PUC

ISSN 0103-9741

Monografias em Ciência da Computação
n° 01/07

Fundamentos de Sistemas Multimídia

Part 1 – Aquisição, Codificação e Exibição de Dados

Luiz Fernando Gomes Soares

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22451-900

RIO DE JANEIRO - BRASIL

Monografias em Ciência da Computação, No. 01/07 ISSN: 0103-9741
Editor: Prof. Carlos José Pereira de Lucena Janeiro de 2007

Fundamentos de Sistemas Multimídia

Part 1 – Aquisição, Codificação e Exibição de Dados

Luiz Fernando Gomes Soares

Laboratório TeleMídia DI – PUC-Rio
Rua Marquês de São Vicente, 225, Rio de Janeiro, RJ - 22451-900.

lfgs@inf.puc-rio.br

Resumo. *Esta monografia compõe a primeira parte do material didático do curso de Fundamentos de Multimídia. Ela diz respeito à aquisição, codificação e exibição de dados, nas mais diversas mídias de representação.*

Palavras chave*: codificação digital, compressão, compactação, padrões de codificação.*

Fundamentos de Sistemas Multimídia

Part 1 – Aquisição, Codificação e Exibição de Dados

Laboratório TeleMídia

Departamento de Informática

Pontifícia Universidade Católica do Rio de Janeiro

Rua Marquês de São Vicente, 225, Prédio ITS - Gávea

22451-900 – Rio de Janeiro – RJ – Brasil

http://www.telemidia.puc-rio.br

Table of Contents

Fundamentos de Sistemas Multimídia

Part 1 – Aquisição, Codificação e Exibição de Dados

Luiz Fernando Gomes Soares

Laboratório TeleMídia DI – PUC-Rio
Rua Marquês de São Vicente, 225, Rio de Janeiro, RJ - 22451-900.

lfgs@inf.puc-rio.br

Resumo. *Esta monografia compõe a primeira parte do material didático do curso de Fundamentos de Multimídia. Ela diz respeito à aquisição, codificação e exibição de dados, nas mais diversas mídias de representação.*

Palavras chave*: codificação digital, compressão, compactação, padrões de codificação.*

1. Introdução

Os sistemas de computação (sistemas de processamento, sistemas operacionais, redes de comunicação etc.) foram desenvolvidos, originalmente, para dar suporte ao processamento e comunicação de dados textuais. Com a evolução tecnológica, não só as redes aumentaram em muito seu desempenho, mas também cresceu muito a capacidade de processamento e armazenamento das estações de trabalho. Isso tornou possível o desenvolvimento de sistemas para processamento e comunicação de informações representadas em várias outras mídias além da textual, como áudio, vídeo etc. O fenômeno da convergência é, em parte, marcado por essa mistura de diferentes tipos de mídia em sistemas integrados de transmissão e processamento de informação. Nesta monografia, apresentaremos os principais conceitos envolvidos na codificação digital dessas informações e na sua comunicação.

2. Informação e Sinal

Os seres humanos adquirem informação através de seus sentidos: visão, audição, tato, olfato e paladar. Esses sentidos são denominados *mídias*[1] *de percepção*. As informações adquiridas são então codificadas em estruturas de dados que denominamos *mídias de representação*, ou simplificadamente *mídias*. São exemplos de mídias de representação as mídias texto, gráfico, áudio e vídeo. Note que não existe uma correspondência biunívoca entre as mídias de percepção e de representação e que ainda é, no mínimo, pouco usual a utilização de mídias representando informações adquiridas pelo olfato, tato e paladar, embora já existam estudos nesse sentido.

Definimos *sinais* como ondas que se propagam através de algum meio físico possuindo, por exemplo, uma amplitude que varia ao longo do tempo, correspondendo à codificação da

[1] Em português, mídia vem da palavra latina medius (de onde derivou a palavra anglo-saxônica medium), e seu plural é mídias (correspondendo à palavra anglo-saxônica media).

informação transmitida. Um sinal pode ser distorcido durante sua transmissão, por terem suas componentes de freqüência atenuações diferentes devido a limitações do meio de transmissão. Pode-se mesmo ter perda ou deformação de parte do sinal por ruídos. Ao transmitir informações esperamos, no entanto, preservar seu significado e recuperar seu entendimento.

Podemos então introduzir informalmente o conceito de *qualidade de sinal* e *qualidade da informação* transmitida, através de um exemplo. Um vídeo transmitido a 30 quadros por segundo (padrão de TV), certamente tem uma qualidade de sinal melhor do que se fosse transmitido a 10 quadros por segundo (imagine, por exemplo, que a cada três quadros dois são perdidos). Se estivéssemos filmando uma paisagem sem qualquer movimento, a qualidade da informação transmitida seria a mesma (nesse caso extremo bastaria até mesmo a transmissão de um único quadro), independente do sinal. Se, no entanto, o vídeo tivesse muito movimento, a qualidade da informação transmitida não seria boa a 10 quadros por segundo e a sensação obtida seria a de várias imagens com movimentos bruscos, sem naturalidade.

Do exemplo anterior, podemos notar que um sinal pode carregar muita informação redundante. Técnicas para redução dessa redundância, denominadas técnicas de compressão e de compactação, podem ser empregadas. Sobre elas, teremos muito o que discutir ao longo desta monografia. Vamos, no entanto, primeiramente, aprofundar um pouco mais a discussão sobre informações e sinais, analógicos e digitais.

Inicialmente os computadores estavam restritos ao processamento e comunicação de dois tipos de dados — palavras e números. Códigos para números (binários, BCD, ponto flutuante IEEE etc.) estão hoje padronizados e estabilizados. Códigos para caracteres alfanuméricos (ASCII, EBCDIC etc.) são também amplamente aceitos. Enfim, a mídia textual é hoje razoavelmente bem entendida como codificação digital.

A mídia gráfica foi a segunda mídia a ganhar representação nos computadores digitais. Ela possui dois formatos: o vetorial e o matricial. O formato vetorial é bastante utilizado em modelagem geométrica e nele as figuras são representadas por um conjunto de segmentos de reta ou curvas, dados pelas coordenadas de seus pontos e pelos atributos das linhas que os unem. Imagens no formato matricial são usualmente chamadas de imagens estáticas. Nesse formato, as imagens são divididas em pequenas regiões, chamadas de elementos de fotografia, ou pixels (picture elements, muitas vezes também chamados de pels). Para cada uma dessas regiões guarda-se sua informação codificada de cor. Quanto maior o número de bits para codificar a cor, mais cores pode-se codificar e mais próximo pode-se chegar da cor original. Temos assim uma matriz de M linhas e N colunas, onde cada elemento representa um dos MxN pixels que compõe a imagem. Na reprodução da imagem, os pixels são reconstruídos utilizando-se a informação de cor armazenada na matriz. Quanto menor for o tamanho do pixel, mais fiel será sua coloração com relação à original, mas maior será a matriz da imagem. Ao tamanho da matriz dá-se o nome de *resolução geométrica* da imagem. A quantidade de bits utilizados para armazenar a cor de um pixel chama-se *resolução de cor* da imagem.

Informações na mídia textual e gráfica são originalmente digitais. Por isso, muitas vezes essas mídias são referidas como *mídias discretas*. Já informações geradas por fontes sonoras e de vídeo apresentam variações contínuas de amplitude, constituindo-se no tipo de informação que comumente é percebida pelos sentidos humanos através de sinais que denominamos analógicos. Devido a isso, as mídias de vídeo e áudio são usualmente referidas como *mídias contínuas*.

É importante que se entenda que qualquer tipo de informação (seja analógica ou digital) pode ser codificada em uma estrutura de dados (mídia de representação) digital, e essa codificação digital pode ser transmitida em um sinal analógico ou digital, como veremos.

A codificação digital de informações tem, em geral, vantagens em uma comunicação. Isso se deve, principalmente, à possibilidade de podermos restaurar, no receptor, a informação codificada original, mesmo na presença de distorções, falhas ou ruídos no sistema de transmissão.

3. Conversão de Sinais

Para utilizarmos as vantagens da codificação digital, devemos transformar as mídias contínuas de áudio e vídeo, normalmente adquiridas através de sinais analógicos. A essa transformação chamamos de conversão analógica digital, ou conversão A/D.

Uma vez processados e transmitidos, sinais digitais[2] podem ter de ser transformados em analógicos para percepção pelos sentidos humanos. A essa transformação chamamos de conversão digital analógica, ou simplesmente conversão D/A.

3.1. Conversão A/D

O teorema de Nyquist assegura que uma taxa de amostragem de $2W$ vezes por segundo é o suficiente para recuperar um sinal com banda passante W Hz. Isso quer dizer que, de um sinal analógico, basta se guardar os valores das amplitudes de suas amostras tomadas a intervalos regulares de $1/2W$ segundos para que se possa ter toda a informação necessária para reconstruí-lo integralmente. O processo de amostrar e guardar os valores dessas amostras, ilustrado na Figura 1, é conhecido como *Pulse Amplitude Modulation (PAM)*.

A partir dos pulsos PAM, podemos produzir os pulsos PCM (*Pulse Code Modulation*) através de um processo conhecido como *quantização*, onde cada amostra PAM é aproximada a um inteiro de n bits. No exemplo da Figura 1, escolhemos $n=3$, dando origem a oito níveis (2^3) de quantização. A saída PCM corresponde ao resultado dessa quantização.

Podemos calcular, a partir desse processo, denominado *conversão A/D*, a taxa gerada pela transmissão de informação analógica através de sinais digitais.[3] Considere o caso de sinais de voz, por exemplo. Se assumirmos que a banda passante necessária desses sinais tem largura igual a 3.100 Hz, a taxa de amostragem de Nyquist é, nesse caso, igual a 6.200 amostras por segundo. Normalmente amostra-se a uma taxa maior, para facilitar a construção dos codecs (codificadores/decodificadores). Se escolhermos uma taxa de 8.000 amostras por segundo e codificarmos cada amostra com oito bits, a taxa gerada será 8.000 x 8 = 64 Kbps, que é a taxa definida pelo padrão ITU-T G.711 [ITU-T G.711] para telefonia digital.

---

[2] Um sinal digital pode ser transformado em um sinal analógico, para transmissão em um dado meio, também pelo processo de modulação.

[3] Nesta monografia consideraremos sempre o sinal digital gerado a partir de uma informação codificada digitalmente como tendo sempre um bit por intervalo de sinalização, ou seja, um sinal onde sua taxa em bauds é a mesma que sua taxa em bits por segundo.

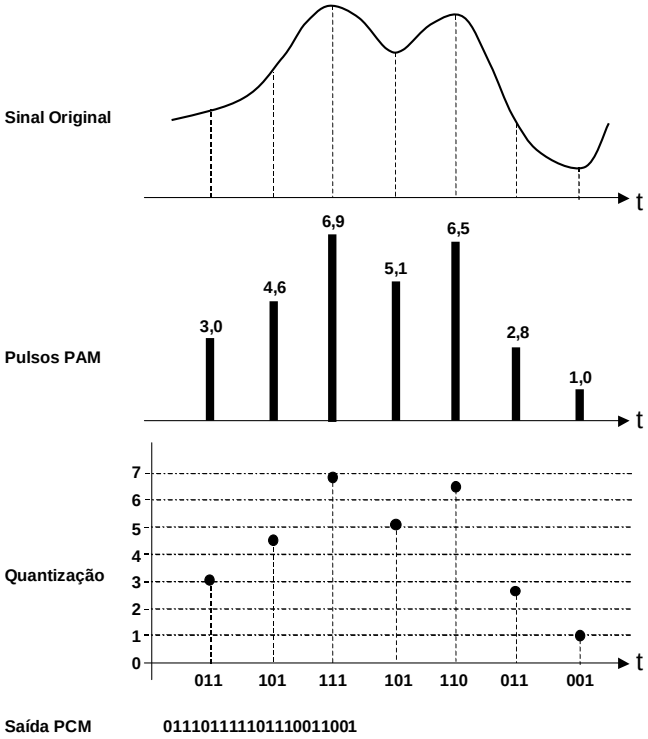


Figura 1: Pulsos PAM e PCM.

A Figura 1 ilustra o caso onde os níveis de quantização são igualmente espaçados, ou seja, o *quantum* ΔQ (diferença entre níveis) é constante. Como conseqüência, o erro máximo de quantização é ΔQ/2. Chamamos esse caso de *quantização linear*. Nele, as amostras pequenas são, em termos proporcionais, mais penalizadas pelo erro de quantização do que as grandes.

Para melhorar a qualidade do sinal amostrado, podemos usar uma quantização logarítmica, onde o sinal é primeiro logaritmicamente transformado de forma a manter o erro máximo de quantização aproximadamente constante, a despeito da amplitude da amostra. Várias funções logarítmicas foram propostas e estudadas com esse intento. Duas dessas funções são largamente utilizadas e padronizadas, sendo denominadas *lei A* e *lei μ.* A primeira é mais utilizada na Europa, enquanto a segunda predomina nos EUA.

