



Departamento de Ciências e Tecnologia da Informação

Ferramenta para Avaliação Subjetiva da Qualidade de Vídeo

Márcio Gonçalves

Dissertação submetida como requisito parcial para obtenção do grau de
Mestre em Engenharia de Telecomunicações e Informática

Orientador:

Prof. Doutor Tomás Gomes da Silva Serpa Brandão,
Professor Auxiliar, ISCTE-IUL

Co-Orientadora:

Prof^ª. Doutora Maria Paula Queluz dos Santos Rodrigues
Professora auxiliar, Universidade Técnica de Lisboa – Instituto Superior Técnico

[Setembro, 2012]

Resumo

A avaliação subjetiva da qualidade de vídeo é um elemento bastante importante para avaliação da Qualidade de Experiência (QoE) dos utilizadores. Para além disso, os dados provenientes de avaliações subjetivas da qualidade têm um papel fundamental no desenvolvimento de técnicas objetivas para avaliação automática da qualidade de vídeo: por um lado, o desempenho de uma métrica objetiva só pode ser validado quando os seus resultados são confrontados com os dados subjetivos; por outro lado, grande parte das métricas objetivas requerem dados subjetivos para calibração ou para efetuar processos de treino. É neste cenário que surgiu o principal objetivo desta dissertação: implementação de uma ferramenta de suporte à realização de testes de avaliação subjetiva de vídeo e utilização experimental dessa mesma ferramenta, com consequente disseminação dos resultados obtidos pela comunidade científica.

A ferramenta desenvolvida proporciona aos utilizadores uma interface gráfica simples e intuitiva de utilizar. Permite configurar e realizar testes subjetivos usando quatro metodologias diferentes: *Single Stimulus* (ACR-HR), *Double Stimulus* (DSIS), comparação (SCACJ) e um método *Single Stimulus*, não normalizado, de avaliação contínua da qualidade (AVCA). Inclui também procedimentos para extração automática dos valores de *Mean Opinion Scores* (MOS) provenientes das avaliações subjetivas e duas metodologias para deteção e rejeição de pontuações incoerentes: uma baseada na Recomendação ITU-R BT.500 e outra baseada em correlação. A apresentação de resultados pode ser feita em forma de tabelas, gráficos ou ficheiros CSV.

Palavras-chave:

Avaliação Subjetiva, Qualidade de Vídeo, Java, MOS.

Abstract

Subjective quality assessment of video is an important factor that is used for measuring the users' Quality of Experience (QoE). Data collected from subjective quality assessment tests plays an important role in objective techniques that automatically assess video quality: in one hand, the performance of an objective metric can only be validated if its results are compared with subjective data; on the other hand, most of the proposed objective metrics need subjective data for calibration or training processes. Considering this context, the main goal of this dissertation is the development of a tool for subjective video quality assessment. This tool has been developed, tested and used. All subjective data collected using this tool is expected to be disseminated among the scientific community.

The developed tool provides a simple and easy to use graphical user interface. It allows to configure and to perform subjective quality assessment tests using four different methodologies: *Single Stimulus* (ACR-HR), *Double Stimulus* (DSIS), *comparison* (SCACJ) and a non-standardized *Single Stimulus* continuous quality evaluation method (AVCA). It includes procedures for extracting *Mean Opinion Scores* (MOS) and two methods for detecting and rejecting inconsistent scores (outliers): one is based on Rec. ITU-R BT. 500 and the other is based on correlation. The results computed by the tool can be presented in the form of tables, plots or CSV files.

Keywords:

Subjective quality assessment, Video quality, Java, MOS.

Índice de Figuras

Figura 1: Artefactos devido a perdas de transmissão	2
Figura 2: Artefactos devido a efeito de bloco	2
Figura 3: <i>Ishihara plate</i>	7
Figura 4: <i>Snellen chart</i>	7
Figura 5: Exemplo da estrutura da sessão de um teste <i>Double Stimulus</i>	8
Figura 6: Escala Continua de Qualidade normalizada	10
Figura 7: Escala Continua de Qualidade normalizada com votação do observador	11
Figura 8: Ferramenta MSU <i>Perceptual Video Quality - task manager</i>	21
Figura 9: Ferramenta MSU <i>Perceptual Video Quality - player</i>	21
Figura 10: Interface principal da aplicação ACR da AccepTV	23
Figura 11: Propriedades da interface ACR da AccepTV	23
Figura 12: Minidemo utilizando tecnologia Java	27
Figura 13: Minidemo utilizando tecnologia C++	27
Figura 14: Fluxo de dados geral	29
Figura 15: Menus da interface gráfica.....	30
Figura 16: Fluxo de dados configuração de um novo teste	30
Figura 17: Configuração de um novo teste ACR-HR	32
Figura 18: Ficheiro de configuração de um novo teste ACR-HR	33
Figura 19: Fluxo de dados realização de um teste.....	34
Figura 20: Exemplo de <i>url</i> inválido de um vídeo.....	34
Figura 21: Painel de informação sobre o participante	35
Figura 22: Descrição do método DSIS.....	36
Figura 23: Teste de avaliação do método DSIS	37
Figura 24: Escala de avaliação do método DSIS	37
Figura 25: Mensagem que indica fim do teste	37
Figura 26: Ficheiro de pontuações de um participante de um teste com o método DSIS	38
Figura 27: Fluxo de dados apresentação de resultados	39
Figura 28: Tabela de resultados com a metodologia ACR-HR.....	41
Figura 29: Resultados de MOS ordenados em função do número da sequência de vídeos	42
Figura 30: Resultados de MOS ordenados em função do valor das pontuações de MOS (ordem crescente).....	42
Figura 31: Configuração de um novo teste AVCA	44
Figura 32: Ficheiro de configuração de um novo teste AVCA	44
Figura 33: Ficheiro de resultados de um participante para um teste com o método AVCA.....	45
Figura 34: Tabela de resultados com o método AVCA	46
Figura 35: Tabela de resultados utilizando a metodologia DSIS e método de rejeição da Correlação	48
Figura 36: Sequências de referência utilizadas nos testes subjetivos.....	49
Figura 37: Resultados de MOS dos testes subjetivos efetuados com método ACR-HR.....	52

Índice de Tabelas

Tabela 1: Condições de visualização recomendadas.....	6
Tabela 2: Escala de Qualidade	9
Tabela 3: Escala de Artefactos	9
Tabela 4: Escala de Comparação SCACJ	12
Tabela 5: Principais características dos métodos de avaliação subjetiva	13
Tabela 6: Excerto da Base de Dados da IRCCyN/IVC do método ACR.....	19
Tabela 7: Características dos computadores e monitores usados nos testes	51
Tabela 8: Condições de visualização dos testes IT-IUL	51

Índice

1	Introdução.....	1
1.1	Enquadramento e motivação.....	1
1.2	Objetivo e contribuição.....	3
2	Avaliação subjetiva da qualidade de vídeo.....	5
2.1	Preparação de testes subjetivos.....	6
2.2	Métodos recomendados.....	8
2.2.1	Métodos <i>Single Stimulus</i>	9
2.2.2	Métodos <i>Double Stimulus</i>	10
2.2.3	Métodos de Comparação.....	12
2.2.4	Comparação entre métodos.....	12
2.3	Cálculo de MOS e rejeição de resultados.....	14
2.4	Trabalho relacionado.....	17
2.4.1	Bases de dados de avaliação subjetiva.....	17
2.4.2	Ferramentas atualmente disponíveis.....	20
3	Implementação da ferramenta.....	25
3.1	Tecnologia de implementação.....	25
3.2	Estrutura da interface gráfica.....	28
3.2.1	Configuração de um novo teste.....	30
3.2.2	Realização de um teste.....	33
3.2.3	Apresentação de resultados.....	38
4	Contribuições adicionais.....	43
4.1	<i>Artifact Visibility Continuous Assessment (AVCA)</i>	43
4.2	<i>Correlation Analysis</i>	47
5	Realização de testes subjetivos.....	49
5.1	Sequências de vídeo utilizadas.....	49
5.2	Condições dos testes e participantes.....	51
5.3	Resultados.....	52
6	Conclusões e trabalho futuro.....	53
6.1	Síntese.....	53
6.2	Trabalho futuro.....	53
	Referências.....	55

1 Introdução

1.1 Enquadramento e motivação

Ao longo dos últimos anos, a avaliação da qualidade de vídeo é um assunto que tem vindo a assumir uma grande relevância, tanto por parte da comunidade científica, como por parte das principais operadoras que oferecem serviços de televisão por cabo, televisão digital terrestre ou televisão móvel. Esta relevância prende-se com a necessidade de obter medidas da Qualidade de Experiência (QoE) dos utilizadores. Entre os vários fatores que influenciam a qualidade de experiência, a qualidade do sinal vídeo percebida pelos consumidores, é sem sombra de dúvida, um dos fatores mais importantes a ter em conta.

No contexto da distribuição de conteúdos vídeo em formato digital, a qualidade do sinal vídeo pode ser vista como uma medida da degradação visual, percebida pelos utilizadores, depois do vídeo ser sujeito a perdas causadas pelos sistemas de codificação e/ou transmissão. O desempenho destes dois tipos de sistemas influencia diretamente a qualidade de vídeo.

Por exemplo, quando se faz a captura de vídeo destinado ao formato *Standard Digital Television* (SDTV), resulta um sinal de vídeo, tipicamente *raw*⁽¹⁾, com um débito binário que pode ir até aos 270Mbits/s. Este sinal é posteriormente codificado, pelos operadores de TV, para que sejam atingidos débitos binários com cerca de 2Mbits/s. Esta compressão do sinal vídeo é feita devido a limitações na Largura de Banda (LB) disponível para transmissão. Quanto mais comprimido for o sinal vídeo num canal com LB limitada, mais conteúdos multimédia se conseguem transmitir por esse canal. Para além disso, a utilização de menos LB por serviço, permite a introdução de serviços adicionais.

No entanto, as formas mais habituais de codificar vídeo – normas H.264 [1] e H.262 [2] (MPEG2) – contém procedimentos que introduzem perdas na informação visual. Associadas a estas perdas originadas na codificação, juntam-se possíveis perdas que ocorrem durante a transmissão (por exemplo, perdas de pacote em transmissão por redes IP). Todas estas perdas podem causar efeitos que poderão ser perceptíveis durante a visualização do vídeo. Estes efeitos são geralmente designados por artefactos. Na Figura 1 e Figura 2 estão ilustrados dois exemplos de artefactos típicos em vídeo: perdas de transmissão e efeito de bloco.

Os artefactos relacionados com a codificação de vídeo têm uma relação muito direta com a técnica de codificação que é utilizada e, tipicamente, quanto mais se comprime o sinal vídeo, mais perceptíveis se tornam os artefactos. Os artefactos originados devido a

⁽¹⁾ Formato de arquivo digital que contém a totalidade da informação visual, tal como foi capturada e sem qualquer tipo de compressão de dados.

perdas de transmissão têm também uma relação com o conteúdo do próprio vídeo: os seus efeitos têm tendência a ser mais perceptíveis em regiões do vídeo com elevado grau de atividade temporal (zonas com mais movimento).



Figura 1: Artefactos devido a perdas de transmissão



Figura 2: Artefactos devido a efeito de bloco

O desempenho de um sistema de codificação pode ser melhorado se forem conhecidos os tipos de artefactos envolvidos no vídeo e qual o seu impacto na qualidade percebida pelos utilizadores. Para tal, são necessárias formas de quantificar esse impacto na qualidade. Essa quantificação pode ser feita segundo dois tipos diferentes de metodologias para avaliação da qualidade de vídeo: avaliação subjetiva e avaliação objetiva.

A avaliação subjetiva consiste na realização de testes em que observadores humanos avaliam por si a qualidade de sequências de vídeo que lhes são exibidas durante uma sessão de teste. Estes processos de avaliação encontram-se normalizados segundo as recomendações ITU (*International Telecommunication Union*) ITU-T P.910 [3] e ITU-R BT.500 [4]. No entanto, esta forma de avaliação da qualidade é um processo que necessita de uma amostra significativa de pessoas e de muitos cuidados na sua preparação. É também um processo demorado e impossível de ser utilizado em tempo real. O facto da avaliação ser subjetiva significa que a qualidade percebida é própria de cada indivíduo, podendo variar de pessoa para pessoa. As pontuações de qualidade provenientes da avaliação subjetiva designam-se habitualmente por *Mean Opinion Scores* (MOS).

Por outro lado, uma métrica de avaliação objetiva avalia a qualidade sequências de vídeo de uma forma automática, sem a intervenção direta de humanos. Esta classe de medidas de qualidade é mais atrativa para as operadoras, uma vez que a rápida obtenção de medidas de qualidade (em tempo-real) possibilita a implementação de aplicações para uma eficiente monitorização e gestão da qualidade.

No entanto, para desenvolver e avaliar o desempenho de uma métrica objetiva de avaliação de qualidade, são necessários dados de MOS provenientes da avaliação subjetiva. Contudo, verifica-se que existe um número reduzido de bases de dados com

dados provenientes de avaliações subjetivas de vídeo, e um número ainda mais reduzido de ferramentas de suporte à realização de testes subjetivos.

É neste contexto que se enquadra o principal objetivo deste trabalho: a implementação de uma ferramenta de suporte à realização de testes de avaliação subjetiva de vídeo e utilização experimental dessa mesma ferramenta, com conseqüente disponibilização dos resultados obtidos. A maior motivação para o desenvolvimento de novas ferramentas para avaliação subjetiva da qualidade de vídeo prende-se com o facto das poucas ferramentas atualmente existentes terem diversas limitações, tais como:

- Alguns métodos importantes para avaliação subjetiva da qualidade de vídeo recomendados em [3] e [4] não se encontram implementados na mesma ferramenta.
- As ferramentas existentes têm pouco suporte para auxílio à preparação dos testes (por exemplo, configurar facilmente um conjunto de sequências a utilizar num teste subjetivo e gerar automaticamente a ordem pela qual as sequências são exibidas durante a realização desse teste).
- Inexistência de interfaces para apresentação de resultados e para deteção automática de participantes que deram pontuações incoerentes durante um teste.
- Impossibilidade de gerar gráficos para análise dos resultados provenientes dos testes.

1.2 Objetivo e contribuição

A avaliação subjetiva da qualidade de vídeo é um elemento importante para medidas de QoE. Para além disso, os dados provenientes de avaliações subjetivas da qualidade têm também um papel fundamental no desenvolvimento de técnicas objetivas para avaliação automática da qualidade de vídeo: por um lado, o desempenho de uma métrica objetiva só pode ser validado quando os seus resultados são confrontados com os dados subjetivos; por outro lado, grande parte das métricas objetivas requerem dados subjetivos para calibração ou para efetuar processos de treino.

Assim, surge a necessidade de implementar uma ferramenta de suporte à realização de testes de avaliação subjetiva da qualidade de vídeo capaz de:

- Proporcionar uma interface gráfica intuitiva e fácil de usar;
- Auxiliar os investigadores na preparação de testes de avaliação subjetiva;
- Ser capaz de exibir de uma sequência de vídeo até uma resolução 1920x1080 (HDTV);
- Ser capaz de exibir duas sequências de vídeo em simultâneo, desde que as dimensões do ecrã o permitam;

- Recolher as pontuações de qualidade, implementando as principais metodologias especificadas nas recomendações [3] e [4];
- Processar os dados recolhidos, de forma a eliminar eventuais *outliers* e detetar inconsistências nas pontuações de qualidade dadas pelos participantes;
- Gerar gráficos e tabelas, para facilitar a interpretação dos resultados obtidos;

A ferramenta desenvolvida no decurso deste trabalho foi utilizada para a realização de alguns testes subjetivos, num ambiente de colaboração com outros alunos a realizar o mestrado no Instituto de Telecomunicações. Os resultados destes testes de avaliação subjetiva serão no futuro disponibilizados à comunidade científica.

Sendo este um trabalho que consistiu essencialmente no desenvolvimento de uma aplicação, a ferramenta desenvolvida constitui por si só o principal contributo desta Dissertação. Foi também feito um esforço no sentido de ampliar as funcionalidades da ferramenta, ao ser implementada uma metodologia adicional, não normalizada, que possibilita a realização de testes de perceptibilidade de artefactos, relacionada com os trabalhos apresentados em [5] e [6]. Outra funcionalidade adicional é a possibilidade de detetar participantes cujas pontuações de qualidade são inconsistentes com as da maioria dos participantes, diferente do método mencionado em [4].

As características da ferramenta desenvolvida visam torná-la apelativa para quem desenvolve trabalho na área da avaliação de qualidade de vídeo. O desenvolvimento da ferramenta foi feito de uma forma modular, o que facilitará a introdução de novas funcionalidades em trabalhos futuros.

Esta Dissertação encontra-se organizada segundo 5 capítulos. Após a Introdução, no Capítulo 2 é feita uma breve revisão dos principais conceitos associados à avaliação subjetiva da qualidade de vídeo, descrevendo-se as principais metodologias normalizadas. Nesse capítulo são também revistas algumas ferramentas para avaliação subjetiva, assim como bases de dados existentes com resultados provenientes de avaliações subjetivas. No Capítulo 3 são descritos os aspetos relacionados com a implementação da ferramenta desenvolvida. No Capítulo 4 são apresentadas duas novas metodologias, uma para deteção e rejeição de pontuações incoerentes e outra para avaliações contínuas da qualidade do vídeo. Os testes de avaliação subjetiva da qualidade realizados com a ferramenta, assim como os seus resultados, são apresentados no Capítulo 5. A Dissertação é finalizada com o Capítulo 6, onde se apresentam as principais conclusões deste trabalho e tópicos para trabalho futuro.

2 Avaliação subjetiva da qualidade de vídeo

A realização de testes de avaliação subjetiva da qualidade de vídeo consiste geralmente na exibição de sequências de vídeo a um conjunto de observadores, sendo pedido a estes que avaliem a qualidade dos conteúdos que vão sendo exibidos. Em [3] e [4] são descritas diversas metodologias que podem ser empregues na avaliação da qualidade subjetiva de vídeo. Estas recomendações definem também um conjunto de parâmetros relacionados com as condições de visualização, sugerem as instruções que devem ser dadas aos participantes, definem o número mínimo de participantes necessários para se obterem resultados fidedignos e repetíveis, e critérios de seleção do material vídeo a utilizar nos testes.

Em [3] descrevem-se os métodos de avaliação da qualidade subjetiva de vídeo para aplicações multimédia ao passo que em [4] descrevem-se os métodos para a avaliação subjetiva da qualidade de imagens televisivas. A recomendação ITU-T P.910 é mais recente que a recomendação ITU-R BT.500, procurando dar resposta ao aparecimento de novos formatos e novas formas de visualizar vídeo em aplicações multimédia fora do cenário televisivo (e.g., videoconferência).

Seguindo a abordagem descrita em [4], podem-se definir duas classes gerais de avaliações subjetivas: *avaliação de qualidade (quality assessment)* e *avaliação de artefactos (impairment assessment)*. Ambas as classes visam avaliar a qualidade de vídeo em sistemas de distribuição onde está prevista a ocorrência de erros relacionadas com a codificação e com a transmissão dos conteúdos (e.g., perdas de pacotes devido a erros de transmissão).

Quanto à forma como é feita a visualização do material vídeo utilizado nos testes, os métodos para avaliação subjetiva da qualidade de vídeo podem ser divididos em três tipos: *Single Stimulus (SS)*, *Double Stimulus (DS)* e métodos de comparação, que serão todos abordados com maior profundidade na secção 2.2. Um teste de avaliação da qualidade deve ser precedido por uma fase de demonstração e calibração, para que cada participante se possa ambientar aos mecanismos de visualização e de pontuação do teste, e para que fique com uma ideia da gama de qualidade das sequências de vídeo que serão mostradas ao longo do teste.

De seguida será feita uma descrição sintética de vários aspetos relacionados com a avaliação subjetiva: procedimentos a ter em conta durante a preparação de um teste, métodos recomendados pela ITU, discutindo-se vantagens e desvantagens de uns em relação aos outros, descrição dos procedimentos de cálculo dos valores *Mean Opinion Score (MOS)* e *Differential Mean Opinion Score (DMOS)* e rejeição de resultados incoerentes por parte dos participantes nos testes. Ainda neste capítulo será abordado o trabalho relacionado existente sobre este tema.

