

Tópicos

- √ Introdução
- √ Representação de Informação Multimídia
- √ Digitalização
- • Mídias Texto, Imagem, Áudio e Vídeo
- Compressão, Padrões de Compressão
- Comunicação Multimídia
- Protocolos de Rede, Redes Locais
- Redes Sem Fio, Bluetooth
- Sincronismo de Mídias
- Qualidade de Serviço
- Tópicos Avançados: Criptografia, Watermarking;
- Realidade Virtual
- Serviços Multimídia: Vídeo Sob Demanda, Videoconferência
- Sistemas Multimídia Avançados: Ambientes Virtuais Colaborativos

Mídia Áudio

Características

- Voz Humana
 - Faixa de frequência de 50Hz a 10kHz
 - Taxa de amostragem de Nyquist de 20ksps (2 x 10 kHz)
 - 12 bits por amostra para manter qualidade
- Música
 - Faixa de frequência de 15Hz a 20kHz
 - Taxa de amostragem de Nyquist de 40ksps (2 x 20 kHz)
 - 16 bits por amostra para manter qualidade (stereo => x2)
- Na prática sinais de áudio são amostrados a uma taxa menor e com um menor número de bits por amostra \leq canal de comunicação.
- Produzido por:
 - Microfone: Sinal analógico (codec de sinal de áudio)
 - Sintetizador de áudio: Sinal digital

Mídia Áudio

Exercício

Calcule a taxa de transmissão (em bits por segundo) necessários para transmissão de um sinal de voz e um de música amostrados à taxa de Nyquist e com, respectivamente, 12 e 16 bits por amostra. Calcule também a taxa de transmissão para a mesma música em stéreo, isto é, com codificação separada para canais direito e esquerdo.

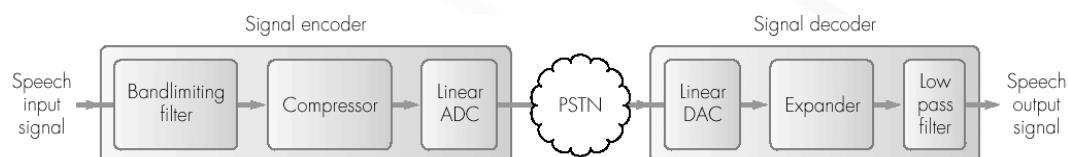
Calcule a quantidade de memória necessária para armazenar 10 minutos de música stéreo como descrito acima.

Calcule o número de minutos de música que caberiam em um CD de 650MB.

Mídia Áudio

Voz => PCM

- A maioria das aplicações para comunicação eram realizadas sobre redes PSTN (*Public-Switched Telephone Network*), que operavam em modo analógico.
- Circuitos digitais foram introduzidos, mas tiveram de conviver com circuitos analógicos por um bom tempo, sendo desenvolvidos tendo como base as características do meio analógico: 200Hz a 3.4kHz de banda de transmissão. Embora a taxa de Nyquist seja 6.8kHz, uma taxa de 8kHz foi escolhida por conta da baixa qualidade dos filtros limitadores de banda à época.
- 7 bits por amostra foram adotados nos EUA e Japão (56kbps) e 8 bits por amostra na Europa (64kbps). 8 bits por amostra prevaleceu.
- Quantização Linear.
- PCM - *Pulse Code Modulation* - Recomendação G.711 da ITU-T (CCITT).

