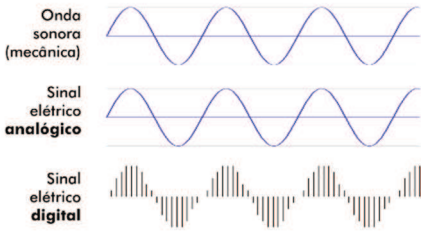
Capítulo 2

Dados Multimídia

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

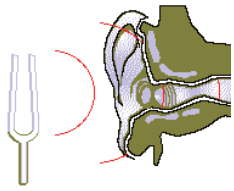
2.1 Representações de Áudio

Esta seção descreve as diversas representações de áudio, seja na forma do fenômeno físico, da onda sonora (mecânica) que se propaga, do sinal elétrico analógico gerado pelo microfone, e finalmente da representação digital do áudio.



2.1.1 Fenômeno: 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 alta e baixa pressão.



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 2, 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 **frequência** (f) é definida como o inverso do período e representa o número de períodos em um segundo. A frequência é normalmente medida em Hz (Hertz) ou ciclos por segundo (cps). A amplitude (A) do som 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 à posição da onda no tempo. Quando os componentes de frequência de som estão na faixa de 20 Hz a 20.000 Hz, o som é audível pelos humanos. A maioria dos sistemas multimídia trabalha com esta faixa de frequência. Usando as variáveis apresentadas anteriormente, uma onda senoidal pode ser representada no tempo por $s(t) = A \sin(2ft+\Phi)$.

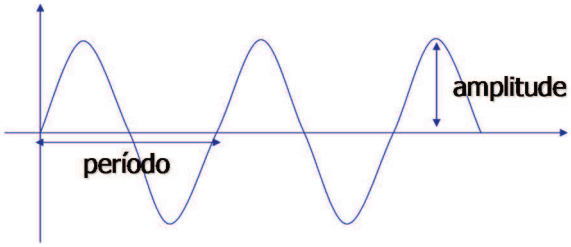


Figura 2. Forma de Onda

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 no tempo (valores dependentes do tempo). A Figura 3 apresenta graficamente diferentes formas de onda com diferentes valores de amplitude, frequência e fase.

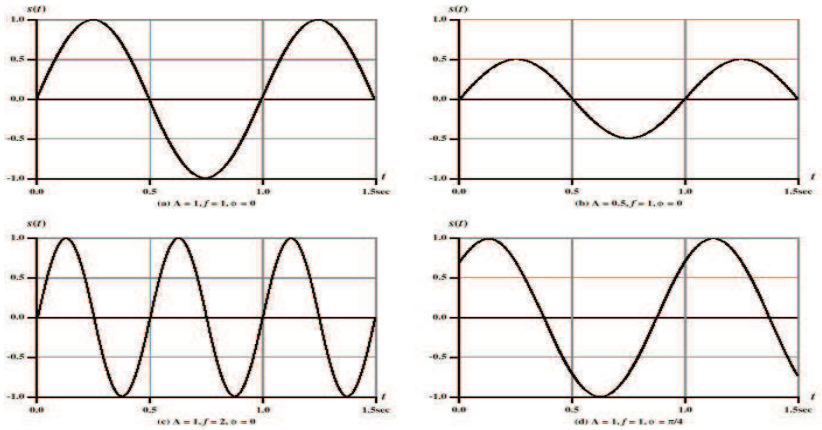


Figura 3. Diferentes formas de onda

2.1.2 Transformação da onda de pressão em sinal elétrico

É muito difícil manipular o som enquanto forma mecânica de energia, por isso é importante transformar onda sonora em outra forma de energia mais conveniente por meio de transdutores. A forma de energia mais adequada é a elétrica, ou seja, em um sinal de áudio. Nesta representação na forma de sinal elétrico, é mais simples de controlar, modificar e armazenar áudios. Esta conversão do áudio para um sinal elétrico contínuo (analógico) por um microfone, como ilustrado na Figura 4. Este sinal elétrico é medido normalmente em volts.

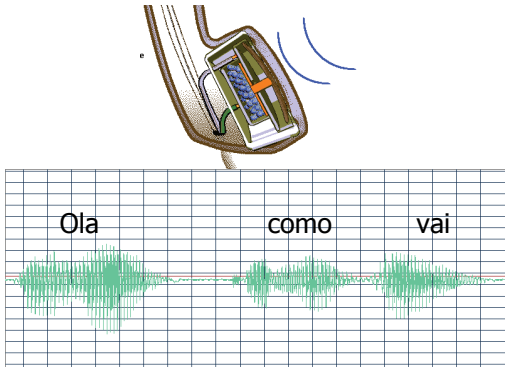
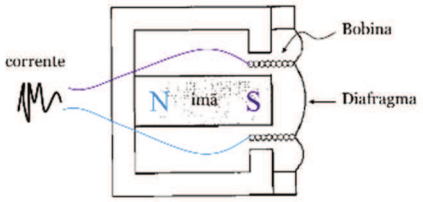
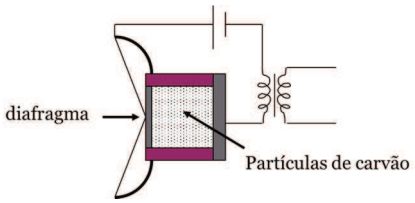


Figura 4. Conversão da onda sonora em sinal elétrico

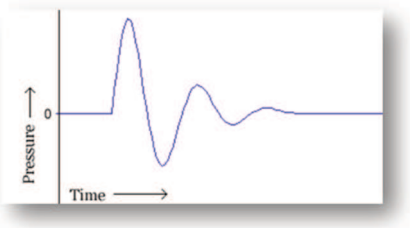
Existem diversos tipos de microfone: microfones dinâmicos (bobina móvel e fixa); microfones capacitivos (condensador); microfones a cristal e microfones cerâmicos; e microfones de carvão (telefone). No microfone dinâmico bobina móvel, a pressão do ar desloca o diafragma, e provoca o movimento da bobina, para dentro e fora do imã, provocando a alteração do campo magnético. Esta alteração induz uma corrente elétrica variável na bobina.