2.1 Preparação de testes subjetivos

Diferentes ambientes com diferentes condições de visualização podem afetar significativamente os resultados dos testes de avaliação subjetiva. Existem fatores, muito importantes, que devem ser considerados durante a realização de testes subjetivos: a intensidade da luz e o ruído ambiente na sala onde são feitos os testes, a qualidade e calibração do ecrã. De acordo com [3] e [4] os testes subjetivos devem estar sujeitos às condições de visualização apresentadas na Tabela 1.

Tabela 1: Condições de visualização recomendadas

Parâmetro	Valor
Distância de visualização	1-8H ⁽²⁾
Pico de luminância do ecrã	100-200 cd/m ²
Rácio entre a luminância do ecrã inativo e pico de luminância	≤ 0.05
Rácio entre a luminância do fundo da imagem no ecrã e o pico de luminância da imagem	≤ 0.2
Crominância do fundo/plano	D ₆₅ ⁽³⁾
Iluminação de fundo da sala	≤ 20 lux
Máximo ângulo de observação relativamente à normal	30°

A escolha e o número de participantes necessários para efetuar testes subjetivos é um aspeto importante para se obterem resultados válidos e coerentes. De acordo com [3], pelo menos 15 participantes, *não-especialistas*, devem participar nos testes. Por outro lado, um número de participantes superior a 40 não traz valor acrescentado aos resultados. Todos os participantes deverão ter uma acuidade visual normal (ou corrigida através de lentes) e deverão ser capazes de distinguir todas as cores (i.e., não podem ser daltónicos). Para testar a capacidade visual dos participantes pode ser utilizado um quadro de *Snellen*, como o representado na Figura 4, e um teste de cores de *Ishihara*, utilizando placas idênticas à representada na Figura 3.

Os participantes podem ser classificados como *especialistas* ou *não-especialistas*. Os *especialistas* são participantes que estão familiarizados com técnicas de processamento de vídeo ou avaliação da qualidade de imagem. Estes participantes são em princípio fáceis de recrutar, pois muitas fazem parte do mesmo grupo de quem organiza os testes subjetivos. Os participantes *não-especialistas* são os que representam a generalidade da população que visualiza conteúdos multimédia (por exemplo, os telespectadores). Como estes não estão familiarizados com técnicas de processamento de imagem/vídeo, avaliam as sequências de vídeo de uma forma mais apreciativa, sem terem à partida uma tendência para procurar distorções em determinadas zonas da imagem.

⁽²⁾ H representa a altura da imagem no ecrã.

⁽³⁾ D₆₅ é uma forma de iluminação normalizada na qual os vários comprimentos de onda que compõem a luz têm uma distribuição de potência que corresponde aproximadamente à luz solar ao meio-dia.

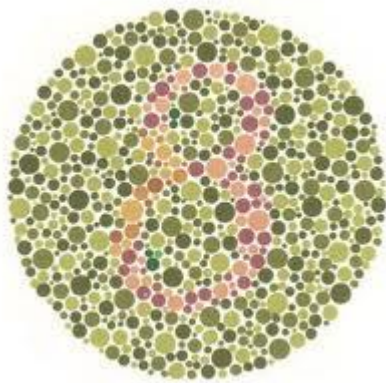


Figura 3: *Ishihara plate*

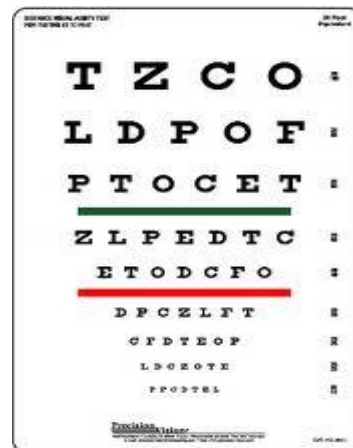


Figura 4: *Snellen chart*

Os testes subjetivos são dispendiosos, em termos de tempo, e muitas vezes não é fácil obter o número de participantes necessários. Existem alguns trabalhos em que se analisa a possibilidade de flexibilizar o número e tipo de participantes. Por exemplo, em [7] apresenta-se um estudo onde se comparam os resultados de participantes *especialistas* com os de participantes *não-especialistas*, chegando-se à conclusão que não há uma diferença significativa entre os resultados produzidos por ambos os tipos de participantes.

Antes de se iniciar cada teste deve ser feita uma sessão de treino, onde as votações não são contabilizadas, para os observadores ficarem com uma percepção mais concreta da qualidade mínima e máxima dos vídeos que vão ser apresentados. Deve-se também dar uma explicação sobre o método subjetivo que vai ser utilizado e qual é a finalidade do teste. Cada sessão de testes não deve demorar mais que do que 30 minutos [4], para evitar que os participantes fiquem fatigados ou aborrecidos.

A seleção das sequências de vídeo, assim como a ordem pela qual é feita a visualização das sequências, são também assuntos que merecem alguma atenção. De modo a encontrar um conjunto de sequências de vídeo com conteúdos representativos, um possível critério para seleção do material a usar nos testes é escolher as sequências de vídeo com base em medidas das suas atividades espacial e temporal. Por outro lado, deve-se também ter em conta o objetivo da aplicação para a qual vão ser feitos os testes.

A ordem da visualização das sequências deve ser ponderada, porque pode ter influência nos resultados dos testes. Quando a pontuação dada a uma sequência é influenciada pela posição e/ou por ter um tipo de codificação muito forte seguida de uma sequência com um tipo de codificação mesmo acentuado, podem ocorrer efeitos contextuais [4] que originam pontuações com maiores desvios face a uma ordem diferente. Por exemplo, se for apresentada uma sequência codificada a baixo débito (qualidade mais baixa) e de seguida se apresentar uma sequência codificada com algumas perdas (qualidade média), os participantes têm tendência a avaliar a qualidade desta última sequência com um valor superior ao real. Devem também ser introduzidas replicações de sequências para

testar se os resultados do participante são consistentes ao longo do teste. Caso se verifique inconsistência nos resultados estes devem ser rejeitados.

A maioria dos métodos de avaliação subjetiva requer a utilização de sequências de vídeo com cerca de 10 segundos de duração. Podem também ser utilizadas sequências de vídeo mais longas nos métodos de avaliação contínua da qualidade de vídeo. A exibição das sequências deve ser intervalada através de uma imagem a cinzento durante 2 segundos [3]. Após a visualização das sequências de vídeo o, pede-se ao observador que atribua a pontuação de qualidade (“votar”), como no exemplo ilustrado na Figura 5.

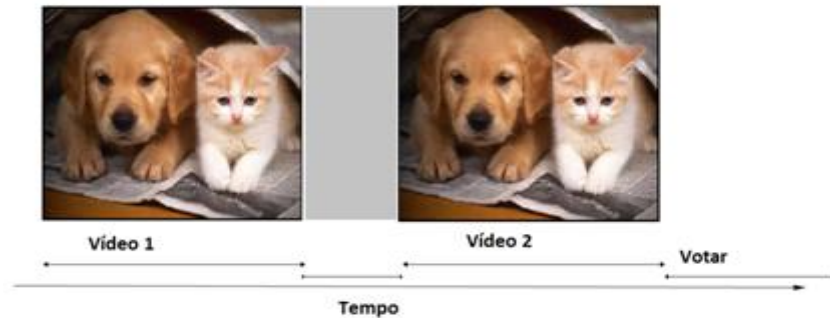


Figura 5: Exemplo da estrutura da sessão de um teste *Double Stimulus*.

2.2 Métodos recomendados

Os métodos para avaliação subjetiva da qualidade de vídeo podem ser divididos nos seguintes tipos de métodos: *Single Stimulus* (SS), *Double Stimulus* (DS) e comparação. Nos métodos SS as sequências de vídeo são apresentadas uma de cada vez, e a sua avaliação é feita individualmente. Nos métodos DS, as sequências de vídeo são apresentadas aos pares, onde uma delas é a *referência*, tipicamente o vídeo original, e a outra é a *condição de teste*, tipicamente uma versão degradada do vídeo original. Ao participante é pedido que avalie a qualidade da condição de teste tendo em conta a qualidade da referência. Existem ainda os métodos de comparação, onde as sequências de vídeo são também apresentadas aos pares mas sem existir referência (i.e., ambas as sequências são condições de teste).

As escalas de avaliação das sequências de vídeo, onde os participantes fazem uma comparação relativa da qualidade entre os vídeos apresentados, variam consoante a metodologia de cada método. Na Tabela 2 e na Tabela 3 apresentam-se dois tipos de escalas utilizadas nos testes subjetivos, escala de qualidade e escala de artefactos respetivamente. Em [3] sugerem-se também escalas de qualidade com mais níveis (9 e 11 níveis), para utilizar quando se pretende uma avaliação de qualidade mais discriminativa. Estas escalas podem posteriormente ser normalizadas para outro tipo de valores, por exemplo entre 0 e 100.

Tabela 2: Escala de Qualidade

5	Excelente
4	Boa
3	Razoável
2	Pobre
1	Má

Tabela 3: Escala de Artefactos

5	Impercetível
4	Percetível
3	Ligeiramente Incómodo
2	Incómodo
1	Muito Incómodo

2.2.1 Métodos *Single Stimulus*

Absolute Category Rating (ACR) – Neste método, as sequências de vídeo são apresentadas ao participante uma de cada vez e este faz a avaliação entre exibições consecutivas. O participante faz a avaliação de cada vídeo com base numa escala de qualidade, tipicamente a que se encontra representada na Tabela 2. Podem ser feitas repetições mostrando as mesmas condições de teste em diferentes períodos de tempo durante o teste. Este é o método mais simples e mais rápido de realizar.

Absolute Category Rating with Hidden Reference (ACR-HR) – Este método é muito semelhante ao método anterior, mas neste caso as pontuações de qualidade são geralmente expressas por valores de *Differential Mean Opinion Scores* (DMOS). Os valores de DMOS são calculados com base nas diferenças entre as pontuações dadas ao vídeo degradado e as dadas ao vídeo de referência (ver secção 2.3). O vídeo de referência é mostrado entre as restantes sequências de vídeos, sem que o observador seja informado que está a pontuar o vídeo de referência. Essa é a razão pela qual o método se designa por “Referência Escondida”. Este método, para além de ter as mesmas características do ACR, é mais robusto à influência da qualidade intrínseca ao vídeo original. As pontuações são atribuídas às sequências de vídeo de acordo com a Tabela 2.

Single Stimulus Continuous Quality Evaluation (SSCQE) – Neste método, a qualidade de cada sequência de vídeo é avaliada pelos participantes de um modo contínuo ao longo do tempo. As pontuações que são dadas à qualidade do vídeo em teste são feitas através de um dispositivo para pontuação designado por *Slider*, no qual o participante pode ir ajustando a posição de um ponteiro que indica a pontuação de qualidade em cada instante de tempo. A avaliação é feita com base numa escala contínua de qualidade, dividida em cinco zonas, as pontuações dadas são normalizadas para valores entre 0 e 100, como se exemplifica na Figura 6. Os dados recolhidos permitem obter a evolução de valores de MOS ao longo do tempo de visualização das sequências de vídeo.

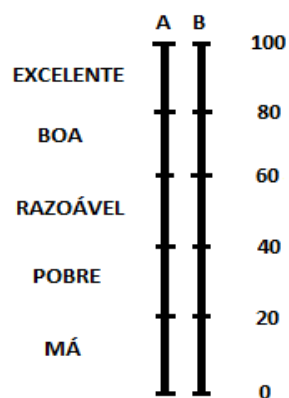


Figura 6: Escala Continua de Qualidade normalizada

2.2.2 Métodos *Double Stimulus*

Degradation Category Rating (DCR) – Método também conhecido como *Double Stimulus Impairment Scale (DSIS)*, no qual as sequências de vídeo são apresentadas aos pares: a sequência vídeo de referência (vídeo original); e a condição de teste, uma versão degradada da sequência de referência. O participante atribui uma pontuação à condição de teste que é feita tendo em conta o conteúdo da sequência vídeo de referência. A pontuação atribuída ao vídeo sob a condição de teste é dada numa escala de artefactos com cinco níveis, de 1 (“muito incómodo”) a 5 (“imperceptível”), conforme a Tabela 3. Se o participante não notar nenhum artefacto no vídeo de teste deve dar nota 5. Quer isto dizer que a diferença entre o vídeo original (referência) e o vídeo sob condição de teste é imperceptível. A referência deverá à partida ter uma qualidade igual ou superior à das condições de teste correspondentes.

Podem ser feitas repetições mostrando as mesmas condições de teste em diferentes períodos de tempo durante o teste. Este método deve ser usado nas situações em que se pretende avaliar a fidelidade do sinal de vídeo original, após sujeito a uma determinada condição de teste (por exemplo, após codificação com perdas).

No método DSIS estão incluídas duas variantes:

- Variante I – cada par referência/condição de teste é mostrado uma única vez, pedindo-se ao participante para votar após a exibição de cada par.
- Variante II – cada par referência/condição de teste é exibido pelo menos duas vezes consecutivas, pedindo-se ao participante para votar após essas exibições.

Double Stimulus Continuous Quality Scale (DSCQS) – À semelhança do método anterior, cada participante avalia a qualidade de ambas as sequências de vídeo, tanto a de referência como a de teste, mas sem que saiba qual delas é a de referência (as sequências são apresentadas com “sequência A” e “sequência B”, por exemplo). Outra diferença em relação ao método DSIS está na escala que é utilizada para recolher as pontuações de qualidade, que neste caso consiste numa escala de qualidade contínua,

conforme ilustrado na Figura 7, onde se apresenta como exemplo a atribuição de um valor de qualidade próximo de 13.

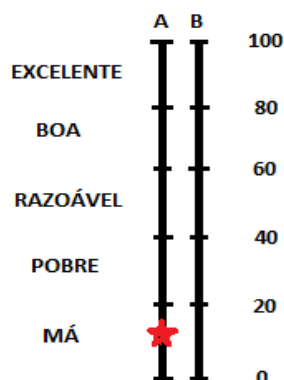


Figura 7: Escala Continua de Qualidade normalizada com votação do observador

Este método, tal como o anterior, tem também duas variantes [4].

- Variante I – é permitido ao participante alternar a visualização entre a sequência de referência e de teste, até que forme uma opinião sólida sobre cada sequência. Depois de formada a opinião é então pedido para ao participante para pontuar a qualidade.
- Variante II – destina-se a múltiplos participantes e é semelhante à variante II, descrita para o método DSIS, mas pedindo-se a cada participante para pontuar ambas as sequências.

No final dos testes são calculados os valores de MOS para ambas as sequências, referência e condição de teste.

Este método é indicado para se medir a qualidade dos sistemas relativamente a uma referência. Segundo [4], é também indicado para medir a qualidade de imagens estereoscópicas codificadas.

Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) – Este método é análogo ao método SSCQE referido na secção 2.2.1. A principal diferença consiste na visualização em simultâneo da sequência de referência e da sequência sob condição de teste, sendo o participante informado sobre qual das duas sequências é a referência.

O participante observa as sequências em simultâneo, verificando as suas diferenças, ao mesmo tempo que avalia a fidelidade do vídeo através do *Slider*, de forma contínua e usando uma escala semelhante à Figura 6. Tal como o método DSIS, o método SDSCE é útil para medir a fidelidade entre duas sequências de vídeo. Neste tipo de métodos são usadas sequências de vídeo longas (com aproximadamente 5 minutos de duração [4]).

Subjective Assessment Methodology for Video Quality Evaluation (SAMVIQ) - Método desenvolvido pela EBU (*European Broadcasting Union*) para avaliar o

desempenho de codificadores de vídeo usados em aplicações multimídia na internet. Este método foi submetido ao ITU-R em 2003 ficando em estado de “*Draft New Recommendation*”.

A principal característica deste método, face aos métodos *Double Stimulus* descritos anteriormente, está no modo como as sequências de vídeo são apresentadas ao participante. Este pode escolher a ordem pela qual as sequências são apresentadas, pode também comparar diretamente as condições de teste entre si e com a referência. A referência pode ser explícita ou escondida. Este método usa uma escala contínua de qualidade idêntica à representada na Figura 6.

2.2.3 Métodos de Comparação

Pair Comparison (PC) – Neste método cada par corresponde a duas condições de teste distintas. A avaliação é feita pedindo aos participantes que indiquem qual das condições de teste de cada par é a que possui melhor qualidade. Neste tipo de métodos não existe uma escala onde os observadores votam num valor específico. Geralmente este teste é feito utilizando todos os pares de condições de teste possíveis, e deve-se assegurar que as sequências de cada par são visualizadas em ambas as ordens possíveis.

Stimulus Comparison Adjectival Categorical Judgement (SCACJ) – De certo modo, este método é semelhante ao método *Pair Comparison*, visto que também são feitas comparações relativas de qualidade entre pares de sequências. A diferença consiste na utilização de uma escala de pontuações idêntica à representada na Tabela 4, onde o observador irá indicar se a segunda imagem tem melhor, pior ou a mesma qualidade da primeira e quão melhor ou pior é essa qualidade.

Tabela 4: Escala de Comparação SCACJ

-3	Muito Pior
-2	Pior
-1	Ligeiramente Pior
0	Igual
1	Ligeiramente Melhor
2	Melhor
3	Muito Melhor

2.2.4 Comparação entre métodos

Os métodos *Single Stimulus* têm a vantagem de conduzirem a testes de avaliação subjetiva mais rápidos de realizar. O participante acaba por ficar menos fatigado ou desinteressado, concentrando-se com mais facilidade no objetivo de avaliar a qualidade dos vídeos. Os métodos *double stimulus* podem ser mais dispendiosos em termos de tempo, caso os pares de sequências não sejam exibidos em simultâneo. De modo a evitar o cansaço e a desconcentração aos participantes, acabam por ser avaliadas menos condições de teste em comparação com as avaliadas usando métodos SS.

Nos métodos onde o participante visualiza uma referência explícita, o participante observa o vídeo original em primeiro lugar (ou em simultâneo com a condição de teste). Poderá assim ficar com uma noção mais concreta da degradação do vídeo de teste, o que lhe possibilita uma maior precisão para quantificar a qualidade. Quando não se usa uma referência explícita, as pontuações dadas pelo participante podem ser influenciadas pela ordem com que as sequências de vídeo são exibidas.

Nos métodos de avaliação contínua, SDSCE e SSCQE, onde as pontuações da qualidade das sequências de vídeo são introduzidas continuamente pelo participante através de um *slider*, podem por vezes surgir resultados incoerentes dados de forma involuntária. Estes podem surgir caso o observador não esteja acostumado à sensibilidade do dispositivo, perdendo tempo a olhar para o dispositivo para ver onde está o ponteiro e para onde o tem que mover, desviando a sua atenção da sequência de vídeo que está a ser exibida. Nestes métodos há ainda que ter em conta o tempo de resposta de cada observador, para que se possam sincronizar as pontuações dadas com os instantes de tempo correspondentes nas sequências de vídeo exibidas. Estes aspetos podem ter impacto na eficiência do método. Estes métodos requerem a utilização de sequências de vídeo longas, sendo mais adequadas para avaliar a fidelidade na reprodução e algoritmos de resiliência a erros.

Tabela 5: Principais características dos métodos de avaliação subjetiva

Parâmetro	ACR-HR	SSCQE	SDSCE	DSIS	DSCQS	SAMVIQ
Referência explícita	Não	Não	Sim	Sim	Não	Sim
Referência escondida	Sim	Não	Não	Não	Sim	Sim
Escala	Tabela 1	Tabela 1	Tabela 1	Tabela 2	Tabela 1	Tabela 1
Tamanho da sequência	10seg	5min	5min	10seg	10seg	10seg
Formato de imagem	Todos	Todos	Todos	Todos	Todos	Todos
Apresentação do teste	1 vez	1 vez	1 vez	1 vez ou 2 vezes seguidas	2 vezes seguidas	Varias vezes
Pontuação	Sequência de teste e referência	Sequência de teste	Diferença entre seq. de teste e referência	Sequência de teste	Sequência de teste e referência	Sequência de teste e referência
Avaliação contínua	Não	Sim	Sim	Não	Não	Não

De acordo com [4], o método DSCQS é aquele que minimiza a influência de efeitos contextuais, pois cada condição de teste é acompanhada em simultâneo pela respetiva referência. Por efeitos contextuais entenda-se a ordem pela qual as sequências são

exibidas durante o teste e pelo nível de degradação associado a cada uma delas. No entanto, verifica-se que os participantes podem trocar ocasionalmente as pontuações atribuídas à referência e à condição de teste.

