

Capítulo 2

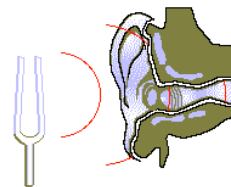
Dados Multimídia

Para estudar os sistemas multimídia distribuídos, é necessário conhecer como as diversas mídias são representadas digitalmente. Este capítulo apresenta a representação digital de áudios, imagens e vídeos. Em seguida, ele apresenta as principais características e requisitos destas mídias.

2.1 Representação Digital de Áudios

Descrevendo sons como forma de onda

O áudio percebido é causado por ondas mecânicas longitudinais que alcança o tímpano. Ela é gerada por qualquer fonte que produz esta vibração do ar. À medida que a onda se propaga as partículas do meio vibram de forma a produzir variações de pressão e densidade segundo a direção de propagação. Estas alterações resultam numa série de regiões de altas e baixas pressões chamadas de condensações e rarefações respectivamente.



A onda sonora é uma onda contínua no tempo e amplitude. A onda apresentada na figura 1a pode ser um exemplo de onda sonora. O padrão de oscilação, como mostrado na Figura 1, é chamado de forma de onda (*waveform*). A forma de onda é caracterizada por um **período** e **amplitude**. O período é o tempo necessário para a realização de um ciclo; intervalo de tempo que, num fenômeno periódico, separa a passagem do sistema por dois estados idênticos. A **freqüência** (f) é definida como o inverso do período e representa o número de períodos em um segundo. A freqüência é normalmente medida em Hz (Hertz) ou ciclos por segundo (cps). A amplitude (A) do som é define um som leve ou pesado. A amplitude, que no caso do som é medido em decibéis - dB, define a intensidade (volume) do som. Por exemplo, o limiar da dor é de 100 a 120 dB. Outro parâmetro é a fase (ϕ) é relativo a posição da onda no tempo.

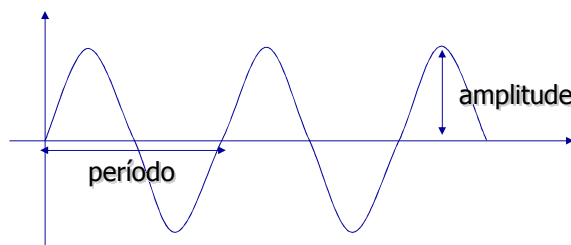


Figura 1. Forma de Onda

O distúrbio da pressão de ar é dependente do tempo e espaço. Na posição de um locutor ou de um detector, os sons podem ser descritos por valores de pressão que variam apenas com o tempo (valores dependentes do tempo). Usando as variáveis apresentadas anteriormente, uma onda senoidal pode ser representada no tempo por $s(t) = A \sin(2\pi ft + \Phi)$. A Figura 2 apresenta graficamente diferentes formas de onda com diferentes valores de amplitude, freqüência e fase.

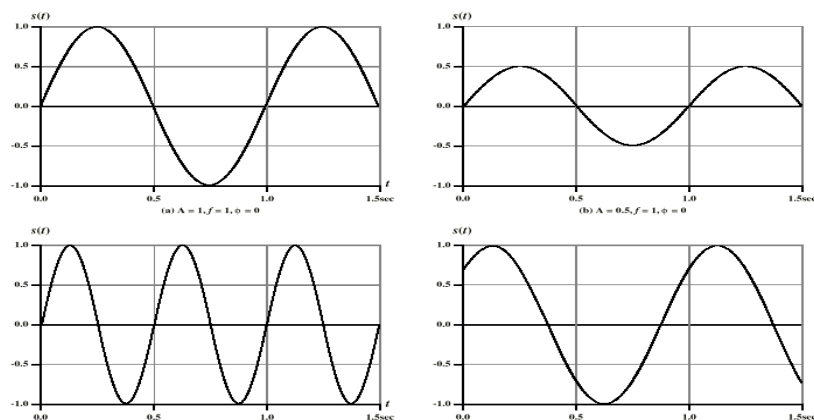


Figura 2. Conversão analógico/digital e digital/análogica

Como a onda de som ocorre naturalmente, ela nunca é perfeitamente suave ou uniformemente periódica como a forma de onda da Figura 1. Os sinais sonoros normalmente são formados por múltiplas freqüências (diferentes sinais). Os diferentes sinais são chamados de componentes de onda senoidal (componentes de freqüência do som). A análise de Fourier diz que qualquer sinal pode ser formado pela combinação de várias ondas ou componentes senoidais. É possível montar uma função baseada no domínio da freqüência para representar os sinais. Por exemplo, a Figura 3c representa um sinal composto de duas senoides (a e b), sendo que a senoide (a), chamada componente de base, é definida por $\sin(2\pi ft)$, e a segunda b), $(1/3)\sin(2\pi (3ft))$ (com 1/3 da amplitude e 3 vezes a freqüência da primeira senoide).

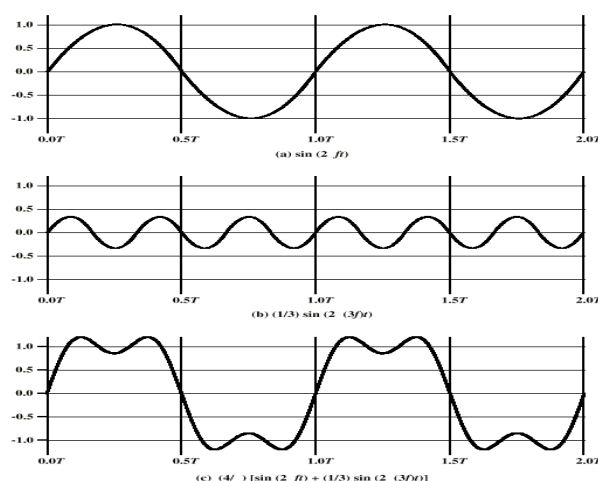


Figura 3. Sinal composto de duas componentes de freqüência.

Muitas vezes é preferível representar o sinal no domínio da freqüência, como ilustrado na Figura 4 para o caso do sinal da Figura 3c.

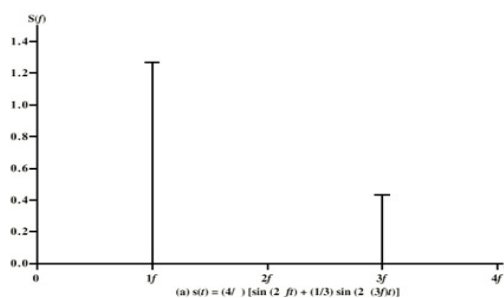


Figura 4. Representação no domínio da freqüência.

Quando os componentes de freqüência do distúrbio de ar estão na faixa de 20 Hz a 20.000 Hz, o som é audível. A maioria dos sistemas multimídia trabalha com esta faixa de freqüência.

2.1.1 Transformação da onda de pressão em sinal elétrico

A forma de onda de áudio é convertida em um sinal elétrico contínuo (analogico) por um microfone, como ilustrado na Figura 5. Este sinal elétrico é medido normalmente em volts.

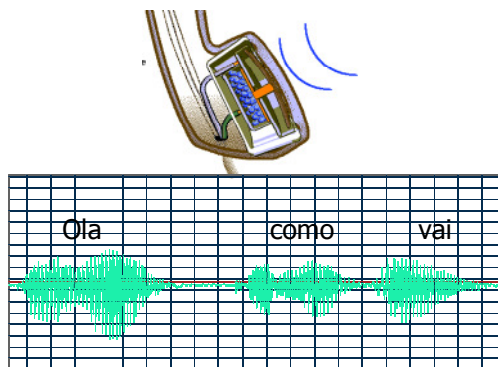


Figura 5. Conversão da onda sonora em sinal elétrico

2.1.2 Digitalização do Áudio

Para que sistemas computacionais processem e comuniquem sinais de áudio, o sinal elétrico deve ser convertido em um sinal digital. O mecanismo que converte o sinal de áudio digital em analógico é chamado de Conversor Analógico para Digital (CAD), ou digitalização. Digitalização aqui é o processo envolvido na transformação de sinais analógicos (sinal elétrico gerado pelo microfone) em sinais digitais. Esta conversão é realizada pelos dispositivos chamados de CODECs (Codificador/Decodificador). Para a conversão de sinais analógicos em digital é necessária a realização de três passos: amostragem, quantificação e codificação. A figura 2 ilustra o processo de digitalização de um sinal analógico no domínio do tempo.