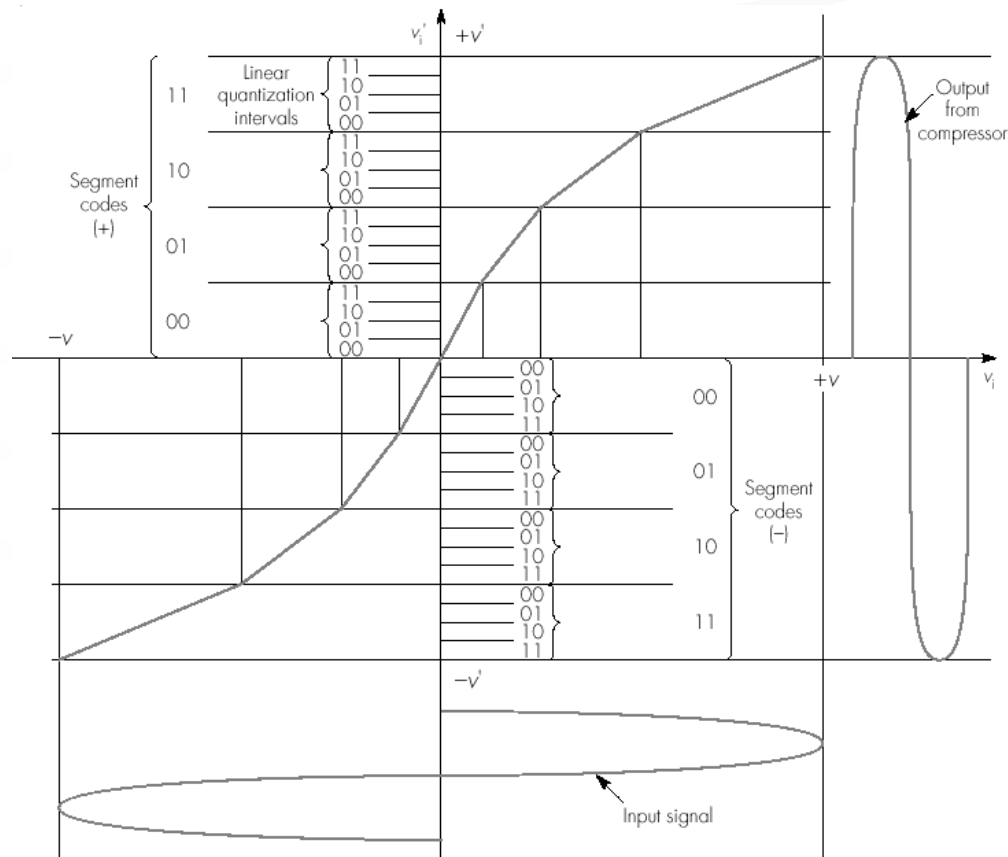


Mídia Áudio

PCM - Compressão

- O ouvido mais sensível a ruídos em sinais baixos que altos.
- Antes do sinal ser amostrado e convertido para uma forma digital pelo ADC, ele passa por um circuito compressor, que comprime a amplitude do sinal.
- O nível de compressão, e portanto dos intervalos de quantização, aumenta conforme a amplitude do sinal aumenta.
- O ADC então executa quantização linear do sinal comprimido.

Característica da Compressão

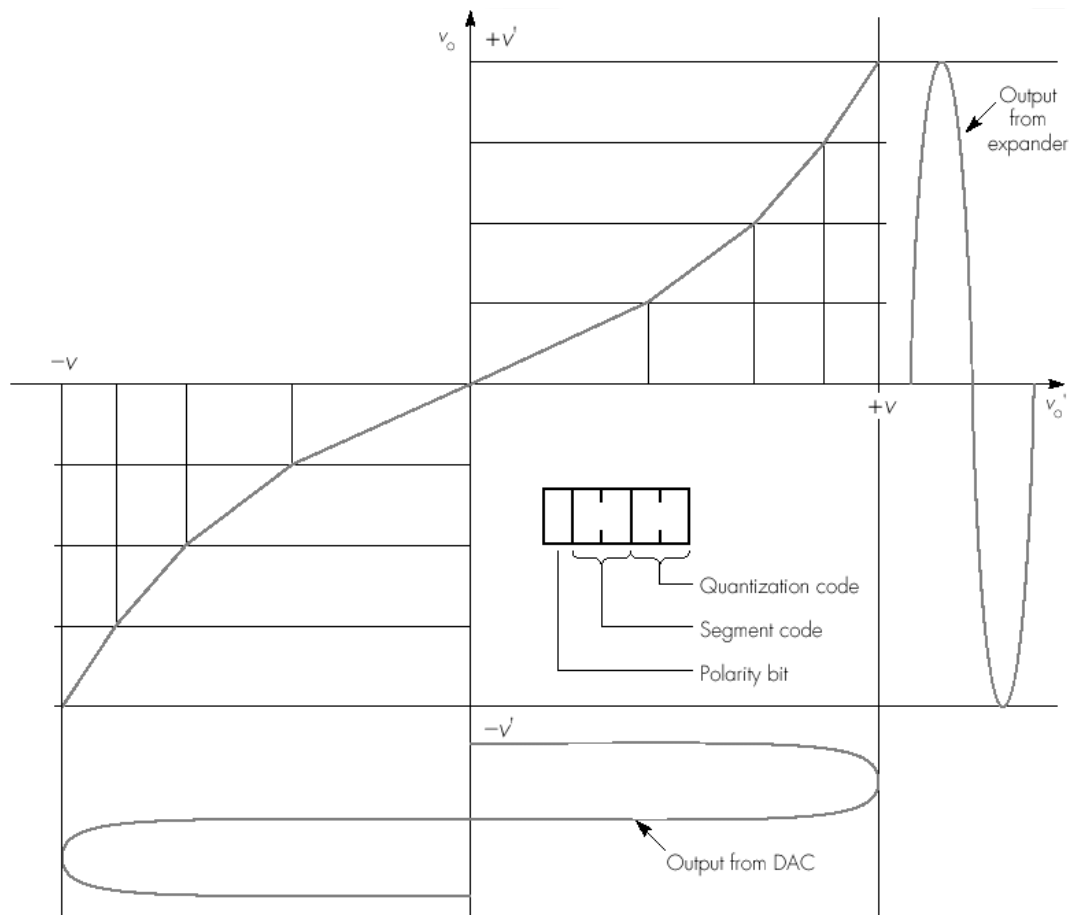


Mídia Áudio

PCM - Expansão

- No lado oposto os sinal digital é primeiro convertido em analógico pelo DAC e é, posteriormente, passado por um circuito expensor.
- Sistemas mais modernos costumam ter a compressão e expansão executadas digitalmente, mas o princípio é o mesmo.
- Compressão/expansão provêm qualidade equivalente a amostragem linear de 12 bits com apenas 8 bits.
- Existem dois padrões de compressão em uso: u-law (μ) e A-law, respectivamente nos EUA/Japão e Europa/outros.

C
a
r
a
c
t
e
r
í
s
t
i
c
a
d
a
E
x
p
a
n
s
ã
o



Note that in the G.711 standard a 3-bit segment code and 4-bit quantization code are used.