O método SAMVIQ é o mais recente, simples de operar e rápido em relação aos métodos de avaliação subjetiva tradicionais [8]. Oferecendo uma maior liberdade ao participante para definir os parâmetros do teste. A Nos métodos onde o participante visualiza uma referência explícita, o participante observa o vídeo original em primeiro lugar (ou em simultâneo com a condição de teste). Poderá assim ficar com uma noção mais concreta da degradação do vídeo de teste, o que lhe possibilita uma maior precisão para quantificar a qualidade. Quando não se usa uma referência explícita, as pontuações dadas pelo participante podem ser influenciadas pela ordem com que as sequências de vídeo são exibidas.

Nos métodos de avaliação contínua, SDSCE e SSCQE, onde as pontuações da qualidade das sequências de vídeo são introduzidas continuamente pelo participante através de um *slider*, podem por vezes surgir resultados incoerentes dados de forma involuntária. Estes podem surgir caso o observador não esteja acostumado à sensibilidade do dispositivo, perdendo tempo a olhar para o dispositivo para ver onde está o ponteiro e para onde o tem que mover, desviando a sua atenção da sequência de vídeo que está a ser exibida. Nestes métodos há ainda que ter em conta o tempo de resposta de cada observador, para que se possam sincronizar as pontuações dadas com os instantes de tempo correspondentes nas sequências de vídeo exibidas. Estes aspetos podem ter impacto na eficiência do método. Estes métodos requerem a utilização de sequências de vídeo longas, sendo mais adequadas para avaliar a fidelidade na reprodução e algoritmos de resiliência a erros.

Tabela 5 sintetiza as principais características de alguns métodos de avaliação subjetiva da qualidade de vídeo.

2.3 Cálculo de MOS e rejeição de resultados

Uma parte muito importante no processo de avaliação do desempenho de algoritmos que implementam métricas objetivas é através de comparação dos resultados de *Mean Opinion Score* (MOS) obtidos com base nas técnicas de avaliação subjetiva. Os observadores atribuem pontuações às sequências de vídeo, com base nestas pontuações é calculada a média dos resultados para cada condição de teste (MOS). De seguida serão apresentados os cálculos para obter este parâmetro, assim como um método de rejeição de pontuações incoerentes segundo [4].

Em [4] é sugerida uma forma de detetar e rejeitar resultados incoerentes recolhidos durante a realização de testes subjetivos. O procedimento proposto começa por calcular a média das pontuações de qualidade, dadas pelos observadores, e desvio padrão das pontuações associadas a cada condição de teste. Assumindo que L é o número total de condições apresentadas e que cada uma dessas condições é avaliada por N

participantes/observadores. A média de pontuações de qualidade para cada condição de teste i , $\mu(i)$, e o seu correspondente desvio padrão, $\sigma(i)$, são definidos por:

$$\mu(i) = \frac{1}{N} \sum_{j=1}^N \varphi(i, j), \quad (2.1)$$

$$\sigma(i) = \sqrt{\sum_{j=1}^N \frac{[\mu(i) - \varphi(i, j)]^2}{N-1}}, \quad (2.2)$$

onde $\varphi(i, j)$ é o valor dado pelo participante j para a condição teste i . Em seguida, calcula-se o coeficiente de *kurtosis*, β_2 , que é a razão entre o quarto momento central e o quadrado do momento central de segunda ordem, i.e.:

$$\beta_2(i) = \frac{m_4(i)}{m_2(i)^2} \quad \text{onde } m_n(i) = \frac{1}{N} \sum_{j=1}^N [\varphi(i, j) - \mu(i)]^n, \quad (2.3)$$

O coeficiente de *kurtosis* permite determinar se a distribuição das pontuações é aproximadamente normal ou não. No caso de uma distribuição normal de pontuações, o valor teórico de β_2 seria 3. Com base no valor do coeficiente de *kurtosis*, define-se o valor de uma margem de erro, $\delta(i)$, de acordo com:

$$\delta(i) = \begin{cases} 2\sigma(i), & \text{se } 2 \leq \beta_2(i) \leq 4 \\ \sqrt{20}\sigma(i), & \text{caso contrário.} \end{cases} \quad (2.4)$$

Define-se o valor $P(j)$ como o número de vezes em que a pontuação de qualidade dada pelo participante j se encontra acima do limite superior dado por essa margem:

$$P(j) = \sum_{i=1}^L p(i, j) \quad \text{onde } p(i, j) = \begin{cases} 1, & \text{se } \varphi(i, j) \geq \mu(i) + \delta(i) \\ 0, & \text{caso contrário.} \end{cases} \quad (2.5)$$

Analogamente, $Q(j)$ é o número de vezes em que a pontuação de qualidade, dada pelo participante j , se situou abaixo do limite inferior dado pela margem $\delta(i)$, i.e.:

$$Q(j) = \sum_{i=1}^L q(i, j) \quad \text{onde } q(i, j) = \begin{cases} 1, & \text{se } \varphi(i, j) \leq \mu(i) - \delta(i) \\ 0, & \text{caso contrário.} \end{cases} \quad (2.6)$$

Depois de calcular os valores de $P(j)$ e $Q(j)$, as pontuações dadas pelo participante j são rejeitadas caso se verifiquem as seguintes condições:

$$\frac{P(j)+Q(j)}{L} > 0.05 \quad \text{e} \quad \left| \frac{P(j)-Q(j)}{P(j)+Q(j)} \right| < 0.3. \quad (2.7)$$

A primeira condição está relacionada com o número de vezes em que um participante atribui pontuações que caem fora do intervalo de confiança. Caso a taxa correspondente seja superior a 5% das vezes, o participante poderá vir a ser rejeitado. A segunda condição tende para zero se os resultados estiverem dispersos em ambas as extremidades da distribuição dos resultados de qualidade. Neste caso as pontuações do observador podem ser rejeitadas. Se os resultados estiverem concentrados apenas numa das extremidades a condição tende para valores próximos de um.

Finalmente, depois de rejeitar os resultados dos participantes com valores de pontuações incoerentes, o valor de MOS para a condição de teste i é dado por:

$$MOS(i) = \frac{1}{N_v} \sum_{j=1}^{N_v} \varphi(i, j), \quad (2.8)$$

onde N_v é o número de participantes considerados válidos. Uma vez calculados os valores de MOS, é habitual definir-se o intervalo de confiança a 95% associado ao cálculo desses valores. Esse intervalo é dado por:

$$\left[MOS(i) - 1.96 \frac{\sigma_v(i)}{\sqrt{N_v}}, MOS(i) + 1.96 \frac{\sigma_v(i)}{\sqrt{N_v}} \right], \quad (2.9)$$

onde $\sigma_v(i)$ representa o desvio padrão das pontuações dadas à condição de teste i pelos observadores considerados válidos.

Caso se pretenda calcular o *Differential Mean Opinion Score* (DMOS), este é expresso através da seguinte diferença:

$$DMOS(i) = MOS_r(i) - MOS_t(i), \quad (2.10)$$

onde $MOS_r(i)$ é o valor de MOS resultante para a sequência de referência correspondente à condição de teste i , e $MOS_t(i)$ é o valor de MOS relativo à condição de teste i .

Foram encontrados alguns aspetos menos positivos relativamente ao modo como são detetados participantes com pontuações incoerentes nos testes de avaliação subjetiva, recomendado em [4].

Em primeiro lugar, os cálculos para rejeição de participantes, descritos anteriormente, são baseados nas médias das pontuações de todos os participantes. Isto significa que com este método as pontuações incoerentes dos participantes também entram para a média total das pontuações.

Outra limitação prende-se com a forma como estão definidas as equações para $p(i, j)$ e $q(i, j)$. Caso todos os participantes atribuam a mesma pontuação a uma determinada condição de teste, essa pontuação será igual à pontuação média (i.e., $\varphi(i, j) = \mu(i), \forall j$), e o desvio padrão correspondente será nulo ($\sigma(i) = 0$). Consequentemente, o resultado do coeficiente de *kurtosis*, $\beta_2(i)$, associado a essas pontuações será um valor indeterminado (o seu cálculo origina uma indeterminação “0/0”). Nessa situação, mesmo que se considere um valor constante para o coeficiente *kurtosis*, os valores de $p(i, j)$ e $q(i, j)$ terão sempre o valor 1 para a sequência de vídeo em causa porque $\varphi(i, j) = \mu(i)$. Fazendo com que os valores de $P(j)$ e $Q(j)$ sejam incrementados para todos os participantes. Caso essa situação ocorra para várias condições de teste, podem ser verificadas ambas as condições de rejeição dadas em (2.7), sendo rejeitados os participantes.

Finalmente, este método de detecção de pontuações incoerentes não rejeita participantes que atribua sempre pontuações muito acima ou muito abaixo da média de todos os participantes (por exemplo, um participante que desse a nota máxima a todas as condições de teste). Isto verifica-se porque o método apenas deteta participantes que, ao longo do teste, atribuem pontuações que umas vezes se situam acima e outras vezes se situam abaixo do limite do intervalo de confiança estabelecido pela margem $\delta(i)$. Caso essa situação não se verifique, a segunda condição em (2.7) não se verifica, e portanto o participante não é rejeitado.

Devido a estas limitações, foi implementado um método adicional de detecção de participantes com pontuações incoerentes, que visa complementar o método da ITU-T descrito nesta secção. Esse método é baseado em correlação linear [9] e será apresentado na secção 4.2.

2.4 Trabalho relacionado

2.4.1 Bases de dados de avaliação subjetiva

Nesta secção serão apresentadas algumas bases de dados com resultados de avaliações subjetivas da qualidade de vídeo, sendo feita referência aos métodos utilizados em cada caso, o número de participantes e o tipo de condições de teste que foram avaliadas nos testes. Será feita uma breve descrição das ferramentas utilizadas para recolher os resultados das avaliações subjetivas. Estas são geralmente interfaces gráficas simples, desenvolvidas pelas equipas de investigação responsáveis pela construção de cada base de dados.

Neste tipo de bases de dados apresentam-se normalmente os resultados de MOS (*Mean Opinion Score*) ou DMOS (*Differential Mean Opinion Score*) extraídos dos testes. Estes resultados desempenham um papel extremamente importante para a comunidade científica ligada à qualidade do vídeo, pois permitem avaliar o desempenho de métricas objetivas para avaliação de qualidade, comparando os resultados obtidos pelos algoritmos com os dados de MOS (ou DMOS) armazenados nas bases de dados.

EPFL-PoliMI Subjective Video Quality Assessment Database - A base de dados EPFL-PoliMI [10][11] foi desenvolvida em conjunto por duas instituições académicas, a Escola Politécnica Federal de Lausanne (EPFL), na Suíça, e o instituto Politécnico de Milão (PoliMI), em Itália. Esta base de dados disponibiliza o resultado da avaliação de 156 condições de teste, que correspondem a sequências de vídeo codificadas em H.264/AVC, corrompidas por perdas de pacotes devido a erros de transmissão na rede. Destas, metade têm uma resolução CIF (352x288) e as restantes uma resolução 4CIF (704x576). A avaliação da qualidade de cada condição de teste foi efetuada por 40 participantes que não eram especialistas em qualidade de vídeo.

Para recolher os valores de MOS provenientes das avaliações subjetivas foi utilizada uma interface gráfica dedicada, desenvolvida em Matlab, sendo invocado um programa

auxiliar para visualizar as sequências de vídeo. Nestes testes foi utilizado o método *Single Stimulus ACR-HR*, descrito na secção 2.2.1, em que as sequências de vídeo eram apresentadas uma de cada vez, não se informando o participante sobre quais eram as sequências de referência.

De referir que as condições de visualização não foram iguais nos dois laboratórios [10], por exemplo os ecrãs eram ambos LCD, mas de marcas diferentes e com dimensões diferentes. Nas avaliações efetuadas no PoliMI, foi usado um LCD da marca Samsung com 19” e na EPFL foi usado um LCD da marca Eizo com 22”. Esta situação encontra-se em desacordo com as recomendações ITU, podendo provocar incoerências na avaliação subjetiva da qualidade de vídeo.

IT-IST MOS Database - A base de dados IT-IST foi construída com base em testes de avaliação subjetiva da qualidade de vídeo realizados no Instituto Superior Técnico (IST), em Lisboa. Esta base de dados inclui 12 sequências de vídeo original em formato CIF. Estas sequências foram codificadas, resultando em 57 condições de teste relativas a codificação H.264/AVC e 32 condições de teste relativas a codificação MPEG-2, com diferentes débitos binários, o que resulta num maior ou menor grau de degradações introduzidas pelo processo de codificação.

Nos testes com sequências codificadas em H.264 estiveram envolvidos 42 participantes e nas sequências codificadas em MPEG-2 participaram 23 pessoas. Todos os participantes eram não-especialistas em qualidade de vídeo, sendo a maioria estudantes de ambos os géneros.

Para obtenção dos resultados presentes nesta base de dados foram utilizadas as ferramentas desenvolvidas na Universidade de Moscovo, que serão apresentadas na secção 2.4.2. O cálculo dos valores de MOS e a rejeição das avaliações inconsistentes foram feitos com auxílio de um script Matlab. A metodologia adotada para os testes subjetivos foi o DCR, descrito anteriormente na secção 2.2.2. Mais detalhes sobre a forma como foi construída esta base de dados encontram-se publicados em [12] e [13].

LIVE Video Quality Database - A base de dados com avaliações subjetivas do Laboratório de Engenharia de Imagem e Vídeo (LIVE), da universidade do Texas em Austin, nos EUA, é provavelmente uma das mais conhecidas pela comunidade científica. Esta base de dados disponibiliza resultados para 150 condições de teste, geradas a partir de 10 sequências vídeo de referência (15 condições de teste por cada referência). As condições de teste foram geradas usando diferentes formas de codificação com perdas, nomeadamente MPEG-2 e H.264, e, para este último caso foram ainda obtidas condições de teste simulando perdas de pacotes devido a erros de transmissão em redes IP e em redes sem fios. Cada vídeo de teste foi avaliado por 38 participantes, estudantes da universidade.

A interface gráfica utilizada para os participantes avaliarem a qualidade do vídeo, foi desenvolvida na própria universidade do Texas, utilizando a XGL *toolbox* para Matlab.

As avaliações subjetivas foram feitas segundo o método ACR-HR, descrito na seção 2.2.1, usando uma escala contínua para recolha das pontuações. Também foi avaliado o desempenho de vários algoritmos de avaliação objetiva de qualidade de vídeo em que ambas as sequências de vídeo, referência e condição de teste, eram disponibilizadas ao observador. A forma como os resultados de DMOS foram obtidos é descrita em [14] e [15].

IRCCyN/IVC Subjective quality assessment database - O grupo de Comunicação de Imagem e Vídeo (IVC) do Instituto de Pesquisa em Comunicações e Cibernética de Nantes (IRCCyN), em França, disponibiliza um número significativo de bases de dados com avaliações subjetivas, tanto para imagens fixas como para vídeo. Irá ser abordada uma destas, a *IRCCyN IVC 1080i Database*, que contém 192 sequências de vídeo em alta definição (HD) e utiliza duas metodologias de avaliação subjetiva [16]: ACR e SAMVIQ, descritos na seção 2.2. Destas 192 sequências de vídeo, 24 são as de referência e as restantes 168 são condições de teste que correspondem a codificação H.264 com diferentes parametrizações em que os débitos de codificação foram escolhidos de modo a abranger toda a escala de qualidade [17].

Para obtenção dos resultados presentes nesta base de dados contribuíram 28 participantes não-especialistas em vídeo. A interface gráfica utilizada para o método ACR foi desenvolvida no laboratório da IVC.

Tabela 6: Excerto da Base de Dados da IRCCyN/IVC do método ACR

<i>Video name</i>	<i>SUBJECTS' RESULTS</i>									<i>MOS</i>	<i>Rejected</i>
	1	2	3	4	5	...	26	27	28		
credits.yuv	5	4	5	3	4		4	5	5	4,4	4
credits-4M.yuv	2	2	4	3	3		3	3	2	2,5	3
credits-6M.yuv	3	3	4	4	3		4	3	5	3,5	4
credits-7M.yuv	3	5	4	5	4		3	4	4	3,9	4
credits-8M.yuv	3	5	4	4	4		4	3	5	3,9	5
credits-9M.yuv	4	4	5	4	5		3	4	4	4	5
credits-10M.yuv	5	5	5	4	5		4	3	5	4,5	4
credits-14M.yuv	5	4	5	4	4		4	5	5	4,4	5

Na Tabela 6 pode-se observar um excerto da base de dados com avaliações subjetivas da qualidade do vídeo para o método ACR. Na primeira coluna do lado esquerdo é indicado o nome do vídeo de referência, “credits.yuv”, e os nomes dos vídeos correspondentes às condições de teste. Nas colunas seguintes, apresentam-se as pontuações de qualidade que os participantes (1 a 28) atribuíram às sequências avaliadas. Assim como a média das pontuações de qualidade para cada sequência avaliada, coluna a verde. Do lado direito encontra-se o único participante rejeitado (participante número 20) e as respetivas pontuações atribuídas que não são consideradas para o cálculo de MOS.

2.4.2 Ferramentas atualmente disponíveis

Nesta secção do documento serão apresentadas características de duas ferramentas existentes que são utilizadas para fazer avaliações subjetivas da qualidade de vídeo. Estas ferramentas são: MSU (*Moscow State University*) *Perceptual Video Quality* [18] e *ACR Test Interface* [19]. A última apenas implementa o método ACR (*Absolute Category Rating*), descrito na secção 2.2.1, sendo por isso limitada neste aspeto. A ferramenta da MSU é mais completa, implementando cinco métodos de avaliação subjetiva da qualidade de vídeo. Estas ferramentas têm o objetivo de permitir elaborar testes de avaliação subjetiva e fazer a coleta dos resultados de MOS e/ou DMOS. De seguida descrevem-se sucintamente estas ferramentas.

MSU Perceptual Video Quality

Esta é provavelmente a ferramenta mais completa para este tipo de avaliações de qualidade. Foi desenvolvida na universidade de Moscovo pelo grupo de vídeo *MSU Graphics & Media Lab*. Atualmente está disponível a versão 1.0 (2006), que consiste em dois módulos distintos. O primeiro módulo (*task manager*) permite realizar as configurações dos testes e o segundo módulo (*player*) permite visualizar os vídeos e atribuir pontuações aos mesmos.

- ***MSU Perceptual Video Quality task manager*** – cria ficheiros de tarefas. Adicionam-se os ficheiros de vídeo a utilizar nos testes, configura-se o conjunto de testes a efetuar e grava-se o ficheiro.
- ***MSU Perceptual Video Quality player*** – mostra as sequências de vídeo e guarda os resultados das avaliações efetuadas pelos observadores, com base na configuração definida num ficheiro de tarefas, gerado pelo módulo anterior.

Esta ferramenta implementa os métodos DSIS, DSCQS (variantes I e II), SCACJ e SAMVIQ, descritos na secção 2.2, e um método alternativo às recomendações ITU desenvolvido pelo *MSU Graphics & Media Lab* – o método *MSU Continuous Quality Evaluation* (MSUCQE). Este método consiste na apresentação em simultâneo de duas sequências vídeo. Se o participante achar que uma das sequências está com pior qualidade que a outra, assinala-o através das teclas direcionais no teclado (prime a seta para a direita se observar que o vídeo da direita tem pior qualidade que o da esquerda e vice-versa). Até certo ponto, este método é semelhante ao *Pair Comparison*, descrito na secção 2.2.3, com a diferença de a avaliação ser feita de forma contínua.

