

SUELEN SARA MOREIRA

**DESENVOLVIMENTO DE UMA METODOLOGIA
COMPARATIVA INTERATIVA PARA AVALIAÇÃO
SUBJETIVA DE VÍDEOS**

Belo Horizonte
12 de maio de 2010

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**DESENVOLVIMENTO DE UMA METODOLOGIA
COMPARATIVA INTERATIVA PARA AVALIAÇÃO
SUBJETIVA DE VÍDEOS**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Engenharia Elétrica.

Área de Concentração: Engenharia de Computação e Telecomunicações

Linha de Pesquisa: Sistemas de Computação

SUELEN SARA MOREIRA

Belo Horizonte
12 de maio de 2010



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Desenvolvimento de uma metodologia comparativa interativa para
avaliação subjetiva de vídeos

SUELEN SARA MOREIRA

Dissertação defendida e aprovada pela banca examinadora constituída por:

Prof. HANI CAMILLE YEHA – Orientador
Departamento de Engenharia Eletrônica / Escola de Engenharia
Universidade Federal de Minas Gerais

Dr. HERMES AGUIAR MAGALHÃES – Co-orientador
DSP Art Hardware e Software Ltda

Prof. MAURÍLIO NUNES VIEIRA
Departamento de Física / Instituto de Ciências Exatas
Universidade Federal de Minas Gerais

Prof. LUCIANO DE ERRICO
Departamento de Engenharia Eletrônica / Escola de Engenharia
Universidade Federal de Minas Gerais

Ma. FABIO DA SILVA LACERDA
Diretor de Tecnologia - ComunIP - Soluções de Mídia em Tempo Real S/A

Belo Horizonte, 12 de maio de 2010

Resumo

A proposta deste trabalho é o desenvolvimento de uma metodologia para a avaliação subjetiva da qualidade de vídeos baseada na análise de pares formados por um vídeo de qualidade conhecida e por um vídeo cuja qualidade deseja-se avaliar. Para isso, foram realizadas duas fases de testes com informantes. Na primeira fase, foram realizados testes MOS (*Mean Opinion Score* - Índice de Opinião Médio), os quais foram usados para a construção de um modelo capaz de estimar o MOS de um vídeo em função da sua taxa de transmissão. Os resultados obtidos indicam uma correlação elevada entre o MOS e o logaritmo da taxa de transmissão (coeficiente de correlação $> 0,9$). Para a segunda fase, foi desenvolvida uma interface interativa através da qual informantes decidiam qual de dois vídeos apresentados possuía qualidade superior. Utilizando-se esta interface, foram realizados testes subjetivos, os quais consistiram de comparações de pares compostos por um vídeo com MOS conhecido e um vídeo de teste. Para cada par, foi solicitado a cada informante que decidisse qual dos vídeos apresentados possuía qualidade superior. Caso o vídeo de teste tivesse qualidade superior à do vídeo de referência, uma nova comparação era realizada entre o vídeo de teste e um vídeo com MOS superior ao MOS do vídeo de referência anterior. Caso contrário, a nova comparação era realizada entre o vídeo de teste e um vídeo com MOS inferior ao MOS do vídeo de referência anterior. Realizando-se este processo de forma iterativa foi possível refinar os testes comparativos tanto quanto possível. A metodologia comparativa mostrou-se eficaz, fornecendo medições individuais com desvio padrão da ordem de um terço daquele obtido através dos testes MOS realizados. Assim, torna-se possível realizar medições de qualidade subjetiva utilizando-se um número reduzido de informantes, sem perda da precisão obtida através de testes MOS convencionais.

Abstract

The purpose of this study is to develop a methodology for assessing subjective quality of videos based on the analysis of pairs formed by a known quality video and a video whose quality is wished to be evaluated. For that, two phases of subjective tests were carried out. In the first phase, MOS (Mean Opinion Score) tests were performed. These tests were used to construct a model to estimate the MOS of a video as a function of its transmission rate. The results indicate a high correlation between MOS and the logarithm of the transmission rate (correlation coefficient > 0.9). For the second phase, we developed an interactive interface through which subjects decided which of two videos presented superior quality. Using this interface, subjective tests, which consisted of comparisons of pairs composed of a MOS-known video and a test video, were performed. For each pair, each subject was asked to decide which of the videos had higher quality. If the test video had higher quality than the reference video, a new comparison was performed between the test video and a video whose MOS was higher than the MOS of the previous reference video. Otherwise, a new comparison was performed between the test video and a video whose MOS was lower than the MOS of the previous reference video. Performing this process iteratively it was possible to refine the comparative tests as much as possible. The comparative methodology was effective, providing individual measurements with a standard deviation of about one third of that obtained through the MOS tests that were carried out. Thus, it becomes possible to perform measurements of subjective quality using a reduced number of subjects, without loss of the accuracy attained by conventional MOS tests.

Dedico este trabalho aos meus pais, ao meu noivo e a minha filha, pessoas que amo muito e que sempre me apoiaram incondicionalmente...

Agradecimentos

Ao meu orientador, Prof. Dr. Hani Camille Yehia, pela paciência, compreensão e persistência dispensados para realização deste trabalho.

Ao meu co-orientador, Prof. Dr. Hermes Aguiar Magalhães, pelo incentivo e auxílio.

À minha família, em especial aos meus pais José Feliciano e Mariângela, e as minhas irmãs Mirellen e Shellen por todo incentivo, auxílio e paciência durante essa longa caminhada.

À minha pequena Luiza pelo carinho e compreensão mesmo com a constante ausência da mamãe.

Ao meu noivo Marçal por toda amizade, paciência e companheirismo nestes anos.

Aos amigos do CEFALA e LABCOM, em especial João Pedro Hallack Sansão e Renata Onety por toda força moral e intelectual, incentivo e auxílio nos vários momentos de dificuldades.

Aos amigos do Synergia, em especial Bárbara e Leonardo pelos incentivos e pensamentos positivos.

À COMUNIP e à FINEP pela confiança e apoio financeiro proporcionados.

À todos que contribuíram para a concretização deste trabalho.

Sumário

1	Introdução	1
2	Fundamentação Teórica	3
2.1	Percepção Visual e Codificação de Vídeo	3
2.2	Princípios da Codificação de Vídeo	4
2.2.1	Compressão do Sinal de Vídeo	4
2.2.2	Padrões para Codificação de Vídeo	7
2.3	Qualidade de Vídeo	9
2.3.1	Medida Subjetiva de Qualidade	9
2.3.2	Medida Objetiva de Qualidade	18
2.4	Síntese	21
3	Metodologia e Configuração dos Experimentos	22
3.1	Primeira Fase de Testes	22
3.1.1	Configuração do Ambiente de Teste	22
3.1.2	Material de Teste	24
3.1.3	Especificações do Teste	27
3.1.4	Execução do Teste	27
3.2	Segunda Fase de Testes	30
3.2.1	Especificações do Teste	30
3.2.2	Execução do Teste	32
3.3	Síntese	34
4	Resultados	35
4.1	Primeira Fase de Testes	35
4.2	Segunda Fase de Testes	42
4.3	Considerações Finais	46
5	Conclusão	48
	Referências Bibliográficas	50

Lista de Figuras

1.1	Escala utilizada para avaliação subjetiva.	2
2.1	Amostragens temporal e espacial de uma sequência de vídeo.	4
2.2	Esquema de realização de testes subjetivos segundo a metodologia DSCQS.	13
2.3	Escala de avaliação utilizada na metodologia <i>Double Stimulus Continuous Quality Scale</i> (DSCQS).	14
2.4	Esquema de realização de testes para metodologia DSIS.	14
2.5	Escala de avaliação utilizada na metodologia DSIS.	15
2.6	Dispositivo de votação automático utilizado em testes segundo a metodologia SS-CQE (Wolf e Pinson, 2007).	16
2.7	Exemplo da exibição dos vídeos durante a realização do teste SDSCE (BT.500-11, 2002).	16
2.8	Sugestão de interface para realização de testes subjetivos utilizando a metodologia SAMVIQ (BT.1788, 2007).	18
2.9	Exemplo onde a qualidade percebida entre imagens com o mesmo PSNR é diferente (Winkler, 2005).	21
3.1	Layout do laboratório onde foram realizados os testes.	23
3.2	Relação entre a altura ou diagonal da tela e a distância em que o observador deve se encontrar.	23
3.3	<i>Pedestrian Area</i> , filmagem de um local destinado a pedestres.	25
3.4	<i>Rush Hour</i> , hora do <i>rush</i> na cidade de Munique.	25
3.5	<i>Riverbed</i> , leito do rio visto através da água.	26
3.6	<i>Station2</i> , vista para a estação de Munique.	26
3.7	T1 = exibição do vídeo de referência (10s) / T2 = intervalo (3s) / T3 = exibição do vídeo a ser avaliado (10s) / T4 = período para avaliação e preenchimento do questionário (5s).	28
3.8	Formulário utilizado para avaliação da sequência <i>Pedestrian Area</i>	29
3.9	Modelo obtido para a sequência <i>Pedestrian Area</i> : $MOS = 2,78 * \log_{10}(taxa) - 5,77$	30

3.10	Interface criada, com auxílio do aplicativo “Kommander” (Laffoon et al., 2002), para realização dos testes subjetivos utilizando uma metodologia de comparação relativa	31
3.11	Interface após a exibição do vídeo relativo ao botão Vídeo 5 . Observe que a decisão do observador define qual o próximo vídeo a ser assistido, ou seja, caso o vídeo relacionado ao botão Vídeo 5 seja melhor que o vídeo relacionado ao botão Tocar vídeo a ser avaliado o botão Vídeo 4 será a próxima escolha do observador. Por outro lado, caso o vídeo relacionado ao botão Vídeo 5 seja pior que o vídeo relacionado ao botão Tocar vídeo a ser avaliado o botão Vídeo 6 deverá ser a próxima escolha.	32
3.12	Formulário utilizado para avaliação das sequências <i>Pedestrian Area</i> , <i>Rush Hour</i> , <i>Riverbed</i> e <i>Station2</i> durante a segunda fase de testes.	33
4.1	Modelo para a sequência <i>Pedestrian Area</i> : $MOS = 2,78 \cdot \log_{10}(taxa) - 5,77$	39
4.2	Modelo para a sequência <i>Rush Hour</i> : $MOS = 2,20 \cdot \log_{10}(taxa) - 3,71$	40
4.3	Modelo para a sequência <i>Riverbed</i> : $MOS = 2,66 \cdot \log_{10}(taxa) - 6,52$	40
4.4	Modelo para a sequência <i>Station2</i> : $MOS = 1,82 \cdot \log_{10}(taxa) - 2,21$	41
4.5	Resultado obtido na avaliação subjetiva da sequência <i>Pedestrian Area</i> utilizando-se a metodologia proposta.	43
4.6	Resultado obtido na avaliação subjetiva da sequência <i>Rush Hour</i> utilizando-se a metodologia proposta.	43
4.7	Resultado obtido na avaliação subjetiva da sequência <i>Riverbed</i> utilizando-se a metodologia proposta.	44
4.8	Resultado obtido na avaliação subjetiva da sequência <i>Station2</i> utilizando-se a metodologia proposta.	44
4.9	Comparação entre os resultados obtidos na primeira e segunda fases de testes para a sequência <i>Pedestrian Area</i>	45

Lista de Tabelas

2.1	Taxas de bits para aplicações que utilizam o codec H.264/AVC (Alvarez et al., 2005).	8
2.2	PVD sugerida pela recomendação BT.500-11.	12
2.3	Comparação entre as metodologias de avaliação subjetiva.	19
3.1	Sequências utilizadas nos testes da primeira e segunda fases.	24
3.2	Parâmetros utilizados na codificação dos vídeos.	27
3.3	Taxas utilizadas na codificação dos vídeos exibidos para os grupos 1, 2, 3 e 4 de observadores, na primeira fase de testes subjetivos.	28
3.4	Taxas utilizadas para codificação dos vídeos comparativos na segunda fase.	34
3.5	Taxas utilizadas na codificação dos vídeos sob avaliação na segunda fase de testes.	34
4.1	Notas obtidas para os testes subjetivos realizados por 16 observadores que avaliaram o vídeo <i>Pedestrian Area</i> codificado a 20 diferentes taxas.	36
4.2	Notas obtidas para os testes subjetivos realizados por 16 observadores que avaliaram o vídeo <i>Rush Hour</i> codificado a 20 diferentes taxas.	36
4.3	Notas obtidas para os testes subjetivos realizados por 16 observadores que avaliaram o vídeo <i>Riverbed</i> codificado a 20 diferentes taxas.	36
4.4	Notas obtidas para os testes subjetivos realizados por 16 observadores que avaliaram o vídeo <i>Station2</i> codificado a 20 diferentes taxas.	37
4.5	Valores do <i>Mean Opinion Score</i> calculados através da média dos valores da Tabela 4.1 para os grupos de observadores 1, 2, 3 e 4 da sequência <i>Pedestrian Area</i>	38
4.6	Valores do <i>Mean Opinion Score</i> calculados através da média dos valores da Tabela 4.2 para os grupos de observadores 1, 2, 3 e 4 da sequência <i>Rush Hour</i>	38
4.7	Valores do <i>Mean Opinion Score</i> calculados através da média dos valores da Tabela 4.3 para os grupos de observadores 1, 2, 3 e 4 da sequência <i>Riverbed</i>	38
4.8	Valores do <i>Mean Opinion Score</i> calculados através da média dos valores da Tabela 4.4 para os grupos de observadores 1 e 2 da sequência <i>Station2</i>	39
4.9	Coefficientes da equação $MOS = A \cdot \log_{10}(taxa) + B$, modelo encontrado para as sequências de vídeo avaliadas.	41

4.10	Coeficientes de correlação entre o MOS e o logaritmo das taxas nas quais os vídeos utilizados foram codificados.	42
4.11	Desvios padrão das notas obtidas para cada uma das sequências da primeira e segunda fases de testes subjetivos.	45
4.12	Comparação entre as metodologias de avaliação subjetiva já existentes e citadas anteriormente, e a metodologia comparativa proposta (MCP).	47

Capítulo 1

Introdução

Nos últimos anos tem sido notável o avanço e a popularização da tecnologia digital. Equipamentos que antes não eram acessíveis à grande maioria da população têm se tornado comuns, tais como televisores com telas de cristal líquido (LCD - *Liquid Crystal Display*) e equipamentos leitores de mídia *blu-ray*. Acompanhando esta tendência, as pessoas cada vez mais têm tido acesso a conteúdos refinados como os exibidos em alta definição, com auxílio, por exemplo, da televisão digital (HDTV - *High Definition Television*). Esses sistemas permitem uma maior fidelidade comparada aquela fornecida pelos sistemas de televisão padrão (SDTV - *Standard Definition Television*) por exemplo. No entanto, todo esse avanço torna os espectadores ou usuários finais cada vez mais exigentes. Com isso, surge a necessidade dos provedores não só garantirem a qualidade do serviço prestado, mas também de alcançar a satisfação do usuário. Portanto, uma forma de tentar prover uma melhor experiência para o espectador é medindo a qualidade da imagem que está sendo transmitida, permitindo que o impacto visual no usuário seja conhecido e conseqüentemente variáveis do sistema (relacionadas a codificação e transmissão) possam ser manipuladas para que sejam obtidos bons resultados quanto à qualidade da imagem final recebida.

Devido à compressão exigida e a exibição em alta definição, os sistemas de imagem digital podem conter artefatos (estruturas ou aparências não naturais, dados indesejáveis). A quantidade e a nitidez destas distorções dependem fortemente do conteúdo mostrado (Winkler, 2005). Portanto, é importante avaliar os sistemas de vídeo determinando a qualidade das imagens mostradas ao espectador e descrevendo assim a sensação causada nesses. Esta avaliação pode ser realizada objetiva ou subjetivamente, sendo a segunda considerada a forma mais fiel de medição, ou seja, a que mais se aproxima da percepção humana.

Métodos de avaliação subjetiva de vídeo, tais como aqueles descritos pelo *International Telecommunication Union* nas recomendações BT.500-11 (2002) e P.910 (1999) são usualmente demorados e necessitam de muitos observadores, o que os torna trabalhosos e conseqüentemente caros (Winkler, 2005). O objetivo desta dissertação é o de apresentar e validar uma nova metodologia de avaliação subjetiva de vídeo que traga mais agilidade ao processo,

mas sem comprometer a precisão dos resultados obtidos.

Sistemas multimídia podem ter sua qualidade mensurada subjetivamente através do Índice de Opinião Média - MOS (*Mean Opinion Score*), o qual fornece uma indicação numérica da qualidade percebida pelo espectador. O MOS é obtido, em testes subjetivos, através da média das pontuações dadas pelos observadores. Estas notas são representadas por um único número, variando de 1 a 5, como pode ser observado na Figura 1.1. Vídeos avaliados com nota 5, excelente, são aqueles onde nenhum defeito pode ser percebido pelo observador, nota 4, bom, são aqueles onde o defeito é percebido e não causa desconforto ao observador, nota 3, aceitável, o defeito é percebido e causa desconforto ao observador, nota 2, ruim, apesar da grande degradação no vídeo o observador consegue visualizar alguma informação, nota 1, o vídeo torna-se ininteligível e o observador fica impossibilitado de extrair alguma informação deste.

MOS	Qualidade
5	Excelente
4	Bom
3	Aceitável
2	Ruim
1	Péssimo

Figura 1.1: Escala utilizada para avaliação subjetiva.

Na metodologia proposta, o MOS é obtido através de comparações, ou seja, a avaliação do vídeo de teste é feita comparando-o com outras sequências de vídeo que já possuem um MOS pré-definido. Sendo assim, é necessário, a princípio, obter um modelo onde a relação entre o MOS e a taxa de compressão das sequências de vídeo fica definida.