Amostragem

Nesta etapa um conjunto discreto de valores analógicos é amostrado em intervalos temporais de periodicidade constante, como apresentado na figura 1a. A frequência de relógio é chamada de taxa de amostragem ou **frequência de amostragem**. O valor amostrado é mantido constante até o próximo intervalo. Isto é realizado através de circuitos *sampling and hold*. Cada uma das amostras é analógica em amplitude: ele tem qualquer valor em um domínio contínuo. Mas isto é discreto no tempo: dentro de cada intervalo, a amostra tem apenas um valor.

Segundo o teorema de Nyquist: se um sinal analógico contém componentes de frequência até f Hz, a taxa de amostragem deve ser ao menos $2f$ Hz. Se o sinal analógico tiver componentes de frequência maiores que a frequência de Nyquist ocorre a pseudonímia (aliasing). Nesta situação, estes componentes maiores que $2f$ são convertidos em frequências mais baixas na reconstrução. Para evitar isto, são utilizados filtros anti-pseudonímia, trata-se de filtros do tipo passa baixa para eliminar as frequências maiores que a de Nyquist. Como filtros não são perfeitos, é necessário filtrar frequências um pouco maior que $2f$. Por exemplo, a taxa de amostragem de CD-audio é de 44,1 kHz, e dos tapes de áudio digital (DAT) é de 48kHz para cobrir uma faixa audível de frequência de 20 kHz.

O sistema telefônico foi projetado para transmitir frequências da voz humana. A voz humana gera frequências entre 15Hz e 14kHz. Na telefonia, por razões econômicas, a faixa de voz escolhida foi entre 300 e 3400 Hz (largura de banda de 3,1kHz), o que garante 85% de inteligibilidade (palavras compreendidas) e 68% de energia da voz humana [Soares, 2002]. No entanto, para evitar a interferência entre sinais que fluem em canais vizinhos, a largura de banda de um canal de voz foi definida em 4KHz, onde as extremidades (0 a 300Hz e de 3,3 a 4 KHz) são usadas como banda de guarda [Soares, 2002]. No sistema telefônico é comum usar uma frequência de amostragem de 8 kHz para converter este sinal em digital.

Um efeito que pode ocorrer na digitalização é a pseudonímia (aliasing). Se o sinal a ser digitalizado tiver componentes de frequência maiores que a frequência de Nyquist ocorre a pseudonímia. Em termos simplificados estes componentes de frequência são convertidos em frequências mais baixas na reconstrução do sinal. Para evitar este efeito deve ser usado filtros anti-pseudonímia, que são filtros

passa-baixa para eliminar as frequências maiores que a de Nyquist. Suponha que você queira digitalizar um som utilizando uma frequência de amostragem de 44.100Hz. Este fenômeno gera distorções toda vez que um sinal de entrada (o som que você quer digitalizar) possui componentes de frequência acima de 22.050Hz é amostrado em 44.100Hz. Para evitar estas distorções, o áudio analógico precisa ser filtrado antes da conversão A/D, impedindo que qualquer conteúdo acima de 22kHz chegue ao conversor e seja amostrado. Pois bem, todo filtro possui uma "curva de atuação", começando a filtrar um pouco antes da frequência de corte, para poder "barrar" efetivamente tudo acima dela. Filtros com curvas "suaves" são mais fáceis de se construir e mais baratos. Filtros de curvas abruptas, além de caros, podem gerar problemas de fase e prejudicar os agudos. A solução é utilizar taxas de amostragens altos, como 88.1 ou 96kHz, e conseguir gravar todo o espectro audível, sem se preocupar com aliasing ou outras distorções causadas pelo filtro. O áudio digital em alta definição pode então ser filtrado na saída (filtro após o conversor D/A) ou então digitalmente por um plugin, caso precise ser convertido para taxas de amostragens mais baixos.

Para contornar o problema dos filtros em 44.1kHz e diminuir o ruído de "digitalização", muitos conversores A/D (e D/A) utilizam oversampling, fazendo amostragens em alto SR. Depois, aplicam filtros digitais precisos para fazer o down-sampling para 44.1kHz, antes de armazenar o áudio. Portanto, existem muitos conversores D/A de 44.1kHz que utilizam técnicas e/ou filtros de alta qualidade, sem que seja necessário gravarmos em alta definição. Não é raro que a frequência de oversampling seja de 64x, 128 vezes ou até mais a frequência de amostragem final.

Quantificação

O processo de converter valores de amostras contínuas em valores discretos é chamado de quantificação. Neste processo nós dividimos o domínio do sinal em um número fixo de intervalos. Cada intervalo tem o mesmo tamanho e recebe um número. Na figura 1c estes intervalos são numerados de 0 a 7. A cada amostra dentro de um intervalo é atribuído o valor do intervalo.

Quando o mesmo tamanho de passo de quantificação é usado na conversão A/D sem olhar para a amplitude do sinal, o processo de conversão é dito uniforme. Este processo de conversão A/D é chamado de **modulação por pulso codificado** (PCM - *Pulse Coded Modulation*). Algumas vezes, nem todos os valores amostrados são retidos após a quantificação. No caso ilustrado pela Figura 6c, todos os valores amostrados foram retidos.

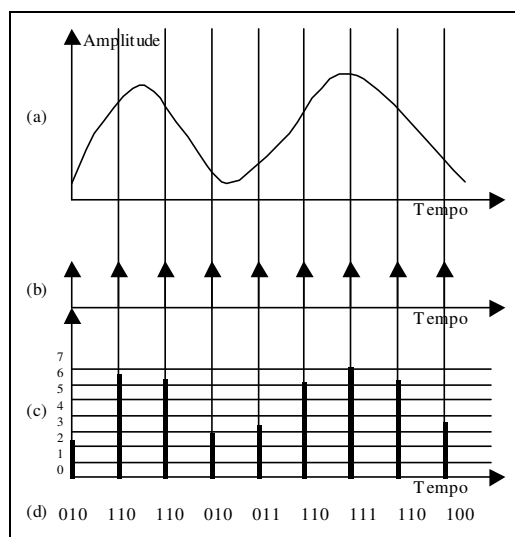


Figura 6. Conversão A/D [Lu, 96]: (a) sinal analógico; (b) pulsos de amostragem; (c) valores amostrados e intervalos de quantificação; (d) sequência digital

O número de intervalos de quantificação (que na codificação vão definir o número de bits por amostra) é limitado. Quanto menor o intervalo (maior o número de bits), melhor é a quantificação da amostra (a sua medida). Como esta medida é limitada pelo tamanho dos intervalos de quantificação, ocorre o chamado erro de quantização. Este erro se traduz auditivamente por um ruído, ouvido na reprodução do som reconstruído (ruído de quantização).

Quantificação não linear

O PCM é simples, mas não é eficiente: a quantificação linear resulta em uma mais elevada SNR na região de amplitude de sinal mais altas que na região de mais baixas amplitudes. Esta elevada SNR na região de amplitude mais altas não aumenta a qualidade percebida. Isto, pois nós somos mais sensíveis às componentes de amplitude mais baixas.

A fim de explorar este fato o tamanho de passo de quantificação que aumenta logarithmicamente com a amplitude do sinal é muito usado na quantificação de sinais de voz. Neste caso, os passos de quantificação são menores quando a amplitude é baixa. Esta técnica de compressão realizada uma transformação de um sinal linear em um sinal não linear.

