

Codificação de Sinais Multimédia

Programa da disciplina

- Introdução aos sistemas multimédia
 - Representação de Texto, imagens, vídeo, áudio e gráficos
- Codificação de fonte e entropia
- Compressão com e sem perdas
 - Taxa de compressão e noção de distorção
- Conceitos fundamentais de imagem
 - Norma JPEG
- Conceitos de video digital
 - Normas H.26x, MPEGx (video)

Bibliografia

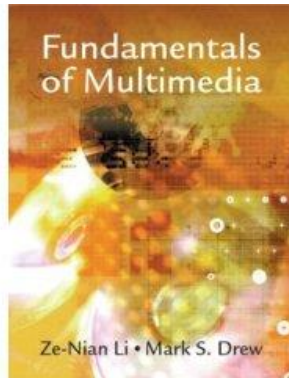


- Fernando Pereira, ``Comunicações Audiovisuais: Tecnologias, Normas e Aplicações," IST PRESS.
ISBN: 978-9728469818

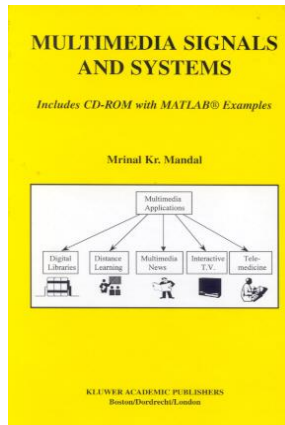


- Nuno Ribeiro e José Torres, ``Tecnologias de Compressão Multimédia“, FCA.
ISBN: 978-9727226337

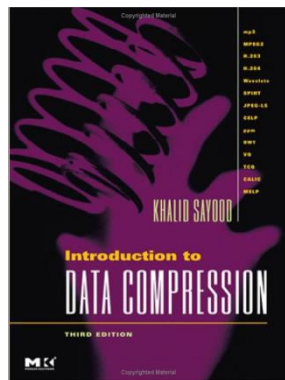
Bibliografia



- Ze-Nian Li, Mark S. Drew, ``Fundamentals of Multimedia," Prentice-Hall.
ISBN: 978-0130618726



- Mrinal Kr. Mandal, ``Multimedia Signals and Systems, " Kluwer.
ISBN: 978-1402072703



- K. Sayood, ``Introduction to Data Compression," Morgan Kaufmann.
ISBN: 978-0126208627

Alguns Marcos Históricos

- publicações em papel, jornais, etc.
- 1890 - 1ª Transmissão via rádio.
- 1900's - Realização de filmes em película.
- 1930's - Transmissão TV comercial.

- 1970 - VCR
- 1980 - PC, protocolo TCP/IP, CD.
- 1990's - WWW (HTML, XML)
- 1991 - MPEG1 - standard para video digital
- 1992 - JPEG - standard para compressão de imagem
- 1994 - Netscape
- 1995 - Java
- 1996 - DVD para comercialização de filmes
- 1998 - MP3 portáteis, XML schema
- hoje - HDTV.

Compressão e Percepção

- FINISHED FILES ARE THE RESULT OF YEARS OF SCIENTIFIC STUDY COMBINED WITH THE EXPERIENCE OF YEARS.
- Par a smana não há auas
- A falta de mao de obra quaificada nas empersas portugeusas proibe a economai de ...

Compressão e Percepção

- Imagem: Tif (786kbytes)



- Jpeg (37kbytes) -“quality” 75%)



compressão 1:20

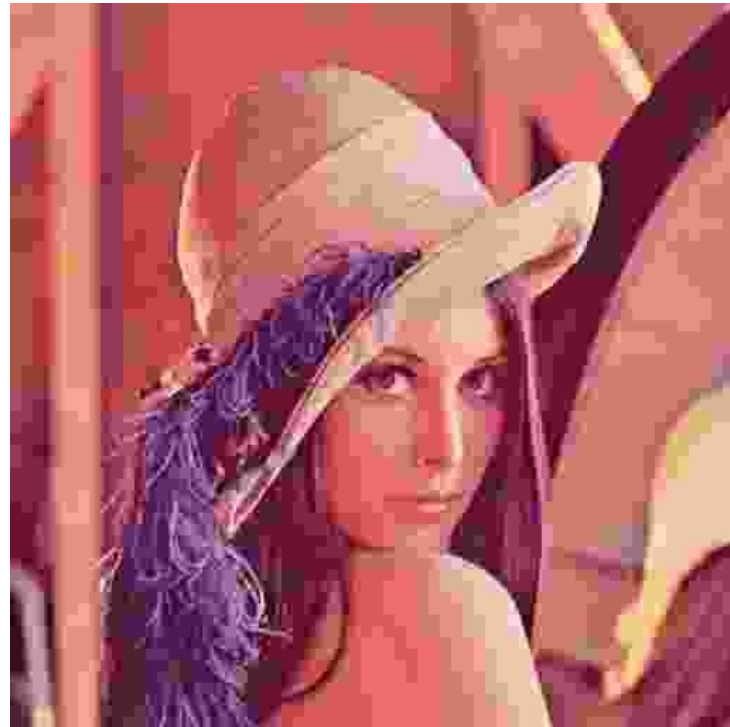
- Audio: 🗣 wav (37Mbytes)
compressão 1:11

🗣 mp3 (3.3Mbytes) – 128kbps

- Video: 1'26'' (352x288) + audio
ficheiro mpg 8,075,264 bytes
compressão ?