Esta dissertação, além desta introdução, Capítulo 1, possui mais quatro capítulos. O Capítulo 2 apresenta uma fundamentação teórica onde são abordados o sinal de vídeo, a compressão deste e a medição da qualidade. As metodologias de avaliação subjetiva e objetiva descritas na literatura também são abordadas neste capítulo. O Capítulo 3 descreve as metodologias e configurações dos experimentos realizados na primeira e segunda fases de testes subjetivos, sendo esta última fase a metodologia comparativa proposta. O Capítulo 4 apresenta os resultados obtidos, tanto na primeira, quanto na segunda fase de testes subjetivos, além de trazer uma discussão comparativa entre eles. Finalmente, no Capítulo 5, são apresentadas as conclusões do trabalho juntamente com as propostas de trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta os fundamentos teóricos necessários para a compreensão da metodologia proposta e dos resultados obtidos. Uma breve descrição do sistema visual humano é colocada no início do capítulo. Em seguida apresentam-se os princípios de codificação de vídeo. Finalmente, as principais técnicas de medição de qualidade de vídeo são mostradas.

2.1 Percepção Visual e Codificação de Vídeo

Os mecanismos e os conceitos associados aos sinais de vídeo se baseiam no processo de percepção de imagens pelo ser humano. O sistema de visão recebe estímulos luminosos e transfere as informações ao cérebro, que as processa criando a percepção de imagens. É um processo dinâmico com dependências temporal e espacial, pois a cada instante e em diferentes posições dentro do campo visual os estímulos e a percepção se renovam.

O sistema visual humano (HVS - *Human Visual System*) possui um mecanismo complexo, considerando-se que grande parte dos neurônios do cérebro humano estão envolvidos na percepção visual (Winkler, 2005). A forma como percebemos cores e movimentos é a chave para a definição da qualidade de vídeo. Sendo assim, o cálculo da diferença de duas imagens como *streams* de dados é além de ineficiente, pouco realista. Em outras palavras, a distorção calculada por uma métrica objetiva de qualidade como RMSE (*Root Mean Squared Error*) pode não casar com a percepção subjetiva do ser humano. O RMSE, por exemplo, pode calcular uma enorme distorção entre a imagem comprimida e a imagem original, embora esse valor possa não representar nenhum problema para o espectador caso o artefato detectado esteja em uma “região” que o HVS não perceba.

O HVS é mais sensível a informações de luminância (contraste claro/escuro) do que a informações de crominância (distinção entre tonalidades de cor com mesma luminosidade aparente). Sendo assim, os padrões de compressão de vídeo podem explorar esta característica humana para aumentar a eficiência de codificação, através da redução da taxa de amostragem dos componentes de crominância em relação aos componentes de luminância. No entanto,

distorções espaciais e temporais podem ser percebidas através de diferenças no espaço e no tempo e pela comparação com a referência (Webster et al., 1993).

2.2 Princípios da Codificação de Vídeo

Para que seja possível transmitir vídeo, utilizando de uma forma mais racional os recursos de transmissão disponíveis, reduzindo assim os dados transportados, usualmente é necessário que o material seja codificado (na fonte) e decodificado (no destino), processo referido pelo acrônimo CODEC. Sendo assim, devido à limitação de banda do canal de transmissão, a codificação de vídeo envolve sua compressão (na fonte) e descompressão (no destino), passos essenciais para aplicações que manipulam vídeos digitais, em função da enorme quantidade de informação necessária para representá-los sem nenhum tipo de compressão.

O vídeo digital é uma representação de cenas visuais naturais amostradas espacialmente e temporalmente, como pode ser observado na Figura 2.1. Uma cena é amostrada em um ponto do tempo para produzir um quadro (representação completa de uma cena visual em um dado momento). As imagens são amostradas em intervalos (por exemplo, 1/25 ou 1/30 segundos) para produzir a sensação de movimento exibida por uma sequência de vídeo (Richardson, 2003).

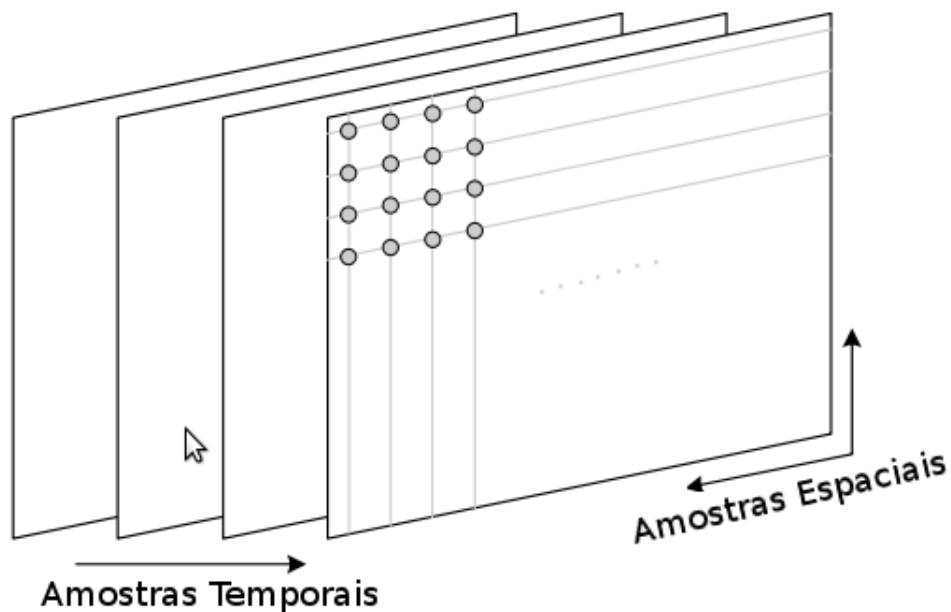


Figura 2.1: Amostragens temporal e espacial de uma sequência de vídeo.

2.2.1 Compressão do Sinal de Vídeo

As informações visuais, em geral, e as de vídeo, em particular, requerem grandes quantidades de largura de banda e espaço de armazenamento. Como exemplo observa-se que uma

sequência de vídeo sem compressão em resolução de SDTV (640x480) gera taxas que têm, tipicamente, centenas de Mbit/s. Já para vídeos em resoluções de HDTV (1920x1080) esse valor pode chegar a Gbit/s, evidenciando que os métodos de compressão são vitais para facilitar o manuseio de tais taxas. Sendo assim, um sistema para compressão de vídeo visa reduzir a taxa de transmissão e trabalha removendo a redundância e/ou informações de menor importância do sinal antes da transmissão. Fazer o processo de codificação de tal forma que a decodificação seja simples e leve também é muito importante, pois em muitas aplicações multimídia o decodificador é um item usado em larga escala pelos consumidores. Esse é um dos motivos pelo qual o decodificador é padronizado, deixando a cargo dos fornecedores, comerciais, por exemplo, a implementação do codificador.

A compressão pode ser realizada sem perdas (*lossless*) ou com perdas (*lossy*). A primeira garante a integridade da informação e, ao ser descomprimida no final do processo, ela é uma cópia exata do conteúdo original. Contudo, esses métodos apresentam desempenhos modestos quando aplicados diretamente a sinais de imagens ou vídeo. A compressão com perdas é importante considerando-se que para obter uma redução considerável da taxa (da ordem de dez a cinquenta vezes), aceita-se uma redução da qualidade. A compressão com perdas apresenta resultados variando desde alta qualidade, com artefatos não detectáveis, até visualização de degradação, tornando em alguns casos o material ininteligível.

As técnicas de compressão de vídeo difundidas são baseadas no paradigma da compressão com perdas, em que grande compressão é atingida às custas de perda de qualidade do sinal tratado, cuja versão decodificada agora passa a ser diferente, mas de certa maneira fiel à versão original. Nesse conjunto de técnicas, quanto maior a degradação inserida no sinal, menor será a versão comprimida; o desafio de um algoritmo de compressão de vídeo é comprimir eficientemente minimizando a distorção inerente ao processo.

Todo algoritmo para compressão de vídeo utiliza idéias comuns como o processamento da diferença entre quadros, estimativa preditiva do movimento e execução de uma transformação do domínio do tempo para o domínio da frequência, simplificando a descrição das partes mais frequentes da imagem. Quando se comprime vídeo, quatro tipos de redundância podem ser exploradas:

- **Redundância temporal e espacial:** valores de pixels não são independentes, mas são em grande parte correlacionados com seus vizinhos, tanto dentro do mesmo quadro (redundância espacial) quanto entre quadros consecutivos (redundância temporal). Assim, dentro de alguns limites, o valor de um pixel pode ser predito a partir dos valores dos pixels vizinhos assim como regiões de um quadro futuro podem ser preditas a partir do quadro atual.
- **Redundância em entropia:** para qualquer sinal digitalizado, alguns valores codificados ocorrem mais frequentemente que outros. Essa característica pode ser explo-

rada através da codificação dos valores que ocorrem mais frequentemente com códigos menores, enquanto que códigos maiores podem ser usados para valores mais raros em ocorrência.

- **Redundância psicovisual:** esta forma de remoção de redundância resulta do princípio de funcionamento do olho e do cérebro humanos (sistema visual humano). Tanto o limite de definição fina de detalhes que o olho pode resolver (limites de resolução espacial), quanto o limite relacionado a habilidade de acompanhar imagens que se movem rapidamente (limites de resolução temporal), são utilizados como limiares para que seja descartado aquele sub-conjunto do fluxo de informação de vídeo que ultrapassa esses limites. Visto que o sistema visual humano não é capaz de perceber esse tipo de informação, não há razão para sua transmissão, resultando assim em compressão.

Para vídeos analógicos, as redundâncias espacial, temporal e psicovisual são exploradas através de técnicas de codificação de cores e técnicas de entrelaçamento. Vídeos digitais oferecem métodos adicionais de compressão, que podem ser classificados como métodos baseados em modelos (compressão fractal, por exemplo) e métodos baseados em transformadas (*Fourier* e *Wavelet*), entre outros. Muitos dos codificadores atuais, e os padrões que os especificam, pertencem à última categoria e abrangem os seguintes estágios:

- Utilização de transformada: a fim de facilitar a exploração de redundâncias psicovisuais, os quadros são convertidos para um domínio onde diferentes faixas de frequências, percebidas distintamente pelo HVS, possam ser separadas, removendo assim, a redundância espacial, através da concentração de energia em poucos coeficientes. Isso pode ser feito através da Transformada Discreta do Cosseno (DCT - *Discrete Cosine Transform*) ou pela transformada *Wavelet*, por exemplo (Winkler, 2005).
- Quantização: após a transformação, a precisão numérica dos coeficientes é reduzida de forma a decrementar o número de bits do fluxo de bits. O grau de quantização aplicado a cada coeficiente é normalmente determinado pela percepção da distorção resultante por um observador humano. Coeficientes de alta frequência podem ser quantizados mais grosseiramente que os coeficientes de baixa frequência. O estágio de quantização acarreta perda de informação (Winkler, 2005).
- Codificação: ademais à quantização dos dados em um conjunto finito de valores, esses podem ser codificados, sem perdas, através da exploração de redundância entre os coeficientes do *stream* de bits quantizados. A codificação de entropia, que se apoia no fato de que certos símbolos ocorrem com mais frequência que outros, é normalmente utilizada neste processo (Winkler, 2005).

2.2.2 Padrões para Codificação de Vídeo

O MPEG - *Moving Picture Experts Group* é um grupo de estudos do ISO - *International Organization for Standardization* que desenvolve padrões internacionais para compressão de vídeo. Esse grupo foi criado em Janeiro de 1988 e engloba os padrões mais populares para codificação de vídeo, sendo eles:

MPEG-1: padrão mais antigo, lançado em 1992, foi desenvolvido para armazenar sinais digitais de áudio e vídeo com qualidade VCR (*Vídeo Cassete Recorder*), com uma taxa de transmissão máxima de 1,5 Mbps.

MPEG-2: padrão para televisão digital aprovado em 1994. Esse é a evolução do padrão MPEG-1 permitindo resoluções mais altas e conseqüentemente exigindo recursos computacionais melhores. Apesar do algoritmo de compressão no MPEG-1 e MPEG-2 ser fundamentalmente o mesmo, o MPEG-2 trouxe algumas melhorias importantes. Como exemplo podemos citar a capacidade de misturar trechos de qualidades diferentes dentro de um mesmo vídeo, permitindo que sejam usados menos bits nas cenas estáticas e mais bits em cenas com muito movimento (taxa de bits variável, mantendo o vídeo uniforme e diminuindo o tamanho do arquivo final), redução da redundância temporal utilizando a técnica preditiva de compensação de movimento, estruturação de perfis e níveis possibilitando maior compatibilidade e eficiência. Um perfil é definido como um subconjunto da sintaxe completa do fluxo de bits ou seja, um subconjunto de ferramentas de codificação. Todos os decodificadores compatíveis com determinado perfil devem suportar as ferramentas de codificação correspondentes àquele. Além disso foram introduzidas funcionalidades tais como escalabilidade, tratamento diferenciado para fontes entrelaçadas e maior eficiência na decodificação.

MPEG-4: padrão desenvolvido para aplicações multimídia aprovado em 1998. O padrão provê um conjunto de tecnologias para satisfazer as necessidades de autores, provedores de serviços e usuários finais tais como: robustez em ambientes suscetíveis a erros, funções de interatividade para conteúdos multimídia e alta eficiência de compressão a baixas taxas de transmissão. Esse padrão fornece as seguintes funcionalidades: melhor eficiência de codificação (ou seja, resultados melhores, em termos da qualidade da sequência codificada (Koenen, 2002)), flexibilidade a erros possibilitando transmissões robustas e possibilidade de interação áudio-visual. Avaliações subjetivas mostraram que o ganho de eficiência de codificação do MPEG-4 em relação ao MPEG-2 varia de 15 a 20%, não sendo esse suficiente para justificar a substituição do MPEG-2 dada a incompatibilidade entre estes dois padrões (Sunna, 2005).

MPEG-4 parte 10: As aplicações atuais e as emergentes, como vídeos em alta definição, necessitam de uma qualidade melhor e ao mesmo tempo uma compressão mais eficiente

do que aquelas obtidas com os padrões MPEG-2 e MPEG-4, por exemplo. Sendo assim, o MPEG-4 parte 10, também conhecido como H.264 ou AVC - *Advanced Video Coding*, foi definido com objetivo de prover uma maior eficiência na codificação de vídeo através do uso de técnicas de compressão. Com esse propósito foram realizadas alterações relativas a compensação de movimento, redução da redundância espacial, quantização e codificação de entropia (Sunna, 2005). Além dessas, outras melhorias podem ser visualizadas em de Queiroz et al. (2006).

Este padrão de compressão de vídeo foi desenvolvido em conjunto pelo ISO MPEG e pelo grupo VCEG - *Video Coding Experts Group* do ITU-T - *International Telecommunication Union*, que juntos formaram o JVT - *Joint Video Team*. O H.264 atinge as mais elevadas taxas de compressão dentre todos os padrões de codificação de vídeo existentes (pode-se obter vídeos com a mesma qualidade obtida com o MPEG-2 mas com praticamente a metade da taxa). A Tabela 2.1 mostra as taxas de bits mais comuns para aplicações que utilizam o H.264. Assim como no MPEG-2, a especificação H.264 define perfis e níveis com o objetivo de maximizar a inter-operabilidade, tornando-o capaz de atender vários tipos de aplicações, taxas, resoluções, qualidade e serviços e consequentemente proporcionando uma boa aceitação do padrão.

Tabela 2.1: Taxas de bits para aplicações que utilizam o codec H.264/AVC (Alvarez et al., 2005).

Aplicação	Resolução	Taxa de Quadros	Taxa de bits sem compressão	Taxa de bits com compressão
HDTV	1280x720	25	270 Mbps	2-8 Mbps
DVD	720x576	25	121 Mbps	1-2 Mbps
Vídeo Conferência	352x288	25	30 Mbps	128-1000 kbps
Vídeo Celular	176x144	15	9 Mbps	50-1000 kbps

Ao avaliar o desempenho do decodificador de vídeo H.264/AVC, com ênfase para vídeos em alta resolução, percebe-se que esse exige uma capacidade de processamento muito maior que os codecs MPEG-4 e MPEG-2, necessitando assim de recursos computacionais mais robustos (Alvarez et al., 2005). Apesar disso, testes subjetivos utilizando seqüências de vídeo em SD - *Standard Definition* e HD - *High Definition*, comparando o H.264 e o MPEG-2, mostram a superioridade do primeiro. Isto porque o H.264 possibilita a entrega de vídeos em boa qualidade a taxas tão baixas quanto 1,5Mbps para seqüências SD e 6Mbps para seqüências HD (essa melhoria foi obtida devido a aprimoramentos na compensação de movimento, por exemplo, como pode ser observado em Sunna (2005)). Estas são taxas onde o MPEG-2 não entregaria seqüências com qualidade aceitável (Oelbaum et al., 2004).

É importante notar que os padrões MPEG especificam somente a sintaxe do fluxo de bits e o processo de decodificação, não especificando portanto o processo de codificação, estando esse livre para propostas de novas técnicas de codificação que visem melhorar o desempenho do sistema.

2.3 Qualidade de Vídeo

Para especificar, avaliar e comparar sistemas de vídeo é necessário determinar a qualidade das imagens mostradas ao espectador. A qualidade visual é inerentemente subjetiva, tornando sua medida difícil e frequentemente imprecisa devido a vários fatores que podem afetar os resultados, como, por exemplo, a opinião de um espectador. Isto ocorre porque esta pode depender diretamente do tipo de vídeo que ele está assistindo: um filme em um DVD passivamente, ativamente participando de uma videoconferência ou se comunicando usando linguagem de sinais.

A avaliação da qualidade de vídeo pode ser realizada através de métodos objetivos ou métodos subjetivos. As medidas subjetivas permitem capturar, através de apresentações de sequências de vídeo, a percepção de qualidade pelo usuário, sendo esta a forma mais confiável de se quantificar a qualidade de vídeo. Esse também é método mais eficiente de testar a performance o desempenho de modelos que tentam simular o sistema visual humano e também métodos de avaliação objetiva. Por outro lado, métodos de avaliação objetiva são baseados em modelos matemáticos que objetivam prever, automaticamente, a qualidade das sequências de vídeo. No entanto, a predição da qualidade de vídeo é uma tarefa difícil, devido à complexidade do sistema visual humano.