O aspeto gráfico de ambos os módulos desta ferramenta pode ser observado na Figura 8 e Figura 9. Na interface do *task manager* começa-se por escolher o nome da tarefa, seleciona-se o número de testes pretendidos e adicionam-se os ficheiros de vídeo. No segundo passo seleciona-se o método de avaliação subjetiva da qualidade que será usado para realizar os testes. Pode-se também configurar algumas opções adicionais

(passos 3 e 4), como por exemplo opções de pausa e de recuo. Por fim grava-se o ficheiro com as configurações dos testes, que fica escrito num formato próprio da ferramenta (“.tsk”) – o *ficheiro de tarefas*.

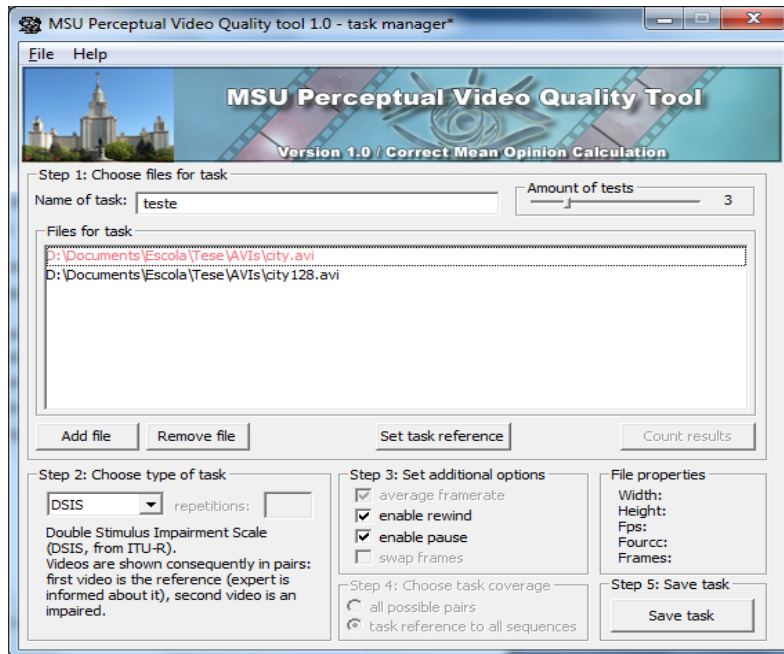


Figura 8: Ferramenta *MSU Perceptual Video Quality - task manager*

Na interface do *player* é pedido para introduzir os dados do participante, que é usado como nome do ficheiro com os resultados (em formato “csv”). Depois carrega-se a tarefa anteriormente gravada. Após a visualização dos vídeos o participante pontua a qualidade. Acabado a sequência de testes os resultados são guardados num ficheiro em formato “.csv”.

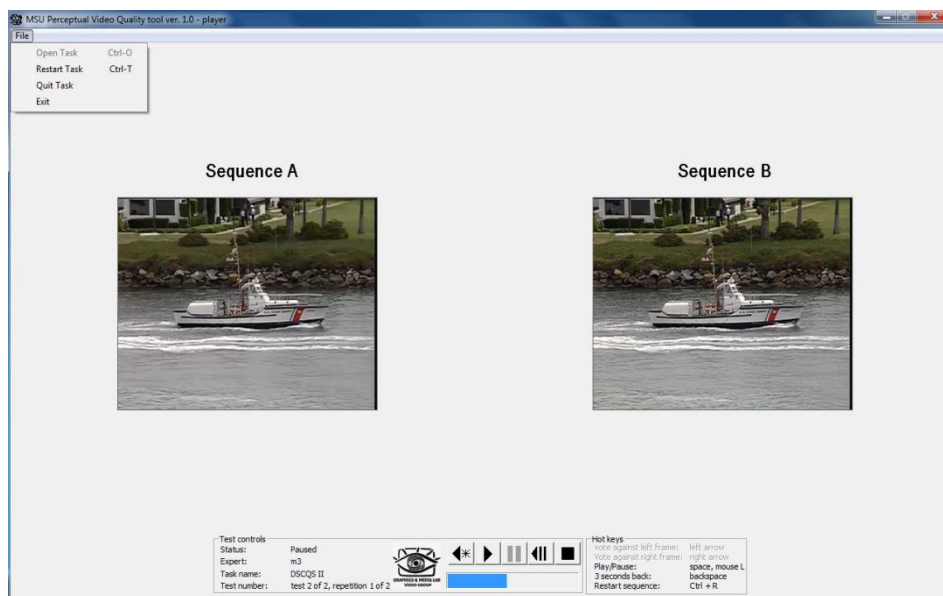


Figura 9: Ferramenta *MSU Perceptual Video Quality - player*

Esta ferramenta tem as seguintes características:

- Implementação de vários métodos de avaliação subjetiva da qualidade.
- Reprodução de ficheiros de vídeo “.avi” ou “.avs”, com resoluções até HD. Opções de pausa, atraso (3seg) e repetição durante a visualização dos testes.
- Três modos de processar os resultados e calcular os valores de MOS:
 - média simples – é feito simplesmente a média dos resultados dos observadores, sem nenhum cálculo adicional.
 - média da Rec. ITU-R BT.500 – os resultados são calculados segundo [4], onde pontuações dos observadores podem ser rejeitadas (ver secção 2.3).
 - média da MSU – cada observador obtêm uma “marca” de 0 a 1, consoante a consistência das suas pontuações. Os resultados dos participantes, que obtenham uma “marca” perto de zero, devem ter menos peso no cálculo do MOS.

Apesar de ser a ferramenta mais completa e de ser de acesso gratuito, a ferramenta *MSU Perceptual Video Quality* apresenta algumas limitações, nomeadamente:

- O facto de ter duas interfaces diferentes, uma para configurar o conjunto de testes e outra para realizar a exibição das sequências de vídeos e recolher as avaliações.
- A interface de configuração dos testes poderia ser aperfeiçoada de forma a tornar mais simples a preparação dos testes. Da forma que está feita obriga, na prática, a definir uma configuração por tarefa, em que uma tarefa corresponde à avaliação de uma única uma condição de teste.
- O cálculo dos valores de MOS não é feito de uma forma global para um teste completo (tem que ser feito tarefa a tarefa).
- Não implementa métodos *Single Stimulus*.

ACR Test Interface

Desenvolvida pela *Acceptv*, esta interface permite fazer avaliações subjetivas da qualidade de vídeo usando o método ACR. As ferramentas da *Acceptv* são bastantes conceituadas e utilizadas nos operadores de televisão em Portugal, especialmente na avaliação objetiva da qualidade de vídeo.

Ao contrário da ferramenta *MSU task manager*, a configuração do material vídeo a usar no teste não é feita através de uma interface gráfica. Assim, os nomes das sequências vídeo a usar no teste subjetivo são definidos num ficheiro de texto. Nesse ficheiro, as

sequências de vídeos são agrupadas, com o vídeo de referência a aparecer em primeiro lugar do despectivo grupo.

A partir da interface gráfica do programa deve-se escolher o ficheiro de texto que contém a definição do material de teste. Na Figura 10 está ilustrada a interface principal da aplicação.

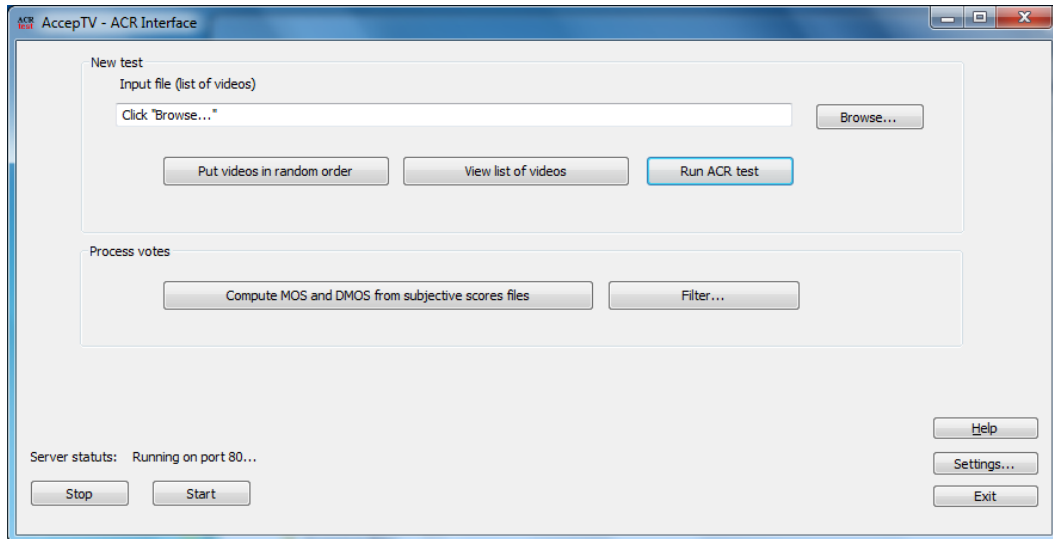


Figura 10: Interface principal da aplicação ACR da AccepTV

Antes de se iniciarem os testes, é possível visualizar a lista das sequências vídeo a exibir e alterar a ordem da sua apresentação. Na janela de configurações, representada na Figura 11, escolhe-se o diretório onde se encontram as sequências vídeo a usar nos testes, o diretório onde se pretendem guardar os resultados obtidos, configura-se o número de observadores que vão pontuar cada sequência exibida, e podem-se indicar dispositivos externos para utilizar durante a visualização do vídeo.

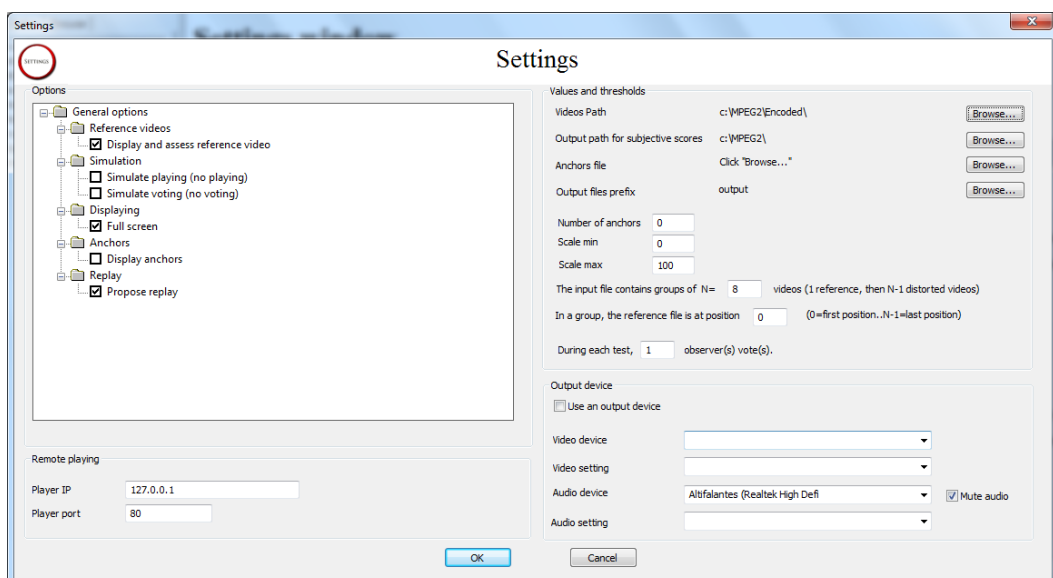


Figura 11: Propriedades da interface ACR da AccepTV

Com a configuração concluída, podem-se iniciar os testes de avaliação da qualidade. No final de cada sequência avalia-se a qualidade da sequência vídeo que foi visualizada. Terminado o teste, os resultados são guardados no diretório previamente configurado. Os ficheiros com os resultados são identificados com o nome do participante.

Esta ferramenta tem a seguinte lista de características implementadas:

- Implementa o método de avaliação subjetiva ACR.
- Suporta ficheiros de vídeo em formato MPEG-2 e H.264 com resoluções de vídeo até HDTV, podendo ser exibidas em ecrã inteiro.
- Preparação dos observadores para a qualidade dos vídeos que vai visualizar.
- Rejeição de resultados incoerentes dos observadores.
- Utiliza ficheiros de texto para configuração e para guardar os resultados.
- Possibilidade de se usarem dispositivos externos para visualizar o vídeo.
- Possibilidade de ter dois modos de funcionamento: simples, onde o mesmo computador (PC) mostra os vídeos e guarda os resultados; e remoto, onde um PC mostra os vídeos e outro guarda os resultados.

À semelhança do que acontecia com as ferramentas da MSU, a aplicação da *AcceptTV* possui também um fraco suporte para configuração do material a usar nos testes e definição da ordem com que vai ser apresentado. Para além disso, esta ferramenta tem ainda as seguintes limitações:

- Apenas implementa o método ACR, o que acaba por ser limitativo em termos de flexibilidade de realização de testes.
- A rejeição de pontuações incoerentes não é feita de forma automática. Escolhe-se um valor para o Máximo Desvio Padrão (MDP) e se a diferença entre a pontuação do observador e a média das pontuações para uma determinada condição de teste for superior ao valor de MDP. Então a pontuação deste participante deve ser rejeitada.

Ao longo deste capítulo foram abordados diversos aspetos importantes para a avaliação subjetiva de vídeo, tendo sido feita também uma descrição sucinta de algumas ferramentas existentes para esse efeito. O próximo capítulo descreve os aspetos gerais relacionados com a implementação da nova ferramenta proposta neste trabalho.

3 Implementação da ferramenta

Um dos principais objetivos que motivaram o desenvolvimento desta nova ferramenta é o de proporcionar aos investigadores uma forma simples de configurar um teste de avaliação subjetiva da qualidade de vídeo. Pretende-se também que a nova ferramenta contenha uma interface gráfica intuitiva, inclua mecanismos automáticos para rejeição de participantes com pontuações incoerentes e implemente métodos *Single Stimulus*, *Double Stimulus* e de comparação. Algumas destas características visam colmatar as limitações das ferramentas existentes que foram identificadas no capítulo anterior.

Neste capítulo são apresentados diferentes aspetos relacionados com a implementação da ferramenta desenvolvida, nomeadamente o tipo de tecnologia utilizada, a estrutura da interface gráfica, e as potencialidades da ferramenta no que diz respeito à configuração, realização e apresentação dos resultados de um teste de avaliação subjetiva.

3.1 Tecnologia de implementação

Antes de se iniciar o desenvolvimento da ferramenta propriamente dita, foi feita uma análise sobre qual a tecnologia em que esta deveria ser programada/implementada. Para tal, foram implementados três protótipos, correspondendo cada um deles a uma biblioteca que permite a visualização de conteúdos multimédia, em particular conteúdos vídeo. Das bibliotecas utilizadas, duas estão associadas à linguagem de programação Java e outra à linguagem C++:

- *Java*
 - *VideoLan Client Java (VLCJ)*
 - *Java Media Framework (JMF)*
- *C++*
 - *Open Source Computer Vision (OpenCV)*

O projeto VLCJ⁽⁴⁾ permite estabelecer uma ligação entre a tecnologia *Java* e o programa VLC (*VideoLan Client*) - este programa permite a visualização de conteúdos multimédia. Permitindo que instâncias do VLC sejam embebidas em componentes gráficas de *Java*, por exemplo em Janelas, Quadros e Painéis. A biblioteca incluída no VLC, *libclv*, pode ser usada para construir aplicações cliente/servidor com visualização de conteúdos multimédia com diferentes tipos de formato disponíveis⁽⁵⁾.

A biblioteca JMF [20] permite que conteúdos vídeo e áudio sejam adicionados a aplicações construídas em tecnologia *Java*. Com a utilização desta biblioteca pode-se capturar, reproduzir e transcodificar vídeo e áudio em vários formatos.

⁽⁴⁾ Java Bindings for vlc, <http://code.google.com/p/vlcj/>

⁽⁵⁾ http://wiki.videolan.org/VLC_Features_Formats

O *OpenCV* [21] é uma biblioteca multiplataforma para o desenvolvimento de aplicações na área de Visão por Computador. Possui módulos de processamento em tempo real de imagens e vídeo, diversos procedimentos usados em álgebra linear, e uma interface gráfica com sistema de janelas independentes. Para além disto, inclui bastantes algoritmos usados em aplicações de Visão por Computador, tais como filtros de imagem, procedimentos para calibração de câmaras e algoritmos de reconhecimento de objetos. Esta biblioteca suporta diferentes formatos de codificação de vídeo disponíveis em [21].

Todos os protótipos foram implementados num computador com o sistema operativo *Windows 7* de 32 bits. Os programas para edição e compilação, IDE (*Integrated Development Environment*), que foram escolhidos para elaboração destes protótipos foram o *Eclipse SDK* e *Microsoft Visual C++*, para o desenvolvimento em linguagem *Java* e *C++*, respetivamente.

Os protótipos foram construídos de modo a implementarem uma interface gráfica bastante simples, contendo apenas com um botão para iniciar o vídeo e um painel para a sua visualização, como se pode observar na Figura 12 e Figura 13. Com a aplicação iniciada, após clicar no botão o vídeo é ajustado aos limites do painel e visualizado.

Como referido, usando tecnologia *Java* foram experimentadas duas bibliotecas (*JMF* e *VLCJ*). Ambas cumpriram o requisito de permitir a visualização de vídeo com resoluções, até 1920x1080 (*Full HD*). Nas páginas de internet das bibliotecas encontram-se disponibilizados pequenos programas de exemplo cujo objetivo é visualizar uma sequência de vídeo usando poucas linhas de código. Por exemplo, usando a biblioteca *JMF*, a visualização de vídeo pode ser feita usando o seguinte código:

```
Player mediaPlayer = Manager.createRealizedPlayer( url );
mediaPlayer.start();
```

onde *url* corresponde ao vídeo e respetivos diretórios. Enquanto a biblioteca *JMF* é mais antiga (já não é fornecido suporte), a biblioteca *VLCJ* é atual (a última versão, 2.1, data de 20 de Maio de 2012). Note-se que para ser possível utilizar a biblioteca *VLCJ*, é necessário ter instalado no computador o *software VLC*.

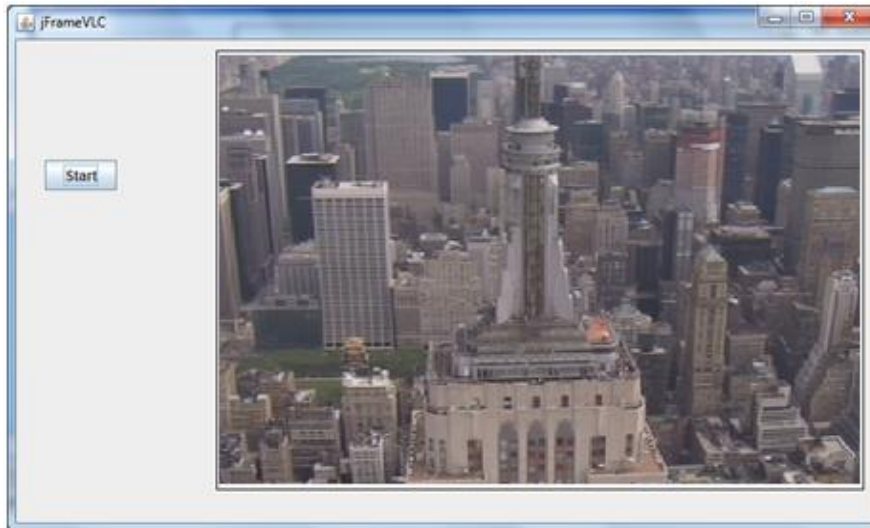


Figura 12: Minidemo utilizando tecnologia Java

Na tecnologia C++ foi experimentada a biblioteca *OpenCV*, que também cumpriu os requisitos propostos. São também disponibilizados vários exemplos na página oficial que explicam como visualizar vídeos. A versão desta biblioteca que foi experimentada data de Maio de 2012.



Figura 13: Minidemo utilizando tecnologia C++

Também foi ponderada a hipótese de implementação de um protótipo em Matlab. O Matlab é uma ferramenta extremamente poderosa para desenvolvimento de programas com algoritmos de processamento de imagem, vídeo e algoritmos que envolvem cálculo matricial. No entanto, esta possibilidade foi abandonada, pois obrigaria ter o Matlab instalado, com os custos associados às licenças deste *software*.