Compressão de sinais multimédia

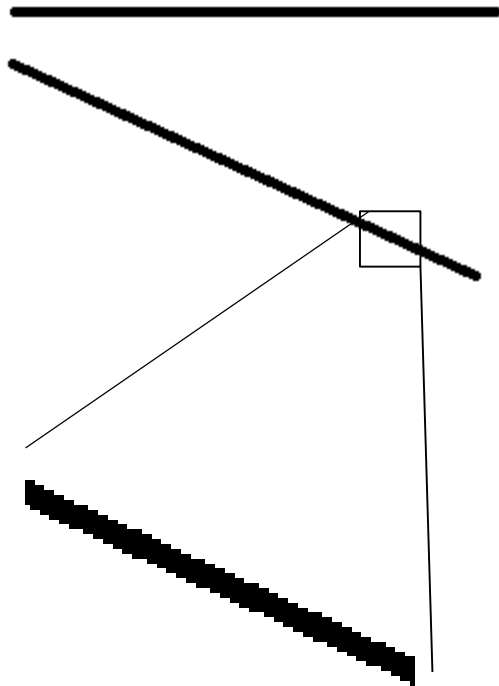
- Distorção na compressão
- BMP (512 x 512 pixels)
786 814 bytes
- JPEG (512x512 pixels)
12 838 bytes



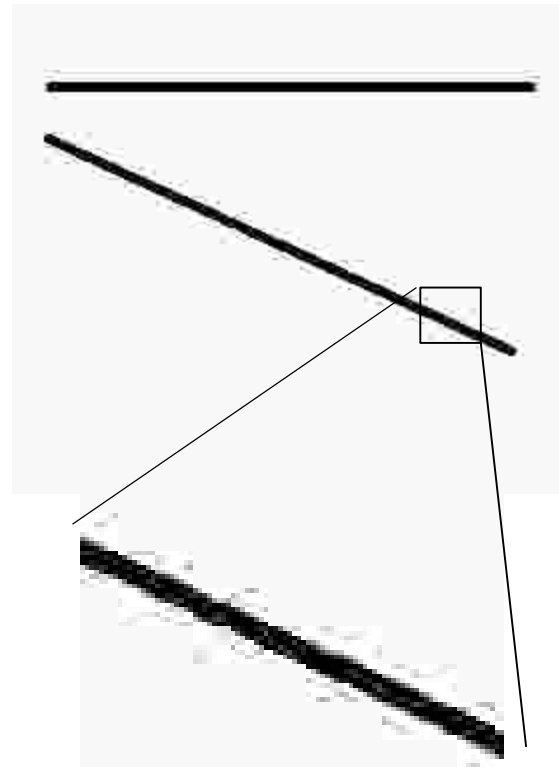
Compressão de sinais multimédia

- Distorção na compressão

- BMP (280x245 pixels)
205 854 bytes



- JPEG (280x245 pixels)
26 954 bytes



Compressão de sinais multimédia

- Compressão de sinais:
 - Audio (música, fala)
 - Imagem
 - Video
- Compressão com e sem perdas (lossy and lossless)

Baseia-se na capacidade de suprimir partes dos sinais mantendo as mesmas características perceptíveis ao homem.

A compressão baseia-se em redundâncias estatísticas, temporais, espaciais, espectrais e estruturais.

Espectro Visível

- A luz observada é função da luz incidente no objecto e da reflectividade ou tranmissividade do objecto

$$I(\lambda) = \rho(\lambda)L(\lambda)$$

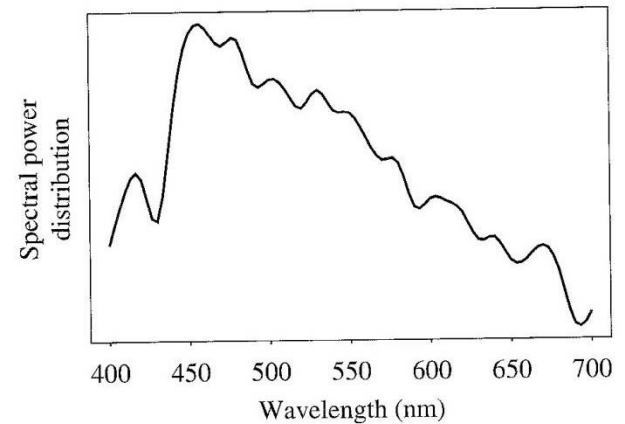
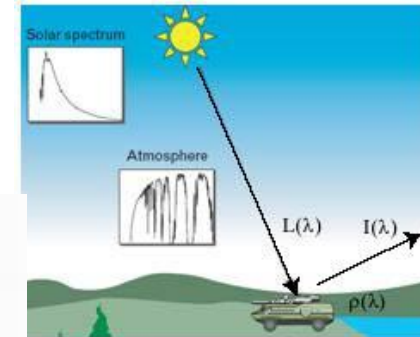
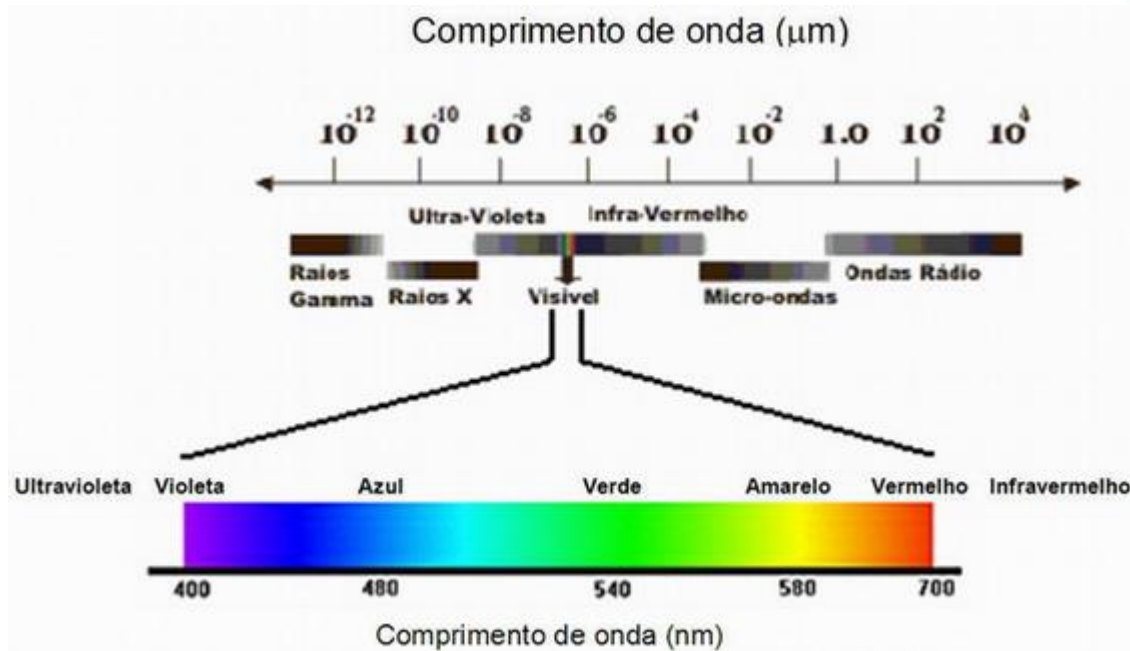
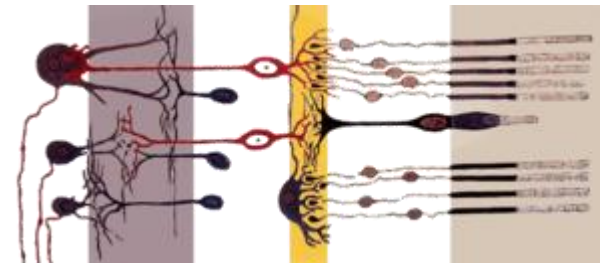
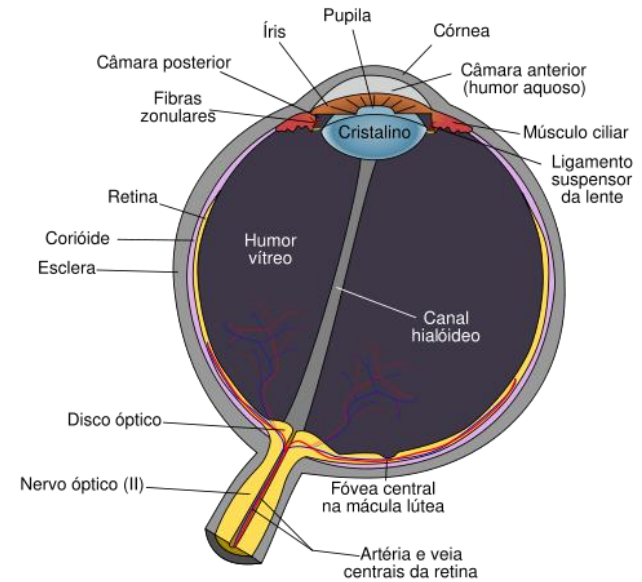