Por sua vez, no microfone a carvão (de telefone), a pressão do ar desloca o diafragma, que faz varia a densidade de partículas de carvão. Esta variação de densidade faz variar a resistência elétrica, que faz a corrente variar.

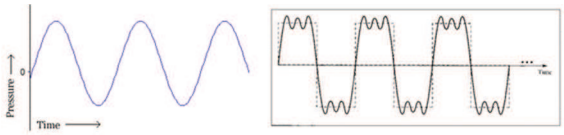


2.1.3 Sinal analógico do áudio



A definição tradicional de sinal é "uma grandeza que varia no tempo e/ou espaço. No caso do áudio, o sinal considera geralmente a dimensão tempo, considerando o microfone um referencial no espaço. Note que os sinais reais são analógicos, que variam continuamente no espaço e amplitude.

Os sinais podem ser classificados em simples ou compostos (). Os **sinais simples**, não podem ser decompostos em componentes (uma senoide). Um sinal de áudio simples (uma senoide), pode ser modelada pela função s(t) = A sin(2πft +Φ). São raros os objetos que produzem sons com frequência única (tons). Os sinais normalmente são compostos, formados por componentes de múltiplas frequências (diferentes sinais), chamados de **componentes de frequência do som.** Combinação das frequências geradas por instrumentos musicais é chamada de timbre.

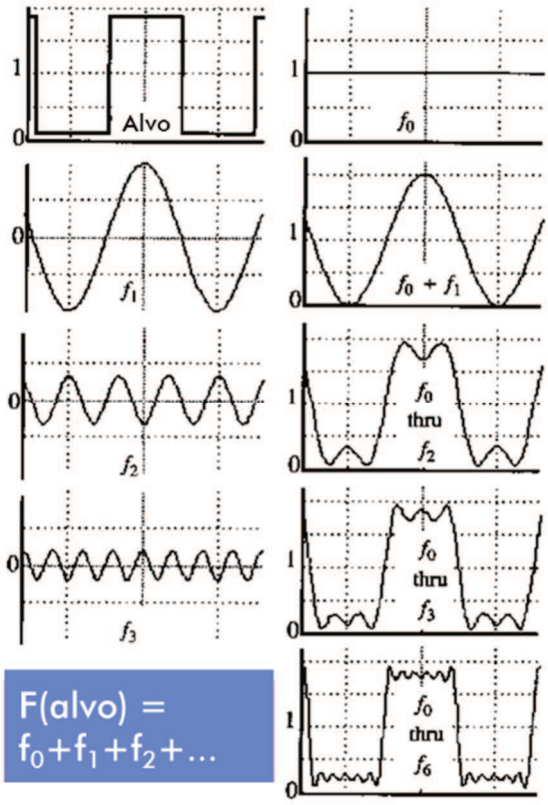


a) Sinal Simples b) Sinal Composto

Figura 5. Tipos de Sinais

Jean Baptiste Joseph Fourier (1768-1830) teve uma ideia (1807), de que Qualquer função periódica pode ser reescrita como uma soma ponderada de senos e cossenos de diferentes frequências. Muitos não acreditaram (Lagrange, Laplace, Poisson e outros), inclusive, esta teoria foi apenas foi traduzida para Inglês em 1878! Ela chama-se de Série de Fourier.

Para ilustrar a ideia de série de Fourier, considere o bloco de construção A sin(2πft +Φ), e o sinal alvo a onda quadrada (a direita). Somando-se (possivelmente soma infinita) várias senoides com diferentes amplitudes, frequências e fases, pode-se representar a onda quadrada.



Como a onda de som ocorre naturalmente, ela nunca é perfeitamente suave ou uniformemente periódica como a forma de onda da Figura 2. Os sinais sonoros normalmente são compostos, formados por múltiplas frequências (diferentes sinais). Os diferentes sinais são chamados de componentes de onda senoidal (componentes de frequência do som). A análise de Fourier se baseia na ideia de que qualquer sinal pode ser formado pela combinação de várias ondas ou componentes senoidais. É possível montar uma função baseada no domínio da frequência para representar os sinais. Por exemplo, a Figura 6c representa um sinal composto de duas senoides (a e b), sendo que a senoide (a), chamada componente de base, é definida por sin(2πft), e a segunda b), (1/3)sin(2π (3ft)) (com 1/3 da amplitude e 3 vezes a frequência da primeira senoide).

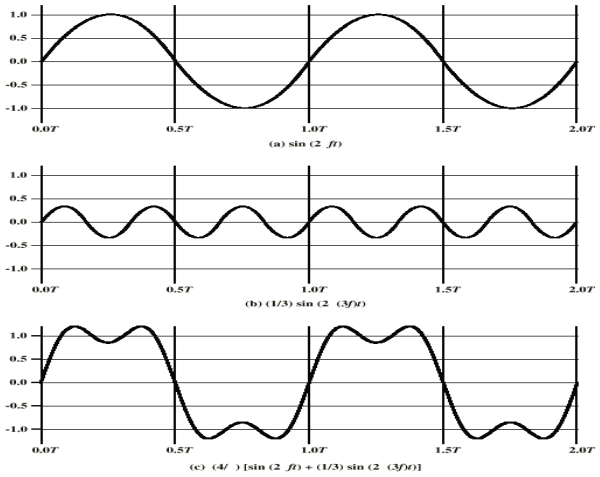


Figura 6. Sinal composto de duas componentes de frequência.

Muitas vezes é preferível representar o sinal no domínio da frequência, como ilustrado na Figura 7 para o caso do sinal da Figura 6c.