Na prática, uma quantificação uniforme é aplicada a um sinal não linear transformado em vez de aplicar uma quantificação não uniforme ao sinal linear. Os resultados destas duas abordagens são o mesmo. O processo de transformação de um sinal linear em não linear é chamado de *companding*. A digitalização uniforme de um sinal *companded* é chamada de *companded PCM*. Esta é na realidade uma técnica de compressão analógica realizada antes da conversão A/D e expandida após a conversão D/A. Usando esta técnica, o sinal de 8 bits pode produzir um sinal de qualidade equivalente aquele sinal codificado PCM de 12 bit.

Áudio na telefonia

Na área da telefonia digital, utiliza-se um método de transformação de natureza logarithmica para comprimir áudios. Ele mapeia 13 ou 14 bits dos valores linearmente quantificados para códigos de 8 bits. O mapeamento de 13 para 8 é conhecido como transformação A-law, e o mapeamento de 14 para 8 é conhecido como transformação μ -law. Usando esta transformação, a SNR da saída transformada é mais uniforme na faixa de amplitude do sinal de entrada. A transformação A-law é usada normalmente em redes ISDN (Redes Digitais de Serviços Integrados) na Europa, e μ -law na América do Norte e Japão. A recomendação ITU (antiga CCITT) G.711 (vista mais adiante), especifica as transformações A-law e μ -law.

Codificação

A codificação consiste em associar um conjunto de dígitos binários, chamado de *code-word*, a cada valor quantificado. No caso da figura 1d, oito níveis de quantificação são usados. Estes níveis podem ser codificados usando 3 bits, assim cada amostra é representada por 3 bits.

Em algumas aplicações de telefonia, a digitalização da voz humana utiliza 16 bits por amostra, que então leva a 2^{16} ou 65.536 passos de quantificação. Em outras aplicações de compressão de voz, algumas vezes, apenas 8 quantificações por bits são necessários, produzindo apenas 256 passos de quantificação.

Taxa de bits

Taxa de bits é definida como o produto entre taxa de amostragem e o número de bits usados no processo de quantificação. Por exemplo, supondo uma frequência de 8k Hz e 8 bits por amostra, a taxa de bits necessária à telefonia é igual a $8000 \times 8 = 64$ kbps.

2.1.3 Exemplos de qualidade de áudio digital

A tabela abaixo mostra a taxa de amostragem e o número de bits usados para cada amostra para várias aplicações de áudio. Relembrando, quanto maior a taxa de amostragem e maior o número de bits por amostragem, maior é a qualidade do áudio restituído, mas com isso maior é a taxa de bits. Note na tabela que para áudio estéreo, tal como CD-Audio, dois canais são necessários.

Aplicações	Nº de canais	Largura de banda (Hz)	Taxa de amostragem	Bits por amostra	Taxa de bits
CD-Audio	2	20-20000	44.1 kHz	16	1,41 Mbps
DAT	2	10-22000	48 kHz	16	1,53 Mbps
Telefone Digital	1	300-3400	8 kHz	8	64 Kbps
Rádio digital, long play DAT	2	30-15000	32 KHz	16	1,02 Mbps

2.1.4 Apresentação do áudio

Em sistemas multimídia todas as informações multimídia são representadas internamente no formato digital. Mas humanos reagem a estímulos sensoriais físicos, assim a conversão digital-para-analógico (ou conversão D/A) é necessária na apresentação de certas informações (figura 2).

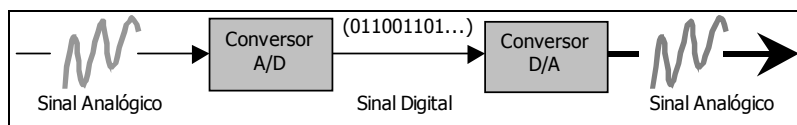
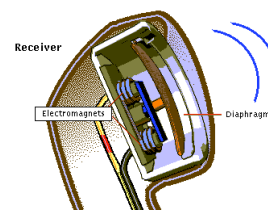


Figura 7. Conversão analógico/digital e digital/analogica

Para a apresentação do áudio digitalizado é necessário realizar a transformação de uma representação artificial do som em uma forma de onda física audível pelo ouvido humano. Para isto, são utilizados Conversores Digital-para-Analógico (CDA).

Normalmente os conversores CAD e CDA são implementados em uma única placa. Um exemplo de placa de áudio é Creative Sound Blaster AWE64, possibilitando até 16 bits por amostras, produzindo áudio qualidade CD.



2.1.5 Problemas da Representação digital

Apesar de aportar várias vantagens, a digitalização de informações multimídia apresenta algumas deficiências que são apresentadas nesta seção.

Distorção

O maior problema da utilização de informações multimídia na forma digital é a distorção de codificação (amostragem, quantificação e codificação dos valores) introduz distorção ao sinal analógico restituído. O sinal gerado após a conversão D/A não é idêntico ao original (como ilustrado na figura 3), assim a informação apresentada ao usuário não é idêntica àquela capturada do mundo real.

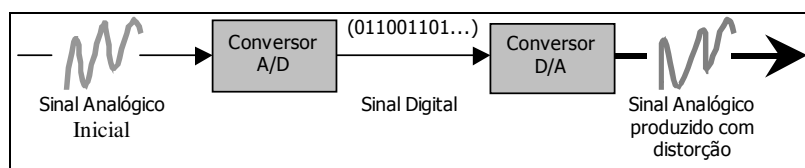


Figura 8. Conversão analógico/digital e digital/analogica

Aumentando a taxa de amostragem e o número de bits usado para codificação reduz esta distorção. Segundo o teorema de Nyquist: se um sinal analógico contém componentes de frequência até f Hz, a taxa de amostragem deve ser ao menos $2f$ Hz. Mas há uma clara limitação tecnológica neste ponto: a capacidade de armazenamento não é infinita, e os sistemas de transmissão têm largura de banda¹ limitadas.

Na maior parte dos sistemas multimídia, os usuários finais das informações multimídia são os humanos. Como nem todas as componentes de frequências são percebidas pelos humanos, a solução para reduzir a distorção é escolher um balanço apropriado entre a precisão da digitalização e a distorção percebida pelo usuário.

¹ Em transmissão, a faixa de frequências que o sistema pode transmitir sem excessivas atenuações. Em redes de dados, algumas vezes este termo é usado para referenciar a capacidade total de uma rede expressa em bits por segundo. Em tais casos, o termo apropriado é taxa de bits.

Necessidade de grandes capacidades de armazenamento

Outro problema gerado pela digitalização de informações multimídia é a necessidade de meios de armazenamento digital com grandes capacidades principalmente para o armazenamento de vídeos, imagens e áudios. Por exemplo, oito minutos de som estereofônico de qualidade CD são suficientes para completar 80 megabytes do disco duro de um PC padrão. Para reduzir este problema, faz-se uso de algoritmos de compressão reduzem este requisito.

2.1.6 Representação simbólica da música: o padrão MIDI

Como visto anteriormente, qualquer som pode ser representado como um sinal de som digitalizado, que é uma seqüência de amostras, cada uma codificada por dígitos binários. Esta seqüência pode ser descompactada como nos discos compactos de áudio ou compactados. Uma característica deste modo é que ele não preserva a descrição semântica do som. A menos que sejam utilizadas técnicas de reconhecimento complexas, o computador não sabe se a seqüência de bits representa, por exemplo, uma fala ou música, e se música que notas são usadas e por que instrumentos.

Algumas representações de som preservam a semântica da informação. No caso da codificação da fala, pode-se usar um texto e atributos como voz masculina ou feminina, sotaque e taxa de palavras. A música também pode ser descrita de uma maneira simbólica usando técnicas similares às pautas musicais. O formato mais utilizado para isto é aquele definido no padrão MIDI (*Musical Instrument Digital Interface*). Este padrão define como codificar todos os elementos musicais, tais como seqüências de notas, condições temporais, e o “instrumento” que deve executar cada nota (são 127 instrumentos e outros sons como aqueles produzidos por helicóptero, telefone, aplausos, etc.).

Arquivos MIDI são muito mais compactos que amostragens digitalizadas: um arquivo MIDI pode ser 1000 vezes menor que um arquivo CD áudio. Além disso, a representação MIDI é revisável (modificáveis). Mas MIDI apresenta algumas desvantagens, como: necessidade de um processamento extra de informação, e imprecisão dos instrumentos de som (variam com o dispositivo usado para a apresentação).