FIGURE 4.2: Spectral power distribution of daylight.

- Nota: há cores que não têm apenas um comprimento de onda (ex: côr purpura)

Sistema Visual Humano

- Aquisição de ondas electromagnéticas é responsável pela 70% da informação adquirida pelo 5 sentidos.
- Íris: diafragma que controla a abertura da pupila
- Cristalino: Funciona como lente
- cones (região fovea) sensíveis à cor (sensação cromática)
- bastonetes sensíveis à luminosidade (sensação acromática)
- Nota: À noite não vemos a cor!

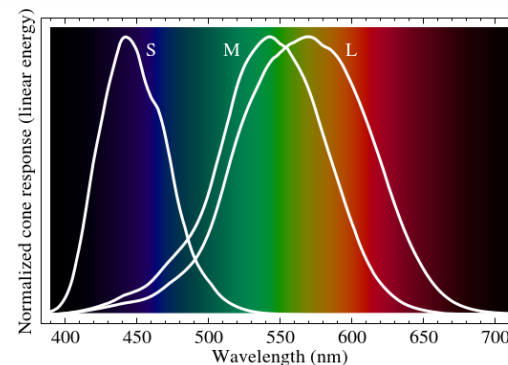
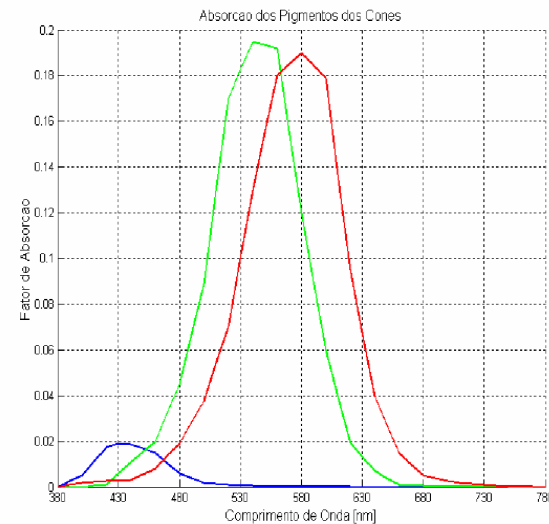
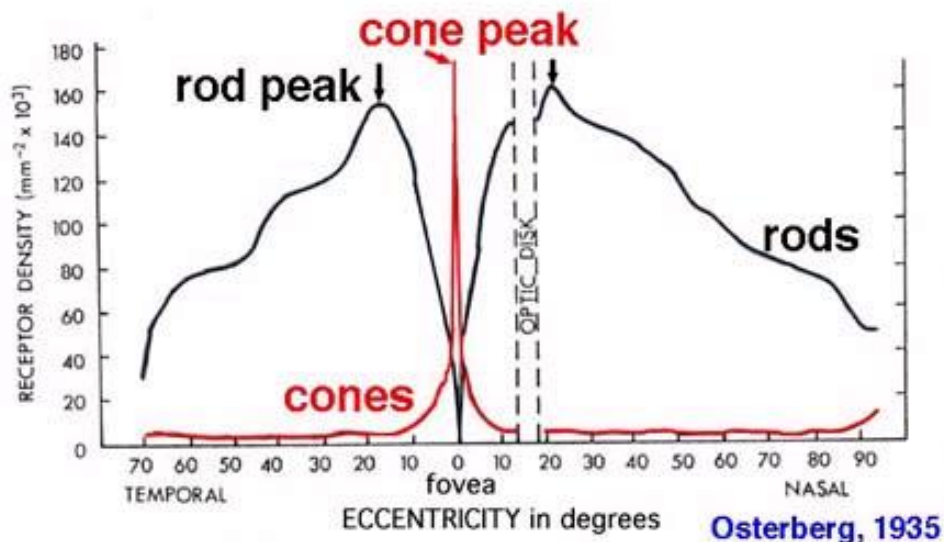


Sistema Visual Humano (percepção cromática)

- Região Fovea: 1-2° (1mm²)
- Cones abundantes só em cerca de 5° (4.5-6 milhões)
- Bastonetes em cerca de 20° (90-120 milhões)
- Bastonetes 1000 vezes mais sensíveis que cones (basta um único fóton)