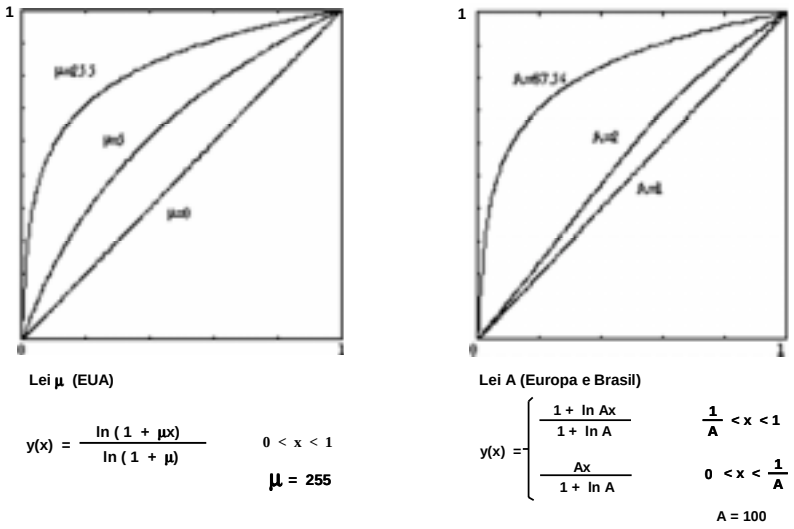


Figura 2: Lei A e Lei μ.

A Tabela 1 apresenta o resultado da conversão A/D de alguns sinais de áudio e vídeo.

Tabela 1: Conversão A/D para alguns sinais de vídeo e áudio.

Sinais	Faixa passante	Freqüência de amostragem	Codificação bits/amostra (b/a)	Taxa de bits
Voz (ITU-T G711)	300- 3.400Hz	8.000 Hz	Log PCM (8b/a)	64 Kbps
Qualidade CD (estéreo)	0 - 21KHz	44,1 KHz	Log PCM (16 b/a por canal)	1,41 Mbps (2 x 720,6 Kbps)
Vídeo (NTSC - luminância)	0 - 4,2 MHz	10 MHz	8 b/a	80 Mbps

3.2. Conversão D/A

Pode-se demonstrar que um trem de pulsos PCM, obtido pela amostragem de um sinal em uma freqüência maior ou igual à dada pelo teorema de Nyquist, tem o mesmo espectro de freqüência que o sinal amostrado, no intervalo de freqüências dado pela banda passante desse sinal. A conversão D/A se faz, então, pela simples passagem do trem de pulsos PCM por um filtro na faixa passante (e, assim, com a largura de banda) do sinal originalmente amostrado.

Não fosse pelo erro de quantização, o sinal obtido da saída do filtro seria idêntico ao sinal analógico original.

Note que o sinal de saída é tão mais próximo do sinal original quanto menor for o erro de quantização. O erro de quantização, por sua vez, é tão menor quanto maior o número de níveis de quantização, ou seja, quanto maior o número de bits utilizados na codificação.

3.3. Outros Codificadores de Onda

A codificação PCM representa cada amostra pelo seu valor absoluto, mas essa não é a única representação possível. Alternativamente, podemos representar apenas a diferença entre o valor de uma amostra e o valor de sua antecessora. Esse esquema de codificação é denominado *DPCM* (*Differencial Pulse Code Modulation)*. Quando os valores das amostras são muito próximos uns dos outros, necessitaremos de um menor número de níveis para representar as diferenças do que o necessário para representar os valores absolutos das amostras, para um mesmo erro de quantização. Como um menor número de níveis pode representar um menor número de bits para codificar todos os níveis, utilizando-se DPCM, poderemos ter um menor número de bits gerados pela codificação. DPCM é um caso particular de *codificação preditiva*, em que o valor predito da amostra corrente é o valor da amostra anterior, guardando-se (codificando-se) então o erro (diferença) de predição.

A idéia do DPCM pode ser ainda refinada um pouco mais, variando-se dinamicamente os níveis de quantização, dependendo se o sinal varia muito ou pouco. Dessa forma, prevê-se não apenas o valor da amostra corrente baseado na amostra anterior, mas também o valor do quantum, baseado em uma função, bem conhecida, dos valores de amostras anteriores. Esse esquema é denominado *ADPCM*, de *Adaptative Differencial Pulse Code Modulation.*

Existem ainda outras formas de codificação que independem do tipo do sinal analógico. Vamos citar apenas mais uma, a *SBC (SubBand Coding)*. Na codificação por sub-bandas, o espectro de freqüência do sinal é dividido em várias bandas de freqüência. Cada banda é então tratada como se representasse um sinal distinto, e nela é aplicada qualquer uma das técnicas anteriores. A vantagem da SBC é que, através da análise de um sinal, pode-se identificar suas bandas mais importantes no transporte da informação. Para essas bandas, pode-se utilizar um erro de quantização menor do que aquele usado nas bandas menos importantes, ou seja, pode-se codificar as bandas menos importantes utilizando um número menor de bits por amostras.

4. Compressão e Compactação

Um sinal digital, em geral, carrega muita informação redundante. Se eliminarmos essa redundância conseguiremos reduzir em muito a quantidade de bits gerados, que em alguns casos pode ser muito grande; na Tabela 1, por exemplo, pode ser observado que 1 minuto de vídeo preto e branco gera 600 Mbytes.

Quando eliminamos apenas a redundância de um sinal, não há perda de informação e dizemos que fizemos uma *compactação*, ou *compressão sem perdas*. No entanto, podemos também diminuir a quantidade de bits com alguma perda de informação. Dependendo de quem for o usuário da informação, parte dela pode ser considerada pouco útil. Raramente é necessário manter o sinal original intacto no caso das mídias vídeo, áudio e imagens estáticas, uma vez que o usuário final perderia de qualquer forma parte da informação por limitações físicas; tal é o caso do ouvido e olho humano. Vemos assim que a quantidade de informação que podemos perder pode ser dependente do usuário, mas ela também pode depender da tarefa em desenvolvimento: por exemplo, perder um pouco da nitidez de um vídeo em uma videotelefonia pode ser perfeitamente aceitável, enquanto a perda da qualidade do vídeo pode ser inadmissível em uma aplicação médica. Quando na redução dos dados gerados há perda de informação, dizemos que fizemos uma *compressão com perdas*, ou simplesmente *compressão*.

Existem várias técnicas de compressão sem perdas (compactação) que podem ser aplicadas a qualquer tipo de dados, independente da mídia representada. As Seções 4.1 à 4.5 são dedicadas a algumas dessas técnicas mais usuais. As técnicas de compressão com perdas serão estudadas para cada mídia em particular nas Seções 4.6 à 4.8.

4.1. Codificação por Carreira

O desempenho da codificação por carreira (run length coding) depende muito da estatística dos dados de entrada. Ela consiste simplesmente em representar os dados pelo seu valor e o número de vezes que ele se repete. A unidade para codificação pode ser um bit, um byte, um caractere, um pixel, uma amostra etc. A Figura 3 ilustra o caso da unidade ser um pixel de 8 bits e o caso da unidade ser o bit.

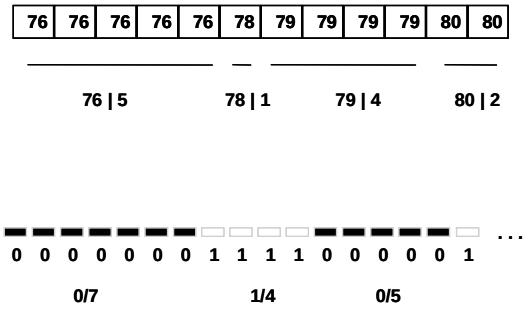


Figura 3: Codificação por carreira.

4.2. Codificação de Shannon-Fano

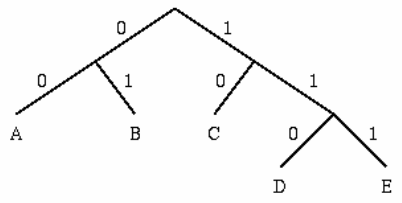
Para as duas próximas técnicas de codificação, imagine que nossos dados sejam apenas os símbolos A, B, C, D e E (que podem representar um pixel, uma amostra de áudio ou vídeo, um caractere etc.) e que eles ocorram na freqüência dada pela tabela a seguir.

Símbolo	A	B	C	D	E
Número de Ocorrências	15	7	6	6	5

A codificação de Shannon-Fano constrói a árvore de codificação seguindo o seguinte algoritmo:

1. Alinhe os símbolos de acordo com suas freqüências/probabilidades, e.g., ABCDE.
2. Divida recursivamente em duas partes, cada uma com aproximadamente o mesmo número de contagem (soma das freqüências).

A árvore gerada fica então:



E temos a seguinte codificação, gerando uma compactação de 117:89.

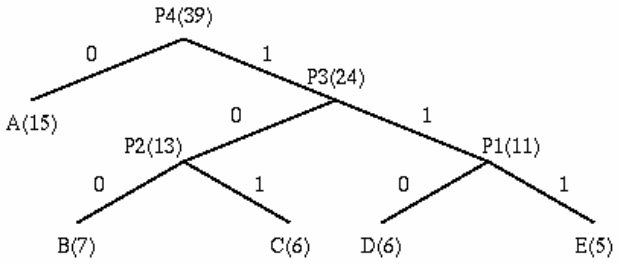
Símbolo	Freqüência	Código	Subtotal (#bits)
A	15	00	30
B	7	01	14
C	6	10	12
D	6	110	18
E	5	111	15

4.3. Codificação de Huffman

Tomando o mesmo exemplo da seção anterior, a codificação de Huffman constrói a árvore de codificação seguindo o seguinte algoritmo:

1. Iniciação: Ponha todos os nós em uma lista ABERTA. Mantenha a lista alinhada todo o tempo de aplicação do algoritmo (e.g., ABCDE).
2. Repita até que a lista ABERTA contenha apenas um nó:
    a. Pegue os dois nós de mais baixas freqüências/probabilidades e crie um nó pai para ambos.
    b. Atribua ao nó pai a soma das freqüências/probabilidades dos filhos e o insira na lista ABERTA.
    c. Atribua códigos 0 e 1 aos dois ramos da árvore, e retire os filhos da lista ABERTA.

A árvore gerada fica então:



E temos a seguinte codificação, gerando uma compactação de 117:87.

Símbolo	Freqüência	Código	Subtotal (#bits)
A	15	0	15
B	7	100	21
C	6	101	18
D	6	110	18
E	5	111	15

Note que, quer usando a codificação de Huffman ou de Shannon-Fano, o decodificador deve usar o mesmo dicionário de códigos gerado pelo codificador para recuperar os símbolos originais.

4.4. Codificação de Lempel-Ziv-Welch

Um procedimento diferente dos dois anteriores é processar símbolo a símbolo e ir construindo o dicionário de códigos passo a passo. À medida que o dicionário vai sendo construído, ele pode ser usado na codificação do próximo símbolo, dinamicamente.

No esquema de Lempel e Ziv, posteriormente estendido por Welch, o dicionário é construído como uma estrutura de dados, uma tabela, que mantém seqüências de símbolos, em conjunto com um identificador único (código) para toda a seqüência. A tabela contém até, digamos, 2^j posições (seqüências). Ela é iniciada simplesmente com o conjunto dos 2^k possíveis símbolos, isto é, todas as seqüências de tamanho 1. Os melhores desempenhos são conseguidos quando $k << j$, dependendo, obviamente, do grau de redundância dos dados.

A codificação se inicia definindo a seqüência de símbolos corrente *S* como o primeiro símbolo a codificar. Note que *S* é membro do dicionário. A codificação então continua como a seguir:

1. Se não existem mais símbolos para codificar, dê como saída o código da seqüência (j bits) para S. Em caso contrário,
2. Pegue o próximo símbolo P e concatene a S, obtendo a nova seqüência SP.
3. Se SP já estiver no dicionário, faça S = SP e volte para o passo 1. Em caso contrário,
4. Dê como saída o código da seqüência (j bits) para S.
5. Adicione SP ao dicionário, se ainda houver espaço.
6. Faça S = P e volte para o passo 1.

Note que assim procedendo, o decodificador não tem a necessidade de conhecer a priori todo o dicionário, pois pode reconstruí-lo passo a passo, dinamicamente, a partir dos dados codificados. A decodificação se inicia definindo a seqüência de símbolos corrente *S* como a entrada no dicionário correspondente ao primeiro código a decodificar e dando como saída o símbolo *S*. Se houver mais códigos a decodificar faça:

1. Leia próximo código X

2. Se houver a entrada no dicionário (P) correspondente a X:
   a) Dê como saída P.
   b) Adicione S concatenado ao primeiro símbolo de P no dicionário, caso a entrada não exista.

   Se não houver a entrada no dicionário (P) correspondente a X:
   a) Faça P igual a S concatenado ao primeiro símbolo de S e adicione ao dicionário.
   b) Dê como saída P.

3. Faça S igual a P e volte para o passo 1, se houver mais símbolos a decodificar.

4.5. Outras Técnicas de Compactação

Além das técnicas anteriormente mencionadas, outras são encontradas, bem como variantes das primeiras. Uma, no entanto, merece destaque por ser comumente utilizada: a codificação aritmética.

A codificação aritmética também parte do conhecimento da freqüência de ocorrências dos símbolos, tal qual as codificações de Shannon-Fano e de Huffman. Baseado na freqüência de ocorrências, intervalos são associados aos símbolos e, a partir desses intervalos, novos intervalos são construídos para seqüências de símbolos. Uma seqüência pode então ser codificada por qualquer número dentro do intervalo calculado para a seqüência, garantindo sua decodificação posterior sem ambigüidade. Na referência [Cormen 02] pode-se encontrar a especificação detalhada do algoritmo.

