

Qualidade de Vídeo em Ambientes de Ensino e Aprendizagem Não-Presenciais

Luiz Fernando Gomes Soares, Marco Antonio Casanova

Depto. de Informática, PUC-Rio
R. Marquês de São Vicente 225
22453-900 - Rio de Janeiro, Brasil

lfgs@inf.puc-rio, casanova@inf.puc-rio.br

Tel.:+55-21-512-2299 ext.4347
Fax: +55-21-511-5645

1 - Introdução

As aplicações de ensino e aprendizagem não-presencial estão contidas na classe de *aplicações em banda larga*, segundo classificação utilizada no mundo das telecomunicações. O termo banda larga vem do fato de tais aplicações exigirem do sistema de comunicação uma maior largura de banda do que as tradicionais aplicações de comunicação de dados.

As aplicações de banda larga, por sua vez, se subdividem em quatro classes de serviços, três das quais inserem as diversas formas de ensino e aprendizagem não-presencial: síncrono e assíncrono.

A classe de *serviços conversacionais* define um conjunto de aplicações que provêm os meios para transferência de informação fim a fim em tempo real. O ponto chave dessas aplicações é prover uma comunicação *interativa* entre os usuários. Dentre essas aplicações encontram-se os sistemas de teleconferência, mais especificamente, os vários tipos de teleconferência, como discutiremos posteriormente. Os sistemas de teleconferência oferecem suporte ao ensino e aprendizagem não-presencial síncrono.

A classe de *serviços de recuperação* provê meios para recuperação de informação armazenada remotamente. O termo remoto pode ser entendido em vários níveis de distribuição da informação: um servidor em uma rede local, em uma rede

metropolitana ou mesmo em uma inter-rede como a Internet. É nessa classe de serviços que encontram-se as aplicações que provêm suporte ao ensino e aprendizagem não-presencial assíncrono. O ponto chave dessas aplicações é fornecer uma largura de banda que possa atender a diversos usuários compartilhando seus serviços, mas defasados no tempo. Entre as aplicações dessa classe estão os serviços de vídeo sob demanda.

A classe de *serviços de distribuição* provê meios para difusão de informação, geradas em tempo real, mas sem a interatividade com o usuário. Exemplo de serviços nessa classe são as difusões de vídeo (TV ou HDTV) e áudio. Como pode ser facilmente percebido, essa classe de serviço também provê suporte para o ensino e aprendizagem não-presencial síncrono.

Diversas características dessas classes de serviços vão demandar requisitos bastante difíceis de serem cumpridos pelo sistema de comunicação.

Os serviços conversacionais vão exigir um sistema de comunicação com pequeno retardo de transferência e pequena variação estatística do retardo, sob pena de perder a interatividade. Se formos rígidos e seguirmos os padrões, o retardo de transferência, por exemplo, não pode exceder 80ms. Uma comunicação de satélite com apenas um salto introduz um retardo de 600ms. Uma conversação nesse sistema já começa a apresentar problemas de interatividade entre os participantes (facilmente notados em uma conversação telefônica via satélite). Uma interatividade “perfeita”, vai exigir um sistema de comunicação onde a qualidade de serviço (QoS) em termos de retardo e variação estatística de retardo possa ser garantida.

Os serviços de recuperação apresentam uma forte demanda de banda passante do sistema de comunicação. Dependendo de quão esparsos no tempo estejam seus usuários e dependendo do atraso que podem tolerar no acesso à informação remota, os requisitos de banda passante, tanto do sistema de comunicação quanto do servidor remoto, podem ser extremamente elevados.

Tanto os serviços conversacionais, quanto os de recuperação, quanto os de difusão, apresentam requisitos de banda passante dependente da qualidade das informações transmitidas. Normalmente, a qualidade do vídeo transmitido é a mais

exigente e é onde centraremos o restante desse artigo.

Este trabalho está organizado da seguinte forma. A seção 2 define os diferentes serviços de teleconferência envolvendo a transmissão de vídeo. A seção 3 discute as várias qualidades de sinais de vídeo e resume as principais características de alguns sinais, entre eles alguns padrões. Finalmente, a seção 4 apresenta algumas considerações finais.

2 - Vídeo em Sistemas de Teleconferência

Tanto os serviços de recuperação quanto os serviços de distribuição, quando aplicados ao ensino e aprendizagem não-presencial, usualmente apresentam uma única fonte de vídeo (servidor) que é distribuída aos vários usuários (clientes) do serviço. Normalmente, o vídeo, nesses casos, apresenta uma qualidade igual ou superior à qualidade de um videocassete (a ser discutida na próxima seção).