A aplicação a implementar neste trabalho poderia ser desenvolvida em qualquer uma das tecnologias mencionadas. No entanto, e tendo em conta que o *Java* é uma tecnologia portátil e sendo expectável um desenvolvimento mais rápido da aplicação,

dado o autor deste trabalho estar mais familiarizado com esta tecnologia, optou-se por utilizar a tecnologia *Java*.

Das duas bibliotecas experimentadas para *Java*, a biblioteca VLCJ seria à partida a candidata mais forte, pois é a mais recente e suporta uma maior diversidade de formatos de vídeo. No entanto, foram identificadas várias limitações da biblioteca VLCJ, nomeadamente:

- Impossibilidade de se obterem as dimensões do vídeo e sua correta duração antes do vídeo ser visualizado;
- Dificuldade em capturar eventos ao ser utilizado o rato;
- Impossibilidade de se adicionarem sentinelas aos eventos do vídeo (por exemplo quando o vídeo termina);
- *Flash* a negro ao ser iniciado o vídeo.

O JMF tem a limitação de já não ser suportado e apenas foi desenvolvido para um sistema de 32 bits. Contudo, optou-se pela utilização da biblioteca JMF por ser esta a que desempenhou as funções pretendidas de uma forma mais simples, sem apresentar as limitações do VLCJ.

Em resumo a ferramenta foi desenvolvida usando a linguagem de programação *java*. A programação do código foi feita usando IDE *Eclipse Helios*, num PC com o sistema operativo *Windows 7* de 32 bits. Para permitir visualizar as sequências de vídeo utilizou-se a biblioteca *Java Media Framework* (JMF). Para auxílio no desenho das componentes gráficas foi utilizado o *plugin Visual Swing* para o *Eclipse* e para gerar gráficos foi usado o *plugin JFreeChart*.

Foi gerado um ficheiro executável (.jar) com cerca de 7Mb para que a aplicação possa ser utilizada em qualquer computador. Bastando para isso que esse PC tenha instalado um *Java Runtime Environment* (JRE) 32bits e o JMF. Esta aplicação foi testada com sucesso em *Windows* nos seguintes sistemas operativos: XP, Vista e *Seven*.

3.2 Estrutura da interface gráfica

Nesta secção será apresentada a estrutura da interface gráfica da ferramenta. Esta aplicação prevê dois tipos de utilizadores: os *organizadores* e os *participantes*. Os *organizadores* são os utilizadores que projetam e configuram os testes de avaliação subjetiva, analisando posteriormente os resultados obtidos. Necessitam por isso de maior conhecimento sobre as características e potencialidades da ferramenta. Os *participantes* são as pessoas que irão avaliar a qualidade das sequências de vídeo exibidas durante um teste. Devem seguir as instruções fornecidas pelos organizadores e disponibilizadas pela ferramenta. Este tipo de utilizadores não necessita de ter um

conhecimento profundo sobre a ferramenta, mas deverá estar familiarizado com a interface para atribuição das pontuações às sequências de vídeo.

A Figura 14 representa o fluxo de dados geral da aplicação, que engloba os três principais intervenientes da ferramenta: sequências de vídeo, aplicação e ficheiros de texto. As sequências de vídeo são visualizadas através da aplicação. É também através da aplicação que se criam e lêem ficheiros de texto referentes à configuração dos testes, assim como ficheiros que guardam as pontuações de qualidade de vídeo provenientes dos testes.



Figura 14: Fluxo de dados geral

Na janela principal da interface gráfica, representada na Figura 15, está disponível uma barra com três menus. No menu *Subjective Test* pode-se obter uma breve descrição dos métodos de avaliação subjetiva implementados (*Methods Info*). Recorde-se que os métodos de avaliação subjetiva implementados foram: ACR-HR, DSIS, SCACJ e AVCA (abordado na secção **Erro! A origem da referência não foi encontrada.**).

No submenu *New Test* pode-se efetuar a configuração de um teste de avaliação subjetiva, escolhendo inicialmente entre os métodos recomendados pela ITU ou o método de avaliação contínua proposto no âmbito deste trabalho (AVCA). No submenu *Load Test*, pode-se carregar um ficheiro de texto, com a configuração do teste, gerado anteriormente na opção *New Test*. Após validação do ficheiro de configuração será apresentado ao observador um painel para este introduzir os seus dados pessoais. De seguida inicia-se o teste de avaliação subjetiva propriamente dito. Adiante serão explicadas com maior detalhe estas duas últimas opções, assim como a parte referente aos resultados do teste.

No menu *Data Analysis* são efetuados os cálculos e apresentação de resultados, com base nos ficheiros de pontuações escolhidos. Inicialmente escolhe-se o tipo de visualização pretendida: visualizar os resultados de *Mean Opinion Score (Compute MOS)* ou os resultados do método de avaliação contínua (*AVCA Results*). Para além de se realizarem os cálculos dos valores de MOS, a opção *Compute MOS* permite escolher a forma como é feita a deteção de participantes com pontuações incoerentes: através do método definido pela ITU (*ITU-T method*), descrito na secção 2.3, ou através do método da correlação (*Correlation Analysis*), método que será apresentado na secção 4.2.

No menu *Help* são disponibilizadas informações sobre a ferramenta (*About*).

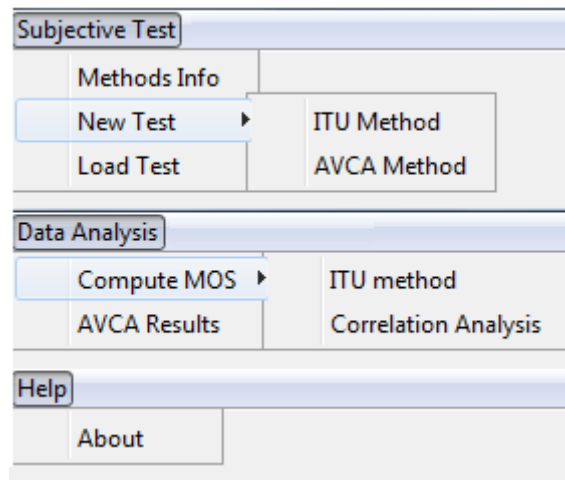


Figura 15: Menus da interface gráfica

Em seguida serão explicadas, detalhadamente, as seguintes funcionalidades da ferramenta: configuração de um novo teste, realização de um teste e apresentação de resultados.

3.2.1 Configuração de um novo teste

O fluxo de dados para configuração de um novo teste de avaliação subjetiva pode ser observado na Figura 16. Um novo teste de avaliação subjetiva configura-se escolhendo a opção *New Test* do menu *Subjective Test* da interface gráfica principal (GUI). Depois de selecionado o tipo de método de avaliação subjetiva, incluem-se as seqüências de vídeo pretendidas. A configuração do teste é de seguida armazenada num ficheiro de texto com as informações necessárias para posterior realização do teste de avaliação subjetiva.

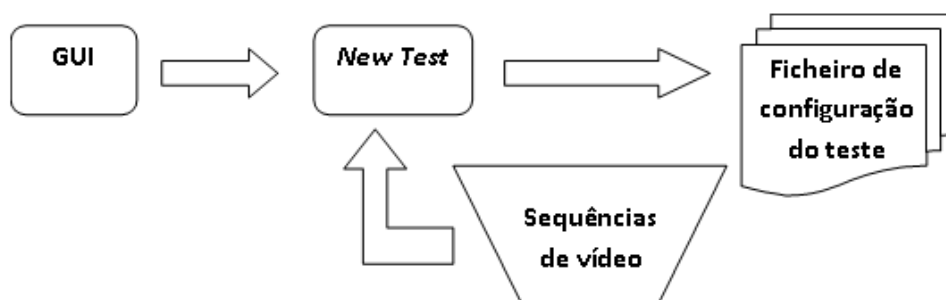


Figura 16: Fluxo de dados configuração de um novo teste

A janela de configuração dos testes para um método recomendado pela ITU, é ilustrada na Figura 17. Esta janela contém uma *combobox* que permite escolher o tipo de método de avaliação subjetiva, diversos botões para diferentes opções, duas listas, designadas por *Video Group List* (à esquerda) e *Display Order List* (à direita), onde serão incluídas as seqüências de vídeo escolhidas pelo organizador a figurar no teste.

Na primeira lista, *Video Group List*, as sequências de vídeo ficam agrupadas segundo ‘famílias’ ou grupos de vídeos. Entende-se por ‘família’ ou grupo de vídeos um conjunto de sequências de teste obtidas com base na mesma sequência de vídeo original. Estes conjuntos de vídeos são caracterizados pelo identificador do grupo (G1, G2, ..., Gn). Assim cada sequência de vídeo fica referenciada por um identificador de grupo. Estes vídeos são selecionados através do gestor de ficheiros apresentado após clicar no botão *Get Files*.

Devido às características dos métodos ACR-HR e DSIS, quando se utiliza um destes métodos é necessário identificar uma das sequências do grupo como sequência de referência, antes de se adicionar o grupo de vídeos à segunda lista. Para os métodos DSIS e SCACJ é necessário que cada grupo contenha pelo menos duas sequências de vídeo, pois os vídeos são exibidos aos pares.

Os grupos de vídeos são adicionados à segunda lista, *Display Order List*, através do botão *Add Group*. Nesta segunda lista é apresentada a configuração final, onde as sequências estão organizadas pela ordem com que serão exibidas durante a realização do teste. Por exemplo, se o método ACR-HR for selecionado na *Display Order List*, cada sequência de vídeo deverá ser descrita pelo diretório da sequência de vídeo, seguido da identificação do grupo e identificador de referência ou condição de teste, como por exemplo:

```
D:\Documents\AVIs\city.avi G1 R,
```

em que G1 é o identificador do grupo e ‘R’ significa que a sequência de vídeo é uma referência (caso fosse uma condição de teste, seria utilizada a designação ‘T’).

Se for selecionado o método DSIS, a configuração será a seguinte: identificação do grupo, seguido do caminho para o ficheiro que contém o vídeo de teste e finalmente o caminho para o ficheiro que contém o vídeo de referência.

No caso de se pretender usar o método SCACJ, é necessário apresentar todas as combinações possíveis entre as sequências de vídeo escolhidas. Sendo assim, no ficheiro de configuração figura a identificação do grupo, precedido pelas duas sequências de vídeo para as quais o participante irá comparar a qualidade. Na linha seguinte constará o mesmo par de sequências de vídeo, mas apresentadas por ordem inversa.

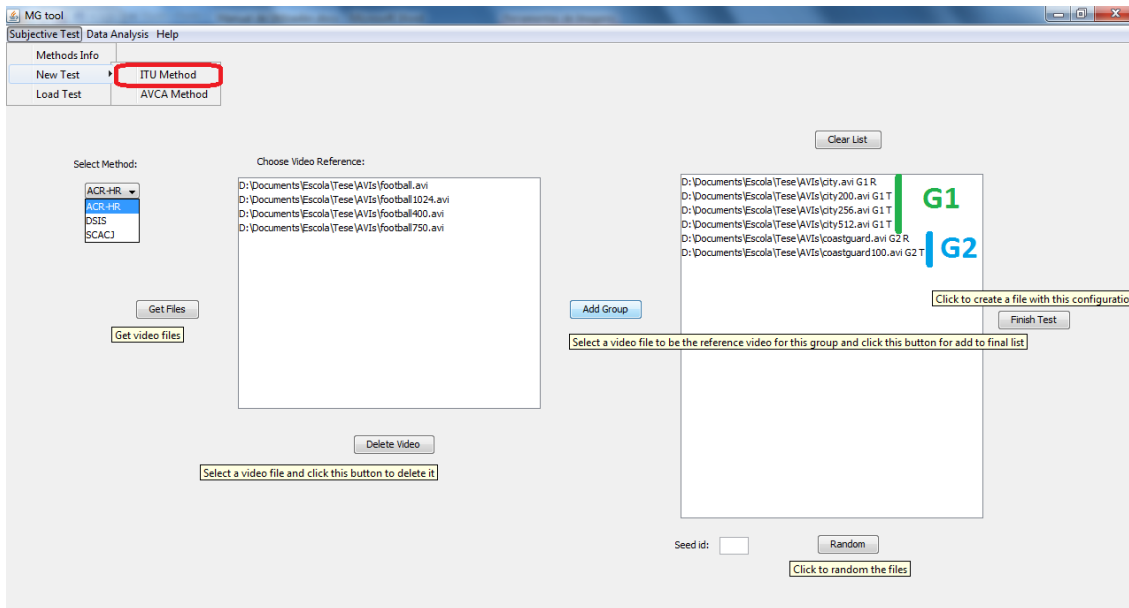


Figura 17: Configuração de um novo teste ACR-HR

Como o tipo de método influencia a configuração dos ficheiros de teste, nomeadamente o modo como são criados e lidos, não é possível modificar o tipo de método utilizado no teste quando a lista *Display Order List* não está vazia.

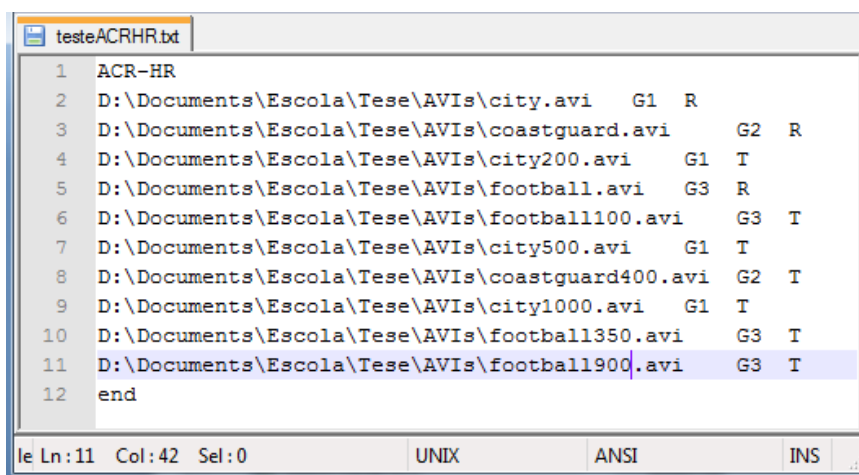
Na lista *Display Order List* é apresentada a configuração final, onde as sequências estão organizadas pela ordem com que serão exibidas durante a realização do teste. Nesta lista apresentam-se todas as famílias de vídeos adicionadas. As sequências de vídeo aparecem inicialmente segundo grupos, pela ordem com que foram incluídas. Mas a ordem com que serão visualizadas durante os testes pode ser gerada de uma forma pseudoaleatória através do botão *Random*.

No entanto, a ordem pseudo-aleatória que é gerada segue algumas regras, de forma a evitar a visualização consecutiva de sequências de vídeo pertencentes à mesma família. Esta regra pretende evitar o aborrecimento dos participantes durante a realização do teste, fazendo com que a visualização de sequências de vídeo pertencentes à mesma família fique bem distribuída ao longo da duração do teste. O algoritmo implementado é o seguinte: escolhe-se aleatoriamente uma sequência de vídeo da lista *Display Order List* e volta-se a colocar esse vídeo na mesma lista, mas numa posição que garanta um espaçamento mínimo entre sequências da mesma família. Escolhe-se novamente uma sequência, mas de outra família, e volta-se a colocar na lista numa posição que obedeça à regra do espaçamento mínimo. O processo repete-se até todas as sequências terem sido recolocadas. O espaçamento mínimo entre sequências do mesmo grupo foi definido como metade do número total de grupos. Por exemplo, se o teste incluir oito grupos de sequências, o espaçamento mínimo entre sequências do mesmo grupo é igual a quatro. Com este algoritmo, só no caso dos grupos de sequências terem dimensões diferentes entre si é que existe a possibilidade de ser exibidas sequências de vídeo da mesma família em posições consecutivas no final do teste.

O gerador de números pseudoaleatórios pode ser iniciado com base num temporizador ou com base numa ‘semente’. Se for preenchido o campo *Seed id* com um número inteiro, a ordem gerada para os elementos contidos na *Display Order List* será a mesma desde que essa semente seja a mesma. Isto permite gerar a mesma ordem, para o mesmo conjunto de sequências, em momentos de tempo diferentes.

Antes de terminar a configuração do teste, e caso se pretenda ajustar a ordem das sequências na *Display Order List*, pode-se utilizar o mecanismo de selecionar uma linha desta lista e arrasta-la para a posição desejada (*drag and drop*). Este mecanismo facilita bastante nos casos em que se pretende alterar apenas algumas posições entre as sequências de vídeo sem alterar toda a ordem de visualização como faz o *Random*.

Após terminadas as ações, clica-se no botão *Finish Test* para criar o ficheiro de configuração do teste. É gerado um ficheiro de texto com o seguinte conteúdo: nome do método, sequências de vídeo com os respetivos grupos e identificadores de sequências. Como exemplo, na Figura 18 encontra-se o conteúdo de um ficheiro de configuração relativo ao método de avaliação ACR-HR. A palavra ‘end’ assinala o fim do ficheiro de configuração e é posteriormente usada durante a realização do teste para se saber que este chegou ao fim.



```
1 ACR-HR
2 D:\Documents\Escola\Tese\AVIs\city.avi G1 R
3 D:\Documents\Escola\Tese\AVIs\coastguard.avi G2 R
4 D:\Documents\Escola\Tese\AVIs\city200.avi G1 T
5 D:\Documents\Escola\Tese\AVIs\football.avi G3 R
6 D:\Documents\Escola\Tese\AVIs\football100.avi G3 T
7 D:\Documents\Escola\Tese\AVIs\city500.avi G1 T
8 D:\Documents\Escola\Tese\AVIs\coastguard400.avi G2 T
9 D:\Documents\Escola\Tese\AVIs\city1000.avi G1 T
10 D:\Documents\Escola\Tese\AVIs\football1350.avi G3 T
11 D:\Documents\Escola\Tese\AVIs\football1900.avi G3 T
12 end
```

Figura 18: Ficheiro de configuração de um novo teste ACR-HR

3.2.2 Realização de um teste

Na Figura 19 é representado o fluxo de dados referente à realização de um teste de avaliação subjetiva. Baseado nas informações contidas no ficheiro de configuração é definida a forma de apresentação das sequências de vídeo e a escala de avaliação usada no teste. Após a visualização das sequências de vídeo os participantes avaliam a sua qualidade através de uma escala de pontuações. Antes de se dar início ao teste subjetivo são introduzidas as informações dos participantes, que servem também para diferenciar as suas pontuações dos outros participantes quando se apresentam os resultados. No final as pontuações atribuídas pelos participantes são guardadas num ficheiro de texto.

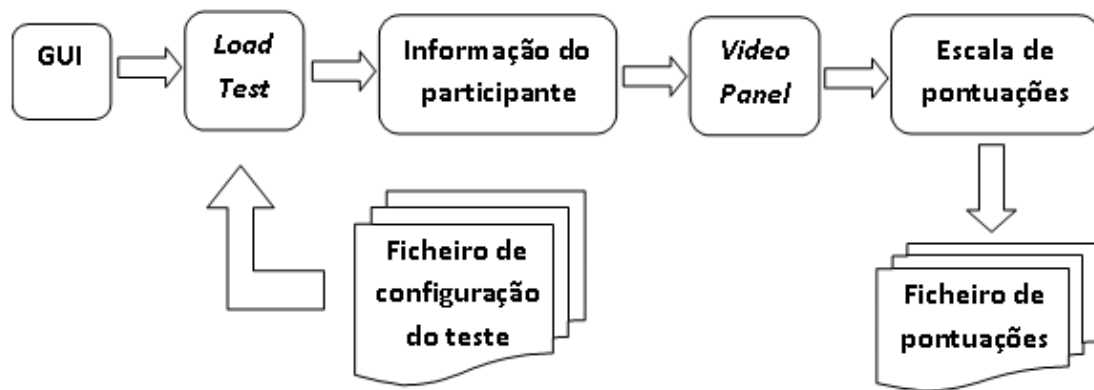


Figura 19: Fluxo de dados realização de um teste

Conforme referido anteriormente, para a realização de um teste de avaliação subjetiva utiliza-se a opção **Load Test** do menu **Subjective Test**. É mostrado um gestor de ficheiros para escolher o ficheiro de configuração com o teste pretendido.