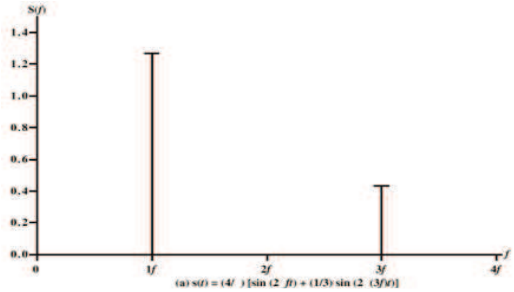


Figura 7. Representação no domínio da frequência.

2.1.4 Digitalização do Áudio

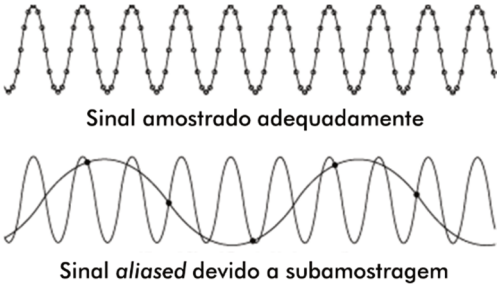
Para que sistemas computacionais processem e comuniquem sinais de áudio, o sinal elétrico deve ser convertido em um sinal digital. O mecanismo que converte o sinal de áudio digital em analógico é chamado de Conversor Analógico para Digital (CAD), ou digitalização. Digitalização aqui é o processo envolvido 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 8 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 8a. 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. Por exemplo, 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. Já a taxa de amostragem de CD de áudio 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.



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) com frequências acima de 22.050Hz é amostrado a 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 altas taxas de amostragens, 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.

Quantificação

O processo de converter valores de amostras contínuas em valores discretos é chamado de quantificação. Neste processo, o domínio do sinal é dividido em um número fixo de intervalos, chamados de **intervalos de quantificação**. Estes passos de quantificação são utilizados para medir os valores de amplitudes analógicos amostrados na etapa de amostragem. Quando o mesmo tamanho de passo de quantificação é usado na conversão A/D sem considerar a amplitude do sinal, o processo de conversão é dito uniforme Esta é a forma usada na ***modulação por pulso codificado*** (PCM - *Pulse Coded Modulation*). Na Figura 8c estes intervalos são numerados de 0 a 7. A cada amostra dentro de um intervalo é atribuído o valor do intervalo.

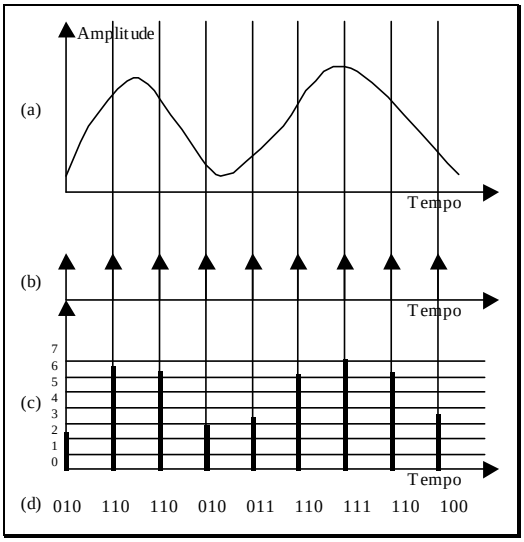


Figura 8. Conversão A/D [Lu, 96]: (a) sinal analógico; (b) pulsos de amostragem; (c) valores amostrados e intervalos de quantificação; (d) codificação das amostras

O tamanho do passo de quantificação (que na codificação vão definir o número de bits por amostra) define a qualidade da digitalização, pois de certa maneira define a precisão na medida das amostras. Quanto maior o passo (menor o número de bits), pior é a quantificação da amostra (a sua medida). Esta imprecisão na quantificação das amostras analógicas provocam o chamado **erro de quantificação**. Este

erro se traduz auditivamente por um ruído, ouvido na reprodução do som reconstruído (ruído de quantização). Este efeito pode ser visto na Figura 9, que apresenta uma senoide sendo digitalizada (Figura 9a), que passa pela etapa da amostragem (Figura 9b) para capturar amplitudes analógicas do sinal. Na fase de quantificação (Figura 9c), os valores analógicos de amplitude devem ser discretizados, e quanto maior o passo de quantificação, menor é a resolução da medida da amplitude, e assim ocasionando o ruído.

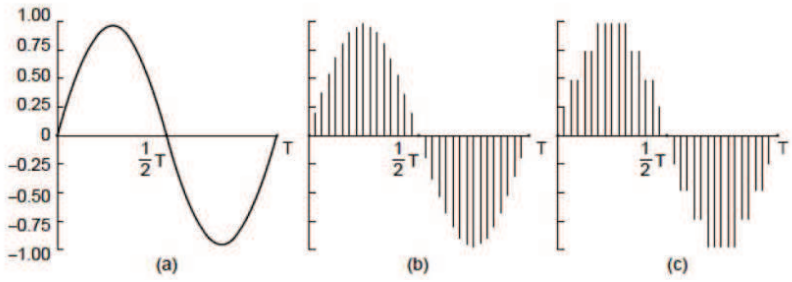


Figura 9. Erro de quantificação

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 logaritmicamente 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 realiza 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 bits.

Áudio na telefonia

Na área da telefonia digital, utiliza-se um método de transformação de natureza logarítmica para comprimir áudios. Ele mapea 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 necessárias, 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 8000x8 = 64 kbps.

2.1.5 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, melhor 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 de áudio, 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

2.1.6 Apresentação do áudio

Em sistemas multimídia, todas as informações multimídia são representadas internamente no formato digital. Contudo, os 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).

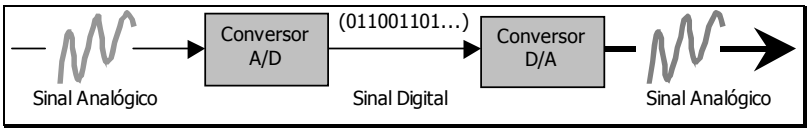
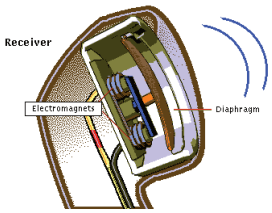


Figura 10. Conversão analógico/digital e digital/analógica

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.