4.6. Compressão em Imagem Estática

As imagens geram, normalmente, uma quantidade de informação muito grande. Se levarmos em consideração a correlação do valor (cor) de cada pixel, podemos reduzir a quantidade de informação gerada.

Existe um grande número de formatos para imagens estáticas, baseados em esquemas diferentes de compressão. Dentre os mais usuais atualmente podemos citar os formatos *TIFF* e *GIF*, que são baseados no algoritmo de Lempel-Ziv, apresentado na Seção 4.4, e o padrão ISO para imagens estáticas, chamado JPEG [ISO 94], baseado em transformadas, como veremos mais adiante.

A forma mais simples de compressão de uma imagem é a redução da sua resolução geométrica. Isso implica em aumentarmos o tamanho da região de um pixel. Tal procedimento pode ser feito até um certo limite, dependendo do usuário final e do dispositivo de exibição, para se evitar o efeito apresentado na Figura 4, onde vemos primeiro a imagem original dividida em pixels e, em seguida, a imagem reproduzida; a diferença se deve ao fato do pixel ter uma região grande.

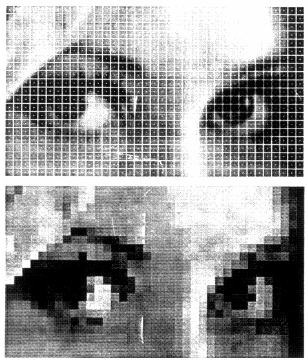


Figura 4: Efeito do tamanho da região de um pixel na codificação/decodificação.

Outra forma de compressão é a redução do espaço de cor pela simples utilização de um número menor de bits para representação de cada pixel. Cabe antes, no entanto, uma discussão sobre como as cores são representadas.

Através das adições das cores vermelha, verde e azul, podemos obter *quase todas* as cores visíveis pelo olho humano. Assim, uma representação bem comum é a RGB (de Red – Green – Blue), onde um pixel é representado pelos valores dessas componentes. É comum encontrarmos o padrão RGB 8-8-8, onde se utilizam 8 bits para codificação de cada componente; e o padrão 5-6-5, onde é reservado um número maior de bits (6) para a componente verde, por ser o olho humano mais sensível a essa componente. Outra representação bastante utilizada é o sistema YC_rC_b. A componente Y é denominada *luminância*, e é uma medida da sensibilidade do olho humano às várias componentes de freqüência de uma cor. Para as fontes usuais de luz provenientes de dispositivos de vídeo, Y é dada por:

$$Y = 0{,}299R + 0{,}587G + 0{,}114B$$

As componentes C_r e C_b são chamadas de *crominância* e sua definição varia de padrão para padrão, bem como seu nome (componentes I e Q no padrão NTSC de TV, e componentes U e V no padrão PAL de TV). C_r e C_b são os nomes utilizados pelos padrões MPEG e JPEG, e têm seus valores dados por:

$$C_r = ((R - Y) / 1{,}6) + 0{,}5$$

$$C_b = ((B - Y)/ 2) + 0{,}5$$

O leitor deve observar que o conjunto de equações para Y, C_r e C_b é linearmente independente (isso também acontece para as outras definições de C_r e C_b), ou seja, dado Y, C_r e C_b, obtém-se facilmente R, G e B.

O olho humano é mais sensível à luminância do que à crominância, assim, na compressão pela redução da resolução de cor, podemos utilizar um número menor de bits para as componentes da crominância. Mais comum, no entanto, é utilizarmos uma menor resolução geométrica para as componentes de crominância do que aquela utilizada para a luminância.

Uma outra forma de compressão de imagem estática, geralmente sem perdas, é a codificação preditiva. De forma análoga à explicação da Seção 3.3, na codificação preditiva de uma imagem, realiza-se uma predição do valor do pixel baseada em valores de outros pixels da imagem. A diferença do valor real para o valor predito é então codificada. A Figura 5 ilustra o caso.

Figura 5: Imagem original e imagem com apenas os erros de predição.

Se na imagem os pixels tiverem valores muito próximos, pode-se usar um número menor de bits para armazenar o erro da predição, do que aquele usado para codificar o valor absoluto do

pixel. Além disso, uma imagem com poucos contornos vai gerar muitos valores pequenos, ou mesmo o valor zero, tornando o emprego de um outro método de compressão posterior bem eficiente.

Antes de continuarmos nossa discussão sobre técnicas de compressão, devemos ressaltar que, normalmente, as técnicas de compressão são seguidas pela aplicação de algum esquema de compactação. Muitas vezes, o esquema de compressão simplesmente prepara os dados para que possam sofrer uma maior compactação. Nada impede também que apliquemos várias técnicas de compressão em seqüência.

Um exemplo de codificação preditiva pode ser encontrado no padrão JPEG [ISO 94] no modo sem perdas, onde a codificação de Huffman é aplicada após a codificação preditiva. No esquema, apresentado na Figura 6, o codificador por entropia utiliza a codificação de Huffman. Note também, pela figura, que existem 7 possíveis predições para um pixel *X*.

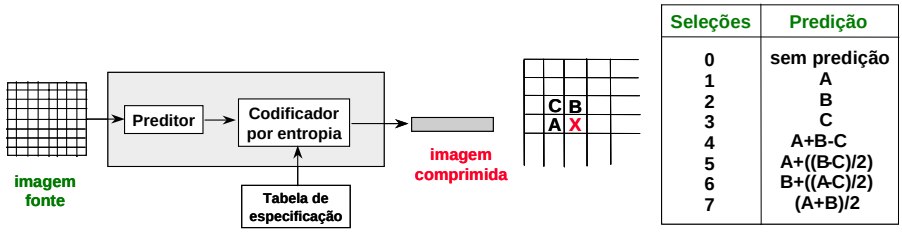


Figura 6: JPEG sem perdas.

Existem ainda outras técnicas para compressão de imagem, tais como a codificação por sub-bandas (similar ao apresentado na Seção 3.3) e a quantização vetorial. Entretanto, nós nos deteremos apenas em mais uma, por ser utilizada nos padrões JPEG e MPEG: a codificação por transformadas.

O leitor já deve ter percebido, pelas várias referências à Seção 3.3, que existem várias similaridades entre amostras no tempo e pixels. Na verdade, podemos considerar os pixels como se fossem amostras do "sinal imagem", só que amostras obtidas não no tempo, mas no espaço. É exatamente por isso que podemos aplicar todas as técnicas da Seção 3.3 nas imagens estáticas. Note também que, em um sinal de vídeo, o grupo de várias amostras temporais formam um quadro (por exemplo, no nosso sistema de TV, existem 30 quadros por segundo). Esse quadro é uma imagem estática onde, de fato, as amostras temporais do vídeo são os pixels (amostras espaciais). Esse fato é que nos permite não somente tratar o vídeo como um sinal contínuo e nele aplicarmos todas as técnicas de compressão conhecidas para sinal contínuo, como também tratá-lo como uma seqüência de imagens estáticas no tempo, aplicando as mesmas técnicas de compressão utilizadas para imagens.

Uma vez que uma imagem estática pode ser considerada uma seqüência de amostras espaciais, nós podemos agora pensar, como fizemos com os sinais analógicos, em aplicar uma transformada (por exemplo Fourier) para descrever o mesmo sinal no domínio da freqüência. Só que agora no domínio das freqüências espaciais. Como estamos com um sinal discreto, precisaremos de uma transformada discreta. Poderíamos usar a transformada discreta de Fourier, como mencionado, mas outra transformada leva a melhores resultados na compressão: a *transformada discreta de co-senos*.

No espaço bidimensional de uma imagem de 8x8 pixels, a transformada discreta de co-senos (*FDCT: Forward Discrete Cosine Transform*) é dada por:

$$F(u,v) = \frac{1}{4} C(u)C(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right]$$

$$C(w) = \frac{1}{\sqrt{2}} \quad para \;\; w = 0$$

$$C(w) = 1 \quad para \;\; w = 1, 2, \ldots, 7$$

E a transformada inversa (*IDCT: Inverse Discrete Cosine Transform*) por:

$$f(x,y) = \frac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C(u)C(v)F(u,v) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right]$$

No domínio da freqüência, as mudanças abruptas que acontecem nos contornos de uma figura estão concentradas nas freqüências mais altas. Assim, uma imagem com poucos contornos deve concentrar seus coeficientes nas freqüências baixas. Mais ainda, os coeficientes das freqüências altas são menos importantes e perdas nesses coeficientes podem diminuir um pouco a nitidez da imagem, mas para muitas aplicações isto pode ser aceitável.

Pelos motivos mencionados no parágrafo anterior, após uma imagem ser transformada, os coeficientes gerados são quantizados de forma diferenciada, usando uma maior precisão (quantum menor) para as freqüências mais baixas. Assim procede o padrão JPEG [ISO 94] no modo seqüencial, dividindo uma imagem em blocos de 8x8 pixels e aplicando uma compressão em cada bloco, conforme o diagrama mostrado na Figura 7. A imagem é varrida uma única vez, da esquerda para direita, de cima para baixo.

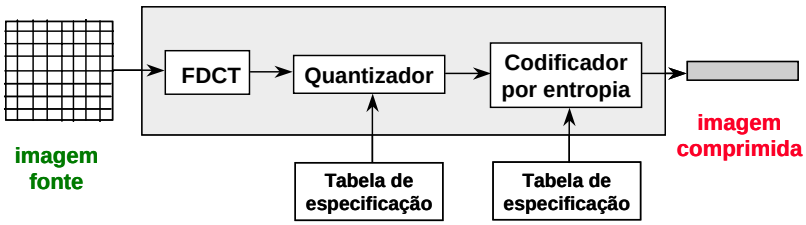


Figura 7: Codificação JPEG modo seqüencial.

No JPEG, após a aplicação da transformada discreta de co-senos e a quantização dos coeficientes, esses são organizados da mais baixa freqüência para a mais alta,[4] quando então é aplicada a codificação por carreira, seguida da codificação de Huffman (ou codificação aritmética), ilustradas na Figura 7 pelo bloco "Codificador por entropia".

[4] Na verdade, um passo adicional existe no JPEG, quando os coeficientes DC (freqüência zero) de um bloco são codificados pela diferença entre seu valor e o valor do coeficiente DC do bloco anterior.

A decodificação JPEG modo seqüencial é ilustrada na Figura 8.

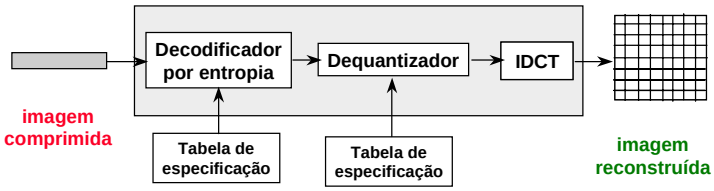


Figura 8: Decodificação JPEG modo seqüencial.

O padrão JPEG ainda possui mais dois modos de compressão: o *progressivo* e o *hierárquico*.

O modo progressivo também utiliza a transformada de co-senos, mas a imagem é codificada em várias varreduras. Na variante denominada *seleção de espectro*, a cada varredura são codificados coeficientes das mesmas freqüências de todos os blocos, da mais baixa freqüência para a maior freqüência. Já na variante denominada *aproximação sucessiva*, primeiro é codificado o coeficiente de mais baixa freqüência de todos os blocos e depois são codificados, paulatinamente, os bits dos demais coeficientes de todos os blocos, do mais significativo para o menos significativo.

Note que no modo progressivo, os primeiros dados da imagem que são decodificados já permitem ter uma visão completa da cena, embora ainda sem muita nitidez. Com o restante dos dados, a nitidez vai se tornando cada vez melhor. Isto pode ser útil em vários casos. Muitas vezes, navegando na Web, queremos apenas ter uma idéia da informação que vem na página a seguir. Certas vezes essa informação nem é aquela que desejamos. Ter uma visão rápida dessa informação, ainda que não com todos os detalhes, pode acelerar em muito a tarefa sendo executada. Outras vezes podemos estar trafegando com a imagem em um meio de pequena banda passante, ou mesmo em um trecho congestionado de uma rede. Nesses casos, o descarte seletivo dos dados que não trazem muita informação pode ser a única forma viável de manter a aplicação em funcionamento. Como veremos, um bom sistema de comunicação deve poder identificar a parte da informação que ele deve manter íntegra e quais partes ele pode perder, em caso de necessidade ou conveniência.

O modo JPEG hierárquico também permite separar da imagem os dados mais relevantes dos menos relevantes, mas através do aumento progressivo da resolução geométrica. Nesse modo, a imagem é codificada com múltiplas resoluções, de forma que a menor resolução pode ser decodificada sem a necessidade de se ter a resolução maior.

4.7. Compressão em Áudio

A compressão de um sinal de áudio depende muito do tipo de sinal. Vamos começar pela voz humana.

Um ser humano falando emite surtos de voz apenas durante 35% a 40% do tempo de fala. O restante do tempo é preenchido com silêncio que existe entre palavras e entre uma sentença e outra. Se pudermos detectar esse silêncio e eliminá-lo da codificação, de forma que ele possa ser recuperado na decodificação, reduziremos muito a quantidade de dados gerados. Essa técnica é aplicada à telefonia digital com o nome TASI (Time Assignment Digital

Interpolation). Ainda como outra característica da voz e do ouvido humano, a perda de surto de voz e de silêncio é muito diferente. Perdas da ordem de 1% da informação do surto de voz são toleráveis,[5] ao passo que podemos tolerar a perda de até 50% do silêncio. Note que, com a detecção de silêncio, transformamos um tráfego de voz contínuo em um tráfego em rajadas.