De forma a prevenir erros durante a realização do teste de avaliação subjetiva é necessário garantir que todas as sequências de vídeo possam ser lidas corretamente. Para tal, durante o carregamento da configuração do teste é feita uma verificação aos caminhos dos ficheiros que contém as sequências de vídeo que são mostradas durante o teste. Se algum destes caminhos estiver incorreto (i.e., a aplicação não encontra o ficheiro), é mostrada uma janela de diálogo com uma mensagem onde se informa que o respetivo ficheiro de vídeo não pode ser lido, como ilustrado na Figura 20.

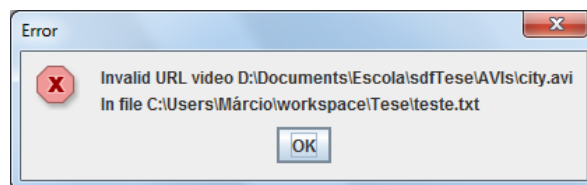
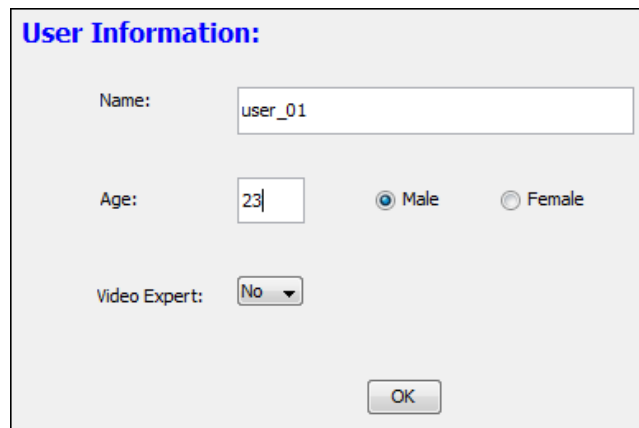


Figura 20: Exemplo de *url* inválido de um vídeo

Note-se que quando se configura um teste subjetivo num determinado PC e depois se pretende realizá-lo noutra PC, é necessário ter em atenção a estrutura de diretórios onde se encontram os ficheiros que contém os vídeos. Se existirem diretórios diferentes de um PC para outro, na altura em que se carrega o teste previamente configurado será mostrada a mensagem de erro ilustrada na Figura 20. Este potencial problema pode ser facilmente contornado armazenando caso o material a utilizar nos testes num dispositivo externo (e.g., disco externo). Em alternativa pode-se utilizar exatamente a mesma estrutura de diretórios, de forma a manter válidos os caminhos para os ficheiros que contém as sequências de vídeo.

Após feita a verificação mencionada, é disponibilizado um painel para o participante inserir os seus dados. A interação do participante com a aplicação começa neste momento.



The image shows a dialog box titled "User Information:" with a light gray background. It contains three input fields: "Name:" with the text "user_01", "Age:" with the number "23", and "Video Expert:" with a dropdown menu showing "No". To the right of the "Age:" field are two radio buttons labeled "Male" (which is selected) and "Female". At the bottom center of the dialog box is an "OK" button.

Figura 21: Painel de informação sobre o participante

Neste painel, ilustrado na Figura 21, o utilizador deverá preencher os campos **nome**, **idade**, **género** e indicar se é especialista em vídeo ou não. Depois de feitas as verificações a estes campos – **nome** não pode estar sem conteúdo e **idade** tem de ser um número inteiro entre 1 e 99 – é dado início à sessão de avaliação subjetiva da qualidade do vídeo. O nome do participante pode ser fictício uma vez que os participantes nem sempre se mostram disponíveis para facultar esse dado. Nestes casos, o nome do participante pode ser substituído por um identificador.

De modo a que as informações relativas ao participante constem no ficheiro de pontuações gerado no final da sessão, estas são guardadas após clicar no botão **OK**. Dá-se então início à sessão de avaliação subjetiva, que começa com uma breve explicação sobre o tipo de método utilizado no teste (exemplo na Figura 22 para um teste com o método DSIS), e de que forma as sequências de vídeo vão ser exibidas. Durante esta breve introdução, aborda-se também a escala de pontuações que vai ser usada para avaliar a qualidade dos vídeos. Todos estes parâmetros dependem do tipo de método utilizado no teste.

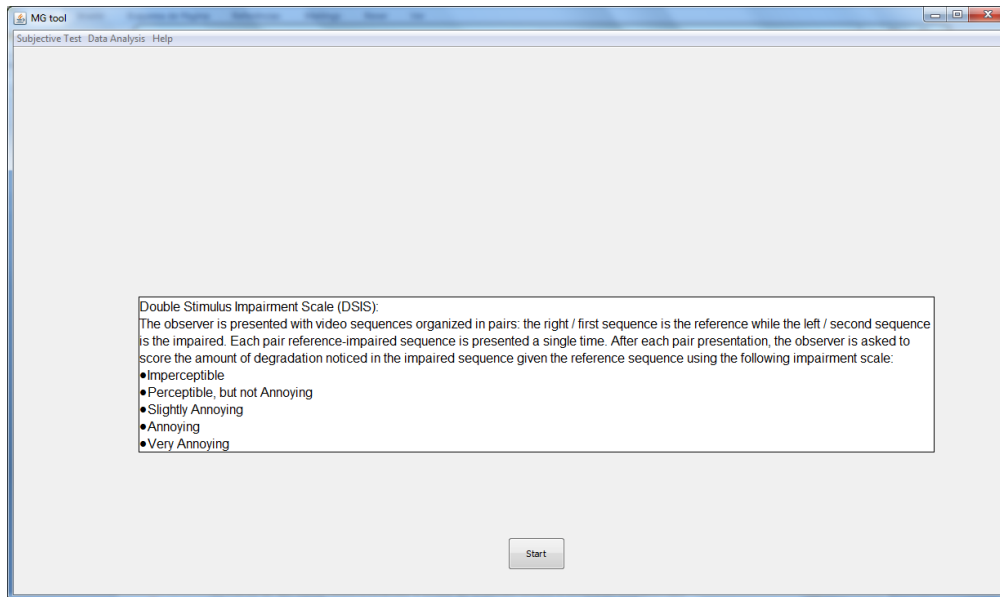


Figura 22: Descrição do método DSIS

Entendido o modo de funcionamento do teste, o participante clica no botão **Start** para dar início à visualização das sequências de vídeo. Enquanto os vídeos estão a ser mostrados, Figura 23, o botão **Start** desaparece e só volta a estar visível depois de efetuada a avaliação de cada sequência de vídeo exibida no teste. Para evitar que os participantes se distraiam e percam a concentração durante a avaliação da qualidade das sequências de vídeo, os menus da barra de ferramentas não estão disponíveis durante a realização do teste. Os menus voltam a ficar disponíveis ao ser terminado o teste.

Dois segundos após terminada a visualização de cada vídeo, a escala de pontuações fica visível. Recorde-se que a escala de pontuações utilizada depende do método seguido para o teste. Na Figura 24 apresenta-se a escala usada num teste DSIS. Na janela correspondente à escala de avaliação apenas é permitido escolher uma das pontuações disponíveis. Depois de se escolher a pontuação desejada e clicar no botão **OK**, volta a aparecer o botão **Start** para que o participante possa dar início à visualização da próxima sequência de vídeo a avaliar. Este procedimento repete-se sucessivamente até aparecer uma janela idêntica à ilustrada na Figura 25, que indica o final que a sessão de testes chegou ao fim.

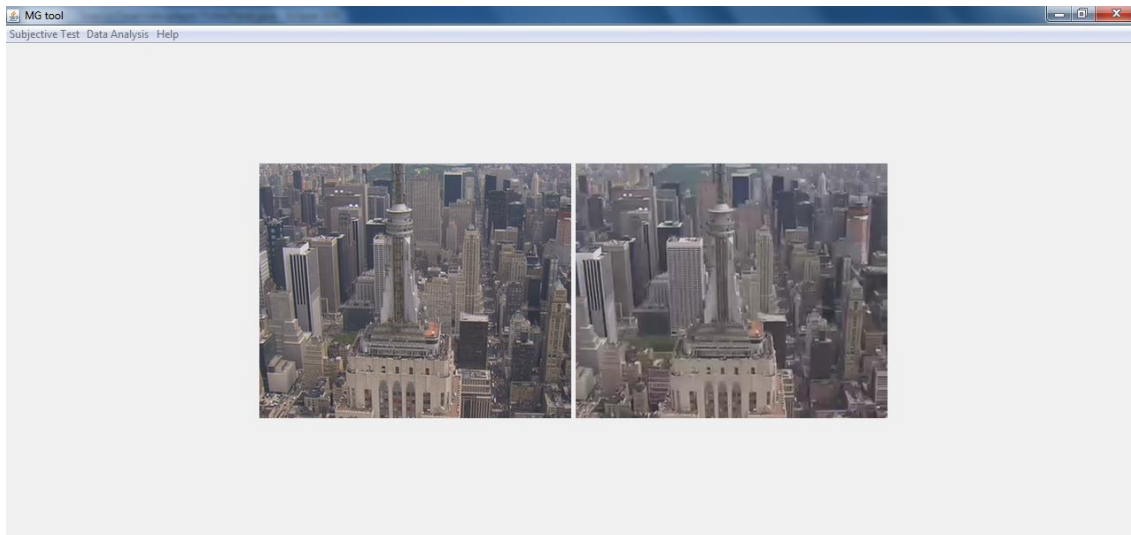


Figura 23: Teste de avaliação do método DSIS

De acordo com [4], quando se utilizam os métodos DSIS ou SCACJ, as sequências de vídeo devem ser exibidas aos pares, e colocadas lado a lado quando as dimensões do ecrã o permitam. Esta opção é influenciada pelo tamanho do ecrã em que está a ser efetuado o teste. Se a soma da largura dos pares de vídeo for inferior à largura total do ecrã, os vídeos são mostrados em simultâneo; caso contrário, as sequências de vídeo são visualizadas uma de cada vez. No caso particular do método DSIS, a sequência de referência é sempre o primeiro vídeo a ser visualizado ou, no caso de as sequências serem exibidas em simultâneo, é sempre o vídeo apresentado no lado esquerdo do ecrã.

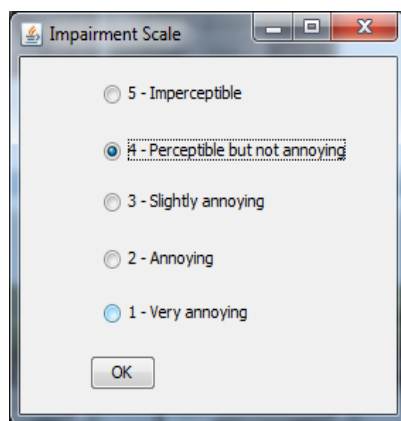


Figura 24: Escala de avaliação do método DSIS

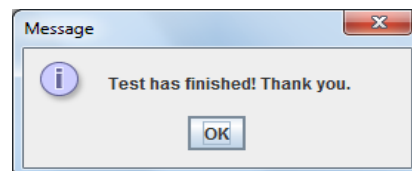
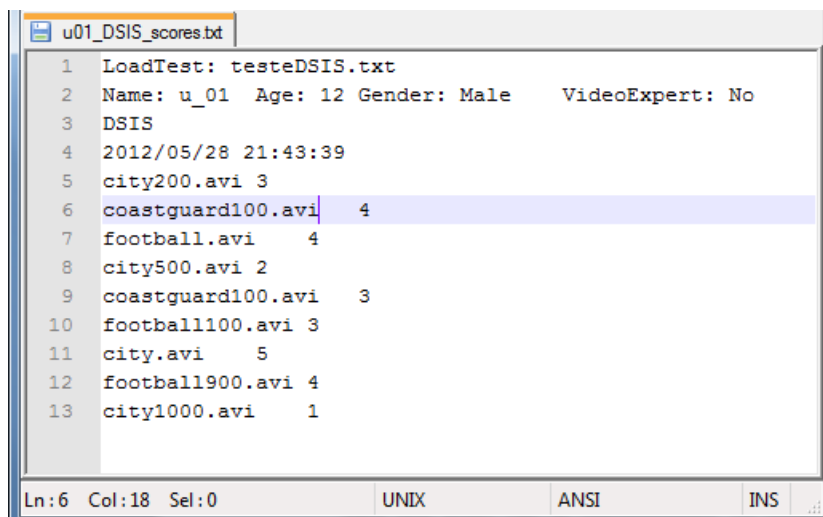


Figura 25: Mensagem que indica fim do teste

Terminado o teste de avaliação subjetiva, as pontuações dadas pelo participante ficam guardadas num ficheiro de texto. O nome dado a este ficheiro é construído automaticamente com base no nome do participante e no tipo de teste realizado. Por exemplo, as pontuações dadas por 'user_01' durante a realização de um teste do tipo DSIS seria: "user_01_DSIS_scores.txt". No ficheiro das pontuações dadas por um participante estão contidos os seguintes dados: o nome do ficheiro de configuração do

teste que foi realizado, a informação relativa ao participante (nome, idade, género e tipo de participante), o nome do método de avaliação subjetiva, a data e hora em que se efetuou o teste, e finalmente o nome das sequências de vídeo avaliadas e respetivas pontuações. A Figura 26 ilustra o conteúdo de um ficheiro de pontuações que corresponde à realização de um teste DSIS por um participante identificado por “u01”.



```
u01_DSIS_scores.txt
1 LoadTest: testeDSIS.txt
2 Name: u_01 Age: 12 Gender: Male VideoExpert: No
3 DSIS
4 2012/05/28 21:43:39
5 city200.avi 3
6 coastguard100.avi 4
7 football.avi 4
8 city500.avi 2
9 coastguard100.avi 3
10 football100.avi 3
11 city.avi 5
12 football900.avi 4
13 city1000.avi 1
Ln:6 Col:18 Sel:0 UNIX ANSI INS
```

Figura 26: Ficheiro de pontuações de um participante de um teste com o método DSIS

No caso de se utilizar o método de avaliação subjetiva SCACJ, o formato do ficheiro das pontuações é ligeiramente diferente. Esta diferença consiste na identificação do par de sequências usadas para comparação, ao contrário dos restantes tipos de teste (ACR-HR ou DSIS), onde apenas consta a informação sobre a condição de teste (sequência de vídeo avaliada).

Os ficheiros com as pontuações dos participantes devem ser gravados com nomes diferentes para que estes não se sobreponham. Para que isto aconteça basta que o nome do participante seja sempre diferente no painel de dados do participante, ilustrado na Figura 21.

3.2.3 Apresentação de resultados

Nesta secção descreve-se o modo como são apresentados os resultados das pontuações dos participantes. O diagrama que ilustra o fluxo de dados envolvidos na apresentação de resultados pode ser observado na Figura 27. Existem dois tipos de análise de dados: apresentação dos cálculos de MOS e apresentação dos resultados provenientes do método proposto AVCA, que será descrito na secção 4.1. Os resultados são obtidos através da análise dos ficheiros de pontuações que forem escolhidos, tendo em comum o mesmo ficheiro de configuração do teste. Estes resultados são apresentados em tabela e podem posteriormente ser exportados para gráficos ou ficheiros CSV.

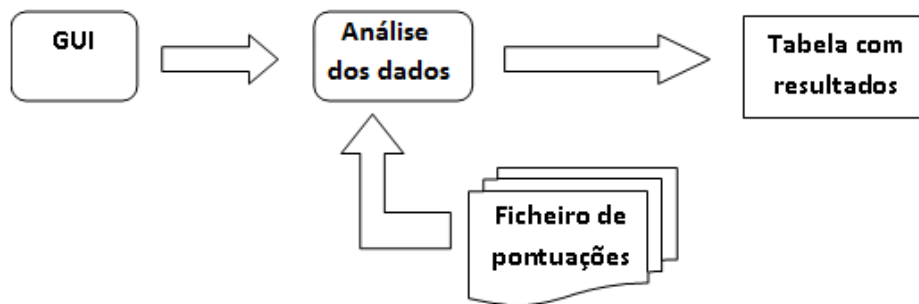


Figura 27: Fluxo de dados apresentação de resultados

No menu **Data Analysis** seleciona-se o tipo de análise de resultados que se pretende. Resultados de um teste realizado segundo o método de avaliação contínua da qualidade do vídeo (**AVCA Results**) ou resultados para obter os valores de MOS (**Compute MOS**). Neste último submenu seleciona-se o tipo de método para a deteção de pontuações incoerentes por parte dos participantes. De seguida os ficheiros com as pontuações dos participantes são selecionados a partir de um gestor de ficheiros.

O cálculo de *Mean Opinion Score* (MOS) e deteção de participantes com pontuações incoerentes, são feitos de duas formas distintas (**ITU method** e **Correlation Analysis**). A apresentação dos resultados, em forma de tabela, depende do tipo de método de avaliação subjetiva usado no teste. Caso se escolha a opção **ITU method** e na realidade o nome do método contido no ficheiro seja o AVCA, ou vice-versa, é mostrada uma janela de erro.

Para um método ITU as colunas da tabela são constituídas pelos seguintes campos, conforme ilustrado na Figura 28:

- nome das sequências de vídeo;
- identificação dos participantes, em que o total de participantes corresponde ao número de ficheiros selecionados anteriormente;
- *RawMOS* – consiste na média das pontuações dadas por todos os participantes, independentemente de terem sido considerados válidos ou não;
- desvio padrão de *RawMOS*;
- valor resultante para o MOS, baseado nas pontuações dos participantes que foram considerados válidos, segundo a metodologia descrita na secção 2.3;
- desvio padrão de MOS.

Caso o método de avaliação subjetiva seja o ACR-HR são ainda acrescentadas duas colunas:

- vídeo de referência – para cada sequência de vídeo, indica se esta foi usada como referência ou não.;
- valor resultante para o *Differential Mean Opinion Score* (DMOS), calculado de acordo com a equação (2.10) da secção 2.3.

As pontuações atribuídas pelos participantes que foram rejeitados para o cálculo do MOS ficam com fundo amarelo. Os resultados são arredondados para 2 casas decimais.