2.1.7 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 distorções 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.

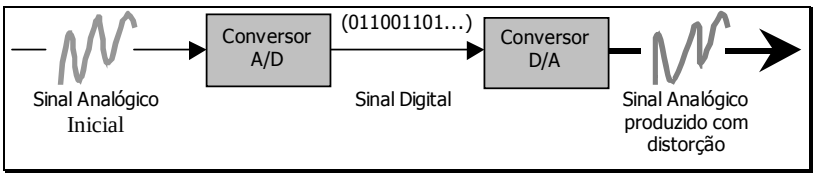


Figura 11. Conversão analógico/digital e digital/analógica

Aumentando a taxa de amostragem e o número de bits usado para codificação reduz estas distorções. 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 banda[1] limitadas.

Na maior parte dos sistemas multimídia, os usuários finais das informações multimídia são os humanos. Como nem todos os 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.

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.

2.1.8 Representação simbólica da música: o padrão MIDI

Como visto anteriormente, qualquer som pode ser representado como um sinal de som digitalizado, que é uma sequência de amostras, cada uma codificada por dígitos binários. Esta sequência 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 sequência de bits representa, por exemplo, uma fala ou música, e se música que notas são usadas e por quais 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 sequências 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).

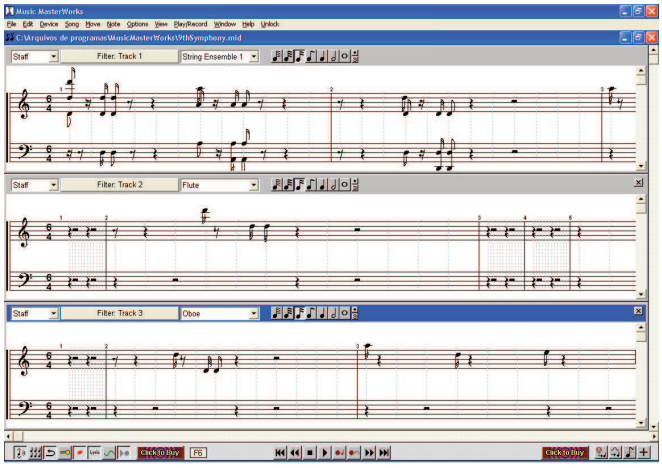


Figura 12. Editor de Midi

[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.

2.2 Representações de Imagens e Vídeos

Esta seção apresenta diferentes "representações" de imagens e vídeo, iniciando pela definição de imagem, de como o sistema visual humano percebe as imagens, e finalmente a representação analógica e digital de um sinal de imagem/vídeo.

2.2.1 Imagem é Luz

Imagens que percebemos do mundo real são na realidade luzes, e portanto, são ondas/radiações eletromagnéticas, vista na Figura 13. As duas principais grandezas físicas de uma onda eletromagnética são: intensidade (ou amplitude); a frequência, ou comprimento de onda (a distância entre valores repetidos sucessivos em um padrão de onda); e a polarização, relacionada com o direcionamento da luz. Na luz natural (não polarizada) o campo elétrico oscila aleatoriamente em todas as direções possíveis.

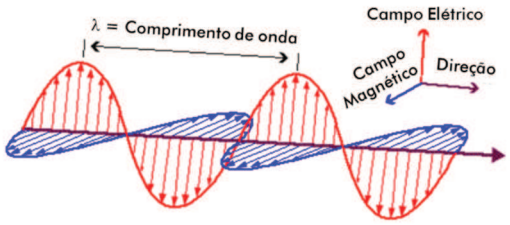


Figura 13. Luz: Radiação Eletromagnética

A luz visível pelos humanos são aquelas cujos comprimentos de onda estão entre 400 e 700 nm, o chamado espectro visível (Figura 14).

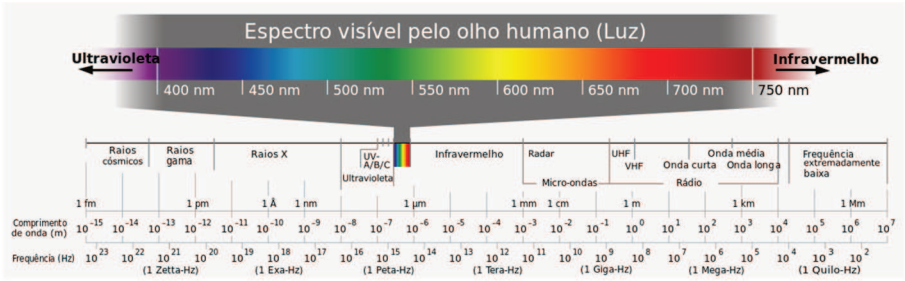
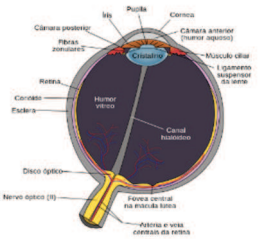


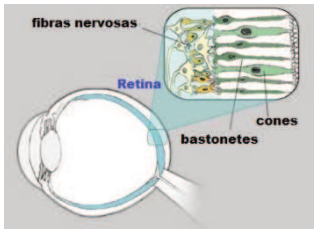
Figura 14. Espectro Visível

2.2.2 Sistema Visual Humano



O sistema visual humano é formado pelo sistema ótico (olho) e o processamento e reconhecimento realizado pelo nosso cérebro. O sistema ótico (olho) possui focagem flexível (realizada pela córnea e lente). A luz que atravessa a córnea, o humor aquoso, o cristalino e o humor vítreo e se dirige para a retina. A retina funciona como o filme fotográfico em posição invertida. O nervo óptico transmite o impulso nervoso provocado pelos raios luminosos ao cérebro que o interpreta e nos permite ver os objetos nas posições em que realmente se encontram. Nosso cérebro reúne em uma só imagem os impulsos nervosos provenientes dos dois olhos.

