

Áudio



PUC-Rio / DI

TeleMídia

Áudio

- ➔ Nós usaremos o termo áudio para nos referirmos à mídia de representação.
- ➔ Nós usaremos os termos voz e fala, indiferentemente, para nos referirmos ao sinal codificado pela mídia áudio, originalmente produzido pela voz humana, ou um sintetizador de voz humana.
- ➔ Nós usaremos o termo som de alta qualidade para nos referirmos ao sinal codificado pela mídia áudio, proveniente de outras fontes sonoras, tais como instrumentos, discos, etc.



PUC-Rio / DI

TeleMídia

Limitações da Mídia

- ▶ Fala é lenta → um orador passa informação numa velocidade muito mais lenta que um bom leitor poderia obter de um livro ou tela.
- ▶ Fala é seqüencial → olhos podem divagar em um menu para sua exploração, o ouvido não.

Limitações da Tecnologia

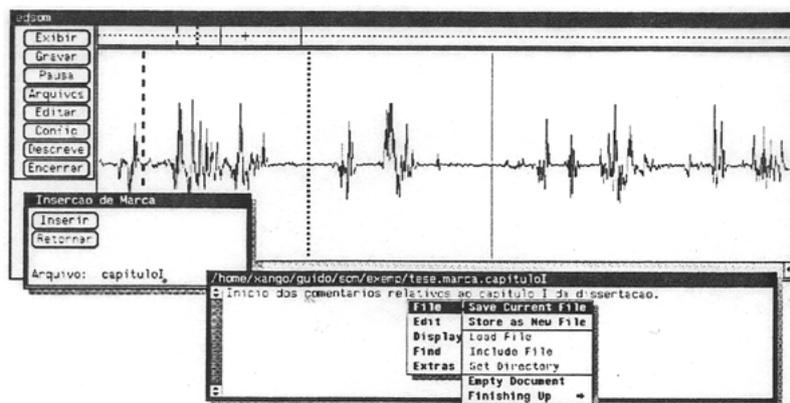
- ▶ Dificuldade ainda de inteligibilidade da fala sintética (cada vez menos).
- ▶ Reconhecimento de voz ainda na infância → trabalhar em ambientes, mesmo com o mínimo ruído acústico, é quase impossível.



PUC-Rio / DI

TeleMídia

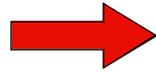
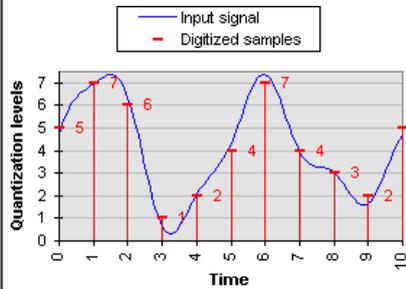
Interfaces de Voz



PUC-Rio / DI

TeleMídia

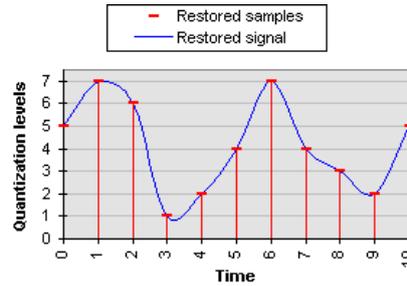
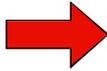
PCM - Digitalização de Sinais



Codificando cada nível com 3 bits:

101 111 110 011 010 100 111 100 ...

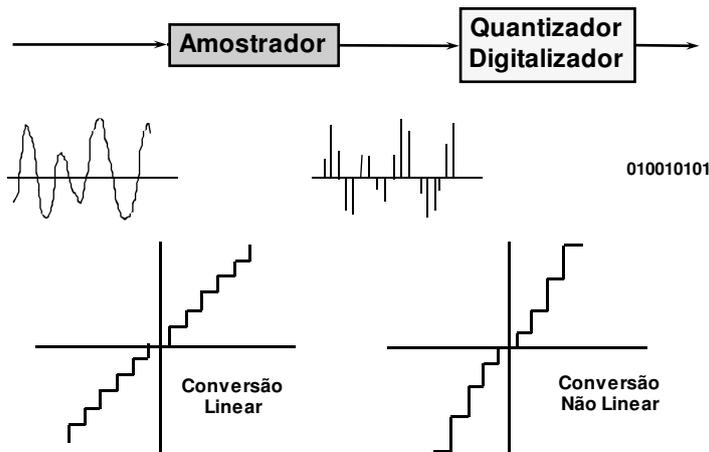
Na recepção...



PUC-Rio / DI

TeleMídia

Conversão Analógica Digital



PUC-Rio / DI

TeleMídia

Codificação PCM Compansão do Sinal

➔ Lei A (Europa)

$$y(x) = \begin{cases} \frac{1 + \ln Ax}{1 + \ln A} & \frac{1}{A} < x < 1 \\ \frac{Ax}{1 + \ln A} & 0 < x < \frac{1}{A} \end{cases}$$

$A = 100$

➔ Lei μ

$$y(x) = \frac{\ln(1 + \mu x)}{\ln(1 + \mu)} \quad 0 < x < 1$$

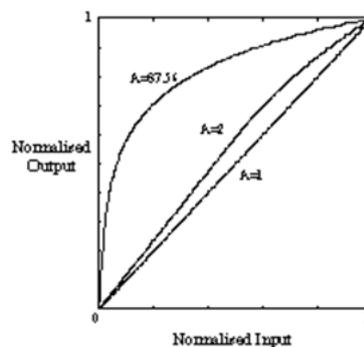
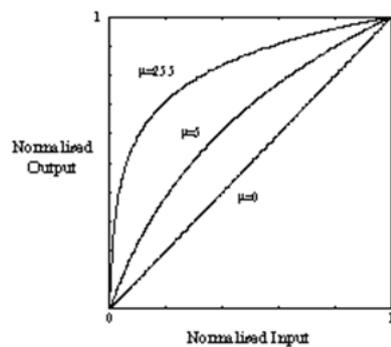
$\mu = 255$



PUC-Rio / DI

TeleMídia

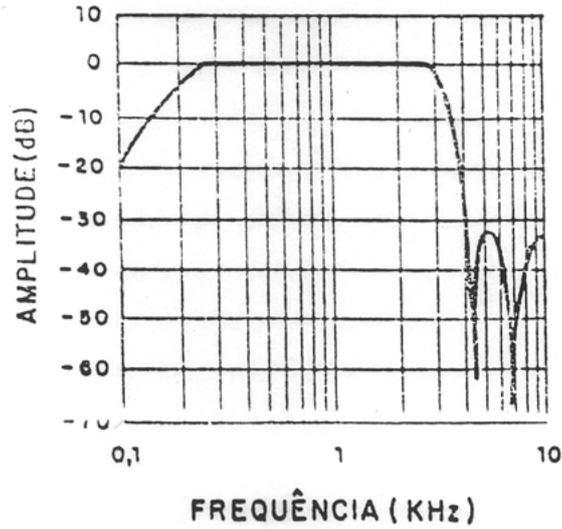
Compansão



PUC-Rio / DI

TeleMídia

Voz



PUC-Rio / DI

TeleMídia

Codificação de Áudio

| Qualidade | Banda (Hz) | Taxa de Amostragem | Bits por Amostra | Mono/ stereo | Taxa |
|-----------|------------|--------------------|------------------|--------------|------------|
| Telefonia | 300 - 3400 | 8000 | 8 (13 linear) | mono | 64 Kbps |
| AM | | 11025 | 8 | mono | 88,2 Kbps |
| FM | | 22050 | 16 | stereo | 705,6 Kbps |
| CD | 0 - 21000 | 44100 | 16 | stereo | 1,411 Mbps |
| DAT | | 48000 | 16 | stereo | 1,536 Mbps |



PUC-Rio / DI

TeleMídia

Codificadores de Onda

- ➔ **PCM** : N bits representando a amplitude do sinal.
- ➔ **DM** : 1 bit indicando o aumento ou a diminuição do sinal de um valor de quantização Q.
- ➔ **DPCM** : s bits representando o quociente entre a variação V entre o sinal de duas amostras consecutivas e o valor de quantização Q.
- ➔ **ADPCM** : Semelhante ao anterior, com Q variando dinamicamente de acordo com a taxa de variação do sinal.



PUC-Rio / DI

TeleMídia

Codificadores de Voz

➔ **LPC**

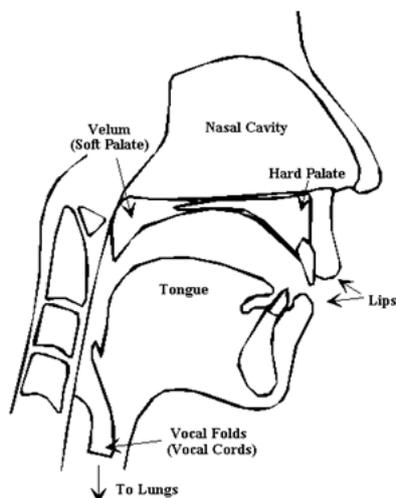
- Modelo do trato vocal humano baseado em uma série de coeficientes calculados a partir do sinal de VOZ.



PUC-Rio / DI

TeleMídia

Trato Vocal



PUC-Rio / DI

TeleMídia

CELP

- O codificador CELP (Code Excited Linear Predictor) gera os mesmos parâmetros LPC, mas então 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.



PUC-Rio / DI

TeleMídia

Sinais de Áudio

➔ Padronização ITU-T para codificação de voz

- **G.711** (1972)
 - Telefonia
 - PCM 64 Kbps
 - MOS 4 → 5
- **G.721** (1986)
 - ADPCM 32 Kbps
 - MOS 4
- **G.722** (1987)
 - Banda passante de 7 KHz
 - 64 Kbps
 - Atraso de 3ms
 - MOS 4 para música
 - MOS 4,3 para voz
- **G.723**
 - ADPCM 24 Kbps e 40 Kbps
- **G.723.1** (03/1996)
 - Aplicações Multimídia
 - 5.3 a 6.3 Kbps (03/1996)
- **G.728** (09/1992)
 - Voz
 - Predição Linear
 - 16 kbps
- **G.729** (03/1996)
 - Voz
 - 8 kbps
 - Codificação algébrica e predição linear



Comparação de CODECs

| 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 de 50 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.728 | LD-CELP | 16 | Muito Alto | Boa | Baixo |
| G.729 | CS-ACELP | 8 | Alto | Boa | Baixo |



CODECs

- ➔ A UIT-T, responsável pela qualidade da voz, definiu o G.729 como seu padrão para a transmissão de voz sobre redes de dados.
- ➔ O IMTC (International Multimedia Teleconferencing Consortium) escolheu o G.723.1 como padrão. O G.729 é o codificador alternativo escolhido pelo IMTC.



PUC-Rio / DI

TeleMídia

Detecção de Silêncio

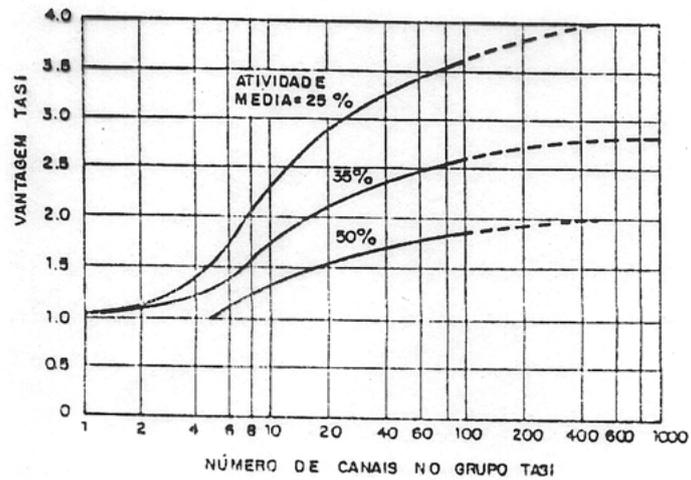
- ➔ Compactação para armazenamento
- ➔ Aumento da eficiência da utilização de canais de comunicação (vantagem TASI)
- ➔ Base para mecanismos de controle especiais
 - por exemplo, controle de acesso a um aplicativo compartilhado.



PUC-Rio / DI

TeleMídia

Vantagem TASI



PUC-Rio / DI

TeleMídia

Deteção de Silêncio

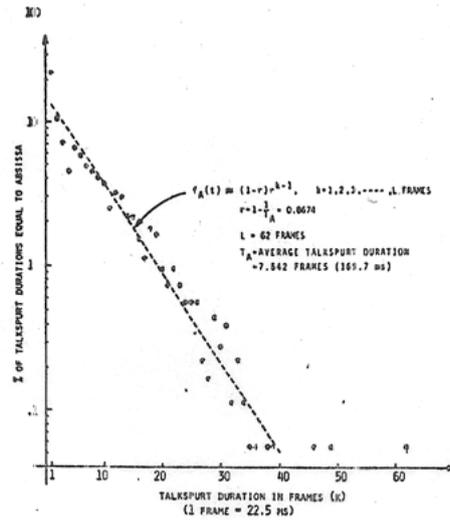
- ➔ Compactação para armazenamento
- ➔ Aumento da eficiência da utilização de canais de comunicação (vantagem TASI)
- ➔ Base para mecanismos de controle especiais
 - por exemplo, controle de acesso a um aplicativo compartilhado.



PUC-Rio / DI

TeleMídia

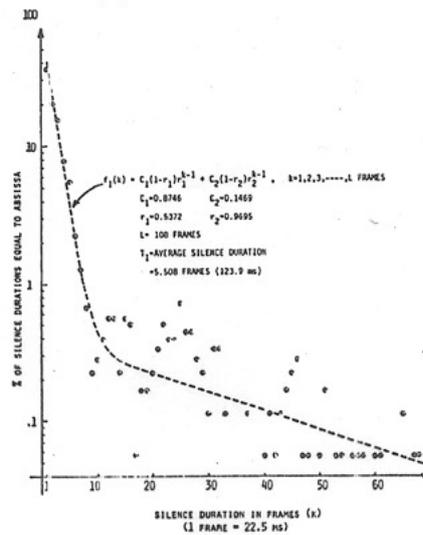
Densidade da Duração do Surto de Voz (sem hangover e sem fill-in)



PUC-Rio / DI

TeleMídia

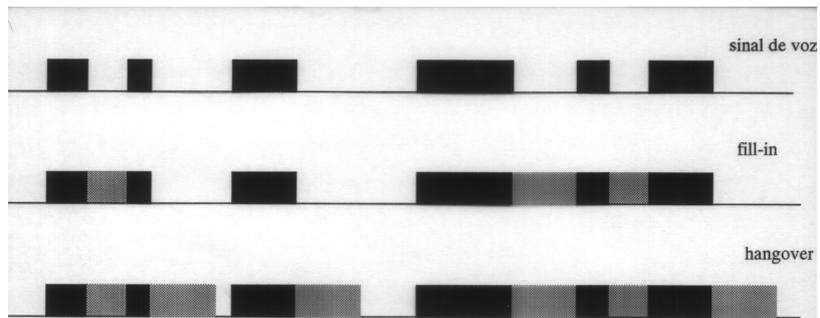
Densidade da Duração do Silêncio (sem hangover e sem fill-in)



PUC-Rio / DI

TeleMídia

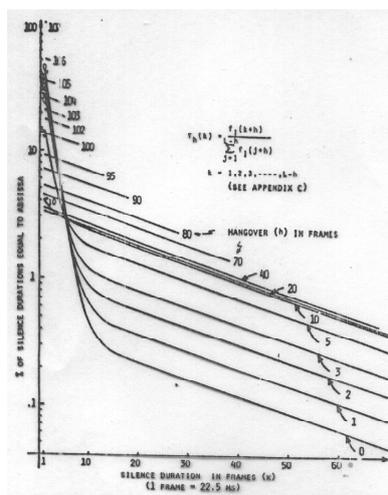
Hangover e Fill-in



PUC-Rio / DI

TeleMídia

Silêncio x Hangover



PUC-Rio / DI

TeleMídia

Hangover e Fill-in

➔ Vantagens:

- Na compensação da variação estatística do retardo, a alteração do período de silêncio só se dá nos períodos relativamente longos (entre frase e sentenças), onde são menos perceptíveis que os períodos pequenos (entre palavras e sílabas).
- Diminuição do corte final das palavras (clipping) de surtos de voz na detecção de silêncio.
- Redução do overhead de chaveamento em redes comutadas por circuito.

➔ Desvantagem:

- Aumento da atividade de voz (pior no hangover).