Uma outra forma de comprimir a voz humana é codificar, ao invés de suas amostras, os parâmetros de um modelo analítico do trato vocal capaz de gerar aquelas amostras. No método conhecido como LPC (Linear Predictive Coding), apenas os parâmetros que descrevem o melhor modelo que se adapta às amostras é codificado. Um decodificador LPC usa esses parâmetros para a geração sintética da voz que é, usualmente, parecida com a original. O resultado é inteligível, mas a tonalidade é aquela de um robô falando.

CELP (Code Excited Linear Predictor) é bastante similar a LPC. O codificador CELP gera os mesmos parâmetros LPC, mas computa os erros entre a fala original e a fala gerada pelo modelo sintético. Tanto os parâmetros do modelo analítico do trato vocal quanto uma representação comprimida dos erros são codificados. A representação comprimida é um índice em um vetor de excitação (que pode ser pensado como um livro de códigos compartilhado pelo codificador e decodificador). O resultado do CELP tem uma qualidade de fala muito boa a uma taxa bem baixa. A Tabela 2 apresenta alguns padrões recomendados pelo ITU-T [ITU-T G.711; ITU-G.722; ITU-T G.723; ITU-T G.726; e ITU-T G.729] para voz.

Tabela 2: Padrões ITU-T para voz.

Padrão	Algoritmo	Taxa de compressão (Kbps)	Recursos de processamento necessários	Qualidade da voz resultante	Atraso adicionado
G.711	PCM	48, 56, 64 (sem compressão)	Nenhum	Excelente	Nenhum
G.722	SBC/ADPCM	64 (faixa passante de50 a 7KHz)	Moderado	Excelente	Alto
G.723	MP-MLQ	5.3, 6.3	Moderado	Boa (6.3) Moderada (5.3)	Alto
G.726	ADPCM	16, 24, 32, 40	Baixo	Boa (40) Moderada (24)	Muito baixo
G.738	LD-CELP	16	Muito Alto	Boa	Baixo
G.729	CS-ACELP	8	Alto	Boa	Baixo

Mais especificamente para áudio, de uma forma geral, um padrão muito importante é o MPEG áudio [ISO 93a e ISO 98].

MPEG áudio leva em conta o modelo psicoacústico humano para realizar uma compressão "perceptualmente sem perdas". O modelo divide o domínio de freqüência audível (entre 20 Hz e 20 KHz) em 32 bandas, chamadas bandas críticas. O sistema de audição tem uma resolução limitada e dependente da freqüência. A medida perceptualmente uniforme de freqüências pode ser expressa em termos das larguras das *bandas críticas*. A Figura 9 mostra a sensibilidade do ouvido humano nas diversas freqüências.

[5] Na realidade, a percentagem de perda depende do tamanho do surto de voz e se a perda ocorre no início ou no meio do surto. Na referência [Gruber 85] é possível encontrar uma discussão sobre o assunto.

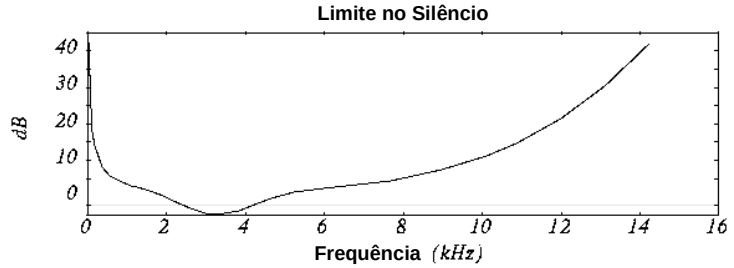


Figura 9: Sensibilidade do ouvido humano.

Note que a sensibilidade do ouvido ilustrada na Figura 9 é medida em *decibéis* (*dB SPL — dB Sound Preassure Level —* ou simplificadamente *dB*). O decibel é uma unidade conveniente para expressar o que se chama de *nível sonoro*. O som, de uma forma geral, tem uma medida de *intensidade* que é a potência transferida por uma onda sonora por área de uma superfície que intercepta essa onda. O decibel nada mais é do que uma forma comparativa de analisar valores. Nesse caso, ao invés de fornecer o valor absoluto da intensidade sonora para uma freqüência, podemos fornecer o seu valor dividido pela menor intensidade perceptível ao ouvido humano[6] e utilizar uma escala logarítmica para representar essa razão, já que o intervalo de intensidades produzido pela voz humana é muito grande. Assim, o nível sonoro em decibéis β é definido como:

$$\beta = 10 \log_{10}\left(\frac{I}{I_0}\right)$$

onde I_0 é o menor valor de intensidade sonora perceptível ao ouvido e I é a intensidade do som sendo medido. Note que, para $I = I_0$, temos que $\beta = 0$ *dB,* ou seja, a nossa referência de menor intensidade perceptível corresponde ao 0 dB.

Voltando ao MPEG áudio, seu modelo leva em conta o mascaramento de freqüências, característica do ouvido humano que quando submetido a um sinal de uma certa amplitude em uma dada freqüência, mascara as outras freqüências ao redor, que possuam uma amplitude abaixo de um certo limite. A Figura 10 ilustra o caso.

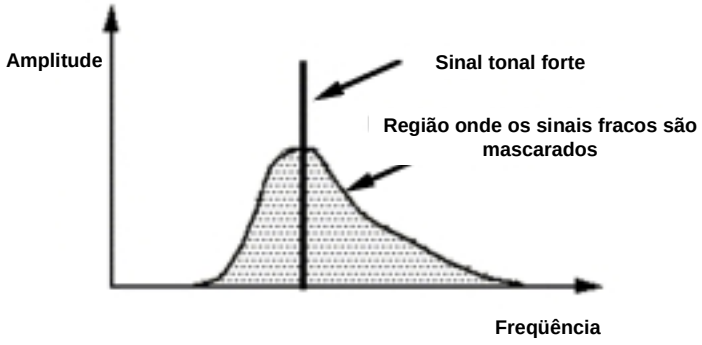


Figura 10: Mascaramento de freqüências.

[6] O valor de referência é definido como $I_0 = 10^{-12}$ Watts/m^2. Ele corresponde à aproximadamente o limiar da audição humana. No entanto, o limiar de audição varia de freqüência para freqüência e com a intensidade do som, como descoberto por Fletcher e Munson em 1933.

MPEG áudio transforma o sinal para o domínio da freqüência e aplica o mascaramento de freqüências, codificando apenas aquelas componentes de freqüência que não são mascaradas. A Figura 11 ilustra o procedimento.

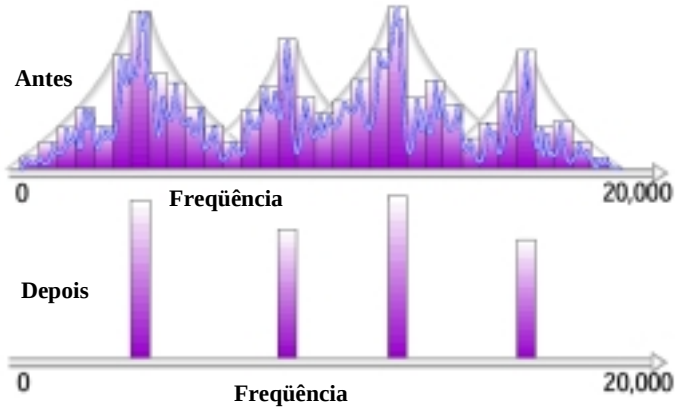


Figura 11: Mascaramento de freqüências nas bandas críticas MPEG.

MPEG áudio também leva em consideração resultados psicoacústicos que mostram que, para freqüências maiores que 2 KHz, o ouvido percebe a imagem estereofônica baseado mais no envelope temporal do áudio do que em sua estrutura mais refinada. Assim, no modo "intensity stereo", o codificador soma as freqüências mais altas do sinal estereofônico em um único sinal. Os canais de esquerda e direita são reconstruídos com a mesma forma, mas com magnitudes diferentes, baseadas em fatores de escala.

Na codificação MPEG, para cada intervalo de tempo de áudio codificado (isto é, para cada conjunto de amostras), existe um número fixo de bits total para todas as 32 sub-bandas. Escolhe-se o número de bits de uma banda de forma a minimizar a percepção auditiva do ruído de quantização, levando-se em conta, como já mencionamos, o mascaramento de freqüências.

MPEG 1 áudio [ISO 93a] define três métodos de compressão, denominados camadas 1, 2 e 3 (MP1, MP2, MP3).

MP1 agrupa 12 amostras para cada uma das 32 sub-bandas. Cada grupo de 12 recebe então os bits para codificação e, se o número de bits não é zero, um fator de escala.

MP2 e MP3 ainda levam em conta uma outra característica psicoacústica, o mascaramento temporal. Quando é emitido um tom em uma dada freqüência e com uma certa amplitude, esse tom mascara os tons, na mesma freqüência, abaixo de uma certa amplitude, que varia no tempo. A Figura 12 ilustra o fato com um tom emitido a 60dB.

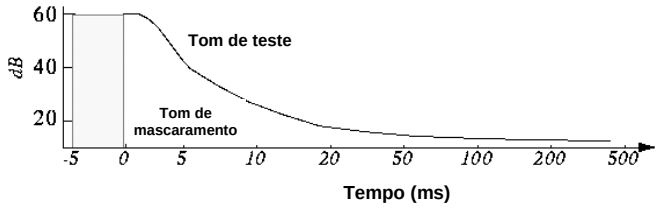


Figura 12: Mascaramento temporal.

MP2 codifica os dados em grupos maiores: para cada sub-banda agrupa 3 grupos de 12 amostras. Isso modela um pouco da máscara temporal, pois se faz uma alocação de bits e se usa três fatores de escala para cada trio de 12 amostras.

Tanto MP1 quanto MP2 usam bandas uniformes, isto é, do mesmo tamanho, que não modelam bem o ouvido humano, como pode ser visto pela Figura 9. MP3 usa bandas não uniformes e faz uma alocação de bits melhor. MP3 faz um melhor cálculo do quantum de cada banda, isto é, uma melhor distribuição de bits, usando inclusive o conceito de reservatório de bits (como mencionado, o número de bits total para as 32 bandas é fixo para cada grupo de amostras, mas no MP3 pode haver empréstimos de bits de um grupo de amostras para outro).

MP2 e MP3 também permitem o *MS stereo mode*, além do *stereo intensity mode*. O modo MS stereo codifica os sinais de freqüências mais altas dos canais direito e esquerdo como a soma (middle-channel) e diferença (side-channel). Técnicas de sintonização são então utilizadas para comprimir o sinal side-channel.

A Tabela 3 apresenta uma comparação das várias camadas MPEG 1 áudio.

Tabela 3: Camadas MPEG 1 áudio (a codificação pode ser realizada com taxas de amostragem de 32, 41.1 e 48 KHz).

Camadas	Taxa de bits alvo (Kbps)	Taxa de compressão	Qualidade	Retardo Máximo (ms)
MP1	32 a 448	4:1		50
MP2	32 a 384	6:1		100
MP3	32 a 320	12:1	Telefonia: 8 Kbps Rádio AM: 32 Kbps Rádio FM: 64 Kbps CD: 128 Kbps	150

As camadas do padrão MPEG 1 áudio, denominada fase 1 (MP1, MP2 e MP3) prevêem apenas o uso de dois canais em um fluxo de áudio, ou seja, apenas o estéreo tradicional. O padrão MPEG 2 [ISO 98] introduz algumas extensões. O padrão MPEG 2 áudio (fase 2) comum, chamado de "BC" (Backward Compatible), tem as mesmas camadas e usa os mesmos algoritmos com os mesmos parâmetros do áudio MPEG 1. A diferença é que o

MPEG 2 prevê a formação de fluxos de áudio com até 6 canais, em vez de implementar apenas o estéreo com dois canais.

A codificação AAC é novidade do áudio MPEG 2, conhecida como NBC (Non Backward Compatible), de não compatível com MPEG 1. Essa codificação é mais eficiente que o MPEG 1 (MP3 etc.), tolera até 48 canais de áudio e 15 canais de freqüências baixas, com taxas de amostragem de 8 a 96 KHz. A AAC necessita de menos processamento e, conseqüentemente, tem retardo menor na (de)codificação.

Além das codificações apresentadas, ainda existem várias outras em uso atual. Entre elas podemos citar a DD e a DTS.

AC-3 era o antigo nome da codificação chamada hoje de Dolby Digital (DD). Essa codificação é proprietária da empresa Dolby, mas foi adotada pelos EUA como codificação de áudio a ser utilizada nos DVDs e em HDTV (High Definition TV). Ela utiliza 6 canais de áudio, sendo 5 com qualidade de CD (20 Hz a 20 KHz) e um apenas para as baixas freqüências, (20 a 120 Hz). A taxa dessa codificação é de cerca de 384 Kbps. A Figura 13 apresenta a distribuição sugerida de auto-falantes para os 6 canais de áudio.