Já os sistemas de teleconferência (serviço conversacional) podem trabalhar com a interação de mais de uma fonte de vídeo, que podem assumir qualidades mais baixas (qualidades de sinal) que os serviços mencionados no parágrafo anterior, sem perda da qualidade da informação que se quer transmitir. Note que aqui foi introduzido o conceito de *qualidade de sinal* e *qualidade da informação* transmitida. Utilizando um caso extremo para exemplificar, um vídeo transmitido a 30 quadros por segundo, certamente tem uma qualidade de sinal melhor do que se fosse transmitido a 10 quadros por segundo; por outro lado, se estivéssemos filmando uma paisagem sem qualquer movimento, a qualidade da informação transmitida seria a mesma (nesse caso extremo bastaria a transmissão de um quadro).

Os sistemas de teleconferência podem ser classificados quanto às mídias que utilizam para transporte de informação. Neste trabalho estamos interessados nos sistemas que transportam a mídia vídeo. Mais precisamente, estamos interessados em conferências em que, além do envio de sinais de controle, áudio, do tratamento de documentos textuais, e do envio de informações visuais estáticas (gráficos, figuras, etc.), há também transmissão de sinal de vídeo. Esses sistemas podem ser classificados como a seguir.

Freeze-frame Videoconferência: onde as imagens dos conferencistas são regularmente distribuídas (a taxas menores que 25 quadros/segundo) a todos os participantes.

Videoconferência: onde as imagens dos conferencistas são transmitidas em tempo real (isto é, 25 a 30 quadros/segundo).

Teleseminários: onde existe a distribuição de áudio e vídeo de uma localidade para todas as demais localidades, contudo com o retorno de apenas um sinal, por exemplo, áudio ou texto.

Obviamente entre esses extremos podemos encontrar outros sistemas, como por exemplo, uma videoconferência onde é difundido o sinal de vídeo do controlador (no nosso caso o professor) da teleconferência e de um dos participantes (por exemplo, o aluno que está colocando uma questão). Ou ainda outro exemplo, um teleseminário onde o vídeo transmitido é a menos de 25 quadros/segundo, uma espécie de freeze-frame teleseminário.

O tipo de serviço necessário para dar suporte ao ensino e aprendizagem não-presencial síncrono vai depender muito do material a ser apresentado. Por exemplo, em uma aula não presencial onde existe muito movimento de cena do apresentador (imagine uma aula de dança ou teatro), uma freeze-frame videoconferência não pode ser utilizada.

Devido a importância do sinal de vídeo transmitido é comum encontrarmos uma outra classificação dos serviços de teleconferência em simplesmente: serviços de videoconferência básicos e serviços de videoconferência de alta qualidade; onde os serviços de videoconferência de alta qualidade fornecem uma qualidade de áudio e vídeo similar ou muito próxima dos parâmetros de qualidade aplicados à difusão de sinais de televisão (padrão CCIR 601, entre outros - ver próxima seção), enquanto os serviços de videoconferência básicos fornecem uma transmissão de sinais de áudio e vídeo com qualidade reduzida (G711 e H261, por exemplo, ver próxima seção).

As mesmas observações sobre a qualidade do sinal de vídeo feitas para os serviços conversacionais valem para os serviços de recuperação e distribuição, no entanto, normalmente nesses casos já é assumida uma qualidade de vídeo pelo menos

igual a de um videocassete, como anteriormente mencionado, embora a maioria dos sistemas de apoio ao ensino e aprendizagem não-presencial não ofereça essa qualidade. Note que oferecer um sinal de vídeo com qualidade inferior à qualidade do videocassete não invalida a utilização desses sistemas no ensino e aprendizagem não-presencial, pois como vimos, a qualidade da informação pode ser excelente, mesmo com uma qualidade do sinal não muito boa, dependendo da informação transmitida.

3 - Qualidade de um Sinal de Vídeo

Quando falamos da qualidade de um sinal de vídeo, estamos nos referindo a qualidade do sinal de saída de um dispositivo de captura do vídeo. É importante notar, embora seja óbvio, que a qualidade do sinal de saída nunca poderá ser melhor que a do sinal de entrada. Assim, de nada adianta um super codificador de vídeo, quando se tem um sinal de entrada de baixa qualidade. Estamos assumindo, em todo esse trabalho, que a qualidade do decodificador seja pelo menos a mesma do codificador.