2.3.1 Medida Subjetiva de Qualidade

A medida subjetiva de qualidade pode ser influenciada por vários fatores, como pode ser observado abaixo:

Interesse individual e expectativas do observador: geralmente os espectadores possuem preferências em relação ao conteúdo de vídeo exibido, seja ele através de sistemas televisivos, DVD ou Internet. Isso implica no fato de que um espectador, fã de futebol, que segue atentamente um jogo, pode ter uma necessidade de qualidade diferente daquele que esporadicamente tem contato com esse tipo de conteúdo. Os observadores também possuem expectativas diferentes em relação à qualidade de vídeo, de acordo com a situação vivida por eles, ou seja, são esperadas qualidades diferentes para vídeos assistidos em um cinema em relação àqueles exibidos por um telefone celular, por exemplo. Ao mesmo tempo, os avanços tecnológicos também podem alterar as expectativas dos

usuários, ou seja, filmes gravados em DVD são considerados por aqueles que possuem um *Blu-ray player* inferiores àqueles gravados em mídias *Blu-ray*, por exemplo.

Tipo de *display* e suas características: atualmente existem varias tecnologias de *displays* disponíveis, entre elas, CRT, LCD, Plasma e LED (*Light Emitting Diode*) . Todos esses *displays* têm características diferentes em termos de brilho, contraste, tempo de resposta, entre outras, que determinam a qualidade do vídeo exibido. Artefatos gerados pela compressão, como o efeito bloco (*blockness*) são mais visíveis em monitores com tecnologias diferentes da CRT, por exemplo (Winkler, 2005). A resolução do *display*, juntamente com seu tamanho, também influenciam na percepção de qualidade. *Displays* de cristal líquido e plasma, por exemplo, deixam distorções introduzidas pela compressão mais visíveis quando comparados a *displays* com tecnologia CRT (Tourancheau et al., 2007).

Segundo Pinson e Wolf (2004), equipamentos com tecnologia CRT com resoluções maiores (720x486) podem ser utilizados para emular a experiência subjetiva de observadores utilizando equipamentos LCD (celulares, por exemplo) com baixas resoluções (352x288). Além disso, a influência da tecnologia de *displays* LCD e CRT e de seus ajustes (tempo de resposta, exibição de artefatos e resolução) pode ser reduzida, significativamente, na avaliação subjetiva de sequências de vídeo, quando utilizadas as resoluções citadas anteriormente. Além disso, Pinson e Wolf (2004) acreditam que o impacto introduzido pela tecnologia do *display* utilizado pode ser reduzido através do emprego de testes subjetivos onde dois estímulos são utilizados (avaliação da diferença entre um vídeo original e outro processado) em detrimento de testes onde apenas um estímulo é utilizado (avaliação apenas do vídeo processado).

Condições de visualização: além da distância entre o observador e o *display*, a iluminação do ambiente também afeta a percepção da qualidade. Apesar do observador estar apto a se adaptar a vários níveis de luz, ambientes onde estes níveis estão sempre mudando diminuem a sensibilidade a pequenas variações de contraste. Além disso, a luz externa ao ambiente onde o observador se encontra pode diminuir a visibilidade da imagem na tela devido as reflexões que ocorrem nesta (Winkler, 2005).

Presença do áudio: o áudio que acompanha o vídeo tem uma grande influência na qualidade percebida pelo usuário. Estudos mostram que as notas dadas por observadores, em testes de avaliação subjetiva, são geralmente maiores quando as cenas são acompanhadas por som de boa qualidade (Winkler, 2005).

Testes subjetivos para avaliação de qualidade visual têm sido formalizados nas recomendações ITU-R BT.500-11 (BT.500-11, 2002) e ITU-T P.910 (P.910, 1999). Essas recomendações sugerem condições gerais sob as quais a avaliação subjetiva de vídeo deve ser realizada

(condições de observação, critérios para seleção de observadores e materiais utilizados nos testes, procedimentos de avaliação e métodos para avaliação dos dados obtidos nos testes, por exemplo). Portanto, as recomendações têm o objetivo principal de proporcionar aos cientistas uma forma de obter resultados homogêneos, ou seja, úteis em futuras comparações.

A recomendação BT.500-11 tem uma história longa e foi escrita com objetivo de padronizar metodologias de avaliação subjetiva para conteúdos televisivos. Ela vem sendo enriquecida ao longo do tempo, sendo que nos últimos anos muitos métodos foram adicionados com o intuito de melhorar a confiabilidade dos resultados (Baroncini, 2006).

No final dos anos 90 (1999), com a evolução da tecnologia dos computadores, surgiu a necessidade de uma nova recomendação. Sendo assim, o setor de Telecomunicações do ITU lançou a recomendação P.910, intitulada "*Subjective video quality assessment methods for multimedia applications*" (Métodos de avaliação subjetiva de qualidade para aplicações multimídia) que, como o próprio nome diz, foi criada com intuito de atender diversos tipos de conteúdos como, por exemplo, videoconferência, vídeos utilizados em equipamentos portáteis e na Internet, etc.

Na recomendação P.910 as metodologias de testes foram descritas utilizando termos mais adequados às necessidades dos serviços de telecomunicações. Afora isso, as metodologias de testes subjetivos mostraram-se uma cópia daquelas descritas na recomendação BT.500. Ou seja, a nova recomendação não resolveu completamente os problemas que surgiram na avaliação subjetiva de vídeos utilizados em serviços multimídia, tais como: descrição do *setup* de laboratórios para realização de testes utilizando monitores de computadores (ou seja, utilizar um *setup* que se aproxime de situações reais) e possibilidade de utilização de *displays* alternativos (projetores, por exemplo) (Baroncini, 2006).

As recomendações BT.500-11 e P.910 sugerem várias condições experimentais para que sejam realizados os testes subjetivos. Algumas dessas condições serão brevemente discutidas a seguir.

Condições de visualização: o *setup* do ambiente onde os testes são realizados pode afetar os resultados experimentais. Sendo assim, alguns fatores devem ser especialmente observados durante a realização do experimento: iluminação do local, presença de ruídos no ambiente, escolha e calibração do *display* utilizado (resolução, brilho, contraste, posicionamento em relação ao espectador).

A distância do observador à tela e o tamanho desta, segundo a recomendação BT.500-11, devem ser selecionados para satisfazer a PVD - *Preferred Viewing Distance*. Esta relação pode ser observada na Tabela 2.2. Os valores exibidos podem ser usados para sistemas SDTV e HDTV. Já a recomendação P.910 determina que a distância do observador à tela deve ser definida não somente levando-se em consideração o tamanho desta, mas também o tipo de tela, o tipo de aplicação e o objetivo do experimento. Para a recomendação P.910 essa distância pode variar de 1 a 8 vezes a altura da imagem exibida.

Tabela 2.2: PVD sugerida pela recomendação BT.500-11.

Diagonal da tela (polegadas)		Altura da tela (H) (cm)	PVD
Tela relação 4/3	Tela relação 16/9		
12	15	18	9H
15	18	23	8H
20	24	30	7H
29	36	45	6H
60	73	91	5H
>100	>120	>153	3-4H

Seleção do material utilizado na realização do teste: os resultados obtidos através de testes subjetivos dependem fortemente do conteúdo das sequências de vídeo assistidas pelos observadores. As informações temporais e espaciais são importantes parâmetros que devem ser observados nessa escolha. Esses parâmetros determinam o esforço realizado para compressão e conseqüentemente o nível de artefatos introduzidos. Além disso, a duração das sequências é também um fator a ser observado. Essa duração depende do objetivo do teste. Por exemplo, a avaliação de CODECs geralmente é realizada utilizando-se sequências de vídeo curtas. Por outro lado, se efeitos produzidos por erros de transmissão precisam ser avaliados continuamente, sequências longas são mais indicadas.

Seleção dos observadores: este é um ponto importante para avaliação subjetiva de vídeo. O número de observadores selecionados para participar dos testes pode variar de 4 a 40 (P.910, 1999), sendo normalmente utilizadas 15 pessoas (BT.500-11, 2002). A consistência dos resultados e também a precisão desses é diretamente proporcional ao número de observadores utilizados.

Antes da realização dos testes os observadores devem ser avaliados, a fim de determinar se esses não possuem alguma anomalia relacionada ao sistema visual humano. Portanto, é desejável que os observadores tenham acuidade visual normal (capacidade de discriminar dois pontos próximos como elementos separados, o que corresponde à nitidez da visão) e percepção normal das cores (ausência de daltonismo).

Os observadores selecionados para participar de testes subjetivos não devem estar envolvidos com avaliação da qualidade de vídeo ou imagens em seu trabalho cotidiano. Ou seja, não devem ser especialistas no assunto. Isso porque os não especialistas fazem parte de um grupo mais representativo, quando comparados com os especialistas. Observadores não especialistas assistem aos vídeos sob teste como se esses não estivessem sendo avaliados, de forma natural.

Por outro lado, quando o teste tem a intenção de avaliar/procurar artefatos específicos presentes em uma sequência de vídeo, um grupo de avaliadores experientes realizaria melhor o teste.

2.3.1.1 Métodos de Avaliação Subjetiva

A recomendação BT.500-11 ilustra vários procedimentos para realização de testes subjetivos, sendo os mais utilizados:

Double Stimulus Continuous Quality Scale (DSCQS) - Escala de Qualidade Contínua por Estímulo Duplo: é o procedimento mais comumente utilizado, onde são apresentados ao observador múltiplos pares de sequências de vídeo, sendo uma delas a sequência de teste e outra a referência. As sequências devem ser, preferencialmente, de curta duração (em torno de 10 segundos).

Na metodologia DSCQS, a ordem de exibição entre as sequências de referência e teste é aleatória, e o observador não é informado de qual tipo de sequência está assistindo, se teste ou referência. As sequências de vídeo são exibidas duas vezes, como pode ser observado na Figura 2.2, e em seguida avaliadas.

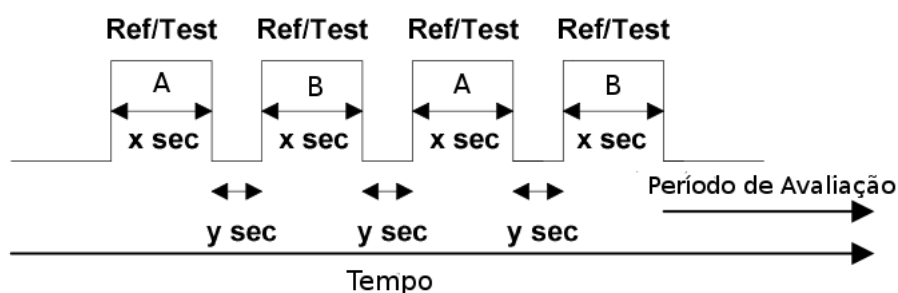


Figura 2.2: Esquema de realização de testes subjetivos segundo a metodologia DSCQS.

Após a apresentação, cada sequência é avaliada separadamente através de uma escala de qualidade contínua variando de Péssimo a Excelente, de acordo com a Figura 2.3.

A análise dos resultados é baseada na diferença entre a pontuação de cada par avaliado, que é calculada num equivalente numérico na faixa entre 0 (zero) e 100 (cem). Essa diferenciação auxilia na remoção da subjetividade inserida devido à experiência do observador e também ao conteúdo da cena. Além disso, uma outra forma de reduzir a influência desses fatores seria através da repetição do teste, utilizando-se grupos diferentes de observadores e sequências, o que aumenta o custo do experimento.

Esta metodologia é tipicamente utilizada para avaliações onde a diferença entre as sequências de teste e referência não é muito grande (Alpert et al., 1997), (Winkler, 2005).

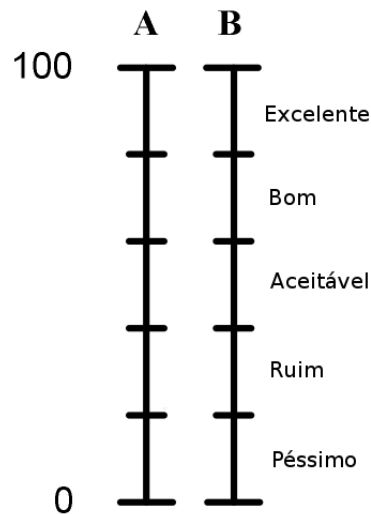
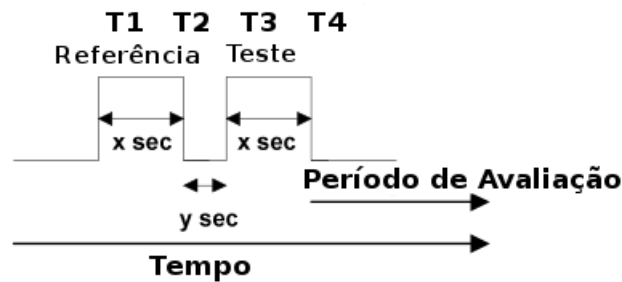


Figura 2.3: Escala de avaliação utilizada na metodologia *Double Stimulus Continuous Quality Scale* (DSCQS).

Double Stimulus Impairment Scale (DSIS) - Escala de Deficiência por Estímulo Duplo: nesta metodologia, diferentemente da anteriormente mostrada (DSCQS), a referência é sempre mostrada antes da sequência de teste, não havendo repetição. O esquema do teste, mostrado na Figura 2.4, prevê tempos definidos para cada apresentação, assim como tempo de recuperação e descanso durante a transição (imagem cinza), e um tempo final para o julgamento.



Legenda:

- T1 = 10s Exibição da Referência (x sec);
- T2 = 3s Imagem cinza (y sec);
- T3 = 10s Exibição Sequência de Teste (x sec)
- T4 = 5 - 11s Imagem cinza (y sec);

Figura 2.4: Esquema de realização de testes para metodologia DSIS.

Os observadores avaliam a qualidade da sequência de teste utilizando uma escala discreta de cinco níveis, variando de Péssimo a Excelente tal qual mostrado na Figura 2.5.

Os níveis da escala da Figura 2.5 devem ser interpretados como:

- EXCELENTE
- BOM
- ACEITÁVEL
- RUIM
- PÉSSIMO

Figura 2.5: Escala de avaliação utilizada na metodologia DSIS.

- Excelente ou Imperceptível - a sequência de teste mostrada ao observador não aparenta nenhuma diferença em relação à sequência de referência.
- Bom ou Perceptível, mas não irritante - neste caso o observador consegue perceber alguma diferença entre a sequência de teste e a referência mas esta não o incomoda.
- Aceitável ou Levemente irritante - o observador percebe a presença de degradação na sequência de teste, em relação à referência, e essa o incomoda.
- Ruim ou Irritante - o observador percebe que a sequência está bastante degradada mas ainda visualiza alguma informação.
- Péssimo ou Muito irritante - neste caso, o conteúdo torna-se ininteligível, ou seja, o observador não consegue extrair nenhuma informação do vídeo exibido.

Esse método é comumente utilizado para testar a fidelidade da transmissão em relação ao sinal original, sendo um fator importante na avaliação de sistemas de alta qualidade (P.910, 1999).

Single Stimulus Continuous Quality Evaluation (SSCQE) - Avaliação de Qualidade Contínua por Estímulo Simples: neste método são apresentados aos observadores trechos de programas televisivos que tenham sido processados pelo sistema sob teste. A duração mínima de cada trecho é de 5 minutos. A referência não é mostrada. Utilizando um dispositivo armazenador de valores de qualidade (um sensor), enquanto assistem, os observadores avaliam continuamente a qualidade instantânea percebida. Um exemplo desse dispositivo pode ser observado na Figura 2.6. Nesse caso é utilizada uma escala contínua, como no método DSCQS. Porém, naquele a nota era única no final de uma apresentação curta, neste as notas são dadas de forma contínua durante o tempo da demonstração. Esse método está mais adequado para medidas de qualidade de vídeo em sequências longas e reproduz situações reais com a ausência da referência.

Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) - Estímulo Duplo Simultâneo para Avaliação Contínua: nesse teste os observadores assistem a duas



Figura 2.6: Dispositivo de votação automático utilizado em testes segundo a metodologia SSCQE (Wolf e Pinson, 2007).

sequências de vídeo ao mesmo tempo, sendo uma a referência e a outra o vídeo processado (sob teste). Caso o formato das sequências seja SIF (*Standard Image Format - 320x240*) ou menor, as duas podem ser exibidas lado a lado no mesmo monitor, como pode ser observado na Figura 2.7. Caso contrário dois monitores devem ser utilizados.

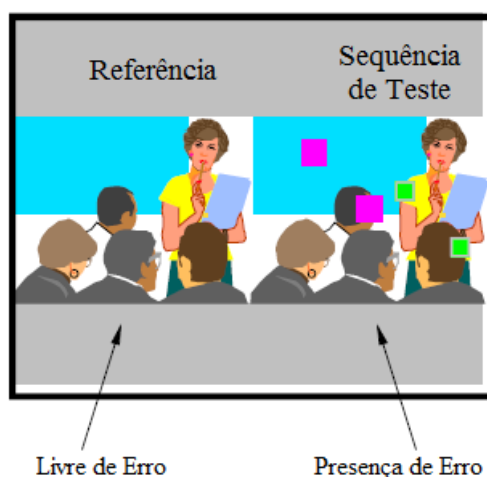


Figura 2.7: Exemplo da exibição dos vídeos durante a realização do teste SDSCE (BT.500-11, 2002).

Os observadores avaliam as diferenças entre as duas sequências de vídeo continuamente utilizando um dispositivo, semelhante ao utilizado pela metodologia SSCQE. Embora o método seja simples, os resultados podem variar de acordo com os observadores (caso sejam otimistas ou pessimistas, mais ou menos comprometidos). Isso pode ser evitado através de um processo

mais elaborado de treinamento.