Mídia Áudio

Música com Qualidade de CD

- CDs são dispositivos de armazenamento digital que são utilizados para armazenar dados e música (padrão CD-DA)
- Áudio com banda de frequência de 15Hz a 20kHz => taxa de Nyquist de 40ksps. Na prática 44.1kHz é usado para lidar com imperfeições de filtros limitantes de banda, bem como para gerar taxa de transmissão de dados coincidente com padrões existentes (sinal amostrado a cada 23 microsegundos).
- O padrão prevê a utilização de 16 bits por amostra com 65536 intervalos de quantização lineares.
- Som estéreo requer dois canais separados, com um total de:
 - Taxa de bits = taxa de amostragem \times número bits por amostra
= $44.1 \times 10^3 \times 16 = 705.6$ kbps
= 2×705.6 kbps = 1.411 Mbps

Mídia Áudio

Exercício

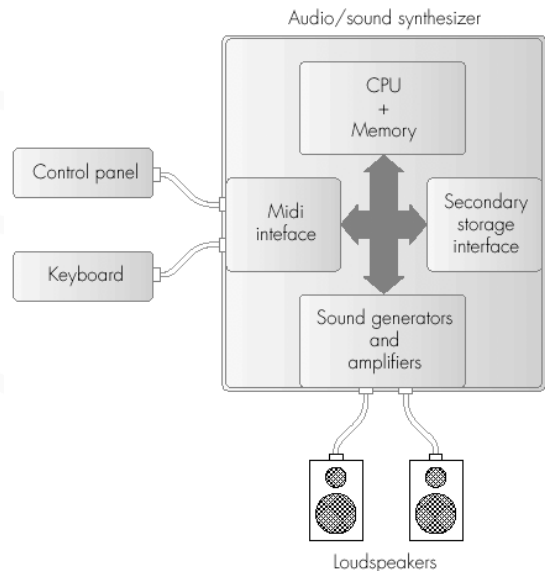
Assumindo que o padrão CD-DA está sendo usado, calcule:

- A capacidade do CD necessária para armazenar 60 minutos de um título multimídia.
- Tempo necessário para transmitir 30 segundos do título multimídia supra-citado usando um canal de transmissão com:
 - 64 kbps
 - 1.5 Mbps

Mídia Áudio

Áudio Sintetizado

- Componentes principais
 - Computador
 - Teclado (como de um piano)
 - Geradores de som
- O computador recebe comandos pelo teclado (ou outro dispositivo) e ativa geradores de som apropriados, que são posteriormente convertidos por um circuito DAC e enviados às caixas de som.
- Painel de controle define volume, tipo de som a ser emitido, etc.



Mídia Áudio

Padrão MIDI

- Padrão MIDI (*Music Instrument Digital Interface*) define comandos de entrada e saída para geradores de som (conjunto de mensagens p/ comunicação, conectores, cabos e sinais elétricos para conectar dispositivos ao sintetizador).
- O formato MIDI consiste de:
 - Um byte de status, que define o evento que gerou a mensagem.
 - Um número de bytes de dados, que define os parâmetros associados com o evento.
- Parâmetros incluem a pressão exercida na tecla do teclado, etc.
- A quantidade de dados gerado por uma sequência MIDI consiste de um conjunto de bytes como descrito acima, I.e. a quantidade de dados é pequena.
- É necessário ter uma placa de som (ou software) que interprete a sequência de comandos MIDI.
- Placas de som utilizam ou síntese FM ou amostras produzidas pelo instrumento real (síntese *wavelet*) para sintetizar o som solicitado.

Mídia Vídeo

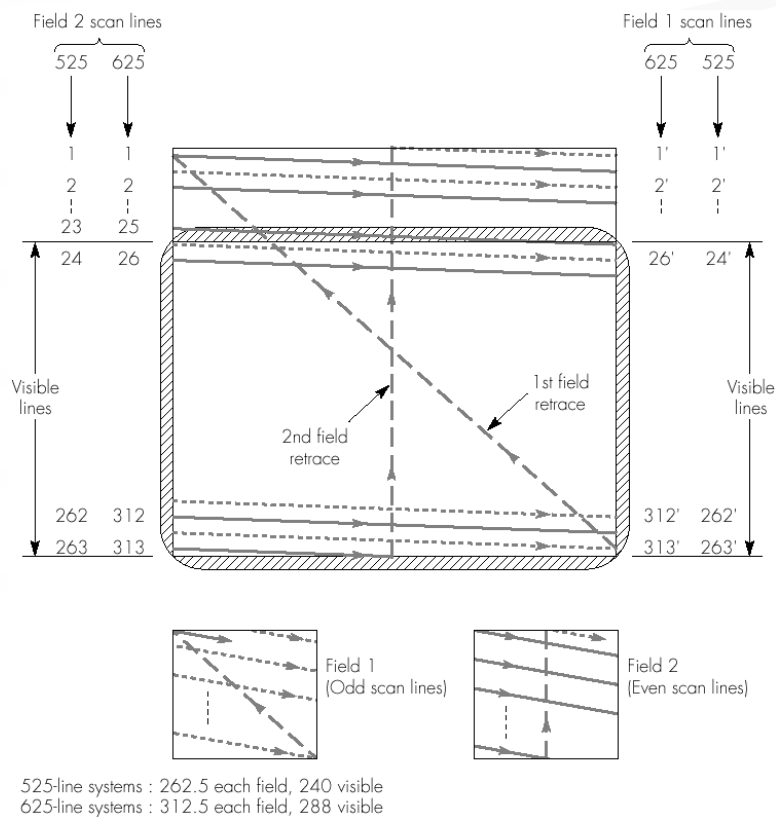
Introdução

- Vídeo é usado em um grande número de aplicações:
 - Lazer: TV, VCR/DVD.
 - Comunicação: Videotelefonia, videoconferência.
 - Interação: Janela contendo video clip.
- A qualidade do vídeo utilizada depende do tipo de aplicação multimídia utilizada.
 - Existe um conjunto de padrões para codificação de vídeo, um para cada tipo específico de aplicação.
 - Estudaremos primeiro padrões de difusão de sinais de televisão, que são a base para os demais padrões.