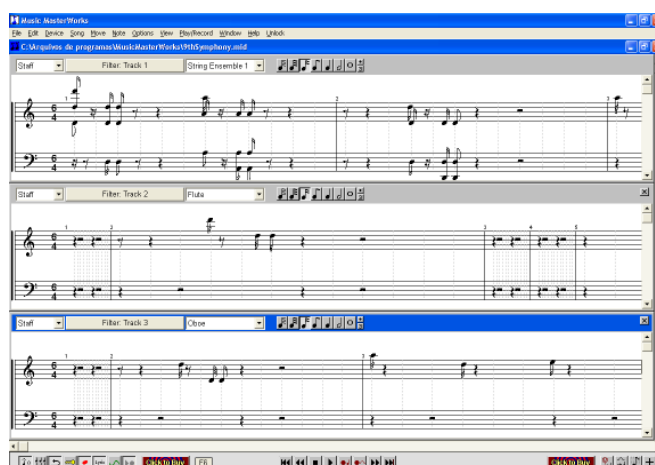


Figura 9. Editor de Midi

2.2 Captura de Imagens e Vídeos

Esta seção apresenta como imagens e vídeos são representados na forma analógica e como os dispositivos de captura operam (câmeras).

2.2.1 Representação analógica de imagens e vídeos

Descrevendo imagens monocromáticas com variáveis físicas

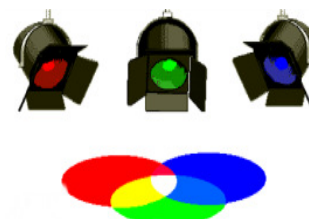
As imagens refletem radiações eletromagnéticas (luz) incidentes que estimulam os olhos do observador. A intensidade de luz é uma função da posição espacial do ponto refletido sob a imagem. Portanto, a imagem pode ser descrita pelo valor da intensidade de luz que é uma função de duas coordenadas espaciais. Se a cena observada não foi plana, uma terceira coordenada espacial é necessária.

Descrevendo imagens coloridas com formas de onda

Se a imagem não é monocromática, ela reflete diferentes comprimentos de onda. Assim uma função simples não é suficiente para descrever imagens coloridas, é necessário um espectro completo de comprimento de onda refletida, cada um com sua própria intensidade. Assim as imagens teriam que ser descritas pela conjunção de várias funções bidimensionais. Felizmente, o sistema visual humano tem certas propriedades que simplificam a descrição de imagens coloridas.

A Luz que consiste em uma distribuição espectral de intensidade estimula o sistema visual e cria uma resposta. A resposta nos olhos depende da sensibilidade do sistema visual aos comprimentos de onda. Testes realizados mostram que diferentes distribuições espectrais da luz podem dar a mesma resposta visual. Em outras palavras, é possível criar sensações de cores idênticas com diferentes combinações de comprimentos de onda (isto é, diferentes combinações de cores).

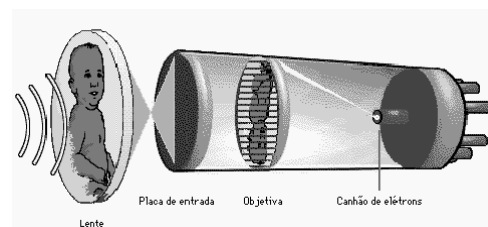
A teoria da cor foi desenvolvida por Thomas Young no início de 1802 e afirma que qualquer sensação de cor pode ser reproduzida pela mistura em proporções apropriadas de três luzes coloridas monocromáticas primárias. Esta é a teoria *Tristimulus*. Cores primárias são independentes no sentido que uma cor primária não pode ser obtida misturando outras duas cores primárias. A *Commission Internationale de l'Eclairage* (CIE) recomendou o uso de uma tripla particular de luz monocromática. Cada fonte de luz é definida pelo seu comprimento de onda ($\lambda_1 = 70$ nm, vermelho; $\lambda_2 = 546.1$ nm, verde; $\lambda_3 = 435.8$ nm, azul). Em vez de ser descrita por uma infinidade de funções bidimensionais, qualquer imagem colorida plana pode ser representada por um conjunto de três funções bidimensionais.



2.2.2 Captura e reprodução de imagens e vídeos

Captura e reprodução de imagens e vídeos monocromáticos

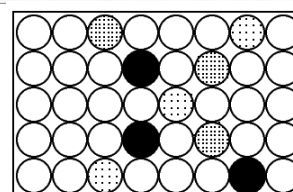
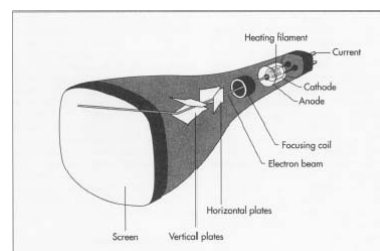
As imagens são capturadas usando câmeras da seguinte maneira: as lentes da câmera focam uma imagem de uma cena em uma superfície foto-sensível de sensores CCD (*Charge-Coupled Device*); o brilho de cada ponto é convertido em uma carga elétrica por uma camada foto-sensível, estas cargas são proporcionais ao brilho nos pontos; a superfície foto-sensível é rastreada por um feixe de elétrons para capturar as cargas elétricas, devendo ser feito rapidamente antes que a cena mude. Desta maneira a imagem ou cena é convertida em um sinal elétrico contínuo.



Nesta seção, por simplificação assume-se a captura e reprodução de vídeos monocromáticos, onde apenas um sinal de luminância é produzido (apenas a luminosidade é capturada, temos a imagem em tons de cinza). Neste caso são usadas **câmeras de Luminância**, que captam a imagem em tons de cinza, e gera um sinal só com a luminância da imagem. A imagem é gerada por um CCD monocromático que capta o tom de cinza que incide em cada célula do circuito. Este tipo de câmera é utilizado em geral para aplicações em visão computacional e nos casos onde a informação sobre a luminosidade da imagem é suficiente.

O sinal elétrico gerado pela câmera pode ser digitalizado através da amostragem, quantificação e codificação, da mesma forma que o sinal elétrico gerado pelo microfone. A diferença é que a frequência de amostragem é no domínio do espaço.

O dispositivo de apresentação de imagens mais utilizado é o tubo de raios catódicos (CRT). Eles são usados nos aparelhos de TV e monitores de computadores. Nas TVs e monitores monocromáticos, há uma camada de fósforo fluorescente no interior do CRT. Esta camada é rastreada por um feixe de elétrons na mesma forma do processo de captura na câmera. Quando o feixe toca o fósforo ele emite luz durante um curto instante. O brilho da luz depende da força do feixe. Quando quadros



repetem-se suficientemente rápidos a persistência da visão resulta na reprodução de um vídeo. Na prática, o sinal elétrico enviado da câmera para o dispositivo de apresentação deve conter informações adicionais para assegurar que o rastreamento esteja sincronizado com o rastreamento do sensor na câmera. Esta informação é chamada *sync information*.

Vídeos e Imagens Coloridos

Os sistemas de captura e apresentação de imagens coloridas (Por exemplo, TVs e monitores de computador) são baseados na teoria *Tristimulus* de reprodução da cor. Para capturar imagens coloridas, uma câmera divide a luz nos seus componentes vermelho, verde e azul. Estas três componentes de cor são focalizadas em sensores de vermelho, verde e azul, que convertem estas três componentes em sinais elétricos separados, o chamado sinal RGB.

Existem vários tipos de câmeras que geram imagens coloridas, entre elas temos:

- Câmera de cromaticidade (1 passo - 1 CCD) - Capta a imagem em cores, e gera um sinal de vídeo composto colorido, em apenas uma passagem. A imagem, em geral, não é profissional, pois é usado um único CCD com filtros RGB em cada célula, como pode ser visto na Figura 10. Este tipo de câmera é utilizado em aplicações multimídia ou em casos onde não é necessária uma imagem com muita qualidade. Esta é uma câmera do tipo doméstica (VHS, 8mm, VHS-C, etc), desta forma tem um custo baixo.

