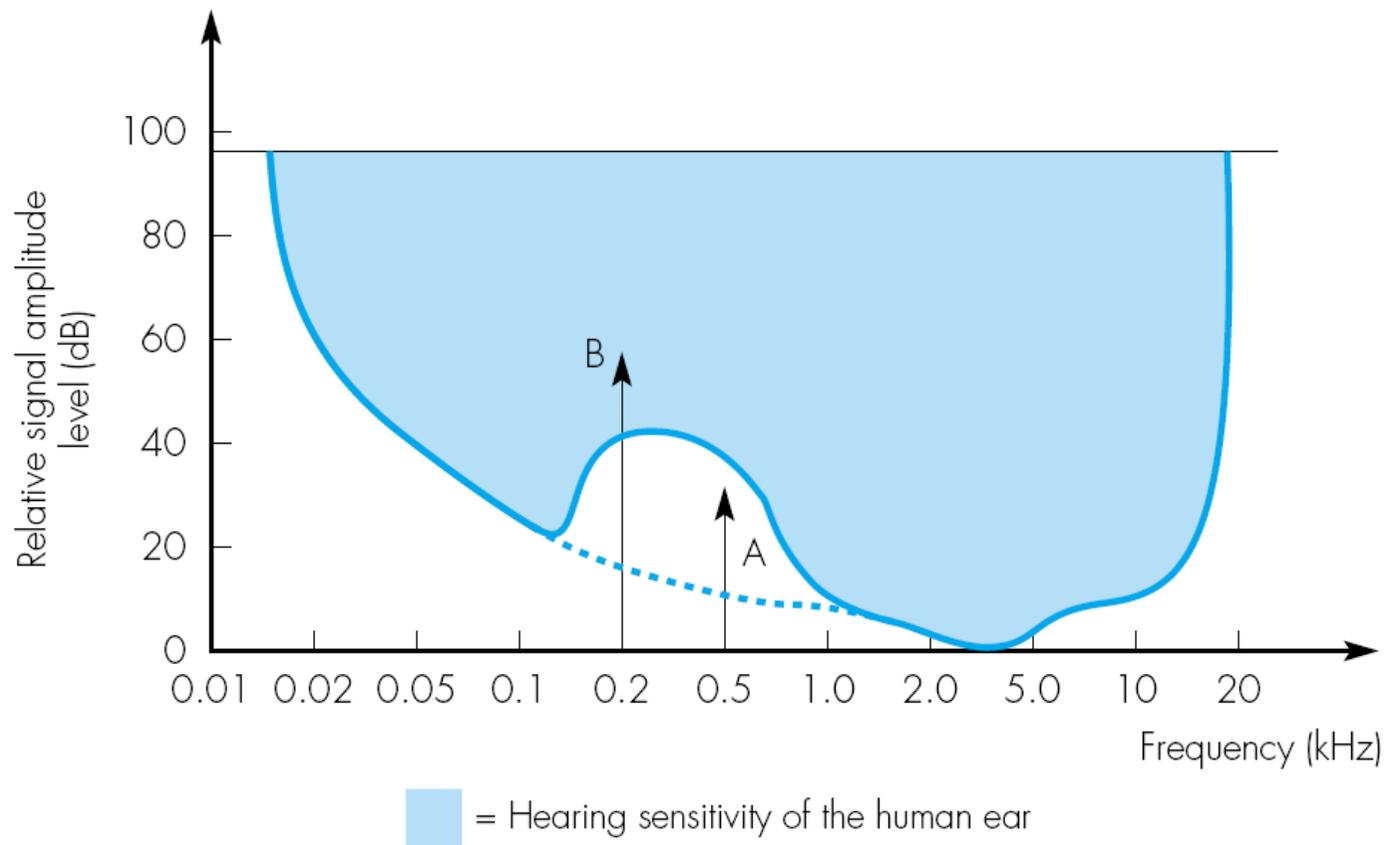


Mascaramento de Frequências

- ✓ Para um sinal de áudio que consiste de sinais com múltiplas frequências, a sensibilidade do ouvido humano varia com a amplitude relativa dos sinais
- ✓ Exemplo:
 - *Sinal B tem amplitude maior que A, o que causa uma distorção na curva de sensibilidade*
 - *Sinal A seria ouvido sozinho, mas próximo do sinal B, não é ouvido*