PUC-Rio / DI

TeleMídia

Detecção de Silêncio - Considerações

➔ Ruído de fundo

➔ Período de classificação → atua sobre segmentos, pois uma amostra isolada não é suficiente para a classificação do sinal. O ideal é que os intervalos de silêncio sejam múltiplos do pacote usado para transmissão do sinal)

➔ Hangover

➔ Reprodução do sinal

- Duração dos intervalos de silêncio devem ser mantidos na reprodução
- Reprodução de algum som no silêncio (reprodução do ruído de fundo)
- Caso se esteja utilizando a técnica de hangover, o final do período de fala (composto de segmentos de silêncio) é repetido durante o tempo de duração do intervalo de silêncio.
- Armazenamento de ruído branco, que é reproduzido multiplicado por um fator que indica o nível de ruído do ambiente (calculado a partir de amostras de silêncio transmitidas periodicamente)



PUC-Rio / DI

TeleMídia

Detecção de Silêncio - Indicadores

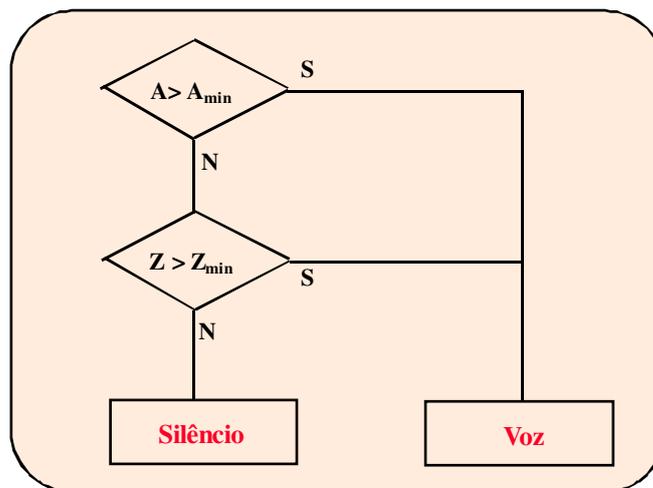
- ▶ Nível de energia → Utiliza-se normalmente o modulo da amplitude ao invés da função quadrada, a fim de simplificar a implementação, diminuir o tempo de execução e realçar menos as amostras de grande amplitude, quando comparadas às amostras de pequena amplitude.
- ▶ Taxa de cruzamento do zero → Expressa pelo número de vezes que as amostras de voz trocam de sinal durante um determinado período. (É um indicador eficiente para fonemas que apesar de terem um nível baixo de energia, possuem o espectro concentrado nas frequências altas - como as consoantes fricativas fracas, no começo e fim de palavras).
- ▶ Taxa de variação da amplitude do sinal → sinal de voz é não estacionário - energia de segmentos sucessivos tem grande variação - enquanto o ruído costuma ser estacionário.
- ▶ Outros.



PUC-Rio / DI

TeleMídia

Detecção de Silêncio Algoritmo Básico



PUC-Rio / DI

TeleMídia

Algoritmos para Detecção de Silêncio Abordagem Estatística

- Reconhecimento de padrão → determina-se o padrão de silêncio e verifica-se se uma determinada amostra (segmento) se encaixa ou não no padrão.
- Um segmento de voz é representado por p parâmetros, através de um vetor de p dimensões.
- A partir de n segmentos de silêncio (x_1, x_2, \dots, x_n) obtém-se a média e a covariância:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$



PUC-Rio / DI

TeleMídia

Algoritmos para Detecção de Silêncio Abordagem Estatística

- Seja x um segmento do sinal de voz codificado. Calcula-se o teste estatístico:

$$F = \frac{n \cdot (n - p)}{p \cdot (n^2 - 1)} \cdot (x - \bar{x})' S^{-1} (x - \bar{x})$$

- Quanto menor o valor de F , mais o segmento x se aproxima do padrão de silêncio. Um segmento será considerado silêncio se F não ultrapassar um valor máximo especificado.
- Os segmentos de silêncio para representar o padrão são continuamente renovados a partir do sinal sendo processado com cada k segmentos contíguos de silêncio, substituindo os segmentos mais antigos dos n anteriores.
- Quanto aos segmentos iniciais para compor o padrão, o algoritmo admite que o período inicial de um sinal de voz é silêncio.



PUC-Rio / DI

TeleMídia

Detecção de Silêncio – Algoritmo 1

- ➔ Seja IA o segmento de voz corrente, e IB e IC os segmentos imediatamente anteriores.
- ➔ Seja ainda:
 - THU - limite superior do nível de amplitude
 - THL - limite inferior do nível de amplitude
 - NTHU - limite superior da taxa de cruzamento do zero
 - NTHL - limite inferior da taxa de cruzamento do zero



PUC-Rio / DI

TeleMídia

Detecção de Silêncio – Algoritmo 1

1. Inicia $IB = IC = \text{silêncio}$
2. Para cada segmento codificado:
 - 2.1 Calcula A e Z para o segmento
 - 2.2 Se $(A > THU \text{ ou } Z > NTHU)$ ou $(A > THL \text{ e } Z > NTHL)$
então IA = fala
senão IA = silêncio
 - 2.3 Se pelo menos dois entre IA, IB e IC forem fala
então classifica o segmento como fala
senão classifica o segmento como silêncio
 - 2.4 $IC = IB, IB = IA$



PUC-Rio / DI

TeleMídia

Cálculo dos Indicadores

$$A = \frac{1}{100} \sum_{i=1}^{100} |x_i|$$

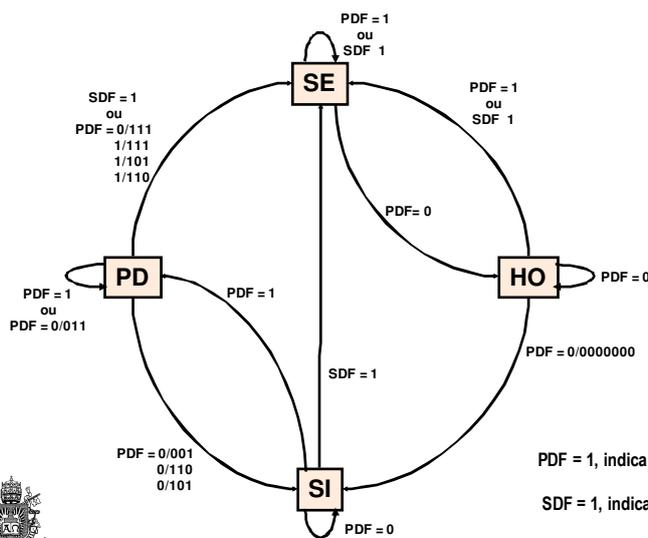
$$Z = \sum_{i=1}^{100} \left[1 - \frac{x_i \cdot x_{i-1}}{|x_i \cdot x_{i-1}|} \right]$$



PUC-Rio / DI

TeleMídia

Detecção de Silêncio – Algoritmo 3 Verificar na tese da Ana

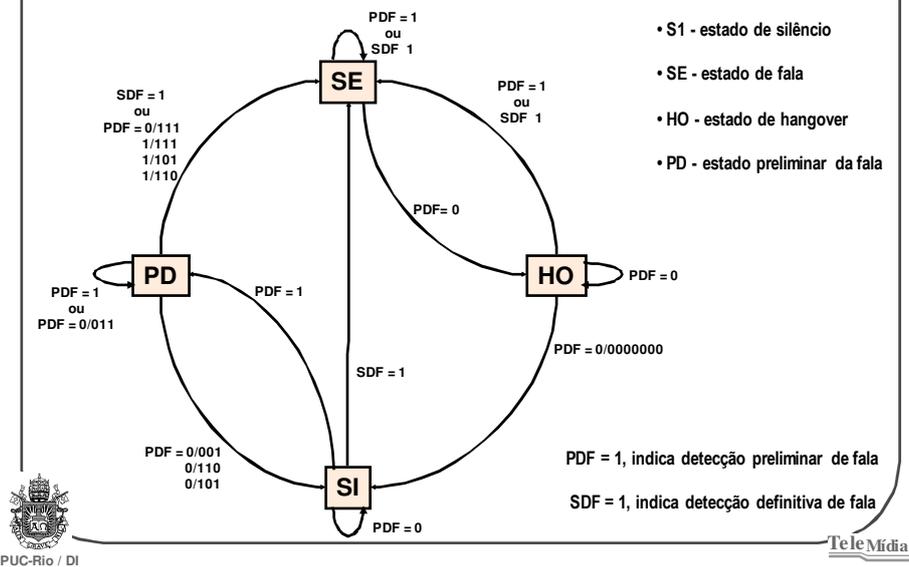


PUC-Rio / DI

TeleMídia

Detecção de Silêncio – Algoritmo 3

Verificar na tese da Ana

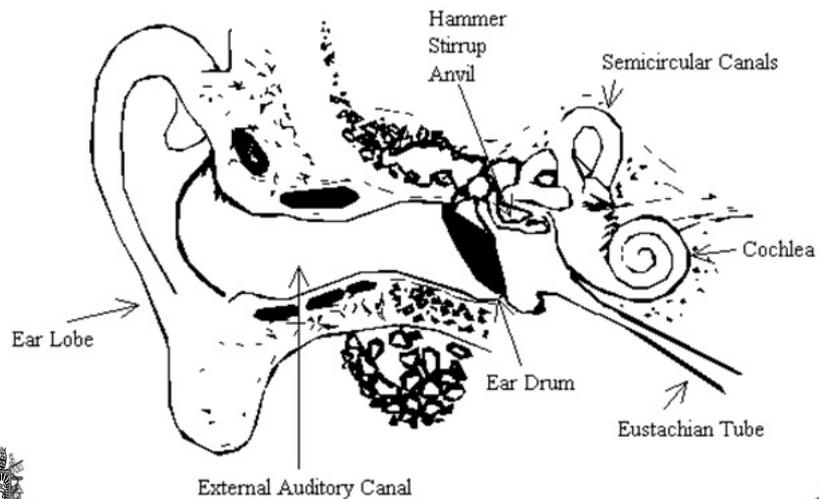


Detecção de Silêncio – Algoritmo 3

➔ Por exemplo, para análise do indicador de energia, são definidos quatro níveis de energia, conforme definição do PDF e SDF como mostrado abaixo:

- $E < S_1$ → PDF = 0
- $S_1 < E < S_2$ → valor dos flags depende de outros indicadores
- $S_2 < E < S_3$ → PDF = 1
- $S_3 < E$ → SDF = 1

Modelo Psicoacústico



PUC-Rio / DI

TeleMídia

Modelo Psicoacústico

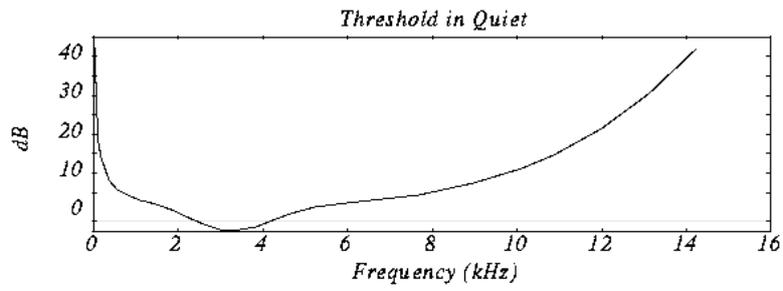
- ➔ Domínio de frequência audível entre 20 Hz e 20 kHz, maior sensibilidade de 2 a 4 KHz.
- ➔ Domínio dinâmico (do silêncio absoluto ao mais alto som) de aproximadamente 96 dB.
- ➔ Sistema de audição tem uma resolução limitada e dependente da frequência. A medida perceptualmente uniforme de frequências pode ser expressa em termos das larguras das *bandas críticas*.



PUC-Rio / DI

TeleMídia

Sensibilidade do Ouvido Humano



PUC-Rio / DI

TeleMídia

Bandas Críticas de Filtragem

Bandas dos bancos de filtros MPEG/Audio



PUC-Rio / DI

TeleMídia

Modelo Psicoacústico

- ➔ Domínio de frequência audível entre 20 Hz e 20 kHz, maior sensibilidade de 2 a 4 KHz.
- ➔ Domínio dinâmico (do silêncio absoluto ao mais alto som) de aproximadamente 96 dB.
- ➔ Sistema de audição tem uma resolução limitada e dependente da frequência. A medida perceptualmente uniforme de frequências pode ser expressa em termos das larguras das *bandas críticas*.



PUC-Rio / DI

TeleMídia

Modelo Psicoacústico

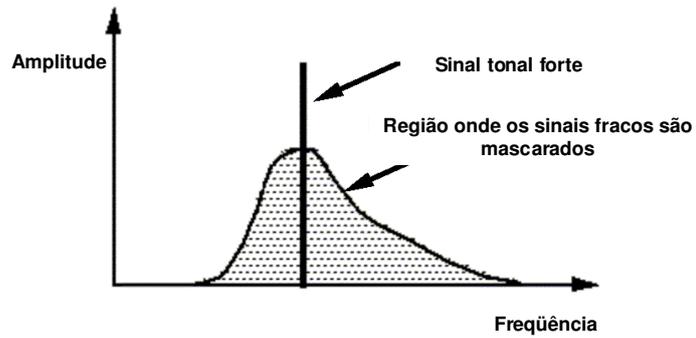
- ➔ Domínio de frequência audível entre 20 Hz e 20 kHz, maior sensibilidade de 2 a 4 KHz.
- ➔ Domínio dinâmico (do silêncio absoluto ao mais alto som) de aproximadamente 96 dB.
- ➔ Sistema de audição tem uma resolução limitada e dependente da frequência. A medida perceptualmente uniforme de frequências pode ser expressa em termos das larguras das *bandas críticas*.
- ➔ Mascaramento de frequências.



PUC-Rio / DI

TeleMídia

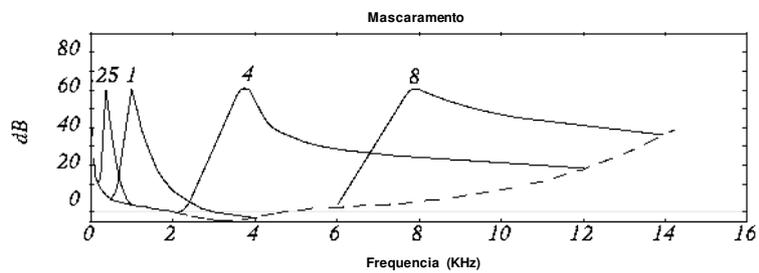
Mascaramento de Frequências



PUC-Rio / DI

TeleMídia

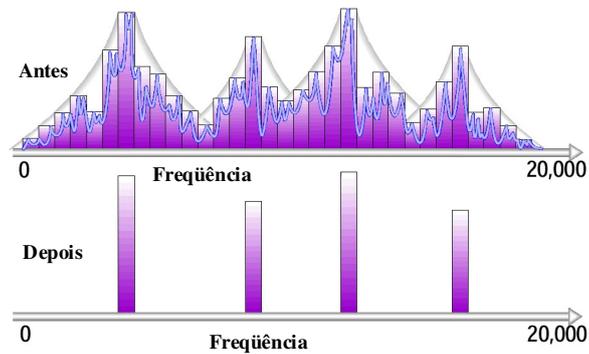
Mascaramento de Frequências (para várias frequências)



PUC-Rio / DI

TeleMídia

Mascaramento de Sub-banda



PUC-Rio / DI

TeleMídia

Modelo Psicoacústico

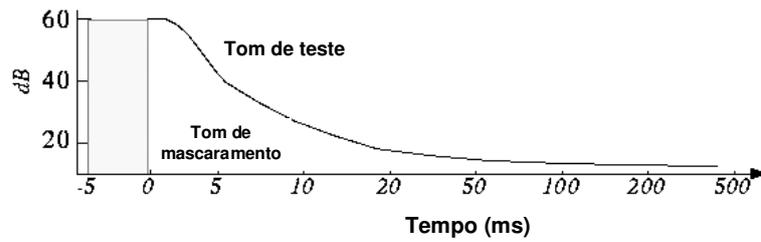
- ➔ Domínio de frequência audível entre 20 Hz e 20 kHz, maior sensibilidade de 2 a 4 KHz.
- ➔ Domínio dinâmico (do silêncio absoluto ao mais alto som) de aproximadamente 96 dB.
- ➔ Sistema de audição tem uma resolução limitada e dependente da frequência. A medida perceptualmente uniforme de frequências pode ser expressa em termos das larguras das *bandas críticas*.
- ➔ Mascaramento de frequências.
- ➔ Mascaramento Temporal.



PUC-Rio / DI

TeleMídia

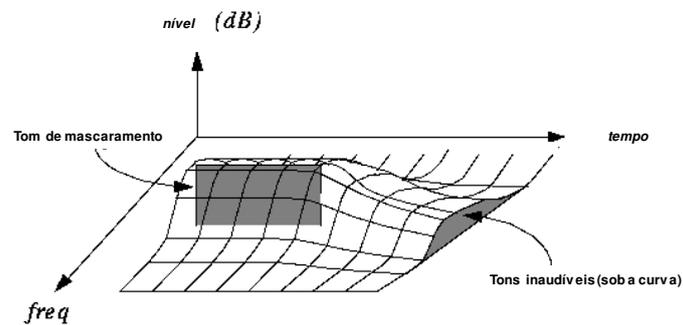
Mascaramento Temporal



PUC-Rio / DI

TeleMídia

Efeito do Mascaramento Temporal e de Frequências



PUC-Rio / DI

TeleMídia

Áudio MPEG

- ➔ Qualquer tipo de áudio.
- ➔ Compressão perceptualmente sem perdas.
- ➔ Faz uso do mascaramento de frequência em todas as camadas.
- ➔ Faz uso da acuidade seletiva do ouvido referente às frequências dividindo o espectro em bandas que tentam refletir o poder de resolução do ouvido em função da frequência.
 - O “threshold” para mascaramento é dependente apenas banda de frequência.