Subjective Assessment Methodology for Video Quality (SAMVIQ) - Metodologia de Avaliação Subjetiva para Qualidade de Vídeo: nos últimos anos, têm sido realizados esforços no desenvolvimento de um método de avaliação subjetiva interativo capaz de avaliar aplicações multimídia (Baroncini, 2006). A metodologia de avaliação subjetiva SAMVIQ foi desenvolvida pela *France Telecom R&D* e padronizada pelo *European Broadcasting Union* (EBU). Essa metodologia utiliza vários estímulos, os quais podem ser assistidos aleatoriamente, diferindo significativamente dos testes subjetivos tradicionais em vários aspectos. O teste é realizado através da utilização de uma interface (Blin, 2002); (BT.1788, 2007) onde são apresentadas para avaliação subjetiva várias sequências de vídeo disponíveis em vários níveis de qualidade, incluindo a referência explícita e a mesma referência oculta entre os vídeos apresentados. O observador pode avaliar cada sequência na ordem desejada, podendo ainda assisti-la várias vezes, o que possibilita uma comparação entre sequências e também com a referência (essa liberdade auxilia na detecção de pequenos defeitos). A avaliação da qualidade de cada sequência pode ser expressa através de uma escala contínua exibida na tela do software (0-100), sendo que o observador pode alterar a nota atribuída várias vezes, ou seja, a cada exibição da sequência de teste. Após a avaliação de todas as sequências de um mesmo vídeo, um novo grupo é apresentado para avaliação.

A Figura 2.8 mostra a interface utilizada nesses experimentos. Ela apresenta uma série de botões que permitem que as sequências de vídeo sejam exibidas.

Alguns testes foram realizados com o objetivo de avaliar esta nova metodologia; estes podem ser encontrados em Huynh-Thu et al. (2007). Em Blin (2006) foram executados testes com o objetivo de comparar as metodologias SAMVIQ e DSCQS. Mostrou-se que testes subjetivos realizados utilizando-se SAMVIQ proporcionam resultados melhores do que aqueles realizados utilizando-se a metodologia DSCQS para vídeos no formato CIF (resolução 352x288).

2.3.1.2 Comparação entre Métodos de Avaliação Subjetiva

As metodologias de avaliação subjetiva de vídeo mencionadas anteriormente são largamente utilizadas e consideradas confiáveis. No entanto existem algumas diferenças importantes entre estas metodologias, como por exemplo o uso de uma sequência de referência explícita ou oculta, frequência da apresentação das sequências, exibição de uma ou duas sequências de vídeo simultaneamente, votação contínua ou avaliação única, etc, as quais podem ser observadas na Tabela 2.3 (Kozamernik et al., 2005).

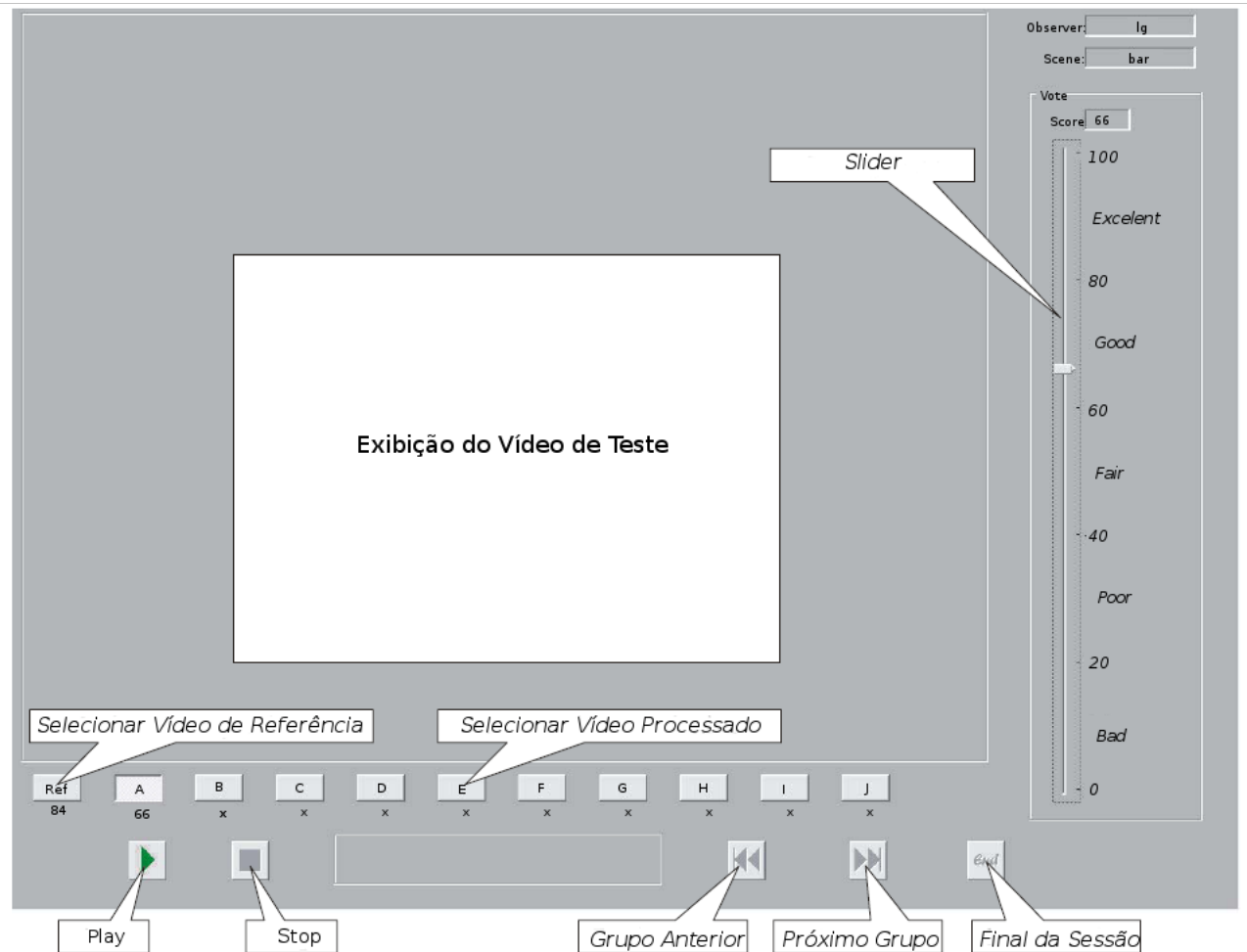


Figura 2.8: Sugestão de interface para realização de testes subjetivos utilizando a metodologia SAMVIQ (BT.1788, 2007).

2.3.2 Medida Objetiva de Qualidade

Muitos métodos computacionais têm sido desenvolvidos com intuito de medir objetivamente a qualidade de vídeo. Esses métodos são ditos objetivos por não haver interação humana, ou seja, a sequência de vídeo original e a sequência a ser avaliada (vídeo que sofreu compressão, por exemplo) alimentam um algoritmo computacional que calcula a distorção entre os dois. A medida de qualidade visual utilizando critérios objetivos possibilita resultados precisos, apesar de nenhum sistema de medida objetiva ser capaz de reproduzir a experiência do observador. No entanto, a complexidade e o custo das medidas subjetivas tornam atrativas as medidas realizadas com auxílio de algoritmos. Sendo assim, nos últimos anos têm sido desenvolvidas várias pesquisas com o intuito de aprimorar esse tipo de medida de qualidade. Em Bovik et al. (2003) podem ser encontrados vários algoritmos utilizados na obtenção da qualidade de vídeo.

As métricas de qualidade objetiva de vídeo podem ser classificadas segundo a disponibili-

Tabela 2.3: Comparação entre as metodologias de avaliação subjetiva.

Parâmetros	DSIS	DSCQS	SSCQE	SDSCE	SAMVIQ
Referência Explícita	Sim	Não	Não	Sim	Sim
Referência Oculta	Não	Sim	Não	Não	Sim
Duração da Sequência	10s	10s	5min	>10s	10s
Dois estímulos simultâneos	Não	Não	Não	Sim	Não
Frequência de apresentação do Material de Teste	Uma vez ou Duas vezes sucessivas	Duas vezes sucessivas	Uma vez	Uma vez	Várias vezes
Votação	Sequência de Teste	Sequência de Teste e Referência	Sequências de Teste	Diferença entre Sequência de Teste e Referência Mostradas Simultaneamente	Sequência de Teste e Referência
Possibilidade de Mudança da Nota	Não	Não	Não	Não	Sim
Avaliação Contínua da Qualidade	Não	Não	Sim	Sim	Não

dade do sinal de vídeo original, o qual é considerado livre de distorções e pode ser utilizado como referência; quando há presença deste, o sistema é conhecido como Referência Total (*Full-Reference* - FR). Quando não há disponibilidade do vídeo original, a avaliação objetiva é dita Sem Referência (*No-Reference* - NR). Existe ainda a avaliação objetiva de vídeo chamada Referência Reduzida (*Reduced-Reference* - RR), onde certas características são extraídas do sinal original e transmitidas para o sistema de avaliação de qualidade, com objetivo de auxiliar na detecção de falhas.

As métricas FR são as mais desenvolvidas e estudadas. Elas são tipicamente baseadas na comparação quadro a quadro entre a referência e a sequência distorcida, necessitando de um alinhamento preciso dos dois vídeos, tanto espacialmente quanto temporalmente. Esse alinhamento é importante para que os pixels de cada quadro do vídeo distorcido estejam perfeitamente casados com os respectivos pixels dos quadros do vídeo de referência, isso para permitir uma comparação quadro a quadro de boa qualidade.

Para as métricas NR, a obtenção objetiva de qualidade é realizada apenas através de informações disponíveis na ponta do receptor (decodificador, por exemplo). Diferentemente das métricas FR, as métricas NR não necessitam dos alinhamentos temporais e espaciais, entre a referência e o vídeo distorcido, sendo que nenhuma comparação quadro a quadro é realizada. A falta de referências normalmente se traduz em processos mais complexos e menos

eficientes.

Na avaliação objetiva com referência reduzida certas características do sinal de vídeo original assim como algumas medidas relacionadas a esse são extraídas e transmitidas com intuito de auxiliar na avaliação de qualidade do vídeo sob teste. Conseqüentemente, esse tipo de avaliação faz com que o sistema necessite de uma largura de banda adicional para que a informação extra, sobre o vídeo original, seja enviada.

A métrica objetiva com referência total mais utilizada é o PSNR - *Peak Signal-to-Noise Ratio*. O PSNR é medido em uma escala logarítmica e depende do erro quadrático médio (MSE - *Mean Squared Error*), entre duas imagens (uma original e uma processada), relativo ao valor $(2^n - 1)^2$, que representa o quadrado do maior valor possível que um pixel pode ter (255 para uma imagem de 8 bits). O PSNR é expresso por:

$$PSNR_{dB} = 10 \log_{10} \frac{(2^n - 1)^2}{MSE} \quad (2.1)$$

Onde:

n = número de bits por pixel.

MSE = erro quadrático médio entre o vídeo original e vídeo avaliado.

Formalmente o MSE é dado por:

$$MSE = \frac{\sum_{i=1}^M \sum_{j=1}^N [f(i, j) - F(i, j)]^2}{M \cdot N} \quad (2.2)$$

Onde $f(i, j)$ é o valor do pixel na imagem original, $F(i, j)$ é o valor do pixel na imagem avaliada, M é o número de linhas e N o número de colunas do quadro.

O PSNR pode ser calculado de forma fácil e rápida e por isso é uma métrica de avaliação objetiva de qualidade popular e bastante utilizada para avaliar a qualidade de imagens de vídeo com e sem compressão (Richardson, 2003). Geralmente, um PSNR alto indica uma alta qualidade e um PSNR baixo indica baixa qualidade. No entanto, nem sempre essa relação pode ser considerada. Por ser baseado em comparações de imagens pixel a pixel, o PSNR tem uma relação limitada com a qualidade percebida pelo sistema visual humano. Em certas situações a qualidade subjetiva de uma imagem pode ser melhorada através da adição de ruído, diminuindo, conseqüentemente, o PSNR. A Figura 2.9 mostra um exemplo onde imagens (b e c) com mesmo PSNR possuem qualidades percebidas diferentes, ou seja, a imagem b parece ter uma qualidade inferior a imagem c, observando a região superior ao telhado da casa exibida.

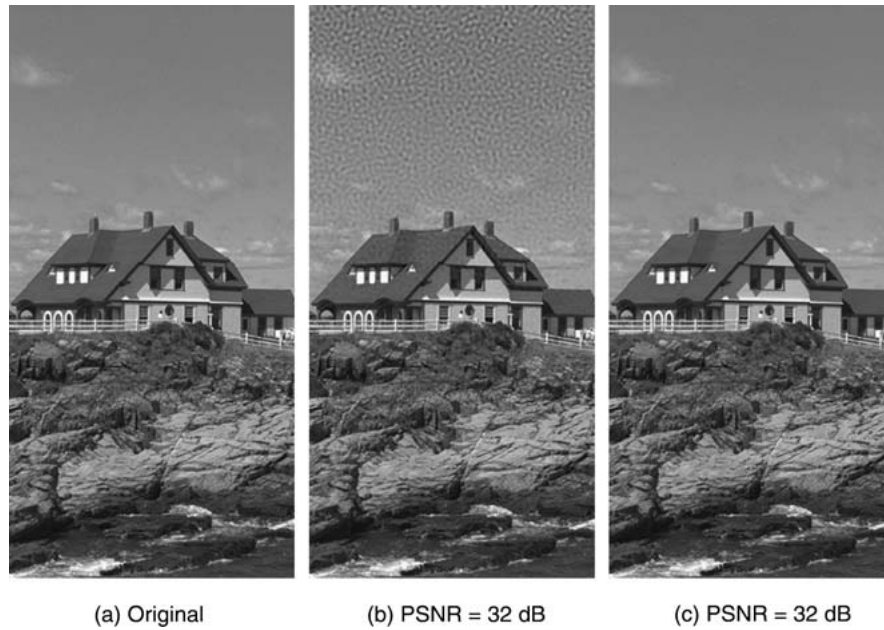


Figura 2.9: Exemplo onde a qualidade percebida entre imagens com o mesmo PSNR é diferente (Winkler, 2005).

As limitações do PSNR, como necessidade de utilização do vídeo original - que deve ser de boa qualidade - e baixa correlação com medidas subjetivas, incentivam o desenvolvimento de uma medida que se aproxime da resposta dos observadores humanos. Sendo assim nas últimas décadas um grande esforço tem sido feito para desenvolver métodos de avaliação objetiva de vídeo que incorporem a medida de qualidade percebida. Ou seja, métodos que considerem as características do sistema visual humano, implementados através do desenvolvimento de um sistema de avaliação objetiva da qualidade de vídeo que emule a percepção humana, utilizando como base testes subjetivos previamente realizados (Webster et al., 1993).

Apesar das limitações da PSNR, esta pode ser usada como um bom indicador da variação da qualidade de vídeo quando o conteúdo da sequência e o CODEC são fixos para uma dada condição de teste, como por exemplo na comparação das características de otimização do CODEC para um dado conteúdo de vídeo (Huynh-Thu e Ghanbari, 2008).

2.4 Síntese

Neste capítulo, foram apresentados métodos de medição objetiva e subjetiva de qualidade de vídeo, os quais fornecem a fundamentação teórica para o desenvolvimento da metodologia descrita no próximo capítulo.

Capítulo 3

Metodologia e Configuração dos Experimentos

Este capítulo apresenta a nova metodologia para avaliação subjetiva de vídeo proposta assim como os testes realizados para sua validação.

Os testes foram conduzidos em duas fases, sendo que em um primeiro momento, a partir de um conjunto de vídeos de teste (Muniqué, 2002), comprimidos a diferentes taxas e apresentados a uma audiência segundo a metodologia DSIS (apresentação do vídeo de referência - apresentação do vídeo a avaliar - nota), determinou-se um modelo relacionando a taxa de bits (kbit/s) a seus respectivos MOS (média das notas obtidas para vídeos codificados a uma determinada taxa). Em um segundo momento, tal modelo foi usado para, a partir de alguns MOS - coincidentes e não coincidentes com os obtidos na etapa anterior - determinar suas taxas de compressão correspondentes. Os vídeos foram comprimidos nestas novas taxas e, a seguir, apresentados à audiência para serem comparados com os vídeos de teste, também apresentados. Implementou-se então desta forma a metodologia comparativa proposta neste trabalho, detalhada a seguir.

3.1 Primeira Fase de Testes

O objetivo desta etapa é realizar a medição dos MOS para os vídeos de teste codificados a taxas que varrem toda a escala MOS.

3.1.1 Configuração do Ambiente de Teste

Os testes subjetivos da primeira fase foram conduzidos em dois dias no Centro de Estudos da Fala, Acústica, Linguagem e música (CEFALA) na Universidade Federal de Minas Gerais. Foi configurada uma estação de teste no laboratório, cujo *layout* pode ser observado na Figura 3.1.

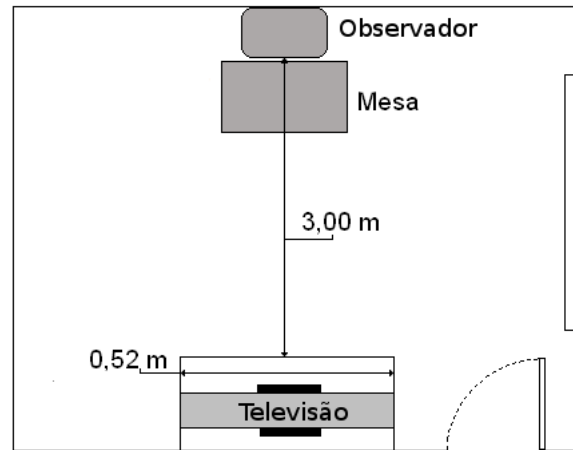


Figura 3.1: Layout do laboratório onde foram realizados os testes.

Neste trabalho, para a realização dos experimentos, a distância entre o observador e a tela utilizada foi calculada para satisfazer as regras da *Preferred viewing distance* (PVD), de acordo com a Figura 3.2, como sugerido pela recomendação do ITU-R BT.500-11 (BT.500-11, 2002).