Mídia Vídeo

Difusão de Televisão

- Cores são criadas pela mistura de cores primárias vermelho, verde e azul
- O ser já tem a sensação de movimento a partir de 25 quadros por segundo, embora pelo menos 50 atualizações por segundo sejam necessárias para evitar *flickering*.
- Para minimizar a banda passante para a transmissão do sinal de vídeo, cada imagem é enviada em duas metades:
 - linhas pares
 - linhas ímpares
- Cada metade é conhecida como um campo. Dois campos montam um quadro através de uma técnica denominada varredura interlaçada.



Mídia Vídeo

Cor em TVs

- Sinais de difusão de TV colorida tinham de possibilitar a apresentação do mesmo em aparelhos em preto e branco.
- Aparelhos em cores tinham de poder apresentar imagem em preto e branco quando o sinal assim o for.
- Pelos motivos acima um padrão diferente do RGB foi adotado para difusão de sinais de TV.

Mídia Vídeo

Propriedades da Cor

- Brilho: representa a quantidade de energia que estimula o olho (intensidade). Varia em tons de cinza de preto para branco. É independente da cor da fonte em questão.
- Hue: Representa a cor da fonte, cada cor tem uma frequência / tamanho de onda diferente, o que é usado pelo olho humano para detectar diferentes tons de cor.
- Saturação: Representa a força ou intensidade da cor. Uma cor desbotada tem nível baixo de saturação, enquanto que uma cor reluzente, como o vermelho, tem maior saturação. Uma cor saturada não tem muito branco adicionada a ela.

Mídia Vídeo

Sinais de Cor

- Luminância é um termo usado para se referir ao brilho de uma fonte de cor, ou quantidade de branco.
- Hue e Saturação são usados para definir Crominância, que é referente à cor da fonte de luz.
- Luminância $Y = 0.299R + 0.587G + 0.114B$
- Duas componentes de crominância C_R (crominância vermelha) e C_B (crominância azul) são utilizadas para representar a cor da fonte em questão.
 - $C_B = B - Y$
 - $C_R = R - Y$
- Pode-se calcular G a partir das três equações, portanto $Y C_R C_B$ contém toda a informação necessária para calcular RGB, com a vantagem de ter o sinal em preto e branco (Y) disponível para aparelhos em P&B.

Mídia Vídeo

Componentes de Crominância

- Com o objetivo de garantir que sinais de difusão em cores ocupam a mesma largura de banda de sinais monocromáticos, os sinais $Y C_R C_B$ precisam ser combinados para transmissão. O sinal resultante é conhecido como sinal de vídeo composto.
- Para evitar degradação no sinal composto, as componentes de crominância são reduzidas.
- No sinal PAL, as componentes $Y C_R C_B$ são, portanto, convertidas nas componentes YUV como segue:
 - $Y = 0.299R + 0.587G + 0.114B$
 - $U = 0.493 (B - Y)$ e $V = 0.877 (R - Y)$
- No sinal NTSC, as componentes $Y C_R C_B$ são convertidas em YIQ:
 - $I = 0.74 (R - Y) - 0.27 (B - Y)$
 - $Q = 0.48 (R - Y) + 0.41 (B - Y)$