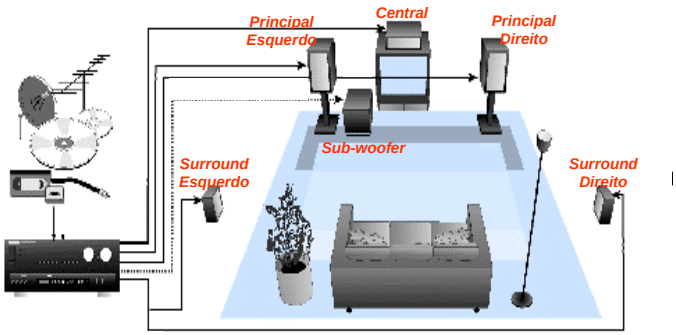


Figura 13: Áudio multicanal.

A Europa usa ainda uma outra codificação proprietária, a DTS. Essa também é multicanal, mas a taxa gerada é de cerca de 1,536 Mbps. Tanto o DD quanto DTS trabalham com codificação por sub-bandas (até 576 na DD). Testes exaustivos com especialistas em áudio não conseguem chegar a uma definição sobre qual codificação é a melhor. No entanto, trilhas DD obviamente ocupam menos espaço de armazenamento (e de banda, quando transmitidas) e, por isso, a codificação DD tem sido preferida pelos fabricantes de DVD.

4.8. Compressão em Vídeo

Como vimos na Seção 3, o sinal de vídeo pode ser codificado usando qualquer uma das técnicas lá mencionadas: PCM, DPCM, ADPCM, SBC etc. Também conforme vimos na Seção 4.6, um vídeo pode ser considerado como uma seqüência de imagens estáticas, ou quadros. Cada um desses quadros pode ser codificado usando as mesmas técnicas empregadas para as imagens estáticas. Em particular, poderíamos empregar a codificação JPEG em cada quadro. Essa técnica constitui a base da codificação chamada MJPEG (Motion JPEG). Entretanto, ao empregarmos essa codificação, estaremos levando em conta apenas a redundância de informação contida em um quadro (redundância *intra-quadro*), quando a

maior redundância pode estar nas informações contidas em quadros consecutivos (redundância *inter-quadros*).

Nesta seção nos deteremos na análise de dois padrões de codificação de vídeo que levam em conta não apenas a redundância intra-quadro, mas também a existente inter-quadros: os padrões H.261 e MPEG vídeo. Antes, porém, vamos analisar alguns padrões de sinal de vídeo.

Sinais de TV são, em geral, adquiridos no formato RGB, mas transmitidos no formato YC_rC_b, onde a resolução dos canais de crominância é menor que a de luminância, levando em conta a maior sensibilidade do olho humano à luminância, como discutimos na Seção 4.6. Os sinais são então multiplexados e modulados, gerando um sinal chamado vídeo composto modulado, conforme ilustrado na Figura 14.

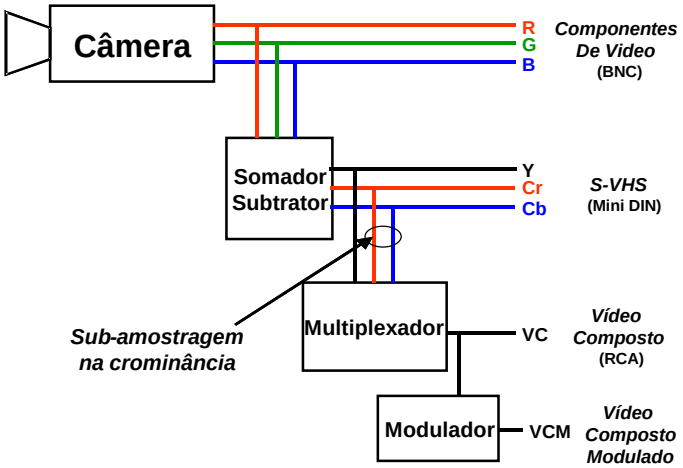


Figura 14: Geração de sinal de vídeo de TV.

Sistemas de vídeo apresentam informações como uma seqüência de quadros, sendo cada quadro composto de linhas. Um dos sistemas de distribuição de televisão mais utilizado usa 525 linhas por quadro, a uma taxa de 30 quadros/segundo (é o padrão M de TV, utilizado tanto pelo padrão americano NTSC, quanto pelo padrão brasileiro PAL-M).[7] As televisões têm uma relação de aspecto de 4:3, dando ao padrão M uma resolução para a luminância de 700 x 525 pixels por quadro.

Nem todas as linhas da televisão são visíveis. A maioria dos formatos de vídeo digital estão relacionados com a área visível para cada padrão de televisão. A recomendação BT 601-4 [ITU-R BT.601-4] especifica 483 linhas ativas, com 720 pixels por linha. Apenas 480 das 483 linhas e 704 dos 720 pixels (os primeiros e últimos 8 pixels são descartados) são usados para codificação. O padrão especifica a sub-amostragem de crominância 4:2:2, conforme dado pela Figura 15, indicando que, para cada dois valores de luminância na horizontal, apenas um de cada crominância deve ser gerado. Isso implica em uma taxa gerada de:

$$704 \times 480 \times 29{,}97 \times 16 = 162 \text{ Mbps}$$

[7] Na verdade, a taxa de quadros é de 29,97 quadros/segundo para TV colorida. Os padrões europeus, em geral, usam 25 quadros/segundo e 625 linhas por quadro.

A Figura 15 apresenta também outras sub-amostragens de crominância utilizadas em outros padrões de codificação, como veremos.

Figura 15: Sub-amostragem de crominância MPEG 2.

O padrão H.261 do ITU-T, discutido mais adiante, usa os formatos CIF (*Common Interchange Format*), com 288 linhas e 352 pixels por linha, e QCIF (*Quarter CIF*), com 144 linhas e 176 pixels por linha, para a luminância. Sua extensão, H.263, acrescenta três novos formatos: o sub-QCIF (128 x 96), o 4CIF (704 x 576), também conhecido como SCIF, e o 16CIF (1408 x 1152). Todos os formatos citados possuem sub-amostragem de crominância (relação de aspecto) 4:2:0, conforme ilustrado na Figura 15.

O padrão MPEG 1 vídeo pode codificar imagens de até 4096 linhas x 4096 pixels x 60 quadros/segundo. No entanto, a maioria das aplicações usam o formato SIF, com 240 linhas, 352 pixels por linha e sub-amostragem de crominância 4:2:0.

O padrão MPEG 2 pode codificar imagens de até 16.383 linhas x 16.383 pixels. O padrão é organizado, como veremos, em diversos perfis e níveis, que especificam o formato utilizado. Exemplos de formato são: nível baixo (240 linhas x 352 pixels/linha x 30 quadros/segundo – idêntico ao SIF MPEG 1), nível principal, visando a codificação com qualidade de TV (720 x 480 x 30), e os níveis alto, visando a TV de alta resolução - HDTV, e a produção de filmes (em geral 1280 x 720 x 30; 1920 x 1080 x 30 ou 1440 x 1152 x 30). O padrão permite sub-amostragem de crominância 4:2:0, 4:2:2 e 4:4:4.

Note que vários formatos são menores que os tamanhos de TV atuais. Note também que as imagens de TV são significativamente menores do que as telas atuais das estações de trabalho. Quase todos os formatos de vídeo digital apresentam a imagem em uma área menor do que a tela da estação. Alguns padrões, no entanto, chegam a resoluções suficientes para atender a qualidade das TVs de alta resolução, a HDTV.

4.8.1. H.261

H.261 [ITU-T H.261] é o padrão de compressão mais usado em sistemas de videoconferência. Seu objetivo inicial foi as aplicações para redes comutadas por circuito, mais especificamente a RDSI-FE. Daí decorre sua geração de tráfego CBR (de Constant Bit Rate, isto é, taxa constante) nas taxas de *p x 64* Kbps, *p* variando de 1 a 30.

H.261 divide cada quadro (QCIF ou CIF) em macroblocos de 16 x 16 pixels, gerando 6 blocos de 8 x 8 pixels (4 de luminância e 2 de crominância), conforme ilustra a Figura 16.

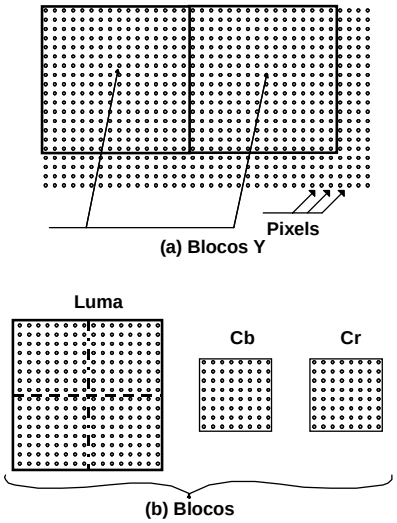


Figura 16: Blocos H.261.

O padrão define dois tipos de codificação. Na codificação intra-quadro, o macrobloco é codificado levando em conta apenas a redundância espacial do bloco. Cada bloco passa por um processo muito parecido com o descrito no JPEG modo seqüencial, gerando os blocos codificados. Na codificação preditiva, é realizada uma pesquisa no quadro anterior à procura de um macrobloco o mais parecido possível com o macrobloco que se quer codificar (a pesquisa é realizada apenas na componente de luminância e apenas em uma região que circunda o macrobloco). O erro de predição (diferença entre os macroblocos) é então enviado para codificação, sofrendo então o mesmo processo descrito para a codificação intra-quadro. No caso da codificação preditiva, deve também ser codificado o vetor (chamado de vetor de deslocamento), que especifica o deslocamento entre o macrobloco corrente e o macrobloco do quadro anterior usado para a predição.

Todo macrobloco sofre uma codificação intra-quadro e preditiva. A que gera a maior compressão é escolhida. Uma vez codificados, os quadros são enviados a um buffer que vai regular o fluxo de informação para uma taxa de bits constante. Lembre-se que o H.261 foi pensado para uma rede comutada por circuitos. Como a taxa de entrada no buffer é VBR (Variable Bit Rate, isto é, taxa variável), se não fosse tomada nenhuma providência poderia ocorrer estouro de buffer ou falta de bits codificados. Para que isso não aconteça, o tamanho do passo do quantizador (o quantum), dos coeficientes gerados a partir da transformada de co-senos, é ajustado, quando necessário, conforme a quantidade de dados no buffer, para se chegar à taxa CBR desejada de saída.

O ajuste no passo do quantizador afeta a qualidade do vídeo gerado, privilegiando os vídeos com poucas mudanças de cena. Pouca mudança de cena implica em maior compressão, isto é, menos bits gerados na codificação que entram no buffer, o que faz o processo de realimentação diminuir o quantum para aumentar a quantidade de bits gerados. Como conseqüência da diminuição do quantum, tem-se uma imagem de melhor qualidade. Como em aplicações de videoconferência e videotelefonia não existe, em geral, muitas mudanças de cenas, o padrão é bem apropriado para elas.

O padrão H.263 [ITU-T H.263] estende o padrão H.261, introduzindo novos formatos de quadros, como discutimos anteriormente, otimizando o H.261 para codificação de vídeo a taxas inferiores a 64 Kbps e adicionando facilidades para maior qualidade e melhores serviços. Contudo, as idéias básicas de compressão são as mesmas.

4.8.2. MPEG Vídeo

Idêntico ao H.261, o padrão MPEG vídeo [ISO 93c e ISO 00a] divide um quadro em macroblocos, como apresentado na Figura 16, válida também para o MPEG com amostragem de crominância 4:2:0.

Cada bloco pode ser codificado usando apenas a informação intra-quadro, de forma idêntica ao que foi apresentado para o JPEG. Quadros em que todos os blocos são codificados dessa forma são denominados *quadros I*.

Um macrobloco pode também ser codificado de forma preditiva, semelhante ao H.261. No entanto, a predição MPEG é mais rica, uma vez que pode ser feita baseada em quadros passados e em quadros futuros da seqüência de um vídeo. Quando a predição é feita baseada em um quadro passado, tal qual o H.261, é codificado o erro de predição (diferença do macrobloco que se quer codificar para o macrobloco de referência do quadro passado), usando os mesmos procedimentos usados para os macroblocos intra-quadros, e o vetor de movimento (que dá a posição relativa do macrobloco que se quer codificar para o macrobloco de referência do quadro passado). Quadros codificados usando esse tipo de predição são chamados *quadros P*. A predição é sempre baseada no primeiro quadro do tipo I ou P, anterior.

A estimação do movimento (estimação do macrobloco mais próximo daquele que se quer codificar) se dá dentro de uma região do quadro de referência, conforme ilustra a Figura 17. Para TV, por exemplo, obtém-se bom desempenho com p=15 pixels em cenas comuns de noticiário e p=63 em cenas de muito movimento, como por exemplo, cenas de esporte.

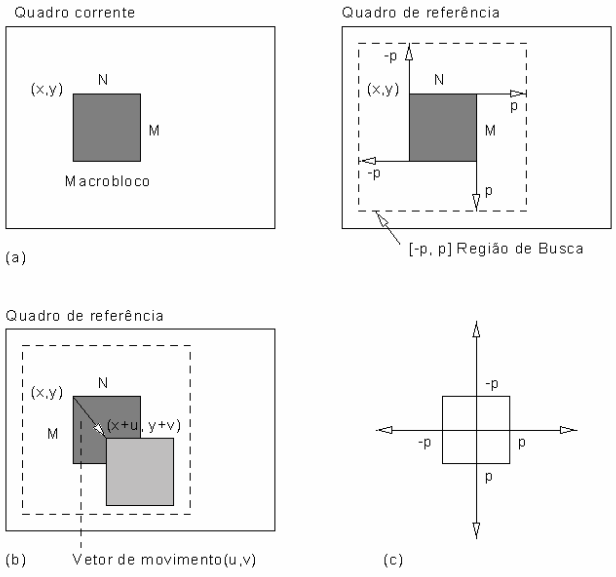


Figura 17: Estimação de movimento.