A qualidade do sinal de vídeo pode ser medida em termos da resolução geométrica, da resolução de cor, da resolução temporal e das perdas introduzidas pelo processo de codificação. Todos esses parâmetros são importantes na determinação da banda passante necessária do sistema de comunicação que dará suporte à transmissão do sinal. A escolha da qualidade do sinal dependerá da qualidade da informação que se quer transmitir.

A resolução geométrica é medida pela quantidade de pixels (resolução horizontal e vertical) de cada componente (RGB ou YCrCb) de um quadro de vídeo. Normalmente, como o olho humano é menos sensível a cor que a luminância, a resolução da componente Y (luminância) é a mais importante e a que usaremos nas tabelas de comparação. Tabelas mais completas utilizando a resolução de todos os componentes podem ser obtidas da literatura.

A resolução de cor é medida pela quantidade de bits utilizada para codificar cada componente. A resolução temporal é medida em termos de quadros/segundo.

A taxa de perda introduzida pela codificação depende dos algoritmos de compressão utilizados e do vídeo sendo transmitido. A tolerância a perdas depende da

informação a ser transmitida. Por exemplo, em uma aula sobre medicina, pode ser que uma imagem sendo transmitida não possa tolerar perda nenhuma de informação, pois a perda poderia representar exatamente a anomalia que se quer estudar. Nesse caso, um algoritmo de codificação sem perdas deve ser utilizado. Em todos os caso citados abaixo, o único algoritmo que não introduz perdas é o MJPEG sem perdas. No entanto, por ser robusto, o algoritmo elimina muito pouco da redundância de um sinal de vídeo, gerando uma alta taxa de bits, ou seja, exigindo uma grande banda passante do sistema de comunicação.

Ainda com relação às perdas devido a compressão do sinal de vídeo, algumas observações ainda se tornam necessárias para compreender qual codificador utilizar.

O sinal de vídeo de entrada em um codificador é um sinal CBR (constant bit rate), isto é, apresenta uma taxa de bits constante. Na saída do compressor (parte do codificador) o sinal se torna VBR (variable bit rate), devido a eliminação das redundâncias espaciais e temporais que um sinal de vídeo possui. Quando existe muita mudança de cena, a taxa de saída do compressor tende a ser maior do que quando as mudanças são pequenas.

Alguns codificadores foram projetados para utilizarem sistemas de comunicação onde a taxa de bits contratada é fixa (caso das LPs e da RDSI-FE). Neste caso, a taxa de bits (banda passante) contratada deveria ser a taxa de pico do sinal VBR de saída do compressor. Neste caso, quando o sinal VBR não estiver na taxa de pico, banda passante estaria sendo desperdiçada. Para melhor aproveitar a banda contratada, alguns codificadores implementam buffers na saída de forma a se poder contratar uma banda menor que a exigida pela taxa de pico. Caso o buffer tenda a esvaziar, uma realimentação aumenta a qualidade (diminui as perdas de compressão) do sinal de vídeo, aumentando a taxa de entrada no buffer. Caso o buffer tenda a saturar, a qualidade do sinal de vídeo é deteriorada de forma a diminuir a taxa de entrada no buffer. Agindo dessa forma, a saída do codificador é CBR, embora a saída do compressor seja VBR (este é o caso do codificador de vídeo H.261). Como o buffer tende a saturar quando são geradas muitas informações, isto é, quando há muita mudança de cena no vídeo, tais codificadores CBR se aplicam melhor em vídeos com poucos movimentos. A utilização de um desses codificadores em uma aula com muito

movimento, por exemplo, aula de teatro ou dança, não seria recomendado.

Quando o sistema de comunicação não reserva uma banda fixa para o canal de vídeo, codificadores de vídeo com saída VBR podem tirar um melhor proveito do meio do que codificadores com saída CBR, para uma mesma qualidade vídeo. Codificadores baseados nos padrões MPEG e MJPEG geram saída VBR. Estes codificadores também são os preferidos para os vídeos com muitas mudanças de cena, conforme explicado no parágrafo anterior.

O mesmo princípio do buffer com realimentação para o compressor pode ser utilizado não mais com a finalidade de gerar uma saída CBR, mas para adaptar a saída VBR do compressor à taxa disponível na rede. Estes compressores podem ter um bom desempenho em sistemas de comunicação que apresentam a qualidade de serviço de melhor esforço (best effort), como é o caso da Internet I atual, mas não apresentam nenhuma vantagem em redes onde a banda passante contratada é fixa (exemplificado por LPs e pela RDSI-FE). Em redes de pacotes com garantias de qualidade de serviço (exemplificado por redes ATM e pelas implementações atuais da Internet II) tais codificadores apresentam um desempenho pior que todos os outros tipos de codificadores mencionados, pois não tira proveito de uma banda passante garantida que pode ser negociada.