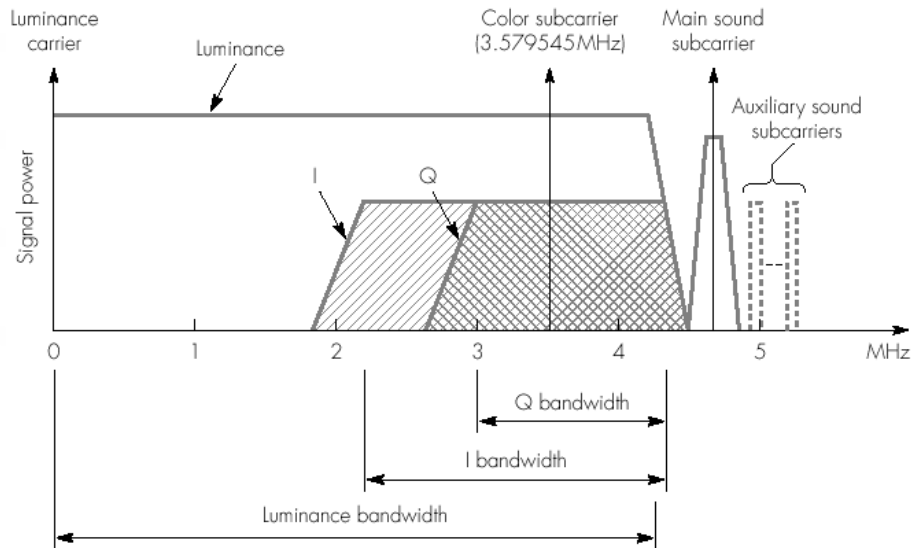
Mídia Vídeo

Exercício

- Calcule as expressões para YUV e YIQ em termos das componentes básicas RGB

Mídia Vídeo

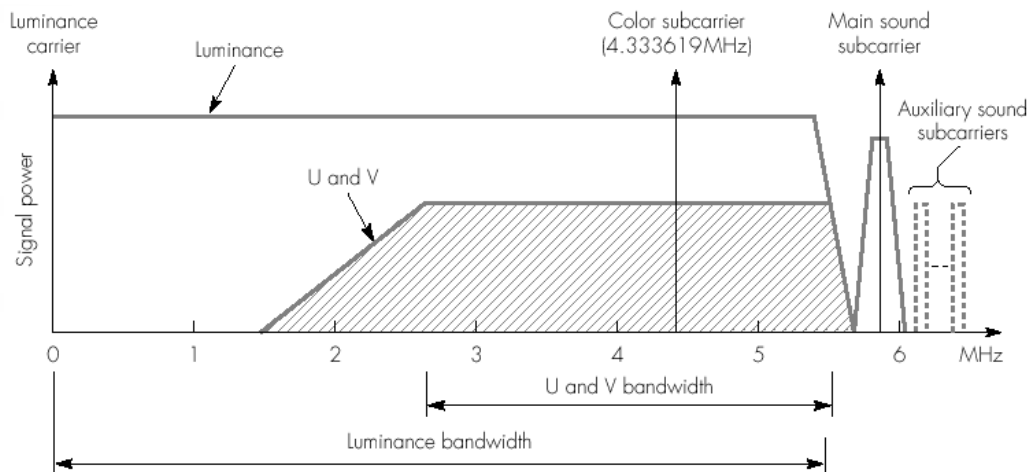
Largura de Banda do Sinal NTSC



Olho mais responsivo à componente I que Q, cada um utiliza uma banda de 2MHz e 1MHz respectivamente.

Mídia Vídeo

Largura de Banda do Sinal PAL



Maior banda passante total (5.5 x 4.2) permite que ambas as componentes utilizem uma banda de 3MHz

Mídia Vídeo

Largura de Banda do Sinal

- O sinal de áudio segue após a o vídeo composto em um ou mais faixas de frequência. Usualmente a primeira faixa (principal) contém som monoaural, enquanto que as demais acrescentam som estéreo. O sinal completo, incluindo faixas frequência utilizadas por sinais de áudio, é denominado Sinal “*Complex Baseband*”.

Mídia Vídeo

Vídeo Digital

- Embora em transmissão analógica o sinal de vídeo precisa ter suas componentes combinadas em um único “sinal composto”, no meio digital costuma-se codificar cada componente separadamente, o que permite maior facilidade na execução de operações de processamento do sinal.
- Já que os sinais são codificados separadamente, poderia-se utilizar as componentes RGB, porém o sinal RGB requer que similar resolução seja utilizada para cada componente. Estudos sobre a percepção visual do olho humano tem demonstrado que o olho é menos sensível à informação de cor do que ele é para a informação de luminância. I.e. os sinais de crominância toleram menor resolução na codificação sem perda significativa de informação para o olho humano. Pode-se, portanto, diminuir a taxa de bits na codificação de um sinal de vídeo se se utilizar componentes de luminância e crominância ao invés de RGB.

Mídia Vídeo

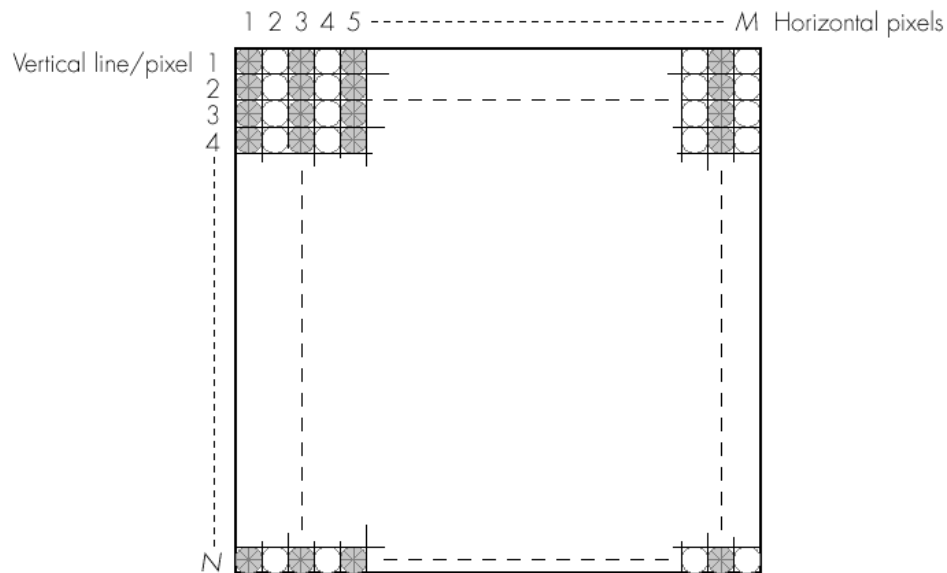
Vídeo Digital

- Estações de TV tem frequentemente utilizado digitalização de vídeo para converter um formato em outro. Com o objetivo de padronizar este processo e facilitar a troca de programas de TV entre estações ao redor do mundo o ITU-R (União Internacional de Telecomunicações - Divisão de Comunicação de Rádio), antiga CCIR, definiu o padrão CCIR-601.
- Variações do padrão foram posteriormente apresentadas objetivando outras aplicações, tais como TV digital, videofone, videoconferência, etc. Tais padrões são conhecidos como “*Digitation Formats*” e todos se baseiam no fato de que as componentes de crominância toleram resolução bem mais reduzida que o componente de luminância.