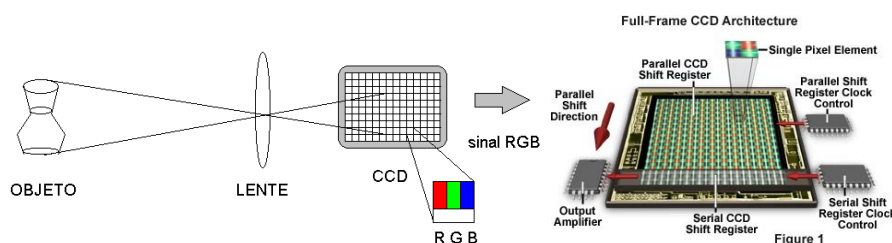


Figura 10. Câmera de cromaticidade (1 passo - 1 CCD)

- Câmera de cromaticidade (1 passo - 3 CCD) - Capta a imagem em cores, e pode gerar sinal de vídeo composto colorido, S-vídeo ou sinal RGB. Tem uma qualidade de imagem profissional, pois são usados 3 CCDs com filtros separados R, G e B em cada um, como pode ser visto na Figura 11. Por ter 3 CCDs independentes, cada um pode ter uma resolução maior, garantindo uma melhor resolução da imagem. É utilizada em aplicações profissionais, onde é necessária uma imagem com boa qualidade. Esta é uma câmera do tipo usado em produtoras e emissoras de TV (U-matic, BetaCAM, SVHS, Hi8, etc), desta forma tem um custo elevado.

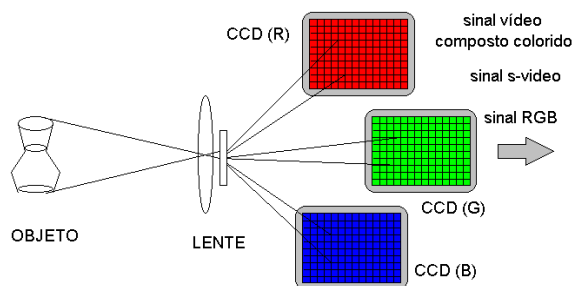


Figura 11. Câmera de cromaticidade (1 passo - 3 CCD)

- Câmera de cromaticidade (3 passos - 1 CCD) - Capta a imagem em cores, porém este processo é feito em 3 passos. É utilizado um único CCD para captar a imagem, sendo que para gerar uma imagem colorida é colocado um filtro externo para cada componente R, G e B (Figura 12). A digitalização da imagem então é feita em 3 passos, ou seja, para cada filtro é feita uma digitalização. Assim temos a informação das intensidades de cada componente RGB. Com esta informação é composta uma imagem colorida, pois para cada ponto temos a contribuição R, G e B. Este processo tem uma desvantagem pelo fato de que as imagens devem ser estáticas, pois é preciso trocar os filtros e fazer nova captação para os outros filtros. Tem uma boa qualidade de imagem, pois este CCD pode ter uma boa resolução, proporcionando uma melhor resolução da imagem. É utilizada em geral para aquisição de imagens de telescópio, onde é necessário uma imagem com alta definição e as imagens são relativamente estáticas. Esta é uma câmera que

pode ter um custo baixo, no caso de CCDs de pouca qualidade (baixa resolução), ou de alto custo se o CCD tiver alta resolução.

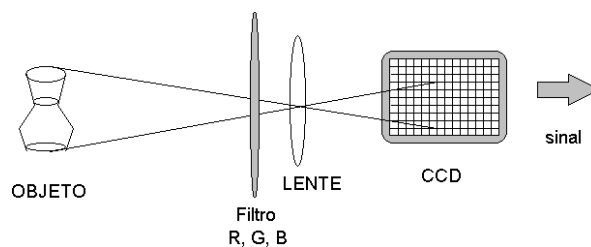


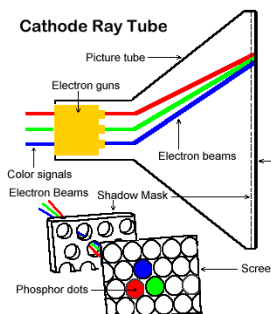
Figura 12. Câmera de crominância (3 passos - 1 CCD)

Como comentado anteriormente, os três componentes de cor de uma imagem são focalizados em sensores de vermelho, verde e azul, que convertem estes três componentes em sinais elétricos separados. Estes três sinais é o que é chamado de sinal RGB. Para digitalização das imagens coloridas, é necessário realizar a digitalização (amostragem, quantificação e codificação) destes três sinais separados.

Na realidade, o sinal analógico pode ser gerado da seguinte maneira [França Neto, 98]:

- Sinal RGB (red, green, blue): O sinal é separado pelas cores básicas, com isso é possível ter uma imagem mais pura. Ele é utilizado em câmeras e gravadores profissionais, imagens geradas por computador, etc.
- Sinal de vídeo composto colorido: os sinais das cores (RGB) são codificados em um único sinal seguindo um determinado padrão (NTSC, PAL-M, SECAM, etc) ;
- Sinal de luminância e crominância ou Y/C (S-video): o sinal é composto por duas partes, a luminância e a crominância; como isso a imagem tem uma melhor qualidade do que no vídeo composto. Muito usado por vídeos SVHS, laser disc, DVD e outros aparelhos que geram imagens de boa qualidade (acima de 400 linhas);

Em um monitor colorido, há 3 tipos de fósforos fluorescentes que emitem luzes vermelha, verde e azul quando tocadas por 3 feixes de elétrons. Estes fósforos são arranjados de tal forma que cada posição do vídeo tem 3 tipos de fósforo. A mistura da luz emitida destes 3 fósforos produz um ponto de cor.



Outros dispositivos de captura de imagens muito populares são as câmeras digitais e scanners.

A câmera fotográfica digital é um dispositivo de funcionamento semelhante a uma câmera fotográfica tradicional, porém a imagem não é armazenada em um filme e sim de forma digital em memória. Esta imagem é digitalizada através de um CCD, e armazenada de forma compactada ou não em um dispositivo de memória. A qualidade da imagem depende da qualidade e resolução do CCD e da compressão utilizada para armazenar a imagem digitalizada. A resolução, atualmente, varia entre 320x240 até algo em torno de 1600x1200 pontos por imagem. A imagem pode ser armazenada em vários tipos de memória como memórias não voláteis, cartões de memória, disquetes magnéticos, etc. As imagens podem ser transferidas para um computador por cabos ou leitores dos dispositivos de memória, e então são processadas, como é visto na Figura 13. O custo deste dispositivo pode ser baixo no caso de câmeras domésticas com poucos recursos e baixa resolução, ou muito alto quando a câmera possui recursos profissionais e alta resolução.

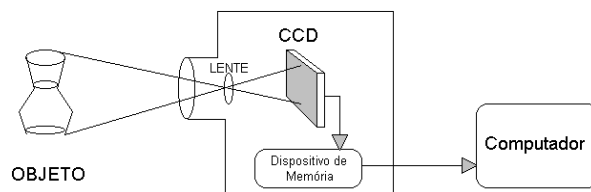


Figura 13. Câmera fotográfica digital

O scanner digitaliza a partir de imagens em papel. A imagem é colocada sobre uma superfície transparente, em geral plana ou cilíndrica, que se move numa direção ortogonal à um elemento de digitalização de linha (Figura 14). Este elemento se compõe de uma fonte de luz e de um sensor que mede a luz refletida linha por linha, em sincronismo com o deslocamento da imagem, ou do sensor. A resolução deste dispositivo está situada entre 50dpi a 4000dpi (pontos por polegada). Um dispositivo semelhante, é o film scanner, que obtém uma imagem digital a partir de imagens em transparências, utilizando o laser para maior resolução. Este dispositivo também atinge resoluções superiores a 2000 dpi.