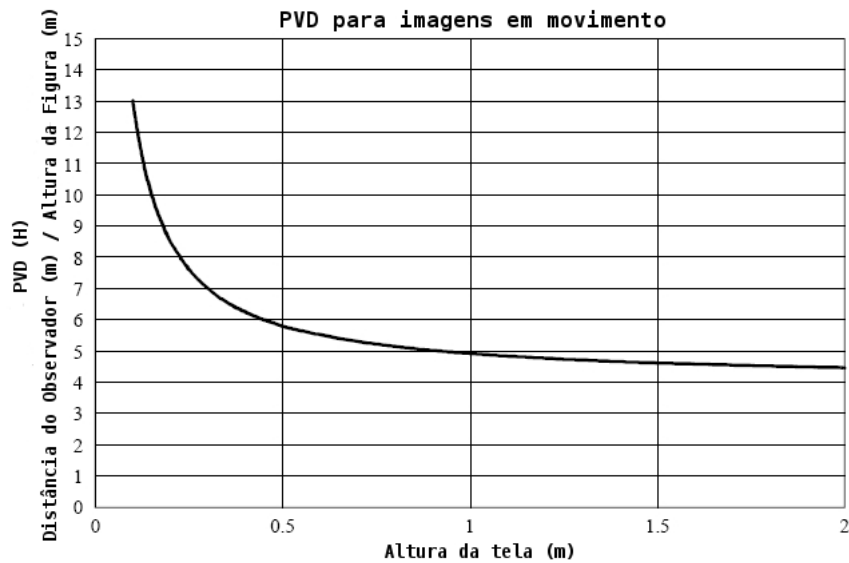


Figura 3.2: Relação entre a altura ou diagonal da tela e a distância em que o observador deve se encontrar.

O equipamento utilizado consiste em um televisor da marca PHILIPS, modelo 42PF7321/78, com tela de plasma 42" e diagonal visual aproximada de 107 centímetros, e um computador Intel Core2 Quad CPU Q9300 @ 2.50GHz. A conexão entre o televisor e o computador foi realizada através da entrada HDMI do televisor. Os ajustes de cor, contraste, nitidez, níveis de brilho, matiz e tom de cor do televisor foram obtidos automaticamente através da função *AutoPicture™* desse, onde estes parâmetros são configurados no processo de fabricação do

equipamento. Utilizou-se a opção *Natural* da função *AutoPictureTM*.

3.1.2 Material de Teste

As sequências utilizadas na realização dos testes subjetivos foram obtidas de vídeos na resolução HD-720 (1280x720) disponíveis no arquivo da Universidade Técnica de Munique (Munique, 2002). Estes foram capturados com a utilização de uma câmera digital Sony HDW-F900 com resolução de 1920x1080 pixels, a uma taxa de 25 quadros por segundo, em varredura progressiva (*progressive scan*) e usando uma subamostragem de cores de 4:2:0. A Tabela 3.1 sumariza as principais características dos vídeos utilizados.

Tabela 3.1: Sequências utilizadas nos testes da primeira e segunda fases.

Sequências de Teste	Resolução	Nº de Quadros	Quadros/Segundo	Descrição
<i>Pedestrian Area</i>	1280x720	250	25	Filmagem de um local destinado a pedestres. Câmera em posição baixa, pessoas passando bem próximas a ela. Alta profundidade de campo. Câmera estática.
<i>Rush Hour</i>	1280x720	250	25	Hora do <i>rush</i> na cidade de Munique. Muitos carros se movimentando lentamente. Alta profundidade de foco. Câmera estática.
<i>Riverbed</i>	1280x720	250	25	Leito do rio visto através da água. Muito difícil de codificar.
<i>Station2</i>	1280x720	250	25	Vista para a estação de Munique. Filmagem noturna. <i>Zoom out</i> longo. Muitos detalhes, estruturas regulares (trilhos).

As figuras 3.3, 3.4, 3.5 e 3.6 exibem *snapshots* dos vídeos utilizados nos testes.

Os vídeos utilizados nos testes também foram avaliados em Alvarez et al. (2007), com exceção da sequência *Station2*, que nesse caso foi utilizada ao invés da sequência *Blue Sky*, dado que essa última possuía um número insuficiente de quadros em relação ao desejado para duração do teste.

Para obtenção das sequências de teste, os vídeos originais que se encontravam no formato .y4m (extensão de vídeos sem codificação) foram editados com auxílio da ferramenta FFmpeg - software livre / de código aberto que produz bibliotecas e programas para manipulação de dados multimídia (FFmpeg, 2005), com objetivo de codificá-los com as características observadas na Tabela 3.2, que também relaciona todas as taxas utilizadas na primeira fase de testes.



Figura 3.3: *Pedestrian Area*, filmagem de um local destinado a pedestres.



Figura 3.4: *Rush Hour*, hora do *rush* na cidade de Munique.



Figura 3.5: *Riverbed*, leito do rio visto através da água.



Figura 3.6: *Station2*, vista para a estação de Munique.

Tabela 3.2: Parâmetros utilizados na codificação dos vídeos.

Taxas (bit/s)	300k 400k 500k 600k 700k 800k 900k 1000k 1100k 1300k 1500k 1700k 2000k 2500k 3000k 3500k 4000k 6000k 8000k 10000k
CODEC	H.264
Resolução	1280x720
Quadros/Segundo	25
Duração	10 segundos

Na codificação com o FFmpeg utilizou-se o x264 (VideoLAN, 2006) - um software de código aberto para codificação de fluxos de vídeo para o formato H.264 - com objetivo de obter vídeos com o padrão de compressão H.264 com o perfil *High Profile* (este perfil foi criado para transmissão e armazenamento de dados, geralmente em alta definição). Desde a padronização do H.264, vários codificadores têm sido implementados. Neste caso optou-se pelo x264 por ele ter código aberto (*open source*), ter um ótimo desempenho em relação a outros codificadores comerciais (Vatolin, 2009) e por ser usado em várias aplicações como FFmpeg, MEncoder, etc. Apesar da taxa de codificação ser um dos parâmetros de entrada para o FFmpeg, este nem sempre consegue alcançá-la, resultando em valores aproximados.

3.1.3 Especificações do Teste

A primeira fase de testes teve como objetivo a obtenção de dados (*Opinion Scores*) para o cálculo do MOS relativo a sequências de vídeo codificadas a diferentes taxas. Para aquisição destas “notas ” foi utilizada a metodologia de avaliação subjetiva DSIS (BT.500-11, 2002). Esta foi escolhida devido as características dos vídeos utilizados e também por ser mais direta e rápida.

Seguindo este procedimento foram apresentadas aos observadores pares de sequências de vídeo, sendo que a referência, onde não ocorrem artefatos ou outros defeitos, sempre é exibida antes da sequência em teste. Após a exibição de cada par, foi dado ao observador um período de 5 (cinco) segundos para que ele pudesse avaliar a sequência em teste, tendo como base a referência. Nesse momento uma imagem cinza foi exibida no televisor. Para a avaliação foi utilizada a escala 1 (Péssimo), 2 (Ruim), 3 (Aceitável), 4 (Bom) e 5 (Excelente). A Figura 3.7 exemplifica esta metodologia adotada nos testes subjetivos da primeira fase.

3.1.4 Execução do Teste

Os testes subjetivos da primeira fase foram conduzidos em uma única sessão, por observador, cuja duração foi de aproximadamente 19 (dezenove) minutos (a duração do teste foi a mesma para todos observadores). As sessões foram iniciadas com uma fase de treinamento, onde foram repassadas instruções, sendo que cada participante teve a oportunidade de conhecer

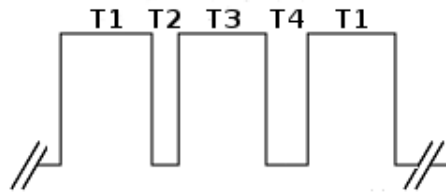


Figura 3.7: T1 = exibição do vídeo de referência (10s) / T2 = intervalo (3s) / T3 = exibição do vídeo a ser avaliado (10s) / T4 = período para avaliação e preenchimento do questionário (5s).

o teste, o formulário e tirar dúvidas. Para completar, uma breve prática da sessão de teste foi conduzida como forma de familiarizar os participantes. A sequência de vídeo utilizada no treinamento foi diferente das exibidas nas sessões de teste propriamente ditas.

As sessões foram assistidas, individualmente, por 16 (dezesseis) observadores. Cada um deles preencheu quatro formulários, um para cada sequência de vídeo avaliada. Esses observadores tinham idades entre 20 e 44 anos e possuíam formação em Ciência da Computação ou Engenharia. Nenhum deles possuía experiência na avaliação de vídeos. A Figura 3.8 mostra um dos formulários utilizados durante a realização da primeira fase de testes.

Cada observador assistiu ao todo a 40 sequências de vídeo (quatro formulários preenchidos), não sendo permitida a repetição de nenhuma delas.

Para não estender as sessões os observadores foram divididos em quatro grupos de quatro pessoas. Os grupos 1 e 2 avaliaram vídeos codificados a taxas diferentes dos grupos 3 e 4. Além disso, dentro dos respectivos pares de grupos, a ordem de apresentação também sofreu alterações, como pode ser observado nas quatro colunas da direita da Tabela 3.3.

Tabela 3.3: Taxas utilizadas na codificação dos vídeos exibidos para os grupos 1, 2, 3 e 4 de observadores, na primeira fase de testes subjetivos.

Videos Apresentados		Ordem de Apresentação			
Grupos 1 e 2 (kbit/s)	Grupos 3 e 4 (kbit/s)	Grupo 1 (kbit/s)	Grupo 2 (kbit/s)	Grupo 3 (kbit/s)	Grupo 4 (kbit/s)
300	400	8000	4000	1300	1000
500	600	1100	700	10000	1300
700	800	500	2000	1700	800
900	1000	2000	1100	400	2500
1100	1300	3000	500	1000	3500
1500	1700	300	300	2500	400
2000	2500	4000	900	6000	600
3000	3500	700	1500	600	10000
4000	6000	1500	3000	3500	1700
8000	10000	900	8000	800	6000

Avaliação Subjetiva de Vídeo - MOS (Mean Opinion Score)

Nome: _____ Idade: _____

Formação: _____

Vídeo 1 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 2 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 3 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 4 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 5 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 6 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 7 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 8 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 9 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Vídeo 10 - Pedestrian_area

EXCELENTE BOM ACEITÁVEL RUIM PÉSSIMO

Figura 3.8: Formulário utilizado para avaliação da sequência *Pedestrian Area*.

3.2 Segunda Fase de Testes

Na segunda fase foram realizados testes subjetivos utilizando a mesma configuração física do ambiente de testes e também o mesmo material da primeira fase de testes, dado que as sequências de vídeo avaliadas foram as mesmas. O objetivo agora é obter, iterativamente, a qualidade subjetiva por meio de comparações de vídeos cuja qualidade deseja-se medir com vídeos cujo MOS foi estimado com base nos modelos desenvolvidos a partir da primeira fase de testes.

3.2.1 Especificações do Teste

A metodologia de avaliação subjetiva utilizada na segunda fase de testes foi proposta com objetivo de agilizar o processo de avaliação, reduzindo o número de observadores e o tempo utilizado para realização do teste.

Nesta nova metodologia aqui proposta, o observador realiza comparações relativas entre o vídeo de teste e outros vídeos, que foram codificados a taxas segundo o modelo obtido na primeira fase, o qual relaciona a taxa de codificação do vídeo com seu MOS. Como exemplo, a Figura 3.9 mostra a relação MOS *versus* logaritmo da taxa de codificação obtida na primeira fase de testes para o vídeo *Pedestrian Area*. Como pode ser observado, cada MOS possui, segundo este modelo, uma taxa de codificação relacionada.

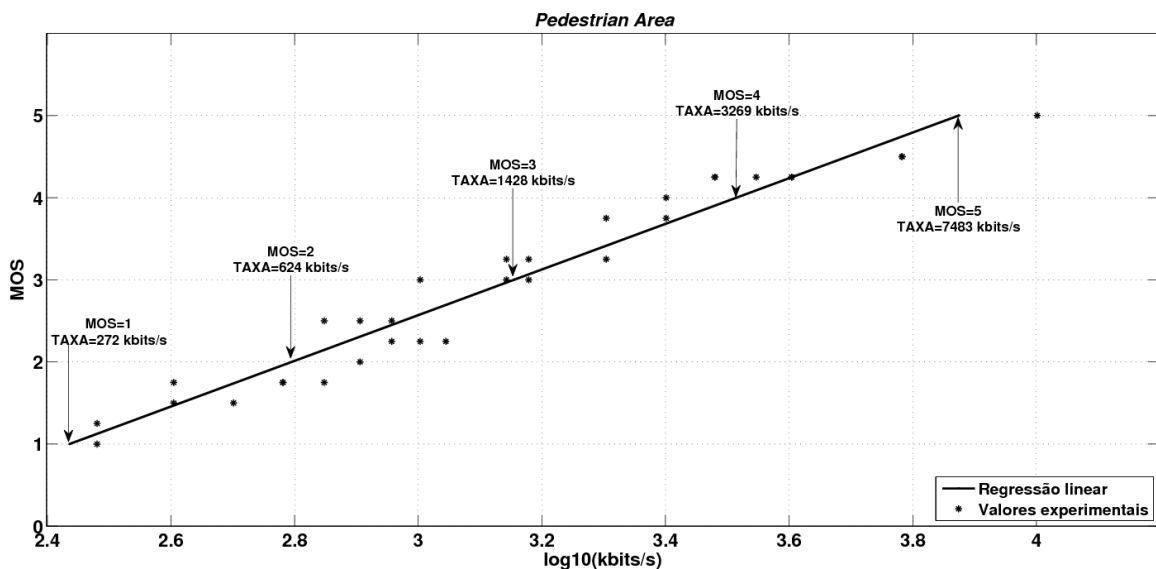


Figura 3.9: Modelo obtido para a sequência *Pedestrian Area*: $MOS = 2,78 * \log_{10}(taxa) - 5,77$.

As comparações são feitas até que o observador encontre o vídeo que mais se assemelha ao vídeo a ser avaliado, definindo assim o seu MOS. A Figura 3.10 ilustra a interface construída para realização dos testes. Observe que segundo os botões habilitados, inicialmente, o obser-

vador assiste ao vídeo a ser avaliado e depois ao vídeo relativo ao MOS=3 (**Vídeo 5**) cabendo a ele avaliar se o vídeo de teste é melhor ou pior que o vídeo exibido ao clicar no botão **Vídeo 5** (MOS=3). Optou-se por iniciar pelo **Vídeo 5** por estar este no meio da escala.



Figura 3.10: Interface criada, com auxílio do aplicativo “Kommander” (Laffoon et al., 2002), para realização dos testes subjetivos utilizando uma metodologia de comparação relativa

Após assistir ao vídeo relacionado ao botão **Vídeo 5**, outros dois botões são evidenciados (habilitados) para o observador. Sendo assim, ao realizar a comparação entre os dois vídeos assistidos (**Vídeo 5** e **Tocar vídeo a ser avaliado**) o observador deve decidir se assiste a um vídeo melhor (vídeo relacionado ao botão **Vídeo 6**) ou pior (vídeo relacionado ao botão **Vídeo 4**), quando comparados ao vídeo relacionado ao botão **Vídeo 5**. A Figura 3.11 mostra a sequência em que os botões são habilitados, de acordo com o descrito anteriormente.

Esse procedimento comparativo é realizado pelo observador até que ele encontre, entre os vídeos dos botões de 1 a 9, aquele que mais se aproxima do vídeo relacionado ao botão **Tocar vídeo a ser avaliado**.

As taxas de codificação dos vídeos relacionados aos botões de 1 a 9 exibidos na interface foram escolhidos a partir dos modelos MOS *versus* taxa de codificação resultantes das avaliações realizadas na primeira fase de testes. Os modelos para todos os vídeos utilizados são mostrados no Capítulo 4, figuras 4.1, 4.2, 4.3 e 4.4.

Os valores das taxas de codificação das sequências de vídeo para cada MOS específico, podem ser observados na Tabela 3.4.



Figura 3.11: Interface após a exibição do vídeo relativo ao botão **Vídeo 5**. Observe que a decisão do observador define qual o próximo vídeo a ser assistido, ou seja, caso o vídeo relacionado ao botão **Vídeo 5** seja melhor que o vídeo relacionado ao botão **Tocar vídeo a ser avaliado** o botão **Vídeo 4** será a próxima escolha do observador. Por outro lado, caso o vídeo relacionado ao botão **Vídeo 5** seja pior que o vídeo relacionado ao botão **Tocar vídeo a ser avaliado** o botão **Vídeo 6** deverá ser a próxima escolha.

3.2.2 Execução do Teste

Os testes subjetivos da segunda fase também foram conduzidos em uma única sessão, cuja duração por indivíduo foi de 10 a 14 minutos dependendo do observador. As sessões foram iniciadas com uma fase de treinamento, onde foram repassadas instruções. Cada participante teve a oportunidade de conhecer o teste, o formulário e tirar dúvidas. Para completar, uma breve prática da sessão de teste foi conduzida como forma de familiarizar os participantes.

As sessões foram assistidas individualmente por 10 observadores, sendo que cada um deles preencheu um formulário onde as sequências *Pedestrian Area*, *Rush Hour*, *Riverbed* e *Station2* foram avaliadas duas vezes. A Figura 3.12 mostra o formulário utilizado durante a realização da segunda fase de testes. Nesse formulário o título “Vídeo 1 - Pedestrian_area” está relacionado ao primeiro vídeo a ser avaliado pelo observador o qual foi codificado a taxa de 1100 kbit/s como mostra a Tabela 3.5 com todas as taxas em que os vídeos avaliados na segunda fase de testes foram codificados.

Como pode ser observado na Figura 3.12, a escala utilizada para avaliação das sequências diz respeito aos vídeos relacionados aos botões **Vídeo 1** a **Vídeo 9** que foram assistidos durante as comparações. Caso o vídeo em avaliação no momento seja igual ao vídeo relacionado ao botão **Vídeo 2**, por exemplo, em termos de qualidade, o quadrado número 2 da escala do formulário deverá ser marcado.