Mídia Vídeo

Formato 4:2:2

- Este é o formato original utilizado na recomendação CCIR-601 para uso em estúdios de TV.
- Um sinal analógico de TV pode ter uma faixa de frequência de até 6MHz para a componente luminância e 3MHz para os sinais de crominância. Para digitalizar este sinal é necessário utilizar um filtro limitador de banda de 6MHz para a luminância e 3MHz para os componentes de crominância, com taxas de amostragem de 12Msps e 6Msps, respectivamente. No padrão utiliza-se 13.5MHz para a luminância e 6.75MHz para ambos os sinais de crominância.
- O número de amostras por linha é de 720 para luminância.
- Para cada 4 amostra de luminância, 2 amostras de cada componente de crominância é feito.



O = Y, + = C_b, X = C_r sample positions

525-line systems: $M = 720$, $N = 480$, 60Hz refresh rate (interlaced)

$Y = 720 \times 480$, $C_b = C_r = 360 \times 480$

625-line systems: $M = 720$, $N = 576$, 50Hz refresh rate (interlaced)

$Y = 720 \times 576$, $C_b = C_r = 360 \times 576$

Mídia Vídeo

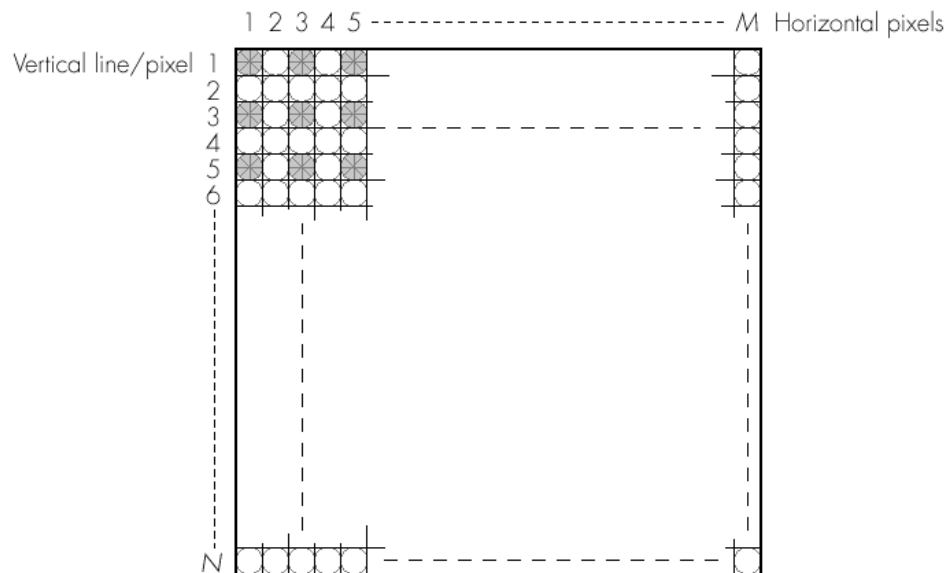
Exercício

- Calcule a taxa de transmissão e os requisitos de memória para cada quadro de vídeo digital resultante da digitalização de ambos formatos com 525 e 625 linhas assumindo uma codificação 4:2:2.
- Calcule a quantidade total de memória necessária para armazenar 1.5 horas de vídeo neste formato digital.

Mídia Vídeo

Formato 4:2:0

- Desenvolvido para aplicações em difusão que não usam a técnica de alternamento de linhas (quadros e campos).
- Reduz-se a resolução das componentes de crominância em ambas as direções, resultando em uma componente de crominância (de cada tipo) para cada 4 amostras de luminância.
- Resulta na mesma resolução de luminância, mas metade das resoluções de crominância, quando comparado com 4:2:2
- Valor da crominância é repetido para as amostras de luminância sem amostras de crominância.



O = Y, + = C_b , X = C_r sample positions

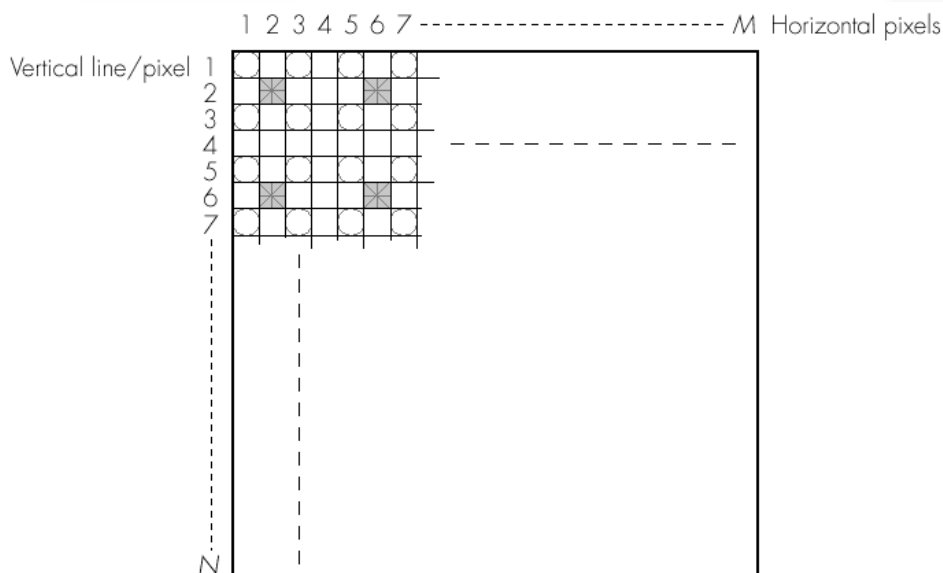
525-line systems: $M = 720$, $N = 480$, 60Hz refresh rate (interlaced)
 $Y = 720 \times 480$, $C_b = C_r = 360 \times 240$