PUC-Rio / DI

TeleMídia

Codificação do Sinal Estereofônico

- ➔ Resultados psicoacústicos mostram que, para frequências maiores que 2KHz, o ouvido percebe a imagem estereofônica baseado mais no envelope temporal do áudio do que em sua estrutura mais refinada.
- ➔ Dois modos:
 - Intensity mode
 - MS stereo mode



PUC-Rio / DI

TeleMídia

Intensity Stereo Mode

- ➔ Soma as frequências mais altas do sinal estereofônico em um único sinal.
- ➔ Os canais de esquerda e direita são reconstruídos com a mesma forma, mas com magnitude diferente, baseadas em fatores de escala.



PUC-Rio / DI

TeleMídia

MS Stereo Mode

- ➔ Codifica os sinais direito e esquerdo em certas faixas de frequência como a soma (middle channel) e diferença (side channel).
- ➔ Técnicas de sintonização são utilizadas para comprimir o sinal side-channel.



PUC-Rio / DI

TeleMídia

Áudio MPEG Layer I

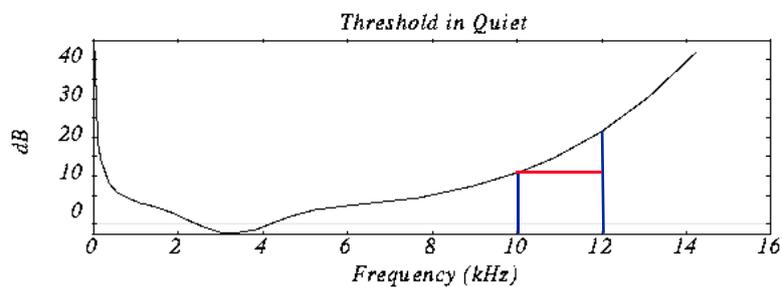
- ➔ Fator de compressão entre 2.7 e 24.
- ➔ Com taxa de compressão de 6:1 (16 bits stereo amostrado em 48 KHz é reduzido para 256 kbits/sec) e nas condições ótimas de audição, especialistas em audição não conseguem distinguir entre o sinal codificado e o original.
- ➔ MPEG áudio pode usar frequências de amostragem de 32, 44.1 e 48 KHz.



PUC-Rio / DI

TeleMídia

Sensibilidade do Ouvido Humano



PUC-Rio / DI

TeleMídia

MPEG

- ➔ Na codificação MPEG existe um número 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 frequências.



PUC-Rio / DI

TeleMídia

MPEG Layer I

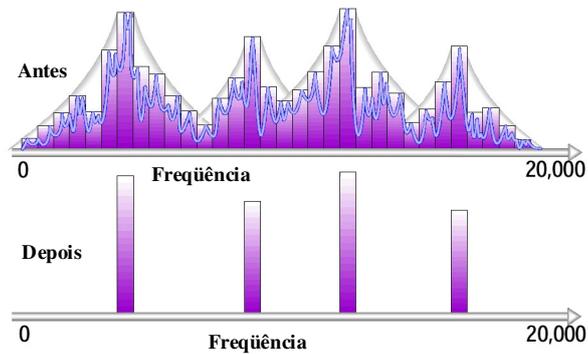
- ➔ Agrupa 12 amostras para cada uma das 32 sub-bandas.
- ➔ Cada grupo de 12 recebe então os bits para codificação e, se o número de bits não é zero, o fator de escala.



PUC-Rio / DI

TeleMídia

Mascaramento de Sub-banda



PUC-Rio / DI

TeleMídia

Exemplo

➔ Depois da análise os primeiros 16 níveis das 32 bandas são:

| | | | | | | | | | | | | | | | | |
|------------|---|---|----|----|---|---|----|----|----|----|----|----|----|----|----|----|
| Band | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Level (db) | 0 | 8 | 12 | 10 | 6 | 2 | 10 | 60 | 35 | 20 | 15 | 2 | 3 | 5 | 3 | 1 |

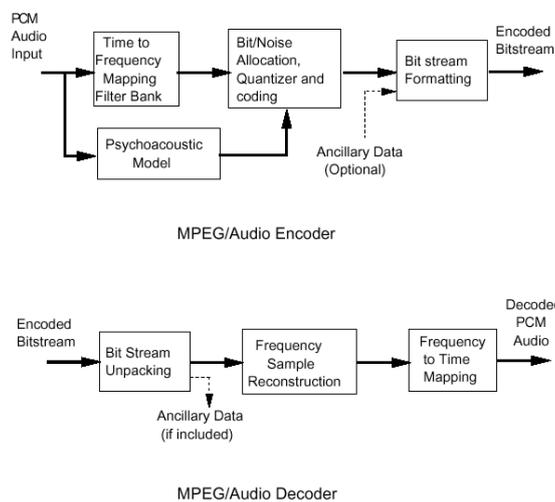
- ➔ Suponha que o nível da 8a. Banda mascara 12 dB na banda 7 e 15dB na 9.
- ➔ Nível na 7a. banda é 10 dB (< 12 dB), então ignore.
- ➔ Nível na 9a. banda é 35 dB (> 15 dB), então codifique, mas apenas a quantidade acima no nível de mascaramento, assim ao invés de usar, por exemplo, 6 bits para codificação, pode-se usar 4 bits -- economizando 2 bits (= 12 dB).



PUC-Rio / DI

TeleMídia

Compressão de Áudio MPEG



PUC-Rio / DI

TeleMídia

MPEG Layer I

➔ Concessões de implementação:

- 32 sub-bandas de largura fixa.
- A filtragem e sua transformada inversa introduzem perdas.
- Bandas adjacentes têm superposição de frequência.

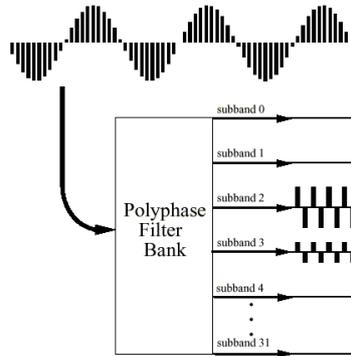


PUC-Rio / DI

TeleMídia

Aliasing

Input audio: 1,500 Hz sine wave sampled at 32 kHz, 64 of 256 samples shown



Subband Outputs:
8x32 samples; both subband 3 and 4
have significant output values



PUC-Rio / DI

TeleMídia

Áudio MPEG I

- ➔ Fator de compressão entre 2.7 e 24.
- ➔ Com taxa de compressão de 6:1 (16 bits stereo amostrado em 48 KHz é reduzido para 256 kbits/sec) e nas condições ótimas de audição, especialistas em audição não conseguem distinguir entre o sinal codificado e o original.
- ➔ MPEG áudio pode usar frequências de amostragem de 32, 44.1 e 48 KHz.
- ➔ Suporta um ou dois canais em um dos três modos:
 - Monofônico -- canal único de áudio.
 - Dual-monofônico -- dois canais independentes, e.g., inglês e francês.
 - Stereo -- para canais stereo que compartilham bits, mas sem usar "Joint-stereo coding".



PUC-Rio / DI

TeleMídia

MPEG Layer I

- ➔ Concessões de implementação:
 - 32 sub-bandas de largura fixa.
 - A filtragem e sua transformada inversa introduzem perdas.
 - Bandas adjacentes têm superposição de frequência.
 - Apenas o modo Intensity Stereo.



PUC-Rio / DI

TeleMídia

Modelo Psicoacústico

- ➔ Domínio de frequência audível entre 20 Hz e 20 kHz, maior sensibilidade de 2 a 4 KHz.
- ➔ Domínio dinâmico (do silêncio absoluto ao mais alto som) de aproximadamente 96 dB.
- ➔ Sistema de audição tem uma resolução limitada e dependente da frequência. A medida perceptualmente uniforme de frequências pode ser expressa em termos das larguras das *bandas críticas*.
- ➔ Mascaramento de frequências.
- ➔ Mascaramento Temporal.



PUC-Rio / DI

TeleMídia

MPEG Layer II

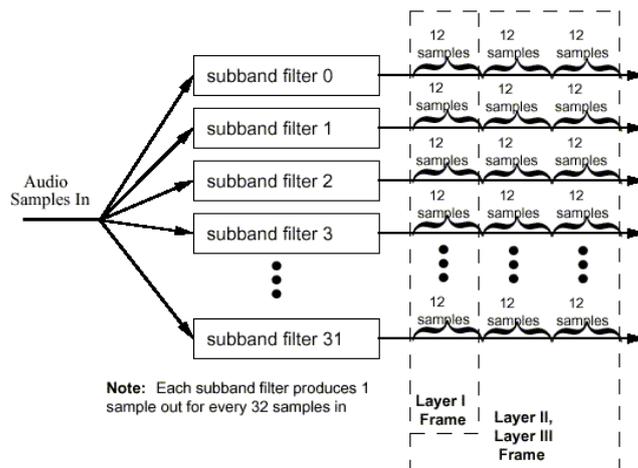
- ➔ Codifica os dados em grupos maiores: para cada sub-banda agrupa 3 grupos de 12 amostras. Isto modela um pouco da máscara temporal.
- ➔ Faz uma aloção de bits e três fatores de escala para cada trio de 12 amostras.



PUC-Rio / DI

TeleMídia

Grupo de Amostras de Sub-bandas - MPEG



PUC-Rio / DI

TeleMídia

MPEG Layer III - Melhoras

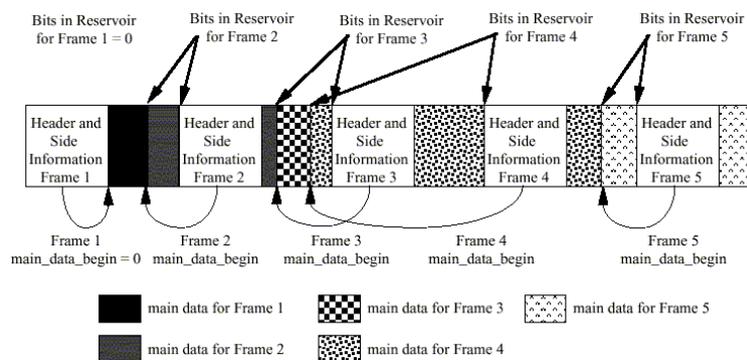
- Redução da sobreposição de bandas.
- Quantização não uniforme (bandas não uniformes).
- Modelo psicoacústico inclui o mascaramento temporal.
- Codificação por entropia
- Uso de “reservatório de bits”.
- Noise allocation (melhor cálculo do quantum de cada banda).



PUC-Rio / DI

TeleMídia

Quadros Layer 3



PUC-Rio / DI

TeleMídia

Comparação – Áudio MPEG 1

| Camadas | Taxa de bits alvo (Kbps) | Taxa de compressão | Qualidade a 64 kbps (MOS) | Qualidade a 128 kbps (MOS) | Retardo Máximo (ms) |
|---------|--------------------------|--------------------|---------------------------|----------------------------|---------------------|
| MP1 | 32 a 448 | 4:1 | --- | --- | 50 |
| MP2 | 32 a 384 | 6:1 | 2.1 a 2.6 | 4+ | 100 |
| MP3 | 32 a 320 | 12:1 | 3.6 a 3.8 | 4+ | 150 |



PUC-Rio / DI

TeleMídia

MPEG - Áudio

➔ MPEG-1

- Um canal (mono) ou dois canais (estéreo ou duo mono)
- Codificação a taxa de amostragem 32, 44.1 e 48 KHz
- Taxas variando de 32 a 448 Kbps para a Camada I, de 32 a 384 Kbps para a Camada II e de 32 a 320 Kbps para a Camada III.



PUC-Rio / DI

TeleMídia

MPEG - Áudio

➔ MPEG-2 BC

- Extensão multicanal compatível com o MPEG-1.
- Até 5 canais principais mais um canal de “melhora de baixa frequência”.
- Extensão da taxa de bits até 1 Mbps.
- Extensão do MPEG-1 visando menores taxas de amostragem (16, 22 e 24 KHz) e de bits (32 a 256 Kbps para a Camada I e 8 a 160 Kbps para as Camadas II e III).

➔ MPEG-2 AAC

- Padrão para codificação de áudio de altíssima qualidade para até 48 canais e taxas de amostragem de 1 a 96 KHz.
- Taxas de 8 Kbps a 160 bps por canal.



PUC-Rio / DI

TeleMídia

MPEG AAC

- ➔ MPEG-2 Advanced Audio Coding (AAC), também conhecido como MPEG-2 Parte 7, ou MPEG-4 Parte 3
- ➔ AAC é considerado o estado da arte para áudio de alta qualidade em uma taxa de bits típica de 128 Kbps. Abaixo dessa taxa, a qualidade do áudio começa a degradar, o que pode ser compensado por técnicas de melhoramento, como SBR (Spectral Band Replication) e PS ([Parametric Stereo](#)).
- ➔ SBR é uma técnica de extensão que permite a mesma qualidade de som à aproximadamente metade da taxa de bits. A combinação de AAC e SBR é chamada de HE-AAC (High efficiency-AAC) versão 1 [ISO/IEC 14496-3 2004], também conhecida como "aacPlus v1".
- ➔ PS aumenta a eficiência de codificação ainda mais, através de uma representação paramétrica da imagem estéreo de um sinal de entrada. A combinação de AAC, SBR e PS é chamada de HE-AAC (High efficiency-AAC) versão 2 [ISO/IEC 14496-3 2004].
- ➔ Note que HE-AAC, definido no padrão MPEG-4, é um super conjunto do núcleo AAC, que estende a alta qualidade de áudio AAC para taxas de bits mais baixas. Decodificadores HE-AAC v2 são capazes de decodificar fluxos HE-AAC v1 e fluxos AAC. Por sua vez, decodificadores HE-AAC v1 são também capazes de decodificar fluxos AAC.
- ➔ O sistema brasileiro de TV digital adotou o padrão MPEG-4 para a codificação do áudio principal de um programa [[ABNT NBR 15602-2 2007](#)], com as características apresentadas na Tabela 1.1.



PUC-Rio / DI

TeleMídia

SBTV D

| | Receptores Fixos e Móveis | Receptores Portáteis |
|---------------------------|---|------------------------------|
| Padrão | ISO/IEC 14496-3 (MPEG-4 AAC) | ISO/IEC 14496-3 (MPEG-4 AAC) |
| Nível e Perfil | AAC@L4 (para multicanal 5.1) HE-AAC v1@L4 (para estéreo) | HE-AAC v2@L3 (dois canais) |
| Taxa de amostragem | 48kHz | 48kHz |



PUC-Rio / DI

TeleMídia

Áudio: Dolby Digital

➔ Dolby Digital:

- 5 canais de áudio (20 Hz a 20 KHz) e um canal para as baixas frequências (20 a 120 Hz), reunidos em um fluxo de 384 Kbps
- Pode trabalhar com amostragens de 32 a 48 KHz
- Realiza codificação por sub-bandas (até 576), obtendo proveito da seletividade da audição humana
- É utilizado no DVD e na HDTV (USA)
- www.dolby.com



PUC-Rio / DI



TeleMídia

Padrão AC-3

- ➔ Blocos de 512 amostras com superposição de 256 amostras, para diminuir os efeitos de blocagem.
- ➔ Cada 6 blocos forma um quadro *synch* de áudio.
- ➔ Troca para o domínio da frequência através de uma transformada rápida de Fourier, utilizando a técnica TDAC (*Time Domain Aliasing Cancelation*).
- ➔ Gera 256 coeficientes de frequências a partir das 512 amostras lidas em sua entrada.



PUC-Rio / DI

TeleMídia

Padrão AC-3

- ➔ Coeficientes gerados são convertidos para números binários em ponto flutuante.
- ➔ Os expoentes, que dificilmente vão diferir de 2 (mais ou menos) para frequências adjacentes podem ser reduzidos a até um expoente por conjunto de 6 blocos (um quadro).
- ➔ Mantissa são quantizadas com uma precisão entre 0 e 16 bits.
- ➔ O codificador pode combinar sinais de dois ou mais dos 6 canais, para uma determinada sub-banda de frequências, caso a taxa necessária para transmissão do áudio for maior que a disponível.