Após visualização da tabela com os resultados dos participantes ficam disponíveis dois novos botões. Um deles (***Export CSV***) permite exportar os resultados para um ficheiro CSV e o outro (***Generate Graph***) pode ser utilizado para gerar gráficos com os valores de MOS obtidos para cada condição de teste. Este último gera dois tipos de gráficos: um com os resultados de MOS ordenados em função do número da sequência de vídeos, ilustrado na Figura 29; outro com os resultados de MOS ordenados em função do valor das pontuações de MOS (ordem crescente), ilustrado na Figura 30;

MG tool
Subjective Test Data Analysis Help

Video Sequence	m01	m02	m03	m04	m05	m06	m07	m08	m09	RawMOS	SD RawMOS	MOS	SD MOS	Video Reference	DMOS
soccer_480p30_3mbs_corrupted_plr_3.2325_.avi	2	2	1	2	2	1	1	2	2	1.67	0.5	1.67	0.5	NO	2.89
soccer_480p30_3mbs_corrupted_plr_1.7177_.avi	1	1	1	2	1	1	1	1	1	1.11	0.6	1.11	0.6	NO	3.45
soccer_480p30_3mbs_corrupted_plr_1.3832_.avi	2	2	2	3	2	2	2	2	2	2.11	0.69	2.11	0.69	NO	2.45
snow_mnt_480p30_3mbs_corrupted_plr_0.60756_.avi	4	3	4	1	4	4	4	3	3	3.33	1.21	3.33	1.21	NO	1.67
old_town_cross_480p25_2mbs_corrupted_plr_0.563_.avi	4	5	4	1	4	4	4	4	5	4.0	1.72	4.0	1.72	NO	0.56
station_480p25_2mbs_corrupted_plr_0.22712_.avi	4	4	5	1	5	5	5	4	4	4.11	2.14	4.11	2.14	NO	-1.11
sunflower_480p25_1mbs_corrupted_plr_7.7564_.avi	1	1	1	5	1	1	1	1	1	1.44	2.52	1.44	2.52	NO	3.12
park_joy_480p25_5mbs.avi	5	5	5	1	5	5	5	5	5	4.56	2.85	4.56	2.85	YES	0.0
raven_480p30_2mbs_corrupted_plr_5.1629_.avi	3	2	3	3	3	3	3	2	2	2.67	2.9	2.67	2.9	NO	1.89
snow_mnt_480p30_3mbs_corrupted_plr_2.2527_.avi	3	3	3	3	3	3	3	3	3	3.0	2.9	3.0	2.9	NO	2.0
big_ships_480p30_2mbs.avi	5	5	5	1	5	5	5	5	5	4.56	3.19	4.56	3.19	YES	0.0
sunflower_480p25_1mbs_corrupted_plr_2.4583_.avi	2	2	2	2	2	2	2	2	2	2.0	3.19	2.0	3.19	NO	2.56
station_480p25_2mbs_corrupted_plr_1.9709_.avi	3	3	3	3	3	3	3	3	3	3.0	3.19	3.0	3.19	NO	0.0
raven_480p30_2mbs_corrupted_plr_1.7177_.avi	3	4	3	3	3	3	3	4	4	3.33	3.23	3.33	3.23	NO	1.23
old_town_cross_480p25_2mbs_corrupted_plr_2.5564_...	2	2	2	2	2	2	2	2	2	2.0	3.23	2.0	3.23	NO	2.56
park_joy_480p25_5mbs_corrupted_plr_3.8627_.avi	2	2	2	2	2	2	2	2	2	2.0	3.23	2.0	3.23	NO	2.56
soccer_480p30_3mbs_corrupted_plr_1.3028_.avi	5	5	5	1	5	5	5	5	5	4.56	3.49	4.56	3.49	NO	0.0
sunflower_480p25_1mbs_corrupted_plr_0.53605_.avi	4	4	4	1	4	4	4	4	4	3.67	3.63	3.67	3.63	NO	0.89
station_480p25_2mbs_corrupted_plr_0.21373_.avi	5	5	5	1	5	5	5	5	5	4.56	3.87	4.56	3.87	NO	-1.56
snow_mnt_480p30_3mbs_corrupted_plr_0.59624_.avi	5	5	5	1	5	5	5	5	5	4.56	4.09	4.56	4.09	NO	0.44
big_ships_480p30_2mbs_corrupted_plr_5.1629_.avi	2	2	2	2	2	2	2	2	2	2.0	4.09	2.0	4.09	NO	2.56
park_joy_480p25_5mbs_corrupted_plr_0.21373_.avi	2	2	2	2	2	2	2	2	2	2.0	4.09	2.0	4.09	NO	2.56
raven_480p30_2mbs.avi	5	5	5	1	5	5	5	5	5	4.56	4.3	4.56	4.3	YES	0.0
old_town_cross_480p25_2mbs_corrupted_plr_7.0338_...	2	2	2	2	2	2	2	2	2	2.0	4.3	2.0	4.3	NO	2.56
station_480p25_2mbs_corrupted_plr_1.9848_.avi	3	3	3	3	3	3	3	3	3	3.0	4.3	3.0	4.3	NO	0.0
snow_mnt_480p30_3mbs.avi	5	5	5	1	5	5	5	5	5	4.56	4.51	4.56	4.51	YES	0.44
sunflower_480p25_1mbs_corrupted_plr_7.0338_.avi	1	1	1	5	1	1	1	1	1	1.44	4.7	1.44	4.7	NO	3.12
big_ships_480p30_2mbs_corrupted_plr_0.19048_.avi	4	4	4	1	4	4	4	4	4	3.67	4.8	3.67	4.8	NO	0.89
raven_480p30_2mbs_corrupted_plr_5.0147_.avi	3	3	3	3	3	3	3	3	3	3.0	4.8	3.0	4.8	NO	1.56
park_joy_480p25_5mbs_corrupted_plr_1.5705_.avi	2	2	2	2	2	2	2	2	2	2.0	4.8	2.0	4.8	NO	2.56
station_480p25_2mbs.avi	5	5	5	1	5	5	5	5	5	4.56	4.99	4.56	4.99	YES	-1.56
old_town_cross_480p25_2mbs_corrupted_plr_7.7564_...	3	3	3	3	3	3	3	3	3	3.0	4.99	3.0	4.99	NO	1.56
big_ships_480p30_2mbs_corrupted_plr_0.16803_.avi	4	4	4	4	4	4	4	4	4	4.0	4.99	4.0	4.99	NO	0.56
snow_mnt_480p30_3mbs_corrupted_plr_7.8129_.avi	2	2	2	2	2	2	2	2	2	2.0	4.99	2.0	4.99	NO	3.0
soccer_480p30_3mbs.avi	5	5	5	5	5	5	5	5	5	5.0	4.99	5.0	4.99	YES	-0.44
park_joy_480p25_5mbs_corrupted_plr_7.0338_.avi	1	1	1	1	1	1	1	1	1	1.0	4.99	1.0	4.99	NO	3.56
sunflower_480p25_1mbs_corrupted_plr_0.563_.avi	4	4	4	4	4	4	4	4	4	4.0	4.99	4.0	4.99	NO	0.56
raven_480p30_2mbs_corrupted_plr_1.7061_.avi	3	3	3	3	3	3	3	3	3	3.0	4.99	3.0	4.99	NO	1.56
big_ships_480p30_2mbs_corrupted_plr_5.0147_.avi	1	1	1	1	1	1	1	1	1	1.0	4.99	1.0	4.99	NO	3.56
old_town_cross_480p25_2mbs_corrupted_plr_2.4583_...	1	1	1	1	1	1	1	1	1	1.0	4.99	1.0	4.99	NO	3.56

Generate Graph Export CSV

Figura 28: Tabela de resultados com a metodologia ACR-HR

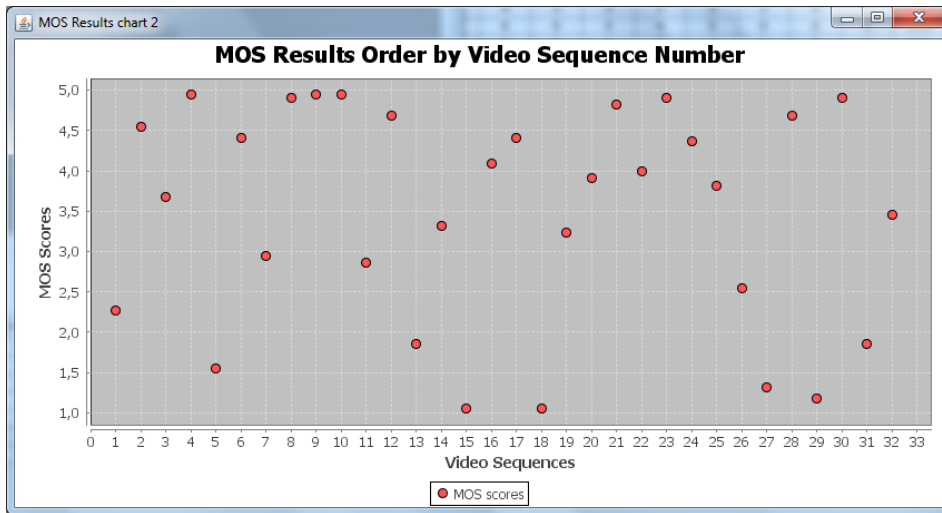


Figura 29: Resultados de MOS ordenados em função do número da sequência de vídeos

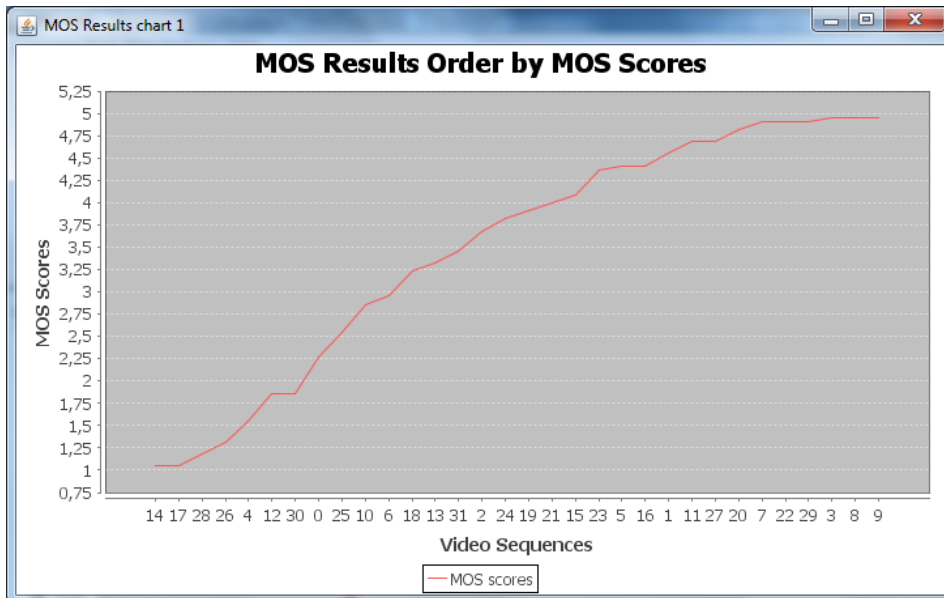


Figura 30: Resultados de MOS ordenados em função do valor das pontuações de MOS (ordem crescente)

4 Contribuições adicionais

Depois de concluída a implementação base da ferramenta, descrita no Capítulo 3, decidiu-se ampliar as suas potencialidades implementando duas novas funcionalidades.

A primeira funcionalidade adicional consiste na implementação de um método da avaliação subjectiva da qualidade, não normalizado. Este método consiste em pedir aos participantes para assinalarem, com auxílio do rato, os artefactos visualizados durante a execução do vídeo. É portanto um método de avaliação contínua, *single stimulus*, e foi denominado por *Artifact Visibility Continuous Assessment (AVCA)*.

A outra funcionalidade implementada diz respeito ao modo como são detetadas e rejeitadas as pontuações incoerentes dos participantes. Este procedimento foi denominado de *Correlation Analysis* e utiliza cálculos de correlação para analisar o grau de associação linear entre as pontuações dadas pelos vários participantes.

4.1 *Artifact Visibility Continuous Assessment (AVCA)*

Neste método, assume-se que existem artefactos que foram introduzidos de forma artificial nas sequências de vídeo a avaliar. O objetivo do teste é determinar se esses artefactos introduzidos são ou não perceptíveis para um observador humano. É possível que nem todas as perdas introduzidas no vídeo, por processos de codificação ou transmissão, seja perceptíveis. Como exemplos de situações que podem levar a que um observador não se aperceba de um artefacto tem-se: perdas isoladas de pacotes que ocorrem em zonas estáticas do vídeo, perdas que ocorrem fora da região de interesse do vídeo e efeitos de mascaramento provocados pelo próprio conteúdo do vídeo.

O método que foi desenvolvido designou-se por *Artifact Visibility Continuous Assessment (AVCA)* e pretende avaliar o impacto causado por um determinado artefacto introduzido no vídeo. Consiste em apresentar ao observador uma sequência de vídeo mais longa, que poderá ser uma composição de sequências de vídeo, de forma a diversificar o conteúdo (cenas) que é exibido. Sempre que o observador se aperceba de um artefacto durante a apresentação da sequência de vídeo, deverá assinalá-lo (usando o rato). Devido à natureza deste teste, não existe uma escala de avaliação de qualidade.

Na Figura 31 é ilustrado o conjunto de parâmetros a configurar para este tipo de teste (**AVCA Method**). Inicia-se selecionando a sequência de vídeo, a figurar no teste, através do botão **Get File**. Os tempos de artefactos são adicionados à lista final de tempos através do botão **Add time**. Estes tempos devem estar em segundos e servem para posteriormente comparar com os tempos dos artefactos assinalados pelos participantes. Para eliminar um tempo da lista de tempos basta seleccioná-lo e clicar no botão **Delete time**.

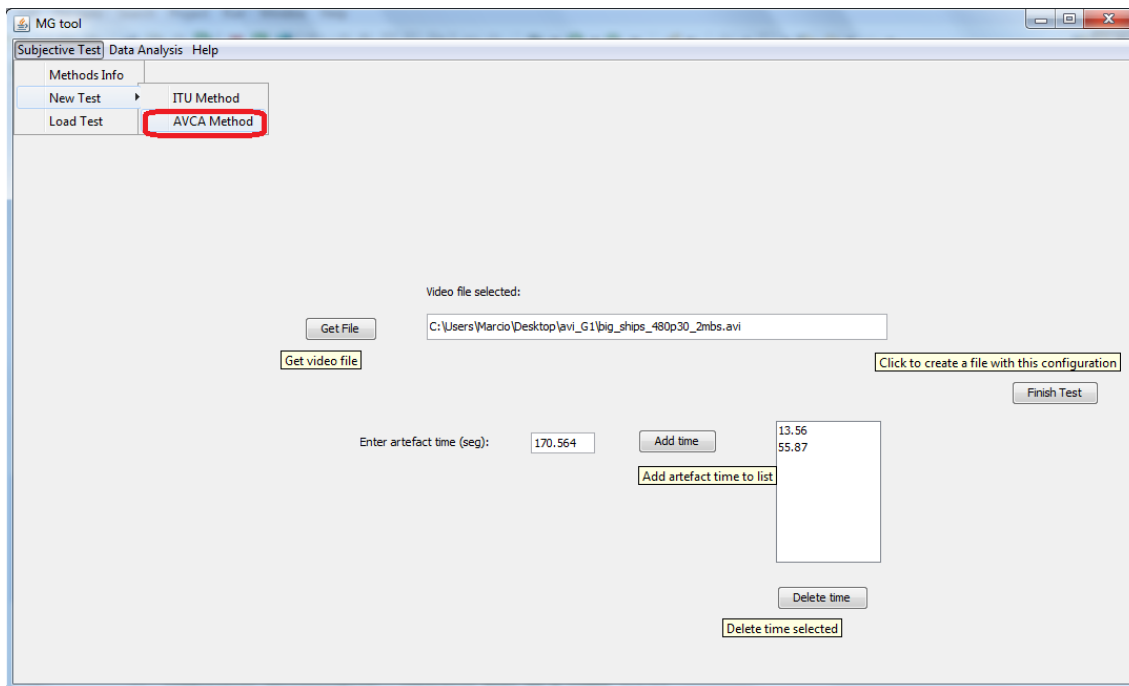


Figura 31: Configuração de um novo teste AVCA

Após terminar a configuração, ao clicar no botão **Finish Test** será criado um ficheiro de configuração do teste. É gerado um ficheiro de texto com o seguinte conteúdo: nome do método, sequência de vídeo, número de tempos de artefactos introduzidos e respetivos tempos, conforme ilustrado na Figura 32. Tal como na configuração dos restantes testes, a palavra 'end' assinala o fim do ficheiro de configuração.

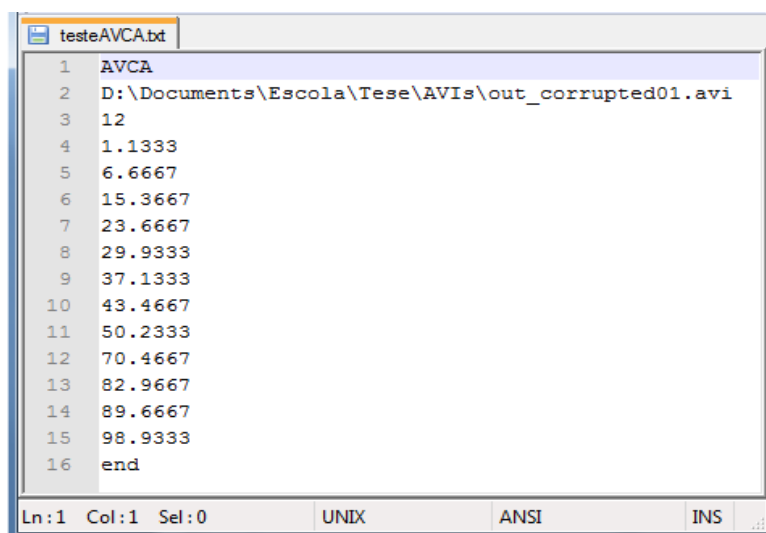
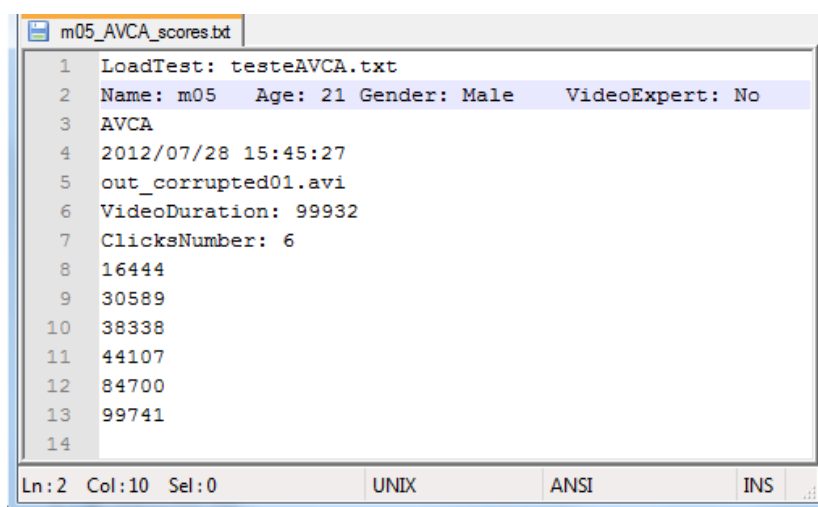


Figura 32: Ficheiro de configuração de um novo teste AVCA

O ficheiro de texto gerado com os resultados de cada observador contém as seguintes informações: o nome do ficheiro de configuração do teste, a informação relativa ao participante (nome, idade, género e tipo de participante), o nome do método de avaliação subjetiva, a data e hora em que se efetuou o teste, o nome da sequência de

vídeo e sua duração, o número total de *clicks* dados pelo participante e respetivos tempos (em milissegundos), conforme ilustrado na Figura 33.



```
m05_AVCA_scores.txt
1 LoadTest: testeAVCA.txt
2 Name: m05 Age: 21 Gender: Male VideoExpert: No
3 AVCA
4 2012/07/28 15:45:27
5 out_corrupted01.avi
6 VideoDuration: 99932
7 ClicksNumber: 6
8 16444
9 30589
10 38338
11 44107
12 84700
13 99741
14
```

Figura 33: Ficheiro de resultados de um participante para um teste com o método AVCA

Para analisar os resultados provenientes de um teste efetuado com este método é necessário escolher a opção **AVCA Results** no menu **Data Analysis**. São ordenados por ordem crescente os tempos de artefactos contidos no ficheiro de configuração e caso haja um tempo superior à duração do vídeo, este é descartado.

Para este método, é interessante analisar a correspondência entre as sinalizações de artefactos dadas pelo participante, e os artefactos de facto presentes no vídeo. Esta correspondência é quantificada através do cálculo das taxas de *verdadeiros positivos*, *falsos negativos* e *falsos positivos*.

Um *verdadeiro positivo* (TP) ocorre quando um participante assinala um artefacto e de facto esse artefacto está presente num instante de tempo próximo do assinalado. Por outro lado, quando um participante não assinala um artefacto existente na sequência vídeo, está-se em presença de um *falso negativo* (FN). Existe também a possibilidade do participante assinalar um artefacto quando de facto não existe nenhum artefacto introduzido próximo desse instante de tempo – esse caso será um *falso positivo* (FP).

Para se poder fazer a associação entre os instantes de tempo assinalados pelo participante e os instantes de tempo que correspondem aos artefactos, é importante definir uma margem de erro temporal que tenha em conta o tempo de resposta do participante, isto é, o tempo que decorre entre o instante em que surge o artefacto e o instante em que o participante o assinala. Para definir a margem de erro temporal foram realizados alguns testes onde se mediu o tempo de resposta dos participantes. Utilizou-se uma amostra com 10 participantes, de ambos os géneros, com idades compreendidas entre os 10 e 50 anos. Foi calculado o tempo de resposta médio, assim como seu desvio padrão. A margem foi definida somando a média ao dobro do desvio padrão, tendo-se

obtido um valor próximo de 1.5 segundos. Esta margem contempla hesitações e falta de experiência em testes de visualização de artefactos por parte do participante.