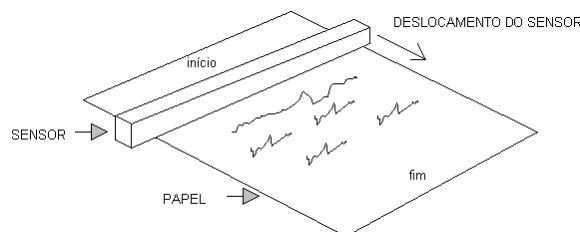


Figura 14. Esquema de funcionamento do scanner

Existem scanners de vários modelos, desde simples scanners de mão, até potentes scanners utilizados em grandes gráficas para captar imagens com um grau de detalhe muito grande. O scanner de mão, é geralmente usado em aplicações domésticas onde não se tem a necessidade de muita qualidade, é composto por um sensor que é arrastado sobre a imagem. Por outro lado os scanners de mesa proporcionam uma melhor qualidade na aquisição das imagens.

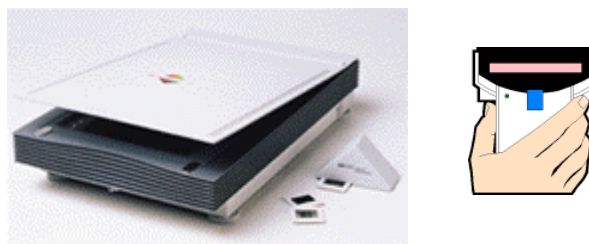


Figura 15. Exemplos de scanners

Vários fatores influenciam na qualidade dos scanners, entre eles temos:

- **Resolução óptica:** É a resolução via hardware que o scanner pode atingir. Outras resoluções acima desta podem ser obtidas com técnicas de interpolação, mas com perda de qualidade, ou seja, a máxima resolução obtida com toda a qualidade é a sua resolução óptica. A escolha da resolução de um scanner depende diretamente da aplicação para qual este vai ser usado, quanto mais detalhes forem necessários na captação das imagens, mais resolução óptica este scanner deve ter. Geralmente esta resolução começa de 300dpi, o que atende a uma boa parte das aplicações, até algo em entre 1200dpi a 2000dpi, que são dispositivos mais sofisticados para aplicações mais específicas.
- **Quantidade de bits para representar cada componente de cor:** este número é diretamente proporcional à quantidade de intensidade de cada componente. Geralmente temos cada componente representada por 8 bits, o que dá 256 níveis de intensidade para cada componente e um total de 24 bits o que corresponde a 16.777.216 cores. Alguns scanners já trabalham com 10 bits, ou seja 30 bits no total, e podem representar mais de um trilhão de cores.
- **Tamanho da área de leitura:** É a área máxima que pode ser usada para digitalizar uma imagem ou documento. A maioria dos modelos trabalha com os formatos A4 ou CARTA.
- **Velocidade de captação da imagem:** É o tempo para realizar a digitalização de uma imagem. Esta velocidade pode variar de acordo com o mecanismo de captação e da forma como os dados são transferidos para o computador. Geralmente são especificados alguns padrões para aferir esta velocidade, como uma imagem de tamanho padrão tipo A4, resoluções e número de cores fixas. Com isso é possível fazer uma análise do desempenho de um determinado scanner em relação a outro, e de acordo com a aplicação desejada, escolher qual o melhor. Outro fator que deve ser levado em conta é a forma como os dados são transferidos para o computador. Geralmente é feito através de interface paralela ou interface padrão SCSI (*Small Computer System Interface*).

No caso da interface paralela, a velocidade é bastante limitada, porém é mais fácil a instalação, pois a maioria dos computadores já vem com uma interface paralela, bastando conectar o cabo. No caso da interface SCSI, a velocidade pode ser bem maior dependendo da qualidade dos dispositivos, este tipo de interface é mais trabalhosa de se instalar, pois caso o computador não possua uma porta SCSI, sendo necessário a instalação de uma placa SCSI para fazer a ligação do scanner ao computador. A velocidade do sistema (cpu, memória, disco rígido, sistema operacional, etc) pode ter importância, dependendo da velocidade do scanner, ou seja, no caso de scanners mais rápidos o sistema pode tornar o processo de aquisição lento. A velocidade do scanner é geralmente proporcional ao seu custo, com isso deve ser levado em conta na hora de escolher o scanner ideal para a aplicação.

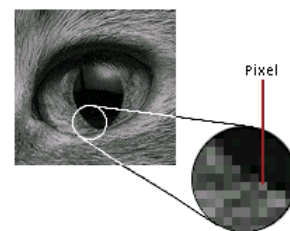
- **Qualidade do sensor:** Esta qualidade diz respeito a capacidade de representar as cores de forma mais fiel possível. Quanto melhor o sensor, mais semelhante a imagem gerada será da imagem real.

2.3 Representação digital de imagens

A seção anterior discutiu muitos conceitos e nomenclaturas de vídeos analógicos. Como nosso interesse é manipulação de imagens e vídeos digitais em sistemas multimídia, esta seção discute a representação digital de imagens e vídeos.

2.3.1 Imagens Digitais

Imagens não são revisáveis porque seu formato não contém informações estruturais. Elas podem resultar de capturas do mundo real (via escaneamento de uma página impressa ou foto, câmeras digitais) ou elas podem ser sintetizadas pelo computador (via programas de *paint*, captura da tela, conversão de gráficos em imagens bitmap). Depois de digitalizadas, as imagens podem ser manipuladas com editores de imagens (por exemplo, Photoshop), que não produzem documentos que retêm a estrutura semântica.



Formatos de Imagens

Imagens no computador são representadas por *bitmaps*. Um bitmap é uma matriz bidimensional espacial de elementos de imagem chamados de pixels. Um pixel é o menor elemento de resolução da imagem, ele tem um valor numérico chamado de amplitude. O número de bits disponíveis para codificar um pixel é chamado de profundidade de amplitude (ou de pixel). Exemplos típicos de profundidade de pixel são 1 (para imagens preto&branco), 2, 4, 8, 12, 16 ou 24 bits. O valor numérico pode representar um ponto preto e branco, um nível de cinza, ou atributos de cor (3 valores) do elemento de imagem em imagens coloridas.

O número de linhas da matriz de pixels (m) é chamado de resolução vertical da imagem, e o número de colunas (n) é chamado de resolução horizontal. Denominamos resolução espacial, ou resolução geométrica, ao produto $m \times n$ da resolução vertical pela resolução horizontal. A resolução espacial estabelece a frequência de amostragem final da imagem. Dessa forma, quanto maior a resolução mais detalhe, isto é, altas frequências, da imagem podem ser captadas na representação matricial. A resolução espacial dada em termos absolutos não fornece muita informação sobre a resolução real da imagem quando realizada em dispositivo físico. Isso ocorre porque ficamos na dependência do tamanho físico do pixel do dispositivo. Uma medida mais confiável de resolução é dada pela densidade de resolução da imagem que fornece o número de pixels por unidade linear de medida. Em geral se utiliza o número de pixels por polegada, ppi ("pixels per inch") também chamada de dpi ("dots per inch").

Formatos bitmap necessitam mais capacidade de armazenamento do que gráficos e textos. Como bitmaps ignoram a semântica, duas imagens de mesma dimensão (altura e largura) ocupam o mesmo espaço. Por exemplo, um quadrado ou uma foto digitalizada com dimensões idênticas ocupam o mesmo espaço. Os gráficos, como eles consideram a semântica, ocupam menos espaço.

2.3.2 Sistema RGB

No sistema RGB de representação de cor, uma cor é representada pela intensidade de três cores primárias (teoria Tristimulus): vermelho (Red), verde (Green) e azul (Blue), com cada valor variando de 0 a 255. Exemplos de cores familiares são apresentados abaixo:

- Branco = 255,255,255; Vermelho = 255,0,0; Verde = 0,255,0
- Azul = 0,0,255; Amarelo = 255,255,0; Preto = 0,0,0

A representação de imagens coloridas pode ser feita através de cores por componente (*true color*), cores indexadas, ou cores fixas. Essa representação vai depender do propósito e dos dispositivos que vão ser usados para trabalhar com essas imagens.