A nossa retina possui, entre outros, cones e bastonetes. Os bastonetes mede a intensidade da luz (luminosidade). Possuímos entre 75 a 150 milhões de bastonetes na retina, sendo que vários são ligados a um nervo, produzindo baixa definição. Os bastonetes são úteis para detectar movimentos e para visualização em baixa luminosidade (percepção de sombras). Por sua vez, os cones medem frequência da luz (cor), e são em média de 6 a 7 milhões. Eles possuem grande definição (nervo único), mas para serem ativados exigem maior luminosidade que os bastonetes.



Existem três tipos de cones, cada um especializado em comprimentos de luz curtos (S), médios (M) ou longos (L), chamados de cone azul, verde e vermelho. Eles definem o espectro de frequência visível (400nm a 700nm).

2.2.3 Representação analógica de imagens e vídeos

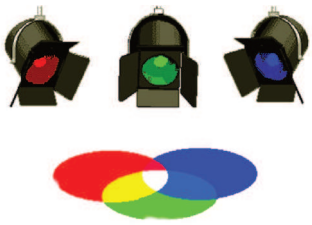
Descrevendo imagens monocromáticas com variáveis físicas

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, a luz refletida possui 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).

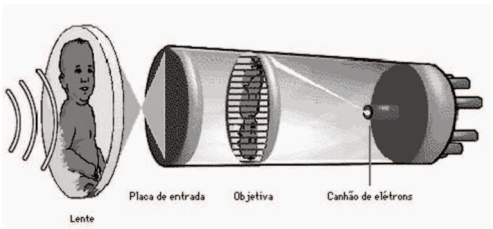


A teoria da cor foi desenvolvida por Thomas Young no início de 1802 e afirma que qualquer sensação de cor pode ser reproduzida pela mistura em proporções apropriadas de três luzes coloridas monocromáticas primárias. Esta é a teoria *Tristimulus*. Cores primárias são independentes no sentido que uma cor 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.

2.2.4 Captura e reprodução de imagens e vídeos

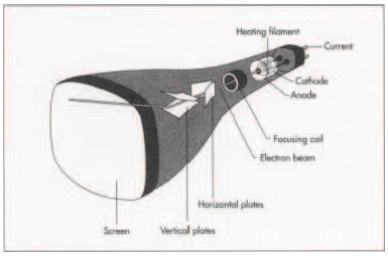
Captura e reprodução de imagens e vídeos monocromáticos

As imagens são capturadas usando câmeras da seguinte maneira: as lentes da câmera focam uma imagem de uma cena em uma superfície fotossensível de sensores CCD (*Charge-Coupled Device*); o brilho de cada ponto é convertido em uma carga elétrica por uma camada fotossensível, estas cargas são proporcionais ao brilho nos pontos; a superfície fotossensível 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.

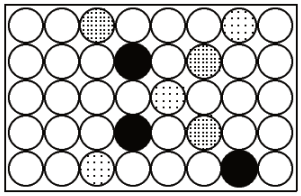


Nesta seção, por simplificação assume-se a captura e reprodução de vídeos monocromáticos, onde apenas um sinal de luminância é produzido (apenas a luminosidade é capturada, temos a imagem em tons de cinza). Neste caso são usadas **câmeras de luminância**, que captam a imagem em tons de cinza, e gera 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.



Um dos antigos dispositivos de apresentação de imagens 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. Essas 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 crominância (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 15. 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.

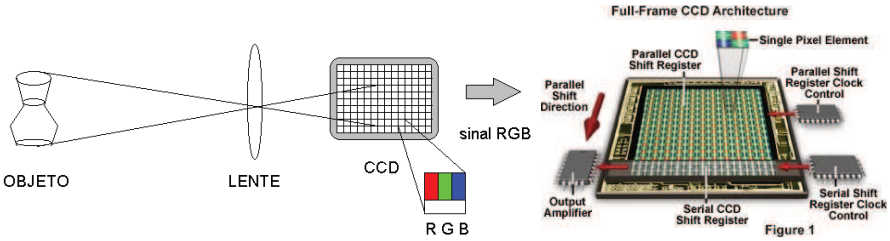


Figura 15. Câmera de crominância (1 passo - 1 CCD)

- Câmera de crominância (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 16. 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.

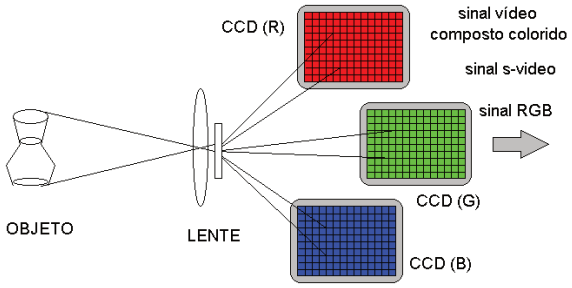


Figura 16. Câmera de crominância (1 passo - 3 CCD)

- Câmera de crominância (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 17). A digitalização da imagem então é feita em 3 passos, ou seja, para cada filtro é feito 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.

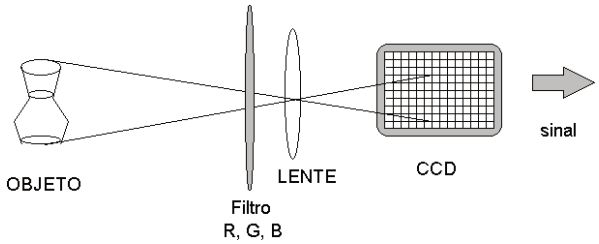
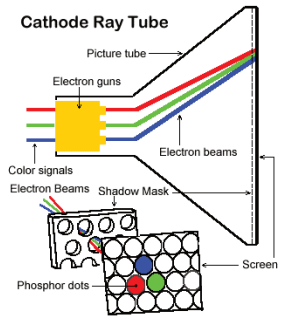


Figura 17. Câmera de crominância (3 passos - 1 CCD)

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; assim 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 120x160 até algo em torno de 4200x2690 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 18. 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.