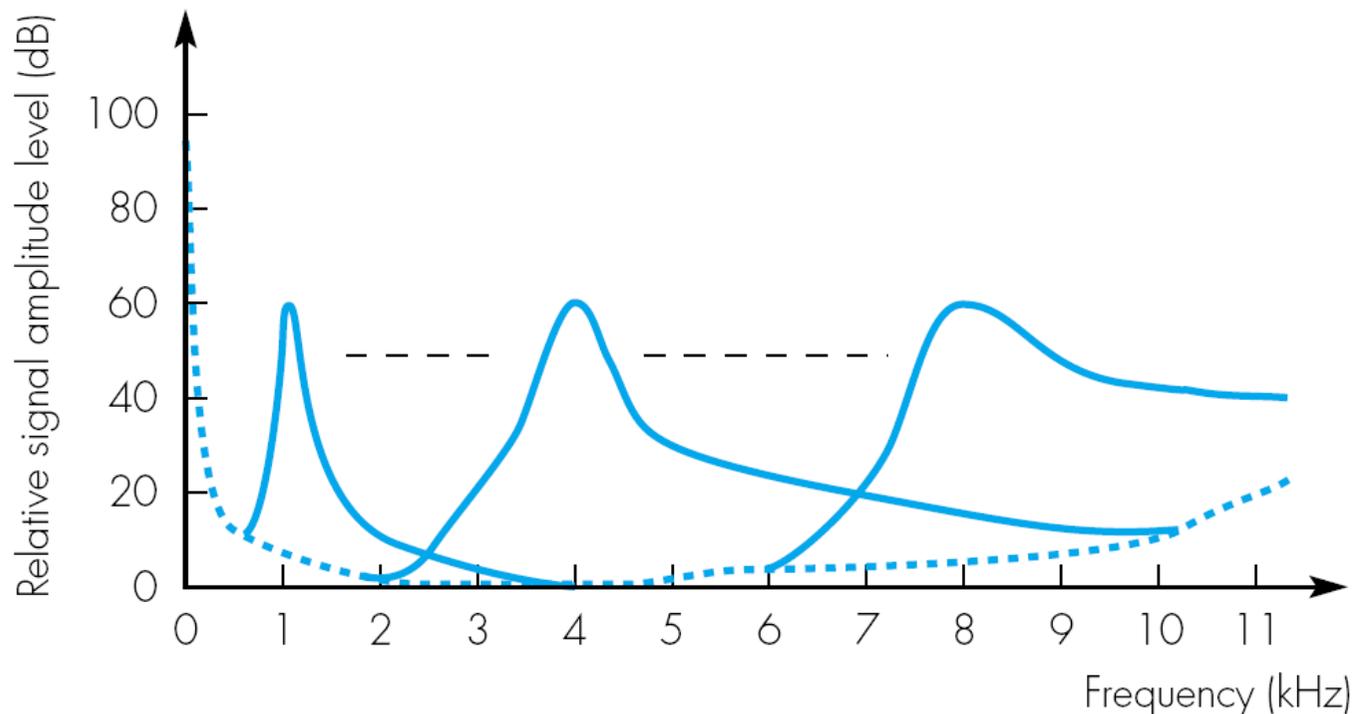
Mascaramento de Frequências

**Sinal A seria ouvido sozinho,
mas próximo do sinal B, não é ouvido**



Mascaramento de Frequências

- ✓ **Efeito do mascaramento de frequências variando com a frequência (sinais de 1, 4 e 8 KHz)**
- ✓ **O intervalo de frequências afetadas (largura da curva) é proporcional à frequência**



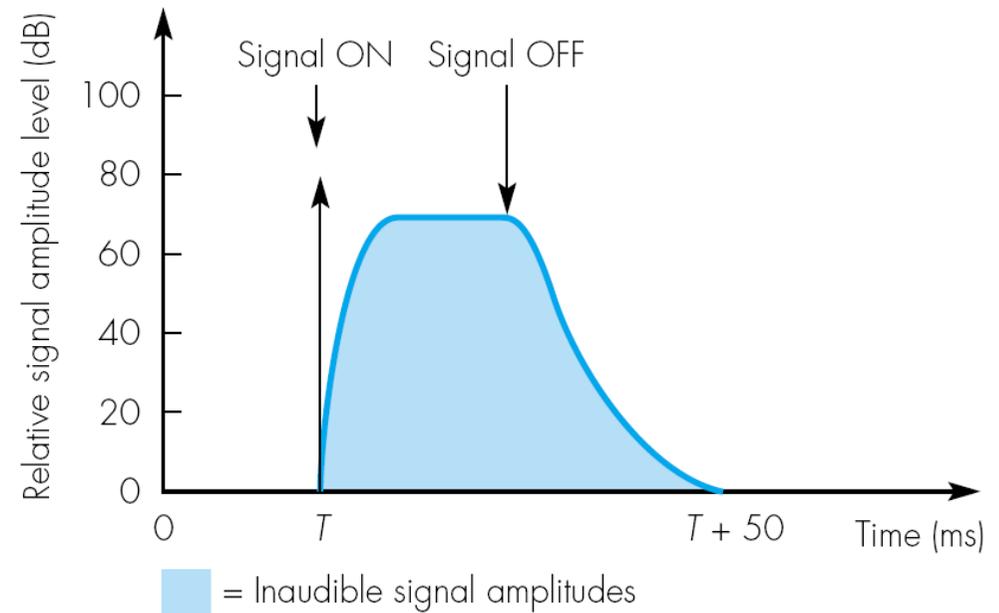
Mascaramento de Frequências

- ✓ **A largura da curva é chamada de banda passante crítica daquela frequência**
 - **$F < 500\text{Hz}$**
 - Banda passante crítica é constante $\sim 100\text{Hz}$
 - **$F > 500\text{Hz}$**
 - Banda passante crítica aumenta aproximadamente linearmente em múltiplos de 100Hz
 - *sinal de 1KHz ($2 \times 500\text{ Hz}$) \Rightarrow banda crítica de 200Hz ($2 \times 100\text{Hz}$)*
 - *sinal de 5KHz ($10 \times 500\text{ Hz}$) \Rightarrow banda crítica de 1000Hz ($10 \times 100\text{Hz}$)*
- ✓ **Conclusão**
 - ***Se a magnitude dos componentes de frequências que compõem o sinal forem determinadas, é possível determinar que frequências serão mascaradas e não precisam ser codificadas/transmitidas***

Mascaramento Temporal

- ✓ Quando o ouvido escuta um som alto, ele leva um tempo curto (dezenas de ms) para conseguir escutar um som mais baixo

- *Mascaramento temporal*



- ✓ Para identificar amostras mascaradas, é necessário processar o sinal de áudio durante um período comparável ao do mascaramento