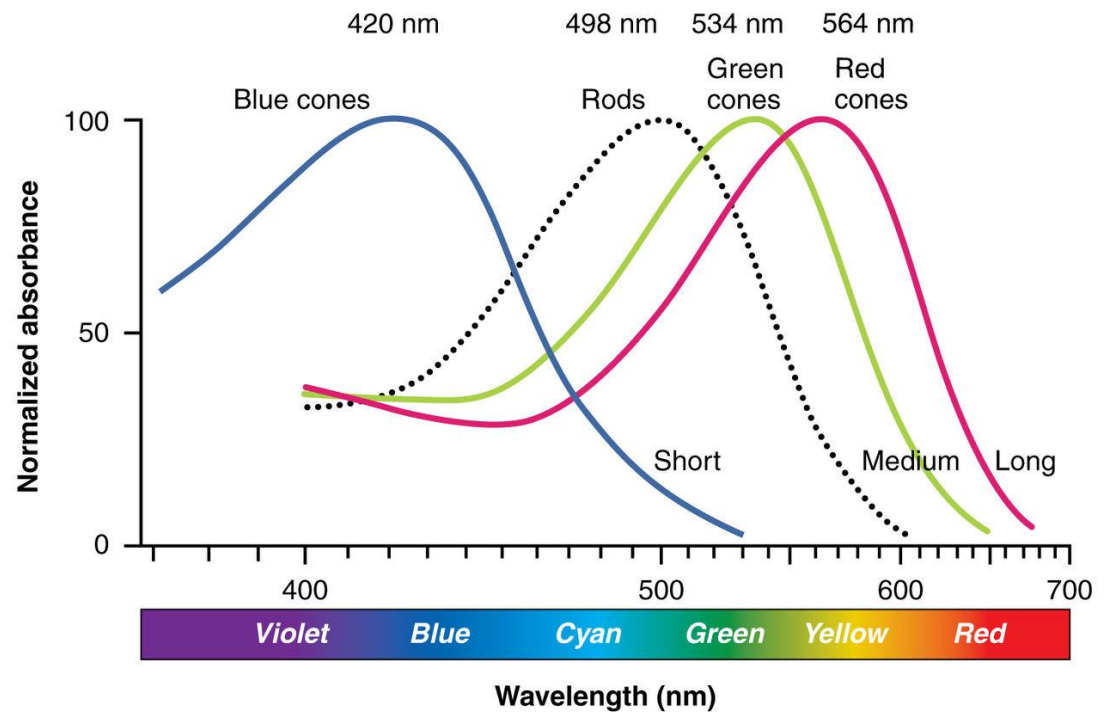
■ Cones

- verde (540nm): 33%
- vermelho (580 nm): 65%
- azul (440 nm) : 2%



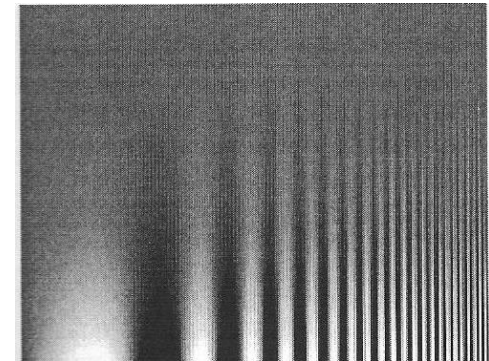
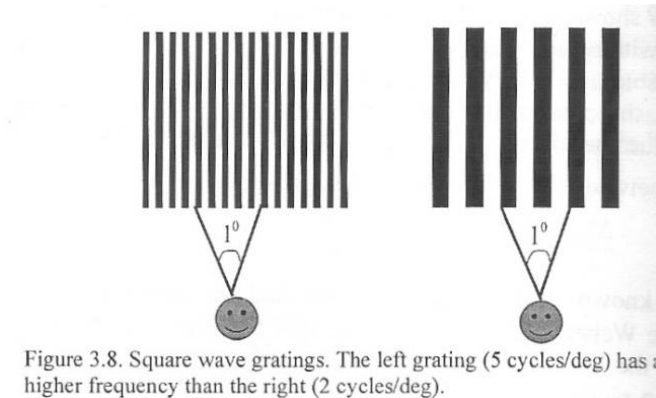
Sistema Visual Humano (percepção cromática)

- A visão humana consegue distinguir:
 - 10^6 - 10^7 cores;
 - $\gg 10^7$ diferentes brilhos



Sistema Visual Humano (percepção acromática)

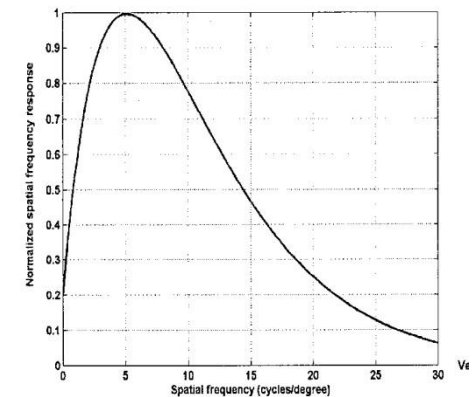
- Resposta em frequência espacial



Contrast versus spatial frequency sinusoidal grating.

- **Nota:**

- O olho tem mais acuidade entre 2 a 10 ciclos por grau.
- A sensibilidade do olho é menor em direcções diagonais.