Macroblocos também podem ser codificados baseados no primeiro quadro I ou P, posterior ou anterior. Nesse caso, teremos dois quadros de referência para a procura do melhor casamento. A codificação pode ser realizada então baseada no quadro anterior, ou baseada no quadro posterior, ou ainda baseada na interpolação do quadro anterior e posterior. Quadros codificados usando esse tipo de predição são chamados *quadros B*.

Quadros B apresentam como vantagem o fato de, normalmente, apresentarem uma compressão maior que os quadros I e P (Tabela 4). São também quadros que, se perdidos, não afetam tanto a qualidade da imagem, pois o erro cometido não se propagará, uma vez que esses quadros não são usados como referência de predição (note que a perda de um quadro I ou P implica na perda de todos os quadros até o próximo quadro I). No entanto, quadros B introduzem um retardo extra no processo de codificação, porque devem ser codificados fora da seqüência, além de exigirem mais processamento para codificação.

Tabela 4: Tamanhos típicos de quadros MPEG 1.

Tipos de Quadro	Tamanho (Kbytes)	Compressão
I	18	2:1
P	6	7:1
B	2,5	50:1
Média	4,8	27:1

Ao contrário do H.261, que escolhe sempre a melhor codificação para o macrobloco, no MPEG, o padrão de codificação de quadros é um parâmetro da escolha do usuário, isto é, o usuário escolhe qual a seqüência de quadros I, P, ou B a ser utilizada. Note que essa escolha vai depender das aplicações. Por exemplo, como veremos na Seção 6, o retardo de transferência pode ser crítico em aplicações onde há uma interatividade em tempo real entre os participantes, como em um sistema de videoconferência. Nesse caso, o número de quadros B em seqüência não pode ser muito grande, devido ao retardo que isso introduz.

O padrão MPEG 1 trabalha, em geral, com o formato SIF, embora uma maior resolução também seja permitida por sua sintaxe. O padrão MPEG 2, como já mencionado, admite vários formatos de quadros e diferentes resoluções para as componentes de crominância. De fato, o MPEG 2 especifica vários conjuntos de parâmetros de restrição, que são definidos nos seus *perfis* e *níveis*. Um perfil especifica as facilidades de codificação que serão utilizadas. Um nível especifica a resolução dos quadros, as taxas de bits etc. Várias combinações de perfis e níveis foram definidas, como veremos mais adiante. O MPEG 2 usa os mesmos tipos de quadro I, P e B, como o MPEG 1, mas introduz outros métodos de predição [Netravali 95] para lidar com vídeo entrelaçado.[8]

O MPEG 2 também apresenta várias extensões de escalabilidade [ISO 00a]. As extensões provêem, basicamente, duas ou mais seqüências de bits, ou *camadas*, que podem ser combinadas para prover um único sinal de vídeo de alta qualidade. A *camada base* pode, por definição, ser totalmente decodificada por si mesma, de forma a prover um vídeo de baixa qualidade. Como veremos a seguir, muitas das técnicas empregadas são semelhantes às codificações progressivas e hierárquica do JPEG.

[8] Vídeo entrelaçado é tipicamente usado em TV, onde são primeiro varridas as linhas ímpares e depois as pares. Aos conjuntos de linhas ímpares e pares chamamos campos. Assim, um quadro é composto de dois campos.

Com a *escalabilidade SNR (Signal to Noise Ratio)*, outra camada pode ser adicionada à camada base, oferecendo uma melhora na precisão dos coeficientes da DCT (Discrete Cosine Transform), adicionando valores de correção para serem utilizados antes da decodificação (aplicação da DCT inversa). Essa extensão também provê a codificação do vídeo na resolução 4:2:2, tendo por camada base o vídeo na resolução 4:2:0.

Escalabilidade por partição de dados é uma versão simplificada da escalabilidade SNR. Com essa extensão, até um certo número de coeficientes de freqüências DCT é enviado pela camada base. Os coeficientes restantes são enviados por outra camada a ser adicionada à básica.

Na *escalabilidade espacial*, a camada base tem uma resolução espacial (resolução geométrica de cada imagem) menor do que a do vídeo original. A camada de melhoramento é então acrescida à camada base para se obter a resolução original.

Na *escalabilidade temporal*, a camada base tem uma resolução temporal (número de quadros por segundo) menor do que a do vídeo original. Novamente, a camada de melhoramento é então adicionada à camada base para a obtenção da resolução original.

Escalabilidade é especialmente útil em redes que permitem distinguir os tipos de fluxos e privilegiar a entrega do mais importante. Assim, em caso de necessidade ou conveniência de perda, um sinal de vídeo com um mínimo de qualidade ainda poderá ser recebido.

No MPEG 2, o perfil principal (*main profile*) utiliza os quadros I, P e B e uma amostragem de cor 4:2:0. O perfil simples (*simple profile*) é basicamente o perfil principal sem os quadros B. O perfil escalável SNR (*SNR Profile*) adiciona a escalabilidade SNR ao perfil principal. O perfil escalável espacialmente (*spatially scalable profile*) adiciona a escalabilidade espacial ao perfil escalável SNR. O perfil alto adiciona cor 4:2:2 ao perfil escalável espacialmente. Todos os perfis são limitados ao máximo de três camadas. O nível principal (*main level*), como mencionado no início desta seção, está definido basicamente para o vídeo ITU-R Rec. 601. O nível simples (*simple level*) é definido para vídeo SIF. Os dois níveis alto para HDTV são o nível alto-1440 (*high-1440 level*), com um máximo de 1440 pixels por linha, e o nível alto (*high level*), com no máximo 1920 pixels por linha. Nem todas as combinações de perfis e níveis foram padronizadas. No futuro, quando necessário, perfis e níveis poderão ser adicionados ao padrão.

Diferente da codificação linear de áudio e vídeo do MPEG 1 e 2, a codificação MPEG 4[9] [ISO 01a] é baseada em objetos, isto é, as cenas audiovisuais são codificadas em termos de objetos. Um objeto pode ser uma imagem, um vídeo ou um áudio.

Objetos codificados separadamente fornecem três benefícios: reusabilidade — a abordagem orientada a objeto permite aos autores reusarem material áudio-visual mais rapidamente; escalabilidade — objetos podem ser codificados usando diferentes resoluções espaciais e temporais (a resolução do objeto pode ser ajustada para casar com a capacidade do meio de transporte); interatividade — porque os objetos audiovisuais são compostos em quadros no decodificador, o usuário pode controlar a saída.

---

[9] MPEG 4, cuja designação formal do ISO/IEC é ISO/IEC 14496, foi finalizado em outubro de 1998 e tornou-se um padrão internacional nos primeiros meses de 1999. No final de 1999, foram acrescentadas novas extensões (MPEG-4 versão 2), tornando-se um padrão internacional formal no começo de 2000.

O objetivo inicial do MPEG 4 era a sua utilização em aplicações com baixas taxas de bits. Com as finalidades originais consideravelmente modificadas, um novo conjunto de aplicações usará o MPEG 4, tais como: videoconferência, comunicações móveis, acesso a vídeo de servidores remotos para aplicações multimídia, jogos etc. Atualmente, o grupo MPEG 4 está direcionando os trabalhos para televisão digital, aplicações gráficas interativas e *World Wide Web*.

O MPEG 4 considera uma cena como sendo composta de *Objetos de Vídeo (Video Objects) — VOs*. Os VOs têm propriedades como forma, movimento, textura etc. Eles vão se constituir nas entidades no fluxo de bits que o usuário pode manipular e ter acesso. Um *Plano de Objetos de Vídeo* (Video Object Plane - VOP) é uma ocorrência de um VO em dado instante de tempo. Cada quadro consiste de vários VOPs. Uma cena que contém somente um VOP pode corresponder aos padrões correntes, tais como MPEG 1 ou 2. Cada VOP tem sua própria resolução espacial e temporal.

Uma cena, dividida em objetos como mencionado, possui uma organização hierárquica. Uma informação adicional é enviada com os VOPs a fim de informar ao receptor como compor a cena. A descrição de cada cena baseia-se em conceitos da Virtual Reality Modeling Language (VRML – Linguagem de Modelagem de Realidade Virtual). Contudo, o padrão MPEG 4 introduziu novos conceitos de modelagem e otimizou os já existentes, dando origem a uma linguagem diferente e mais poderosa: *Binary Format for Scenes* (BIFS).

4.9. Multiplexação e Sincronização

Tanto o H.261 quanto o MPEG padronizam a forma como informações áudio-visuais devem ser multiplexadas (unidas em um único fluxo). No caso do MPEG, a padronização *MPEG System* [ISO 93b e ISO 00b] é responsável por essa especificação. Ambos os padrões adicionam aos fluxos elementares de áudio e vídeo informações para suas exibições sincronizadas. A sincronização é realizada pela adição de selos de tempo (*timestamps*) a conjuntos de amostras codificadas de vídeo e áudio, baseadas em um relógio compartilhado. A Figura 18 ilustra o procedimento.

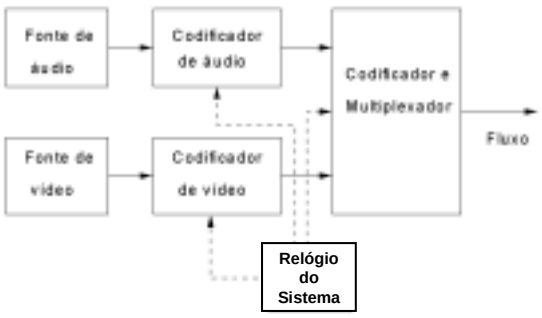


Figura 18: Multiplexação e sincronização dos fluxos de áudio e vídeo.

Um fluxo MPEG é organizado em duas camadas: a camada *pack* e a camada *packet*. A camada *pack* contém informações utilizadas por todos os fluxos elementares e a camada *packet* as informações particulares a cada fluxo. Um fluxo MPEG consiste de um ou mais

packs. O cabeçalho *pack* contém informações de temporização do sistema e sobre as taxas de dados. O cabeçalho *packet* contém a identificação do fluxo elementar, os requisitos de armazenamento e informações de temporização. Os dados *packet* contêm um número de bytes variável de um mesmo fluxo elementar. Assim, depois de remover o cabeçalho *packet*, os dados *packet* de todos os *packets* com o mesmo identificador são concatenados para a recuperação de um fluxo elementar. Até 32 fluxos de áudio e 16 fluxos de vídeo podem ser multiplexados em um fluxo MPEG. A Figura 19 apresenta a estrutura de camadas MPEG 1 System [ISO 93b].

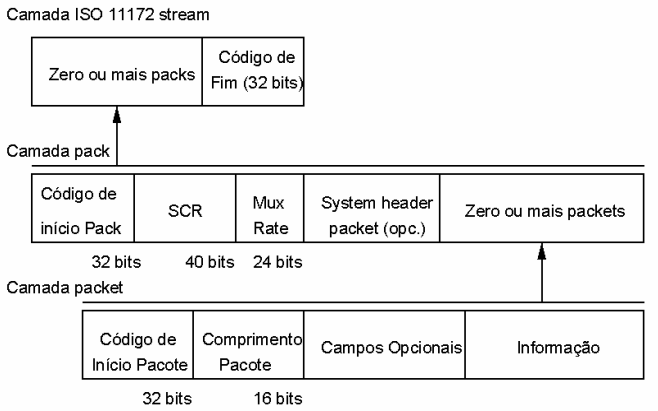


Figura 19: Camadas MPEG 1 System.

A função do MPEG 2 System [ISO 00b] é idêntica à do MPEG 1. Contudo, o MPEG 2 System especifica dois formatos de dados: o fluxo de programa (*program stream*) e o fluxo de transporte (*transport stream*) (Figura 20). O fluxo de programa é similar e compatível ao fluxo MPEG 1 System. Ele foi otimizado para aplicações multimídia e para ser processado por software.

O fluxo de transporte pode transportar múltiplos programas simultaneamente e é otimizado para aplicações onde a perda de dados é comum. O fluxo de transporte consiste de pacotes de tamanho fixo (188 bytes, incluindo 4 bytes de cabeçalho).

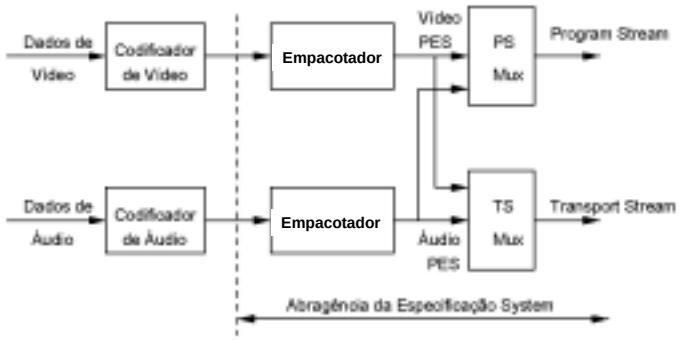


Figura 20: MPEG 2 System

Similar ao MPEG 1 e 2, o MPEG 4 System [ISO 01b] é desenvolvido para fornecer multiplexação de fluxos de dados elementares, sincronização e empacotamento. Adicionalmente, o MPEG 4 System fornece parâmetros de representação/manipulação básicos (translação, rotação e zoom) no cabeçalho da camada de fluxo de dados de cada objeto.