Vários codificadores permitem a escolha dos parâmetros resolução geométrica e resolução temporal. A escolha desses parâmetros determinará a taxa de bits produzida pelo codificador. Como exemplo de codificadores parametrizados, podemos citar os que seguem os padrões de codificação H.261 e MPEG-2 (embora seja possível encontrar codificadores que sigam o padrão com apenas uma opção).

O padrão H.261 é também conhecido como p64, exatamente pela forma como é calculada a taxa de bits gerada. A taxa é dada por $p.64Kbps$, onde o parâmetro p pode variar de 1 a 30 ($p=1$ ou $p=2$ para videofonia e $p > 5$ para videoconferência, segundo recomendação do padrão). A escolha do parâmetro p determina a escolha da resolução temporal (a taxa de quadros pode ser: 30; 15; 10; 7,5 quadros/segundo) e geométrica (dois formatos de imagem CIF e QCIF), que varia conforme a tabela que se segue.

	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)

O padrão MPEG II é organizado em diversos perfis e níveis. Perfis possuem um conjunto de características que atende a uma determinada classe de aplicações. Definem o espaço de resolução de cores e a escalabilidade do fluxo de bits. Níveis definem as resoluções geométricas máximas e mínimas, o número de amostras de luminância e crominância, número de camadas de vídeo e áudio suportadas por perfis escaláveis e a taxa máxima em bps para cada perfil.

O nível baixo (low level) é compatível com o padrão MPEG-1, com resolução geométrica e temporal dada por: SIF (Source Input Format) — 382x 240 x30. A qualidade de vídeo 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. A qualidade de vídeo HDTV está baseada no Perfil Principal e no Nível Alto do padrão MPEG, MP@HL (Main Profile at the High Level).

Terminando esta seção, é apresentada a seguir uma tabela com várias qualidades de sinal de vídeo, a título de exemplo.

	Resolução Geométrica (componente Y em pixels)	Quadros/segundo
H.261 QCIF:	144 x 176	7,5 a 30
H.261 CIF:	288 x 352	7,5 a 30
MPEG-1 MPEG-2 (SIF):	240 x 352	30
CCIR 601 (PAL-M) MPEG-2 (MP@ML)	480 x 720	30
Videocassete:	280 x 330 (≈ MPEG 1)	30
Difusão/Cabo:	335 x 450	30
Laser Vídeo:	425 x 560	30
TV via Satélite:	480 x 544	30
DVD (MPEG-2):	≈ 500 x 600	30
HDTV MPEG-2 (MP@HL):	até 1.100 x 2.000	30

4 - Considerações Finais

A qualidade da informação de vídeo, a qualidade do sinal de vídeo (resolução geométrica, resolução de cor, resolução temporal, algoritmo de compressão - incluindo nesse último a natureza do tráfego gerado, se VBR ou CBR) e a taxa de bits gerada (banda passante a ser contratada) são três parâmetros relacionados. Em uma aplicação de aprendizagem e ensino não-presencial, o tipo de informação a ser transmitida (tipo de aula, por exemplo) caracterizará a qualidade da informação de vídeo e, a partir dela, será possível escolher a qualidade do sinal de vídeo necessária (e assim o leque dos possíveis codificadores a serem utilizados) e, a partir dessa qualidade, será possível determinar a possível largura de banda a ser contratada.

Terminando este estudo, é apresentado a seguir uma tabela identificando a taxa média de bits gerado por codificadores padrão, chamando a atenção que, obviamente, a taxa média é igual a taxa de pico nos codificadores onde a natureza do tráfego gerado é CBR.

Objetos		Taxa (bps)
Videofone	H.261, p=1	64 k
Videofone	H.261, p=2	128 K
Videoconferência	H.261, p=6	384 K
Videoconferência	H.261, p=30	1,92 M
Vídeo MPEG-1	352 x 240 x 30 q/s	1,2M
Vídeo TV MPEG-2	480 x 720 x 30 q/s	~ 8 M
Vídeo TV MJPEG (sem perda)	480 x 720 x 30 q/s	~ 60 M
HDTV	(MPEG-2 MP@HL)	~ 18 M