(a)

Acuidade visual

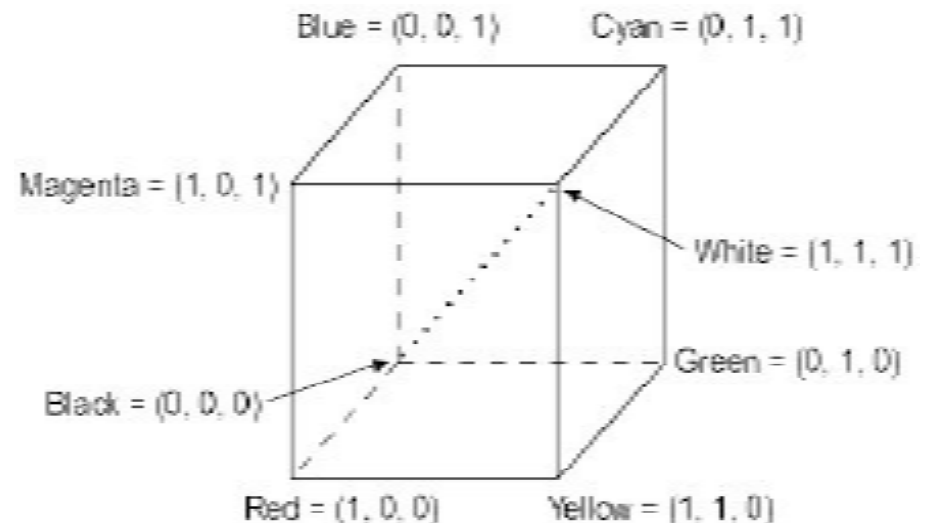
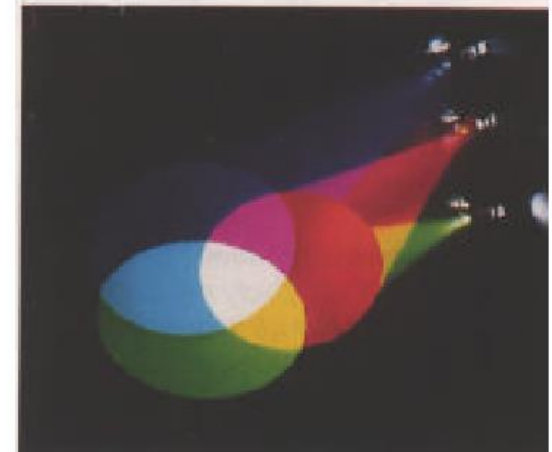
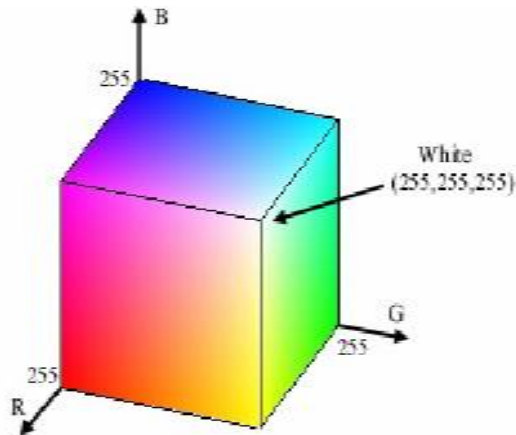
- Ângulo de visão
- Resolução espacial
 - O mais pequeno detalhe perceptível à visão humana é cerca de 1minuto = $1/60^\circ$

Propriedades temporais da visão

- Persistência da visão:
 - Várias frames amostradas a uma determinada velocidade parece um movimento continuo
 - O efeito “flicker ” é tanto maior quanto maior for a intensidade luminosa
 - Este efeito e desprezável quando a frequência fôr superior a 50/60hz.

Representação de Côr em Imagem

- Sistema RGB
 - Red Green Blue
 - Sistema aditivo na mistura de côr
 - Usado nos monitores



Representação de Côr em Imagem

- Sistema CMY
 - Cian Magenta Yellow
 - sistema subtrativo na mistura de côr
 - Conversão CMY – RGB

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$



- Sistema CMYK
 - K – preto (melhora o contraste e poupa tinta)
 - Usado nas impressoras
 - Conversão CMYK - RGB

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} C - K \\ M - K \\ Y - K \end{bmatrix}$$

$$K = \min\{C, M, Y\}$$

Representação de Côr em Imagem

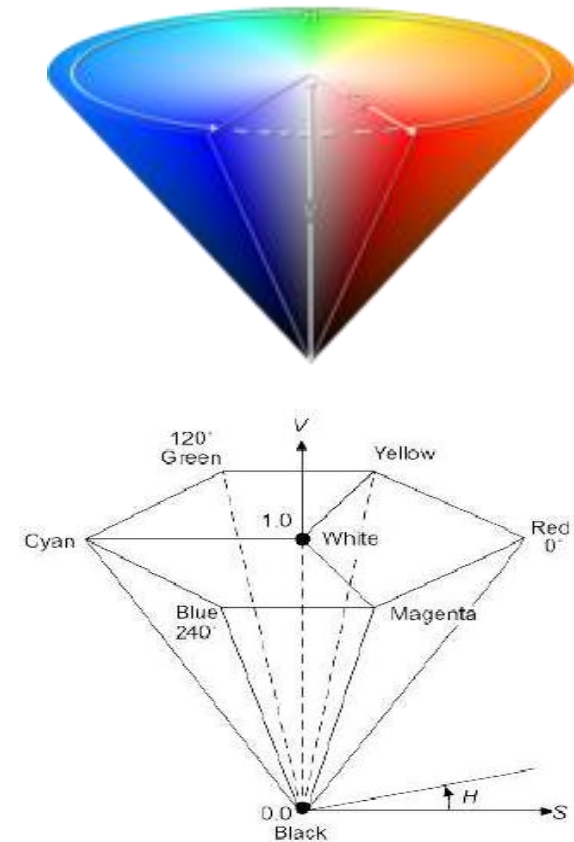
- Sistema HSV
 - Hue (cor dominante)
varia entre 0 e 360°
180° entre cores complementares
 - Saturation (puricidade)
varia entre 0 e 1 (1-saturado)
 - Value (luminância)
varia entre 0 e 1 (1-brilhante)

Branco: $S=0$ $V=1$

Cinzentos: $S=0$, $V[0\ 1]$

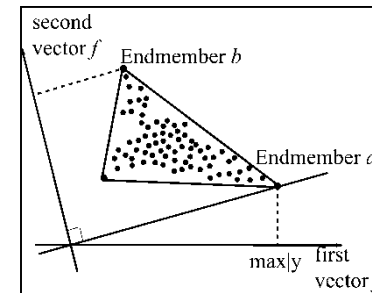
Preto: $V=0$

Cor pura: $S=1$ $V=1$



Representação de Imagens

- Tipos de ficheiros:
 - BMP, GIF, TIFF, JPEG, PS, EPS, PDF, etc
 - Há formatos que fazem compressão (com e sem perdas)
- Resolução espacial da imagem:
 - “aspect ratio” - 4:3 ou 16:9
 - (640x480) (1024x768) ou (1280x720) (1920x1080)
- Representação cor
 - Monocromática: 1bit/pixel (linhas x colunas x 1)