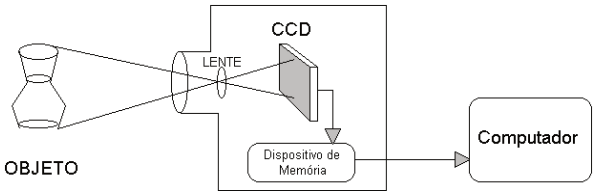


Figura 18. Câmera fotográfica digital

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 a um elemento de digitalização de linha (Figura 19). 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).

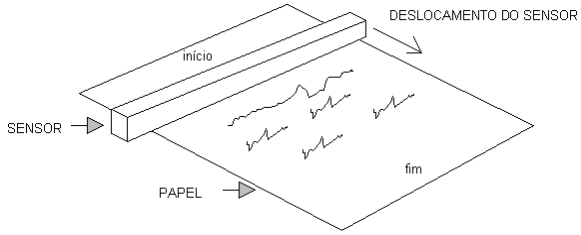


Figura 19. Esquema de funcionamento do scanner

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. Este tipo de scanner 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.

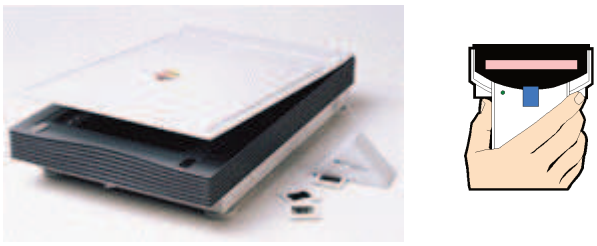


Figura 20. Exemplos de scanners

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

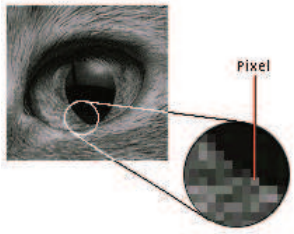
- Resolução óptica: É a resolução via hardware que o scanner pode atingir. Outras resoluções acima desta podem ser obtidas com técnicas de interpolação, mas com perda de qualidade, ou seja, a máxima resolução obtida com toda a qualidade é a sua resolução óptica. A escolha da resolução de um scanner depende 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.
- Qualidade do sensor: Esta qualidade diz respeito à 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 retém 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 pixeis. 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 pixeis (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 x 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 pixeis por unidade linear de medida. Em geral se utiliza o número de pixeis 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, melhor 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. Em uma representação a 8 bits/pixel, pode-se usar 3 bits para quantificar os componentes R e G, e dois 2 bits para o componente B (3-3-2). A quantificação do componente B com menos bits é justificada pois temos menos sensibilidade ao componente azul.

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 x 16 x 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 de uma tabela de cores, a **paleta de cores**, que contém as informações sobre as cores (Figura 21). 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.

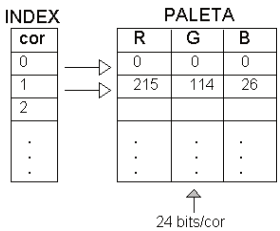


Figura 21. Índice e paleta de cores

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 é 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 Representação de Caracteres

A escrita é a forma mais adequada para transferir informações essenciais de maneira precisa. Sendo que palavras e símbolos, falados ou escritos, são a forma mais comum de comunicação. É a escrita, a principal forma de comunicação assíncrona (defasada no tempo), ou quase tempo-real (como mensagens instantâneas) entre pessoas. E escrita

Esta seção apresenta os principais conceitos relacionados à representação de caracteres.

2.5.1 Tipos possíveis de texto

Os textos podem ser basicamente de dois tipos:

- Texto não formatado (*plain text*), onde o número de caracteres disponíveis é limitado, e os caracteres possuem uma representação simples (dimensão dos caracteres é fixa e não permite diferentes fontes ou estilos);
- Texto formatado (rich text), onde aparência do texto é mais rica, com várias fontes, cores, estilos e dimensões. Estes textos são produzidos por processadores de texto; e,
- Hipertexto, que são textos ao qual se adicionam hiperligações originando texto não linear, permitindo a navegação entre documentos de texto.

2.5.2 Natureza dupla do texto: conteúdo léxico e aparência

Os textos têm uma natureza dupla:

- **Conteúdo léxico**; conteúdo entende-se os caracteres que constituem as palavras e outras unidades de pontuação ou simbólicas, sendo a parte do texto que transmite o seu significado (sua semântica). Não importa a aparência dos caracteres para o entendimento da semântica. Ignorando-se a aparência dos caracteres, temos os caracteres abstratos.
- **Aparência**, que é definida por atributos visuais dos caracteres (fonte, tamanho, disposição na tela, etc.). A representação visual de um caractere denomina-se Glifo. Por exemplo, o caractere abstrato "A" pode ter uma infinidade de representações gráficas, incluindo "𝔸", "𝐀", "𝘈" , "a", "𝖆", "𝒶".

Os caracteres abstratos são caracteres vistos apenas quanto a sua natureza léxica, e são agrupados em alfabetos. Cada idioma ou grupo de idiomas usa um alfabeto.