Avaliação Subjetiva de Vídeo – Segunda Parte

Nome: _____

Formação: _____

Vídeo 1 – Pedestrian_area

1		2		3		4		5		6		7		8		9
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---

Vídeo 2 – Pedestrian_area

1		2		3		4		5		6		7		8		9
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---

Vídeo 3 – Rush_hour

1		2		3		4		5		6		7		8		9
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---

Vídeo 4 – Rush_hour

1		2		3		4		5		6		7		8		9
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---

Vídeo 5 – Riverbed

1		2		3		4		5		6		7		8		9
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---

Vídeo 6 – Riverbed

1		2		3		4		5		6		7		8		9
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---

Vídeo 7 – Station2

1		2		3		4		5		6		7		8		9
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---

Vídeo 8 – Station2

1		2		3		4		5		6		7		8		9
---	--	---	--	---	--	---	--	---	--	---	--	---	--	---	--	---

Figura 3.12: Formulário utilizado para avaliação das sequências *Pedestrian Area*, *Rush Hour*, *Riverbed* e *Station2* durante a segunda fase de testes.

Tabela 3.4: Taxas utilizadas para codificação dos vídeos comparativos na segunda fase.

<i>Pedestrian Area</i>									
	Vídeo 1	Vídeo 2	Vídeo 3	Vídeo 4	Vídeo 5	Vídeo 6	Vídeo 7	Vídeo 8	Vídeo 9
MOS	1	1,5	2	2,5	3	3,5	4	4,5	5
Taxa (kbit/s)	272	412	624	944	1428	2160	3269	4946	7483
<i>Rush Hour</i>									
	Vídeo 1	Vídeo 2	Vídeo 3	Vídeo 4	Vídeo 5	Vídeo 6	Vídeo 7	Vídeo 8	Vídeo 9
MOS	1	1,5	2	2,5	3	3,5	4	4,5	5
Taxa (kbit/s)	138	233	394	665	1122	1894	3196	5393	9101
<i>Riverbed</i>									
	Vídeo 1	Vídeo 2	Vídeo 3	Vídeo 4	Vídeo 5	Vídeo 6	Vídeo 7	Vídeo 8	Vídeo 9
MOS	1	1,5	2	2,5	3	3,5	4	4,5	5
Taxa (kbit/s)	672	1035	1596	2460	3793	5847	9013	13895	21420
<i>Sation2</i>									
	Vídeo 1	Vídeo 2	Vídeo 3	Vídeo 4	Vídeo 5	Vídeo 6	Vídeo 7	Vídeo 8	Vídeo 9
MOS	1	1,5	2	2,5	3	3,5	4	4,5	5
Taxa (kbit/s)	58	109	206	387	729	1372	2583	4862	9153

Tabela 3.5: Taxas utilizadas na codificação dos vídeos sob avaliação na segunda fase de testes.

<i>Pedestrian_area</i>		<i>Riverbed</i>	
Vídeo 1	1100 (kbit/s)	Vídeo 5	6000 (kbit/s)
Vídeo 2	6000 (kbit/s)	Vídeo 6	2000 (kbit/s)
<i>Rush_hour</i>		<i>Station2</i>	
Vídeo 3	700 (kbit/s)	Vídeo 7	1100 (kbit/s)
Vídeo 4	4000 (kbit/s)	Vídeo 8	500 (kbit/s)

Em algumas situações o observador pode concluir que o vídeo sob teste possui qualidade melhor que o vídeo relacionado ao botão **Vídeo 6** e pior que o vídeo relacionado ao botão **Vídeo 7**, ou seja, um valor intermediário, existindo, no formulário de avaliação também esta opção, representada pelo quadrado branco entre os números 6 e 7.

Os testes subjetivos, segundo a metodologia comparativa, não possuem um número fixo de exibições de cada sequência de vídeo, dado que os observadores têm liberdade para assistir novamente tanto ao vídeo relacionado ao botão **Tocar vídeo a ser avaliado** quando os vídeos relacionados aos botões **Vídeo 1** a **Vídeo 9**.

3.3 Síntese

Neste capítulo, foi descrita a metodologia desenvolvida para medição subjetiva da qualidade de vídeo. Os resultados obtidos experimentalmente são descritos a seguir.

Capítulo 4

Resultados

No capítulo anterior foram descritas as metodologias conduzidas nas duas fases de testes subjetivos. Neste capítulo são apresentados os resultados obtidos na primeira fase de testes e a forma como estes foram trabalhados a fim de se obter um modelo, o qual foi utilizado para validar a metodologia de avaliação subjetiva proposta, objeto da segunda fase, cujos resultados também são exibidos aqui.

4.1 Primeira Fase de Testes

Os resultados da primeira fase de testes subjetivos, para a sequência de vídeo *Pedestrian Area*, podem ser visualizados na Tabela 4.1 onde são exibidas notas (*Opinion Scores*), que variam de 1 a 5, para os vídeos codificados nas taxas selecionadas. Note que os 16 observadores que participaram dos testes subjetivos da primeira fase estão divididos em 4 grupos de 4 observadores cada.

O primeiro e segundo grupos de observadores avaliaram a sequência de vídeo codificada a 10 taxas, distintas daquelas avaliadas pelo terceiro e quarto grupos de observadores. Apesar de avaliarem vídeos a taxas iguais, a ordem de exibição desses foi diferente entre o primeiro e segundo grupos assim como entre o terceiro e quarto, conforme exposto anteriormente na Tabela 3.3.

O mesmo procedimento foi adotado para as sequências *Rush Hour*, *Riverbed* e *Station2*, sendo que as notas (*Opinion Scores*) obtidas em relação a estas sequências podem ser visualizadas, respectivamente, nas tabelas 4.2, 4.3 e 4.4. Observe nessas tabelas que as taxas de codificação exibidas são próximas, mas ligeiramente diferentes daquelas mostradas na Tabela 3.3. Essa diferença ocorre devido a tolerância que o CODEC possui em relação a taxa de codificação desejada.

Tabela 4.1: Notas obtidas para os testes subjetivos realizados por 16 observadores que avaliaram o vídeo *Pedestrian Area* codificado a 20 diferentes taxas.

		Taxas (kbit/s)										Taxas (kbit/s)									
		302	503	704	906	1108	1509	2014	3019	4020	8097	402	604	805	1007	1389	1711	2518	3521	6062	10033
Grupo 1	Grupo 1	2	2	3	2	3	4	4	5	5	5	1	2	2	2	2	4	3	4	4	5
		1	1	3	3	2	4	4	5	5	5	2	2	3	3	4	4	4	5	5	5
		1	1	2	2	2	2	3	3	3	3	1	1	1	2	3	3	4	4	5	5
		1	1	2	2	2	2	4	4	4	5	2	2	2	2	3	4	4	4	4	5
Grupo 2	Grupo 2	1	1	2	2	2	4	4	5	4	5	2	2	3	4	4	5	4	5	5	5
		1	1	1	2	1	2	3	3	4	4	2	2	3	3	3	4	4	5	5	5
		1	2	2	3	3	3	3	4	4	4	2	2	2	3	3	3	4	4	5	4
		1	2	2	3	2	4	3	5	5	5	1	1	2	2	3	3	4	4	3	5

Tabela 4.2: Notas obtidas para os testes subjetivos realizados por 16 observadores que avaliaram o vídeo *Rush Hour* codificado a 20 diferentes taxas.

		Taxas (kbit/s)										Taxas (kbit/s)									
		303	504	705	911	1103	1504	2022	3015	4063	8139	404	602	810	1003	1303	1704	2517	3566	6056	10349
Grupo 1	Grupo 1	2	2	4	3	4	4	4	5	5	5	1	2	3	3	5	3	4	4	5	5
		1	2	3	3	3	4	4	4	5	5	2	1	4	3	5	4	5	4	5	5
		1	1	2	2	2	2	3	2	3	3	1	2	3	3	4	3	4	4	5	5
		1	1	1	2	3	3	2	3	4	3	2	2	4	4	5	4	4	4	5	5
Grupo 2	Grupo 2	1	1	3	3	4	3	4	5	4	5	3	4	4	5	5	5	5	5	5	5
		1	2	3	2	2	3	3	4	4	4	1	2	2	3	3	3	4	3	3	4
		2	2	4	4	3	4	4	4	4	5	2	3	4	4	4	4	4	4	5	5
		2	3	4	4	4	4	4	4	5	5	1	2	2	3	3	3	4	3	4	4

Tabela 4.3: Notas obtidas para os testes subjetivos realizados por 16 observadores que avaliaram o vídeo *Riverbed* codificado a 20 diferentes taxas.

		Taxas (kbit/s)										Taxas (kbit/s)									
		544	573	706	912	1109	1514	2026	3038	4042	8075	557	605	809	1011	1307	1715	2531	3542	6063	10093
Grupo 1	Grupo 1	1	1	1	2	2	2	2	3	3	5	1	1	1	1	2	1	2	2	3	3
		1	1	1	1	1	1	2	2	3	5	1	1	1	1	2	2	2	2	4	5
		1	1	1	1	1	1	2	2	2	3	1	1	1	1	2	1	3	4	5	5
		1	1	1	1	1	1	1	2	3	3	1	1	1	1	1	1	1	3	4	5
Grupo 2	Grupo 2	1	1	1	1	1	1	2	3	3	4	1	2	2	2	3	3	4	5	5	5
		1	1	1	1	1	1	2	2	2	4	1	1	1	1	1	2	2	2	3	3
		1	1	1	1	1	2	2	3	3	4	1	1	2	2	2	2	3	2	4	4
		1	1	1	2	2	2	3	4	4	5	1	1	1	2	2	2	2	2	4	4

Tabela 4.4: Notas obtidas para os testes subjetivos realizados por 16 observadores que avaliaram o vídeo *Station2* codificado a 20 diferentes taxas.

		Taxas (kbit/s)												Taxas (kbit/s)										
		307	496	718	917	1155	1590	2163	3118	4009	8687			398	605	798	1017	1334	1813	2524	4009	6386	10397	
Grupo 1	Grupo 1	2	3	4	5	5	4	4	5	5	5	Grupo 3	1	2	2	2	3	3	3	3	3	4	4	4
		1	3	3	5	5	5	4	5	5	5			2	3	3	3	4	5	4	4	4	5	5
		1	2	2	3	3	3	3	3	3	3			2	4	3	5	4	4	4	4	5	5	5
		1	2	2	3	4	3	3	4	5	5			3	4	3	5	3	4	4	4	5	5	5
Grupo 2	Grupo 2	1	2	2	3	3	4	4	5	5	5	Grupo 4	2	3	5	4	4	4	5	5	5	5	5	5
		1	2	2	2	3	3	4	4	4	5			2	2	3	3	3	2	4	5	5	5	5
		2	3	3	4	4	4	4	4	4	4			2	3	4	4	4	4	4	5	5	5	4
		1	3	3	4	5	5	5	4	4	5			1	2	2	2	3	3	3	4	5	5	5

4.1.0.1 Cálculo do MOS e Definição do Modelo

Com base nas notas obtidas para cada sequência de vídeo na primeira etapa de testes, calculou-se o *Mean Opinion Score* - MOS. O MOS é determinado através do cálculo da média dos valores (notas) atribuídos aos vídeos, de acordo com a Equação 4.1:

$$\bar{u}_{ij} = \frac{1}{N} \sum_{k=1}^N u_{ijk} \quad (4.1)$$

onde,

u_{ijk} : pontuação do observador k, para a taxa j, do vídeo i.

N : número de observadores.

A Tabela 4.5 mostra os valores encontrados para o MOS da sequência *Pedestrian Area*. As colunas Δ da Tabela 4.5 exibem a diferença entre o MOS para grupos que analisaram vídeos codificados nas mesmas taxas. Apesar da mudança na ordem de exibição das sequências de vídeo, em cada um dos grupos, este não parece ser um fator determinante para a avaliação do observador.

As tabelas 4.6, 4.7 e 4.8 mostram o valor calculado para o MOS das sequências de vídeo *Rush Hour*, *Riverbed* e *Station2*, respectivamente. Observe que em todas as tabelas foi calculada a diferença Δ entre os MOS para grupos que avaliaram vídeos com as mesmas taxas.

Obtenção do Modelo As figuras 4.1, 4.2, 4.3 e 4.4 mostram retas ajustadas através de regressão linear, fornecendo um modelo que relaciona o logaritmo da taxa de codificação da sequência com o MOS obtido através da primeira fase de testes.

Tabela 4.5: Valores do *Mean Opinion Score* calculados através da média dos valores da Tabela 4.1 para os grupos de observadores 1, 2, 3 e 4 da sequência *Pedestrian Area*.

Taxas (kbit/s)	MOS		Δ	Taxas (kbit/s)	MOS		Δ
	Grupo 1	Grupo 2			Grupo 3	Grupo 4	
302,3	1,25	1,00	-0,25	402,3	1,50	1,75	0,25
502,64	1,25	1,50	0,25	603,94	1,75	1,75	0,00
704,39	2,50	1,75	-0,75	805,13	2,00	2,50	0,50
905,74	2,25	2,50	0,25	1007,2	2,25	3,00	0,75
1108,1	2,25	2,00	-0,25	1388,6	3,00	3,25	0,25
1509,1	3,00	3,25	0,25	1710,7	3,75	3,75	0,00
2014,1	3,75	3,25	-0,50	2518,3	3,75	4,00	0,25
3019,4	4,25	4,25	0,00	3521,2	4,25	4,50	0,25
4019,5	4,25	4,25	0,00	6061,7	4,50	4,50	0,00
8096,8	4,50	4,50	0,00	10033	5,00	4,75	-0,25

Tabela 4.6: Valores do *Mean Opinion Score* calculados através da média dos valores da Tabela 4.2 para os grupos de observadores 1, 2, 3 e 4 da sequência *Rush Hour*.

Taxas (kbit/s)	MOS		Δ	Taxas (kbit/s)	MOS		Δ
	Grupo 1	Grupo 2			Grupo 3	Grupo 4	
303,1	1,25	1,50	0,25	404,1	1,50	1,75	0,25
503,7	1,50	2,00	0,50	602,3	1,75	2,75	1,00
705,0	2,50	3,50	1,00	809,7	3,50	3,00	-0,50
910,5	2,50	3,25	0,75	1003,0	3,25	3,75	0,50
1103,0	3,00	3,25	0,25	1302,9	4,75	3,75	-1,00
1503,9	3,25	3,50	0,25	1704,4	3,50	3,75	0,25
2022,1	3,25	3,75	0,50	2516,6	4,25	4,25	0,00
3015,1	3,50	4,25	0,75	3566,0	4,00	3,75	-0,25
4063,3	4,25	4,25	0,00	6055,8	5,00	4,25	-0,75
8138,7	4,00	4,75	0,75	10349	5,00	4,50	-0,50

Tabela 4.7: Valores do *Mean Opinion Score* calculados através da média dos valores da Tabela 4.3 para os grupos de observadores 1, 2, 3 e 4 da sequência *Riverbed*.

Taxas (kbit/s)	MOS		Δ	Taxas (kbit/s)	MOS		Δ
	Grupo 1	Grupo 2			Grupo 3	Grupo 4	
544,0	1,00	1,00	0,00	556,5	1,00	1,00	0,00
572,9	1,00	1,00	0,00	605,3	1,00	1,25	0,25
706,0	1,00	1,00	0,00	808,7	1,00	1,50	0,50
911,8	1,25	1,25	0,00	1011,0	1,00	1,75	0,75
1108,6	1,25	1,25	0,00	1306,7	1,75	2,00	0,25
1513,9	1,25	1,50	0,25	1715,2	1,25	2,25	1,00
2026,2	1,75	2,25	0,50	2530,9	2,00	2,75	0,75
3037,6	2,25	3,00	0,75	3541,8	2,75	2,75	0,00
4042,3	2,75	3,00	0,25	6063,3	4,00	4,00	0,00
8075,4	4,00	4,25	0,25	10093	4,50	4,00	-0,50

Tabela 4.8: Valores do *Mean Opinion Score* calculados através da média dos valores da Tabela 4.4 para os grupos de observadores 1 e 2 da sequência *Station2*.