Representação de Imagens

- Tipos de ficheiros:
 - BMP, GIF, TIFF, JPEG, PS, EPS, PDF, etc
 - Há formatos que fazem compressão (com e sem perdas)
- Resolução espacial da imagem:
 - “aspect ratio” - 4:3 ou 16:9
 - (640x480) (1024x768) ou (1280x720) (1920x1080)
- Representação cor
 - Monocromática: 1bit/pixel
 - Níveis de cinzento: 8bit/pixel (linhas x colunas x 8)



Representação de Imagens

- Tipos de ficheiros:
 - BMP, GIF, TIFF, JPEG, PS, EPS, PDF, etc
 - Há formatos que fazem compressão (com e sem perdas)
- Resolução espacial da imagem:
 - “aspect ratio” - 4:3 ou 16:9
 - (640x480) (1024x768) ou (1280x720) (1920x1080)
- Representação côr
 - Monocromática: 1bit/pixel
 - Níveis de cinzento: 8bit/pixel
 - Full color: 24 bit/pixel (8 bit cada côr) (linhas x colunas x 8 x 3)
Suporta 256x256x256 cores!

Nota: Imagens com côr a 24 bit
Armazenadas em 32 bit
(1 byte para efeitos especiais)



Representação de Imagens

- Tipos de ficheiros:
 - BMP, GIF, TIFF, JPEG, PS, EPS, PDF, etc
 - Há formatos que fazem compressão (com e sem perdas)
- Resolução espacial da imagem:
 - “aspect ratio” - 4:3 ou 16:9
 - (640x480) (1024x768) ou (1280x720) (1920x1080)
- Representação côr
 - Monocromática: 1bit/pixel
 - Níveis de cinzento: 8bit/pixel
 - Full color: 24 bit/pixel
 - Mapeamento de côr: 8 bit/pixel ou seja 256 cores
 - Usa uma “palette” de cor

Como construir a palette?



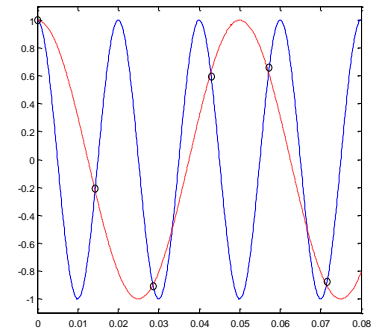
Dithering

- Compromisso entre resolução de cor e resolução espacial
- A ilusão criada deve-se ao facto que a visão humana integra a intensidade em pequenas regiões da imagem
- Melhor que quantificar com 2 níveis
- Diferentes técnicas:
 - Ordered
 - Floyd-Steinberg (error diffusion)
 - Bayer
 - Halftone
 - Random



Digitalização de sinais - PCM

- Amostragem e Quantificação
- Amostragem
 - Frequência de amostragem ($f_s > 2f_{max}$)
 - Aliasing
 - Em audio (1kHz amostado a)
 - 8kHz
 - 1.5kHz



- Em Imagens



- Em video

Medidas de qualidade da imagem

- Subjetivas

ou

- objectivas

$$SNR(dB) = 10 \log_{10} \left[\frac{\sum_l \sum_c I_{ap}(l, c)^2}{\sum_l \sum_c [I_{ap}(l, c) - I_{or}(l, c)]^2} \right]$$

$$MSE = \frac{1}{LC} \sum_l \sum_c [I_{ap}(l, c) - I_{or}(l, c)]^2$$

Modelos de Côr em Video

- Transformação YCbCr – RGB

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.168 & -0.331 & 0.5 \\ 0.5 & -0.418 & -0.08 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix}$$

- Na recomendação ITU-R BT.601-4:

- Codificação deve ser a 8 bit;
- Y varia entre 235 e 16

$$Y = 219(0.299R + 0.587G + 0.114B) + 16$$

- Cb,Cr variam entre 16 e 240

$$C_b = 112(B - Y)/0.866 + 128$$

$$C_r = 112(R - Y)/0.701 + 128$$

Video Digital

- Vantagens
 - Qualidade de imagem (resol. espacial, frame rate, aspect ratio)
 - Qualidade do som (4-6 canais)
 - Processado num PC
 - Encriptação
 - Melhor SNR
 - Fácil armazenamento
 - Fácil acesso para edição
- Desvantagem
 - Requer uma largura de banda elevada
 - Requer muito espaço de armazenamento
- Compromisso
 - **Compressão**
- Exemplo imagem 1920*1080
 - Frame rate 30Hz
 - 24bit por pixel
 - 1492Mbit/seg = **186Mbytes/seg**

Video Digital

- Dado que o video é uma sequência de imagens
 - Pode-se comprimir cada imagem (intraframe)
e
 - Pode-se tirar partido da redundância temporal
e do sistema visual que é menos sensível a objectos em movimento
(compressão interframe)

- Dado que o sistema da visão humana é menos sensível à cor que a intensidade a quantificação de cor pode ser mais pobre.

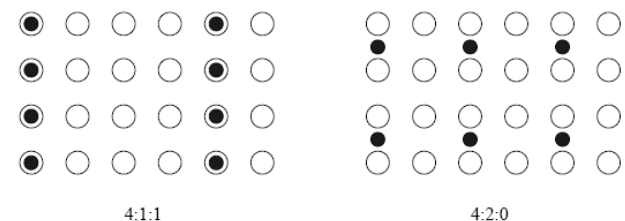
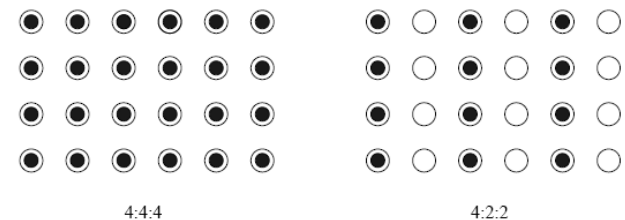
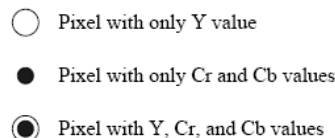
- Chroma subsampling