PUC-Rio / DI

TeleMídia

Áudio: DTS



➔ Digital Theater Systems (DTS):

- É utilizado nas trilhas sonoras de filmes de cinema
- Também utiliza os 6 canais de áudio do Dolby Digital
- Divide o áudio em apenas 4 sub-bandas
- A codificação é realizada com 20 bits
- A taxa gerada é de 1,536 Mbps
- Testes práticos tentam demonstrar a diferença entre DD e DTS mas não obtêm sucesso



PUC-Rio / DI

TeleMídia

Perdas de Segmento de Voz

➔ Início do surto de voz

- Limite de perceptividade ~15ms
- Problemas de inteligibilidade ~50ms

➔ Meio do surto de voz

- Limite de perceptividade ~5ms
- Problemas de inteligibilidade ~50ms

➔ CCIR (Comité Consultatif International des Radios)

- perdas < 0,5 % do total de surto
- Na prática < 5%

- ➔ 50ms → segmento de 400 bytes no PCM
- ➔ 5ms → segmento de 40 bytes no PCM



PUC-Rio / DI

TeleMídia

Variações no Intervalo de Silêncio

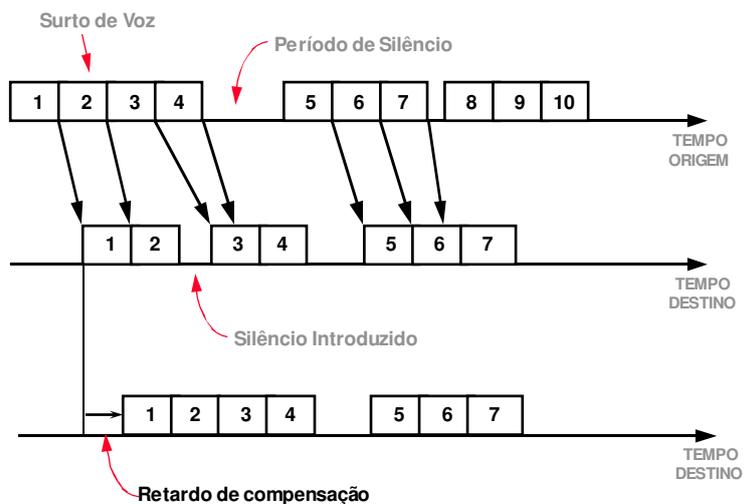
- ➔ Alterações em torno de 50% na duração das pausas são toleráveis.
- ➔ Técnicas de compensação da variação estatística do retardo são possíveis.
- ➔ Necessidade de utilização de técnicas de hangover.



PUC-Rio / DI

TeleMídia

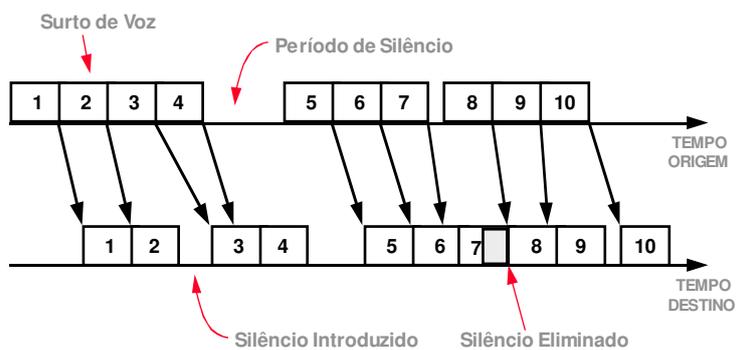
Varição Estatística do Retardo



PUC-Rio / DI

TeleMídia

Variação Estatística do Retardo



PUC-Rio / DI

TeleMídia

Retardo Absoluto Voz com Interatividade

➔ Sistema Telefônico

- Distâncias continentais (< 5.000 Km) → 40 ms
- Distâncias intercontinentais (< 16.0000 Km) → 80 ms

| Estado | Resultado Padrão | 600 ms | 1200 ms |
|------------------|------------------|--------|---------|
| 1 Lado Fala | 40,5 % | 40,0 % | 38,4 % |
| Dois Lados Falam | 4,2 % | 5,9 % | 6,2 % |
| Mútuo Silêncio | 23,2 % | 25,9 % | 29,4 % |



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo

➤ Problemas

- Retardo inicial insuficiente
- Estouro do retardo absoluto

➤ Problemas na Bufferização

- Buffer Overflow
- Buffer Underflow

➤ Underflow

- $> 0.5\%$ → começa a ser perceptível!
- $> 2\%$ → começa a ser intolerável

➤ Overflow

- Sempre que houver a necessidade de se descartar um pacote de voz ou parte dele, deve ser descartado a informação mais velha, pois é a que tem a menor chance de chegar em tempo hábil no destino.



Algoritmo Ideal

- Eliminar toda a variação de retardo na reprodução de pacotes.
- Não ultrapassar o retardo máximo de transferência tolerável para a reprodução do pacote.
- Minimizar a taxa de supressão de voz devido a compensação (informação de voz perdida).
- Não confundir um surto de silêncio com um retardo na rede, preservando durante a reprodução o tempo equivalente ao silêncio na origem.
- É necessário estabelecer limites na carga da sistema para garantir o atendimento dos requisitos acima. A eficiência de um algoritmo pode ser medido pela carga que a rede suporta, sem que os requisitos de qualidade de voz deixem de ser atendidos.



Retardo Total

- Retardo de empacotamento.
- Retardo para o tratamento do pacote pelo sistema operacional e a aplicação, para a entrega para transmissão pela rede.
- Retardo da processamento dos diversos protocolos de rede na estação de origem.
- Retardo de acesso à rede.
- Retardo do transmissão do pacote.
- Retardo de processamento dos diversos protocolos de rede na estação de destino.
- Retardo para o tratamento do pacote pelo sistema operacional e a aplicação, para a entrega a Interface de voz.
- Retardo inserido pelo algoritmo de compensação de retardo.



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo Algoritmo 1

1. Determina-se o retardo do primeiro pacote de um surto de voz.
 2. Retarda o primeiro pacote de Ret_Max - retardo do pacote.
 3. Pacotes que chegam com retardo maior que Ret_Max são descartados.
- **Dificuldade:** determinação dos retardos de transferência dos diversos pacotes.



PUC-Rio / DI

TeleMídia

Determinação do Retardo

- Sincronização dos relógios das várias estações. Cada pacote teria impresso a hora de sua criação na origem, possibilitando o cálculo do retardo no destino.
- Em redes com o tempo de transmissão determinístico, a estação de origem calcularia o tempo desde a criação do pacote até o instante do acesso a rede. Este tempo seria impresso no pacote, possibilitando o destino calcular o retardo total por procedimentos semelhantes aos da origem.
 - necessita-se neste caso, de protocolos que permitam o acesso ao pacote no instante de sua transmissão, para a impressão de seu tempo de vida até então.



PUC-Rio / DI

TeleMídia

Determinação do Retardo

- Não se consegue saber o retardo total sofrido por um pacote. Os algoritmos de compensação tentam estimar este retardo.
- Na estimação do retardo dois tipos de erro podem ocorrer:
 - Um pacote pode ser perdido porque seu atraso estimado ultrapassa o limite, embora o atraso real seja aceitável.
 - Um pacote pode ser aceito apesar do seu atraso real ultrapassar o limite.
- Os erros enunciados acima devem ser levados em consideração quando do cálculo do número de amostras total que podem ser perdidas.



PUC-Rio / DI

TeleMídia

Perda da Informação de Voz

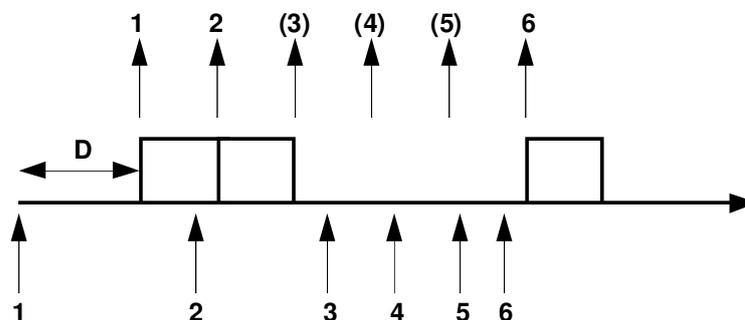
- ➔ Limitação da banda passante.
- ➔ Erro de quantização.
- ➔ Detecção errada de silêncio.
- ➔ Perdas na compensação da variação estatística do retardo
 - Silêncio introduzido indevidamente (não deu para compensar todo o retardo).
 - Perdas próprias do algoritmo de compensação.
 - Segmentos com atraso maior que o máximo permitido.



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo Algoritmo 2 (MIT)



PUC-Rio / DI

TeleMídia

Cálculo de D

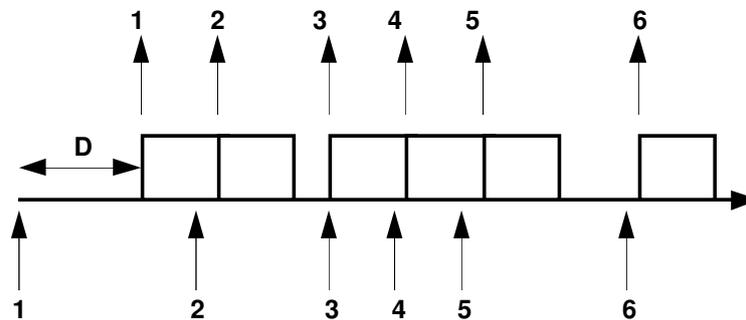
- ➔ D grande pode interferir na natureza interativa de uma conversação.
- ➔ D grande pode causar estouro na área de armazenamento.
- ➔ D pequeno pode resultar em uma probabilidade de perda de pacotes intolerável.



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo Algoritmo 3 (USC)



PUC-Rio / DI

TeleMídia

Cálculo de D

- ➔ D grande pode interferir na natureza interativa de uma conversação.
- ➔ D grande pode causar estouro na área de armazenamento.
- ➔ D pequeno pode resultar em uma probabilidade de geração de “gaps” intolerável.



PUC-Rio / DI

TeleMídia

Cálculo de D

- ➔ Na proposta de Naylor para os algoritmos 2 e 3, a escolha do valor de D deve ser feita baseada no seguinte
 1. Guarda-se a diferença entre os retardos dos pacotes do surto anterior.
 2. O Valor de D escolhido é a máxima diferença entre estes retardos.
 3. Para que a escolha não seja muito pessimista, utiliza-se apenas os retardos de parte dos pacotes do surto anterior.



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo Algoritmo 4 (Cambridge)

1. Retarde o primeiro pacote, após estabelecida a conexão, de um certo número de reservas de amostras D .
2. Cada pacote possui um número de seqüência (silêncio também entra nesta numeração), de tal forma que ao chegar ao destino, se a diferença entre seu número de seqüência e o do pacote anteriormente recebido for igual a 1, então:
 - a) Se ainda houver um certo numero de amostras a serem consumidas (nova reserva), ele espera o tempo correspondente a este consumo, para só então ser reproduzido.
 - b) Se ao chegar o pacote encontrar uma reserva negativa (déficit), ele começa a ser reproduzido imediatamente
3. Se ao chegar ao destino, a diferença entre seu número de seqüência e o do pacote anteriormente recebido for maior que 1, é acrescentado a reserva um número de amostras igual a esta diferença multiplicada pelo número de amostras de um pacote (repare que as novas amostras acrescentadas são silêncio). A partir de então:
 - a) Se ainda houver um certo número de amostras a serem consumidas, ele espera o tempo correspondente a este consumo, para só então ser reproduzido.
 - b) Se a reserva for negativa (déficit), ele começa a ser reproduzido imediatamente.



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo Algoritmo 4 (Cambridge)

- ➔ Nos algoritmos 2 e 3, uma vez estimado o retardo do primeiro pacote de voz, o erro cometido se propaga por todo o surto. Como se espera que em média o primeiro pacote do surto sofra o retardo médio, se a variância do retardo não for grande, os protocolos podem funcionar a contento.
- ➔ No algoritmo 4, o erro cometido se propagaria por toda a comunicação, se não for tomado cuidados adicionais.



PUC-Rio / DI

TeleMídia

Algoritmo 4 Ajuste Dinâmico das Reservas

1. A cada chegada de pacotes, armazene a reserva na fila correspondente de déficit (se a reserva for negativa) ou de crédito (se a reserva for positiva).
2. Acione um mecanismo de ajuste periodicamente que:
 - a) Se houver elemento na fila de déficit, zere as duas filas.
 - b) Se não houver elemento na fila de déficit, então:
 1. Se o número de elementos na fila de créditos (N_c) for igual ao número de pacotes que deveriam ser reproduzidos entre os intervalos de fechamento do mecanismo não faça nada.
 2. Caso contrário, espere o próximo intervalo de silêncio e o diminua da metade do menor crédito encontrado na fila (ou a metade do intervalo de silêncio; o que for menor), e zera a fila.



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo

- No algoritmo 4, não é necessário se escolher o pior caso para a introdução da reserva inicial, pois vai havendo uma convergência do algoritmo durante a comunicação.
- No algoritmo 4, o controle da seqüência do pacote preserva os períodos curtos de silêncio, não deixando estes serem eliminados, o que é bom na ausência de hangover.
- Uma melhora em todos os algoritmos poderia ser efetuada, tentando ajustar a compensação dentro de um próprio surto de voz, como é o caso do algoritmo 5.



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo Algoritmo 5 (TeleMídia -PUC-Rio)

- O objetivo é fazer com que os retardo de transferência de cada pacote, somado a compensação, convirjam para o tempo de transferência médio T_{tr} , somado a compensação inicial D , onde $T_{tr}+D <$ retardo máximo permitido.
- A compensação da variação estatística do retardo varia em torno de D , dentro do intervalo $[0, 2D]$.
- O algoritmo estabelece o tempo máximo de armazenamento de um pacote igual a $2D$, para reduzir a possibilidade de se ultrapassar o retardo máximo permitido.
- O algoritmo tenta garantir, por convergência, que os pacotes com um retardo de transferência dado por:
 $T_{tr} - D < Ret_{pac} < T_{tr} + D$
sejam aceitos.
- Note que o algoritmo pode ser facilmente implementado usando um buffer circular com tamanho igual a $(2D + T_{pac}) * \text{taxa de codificação}$, onde T_{pac} é igual ao tamanho do pacote de voz em unidades de tempo.



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo Algoritmo 5 (TeleMídia -PUC-Rio)

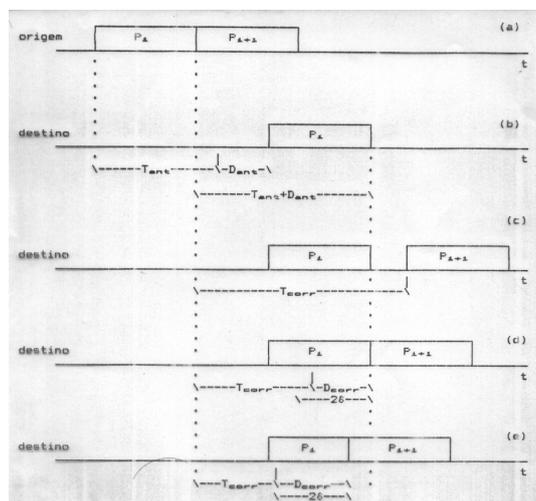
1. Atrase o primeiro pacote do surto de D $R_{pac} = D$
2. Se não for o primeiro pacote do surto, então:
 - a) Se $0 < T_{corr} - T_{ant} < R_{pac}$ e $0 < T_{ant} - T_{corr} < 2D - R_{pac}$, então:
Começa a decodificação do pacote que chegou tão logo termine a decodificação do pacote anterior.
 $R_{pac} = R_{pac} - T_{corr} + T_{ant}$
 - b) Se $0 < T_{ant} - T_{corr} > 2D - R_{pac}$, então:
Descarta os bits não decodificados do pacote anterior a partir do começo do mesmo, até que o número de bits não decodificados corresponda a $2D$ e logo que acabar a decodificação destes bits, comece a decodificação do pacote que acabou de chegar
 $R_{pac} = 2D$
 - c) Se $T_{corr} - T_{ant} > R_{pac}$, então:
Comece a decodificação do pacote assim que ele chegar
 $R_{pac} = 0$



PUC-Rio / DI

TeleMídia

Compensação da Variação Estatística do Retardo Algoritmo 5 (TeleMídia -PUC-Rio)



PUC-Rio / DI

TeleMídia

Tamanho dos Pacotes de Voz

- ➔ Perdas (no início ou meio do surto) de voz menores que 50ms (5ms para serem não perceptíveis).
- ➔ Retardo absoluto menores que da ordem de 100 ms.