5. Aplicações de Banda Larga

Com o grande desenvolvimento da tecnologia digital, os diferentes tipos de informação (texto, áudio, vídeo etc.) passaram a ser processados de forma integrada, dando origem aos *sistemas multimídia*.

Tradicionalmente, os sistemas de comunicação foram desenvolvidos para o transporte de tipos específicos de informação — o sistema telefônico para o tráfego de voz; as redes de comutação de pacotes para dados textuais; vídeo e televisão em redes de radiodifusão ou a cabo. Essas redes foram claramente projetadas para aplicações específicas, normalmente adaptando-se mal a outros tipos de serviço. O ideal de uma única rede capaz de atender a todos esses serviços, para a obtenção de uma economia advinda do compartilhamento dos recursos, veio motivar o conceito das *redes de serviços integrados e diferenciados*.

Vários cenários foram delineados de forma a caracterizar os efeitos de aplicações multimídia em redes de alta velocidade e com integração de serviços. Imagine, por exemplo, médicos em hospitais localizados em regiões diferentes, discutindo o diagnóstico de um paciente usando informações de raio X, histórico da evolução do caso, vídeos, banco de dados sobre acompanhamentos de casos semelhantes etc. Suponha cada participante dessa teleconferência multimídia tendo acesso ao conjunto de tais informações, trocando comentários por voz e apontando os diversos itens sob discussão. Suponha cada participante podendo modificar, organizadamente, o conteúdo visível da tela e seu formato (selecionando uma nova imagem, ampliando-a na tela etc.). Suponha ainda que os fluxos de voz e as informações visuais sejam enviadas a todos os participantes simultaneamente, de forma que todo o sincronismo necessário (por exemplo, uma imagem em movimento sendo referenciada e comentada oralmente) seja preservado. Esse cenário serve para colocar em evidência que a infra-estrutura de comunicação necessária para essa aplicação deve prover serviços para realizar uma transmissão síncrona e em tempo real das informações multimídia. Como veremos na próxima seção, para complicar ainda mais essa aplicação já bastante complexa, as características e requisitos de comunicação dos diferentes tipos de informação variam muito.

O ITU-T define como *serviços de banda larga* (*broadband*) qualquer serviço que requer canais de transmissão capazes de suportar taxas maiores que aquelas do acesso primário das RDSI-FE (T1 = 1,544 Mbps, E1 = 2,048 Mbps).

O ITU-T classifica as aplicações em banda larga em quatro categorias: serviços conversacionais, serviços de recuperação, serviços de mensagem e serviços de distribuição.

Serviços conversacionais provêem meios para transferência fim a fim em tempo real. Entre as aplicações que aí se enquadram podemos citar: videotelefonia, videoconferência, transferência de documentos multimídia em tempo real (fac-símile de alta resolução, anotação de voz etc.), serviços de segurança, supercomputação virtual e teleação (controle por computador de dispositivos físicos remotos, associados a um controle de processo em tempo real). O ponto chave dessas aplicações é a interatividade em tempo real da comunicação entre os usuários dos serviços.

Serviços de recuperação fornecem a facilidade de recuperação de informação armazenada remotamente. Entre as aplicações que aí se enquadram podemos citar: videotexto, livrarias eletrônicas e vídeo sob demanda (tanto para entretenimento, pela substituição da TV a cabo, quanto para educação e treinamento remotos).

Serviços de mensagem oferecem a comunicação entre usuários via unidades de armazenamento, com funções de store-and-forward, mailbox (caixa de correio) ou manipulação de mensagens. Ao contrário dos serviços conversacionais, esses serviços não são em tempo real. Entre as aplicações que aí se enquadram podemos citar: correio de vídeo e correio de documentos multimídia.

Serviços de distribuição são subdivididos em duas classes: *serviços sem controle do usuário* e *serviços com controle do usuário*. Entre as aplicações sem controle do usuário podemos citar: distribuição de áudio e vídeo, distribuição de documentos (jornais, revistas e livros), distribuição de cotação da bolsa de valores e difusão de TV. Entre as aplicações com controle do usuário, podemos citar a substituição de documentos tradicionais (livros, revistas, jornais etc.), pelos equivalentes eletrônicos com distribuição sob controle do usuário.

O que caracteriza as aplicações de banda larga é o fato de terem de lidar com objetos não convencionais (áudio, vídeo etc.). Não convencionais no sentido de serem objetos longos (1 minuto de vídeo colorido não comprimido, qualidade TV, contém 1,8 Gbytes de dados), de exigirem transferência contínua de dados a altas taxas (162 Mbps para sinal de TV não comprimido) e, além de tudo, de exigirem acesso sincronizado aos dados. A próxima seção resumirá as características e requisitos de comunicação das diversas mídias envolvidas nas aplicações de banda larga.

6. Requisitos de Comunicação das Diversas Mídias

As características e requisitos de comunicação exigidos pelos diversos tipos de mídia são muito diferentes. Várias características devem ser consideradas ao classificarmos fontes de tráfego. A *natureza* do tráfego gerado é uma de suas características mais importantes, dando origem a três classes básicas: a classe de tráfego contínuo com taxa constante (Constant Bit Rate — CBR), a classe de tráfego em rajadas (bursty) e a classe de tráfego contínuo com taxa variável (Variable Bit Rate — VBR).

Na *classe de tráfego contínuo com taxa constante*,[10] o tráfego, como o próprio nome diz, é constante e, por conseguinte, sua taxa média é igual a sua taxa de pico. Essa taxa é o único parâmetro necessário para se caracterizar tal fonte.

As fontes cujo tráfego gerado tem característica de *rajadas* apresentam períodos ativos (durante os quais há geração de informação pela fonte, que opera na sua taxa de pico) intercalados por períodos de inatividade (durante os quais a fonte não produz tráfego algum). Para se caracterizar uma fonte com tráfego em rajadas não é suficiente utilizarmos a taxa média de geração de informação, já que essa taxa não representa corretamente o seu

[10] Em geral, os padrões de comunicação utilizam a palavra *contínuo* para caracterizar *sem interrupção* e a *taxas constantes*. Note, no entanto, que temos, além do tráfego em rajadas, o tráfego sem interrupção mas com taxa variável. Em geral, os padrões chamam apenas de *tráfego com taxa variável* (VBR) a ambos os tráfegos (em rajadas e contínuo com taxa variável), independente de serem sem interrupção ou não.

comportamento. A taxa média nem sequer representa uma taxa na qual a fonte opera em algum momento. Muito mais significativas são informações sobre a distribuição das rajadas ao longo do tempo, a duração das rajadas, e a taxa de pico atingida durante as rajadas. Alguns parâmetros comumente utilizados para caracterização desse tipo de tráfego incluem a *duração média dos períodos de atividade* e a *explosividade (burstiness)* da fonte — a razão entre a taxa de pico e a taxa média de utilização do canal.[11]

Por fim, as *fontes de tráfego contínuo com taxa variável* apresentam variações na taxa de transmissão ao longo do tempo. Parâmetros como a média e a variância da taxa de transmissão podem ser utilizados para caracterizar o comportamento de fontes com essa característica. O parâmetro de explosividade (burstiness) também é bastante utilizado na caracterização dessas fontes.

Requisitos sobre a qualidade do serviço de comunicação desejado (QoS), tais como retardo máximo de transferência, variação estatística do retardo (jitter), vazão média, taxas aceitáveis de erro de bit e de pacote de dados, variam muito de uma mídia para outra, e são dependentes da aplicação. De uma forma geral, podemos caracterizar as diversas mídias, quanto aos requisitos de comunicação exigidos, como se segue nos próximos parágrafos.

6.1. Texto

O tráfego gerado por informações em texto é, em sua grande maioria, de rajada. Para compreender essa natureza do tráfego, pense na comunicação de um terminal com um computador durante a execução interativa de um programa. A vazão média dos dados vai depender muito da aplicação, variando desde alguns poucos bits por segundo para aplicações de correio eletrônico, até alguns megabits por segundo em transferência de arquivos. Para texto, excetuando-se algumas aplicações em tempo real, como por exemplo para controle de processos críticos, o retardo máximo de transferência e a variação estatística do retardo não se constituem em problemas, sendo seus requisitos, em geral, facilmente satisfeitos pelo sistema de comunicação. Quanto à tolerância a erros, na grande maioria das aplicações, não se pode tolerar erro nem em um único bit: suponha, por exemplo, o caso da perda de um bit numa transferência eletrônica de fundos.

6.2. Imagem

O tráfego gerado em aplicações gráficas animadas, onde vários quadros são gerados em intervalos regulares de tempo, tem características bem semelhantes às da mídia de vídeo, comentadas mais à frente. Excetuando o caso de imagens animadas, a natureza do tráfego gerado pela mídia gráfica também é de rajadas, com vazões médias chegando a algumas dezenas de megabits por segundo. Como em textos, o retardo máximo e a variação estatística do retardo, em geral, não são relevantes.

Como discutido na Seção 2 as imagens gráficas podem estar no formato vetorial ou matricial. Para imagens no formato matricial e sem compressão, a taxa de erro de bit pode ser bem maior que a taxa de erro de pacote, uma vez que, em geral, não haverá nenhum problema se,

[11] Existem outras definições para a medida da explosividade da fonte: a razão entre o desvio padrão e a taxa média gerada, por exemplo.

por exemplo, um único pixel de uma tela ficar azul em vez de verde. O mesmo não se pode dizer da perda de um pacote, que poderá, por exemplo, apagar um bloco da imagem na tela. Para imagens no formato vetorial e imagens (vetoriais ou matriciais) onde foram utilizadas técnicas de compressão ou compactação, a tolerância à perda depende muito da aplicação e seus usuários. Como discutimos na Seção 4.6, existem métodos de compressão que identificam a porção mais importante dos dados de uma imagem. Para esses dados, deve-se evitar ao máximo as perdas. As porções menos importantes podem ser descartadas, se necessário (seja por erro na transmissão, por congestionamento no sistema de comunicação, ou mesmo porque o usuário final não necessita delas para obter a informação que deseja). Um sistema de comunicação deve poder identificar as porções que ele deve manter íntegras. Outro caso importante, com relação às perdas, são as imagens que não são processadas somente pelo olho humano, mas também pelo computador como, por exemplo, imagens médicas ou cartográficas. Nesse caso, a perda de um único bit (seja devido à comunicação ou ao método de compressão) pode ser intolerável (imagine uma doença que se quer diagnosticar através de uma imagem médica).

6.3. Áudio

A mídia de áudio tem características bem distintas das mencionadas nos dois parágrafos anteriores, principalmente em aplicações de tempo real com interatividade, como os serviços conversacionais do ITU-T. Começando pela natureza do tráfego gerado, a mídia de áudio se caracteriza por gerar um tráfego contínuo com taxa constante. Mesmo quando no sinal de voz é realizada a compactação por detecção de silêncio, por exemplo, passando a se caracterizar agora como um tráfego de rajada [Gruber 82], ele deve ser reproduzido no destino a uma taxa constante. O tráfego gerado para comunicação dessa mídia é do tipo CBR, caso não seja empregada nenhuma técnica de compactação ou compressão. Em caso contrário, o tráfego se caracteriza como VBR e, às vezes, como no caso da voz com detecção de silêncio, como um tráfego em rajadas.

A vazão média gerada pela mídia de áudio depende da qualidade do sinal, da codificação e compactação ou compressão utilizadas. Para sinais de voz, por exemplo, já apresentamos a técnica PCM, que gera 64 Kbps se utilizarmos 8 bits para codificar cada amostra (tomada a cada 125 μseg, isto é, 8.000 amostras por segundo). Com qualidade aproximadamente igual, a codificação ADPCM gera 32 Kbps. Sinais de áudio de alta qualidade (qualidade de CD estéreo, por exemplo) geram taxas bem superiores, como, por exemplo, os CDs de áudio, onde a taxa é de 1,411 Mbps, como vimos na Seção 3.1.

Quanto às perdas, as taxas de erros de bits ou de pacotes podem ser relativamente altas, devido ao alto grau de redundância presente nos sinais de áudio. O único requisito é que os pacotes não sejam muito grandes (no caso da voz, menores que uma sílaba), o que normalmente já é satisfeito para não se perder tempo no empacotamento e assim não aumentar o retardo de transferência. Perdas da ordem de 1% da informação de voz são toleráveis[12] [Gopal 84, Gruber 85]. Uma vez que as redes de alta velocidade utilizam, hoje em dia, meios físicos de alta confiabilidade (como fibra ótica, por exemplo, onde a taxa de erro típica é de 10^{-9} ou menos) a detecção de erros para a voz nessas redes pode ser tranqüilamente

[12] Na realidade, como já comentamos, a percentagem de perda depende do tamanho do surto de voz e se a perda ocorre no início ou no meio do surto.

dispensada, em benefício de um maior desempenho. Apesar da baixa taxa de erros das redes de fibra ótica, nas mídias gráfica e de texto a detecção de erros ainda é, na maioria das vezes, necessária, e em alguns casos até a detecção e correção. Um cuidado adicional deve ser tomado quando, devido às técnicas de compressão utilizadas no áudio, um erro pode se propagar para outros bits. Nesse caso, o erro pode ser intolerável. Ainda com respeito ao áudio, porções da informação podem ser diferenciadas quanto à tolerância às perdas. No caso da voz, por exemplo, perdas nos intervalos de silêncio são muito mais toleráveis do que perdas durante os surtos de voz. Um sistema de comunicação deve poder identificar as porções mais sensíveis a perdas, caso seja necessário o descarte de dados.