True Color

No True Color, cada pixel da imagem é representado por um vetor de 3 componentes de cores (RGB) com um certo número de bits para representar cada componente de cor (resolução de cor). Com isso, quanto maior for a resolução de cor mais qualidade teremos para representar as cores de cada pixel. Geralmente o número de bits para cada componente RGB é igual, ou seja quando temos um pixel sendo representado por 9 bits, usamos 3 bits para cada componente (3-3-3). Mas pode ser feita uma representação com diferentes valores para as componentes, por exemplo, uma representação 8 bits/pixel, pode ser usado 3 para componentes R, 3 para G e 2 para B (3-3-2), tal representação em um byte é comumente usado e tira proveito do fato que a percepção humana da componente azul é menos sensível que as outras componentes.

O número de bits para representar cada componente fornece a quantidade de cores que podem ser representados por essa componente. Ou seja, se n é a resolução de cor então a quantidade de níveis possíveis é de 2^n níveis. Por exemplo, uma imagem colorida representada por 12 bits/pixel, com 4 bits para cada componente RGB. Temos então: $2^4=16$ níveis para cada componente de cor RGB, o que nos possibilita representar até 4.096 cores diferentes ($16 \times 16 \times 16 = 4.096$), o que é equivalente a $2^{12} = 4.096$.

Temos alguns padrões de cores nesse formato que são:

Bits/pixel	Padrão	Componente de cor RGB	Máximo de cores
15 bits/pixel	High Color (15 bits)	5 bits/pixel, 32 níveis por componente	32.768 cores
16 bits/pixel	High Color (16 bits)	5/6 bits/pixel, 32/64 níveis por componente	65.536 cores
24 bits/pixel	True Color, (24 bits)	8 bits/pixel, 256 níveis por componente	16.777.216 cores

O padrão com 24 bits/pixel é o mais usado para representar com fidelidade as cores, pois o número de cores que podem ser representadas com essa resolução de cores é maior do que a visão humana pode reconhecer.

Cores Indexadas

Nas cores indexadas, cada pixel é representado por um índice que aponta para uma tabela de cores (paleta) que contem as informações sobre as cores (Figura 16). Temos então um número de cores que podem ser representadas, que é o número de entradas na paleta. A paleta por sua vez, tem em geral 24 bits para representar cada cor no formato RGB. Dessa forma podemos representar n cores de um conjunto com mais de 16 milhões de cores. Nesse caso, para representar esse tipo de imagem, as informações das cores da paleta devem constar da estrutura além das dimensões e sequência de índices.

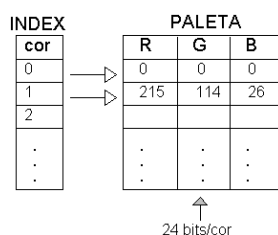


Figura 16. Índice e paleta de cores

O número de cores e a resolução de cor da paleta podem variar. Os dois padrões mais usados são apresentados na tabela abaixo.

Bits/pixel	Padrão	Resolução de cor da paleta (RGB)
4 bits/pixel	16 cores indexadas	24 bits/cor
8 bits/pixel	256 cores indexadas	24 bits/cor

Cores fixas

Nas cores fixas, cada pixel é representado por um índice que aponta para uma tabela de cores fixa. Esse sistema geralmente é usado quando o dispositivo não permite a representação de muitas cores, como no caso de placas de vídeos antigas ou padrões de cores (padrão de cores do MS Windows 3.x, 16 cores). O número de bits para representar um pixel depende do número de cores fixas, ou seja, para representar, por exemplo, 16 cores, são necessários 4 bits/pixel.

Imagens em Tons de Cinza

A representação de imagens em tons-de-cinza é feita discretizando a informação de luminância de cada ponto da imagem. Ou seja, cada pixel contém a intensidade de luminosidade representada em um certo número de bits. Assim, uma imagem com resolução de cor de 8 bits, pode representar até 256 níveis de cinza, variando do preto ao branco.

Os padrões mais usados são de 16 e 256 tons-de-cinza, 4 e 8 bits/pixel respectivamente. Representações com mais que 256 tons-de-cinza não são percebidas pela vista humana, ou seja, representar uma imagem com 256 níveis é suficiente para a maioria das aplicações.

Imagens Binárias

As imagens binárias são imagens com dois níveis, como preto e branco. São muito usadas por dispositivos de impressão e para representar imagens de documentos monocromáticos. Para representar um pixel de uma imagem binária, como o próprio nome diz, é necessário apenas 1 (um) bit. Essa informação é suficiente para representar cada pixel, ou seja, temos uma representação de 1 bit/pixel. Em alguns casos, temos uma informação extra sobre a cor de cada informação, a cor para o bit com valor 0 (zero) e a cor para o bit de valor 1 (um). Essa informação de cor é geralmente representada em 24 bits/cor no padrão RGB, podendo, porém ser representada de outras formas.

2.4 Vídeos e Gráficos Animados

As imagens e os gráficos podem ser apresentados na tela do computador como uma sucessão de imagens/gráficos que podem criar a sensação de movimento.

Quadro e Taxa de Quadro

Uma imagem ou gráfico individual de uma animação é chamado de quadro (ou *frame*). Para ser compreensível, os quadros que compõem a animação devem ser apresentados geralmente em uma taxa aproximadamente fixa. O número de quadros apresentados por segundo é definido como frequência de quadros e é medido em termos de quadros por segundo (fps – *frames per seconds*). A taxa deve ser alta suficiente para produzir a sensação de movimento. Para isto, taxas maiores ou iguais a 25 fps devem ser utilizadas. A tabela abaixo resume as principais frequências de quadro utilizadas atualmente.

Fps	Comentários
<10	Apresentação sucessiva de imagens
10 a 16	Impressão de movimento, mas com sensação de arrancos
>16	Efeito do movimento começa
24	Cinema
30/25	Padrão de TV americano/europeu
60	Padrão HDTV

Imagens Bitmap Animadas (Vídeo)

Na animação de imagens, cenas são registradas como uma sucessão de quadros representados por imagens bitmap possivelmente compactadas. Estas imagens podem ser capturadas da vida real com câmeras ou criadas através do computador. A primeira técnica produz o que é chamado de **vídeo**.

Animação de imagens tem as mesmas características que as imagens: falta de uma descrição semântica e necessidade de uma grande capacidade de armazenamento.

Gráficos Animados

O termo gráfico animado ou animação gráfica é utilizado para referenciar apresentação sucessiva de objetos visuais gerados pelo computador em uma taxa suficiente para dar a sensação de movimento e onde cada atualização é comutada de uma descrição abstrata em tempo de apresentação.

A principal vantagem das animações gráficas é que elas são mais compactas: elas são descritas por um conjunto de objetos com diretivas temporais (em outras palavras um programa a ser executado em tempo de apresentação). Outra vantagem é que animações gráficas são revisáveis. Existe uma desvantagem: é necessário um poder de processamento suficiente para apresentação.

Vídeos Híbridos

Técnicas avançadas, incluindo reconhecimento de padrões, permitem formas híbridas combinando vídeos e animações gráficas. Tais aplicações são suportadas por programas avançados que necessitam de unidades de processamento poderosas.

Um exemplo é quando imagens bitmap individuais providas por uma câmera de TV ao vivo ou videotape são analisados por programas de computadores e modificados de acordo com um critério predefinido. A apresentação de objetos reais ou pessoas pode ser modificada, ou em modo ao-vivo ou *off-line*.

2.5 Principais características e Requisitos das Informações multimídia

Esta seção resume as principais características das várias mídias apresentados neste capítulo e obtém a partir disto os principais requisitos de um sistema multimídia em rede.

2.5.1 Requisitos de armazenamento e largura de banda

Requisito de armazenamento é medido em termos de bytes ou Mbytes. No domínio digital, a largura de banda é medida como taxa de bits em bits/s ou Mbits/s. A unidade para armazenamento é byte e para largura de banda é bit.