4:4:4 - sem subamostragem

4:2:2 - amost. hor. Cb Cr factor 2

4:1:1 - amost. hor. Cb Cr factor 4

4:2:0 - amost. hor. E vert. factor 2

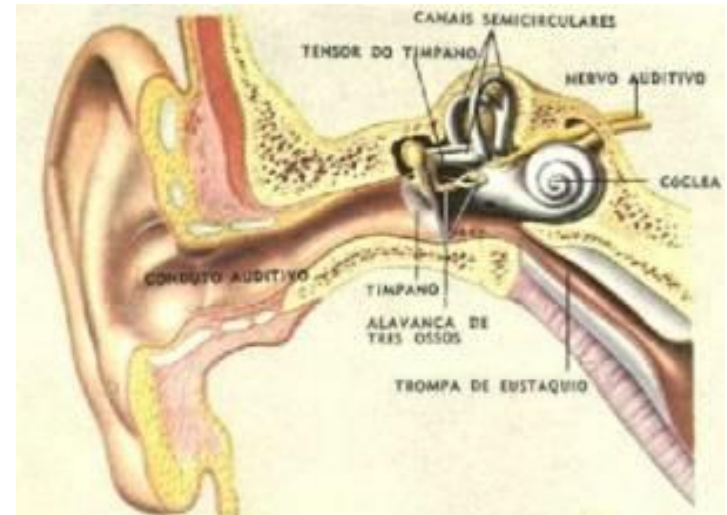


Conceitos básicos do som

- As ondas sonoras são ondas longitudinais (mecânicas) produzidas por deformações provocadas pela diferença de pressão num meio elástico qualquer (gasoso,sólido, liquido. ex: ar) precisando deste meio para se propagar;
- Propagam-se a uma velocidade de 344m/s (no ar a uma temp. 20º);
- Estas ondas gozam das mesmas propriedades que outras ondas mecânicas (refracção, difracção, reflecção, interferência, etc);
- Tom: onda sinusoidal, tem uma só frequência;
- Sons mais ricos contêm mais que um tom;
- Instrumentos musicais têm uma frequência fundamental e harmónicas desta.

Sistema Auditivo Humano

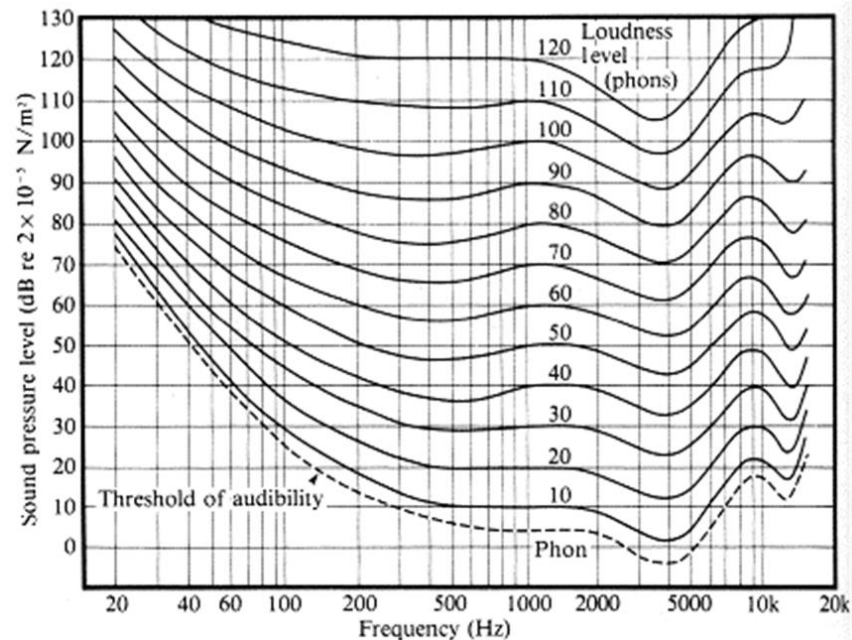
- Ouvido externo:
 - Pavilhão auditivo (orelha)
 - Colecta e filtra o som
 - Conduto auditivo Externo
 - Transmite som para o tímpano
 - Câmara de ressonância
- Ouvido médio
 - Ossículos (martelo, bigorna, estribo)
 - Conduz a vibração do tímpano ao ouvido interno
- Ouvido interno
 - Cóclea (tem líquido no seu interior) este activa o órgão de Corti (transmite impulsos nervosos)
 - Trompa Eustáquio (equilibra a pressão interior/exterior)
 - Aparato vestibular – Equilíbrio e sistema de visão



Sistema Auditivo Humano (percepção)

- Ouvimos o som de qualquer direcção, mas temos dificuldade em saber essa direcção
- Altamente sensível a defeitos e a sinais atractivos
- Ignora defeitos graves em sinais assumidos como irrelevantes
- Sensível a interrupções no som.
- Adaptativo
- Não-linear
 - Resposta na frequência
 - Banda 20Hz-20Khz
 - maior sensibilidade na banda 500-4000Hz
 - Volume
 - Reage ao logaritmo da intensidade

■ curvas de Fletcher-Munson



Sistema Auditivo Humano (percepção)