Taxas (kbit/s)	MOS		Δ	Taxas (kbit/s)	MOS		Δ
	Grupo 1	Grupo 2			Grupo 3	Grupo 4	
307,1	1,25	1,25	0,00	398,4	2,00	1,75	-0,25
495,7	2,50	2,50	0,00	605,4	3,25	2,50	-0,75
717,8	2,75	2,50	-0,25	797,9	2,75	3,50	0,75
916,9	4,00	3,25	-0,75	1017,0	3,75	3,25	-0,50
1155,0	4,25	3,75	-0,50	1334,2	3,50	3,50	0,00
1590,0	3,75	4,00	0,25	1812,6	4,00	3,25	-0,75
2163,1	3,50	4,25	0,75	2523,7	3,75	4,00	0,25
3118,3	4,25	4,25	0,00	4008,6	4,50	4,75	0,25
4008,6	4,50	4,25	-0,25	6386,4	4,75	5,00	0,25
8687,4	4,50	4,75	0,25	10397	4,75	4,75	0,00

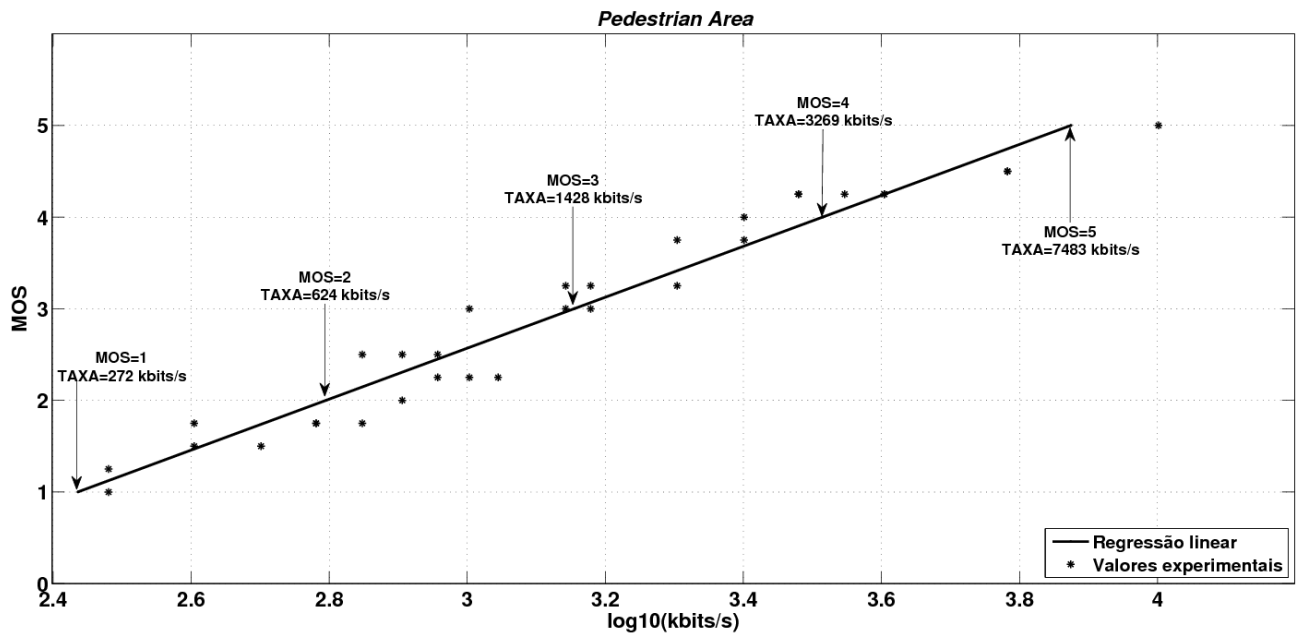


Figura 4.1: Modelo para a sequência *Pedestrian Area*: $MOS = 2,78 \cdot \log_{10}(taxa) - 5,77$

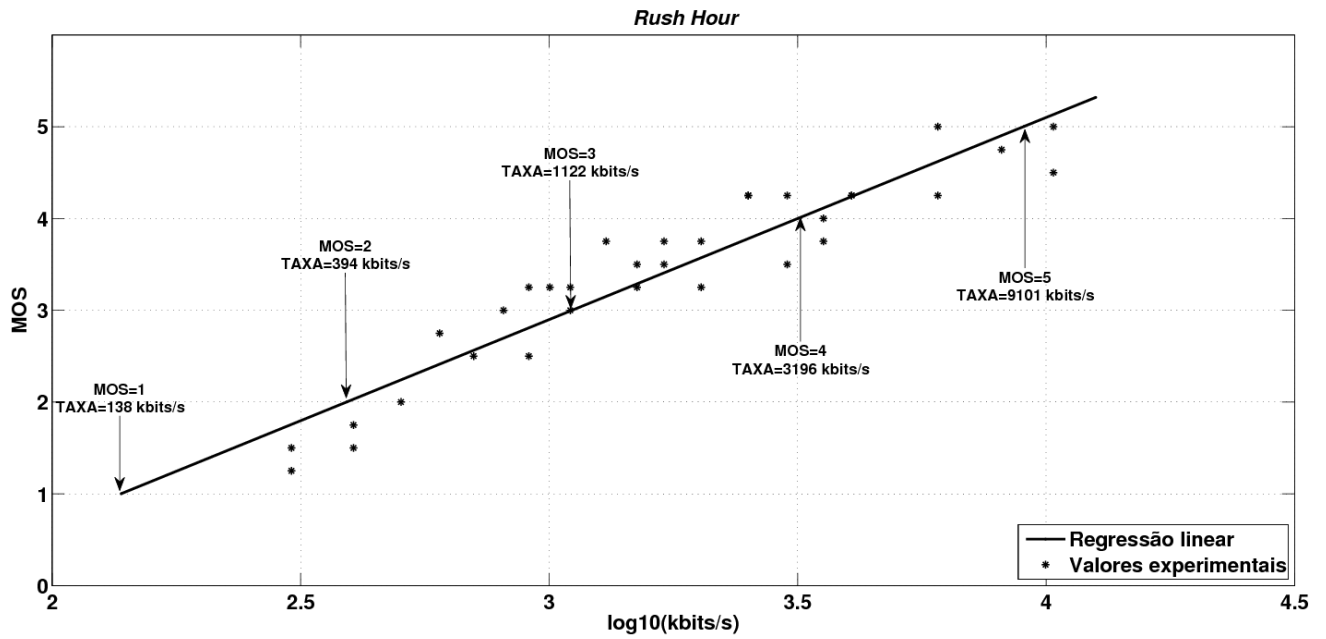


Figura 4.2: Modelo para a sequência *Rush Hour*: $MOS = 2,20 \cdot \log_{10}(taxa) - 3,71$

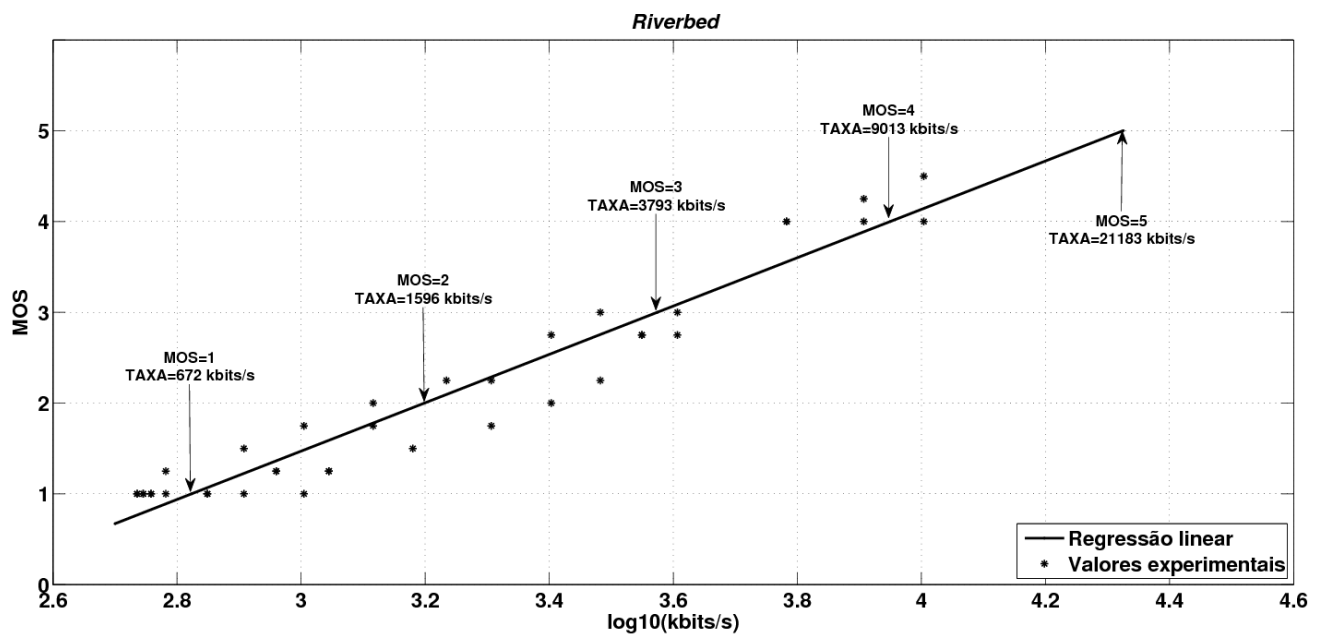


Figura 4.3: Modelo para a sequência *Riverbed*: $MOS = 2,66 \cdot \log_{10}(taxa) - 6,52$

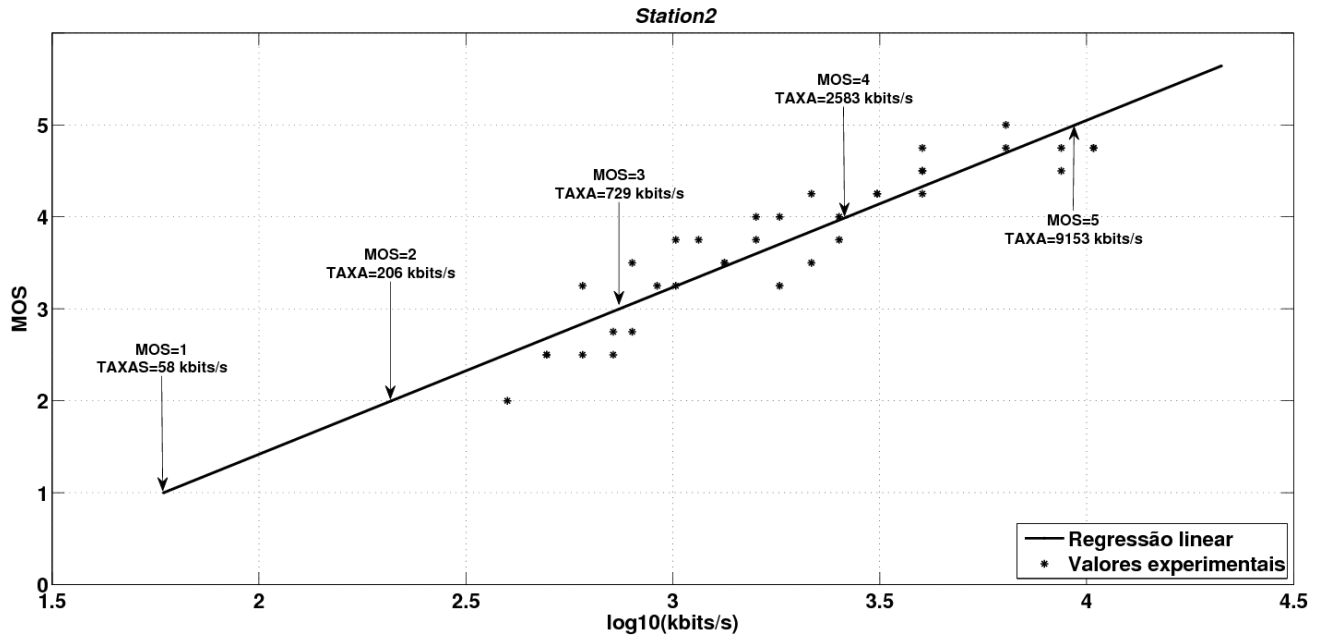


Figura 4.4: Modelo para a sequência *Station2*: $MOS = 1,82 \cdot \log_{10}(taxa) - 2,21$

Todos os gráficos apresentados permitem concluir que, dentro dos limites analisados, existe uma relação de dependência linear entre o logaritmo da taxa de transmissão e o MOS correspondente. A Tabela 4.9 mostra os coeficientes encontrados para os modelos obtidos.

Tabela 4.9: Coeficientes da equação $MOS = A \cdot \log_{10}(taxa) + B$, modelo encontrado para as sequências de vídeo avaliadas.

	A	B
<i>Pedestrian Area</i>	2,78	-5,77
<i>Rush Hour</i>	2,20	-3,71
<i>Riverbed</i>	2,66	-6,52
<i>Station2</i>	1,82	-2,21
Média (Desvio Padrão)	2,37(0,44)	-4,55(1,92)

4.1.0.2 Coeficiente de Correlação

O coeficiente de correlação ρ_{xy} entre duas variáveis aleatórias x e y é uma medida do grau de associação da relação linear entre essas duas variáveis (Bendat e Piersol, 2000) e é definido por:

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}, \text{ onde} \quad (4.2)$$

C_{xy} : covariância de x e y , onde $C_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$.

σ_x : desvio padrão de x , onde $\sigma_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$.

σ_y : desvio padrão de y , onde $\sigma_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$.

O coeficiente de correlação pode ser estimado para dois conjuntos de dados x_i e y_i que são realizações das variáveis aleatórias x e y , respectivamente, onde i vai de 1 até n através da seguinte equação:

$$\hat{\rho}_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4.3)$$

Assim como ρ_{xy} , o valor de $\hat{\rho}_{xy}$ estará sempre entre -1 e $+1$, com o valor de $\hat{\rho}_{xy}$ igual a zero correspondendo à não-associação (não há uma dependência linear), o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita mas negativa, ou seja, quando uma das variáveis aumenta, a outra diminui. Quanto mais próximo estiver de 1 ou -1 , mais forte é a dependência linear entre as duas variáveis.

Observe, através da Tabela 4.10 que, como já observamos nas figuras 4.1, 4.2, 4.3 e 4.4, os valores próximos da unidade demonstram uma forte correlação entre o MOS e o logaritmo das taxas de codificação.

Tabela 4.10: Coeficientes de correlação entre o MOS e o logaritmo das taxas nas quais os vídeos utilizados foram codificados.

Vídeo	$\hat{\rho}_{MOS, \log(taxa)}$
<i>Pedestrian area</i>	0,96
<i>Rush hour</i>	0,87
<i>Riverbed</i>	0,95
<i>Station2</i>	0,90

4.2 Segunda Fase de Testes

Em um segundo momento, foram realizados os testes subjetivos segundo a metodologia proposta, que consistiu em uma série de comparações relativas onde o observador escolheu a melhor entre duas sequências de vídeo. Obtiveram-se 20 avaliações para cada uma das quatro sequências de vídeo utilizadas na primeira fase de testes, sendo 10 relativas a uma taxa pré-definida e 10 relativas a outra. Os resultados obtidos podem ser verificados nas figuras 4.5 a 4.8.

Nas figuras 4.5 a 4.8 os valores indicados pelas setas mostram o MOS da sequência avaliada pelos observadores, na segunda fase de testes subjetivos, que foi obtido pelo modelo. Já as barras simbolizam os valores medidos, ou seja os resultados obtidos através das comparações,

sendo que sua largura nas figuras é irrelevante, consistindo apenas artifício gráfico de apresentação, pois os valores medidos variam em passos de 0,25. A Figura 4.8, por exemplo, mostra os resultados obtidos para a sequência *Station2* onde foram avaliados vídeos com $MOS = 2,7$ e $MOS = 3,3$. Observe que em relação ao vídeo com $MOS = 2,7$ foram obtidas 7 notas com $MOS = 2,75$ e 3 notas com $MOS = 2,5$. Para o vídeo com $MOS = 3,3$ foram obtidas 3 notas com $MOS = 3,25$, 2 notas com $MOS = 3,5$, 1 nota com $MOS = 3,0$ e 4 notas com $MOS = 4,0$. O resultado relativo à sequência *Station2* foi bom, possivelmente devido ao fato dessa possuir cenas com estruturas regulares que se tornam mais evidentes quando codificadas em diferentes taxas.

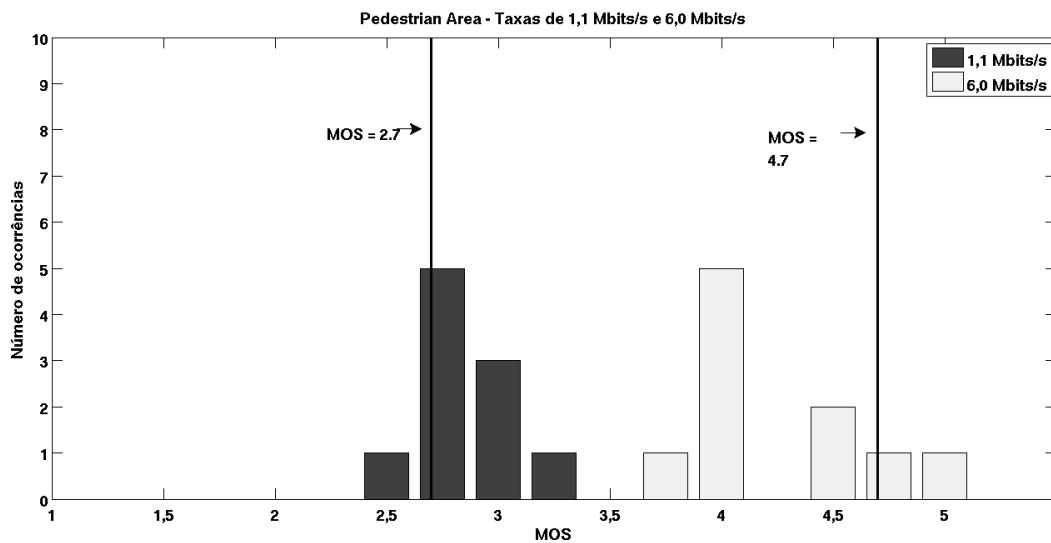


Figura 4.5: Resultado obtido na avaliação subjetiva da sequência *Pedestrian Area* utilizando-se a metodologia proposta.

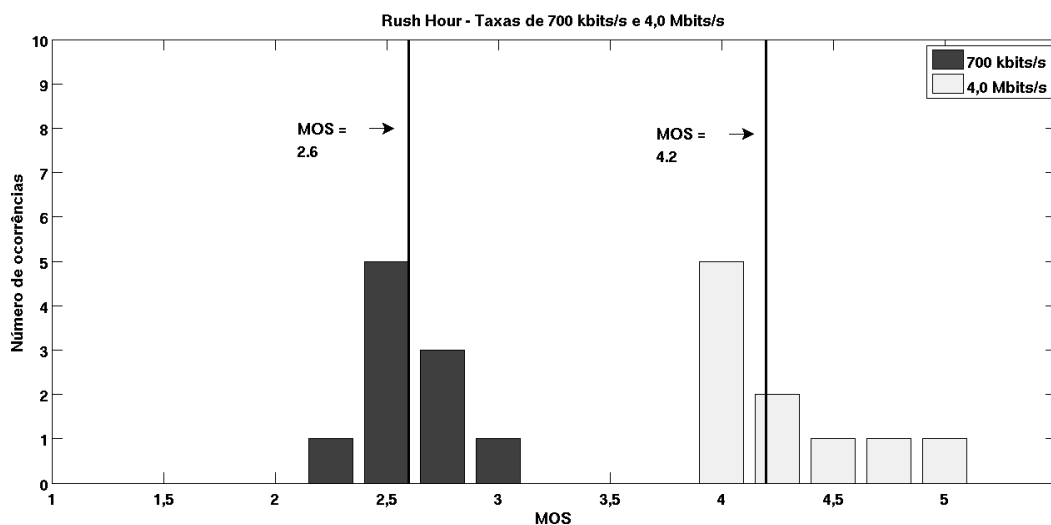


Figura 4.6: Resultado obtido na avaliação subjetiva da sequência *Rush Hour* utilizando-se a metodologia proposta.