Imagens

Para imagens, o requisito de armazenamento pode ser calculado a partir do número de pixels (H) em cada linha, o número de linhas (V) na imagem e o (P) número de bits por pixel, da seguinte forma: **Requisito de armazenamento = HVP/8**. Por exemplo, uma imagem com 480 linhas, 600 pixels cada linha e um número de bits por linha igual a 24 necessita 864 Kbytes para representar a imagem.

A largura de banda necessária para a transmissão da imagem pode ser calculada a partir do seu requisito de armazenamento. Por exemplo, se a mensagem acima (864 Kbytes) deve ser transmitida em 2 segundos, então a largura de banda necessária é 3,456 Mbits/s. Em muitas aplicações, imagens devem ser apresentadas em sincronia com mídias contínuas tal como áudio. Neste caso, a transmissão de imagem impõe tempo restrito e requisitos de largura de banda.

Áudios e Vídeos

Áudios e vídeos são mídias contínuas que são normalmente caracterizadas em bits/s ou Mbits/s. Para áudio este número é calculado baseado na taxa de amostragem e no número de bits por amostragem. Para vídeo este cálculo é baseado na quantidade de dados em cada quadro e no número de quadros por segundo. O resultado especifica a taxa de bits necessária ao canal de transmissão. Caso a duração do áudio ou do vídeo seja conhecida, o montante de requisito de armazenamento pode ser calculado.

A tabela que segue apresenta os requisitos de largura de banda de áudios e vídeos de diferentes qualidades.

Aplicações	Taxa de transmissão (Kbps)
CD-Audio	1.411,2
DAT	1.536
Telefone Digital	64
Radio digital, long play DAT	1.024
Vídeo de qualidade televisão	216.000
Vídeo de qualidade VHS	54.000
HDTV	864.000

A tabela abaixo mostra os requisitos de armazenamento para mídias estáticas e dinâmicas comuns de diferentes durações.

Aplicações	Requisitos de Armazenamento (MBytes)
Livro de 500 páginas	1
100 imagens monocromáticas	7
100 imagens coloridas	100
1h de áudio qual. telefone	28,8
1h de Áudio-CD	635
1h Vídeo qualidade VHS	24,3
1h TV	97000
1h HDTV	389000

2.5.2 Relações temporais e espaciais entre mídias

Em computação e comunicação multimídia, as diversas mídias estáticas e dinâmicas podem estar relacionadas em uma aplicação ou apresentação no domínio do tempo e do espaço. As relações espaciais são definidos no momento da criação da aplicação, e não existem muitos problemas tecnológicos associados.

O objetivo principal das aplicações multimídia é apresentar informações multimídia ao usuário de forma satisfatória, sendo que estas informações podem ser oriundas de fontes ao vivo, como câmeras de vídeo e microfones, ou originária de servidores distribuídos. Para obter uma boa qualidade, as relações temporais dos elementos de mídia devem ser mantidas durante a apresentação dos dados multimídia. Uma das principais problemáticas de sistemas multimídia é a **sincronização multimídia**, especialmente em sistemas distribuídos. Neste contexto, sincronização pode ser definida como o aparecimento (apresentação) temporal correto e desejado dos componentes multimídia de uma aplicação, e um esquema de sincronização define os mecanismos usados para obter a sincronização requerida.

O aparecimento temporal correto e desejado de componentes multimídia em uma aplicação tem três significados quando usado em diferentes situações [Lu, 96]:

- Quando usado para um fluxo contínuo, o aparecimento temporal correto e desejado significa que as amostras de um áudio e quadros de um vídeo devem ser apresentados em intervalos regulares. Este tipo de sincronização é chamado de **sincronização intramídia**.
- Quando usado para descrever os relacionamentos temporais entre componentes multimídia, o aparecimento temporal correto e desejado significa que os relacionamentos temporais desejados entre os componentes devem ser mantidos. Este tipo de sincronização é chamada de **sincronização intermídia**, que está relacionada com a manutenção das relações temporais entre os componentes envolvidos em uma aplicação.
- Quando usado em aplicações interativas, o aparecimento temporal correto e desejado significa que a resposta correta deveria ser fornecida em um tempo relativamente curto para obter uma

boa interação. Este tipo de sincronização é chamada de sincronização de interação, que está relacionada com a manutenção de que o correto evento (resposta) ocorra em um tempo relativamente curto.

Existem basicamente duas categorias de trabalhos em sincronização multimídia: trabalhos na área de especificação das relações temporais entre componentes; e trabalhos que buscam a definição de mecanismos para satisfazer as relações temporais especificadas..

2.5.3 Requisitos de atrasos e variações de atrasos (Jitter)

Para obter uma qualidade razoável na apresentação de áudios e vídeos, amostras de áudio e vídeo devem ser recebidas e apresentadas em intervalos regulares. Por exemplo, se uma peça de áudio é amostrada numa taxa de 8 kHz, ele deve ser apresentado a 8000 amostras por segundo. Como mídias contínuas têm essa dimensão temporal e os componentes do sistema podem atuar assincronamente, suas correções dependem não apenas dos valores das amostras, mas também do tempo de apresentação das amostras.

Atrasos fim-a-fim é a soma de todos os atrasos em todos os componentes de um sistema multimídia, incluindo acesso a disco, conversão A/D, codificação, processamento no hospedeiro, acesso a rede, transmissão, *buffering*, decodificação e conversão D/A. O atraso aceitável é muito subjetivo e é dependente de aplicação:

- Aplicações de conversações ao vivo necessitam a manutenção da natureza interativa, para tal o atraso não pode ser superior a 300ms.
- Para aplicações de recuperação de informação o requisito de atraso não é muito forte desde que o usuário não aguarde muito pela resposta. Em muitas aplicações o atraso de alguns segundos é tolerável.

Para mídias contínuas a variações de atrasos deve ser pequena. Para voz com qualidade de telefone e vídeo com qualidade de televisão a variação de atraso deve ser inferior a 10ms. No caso de áudio de alta qualidade esta variação deve ser muito pequena (<1ms) isto, pois nossa percepção do efeito estéreo é baseado nas diferenças de fase mínimas.

Tanto o atraso quanto a variação de atraso devem ser garantidos em toda a seção de comunicação. Isto não é suportado pelas redes, protocolos de transporte, sistemas operacionais usuais.

2.5.4 Tolerância a erros e perdas em dados multimídia

Diferentes dos dados alfanuméricos, onde perdas e erros na transmissão são na sua grande maioria intoleráveis. Erros ou perdas em dados de áudio, vídeo e imagens podem ser tolerados. Isto, pois estas perdas e erros de bits não são desastrosos e geralmente não são percebidos pelo usuário.

Para voz, nós podemos tolerar uma taxa de erros de bit de 10^{-2} . Para imagens e vídeos pode-se tolerar uma taxa de bits de 10^{-4} a 10^{-6} . Outro parâmetro que mede o erro é a taxa de perda de pacotes. Os requisitos de taxa de perdas de pacote são mais forte que a de erros de bit, porque uma perda de pacote pode afetar a decodificação de uma imagem por exemplo. Quando técnicas de compressão são utilizadas a taxa de erro de bits deve ser pequena, pois um erro de bit pode causar um erro de descompactação de muitos bits.

Técnicas de recobrimento de erros podem ser empregadas para aumentar a qualidade de áudio e vídeo.

2.5.5 Qualidade de Serviço

Como discutido na seção anterior, dados multimídia impõem duros e diversos requisitos a sistemas multimídia. Eles requerem largura de banda, capacidade de armazenamento alto e alta taxa de transferência, limitação de atraso e variação, sincronizações espaciais e temporais. O conceito de **Qualidade de Serviço** (QoS) é utilizado para especificar o conjunto de parâmetros de requisitos. Não há um acordo universal deste conjunto, mas parâmetros de requisitos mais comuns cobrem os requisitos mencionados acima, incluindo largura de banda (taxa de transferência), limitações de atrasos e variações e requisitos de sincronização. Estes parâmetros são especificados em duas grades: qualidade preferível e qualidade aceitável.