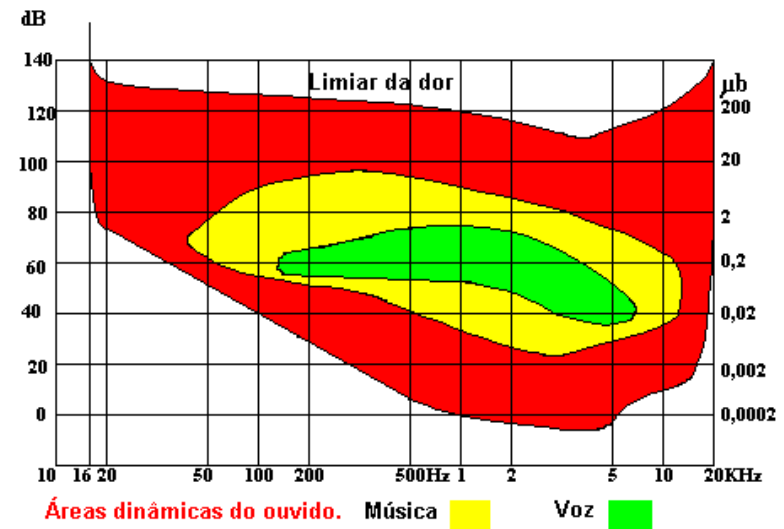
- Nível sonoro (sound pressure level)

$$SPL = 10 \log \frac{P}{P_o} \text{ W/m}^2$$

$$P_o = 10^{-12} \text{ W/m}^2$$

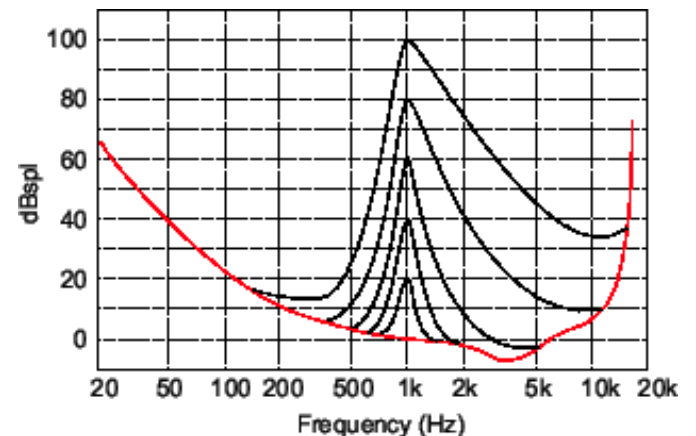
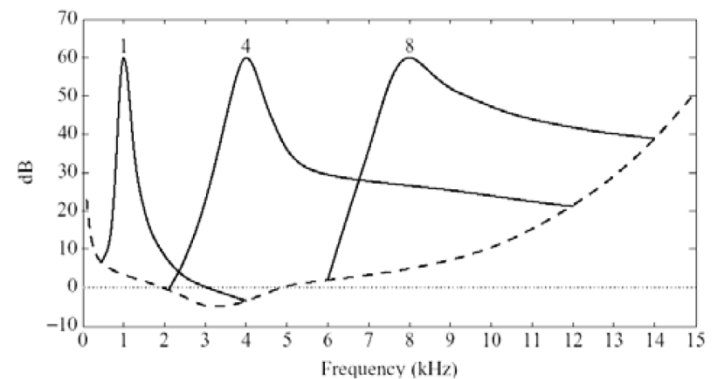
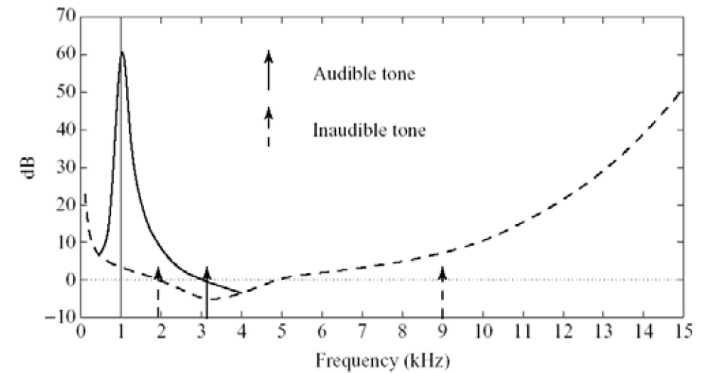
pressão $2 \times 10^{-5} \text{ N/m}^2$ at 1KHz

- 160 dB – Lesão no tímpano
- 140 dB – Limiar da dor
- 130 dB – Barulho do avião
- 120 dB – Limiar de desconforto
- 100 dB – Concerto Rock
- 70 dB – Rua com tráfego
- 60 dB – conversa comum
- 35 dB – casa comum
- 0 db – silêncio (limiar de audição)



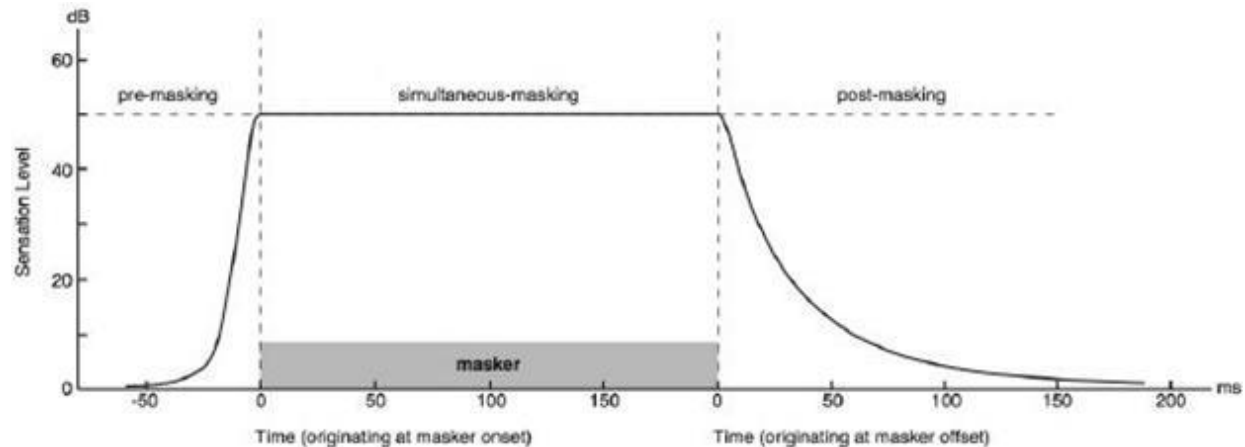
Efeito de máscara na frequência

- Um tom grave pode mascarar outro mais agudo
- O contrário não é verdade
- Quanto maior a frequência
Maior a gama de frequências
afectada
- Se aumentar a potência da
mascara, maior a banda
afectada.



Efeito de máscara temporal

- Um tom máscara outro que ocorra dentro de uma janela temporal:
 - Quanto maior o nível de som, maior a duração da máscara
 - Depende da duração do tom aplicado



- Máscara no tempo e na frequência