2.5.3 Conjunto de Caracteres

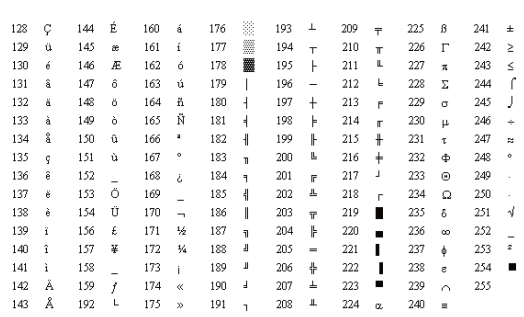
Conjunto de caracteres são tabelas mantidas pelo sistema operacional que consistem em uma correspondência entre os códigos e os caracteres. Eles contém representações de grafemas (unidades fundamentais de um sistema de escrita) ou unidades similares a grafemas (incluindo maiúsculas, minúsculas, sinais de pontuação, números e símbolos matemáticos).

A adoção de conjunto de caracteres na informática trouxe várias vantagens. Primeiro, é vital guardar os caracteres na forma de códigos, permitindo sua edição (alteração) e pesquisa de texto. Esta codificação de caracteres permite, por exemplo, a comparação de caracteres e a associação das teclas do teclados a representação destes caracteres. Por exemplo, quando se pressiona um A no teclado, esse caractere é procurado na tabela de caracteres para depois ser apresentado no monitor.

Nesta área, a normalização é o mais importante, pois os códigos universais podem facilmente ser trocados entre máquinas diferentes e que usam sistemas operacionais diferentes. O primeiro conjunto de caracteres normalizado (isto em 1968) foi o ASCII (*American Standard Code for Information Interchange*). Ele foi criado considerando-se a língua inglesa, onde definiu-se que um conjunto de 128 caracteres era suficiente. Para isso, a representação ASCII adotou 7 bits para codificar cada caracteres (resultando nos 2^7 valores).

Bits 3210	*654 000*	*001*	*010*	*011*	*100*	*101*	*110*	*111*
0000	NUL	DLE	SP	0	@	P	\	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	´	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	I	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	‘	<	L	\	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

Logo percebeu-se que 7 bits não era suficiente para representar outros idiomas. Tentando resolver isso, surgiu a ISSO 8859, que normaliza conjuntos de caracteres de 8 bits. Esta norma é dividida em 10 partes. Particularmente importante para o português, a ISO 8859-1 (ISO Latin1), define caracteres utilizados na maioria dos países da Europa Ocidental, primeiros 128 caracteres são os mesmos do ASCII de 7 bits, os restantes 128 são códigos para os idiomas europeus. Por sua vez, a ISO 8859-2 define a ISO Latin2, para outros idiomas da Europa Oriental (Checo, Eslovaco, Croata).



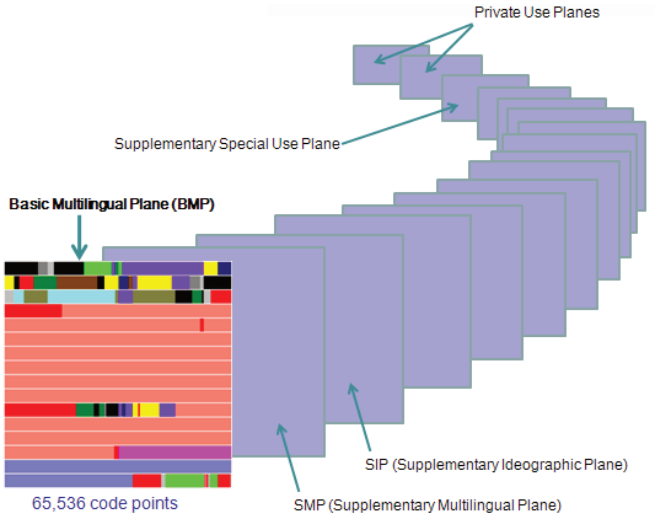
A opção pelas variantes ISO 8859 acaba por não conseguir resolver bem o problema. Obviamente, 7+1 bits são claramente insuficientes para representar todas as línguas (Chinês, japonês etc.). Além disso, tem-se os textos multilíngue, onde a codificação deveria permitir a existência de línguas diferentes no mesmo arquivo. Uma tentativa para resolver o problema foi a SO 10646 (32 bits) de 1991, que permite representar 4.294.967.296 caracteres diferentes (2^{32}). A desvantagem desta última é o espaço ocupado pelo texto, quando se passou de 8 bits para 32 bits na representação dos caracteres.

Mais recentemente, um consórcio de empresas (incluindo Adobe, Apple e Microsoft) definiram o Unicode. Esta representação é amplamente adotada atualmente, incluindo as linguagens HTML, XML e Java. Trata-se de um padrão que permite aos computadores representar e manipular, de forma consistente, texto de qualquer sistema de escrita existente. O Unicode foi desenvolvido em conjunto com um Conjunto Universal de Caracteres (UCS – *Universal Character Set*).

O Unicode consiste de:

- um repertório de mais que 100.000 caracteres cobrindo 100 scripts (coleção de letras e outros signos escritos usado para representar uma informação textual em um ou mais sistemas de escritas),
- uma metodologia para codificação,
- um conjunto de codificações padrões de caracteres,
- uma enumeração de propriedades de caracteres (como caixa alta e caixa baixa),
- um conjunto de arquivos de computador com dados de referência, e
- regras para normalização, decomposição, ordenação alfabética e renderização.

O Unicode codifica caracteres em um espaço numérico entre 0 a 10FFFF, sendo que este espaço é dividido em 17 planos (numerados de 0 a 16). O plano BMP (*Basic Multilingual Plane*) tem um espaço numérico de 0 a FFFF, sendo os valores dos códigos são anotados da seguinte forma: U+aaaa. U+ se refere a valores de código Unicode, e aaaa representa um número de quatro dígitos hexadecimais de um caractere codificado. Por exemplo, LETRA MAIÚSCULA LATINA A é codificado por U+0041



Existem alguns formatos de codificação dos valores Unicode, que são: UTF-8, UTF-16 e UTF-32. O UTF-8 é uma forma de codificação de tamanho variável, requer de um a quatro bytes para expressar cada caractere Unicode. Nesta codificação, "A" é 41 (mesmo que no ASCII!), α é CE 91, Katakana "A" (ア) é E3 82 A2, e Gothic Ahsa (𐌰) é F0 90 8C B0.