PUC-Rio / DI

TeleMídia

Interrupções

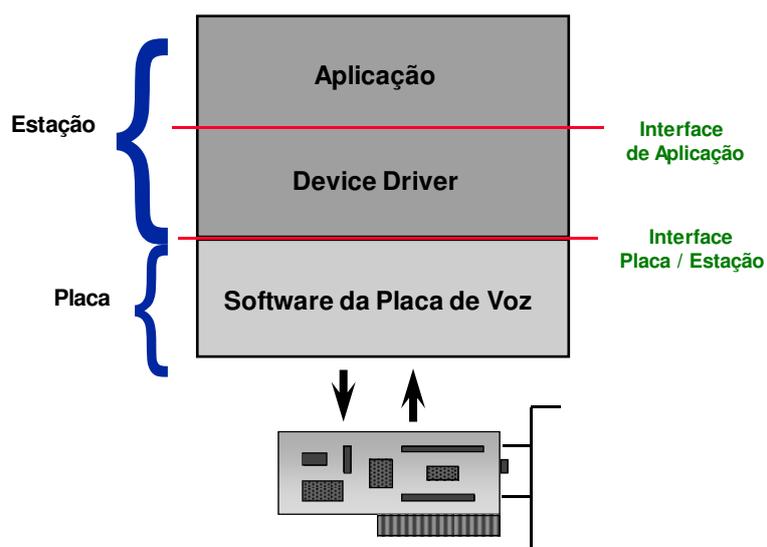
- Número de conexões interrompidas depois que a conexão já tenha sido estabelecida.
- Padrão telefônico atual tolera 0,5%
- Projetos futuros de redes prevêem 0,15%



PUC-Rio / DI

TeleMídia

Áudio



PUC-Rio / DI

TeleMídia

Vídeo

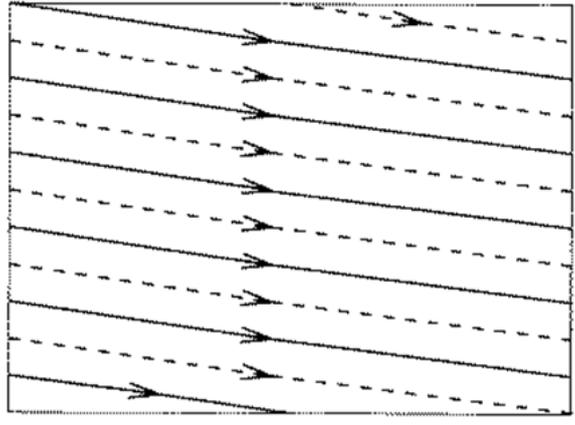


PUC-Rio / DI



TeleMídia

Vídeo



Varredura entrelaçada (reduz cintilação)

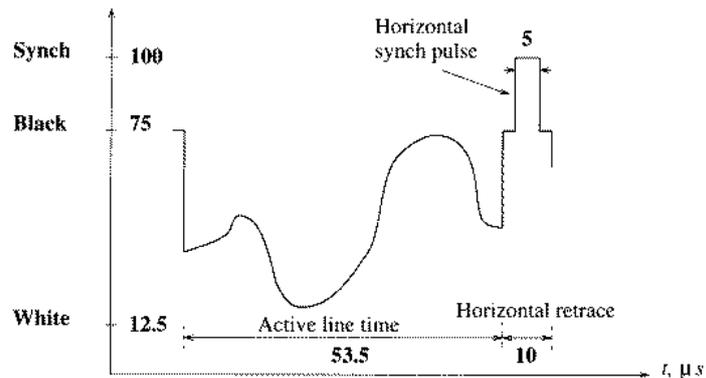


PUC-Rio / DI



TeleMídia

Sinal de Vídeo (uma linha)



PUC-Rio / DI

TeleMídia

Sinal de Vídeo a Cores

- ➔ **RGB** (component video)
 - Cores básicas em separado
 - Sinal de sincronismo no verde (G) ou em separado.

- ➔ **Y/C** (S-video)
 - Luminância e crominância em separado

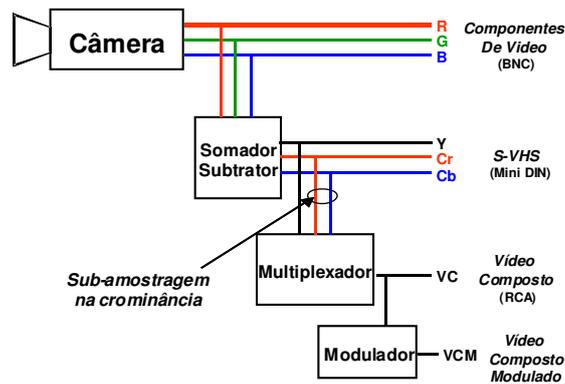
- ➔ **Vídeo composto**
 - Luminância e crominância combinadas no mesmo sinal.



PUC-Rio / DI

TeleMídia

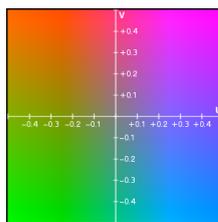
Geração do sinal de vídeo



PUC-Rio / DI

TeleMídia

Padrões para Vídeo Composto

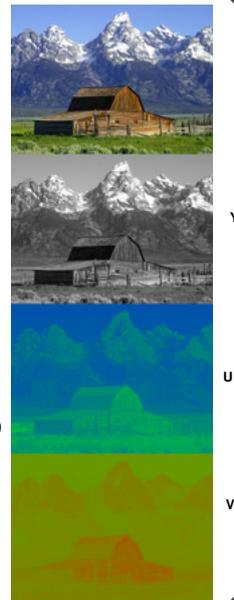


Plano U-V em Y=0.5

→ Sistema PAL

- $Y = 4,2 \text{ MHz}$ $Y = (0,299R + 0,57G + 0,114B)$
- $U = 1,5 \text{ MHz}$ $U = (0,492 (B-Y))$
- $V = 1,5 \text{ MHz}$ $V = (0,877 (R-Y))$

$$\cos 2\pi f_c t + v \sin 2\pi f_c t$$



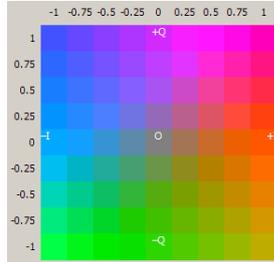
fonte: wikipedia



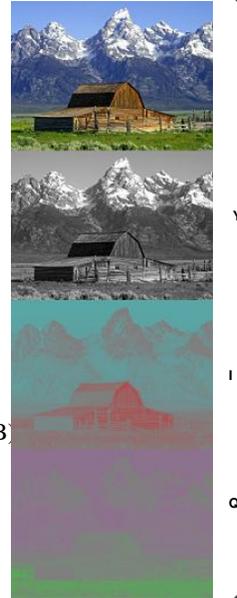
PUC-Rio / DI

TeleMídia

Padrões para Vídeo Composto



YIQ com Y=0.5



→ Sistema NTSC

- Y = 4,2 MHz $Y = (0,299R + 0,587G + 0,114B)$
- I = 1,5 MHz $I = (0,59R - 0,27G - 0,32B)$
- Q = 0,5 MHz $Q = (0,21R - 0,52G - 0,31B)$

$$\cos 2\pi f_c t + v \sin 2\pi f_c t$$

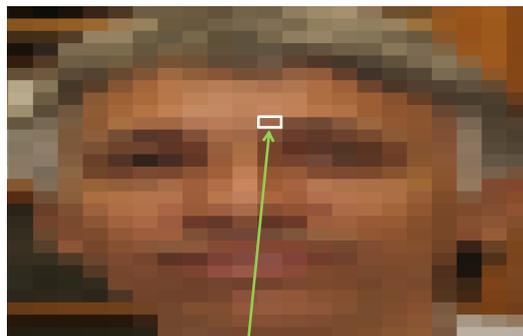


PUC-Rio / DI

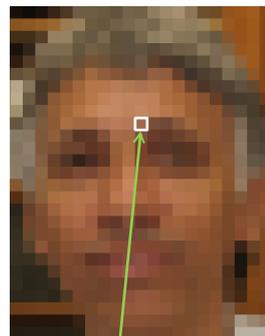
fonte: wikipedia

TeleMídia

Relação de Aspecto do Pixel



Relação de aspecto de pixel 2:1



Relação de aspecto de pixel 1:1



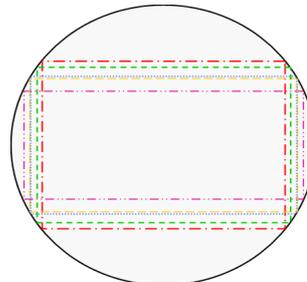
PUC-Rio / DI

Source: wikipedia

TeleMídia

Relação de Aspecto da Imagem

- ➔ SDTV é 4:3 (1.33)
- ➔ HDTV é 16:9 (2.11)
- ➔ Cinema usa 1.85:1 ou 2.35:1



--- 4:3
--- 3:2
--- 16:9
--- 1.85:1
--- 2.39:1



PUC-Rio / DI

TeleMídia

Vídeo

- ➔ O sinal de vídeo pode ser codificado usando qualquer uma das técnicas para sinais contínuos: PCM, DPCM, ADPCM, SBC etc.
- ➔ Um vídeo pode ser também 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. Esta técnica se constitui a base da codificação chamada MJPEG (Motion JPEG).



PUC-Rio / DI

TeleMídia

Sinal de Vídeo Monocromático

➔ Padrão M (Brasil, EUA)

- 30 quadros / seg. (60 campos / seg.)
- 525 linhas / quadro (493 visíveis)
- Resolução de tela 640x480
- Relação de aspecto do pixel 10:11

➔ Padrão Europeu

- 25 quadros / seg.
- 625 linhas / quadro
- Resolução de tela 768x576
- Relação de aspecto do pixel 12:11



PUC-Rio / DI

TeleMídia

TV Digital

➔ A recomendação BT 601-4 [ITU-R BT.601-4] especifica:

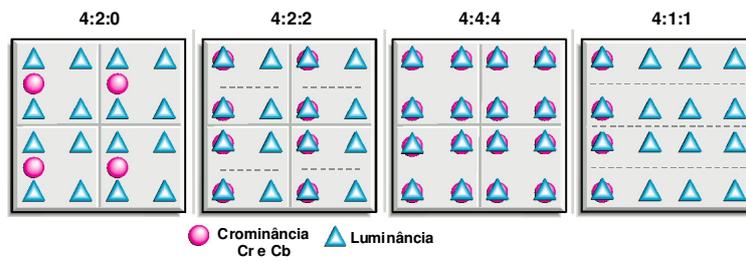
- 483 linhas ativas com 720 pixels por linha.
- Apenas 480 das 483 linhas são usadas para codificação e apenas 704 dos 720 pixels (os primeiros e últimos 8 pixels são descartados).
- Sub-amostragem de croma 4:2:2



PUC-Rio / DI

TeleMídia

Vídeo Digital



PUC-Rio / DI

TeleMídia

Outras Padrões de Resolução



PUC-Rio / DI

TeleMídia

ITU-T H.261

| | CIF | | QCIF | |
|-------------------------|---------------|------------------|---------------|------------------|
| | Linhas/quadro | pixels/linha | Linhas/quadro | pixels/linha |
| Luminância (Y) | 288 | 360 (352) | 144 | 180 (176) |
| Crominância (Cb) | 144 | 180 (176) | 72 | 90 (88) |
| Crominância (Cr) | 144 | 180 (176) | 72 | 90 (88) |



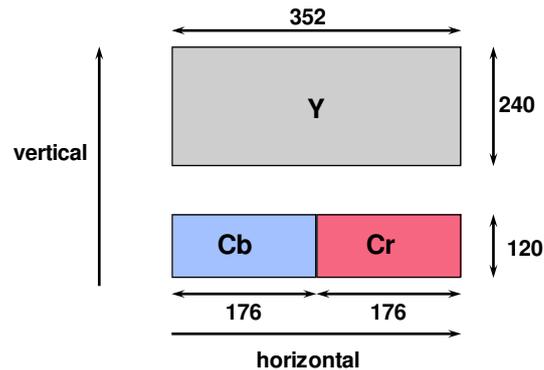
ITU-T H.263

- ➔ H.263 acrescenta três novos formatos:
 - Sub-QCIF (128 x 96),
 - 4CIF (704 x 576), também conhecido como SCIF
 - 16CIF (1408 x 1152).
- ➔ Todos os formatos citados possuem sub-amostragem de crominância 4:2:0.



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:



PUC-Rio / DI

TeleMídia

MPEG 2 Vídeo

- ➔ 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 (352 x 240 x 30 quadros – idêntico ao SIF MPEG 1)
 - nível principal visando a codificação com qualidade de TV (720 x 480 x 30)
 - 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.



PUC-Rio / DI

TeleMídia

ATSC Digital Television Standard

| Vertical Lines | Horizontal Pixels | Aspect Ratio | Picture Rate |
|----------------|-------------------|--------------|-----------------|
| 1080 | 1920 | 16:9 | 60I 30P 24P |
| 720 | 1280 | 16:9 | 60P 30P 24P |
| 480 | 704 | 16:9 & 4:3 | 60I 60P 30P 24P |
| 480 | 640 | 4:3 | 60I 60P 30P 24P |

Para HD (1280x720) e Full-HD (1920x1080) a relação de aspecto do pixel é 1:1



PUC-Rio / DI

TeleMídia

Compressão



PUC-Rio / DI

TeleMídia

Vídeo

- ➔ Poderíamos empregar a codificação JPEG em cada quadro. Esta técnica se 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-quadro*).



PUC-Rio / DI

TeleMídia

Interframe

- ➔ Leva em consideração a redundância temporal
- ➔ Para cada bloco do quadro corrente procura-se o bloco mais próximo do quadro anterior.
- ➔ A estimativa do movimento é mandada ao decodificador através do vetor de movimento do bloco.
- ➔ A diferença dos dados correntes e os dados do quadro anterior compensados pelo movimento (mudados de lugar) são então codificados e comprimidos.



PUC-Rio / DI

TeleMídia

H.261



PUC-Rio / DI

TeleMídia

H-320

- ▶ Codificador de Vídeo: H.261
- ▶ Codificador de Áudio: G.711, G.722, H.200/AV.254, /AV.253
- ▶ Procedimento de Comunicação: H.242
- ▶ Estrutura do Quadro: H.221
- ▶ Controle e Indicação: H.230



PUC-Rio / DI

TeleMídia

ITU-T H.261

- ➔ Aprovado em dezembro de 1990
- ➔ Cobre taxa de $P \times 64$ Kbps ($p = 1, 2, \dots, 30$), isto é :
64 Kbps a 1,92 Mbps
- ➔ $p=1$ ou $p=2$ para videofonia
 $p > 5$ para videoconferência
- ➔ Taxa de quadros: 30, 15, 10, 7,5 quadros/segundo
- ➔ Dois formatos de imagem CIF e QCIF



PUC-Rio / DI

TeleMídia

ITU-T H.261

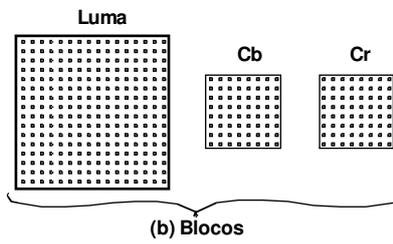
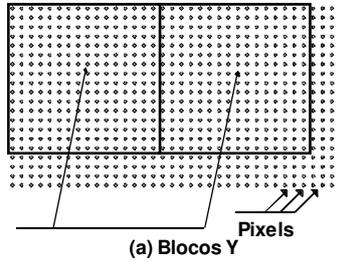
| | CIF | | QCIF | |
|------------------|---------------|--------------|---------------|--------------|
| | Linhas/quadro | pixels/linha | Linhas/quadro | pixels/linha |
| Luminância (Y) | 288 | 360 (352) | 144 | 180 (176) |
| Crominância (Cb) | 144 | 180 (176) | 72 | 90 (88) |
| Crominância (Cr) | 144 | 180 (176) | 72 | 90 (88) |



PUC-Rio / DI

TeleMídia

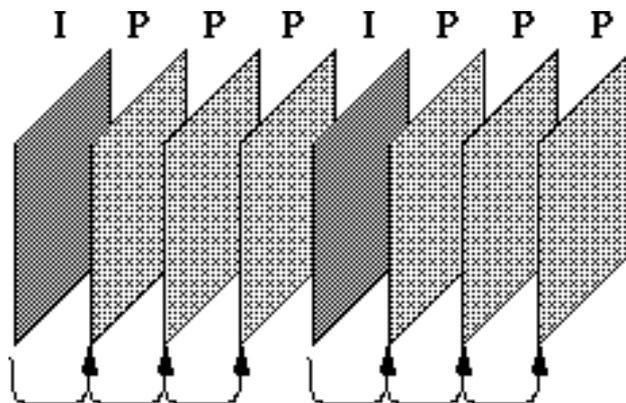
H.261



PUC-Rio / DI

TeleMídia

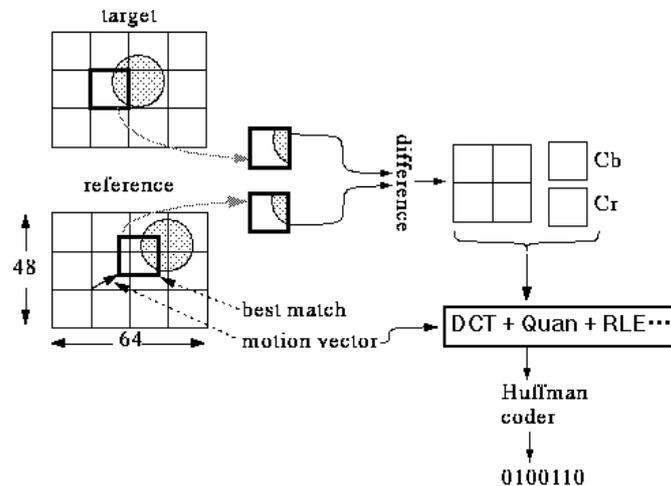
Tipos de Quadro H. 261



PUC-Rio / DI

TeleMídia

Estimação de Movimento



PUC-Rio / DI

TeleMídia

ITU-T H.261

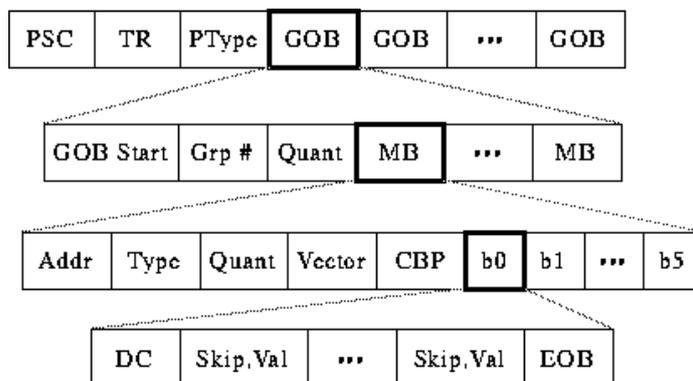
- Cada quadro é dividido em macro blocos disjuntos onde cada macrobloco consiste de um bloco de luminância (Y) de 16 x 16 e dois blocos de crominância (Cr e Cb) de 8 x 8
- Para cada bloco de luminância é calculado o vetor de movimento
- O codificador decide qual modo deverá ser utilizado, intramodo ou preditivo. Se preditivo, usa o DPCM entre o bloco corrente e o predito. No intramodo o bloco é codificado diretamente
- Em qualquer dos dois casos anteriores, o processo de DCT seguido de um quantizador linear é utilizado. Após é realizada a codificação por entropia (Huffman)
- Após o processo acima os dados são enviados a um buffer que vai regular o fluxo de informação para uma taxa de bits constante. O tamanho do passo do quantizador é ajustado periodicamente para chegar às taxas desejadas.