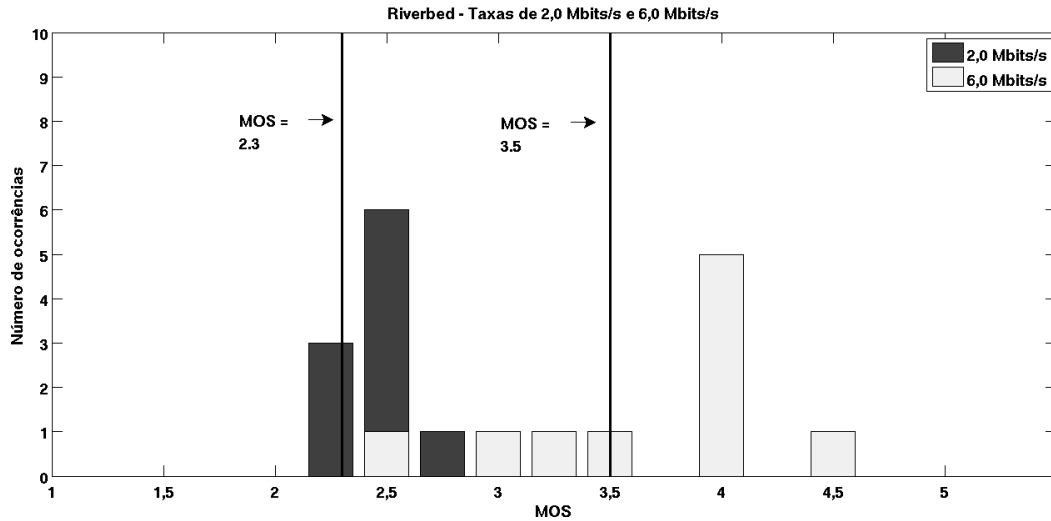


Figura 4.7: Resultado obtido na avaliação subjetiva da sequência *Riverbed* utilizando-se a metodologia proposta.

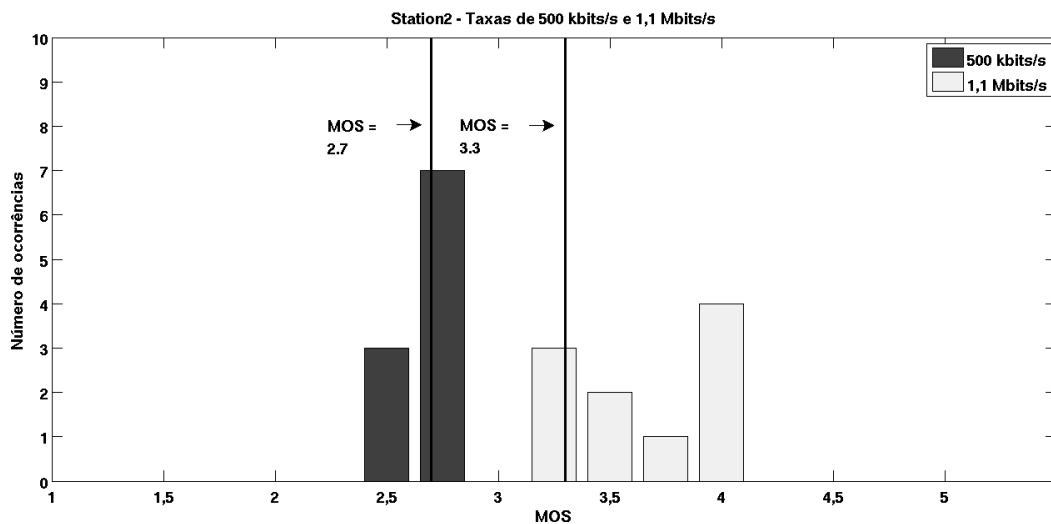


Figura 4.8: Resultado obtido na avaliação subjetiva da sequência *Station2* utilizando-se a metodologia proposta.

Para comparar os resultados obtidos na primeira e segunda fase de testes subjetivos calculou-se os desvios padrão das notas obtidas dos observadores, para cada uma das sequências e taxas utilizadas. Os valores encontrados para a primeira e segunda fase de testes são mostrados na Tabela 4.11.

O desvio padrão indica a dispersão dos dados dentro da amostra, isto é, o quanto estes em geral diferem da média. Quanto menor o desvio padrão, mais parecidos são os valores, ou seja, menor a variabilidade. Sendo assim, a Tabela 4.11 mostra que a metodologia de avaliação subjetiva proposta proporciona uma menor variabilidade dos resultados, ou seja, esta fornece medições individuais com desvio padrão da ordem de um terço daquele obtido através dos

testes MOS realizados na primeira fase.

Tabela 4.11: Desvios padrão das notas obtidas para cada uma das sequências da primeira e segunda fases de testes subjetivos.

Vídeos	Taxas (kbit/s)	Desvios padrão Primeira Fase	Desvios padrão Segunda Fase
<i>Pedestrian Area</i>	1100	0,64	0,21
	6000	0,76	0,41
<i>Rush Hour</i>	700	1,07	0,21
	4000	0,71	0,36
<i>Riverbed</i>	2000	0,53	0,16
	6000	0,76	0,60
<i>Station2</i>	500	0,53	0,12
	1100	0,93	0,34

À primeira vista, pode parecer que os resultados ilustrados pelos histogramas das figuras 4.5 a 4.8 são pouco consistentes pois, devido à representação gráfica escolhida, eles parecem distantes do desejado. Entretanto, ao visualizar o histograma de uma das sequências utilizadas com base nos dados obtidos na primeira fase de testes (medida de MOS convencional), Figura 4.9, observa-se que este último apresenta resultados espalhados ao longo de uma faixa mais extensa do que a observada nos resultados da segunda fase de testes, Figura 4.5. Este fato é sintetizado de forma quantitativa na Tabela 4.11, que compara os desvios padrão dos resultados das duas fases de testes.

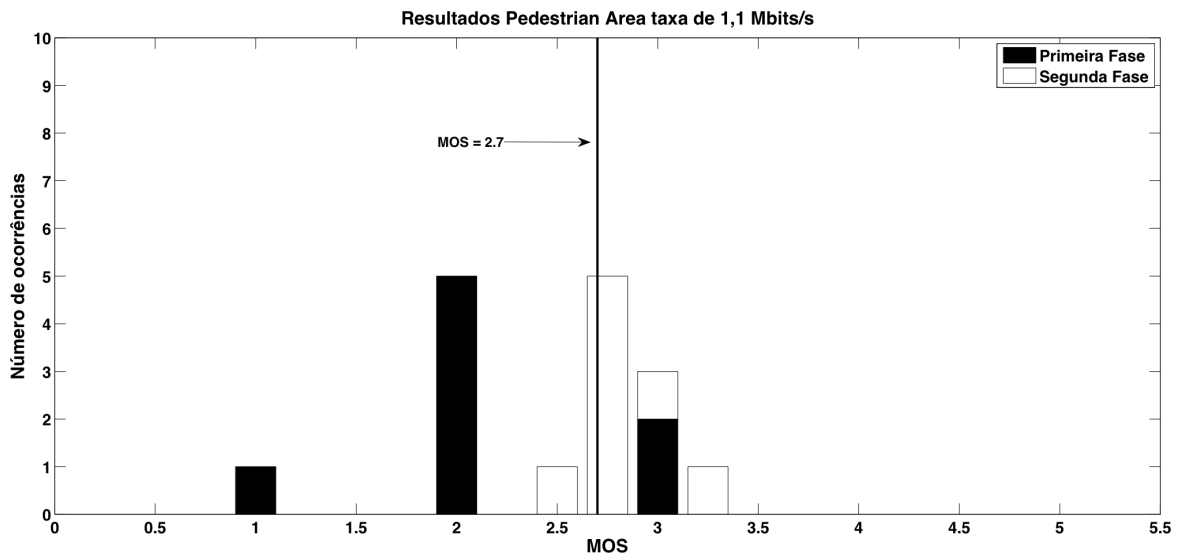


Figura 4.9: Comparação entre os resultados obtidos na primeira e segunda fases de testes para a sequência *Pedestrian Area*.

4.3 Considerações Finais

Na primeira fase de testes alguns observadores deixaram transparecer um certo cansaço nos momentos finais das avaliações (último questionário). Além disso, parte dos observadores disseram sentir dificuldades na avaliação, principalmente nas sequências *Riverbed* e *Pedestrian Area*. Neste caso as duas sequências de vídeo citadas possuem cenas bastante movimentadas, deixando o observador confuso por muitas vezes não conseguir focar em apenas alguns pontos de observação. Salienta-se que na primeira fase de testes não houve repetição de nenhuma sequência, fato esse que fez com que alguns observadores se sentissem pressionados a votar.

Na segunda fase de testes, os observadores mostraram-se mais seguros e menos ansiosos. Possivelmente estes sentimentos foram resultantes do controle que o próprio observador teve da situação durante a realização dos testes, ou seja, não havia um tempo estipulado para a avaliação e nem uma ordem de exibição definida, ficando estes parâmetros por conta do observador.

Para obtenção do modelo utilizado na segunda fase de testes foi necessário excluir alguns pontos, como forma de se alcançar um melhor ajuste: *Pedestrian Area* (20%), *Rush Hour* (17,5%), *Riverbed* (5%) e *Station2* (12,5%).

As metodologias de avaliação subjetiva descritas no Capítulo 2 deste trabalho podem ser novamente comparadas considerando-se a metodologia comparativa proposta, como pode ser observado na Tabela 4.12.

Tabela 4.12: Comparação entre as metodologias de avaliação subjetiva já existentes e citadas anteriormente, e a metodologia comparativa proposta (MCP).

Parâmetros	DSIS	DSCQS	SSCQE	SDSCE	SAMVIQ	MCP
Referência Explícita	Sim	Não	Não	Sim	Sim	Sim
Referência Oculta	Não	Sim	Não	Não	Sim	Não
Duração da Sequência	10s	10s	5min	>10s	10s	10s
Dois estímulos simultâneos	Não	Não	Não	Sim	Não	Não
Frequência de apresentação do Material de Teste	Uma vez ou Duas vezes sucessivas	Duas vezes sucessivas	Uma vez	Uma vez	Várias vezes	Várias vezes
Votação	Sequência de Teste	Sequência de Teste e Referência	Sequência de Teste	Diferença entre Sequência de Teste e Referência Mostradas Simultaneamente	Sequência de Teste e Referência	Sequência de Teste
Possibilidade de Mudança da Nota	Não	Não	Não	Não	Sim	Não
Avaliação Contínua da Qualidade	Não	Não	Sim	Sim	Não	Não

Capítulo 5

Conclusão

Neste trabalho, estabeleceu-se, inicialmente, um modelo onde ficou configurada a relação existente entre o MOS (*Mean Opinion Score*) e a taxa de codificação das sequências de vídeo utilizadas. Foram obtidas equações do tipo $MOS = A \cdot \log_{10}(taxa) + B$ das quais foram retirados os valores de taxa relativos a cada MOS específico.

A validação da metodologia para avaliação subjetiva proposta foi realizada em um segundo momento onde foram utilizados os modelos obtidos na fase inicial de testes. Esta metodologia comparativa foi avaliada e a variabilidade dos resultados comparada com aqueles obtidos através do método DSIS.

Os resultados obtidos na primeira fase deixaram bastante evidente a relação logarítmica existente entre o MOS e a taxa de codificação dos vídeos, fato este observado em gráficos e também através do cálculo de coeficientes de correlação.

Os testes realizados segundo a metodologia comparativa mostraram-se eficazes no que diz respeito ao menor tempo de execução e também a redução do cansaço evidenciado pelos observadores, os quais, em sua maioria, mostraram-se interessados e aparentemente seguros em suas avaliações.

A variabilidade dos resultados obtidos com a metodologia proposta foi menor em relação aos testes subjetivos da primeira fase (cerca de um terço), mostrando ser essa mais precisa na avaliação subjetiva da qualidade dos vídeos. Desse modo, a metodologia interativa comparativa proposta confirma a hipótese de que, para o ser humano, a tarefa de decidir se a qualidade de um conteúdo é inferior, igual ou superior à qualidade de uma referência é mais precisa do que a tarefa de atribuição de notas de 1 a 5 utilizada em testes MOS.

Trabalhos Futuros Neste trabalho, a configuração do ambiente experimental, utilizado na primeira e segunda fase de testes, não foi completamente controlada como contempla a recomendação BT-500 do ITU-R, sendo necessário, em experimentos futuros, melhorar este aspecto.

Percebeu-se uma necessidade de acrescentar funcionalidades ao software desenvolvido para realização dos testes subjetivos segundo a metodologia comparativa proposta, como por exemplo, fornecer ao observador a possibilidade de utilizar um formulário eletrônico e também melhorar a usabilidade do software através de indicadores de qualidade na tela (indicativo da variação de qualidade em relação aos botões **Vídeo 1** → péssimo a **Vídeo 9** → excelente).

Para obtenção de resultados mais precisos, na primeira fase de testes, e conseqüentemente na segunda, seria interessante a realização de testes MOS com um número maior de observadores. Além disso a utilização de um grupo de observadores treinados (pessoas submetidas a treinamento prévio para identificação de artefatos em imagens), poderia produzir resultados mais precisos devido à percepção visual diferenciada.

Referências Bibliográficas

- Alpert, T.; Baroncini, V.; Choi, D.; Contin, L.; Koenen, R.; Pereira, F. e Peterson, H. (1997). Subjective evaluation of MPEG-4 video codec proposals: Methodological approach and test procedures. *Signal Processing: Image Communication*, pp. 305–325.
- Alvarez, M.; Salam, E.; Ramirez, A. e Valero, M. (2005). A performance characterization of high definition digital video decoding using H.264/AVC. *IEEE International*, pp. 24 – 33.
- Alvarez, M.; Salam, E.; Ramirez, A. e Valero, M. (2007). HD-VideoBench - a benchmark for evaluating high definition digital video applications. *IEEE 10th International Symposium on Workload Characterization*, pp. 120–125.
- Baroncini, V. (2006). New tendencies in subjective video quality evaluation. *Institute of Electronics, Information and Communication Engineers*, E89-A(11):2933–2937.
- Bendat, J. S. e Piersol, A. G. (2000). *Random Data - Analysis and Measurement Procedures*. Wiley Series in Probability and Statistics. John Wiley & Sons, INC.
- Blin, J. L. (2002). *SEOVQ software tool for quality, preference and acceptability evaluation of multimedia images*. Development division France Telecom R&D Human interaction's division.
- Blin, J. L. (2006). New quality evaluation method suited to multimedia context - SAMVIQ. *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*.
- Bovik, A. C.; Wang, Z. e Sheikh, H. R. (2003). Objective video quality assessment. In Press, C., editor, *The Handbook of Video Databases: Design and Applications*, pp. 1041–1078.
- BT.1788, R. I.-R. (2007). *Methodology for the Subjective Assessment of Video Quality in Multimedia Applications*.
- BT.500-11, R. I.-R. (2002). *Methodology for the Subjective Assessment of the Quality of Televisions Pictures*. ITU-R.
- de Queiroz, R. L.; Ortis, R. S.; Zaghetto, A. e Fonseca, T. A. (2006). Fringe benefits of the H.264/AVC. *VI International Telecommunications Symposium (ITS2006)*.

- FFmpeg (2005). The ffmpeg libavcoded [online]. <http://ffmpeg.org/>. Acessado em: maio de 2010.
- Huynh-Thu, Q.; Brotherton, M.; Hands, D.; Brunnström, K. e Ghanbari, M. (2007). Examination of the SAMVIQ methodology for the subjective assessment of multimedia quality. *Proceedings of Third International Workshop on Video Processing for Consumer Electronics*.
- Huynh-Thu, Q. e Ghanbari, M. (2008). Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, (44):800–801.
- Koenen, R. (2002). Overview of the MPEG-4 standard. Technical report, WG11 (MPEG).
- Kozamernik, F.; Sunna, P.; Wyckens, E. e Pettersen, D. I. (2005). Subjective quality of internet video codecs - phase 2 evaluations using SAMVIQ. Technical report, EBU Technical Review.
- Laffoon, E.; Mantia, A.; Rudolf, M. e Britton, M. (2002). Kommander 1.3. <http://docs.kde.org/stable/en/kdewebdev/kommander/>.
- Munique, U. T. (2002). HD test sequences. ftp://ftp.ldv.e-technik.tu-muenchen.de/pub/test_sequences/1080p/.
- Oelbaum, T.; Baroncini, V.; Tan, T. K. e Fenimore, C. (2004). Subjective quality assessment of the emerging AVC/H.264 video coding standard. *International Broadcasting Conference (IBC)*.
- P.910, R. I.-T. (1999). *Subjective Video Quality Assessment Methods for Multimedia Applications*. ITU-T.
- Pinson, M. H. e Wolf, S. (2004). The impact of monitor resolution and type on subjective video quality testing. Technical report, NTIA Technical Memorandum.
- Richardson, I. E. G. (2003). *H.264 and MPEG-4 Video Compression- Video Coding for Next-generation Multimedia*. John Wiley & Sons Ltd.
- Sunna, P. (2005). AVC/H.264 - an advanced video coding system for SD and HD broadcasting. Technical report, EBU Technical Review.
- Tourancheau, S.; Callet, P. L. e Barba, D. (2007). Impact of the resolution on the difference of perceptual video quality between CRT and LCD. *IEEE International Conference on Image Processing*.
- Vatolin, D. (2009). MPEG-4 AVC/H.264 video codecs comparison. Technical report, CS MSU Graphics&Media Lab Video Group.

- VideoLAN (2006). x264 - a free H264/AVC encoder. <http://www.videolan.org/developers/x264.html>.
- Webster, A. A.; Jones, C. T.; Pinson, M. H.; Voran, S. D. e Wolf, S. (1993). An objective video quality assessment system based on human perception. In *SPIE Human Vision*, volume 1913, pp. 15–26. Visual Processing and Digital Display.
- Winkler, S. (2005). *Digital Video Quality - Vision Models and Metrics*. John Wiley & Sons Ltd.
- Wolf, S. e Pinson, M. H. (2007). Application of the NTIA general video quality metric (VQM) to HDTV quality monitoring. In *The Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*.