2.5.4 Fontes e Faces

Muitas pessoas confundem os termos fontes e faces. Uma Face é uma família de caracteres gráficos que normalmente inclui muitos tamanhos e estilos de tipos. São exemplos de faces: Arial, Times New Roman e `Courier New`. Já uma fonte é um conjunto de caracteres de um único tamanho e estilo pertencente a uma família de face particular. Um exemplo de fonte é *Times New Roman 12 pontos itálico*. As fontes digitais são versões das fontes tradicionais (algumas do século XV). As fontes podem ser vistas como tabelas de correspondência entre os caracteres abstratos e a sua representação gráfica (grifo).

As fontes têm duas possibilidades de armazenamento. Elas podem ser armazenadas em arquivos e instalados no sistema operacional (exemplo C:\Windows\Fonts) e assim podem ser compartilhadas por todos os arquivos e todas as aplicações. Se por acaso elas são requeridas e não existem, elas devem ser trocadas por fontes alternativas. Outra forma de armazenamento de fontes é embutidas nos próprios arquivos de texto. A vantagem desta última para o designer de uma aplicação multimídia pois é livre de usar qualquer fonte no seu trabalho. Mas elas não podem ser compartilhadas entre documentos que usam as mesmas fontes.

Na definição das fontes, os tamanhos geralmente são expressos em pontos, sendo que um ponto corresponde a 0,0138 polegadas ou aproximadamente 1/72 de uma polegada. Os estilos os estilos normais das fontes são regular, negrito, itálico (oblíquo) e sublinhado. Outros atributos, como contorno de caracteres, podem ser adicionados pelo programa.

2.6 Principais Requisitos das Informações multimídia

Esta seção resume os principais requisitos das várias mídias apresentadas neste capítulo, em termos de taxa de bits, requisitos de armazenamento e sincronismo.

2.6.1 Requisitos de armazenamento e largura de banda

Requisito de armazenamento é medido em termos de Bytes (B), KBytes (KB – 1024 B), Mbytes e assim por diante. A taxa de bits é convencionalmente medida em bits/s (bps), kbps, Mbps (1000 bps), Gbps, e assim por diante. Em suma, 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 pixeis (H) em cada linha, o número de linhas (V) na imagem e o (P) número de bits por pixel, da seguinte forma:

Espaço_ocupado_imagem = HVP/8

Por exemplo, uma imagem com 480 linhas, 600 pixeis cada linha e um número de bits por linha igual a 24 necessita 846,75 KB (480*600*24/8) para representar a imagem.

A taxa de bits necessária para a transmissão da imagem pode ser calculada a partir do do seu espaço ocupado e o tempo limite para sua transferência, da seguinte forma

Taxa_de_bits_imagem = *Espaço_ocupado_imagem/tempo_limite*

Por exemplo, se a mensagem acima (864 KB) deve ser transmitida em 2 segundos, então a taxa de bits necessária é 3,456 Mbps (480*600*24*8/2). Como será visto mais adiante, em muitas aplicações, imagens devem ser apresentadas em sincronia com mídias continuas, tal como áudio. Nestes casos, a transmissão de imagem impõe tempo restrito e requisitos de taxa de bits.

Áudios

Para áudio, a taxa de bits determinada pelo número de canais, taxa de amostragem e número de bits por amostragem, da seguinte forma:

*Taxa_de_bits_áudio = número_de_canais*taxa_de_amostragem*bits_por_amostra*

A tabela que segue apresenta os requisitos de taxa de bits de áudios e vídeos de diferentes qualidades.

Qualidade	Número de canais	Taxa de amostragem	Bits por amostra	Taxa de transmissão (Kbps)
Telefone Digital	1	8000	8	64
CD-Audio	2	44100	16	1.411,2
DAT	2	48000	16	1.536
Radio digital	2	32000	16	1.024

O espaço ocupado por um áudio depende de sua taxa de bits e duração, e pode ser calculado utilizando a seguinte equação:

*Espaço_ocupado_audio = Taxa_de_bits_áudio * duração/8*

Por exemplo, 1 minuto de áudio qualidade telefone ocupa um espaço de 64000*60/8 = 468,75 KB. A divisão por 8 é necessária para conversão de bits em bytes.

Vídeos

A taxa de bits gerada por um vídeo é determinada com base na quantidade de dados em cada quadro, que é o espaço ocupado por uma imagem) e pelo número de quadros por segundo, utilizando a seguinte equação:

*Taxa_de_bits_vídeo = HPV*taxa_de_quadros*

Por exemplo, um vídeo de resolução 720x480 e de 24 bits/pixel gera uma taxa de 249 Mbps (720*480*24*30).

A tabela a seguir apresenta as taxas geradas por diversos tipos de qualidade de vídeos. Notem que o número de bits e taxa de quadros (para HDTV) podem variar. Os números apresentados são opções dependendo até da transmissão.

Qualidade	Resolução	Bits por píxel	Taxa de quados	Taxa de transmissão (Mbps)
DVD (PAL 4x3)	720x576	24	30	298,6
SDTV (HDMI 1.3)	704x480	48	30	486,6
HDTV (HDMI 1.3)	1920x1080	48	30	2.986

O espaço ocupado por um vídeo é determinado pela mesma forma das outras mídias, sendo que

Espaço_ocupado_vídeo = Taxa_de_bits_vídeo * Duração/8.

Por exemplo, o espaço ocupado pelo vídeo do exemplo anterior (249 Mbps) com duração de 1 minuto é de 1,8GB (249*60/8)

2.6.2 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 definidas 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. A sincronização labial é o tipo de sincronização intermídia mais conhecido, que é a sincronização entre o movimento do lábio e a voz gerada no instante deste movimento.
- 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..

2.6.3 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.

O **atraso 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.

2.6.4 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.