PUC-Rio / DI

TeleMídia

H. 261 - Estrutura de Dados



PUC-Rio / DI

TeleMídia

H-324

- ➔ Codificador de Vídeo: H.263 ou H.261
- ➔ Codificador de Áudio: G.723.1.1
- ➔ Controle e Indicação: H.245 (sinalização fim a fim para operação apropriada do H.324)
- ➔ Estrutura do Quadro: H.223



PUC-Rio / DI

TeleMídia

H.263

- ➔ Otimiza o H.261 para codificação de vídeo a taxas inferiores a 64 Kbps
- ➔ Ideal para videotelefonia e comunicações móveis
- ➔ Otimiza:
 - Compensação de movimento
 - Quantização
 - Compactação
- ➔ Deverá substituir o H.261, pois também oferece suporte a taxas e resoluções maiores



PUC-Rio / DI

TeleMídia

ITU-T H.263

- ➔ H.263 acrescenta três novos formatos:
 - Sub-QCIF (128 x 96),
 - 4CIF (704 x 576), também conhecido como SCIF
 - 16CIF (1408 x 1152).
- ➔ Todos os formatos citados possuem sub-amostragem de croma 4:2:0.



PUC-Rio / DI

TeleMídia

Formato de Quadros H.263 e H.261

| Video format | Luminance Image Resolution | Chrominance Image Resolution | H.261 support | H.263 support | Bit-rate (Mbit/s) (if uncompressed, 30 fps) | | Max bits allowed per picture (BPPmaxKb) |
|--------------|----------------------------|------------------------------|---------------|---------------|---|-------|---|
| | | | | | B / W | Color | |
| SQCIF | 128 x 96 | 64 x 48 | Optional | Required | 3.0 | 4.4 | 64 |
| QCIF | 176 x 144 | 88 x 72 | Required | Required | 6.1 | 9.1 | 64 |
| CIF | 352 x 288 | 176 x 144 | Optional | Optional | 24.3 | 36.5 | 256 |
| 4CIF | 704 x 576 | 352 x 288 | n/a | Optional | 97.3 | 146.0 | 512 |
| 16CIF | 1408 x 1152 | 704 x 576 | n/a | Optional | 389.3 | 583.9 | 1024 |



MPEG



MPEG

- ➔ MPEG (Moving Picture Experts Group)
ISO-IEC/JTC1/SC2/WG11
 - MPEG - Video
 - MPEG-Audio
 - MPEG-System

- ➔ Objetivo: Produzir sequências de bits comprimidos, de vídeo e áudio associado, e multiplexação dessas sequências.



PUC-Rio / DI

TeleMídia

MPEG

- ➔ Fundado em 1988
- ➔ MPEG-1 (ISO 11172) publicado em 1992
- ➔ MPEG-2 (ISO 13818) iniciado em 1990
- ➔ Em 1992 MPEG-2 é ampliado para atender a codificação de HDTV (antes assunto do MPEG-3)
- ➔ MPEG-4 iniciado em fins de 1992 e início de 1993



PUC-Rio / DI

TeleMídia

MPEG

- ➔ Objetivo: Produzir seqüências de bits comprimidos, de vídeo e de áudio associado, e multiplexação dessas seqüências.
- ➔ MPEG-1: Tem como objetivo atender aplicações interativas que envolvem dispositivos de armazenamento (CD-ROM, Fitas DAT, ...), canais de comunicação (LAN's, ISDN, ...) e taxas de compressão abaixo de 1,5 Mbps.
- ➔ MPEG-2: Tem como objetivo atingir um número maior de aplicações que exijam maior resolução, como TV digital.



PUC-Rio / DI

TeleMídia

MPEG-1



PUC-Rio / DI

TeleMídia

Resolução

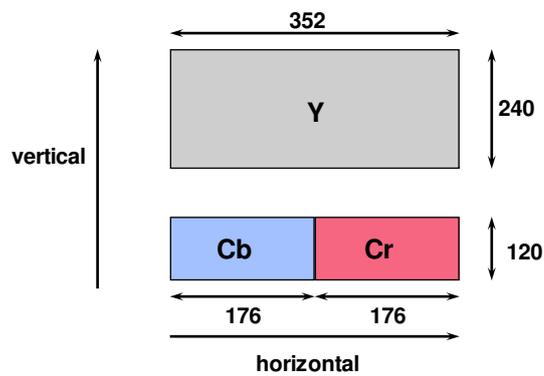
- ▶ Pode codificar imagens de até 4096 x 4096 x 60 quadros/segundo.
- ▶ A maioria das aplicações usam, no entanto, o formato SIF.



PUC-Rio / DI

TeleMídia

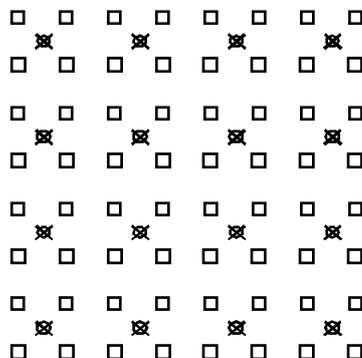
MPEG-1 Vídeo



PUC-Rio / DI

TeleMídia

MPEG-1 Método 4:2:0



Amostragem 4:2:0
 requer a interpolação
 por média aritmética
 simples dos valores
 das quatro amostras
 para obtenção do
 croma

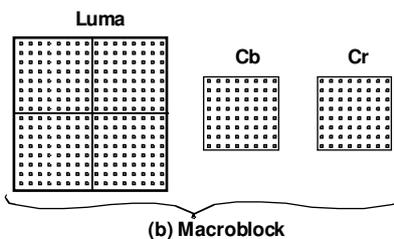
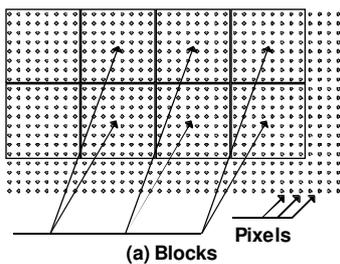
- Luma sample
- × Cb sample
- Cr sample



PUC-Rio / DI

TeleMídia

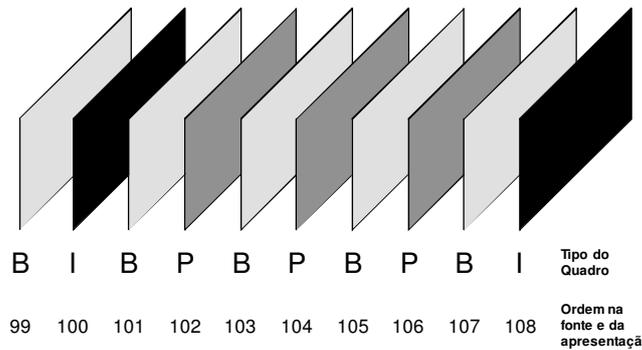
MPEG - 1



PUC-Rio / DI

TeleMídia

Tipos de Quadro MPEG-1



PUC-Rio / DI

TeleMídia

MPEG - 1

- ➔ I utiliza apenas redundância espacial
- ➔ Pelo menos um quadro I a cada 132 quadros
- ➔ Quadros P utilizam a redundância temporal pela predição baseada no quadro I ou P imediatamente anterior
- ➔ Dentro do quadro P podem existir macroblocos oriundos da predição temporal, ou macroblocos oriundos da codificação intraquadros.



PUC-Rio / DI

TeleMídia

MPEG - 1

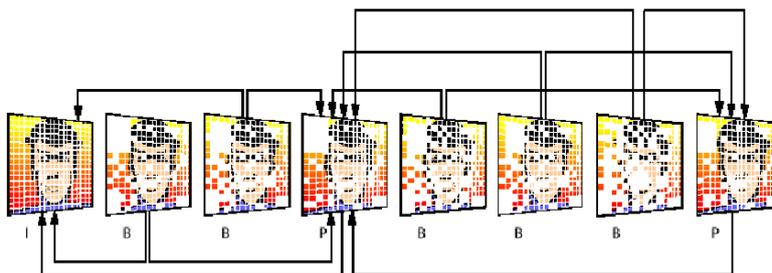
- ➔ Quadros B incluem a predição de dois quadros âncoras, um anterior e outro posterior. Só quadros I e P podem ser quadros âncoras.



PUC-Rio / DI

TeleMídia

MPEG-1



PUC-Rio / DI

TeleMídia

Estimação do Movimento

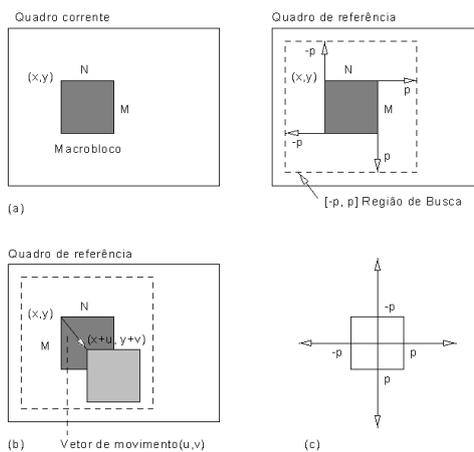
- ➔ Unidades de predição: macroblocos (sem limites de fronteira de localização)
- ➔ Os vetores de movimento de uma fatia são diferenciados.
 - O valor do primeiro vetor é transmitido diretamente
 - Os demais vetores são transmitidos como diferenças em relação ao vetor anterior.



PUC-Rio / DI

TeleMídia

Estimação de Movimento



PUC-Rio / DI

Para TV, obtém-se bom desempenho com $p=15$ em cenas comuns de noticiário e $p=63$ em cenas de muito movimento, por exemplo, cenas de esporte.

TeleMídia

Casamento ruim

Casamento razoável

Casamento bom

Macrobloco a ser codificado

PUC-Rio / DI

TeleMídia

(a)

(b)

(c)

(a) Quadro desejado (b) menos quadro "previsto" (c) quadro de erro residual

PUC-Rio / DI

TeleMídia

Estimação de Movimento Quadro B

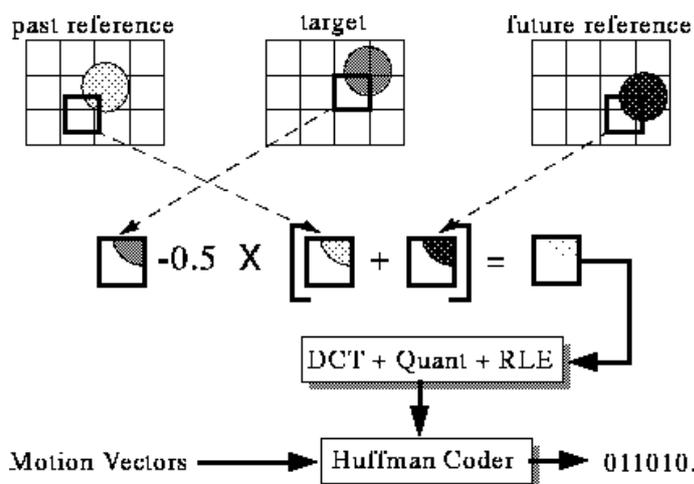
- ➔ Baseada no quadro anterior
- ➔ Baseada no quadro posterior
- ➔ Baseada na interpolação do quadro anterior e posterior



PUC-Rio / DI

TeleMídia

Interpolação



PUC-Rio / DI

TeleMídia

Transformada de Co-senos

Direta

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

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

$$C(w) = 1 \text{ para } w = 1, 2, \dots, 7$$

Inversa

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

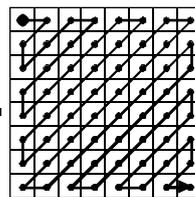


PUC-Rio / DI

TeleMídia

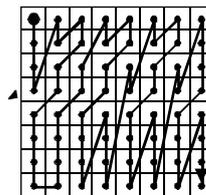
Codificação por Entropia

- Código de Huffman
- Codificação por carreira



(a) Zigzag scan ordering of coefficients

Blocks of DCT coefficients



(b) Alternate scan ordering of coefficients



PUC-Rio / DI

TeleMídia

Quantização

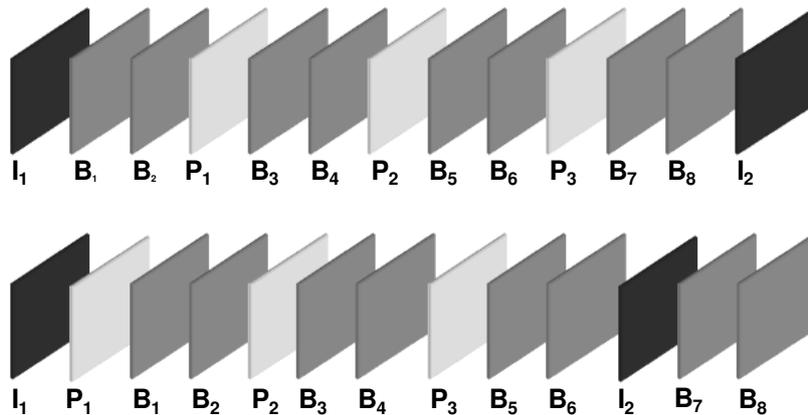
- ➔ Cada valor da matriz de quantização representa a precisão do seu coeficiente correspondente na DCT
- ➔ Uma matriz de quantização para macroblocos com codificação intraquadros e outra para macroblocos com codificação interquadros
- ➔ O nível de quantização pode ser ajustado para cada macrobloco



PUC-Rio / DI

TeleMídia

Decodificação do Vídeo MPEG-1



PUC-Rio / DI

TeleMídia

Tamanhos Típicos de Quadros MPEG-1

| <i>Type</i> | <i>Size</i> | <i>Compression</i> |
|-------------|-------------|--------------------|
| I | 18 KB | 7:1 |
| P | 6 KB | 20:1 |
| B | 2.5 KB | 50:1 |
| Avg | 4.8 KB | 27:1 |



PUC-Rio / DI

TeleMídia

Fatias MPEG

- ➔ Mecanismo para limitar a propagação de erros
 - Sequência de bits transmitida por palavras de tamanho variável
 - Erro não corrigido compromete o alinhamento das palavras em uma fatia
- ➔ Número de fatias é um compromisso entre boa recuperação de erros e boa taxa de compressão.
- ➔ Pelo padrão uma fatia pode ocupar mais de uma linha, pelo padrão americano o primeiro macrobloco de cada linha corresponde sempre ao início de uma fatia.