625-line systems: $M = 720$, $N = 576$, 50Hz refresh rate (interlaced)
 $Y = 720 \times 576$, $C_b = C_r = 360 \times 288$

Mídia Vídeo

HDTV e SIF

- HDTV
 - TV de alta definição.
 - Resolução de 16:9 ao invés de 4:3.
 - 1440 x 1152 pixels (1080 linhas visíveis).
 - Utiliza 4:2:2 com 50/60Hz ou 4:2:0 com 25/30Hz.
- SIF
 - Source Intermediate Format (SIF) produz qualidade de imagem similar àquela produzida por VCRs.
 - Utiliza a metade da resolução espacial em ambas as direções (como 4:2:0), técnica conhecida como sobamostragem, além de utilizar metade da taxa de atualização (*refresh rate*), também conhecido como resolução temporal.
 - Desenvolvido para aplicações de armazenagem com varredura progressiva.
 - Técnica conhecida como amostragem 4:1:1
 - Amostras em falta são calculadas por interpolação.



O = Y , + = C_b , X = C_r sample positions

SIF: 525-line systems: $M = 720$, $N = 480$ with 30Hz refresh rate (non-interlaced)
 $Y = 360 \times 240$, $C_b = C_r = 180 \times 120$

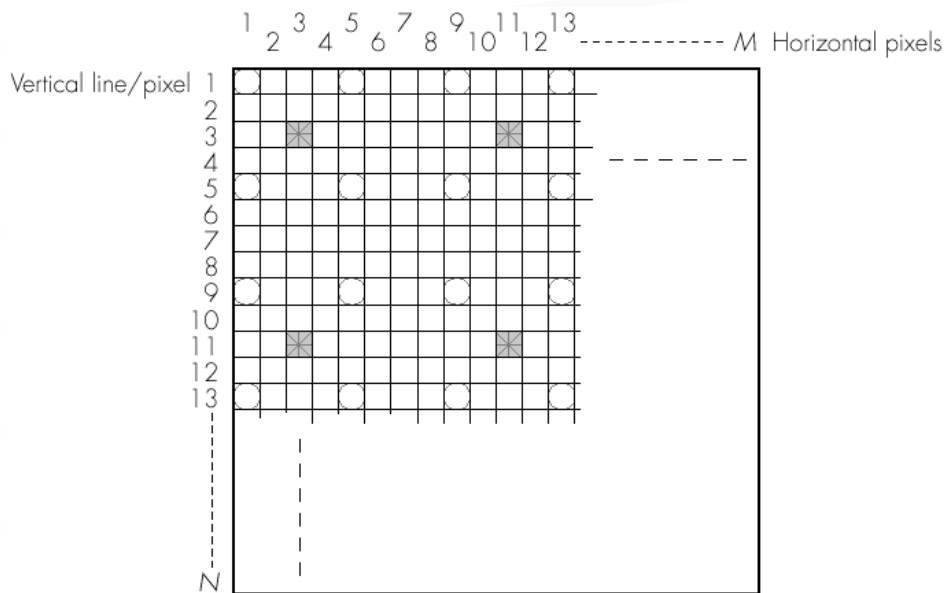
625-line systems: $M = 720$, $N = 576$ with 25Hz refresh rate (non-interlaced)
 $Y = 360 \times 288$, $C_b = C_r = 180 \times 144$

CIF: $M = 720$, $N = 576$ with 30Hz refresh rate (non-interlaced)
 $Y = 360 \times 288$, $C_b = C_r = 180 \times 144$

Mídia Vídeo

CIF, QCIF, SQCIF

- *Common Intermediate Format* (CIF) foi definido para uso em sistemas de videoconferência. O formato é derivado de SIF, com digitalização 4:1:1. Para converter quadros com 525 linhas é necessário utilizar um conversor de linhas, enquanto quadros com 625 linhas precisa de um conversor de taxa de frames (fps).
 - $Y = 360 \times 288$ pixels.
 - $C_R = C_B = 180 \times 144$ pixels.
- Derivativos de mais alta/baixa resolução:
 - 4CIF com $Y=720 \times 576$, $C_R=C_B=360 \times 288$
 - 16CIF com $Y=1440 \times 1152$, $C_R=C_B=720 \times 576$
 - QCIF com $Y=180 \times 144$, $C_R=C_B=90 \times 72$
 - SQCIF (Sub-QCIF) com $Y=128 \times 96$ e $C_R=C_B=64 \times 48$



O = Y, + = C_b , X = C_r sample positions

$M = 720$, $N = 576$ with 15Hz or 7.5Hz refresh rate (non-interlaced)

$Y = 180 \times 144$, $C_b = C_r = 90 \times 72$

Respostas dos Exercícios

Assuming the bandwidth of a speech signal is from 50 Hz through to 10 kHz and that of a music signal is from 15 Hz through to 20 kHz, derive the bit rate that is generated by the digitization procedure in each case assuming the Nyquist sampling rate is used with 12 bits per sample for the speech signal and 16 bits per sample for the music signal. Derive the memory required to store a 10 minute passage of stereophonic music.