No caso da utilização de sistemas de comunicação que apresentam variação estatística do retardo (como as redes comutadas por pacotes), tal variação deve ser compensada. Para entendermos melhor o problema, analisemos a Figura 21.

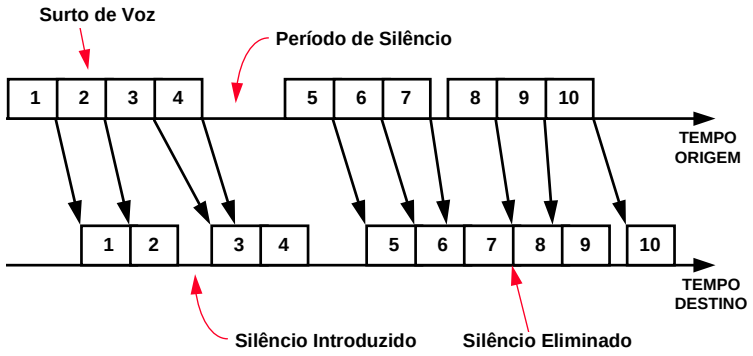


Figura 21: Efeito da variação estatística do retardo na comunicação de voz.

Na linha horizontal superior vemos os surtos de voz e de silêncio sendo gerados na fonte a uma taxa constante. Os surtos de voz são divididos em pacotes, que são as unidades que transitarão no sistema de comunicação (os surtos de silêncio não são transmitidos). Uma vez que o pacote é gerado, ele é imediatamente entregue para transmissão (veremos a seguir que o retardo máximo é um requisito importante). Se os pacotes sofrerem retardos variáveis, chegarão ao destino não mais preservando a continuidade, conforme mostra a linha horizontal inferior da figura, podendo gerar intervalos de silêncio dentro de um surto de voz, ou diminuir, e até mesmo eliminar, intervalos de silêncio, o que pode causar a perda da inteligibilidade da informação no destino. Alguma forma de compensação dessa variação estatística do retardo deve ser realizada.[13] A estratégia utilizada pelos algoritmos de compensação baseia-se fundamentalmente em assegurar uma reserva de pacotes antes de dar início ao processo de reprodução, introduzindo um retardo inicial a cada surto de voz. Aparentemente o problema estaria resolvido se escolhêssemos o retardo inicial bem grande,

[13] Note que a variação estatística do retardo não é necessariamente introduzida só pela rede de comunicação, mas por todo o sistema. Ela é introduzida desde a interação da placa de áudio com o sistema operacional da estação, passando pelos protocolos de comunicação (sistema operacional de rede), até chegar ao sistema de transmissão. No destino, o caminho semelhante, mas em ordem inversa, também pode introduzir aleatoriedade no retardo antes da reprodução. Assim, embora muitas vezes o sistema de transmissão não introduza aleatoriedade no retardo, a compensação ainda deve ser feita.

entretanto, o valor desse retardo está limitado pelo máximo retardo de transferência (desde a geração até a reprodução) permitido para o sinal de voz, sem que haja perda da interatividade da comunicação [Bastos 92]. As referências [Soares 91, Bastos 92, Soares 92, Gopal 84, Adams 85, e Faria 92] discutem com algum detalhe a análise de desempenho de vários algoritmos para compensação da variação estatística de retardo, com o objetivo de manter a continuidade em sinais de voz. Embora apresentados para sinais de voz, os algoritmos podem ser facilmente estendidos para qualquer sinal contínuo (com taxas constante ou variável). O Apêndice C apresenta uma discussão resumida desses algoritmos.

O retardo de transferência máximo é crítico para o áudio, principalmente no caso de conversações. Um dos motivos é devido ao problema do eco [Handel 89], mas, mesmo nos casos em que o eco não cause problemas, ou que canceladores de eco sejam utilizados, o retardo de transferência máximo pode ser crítico. Cada interlocutor, em uma conversação, normalmente espera o fim do discurso do outro para dar início à sua fala; se o retardo de transferência for muito grande, a conversação começa a sentir um efeito de ruptura, podendo até se tornar inviável (se o leitor já utilizou a rede telefônica via satélite, deve ter sentido esse efeito). Um retardo de transferência maior que 200 ms já começa a incomodar os interlocutores [Bastos 92]. Os padrões de telefonia estipulam 40 ms para distâncias continentais e 80 ms para distâncias intercontinentais como limites para o retardo máximo de transferência. É bom frisarmos novamente que os problemas de retardo só são críticos em aplicações que exigem comunicação interativa em tempo real. Nesse caso, como não podemos introduzir um retardo inicial muito grande para compensarmos a variação estatística do retardo, a compensação só poderá ser efetiva se a variação estatística apresentada for pequena.

6.4. Vídeo

Tal qual a mídia de áudio, a mídia de vídeo se caracteriza por gerar um tráfego contínuo com taxa constante. Da mesma forma que no áudio, mesmo quando no sinal é realizada alguma técnica de compactação ou compressão e o tráfego gerado para comunicação se caracterizar como um tráfego com taxas variáveis, o sinal deve ser reproduzido no destino a uma taxa constante. Como na mídia de áudio, o retardo de transferência máximo tem grande importância, e a variação estatística do retardo deve ser compensada. Normalmente, como o vídeo vem acompanhado de (sincronizado com) áudio, uma vez obedecidos os requisitos de retardo desse, estão obedecidos os daquele.

A vazão média gerada por uma fonte de vídeo varia com a qualidade do sinal e os algoritmos de codificação, compactação e compressão empregados, conforme discutido na Seção 4.8.

Em vídeo, a taxa de erro de bit pode ser maior que a taxa de erro de pacote, pelos mesmos motivos explicitados para as imagens gráficas no formato matricial. No entanto, no vídeo, como a imagem não é estática e devem ser gerados vários quadros por segundo, a taxa de erro de pacote não é tão crítica. Mesmo a taxa de erro de bit tolerável é maior do que aquela para imagens estáticas [Hehmann 90]. Na verdade, a discussão sobre a taxa de erro aceitável não é tão simples. Quando utilizamos técnicas de compressão, um erro pode se propagar. Dessa forma, alguns quadros, em que o erro não se propaga, podem tolerar erros de bits e de pacotes. Naqueles em que o erro se propaga, às vezes até um único erro de bit pode ser intolerável. Tal qual nas imagens no formato matricial, quando se utiliza técnicas de compressão ou compactação, a tolerância a perda depende muito da aplicação e seus usuários. Como discutimos na Seção 4.8, existem métodos de compressão que identificam a porção mais importante dos dados de um vídeo. Para esses dados, deve-se evitar ao máximo as

perdas, as porções menos importantes podem ser descartadas, se necessário (ou por erro na transmissão, ou por congestionamento no sistema de comunicação, ou mesmo porque o usuário final não necessita delas para obter a informação que deseja). Mais uma vez, um sistema de comunicação deve poder identificar as porções em que ele deve minimizar as perdas.

7. Considerações Finais

As referências [Hehmann 90, Gruber 82, Soares 92] apresentam várias características das diversas mídias para diferentes aplicações e técnicas de compressão e compactação utilizadas.

O que vimos na Seção 6 foi um grande número de combinações de características e requisitos que um sistema de comunicação e processamento devem satisfazer de forma a tornar possíveis as aplicações multimídia. Naquela seção, foram mostrados apenas alguns requisitos, sem contudo se fazer menção às diferenças no que concerne ao gerenciamento de buffers, controle de fluxo, gerenciamento de conexões. Não foram sequer mencionadas as direções para a otimização de compromissos no atendimento dos vários requisitos de diferentes aplicações. Pelo exposto, fica claro que um desafio para a construção de qualquer sistema de comunicação e processamento é fornecer mecanismos para oferecer o suporte a essas diversas características de tráfego. Esses mecanismos devem permitir negociar a qualidade de serviço desejada (QoS), isto é:

- O máximo retardo de transferência;
- A variação máxima de retardo para o atendimento dos requisitos de especificação de áudio e vídeo;
- Os mecanismos para compensação da variação estatística do retardo;
- O valor da vazão necessária para a abertura de uma comunicação.
- As taxas de erros de bit e pacote toleráveis (separadamente negociadas);
- A especificação de qual estratégia a ser adotada, para todos os tipos de erro: detecção, detecção e correção, ou nada;
- Os mecanismos para controle do fluxo de dados e congestionamento do sistema de comunicação.
- As condições para o fechamento de uma comunicação, caso não seja possível atender aos requisitos.

Referências

[Adams 85] Adam, C. e Ades, S. “Voice Experiments in the UNIVERSE Project”. Proceedings of International Conference on Communications. 29.4.1 - 29.4.9, 1985.

[Bastos 92] Bastos, T.L.P. e Soares, L.F.G. “Análise de Algoritmos para Reprodução em Tempo Real de Voz em Redes de Pacotes”. *Relatório Técnico IBM CCR-141*, Rio de Janeiro. Janeiro, 1992.

[Cormen 02] Cormen, T.H.;Leiserson, C.E.; Rivest, R.L.; Stein, C. *Algoritmos*. Tradução da 2ª edição americana Teoria e Prática. 2002.

[Faria 92] Faria, A.L.A. "Implementação do Mecanismo de Controle de Acesso por Detecção de Silêncio em um Sistema de Teleconferência". *Dissertação de Mestrado, Depto. de Engenharia Elétrica, PUC-Rio*. Março, 1992.

[Gopal 84] Gopal, P.M., Wong, J.W. e Majithia, J.C. "Analysis of Playout Strategies for Voice Transmission Using Packet Switching Techniques". *Performance Evaluation*, n.4. Fevereiro, 1984

[Gruber 82] Gruber, J.G. "A Comparison of Measure and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection". *IEEE Transactions on Communications*, vol. com-30, n.4. Abril, 1982.

[Gruber 85] Gruber, J.G. "Subjective Effects of Variable Delay and Speech Loss in Dinamically Managed Voice Systems". *IEEE Transactions on Communications*, vol. com-33. Agosto, 1985.

[Handel 89] Handel, R. "Evolution of ISDN Towards Broadband ISDN". *IEEE Network Magazine*, pg 7-13. Março, 1989.

[Hehmann 90] Hehmann, D. B., M.G. Salmony, and H.J. Stuttgen. "Transport services for multimedia applications on broadband networks." Computer Communications Vol 13 No. 4, 1990, Pages 197-203.

[ISO 93a] ISO/IEC. "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - Part 3: Audio". ISO/IEC 11172-3. 1993.

[ISO 93b] ISO/IEC. "Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - Part 1: Systems". ISO/IEC 11172-1. 1993.

[ISO 93c] ISO/IEC. "Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - Part 2: Video". ISO/IEC 11172-2. 1993.

[ISO 94] ITU-T Recommendation T.81. "Joint Photographic Experts Group", 1993.

[ISO 98] ISO/IEC. "Information technology - Generic coding of moving pictures associated audio information: Audio". ISO/IEC 13818-3. 1998.

[ISO 00a] ISO/IEC. "Information technology - Generic coding of moving pictures and associated audio information: Video". ISO/IEC 13818-2. 2000

[ISO 00b] ISO/IEC. "Information technology - Generic coding of moving pictures and associated audio information: Systems". ISO/IEC 13818-1. 2000

[ISO 01a] ISO/IEC. "Coding of Audio-Visual Objects – Part 2: Video". ISO/IEC 14496-2. 2º Edition. 2001.

[ISO 01b] ISO/IEC. “Coding of Audio-Visual Objects – Part 1: Systems”. ISO/IEC 14496-1. 2º Edition. 2001.

[ITU-R BT.601-4] ITU-R. “Encoding parameters of digital television for studios”. ITU-R BT.601-4. 1994.

[ITU-T G.711] ITU-T. “Pulse Code Modulation (PCM) of Voice Frequencies”. ITU-T G.711. 1988.

[ITU-G.722] ITU-T. “7 kHz Audio-Coding Within 64 kbit/s”. ITU-G.722. 1988.

[ITU-T G.723] ITU-T. “Speech Coders – Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s”. ITU-T G.723. 1996.

[ITU-T G.726] ITU-T. “40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)”. ITU-T G.726. 1990.

[ITU-T G.729] ITU-T. “Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CSACELP)”. ITU-T G.729. 1996.

[ITU-T H.261] ITU-T. “Video Codec for Audiovisual Services at p x 64 kbit/s”. ITU-T H.261. 1993.

[ITU-T H.263] ITU-T. “Video Coding for Low Bit Rate Communication”. ITU-T H.263. 2005.

[Netravali 95] Netravali, A.N.; Haskell, B.G. “Digital Pictures: Representation, Compression and Standards”. Springer. 1995

[Soares 91] Soares, L.F.G., Martins, S.L. e Bastos, T.L.P. “Um Algoritmo para Compensação da Variação Estatística do Retardo em Redes Comutadas por Pacotes”. *Anais do 8º Simpósio Brasileiro de Redes de Computadores*. 1991.

[Soares 92] Soares, L.F.G. e Bastos, T.L.P. “Análise de Algoritmos para Reprodução em Tempo Real de Voz em Redes de Pacotes”. *Anais do 10º Simpósio Brasileiro de Redes de Computadores*, Recife. 1992.