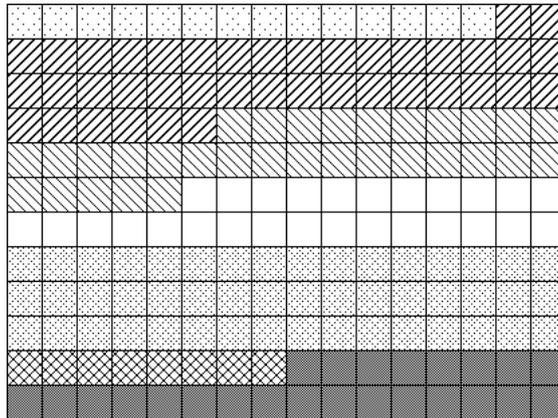


PUC-Rio / DI

TeleMídia

Fatias MPEG

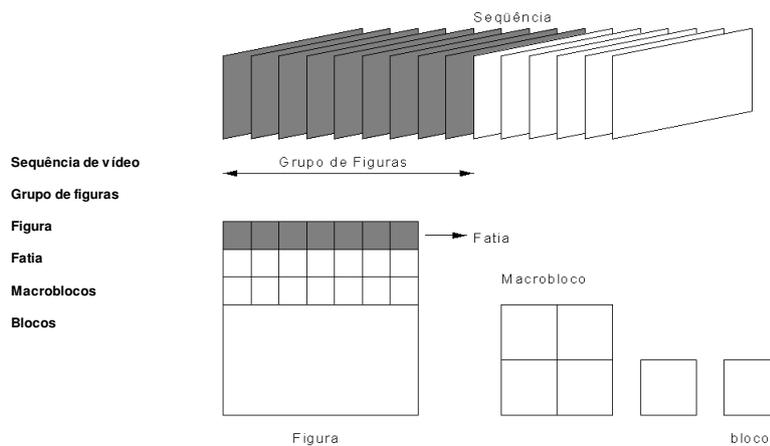
(exemplo: figura com 7 fatias)



PUC-Rio / DI

TeleMídia

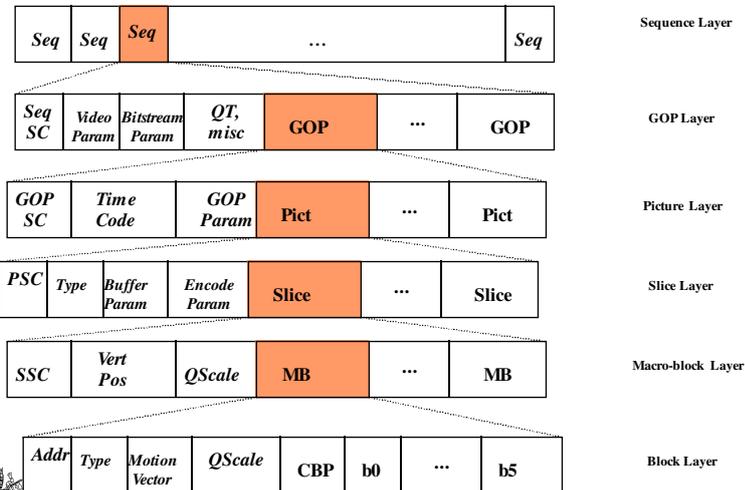
Camadas de Vídeo MPEG-1



PUC-Rio / DI

TeleMídia

Formato MPEG-1



PUC-Rio / DI

CS 414 - Spring 2012

TeleMídia

MPEG System

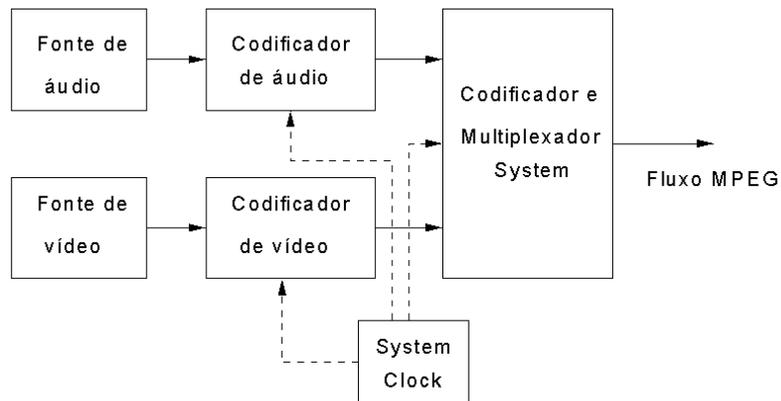
- ➔ Sincronização dos fluxos elementares.
- ➔ Parsing do fluxo multiplexado após acesso randômico.
- ➔ Gerenciamento de buffer no decodificador.
- ➔ Identificação do tempo absoluto da informação codificada.



PUC-Rio / DI

TeleMídia

Fluxo de Dados MPEG 1 System



PUC-Rio / DI

TeleMídia

MPEG 1 System

- ➔ 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.



PUC-Rio / DI

TeleMídia

MPEG 1 System

- O cabeçalho *packet* contém a identificação do fluxo elementar, os requisitos de bufferização e informações de temporização.
- Os dados *packet* contêm um número de bytes variável de um mesmo fluxo elementar.
- 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.



PUC-Rio / DI

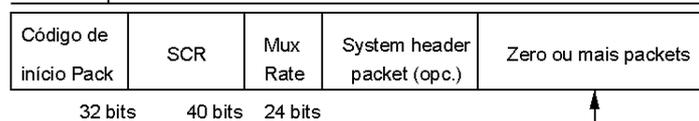
TeleMídia

Camadas MPEG System

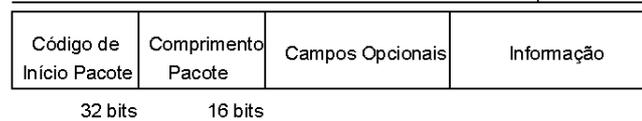
Camada ISO 11172 stream



Camada pack



Camada packet



PUC-Rio / DI

TeleMídia

MPEG System - Sincronização

- ▶ STC (System Time Clock): tempo de referência que opera a 90 KHz.
- ▶ PTS (Presentation Time Stamp): Valor de STC armazenado na chegada de algumas unidades de áudio e vídeo ao chegar no codificador.
- ▶ DTS: Decoding Time Stamp
- ▶ SCR (System Clock Reference): Amostras de STC gerados no momento que o último octeto relativo ao SCR deixa o codificador.
- Usados para sincronismo e gerência de buffers, SCR e PTS, e DTS quando requerido, devem ser inseridos em intervalos de tempo que não excedam 0,7 s.



PUC-Rio / DI

TeleMídia

MPEG-2



PUC-Rio / DI

TeleMídia

MPEG-2 x MPEG-1

- ▶ Sintaxe para quadros até 16383x16383
- ▶ Várias resoluções suportadas.
- ▶ Macroblocos 4:2:0, 4:2:2 e 4:4:4
- ▶ Precisão de 8, 9, 10 e 11 bits pada DCT
- ▶ Quantização não linear
- ▶ Melhor eficiência no tratamento de entrelaçamento
- ▶ Extensões escaláveis



PUC-Rio / DI

TeleMídia

MPEG-2

- ▶ Qualquer decodificador MPEG-2 pode decodificar vídeos MPEG-1
- ▶ Atinge qualidade de vídeo com pequenas degradações visíveis, para o padrão CCIR-601, a 1,5 - 6 Mbps.
- ▶ Suporta quadros de até 16K x 16K.



PUC-Rio / DI

TeleMídia

MPEG 2 Vídeo

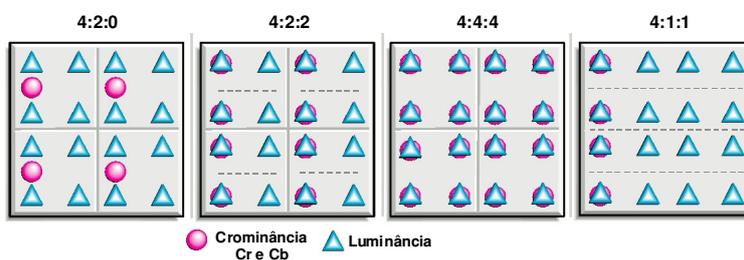
- ▶ O padrão MPEG 2 pode codificar imagens de até 16.383 linhas x 16.383 pixels.
- ▶ O padrão é organizado, como vimos, em diversos perfis e níveis, que especificam o formato utilizado. Exemplos de formato são:
 - nível baixo (352 x 240 x 30 quadros – idêntico ao SIF MPEG 1)
 - nível principal visando a codificação com qualidade de TV (720 x 480 x 30)
 - 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.



PUC-Rio / DI

TeleMídia

Vídeo Digital



PUC-Rio / DI

TeleMídia

Escalabilidade

- ➔ As extensões de escalabilidade provêm 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.
- ➔ Muitas das técnicas empregadas são semelhantes às codificações progressivas e hierárquica do JPEG.



PUC-Rio / DI

TeleMídia

Escalabilidade

- ➔ Escalamento Espacial
- ➔ Escalamento SNR
- ➔ Particionamento de Dados
- ➔ Escalamento Temporal



PUC-Rio / DI

TeleMídia

Escalabilidade

➔ Escalamento Espacial

- Útil para *simulcasting* e para decodificação por software
- Codifica a camada base em uma resolução menor que as camadas superiores
- Camada base é usada para predição das camadas superiores



PUC-Rio / DI

TeleMídia

Escalabilidade

➔ Escalamento Espacial

➔ Escalamento SNR

- Outra camada pode ser adicionada à camada base oferecendo uma melhora na precisão dos coeficientes da DCT (Direct 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.



PUC-Rio / DI

TeleMídia

Escalabilidade

- Escalamento Espacial
- Escalamento SNR
- Particionamento de Dados
 - Bloco de 64 coeficiente é quebrado em duas seqüências:
 - a primeira, de alta prioridade, contém os coeficientes de baixa frequência e informação adicional (tais como valores DC e vetores de movimento)
 - a segunda carrega altas frequências



PUC-Rio / DI

TeleMídia

Escalabilidade

- Escalamento Espacial
- Escalamento SNR
- Particionamento de Dados
- Escalamento Temporal
 - Sequência de prioridade alta codifica o vídeo a uma taxa baixa de quadros
 - Quadros intermediários são codificados em uma segunda seqüência de bits usando a reconstrução da primeira seqüência como predição



PUC-Rio / DI

TeleMídia

Escalabilidade

- Escalamento Espacial
- Escalamento SNR
- Particionamento de Dados
- Escalamento Temporal
- Escalabilidade é especialmente útil em redes que permitem distinguir os tipos de fluxos e privilegiar a entrega do mais importante. Assim, em caso da necessidade ou conveniência de perda, um bom sinal de vídeo ainda poderá ser recebido.



PUC-Rio / DI

TeleMídia

MPEG 2 Vídeo

- nível baixo (352 x 240 x 30 quadros – idêntico ao SIF MPEG 1)
- nível principal visando a codificação com qualidade de TV (720 x 480 x 30)
- 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).



PUC-Rio / DI

TeleMídia

Perfis

- ▶ Principal (*main profile*): utiliza os quadros I, P e B e uma amostragem de cor 4:2:0.
- ▶ Simples (*simple profile*): é basicamente o perfil principal sem os quadros B.
- ▶ Escalável SNR (*SNR Profile*): adiciona a escalabilidade SNR ao perfil principal.
- ▶ Escalável espacialmente (*spatially scalable profile*): adiciona a escalabilidade espacial ao perfil escalável SNR.
- ▶ Alto adiciona cor 4:2:2 ao perfil escalável espacialmente.
- ▶ Todos os perfis são limitados ao máximo de três camadas.



PUC-Rio / DI

TeleMídia

MPEG-2

- ▶ SIF (Source Input Format): 352x 240 x30 (low level)
- ▶ TV está baseada no Perfil Principal e no Nível Principal do padrão MPEG, MP@ML (Main Profile at the Main Level). CCIR 601: 720 x 480 x 30 (main level)
- ▶ HDTV está baseada no Perfil Principal e no Nível Alto do padrão MPEG, MP@HL (Main Profile at the High Level).



PUC-Rio / DI

TeleMídia

Perfil Principal MPEG-2

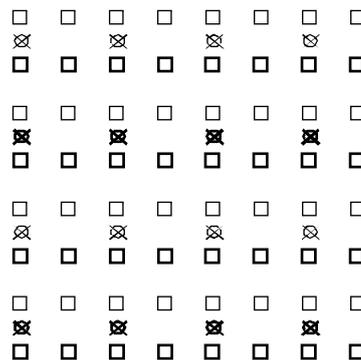
- ➔ Três tipos de quadros para predição: I, B e P
- ➔ Amostragem de luminância e crominância 4:2:0 dentro de cada quadro
- ➔ Não prevê a escalabilidade, isto é, o algoritmo não permite que seja decodificada apenas uma parte de uma unidade de dados MPEG



PUC-Rio / DI

TeleMídia

MPEG-2 Método 4:2:0



Amostragem 4:2:0
 requer a interpolação
 por média aritmética
 simples dos valores
 das duas amostras,
 imediatamente acima
 e abaixo, para
 obtenção do
 croma

- Luma sample
- ⊗ Cb sample
- Cr sample

NOTE: **heavy line** indicates
 bottom field, if interlaced



PUC-Rio / DI

TeleMídia

MPEG-2

- ➔ Bloco 8 x 8
- ➔ Macrobloco contém 4 blocos (luminância 16 x 16, crominância 8 x 8)

| <i>Formato (linhas)</i> | <i>Linhas de Macroblocos</i> | <i>Macroblocos por Linha</i> | <i>Total de Macroblocos</i> |
|-------------------------|------------------------------|------------------------------|-----------------------------|
| 720 | 45 | 80 | 3.600 |
| 1080 | 68 | 120 | 8.160 |



PUC-Rio / DI

TeleMídia

MPEG-2 System

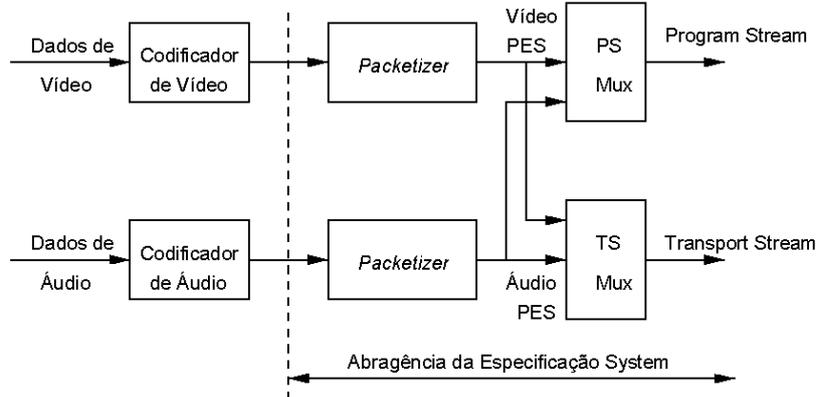
- ➔ Dois formatos de sequência de dados
 - Transport Stream: Otimizada para aplicações onde a perda de dados é comum. Tamanho de pacote de sequência fixado em 188 bytes, incluindo 4 bytes de cabeçalho.
 - Program Stream: Otimizada para aplicações multimídia (e para ser compatível com o MPEG-1)



PUC-Rio / DI

TeleMídia

MPEG-2 System

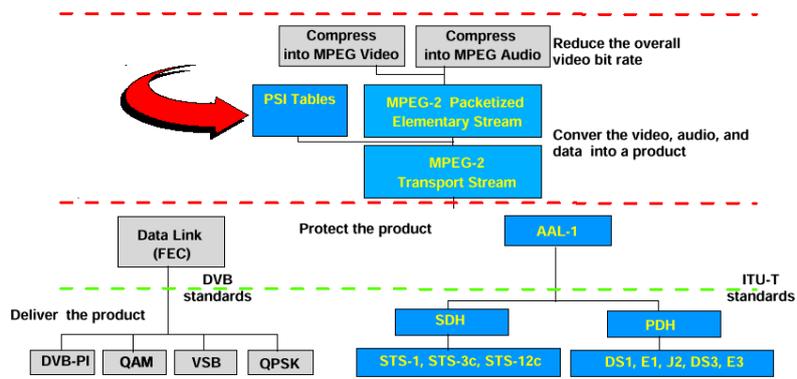


PUC-Rio / DI

TeleMídia

Fluxo de Transporte

Protocol stack



PUC-Rio / DI

TeleMídia

MPEG-4



PUC-Rio / DI

TeleMídia

MPEG 4

- ➔ 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.



PUC-Rio / DI

TeleMídia

MPEG-4

➔ Novas funcionalidades:

- Uso em ambientes sujeitos a falhas (redes móveis, redes sem fio, etc.)
- Não necessidade da decodificação do vídeo para poder manipulá-lo e editar as sequências
- Acessos randômicos para baixa taxa de bits
- Escalabilidade
- Métodos eficientes para combinar cenários sintéticos e naturais em um mesmo vídeo.



PUC-Rio / DI

TeleMídia

MPEG 4

- ➔ Parte 1- Sistema
- ➔ Parte 2 – Visual
- ➔ Parte 3 – Audio
- ➔ Parte 4 – Testes de Conformidade
- ➔ Parte 5 – Software de Referência
- ➔ Parte 6 –Delivery Multimedia Integration Framework (DMIF)
- ➔ Parte 7 – Softwares Otimizados para ferramentas MPEG-4
- ➔ Parte 8 – MPEG-4 sobre IP
- ➔ Parte 9 – Hardware de Referência
- ➔ Parte 10 – Codificação Avançada de Vídeo



PUC-Rio / DI

TeleMídia

MPEG 4

- ▶ Diferente da codificação linear de áudio e vídeo do MPEG 1 e 2, a codificação MPEG 4 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.