Answer:

- (i) Bit rates: Nyquist sampling rate = $2 f_{\max}$
Speech: Nyquist rate = $2 \times 10 \text{ kHz} = 20 \text{ kHz}$ or 20 ksps
Hence with 12 bits per sample, bit rate generated
= $20 \text{ k} \times 12 = 240 \text{ kbps}$
Music: Nyquist rate = $2 \times 20 \text{ kHz} = 40 \text{ kHz}$ or 40 ksps
Hence bit rate generated = $40 \text{ k} \times 16 = 640 \text{ kbps}$ (mono)
or $2 \times 640 \text{ k} = 1280 \text{ kbps}$ (stereo)
- (ii) Memory required: Memory required = bit rate (bps) \times time (s) / 8 bytes
Hence at 1280 kbps and 600 s,
Memory required = $\frac{1280 \times 10^3 \times 600}{8} = 96 \text{ Mbytes}$

Assuming the CD-DA standard is being used, derive:

- (i) the storage capacity of a CD-ROM to store a 60 minute multimedia title,
- (ii) the time to transmit a 30 second portion of the title using a transmission channel of bit rate:
 - 64 kbps
 - 1.5 Mbps.

Answer:

- (i) The CD-DA digitization procedure yields a bit rate of 1.411 Mbps.
Hence storage capacity for 60 minutes

$$= 1.411 \times 60 \times 60 \text{ Mbits}$$

$$= 5079.6 \text{ Mbits or } 634.95 \text{ Mbytes}$$

- (ii) One 30 second portion of the title = $1.411 \times 30 = 42.33 \text{ Mbits}$
Hence time to transmit this data:

$$\text{At 64 kbps} = \frac{42.33 \times 10^6}{64 \times 10^3} = 661.4 \text{ s} \quad (\text{about 11 minutes})$$

$$\text{At 1.5 Mbps} = \frac{42.33 \times 10^6}{1.5 \times 10^6} = 28.22 \text{ s}$$

Derive the scaling factors used for both the U and V (as used in PAL) and I and Q (as used in NTSC) color difference signals in terms of the three R, G, B color signals.

Answer:

$$\text{PAL:} \quad Y = 0.299 R + 0.587 G + 0.114 B$$

$$U = 0.493 (B - Y) \quad \text{and} \quad V = 0.877 (R - Y)$$

$$\text{Hence} \quad U = 0.493 B - 0.493 (0.299 R + 0.587 G + 0.114 B)$$

$$= -0.147 R - 0.289 G + 0.437 B$$

$$\text{and} \quad V = 0.877 R - 0.877 (0.299 R + 0.587 G + 0.114 B)$$

$$= 0.615 R - 0.515 G - 0.100 B$$

$$\text{NTSC:} \quad I = 0.74 (R - Y) - 0.27 (B - Y)$$

$$= 0.74 R - 0.27 B - 0.47 Y$$

$$= 0.599 R - 0.276 G - 0.324 B$$

$$Q = 0.48 (R - Y) + 0.41 (B - Y)$$

$$= 0.48 R + 0.41 B - 0.89 Y$$

$$= 0.212 R - 0.528 G + 0.311 B$$

Derive the bit rate and the memory requirements to store each frame that result from the digitization of both a 525-line and a 625-line system assuming a 4:2:2 format. Also find the total memory required to store a 1.5 hour movie/video.

Answer:

525-line system: The number of samples per line is 720 and the number of visible lines is 480. Hence the resolution of the luminance (Y) and two chrominance (C_b and C_r) signals are:

$$Y = 720 \times 480$$
$$C_b = C_r = 360 \times 480$$

Bit rate: Line sampling rate is fixed at 13.5 MHz for Y and 6.75 MHz for both C_b and C_r , all with 8 bits per sample.

Hence: Bit rate = $13.5 \times 10^6 \times 8 + 2 (6.75 \times 10^6 \times 8) = 216 \text{ Mbps}$

Memory required: Memory required per line = $720 \times 8 + 2 (360 \times 8)$
= 11 520 bits or 1440 bytes

Hence memory per frame, each of 480 lines = $480 \times 11\,520$
= 5.5296 Mbits or 691.2 kbytes

and memory to store 1.5 hours assuming 60 frames per second:
= $691.2 \times 60 \times 1.5 \times 3600 \text{ kbytes}$
= 223.9488 Gbytes

625-line system: Resolution: $Y = 720 \times 576$
 $C_b = C_r = 360 \times 576$

Bit rate = $13.5 \times 10^6 \times 8 + 2 (6.75 \times 10^6 \times 8) = 216 \text{ Mbps}$

Memory per frame = $576 \times 11\,520 = 6.635\,55 \text{ Mbits}$ or 829.44 kbytes

and memory to store 1.5 hours assuming 50 frames per second:
= $829.44 \times 50 \times 1.5 \times 3600 \text{ kbytes}$
= 223.9488 Gbytes

It should be noted that, in practice, the bit rate figures are less than the computed values since they include samples during the retrace times when the beam is switched off. Nevertheless, as we can deduce from the computed values, both the bit rate and the memory requirements are very large for both systems and it is for this reason that the various lower resolution formats have been defined.