PUC-Rio / DI

TeleMídia

Cenas MPEG 4

- ▶ O MPEG 4 considera uma cena como sendo composta de *Objetos de Vídeo - OV*s.
- ▶ Os OV's têm propriedades como forma, movimento, textura etc.
- ▶ Eles vão se constituir das 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 OV em dado instante de tempo. Cada quadro consiste de vários VOPs.
- ▶ Uma cena que contém somente um VOP corresponde aos padrões correntes, tais como MPEG 1 ou 2. Cada VOP tem sua própria resolução espacial e temporal.



PUC-Rio / DI

TeleMídia

Cenas MPEG 4

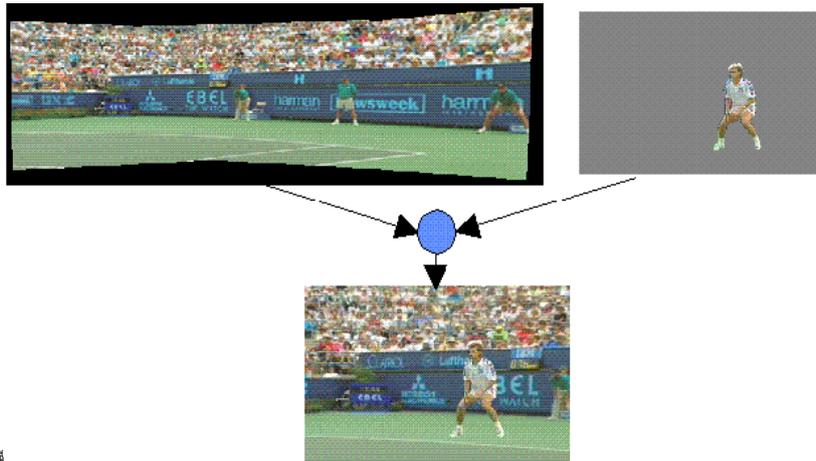
- Uma cena, dividida em objetos como mencionado, possui uma organização hierárquica.
- Uma informação adicional é enviada com os VOPs a fim de informar ao receptor como compor a cena.
- A descrição de cada cena baseia-se em conceitos da Virtual Reality Modeling Language (VRML – Linguagem de Modelagem 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).



PUC-Rio / DI

TeleMídia

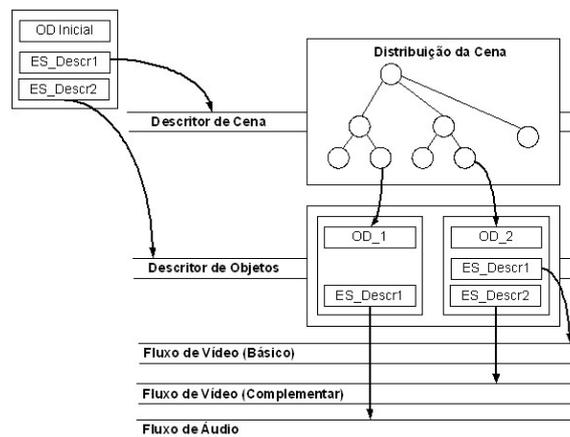
Cenas MPEG 4



PUC-Rio / DI

TeleMídia

Fluxo MPEG 4



PUC-Rio / DI

TeleMídia

Fluxos

➔ Fluxos Elementares (ES):

- Realizam a codificação de todos os objetos audiovisuais utilizados em uma apresentação

➔ Descritor de Objetos (OD):

- Descreve todos os objetos audiovisuais representados pelos fluxos elementares

➔ Descritor de Cena:

- Contém as relações temporais e espaciais dos objetos da apresentação
- Codificações de objetos sintéticos



PUC-Rio / DI

TeleMídia

Descritor de Objetos (OD)

- Fluxos Elementares Individuais (ES)
 - Escaláveis
 - Alternativas de qualidade e conteúdo
- Informações Semânticas sobre o Objeto – Object Content Information (OCI)
- Regras para o acesso ao conteúdo – Intellectual Property Management and Protection (IPMP)
- Permitem ao terminal adaptar a informação a sua capacidade
 - Perfis e níveis



PUC-Rio / DI

TeleMídia

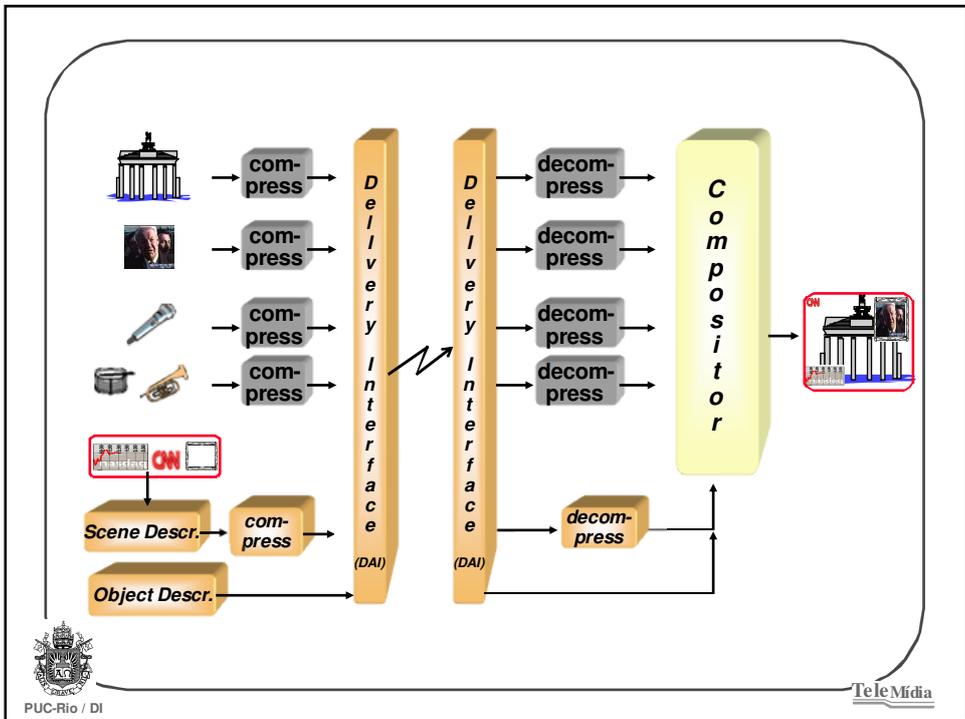
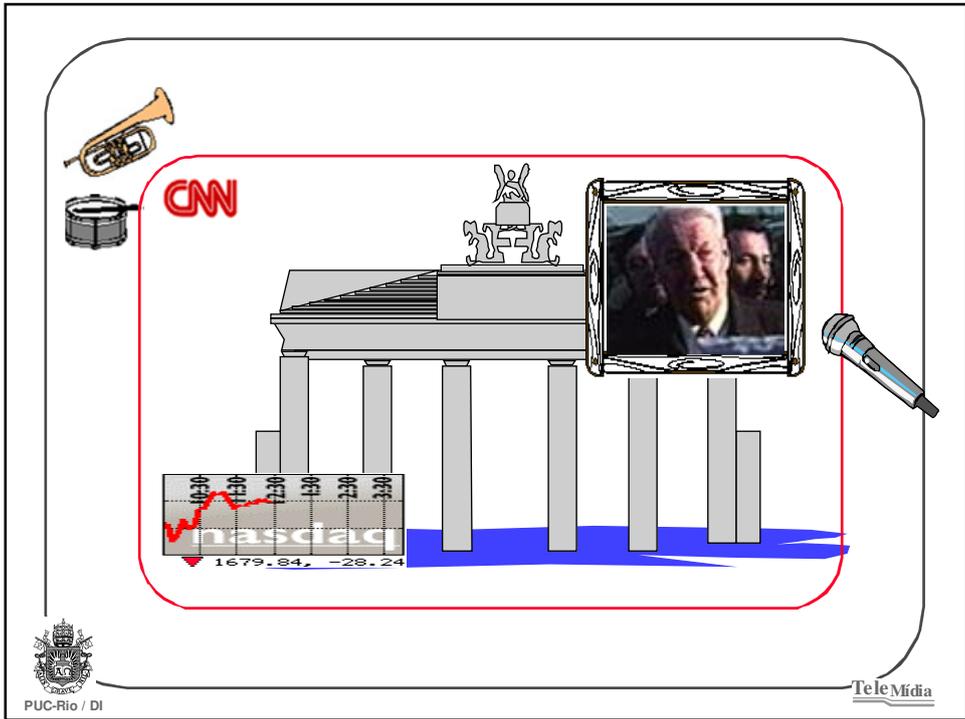
Descritor de Fluxo (ESD)

- Prioridade do fluxo
- Informações para decodificação
- Informações gerais - Taxa, buffers necessários, tipo etc
- Qualidade de Serviço
- Informações para a camada de sincronização
- Ligações do fluxo com OCI e IPMP
- Extensões futuras

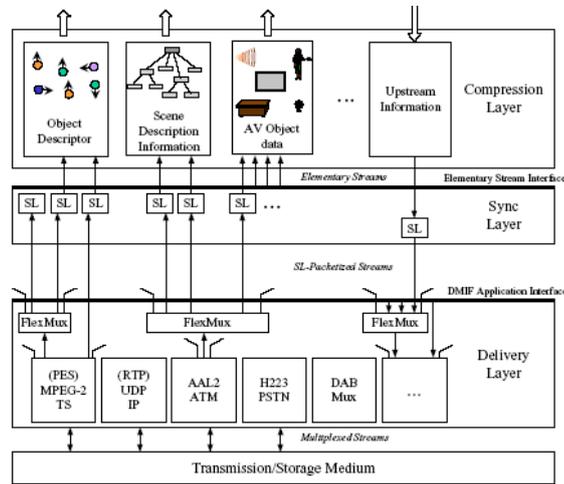


PUC-Rio / DI

TeleMídia



DMIF (DAI)



PUC-Rio / DI

TeleMídia

Sincronização de Fluxos

➔ O tempo para o MPEG-4 é sempre relativo

- Início da descrição da cena
- Eventos definidos
- Intervalos.

➔ Define os relacionamentos entre os objetos

➔ Time Stamps

Descritor da Cena

| | |
|------------------------------|------------------------|
| Mostrar Vídeo em 10 segundos | Encerrar o vídeo agora |
|------------------------------|------------------------|

Fluxo de Vídeo



PUC-Rio / DI

TeleMídia

Sincronização de Fluxos

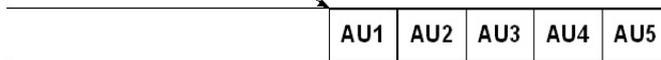
➔ Access Unit – Unidade discreta de informação

- Fluxos de descrição de cenas
- Fluxos OD
- OCI
- IPMP
- Fluxos de Mídia

Descritor da Cena



Fluxo de Vídeo



PUC-Rio / DI

TeleMídia

Camada de Sincronização

➔ Pacotes SL

- Permite a fragmentação de AU
- Cabeçalho especifica informações do ES
- Número de sequência
- Time Stamps
- Object Clock Reference (OCR)



PUC-Rio / DI

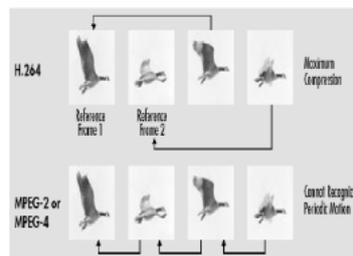
TeleMídia

Camada de Sincronização



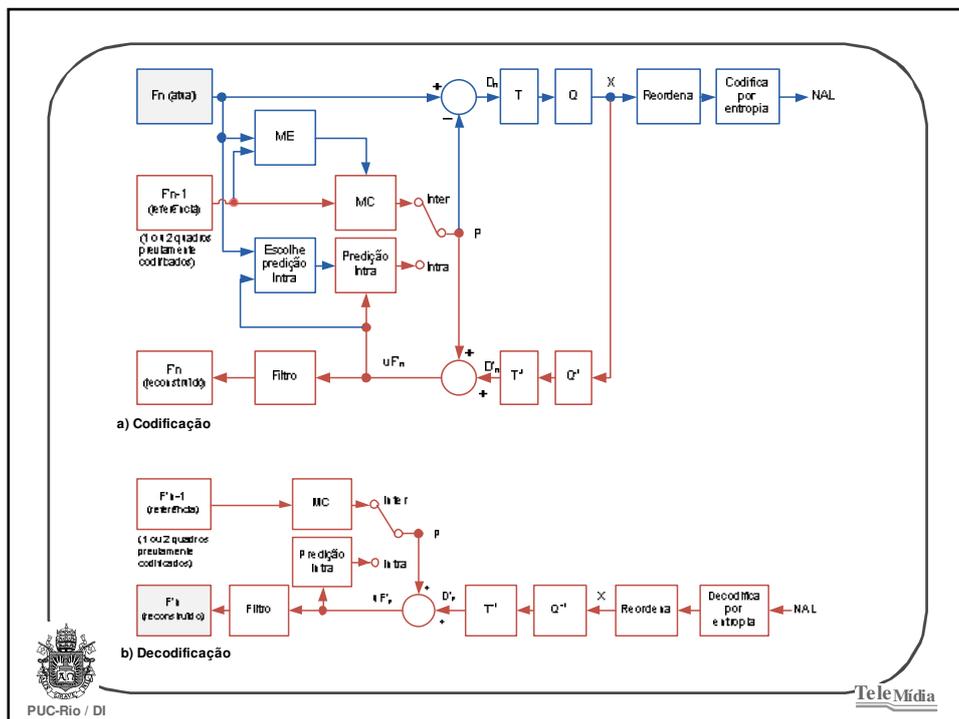
PUC-Rio / DI

TeleMídia



PUC-Rio / DI

TeleMídia



H.264/MPEG-4 AVC Part 10

➔ Esforço conjunto

- ITU- Video Coding Experts Group (VCEG) and
- ISO/IEC Moving Picture Experts Group (MPEG)
- Terminado em 2003

➔ H.264 – codec

- Padrão para **Blu-ray Discs**
- Padrão para video streaming internet no YouTube e **iTunes Store**
- Serviços de Broadcast



H.264 Characteristics

➔ Multi-view Video Coding (MVC)

- Construção de bit-streams que representam **mais de um vídeo** de uma única cena
 - Exemplo: vídeo (two-view) estereoscópico
 - Exemplo: televisão com multi-câmera
 - Exemplo: televisão 3D
- Dis perfis MVC:
 - **Multi-view High Profile** (número arbitrário de visões);
 - **Stereo High Profile** (vídeo estereoscópico com duas visões);
- Terminado em 2009



CS 414 - Spring 2012

TeleMídia

MVC

- ➔ Contém um grande número de **dependências estatísticas** entre visões
 - Câmeras capturam a mesma cena de diferentes pontos de visada
- ➔ Combina predição temporal e entre visadas
 - Ponto chave para uma codificação MVC eficiente
 - Quadros de uma certa câmera pode ser predito não apenas tendo como base quadros temporalmente relacionados da mesma câmera, mas também de **câmeras vizinhas**



PUC-Rio / DI

TeleMídia

Características das Várias Mídias



PUC-Rio / DI

TeleMídia

Característica

- ➔ Natureza do tráfego gerado (CBR ou VBR)
- ➔ Retardo máximo de transferência
- ➔ Variação estatística do retardo
- ➔ Vazão média
- ➔ Taxas aceitáveis de erro de bits e de pacotes de dados



PUC-Rio / DI

TeleMídia

Natureza do 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)
- a classe de tráfego contínuo com taxa variável (Variable Bit Rate VBR).



PUC-Rio / DI

TeleMídia

Requisitos de Comunicação das Mídias

| | <i>Texto</i> | <i>Áudio não Comprimido</i> | <i>Áudio Comprimido</i> | <i>Imagem Estática</i> | <i>Vídeo Comprimido</i> |
|--|--------------|-----------------------------|-------------------------|------------------------|-------------------------|
| <i>Retardo Máximo de Transferência</i> | Não Importa | Importa | Importa | Não Importa | Importa |
| <i>Variação Estatística do Retardo</i> | Não Importa | Importa | Importa | Não Importa | Importa |
| <i>Natureza do Tráfego Gerado</i> | Rajada | CBR | VBR | Rajada | VBR |
| <i>Taxa de Dados Média</i> | Kbps | Kbps a Mbps | Kbps | Kbps a Mbps | Mbps |
| <i>Tolerância a Erros</i> | ZERO ! | Alta Bit = Pacote | Média Bit > Pacote | Alta Bit > Pacote | Alta Bit > Pacote |



PUC-Rio / DI

TeleMídia