Uma vez definida a margem para o tempo de resposta, Δ_R , e designando por t_A e t_S os instantes de tempo em que surge o artefacto e em que este é assinalado, respetivamente, o número de *verdadeiros positivos*, $\#TP$, é contabilizado verificando, por cada artefacto introduzido, se os tempos mencionados satisfazem a condição:

$$t_S \leq t_A + \Delta_R. \quad (3.1)$$

Uma vez contabilizados os *verdadeiros positivos*, o número de *falsos positivos* ($\#FP$) é dado por:

$$\#FP = N_S - \#TP, \quad (3.2)$$

onde N_S designa o número de vezes em que o participante assinalou a existência de um artefacto que poderia ou não estar presente (o número total de sinalizações).

Finalmente, o número de falsos negativos ($\#FN$) é dado por:

$$\#FN = N_A - \#TP, \quad (3.3)$$

em que N_A é o número total de artefactos de facto existentes na sequência vídeo.

As taxas de *verdadeiros positivos*, *falsos negativos* e *falsos positivos* são dadas por:

$$\%TP = \frac{\#TP}{N_S} \times 100; \quad \%FP = \frac{\#FP}{N_S} \times 100; \quad \%FN = \frac{\#FN}{N_A} \times 100; \quad (3.4)$$

A título de exemplo, na Figura 34 apresenta-se uma tabela de resultados relativa a um teste realizado sobre uma sequência de vídeo contendo 12 artefactos. Esta sequência de vídeo consistiu numa compilação de 10 cenas de vídeo com conteúdo espaço-temporal diverso, e com uma duração total de 1:40 minutos.

out_corrupted01.avi	m01	m02	m03	m04	m05	m06	m07	m08	m09	m10
Nº clicks	6.0	4.0	6.0	8.0	6.0	5.0	5.0	3.0	5.0	2.0
% TP	66.67	75.0	100.0	75.0	66.67	80.0	60.0	66.67	80.0	50.0
% FP	33.33	25.0	0.0	25.0	33.33	20.0	40.0	33.33	20.0	50.0
% FN	50.0	66.67	50.0	33.33	50.0	58.33	58.33	75.0	58.33	83.33

Figura 34: Tabela de resultados com o método AVCA

Durante a visualização do vídeo, é importante que não existam atrasos devido a falta de capacidade de processamento do computador. Se a visualização do vídeo não decorrer de uma forma natural e fluida, o participante pode assinalar erradamente artefactos de

visualização e não artefactos introduzidos para efeitos de teste. Numa situação deste género será de esperar um aumento da taxa de falsos positivos.

4.2 Correlation Analysis

De forma a complementar o método de deteção de resultados incoerentes apresentado na secção 2.3, foi implementado um método adicional para o mesmo efeito. Neste novo método a deteção de resultados incoerentes é feita com base nos valores do coeficiente de correlação de Pearson⁽⁶⁾ [9], entre as pontuações dadas por um observador j e a média de todos os observadores. Matematicamente, esta correlação, $\rho(j)$, pode ser expressa por:

$$\rho(j) = \frac{\sum_{i=1}^N (x(i,j) - \mu_x(j)) \times (y(i) - \mu_y)}{\sqrt{\sum_{i=1}^N (x(i,j) - \mu_x(j))^2} \times \sqrt{\sum_{i=1}^N (y(i) - \mu_y)^2}}, \quad (3.5)$$

onde N é o número de sequências de vídeo utilizadas durante os testes subjetivos, $x(i, j)$ é a pontuação dada pelo participante j à sequência de vídeo i e $\mu_x(j)$ representa a média das pontuações do participante j . O valor médio das pontuações dos restantes participantes para cada sequência de vídeo é representado por $y(i)$, e finalmente μ_y representa a média das pontuações dadas pelos restantes observadores a todas as sequencias de vídeo.

De cada vez que se rejeita um participante, efetua-se novamente o cálculo dos coeficientes de correlação para todos os participantes. Deste modo garanta-se que as pontuações incoerentes dos participantes rejeitados não influenciam o cálculo final.

Para visualização de resultados através deste método (*Correlation Analysis*) é necessário aceder ao submenu *Compute MOS* no menu *Data Analysis*. Depois de carregados e verificados os ficheiros de pontuações dos participantes é escolhido um valor de limiar para o coeficiente de correlação. Embora seja sugerido um valor aceitável para este valor, 0.9, o utilizador pode definir um outro valor. Os participantes com coeficientes de correlação inferiores ao valor limiar definido pelo utilizador serão considerados incoerentes e consequentemente rejeitados.

Os resultados são apresentados numa tabela, com algumas diferenças em relação ao método recomendado pela ITU. A primeira coluna contém o nome das sequências de vídeo, as colunas seguintes apresentam as pontuações de cada observador. Nas duas últimas colunas são visíveis os valores de *RawMOS* e *MOS* respetivamente. Para o cálculo do *RawMOS* estão abrangidos todos os participantes, para o cálculo dos valores de *MOS* apenas se consideram os participantes válidos. Caso o método de avaliação subjetiva seja o ACR-HR são apresentadas mais duas colunas com identificação das sequências de referência e valor de *DMOS*. Na última linha da tabela é disponibilizado o valor do coeficiente da correlação para cada observador.

⁽⁶⁾ Este coeficiente é definido, em geral, para duas variáveis e mede o grau de associação linear entre elas.

Conforme ilustrado na Figura 35 os observadores com valores de coeficiente de correlação inferiores ao definido são rejeitados, colunas a amarelo. Para este exemplo definiu-se um coeficiente de correlação mínimo de 0.94. Caso um observador atribua a mesma pontuação a todas as seqüências de vídeo definiu-se que as suas pontuações também deveriam ser rejeitadas. E o valor apresentado para o seu coeficiente de correlação será ‘-‘ (indefinido). Os resultados apresentados são arredondados para duas casas decimais.

The screenshot shows the MG tool interface with a Subjective Test Data Analysis window. The main window displays a correlation matrix for 19 video sequences. The columns are labeled u01 through u09, and the rows are labeled with video sequence names. The diagonal elements are all 1.0. The off-diagonal elements represent the correlation coefficients between sequences. A correlation threshold of 0.94 is set, and sequences with correlation coefficients below this threshold are highlighted in yellow. An 'Input' dialog box is open, prompting the user to enter a value for the correlation threshold (between 0 and 1), with '0.94' entered in the text field. The dialog box has 'OK' and 'Cancel' buttons. Below the dialog box are 'Generate Graph' and 'Export CSV' buttons.

Video Sequence	u01	u10	u02	u03	u04	u05	u06	u07	u08	u09	RawMOS	MOS
australia_32.264	2	2	3	2	2	2	2	2	3	1	2.1	2.12
table_256.264	4	4	5	5	5	4	4	4	4	1	4.0	4.25
container_64.264	3	4	4	4	4	3	4	4	5	1	3.6	3.88
football_2048.264	4	5	3	5	5	5	5	5	5	1	4.3	4.88
mobile_128.264	1	2	2	2	2	1	2	2	1	1	1.6	1.62
coastguard_256...	5	4	3	5	4	4	5	4	5	1	4.0	4.5
foreman_128.264	3	3	3	3	4	2	3	3	3	1	2.8	3.0
stephan_1024.264	5	5	3	5	5	5	5	5	5	1	4.4	5.0
container_512.264	5	5	5	5	5	5	5	4	5	1	4.5	4.88
australia_256.264	5	5	5	5	5	5	5	4	5	1	4.5	4.88
table_128.264	3	3	2	4	3	1	3	3	3	1	2.6	2.88
mobile_512.264	5	5	5	5	5	4	5	4	5	1	4.4	4.75
coastguard_64....	2	2	2	2	2	1	2	2	2	1	1.8	1.88
football_512.264	3	4	3	4	3	2	4	3	4	1	3.1	3.38
stephan_128.264	1	1	1	1	1	1	1	1	1	1	1.0	1.0
foreman_256.264	4	4	4	5	4	4	4	3	4	1	3.7	4.0
australia_128.264	5	4	4	5	4	5	4	4	5	1	4.1	4.5
foreman_64.264	1	1	1	1	1	1	1	1	1	1	1.0	1.0
coastguard_128...	3	3	4	4	4	3	4	3	3	1	3.2	3.38
mobile_256.264	4	4	4	4	4	3	4	4	4	1	3.6	3.88
container_256.264	5	5	5	5	5	4	5	5	5	1	4.5	4.88
football_1024.264	4	4	4	4	4	4	4	4	4	1	3.7	4.0
table_512.264	5	5	5	5	5	5	5	5	5	1	4.6	5.0
stephan_512.264	4	5	5	4	5	4	4	4	4	1	4.0	4.25
container_128.264	3	4	4	4	4	4	5	4	4	1	3.7	4.0
stephan_256.264	3	3	3	2	3	2	2	2	2	1	2.3	2.38
mobile_64.264	1	2	2	2	2	1	1	1	1	1	1.4	1.38
coastguard_512...	5	4	4	5	5	4	5	4	5	1	4.2	4.62
table_64.264	1	2	2	2	1	1	1	2	1	1	1.4	1.38
foreman_512.264	5	4	5	5	5	5	5	4	5	1	4.4	4.75
football_256.264	1	3	3	1	2	1	2	2	2	1	1.8	1.75
australia_64.264	3	4	4	4	4	3	3	4	4	1	3.4	3.62
Correlation:	0.96	0.95	0.87	0.96	0.97	0.95	0.97	0.95	0.97	-	3.4	

Figura 35: Tabela de resultados utilizando a metodologia DSIS e método de rejeição da Correlação

5 Realização de testes subjetivos

Ao longo deste capítulo são descritos os testes subjetivos preparados e realizados utilizando a primeira versão da ferramenta, assim com os resultados provenientes destes. As sequências de vídeo usadas nos testes foram codificadas em H.264 [1] e selecionadas de modo a abranger um conteúdo de vídeo diverso. Depois de codificadas, as sequências foram sujeitas a perdas de pacotes devido a erros de transmissão. O método de avaliação subjetiva utilizado nos testes foi o ACR-HR, descrito na secção 2.2.1.

5.1 Sequências de vídeo utilizadas

Na realização dos testes foram utilizadas 8 sequências de vídeo originais com conteúdo diverso, tanto em termos semânticos como em termos de atividade espaço-temporal. As sequências utilizadas têm todas uma resolução espacial de 720 linhas por 480 colunas, e resoluções temporais de 25 e 30 tramas por segundo, consoante a sequência. Todas as sequências têm 10 segundos de duração. As oito famílias/grupos de sequências de vídeo utilizadas, representadas na Figura 36, foram: *big ships*, *old town cross*, *park joy*, *raven*, *snow*, *soccer*, *station*, *sun flower*.



Figura 36: Sequências de referência utilizadas nos testes subjetivos

As sequências originais foram codificadas, utilizando o *software* de referência JM H.264 *encoder* [22], com diferentes débitos binários (1 a 5 Mbps), sendo o débito ajustado a cada sequência de modo a que fosse obtida uma sequência codificada com uma qualidade percetual semelhante à sequência original. O débito binário usado em cada sequência foi escolhido de modo a que o processo de codificação não introduzisse artefactos visíveis nas versões codificadas. Sendo assim, este valor depende da

complexidade da sequência em causa (atividade espaço-temporal). A gama de débitos binários utilizada corresponde a débitos típicos numa rede IPTV.

No processo de codificação foi definido um *Group of Pictures* (GOP) com tamanho de 15 tramas seguindo uma estrutura IBBPBBP... De 30 em 30 tramas (de dois em dois GOPs) foi inserida uma trama de *Instantaneous Decoding Refresh* (IDR), de forma a impedir dependências para tramas anteriores e desse modo limitar a propagação de eventuais erros de transmissão ao longo da sequência de vídeo. Cada pacote *Real-time Transport Protocol* (RTP) transporta a informação codificada referente a um *slice*. Cada *slice* contém um número fixo de 45 macroblocos (uma linha de macroblocos). As sequências codificadas foram utilizadas como sequências de referência durante os testes subjetivos.

Depois de codificadas as sequências foi gerado um padrão de perdas de pacotes simulando uma rede IPTV. Com base no *software* de simulação que consta na recomendação ITU-T G.1050 [23], foram gerados diferentes padrões de erros com taxas de perda de pacotes (PLR) entre 0.1% e 8%, para cenários de rede IPTV com acesso GPON ou DSL. O modelo implementado em [23] permite a geração de padrões erros numa transmissão por pacotes utilizando cenários reais baseados numa rede *Internet Protocol* (IP). As sequências de vídeo corrompidas com perdas foram posteriormente decodificadas para formato *Raw* (YUV) e encapsuladas em ficheiros *Audio Video Interleave* (AVI).

As sequências correspondentes às condições de teste foram divididas em três grupos. Foi feita esta divisão de modo a não sobrecarregar as sessões de teste com demasiadas condições de teste a avaliar, o que poderia fatigar os participantes. Caso os participantes avaliassem todas as condições de teste, cada sessão de avaliação ficaria a durar mais de 30 minutos, o que contrariaria as recomendações. Com a divisão segundo três grupos, cada sessão de testes foi dimensionada para durar cerca de 15 minutos, com cada participante a avaliar a qualidade de 56 sequências de vídeo. De forma a verificar a consistência dos valores de MOS obtidos nos três locais, algumas condições de teste (uma condição de teste por família de vídeos) foram escolhidas para serem avaliadas em todos os locais. As sequências de referência foram também avaliadas em todos os locais.

5.2 Condições dos testes e participantes

As características dos computadores e monitores utilizados durante os testes subjetivos realizados no IT-IUL correspondem às apresentadas na Tabela 7.

Tabela 7: Características dos computadores e monitores usados nos testes

Computadores	Monitores
Sistema Operativo Windows 7 Pro – 64bit	LCD ASUS ML238H com 23”
Processador Intel Core i7-3770 3.4GHz	Resolução 1920X1080 (<i>Full HD</i>)
Memória RAM 16GB DDR3-1600	Tempo de Resposta de 2 milissegundos
Placa gráfica NVidia Geforce GT 630	16.7 Milhões de cores
	Taxa de Contraste (Max.): 10.000.000 :1

Um aspeto extremamente importante na avaliação da qualidade dos vídeos são as condições de visualização existentes. Os testes subjetivos efetuados no IT-IUL decorreram em ambiente controlado e segundo as recomendações em [3], com os valores de iluminação, calibrados segundo [3], apresentados na Tabela 8.

Tabela 8: Condições de visualização dos testes IT-IUL

Parâmetro	Valor
Distância de visualização	$2H^{(7)}$
Pico de luminância do ecrã	132 cd/m^2
Rácio entre a luminância do ecrã inativo e pico de luminância	0.023
Rácio entre a luminância do fundo da imagem no ecrã e o pico de luminância da imagem	0.05
Crominância do fundo/plano	$D_{65}^{(8)}$
Iluminação de fundo da sala	14 lux
Máximo ângulo de observação relativamente à normal	30°

Antes de iniciarem os testes subjetivos todos os participantes foram submetidos a testes de acuidade visual (usando um quadro de *Snellen*) e foi feito o despiste de daltonismo (usando placas *Ishihara*). Nos testes que decorreram no IT-IUL participaram um total de 15 observadores não especialistas, com idades compreendidas entre os 17 e os 30 anos, de ambos os géneros.

⁽⁷⁾ H representa a altura da imagem no ecrã.

⁽⁸⁾ D_{65} é uma forma de iluminação normalizada na qual os vários comprimentos de onda que compõem a luz têm uma distribuição de potência que corresponde aproximadamente à luz solar ao meio-dia.

Antes de se iniciar cada sessão de testes de avaliação subjetiva, efetuou-se um teste para calibração de qualidade das sequências de vídeo. Aos observadores foram apresentadas seis sequências de vídeo, três com qualidade “excelente” (sem artefactos) e outras três com qualidade “má” (segundo a escala na Tabela 2 na secção 2.2), para que estes ficassem com percepção da melhor e pior qualidade que iriam encontrar durante a realização do teste subjetivo. Os participantes foram também informados sobre o género de artefactos que seriam expectáveis de ocorrer ao longo da sessão de testes.

Foi utilizado o método ACR-HR. As principais razões para escolha deste método foram: o facto de num cenário IPTV o utilizador não ter acesso à referência e o facto de este método ser um dos que leva a uma melhor relação entre o número de condições de teste avaliadas e a duração do teste. Durante a realização dos testes foi possível realizar duas sessões em simultâneo.

5.3 Resultados

Nesta secção do documento são apresentados os resultados dos testes efetuados até ao momento, que correspondem ao primeiro grupo de sequências mencionado na secção 5.1. Os valores de MOS relativos às avaliações feitas pelos 15 participantes e os respectivos intervalos de confiança a 95% podem ser observados na Figura 37.

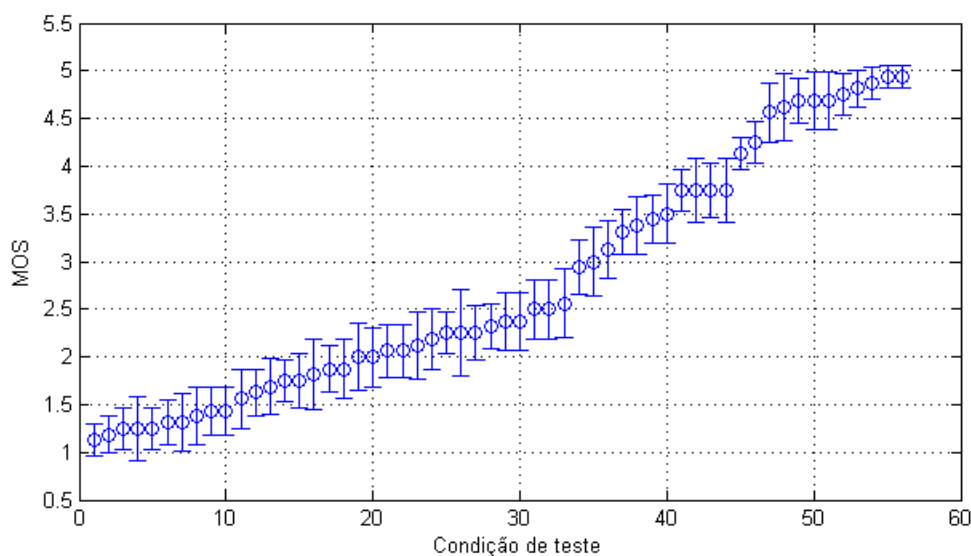


Figura 37: Resultados de MOS dos testes subjetivos efetuados com método ACR-HR

Os resultados de todos os participantes assim como todas as sequências de vídeo usadas durante os testes subjetivos serão brevemente disponibilizados para consulta da comunidade científica.

6 Conclusões e trabalho futuro

6.1 Síntese

O principal objetivo deste trabalho foi implementar uma ferramenta capaz de realizar e preparar testes de avaliação subjetiva da qualidade de vídeo ultrapassando limitações encontradas em ferramentas existentes para o mesmo efeito. Uma das principais funcionalidades implementadas foi a de proporcionar aos utilizadores da ferramenta uma forma simples e rápida de configurar uma sessão de testes completa. Os resultados provenientes das avaliações são disponibilizados em forma de tabela, gráficos e em formato CSV.

A aplicação desenvolvida permite a configuração e realização de quatro formas diferentes de avaliar a qualidade de vídeo. Três dos métodos implementado encontram-se entre os recomendados em [3] e [4], nomeadamente ACR-HR, DSIS e SCACJ. Foi proposto e implementado um outro método (AVCA), de avaliação contínua, que consiste em detetar e assinalar os artefactos visualizados durante a exibição do vídeo. Foram ainda implementados dois modos para deteção de pontuações incoerentes. Um deles foi o apresentado em [4] e, com vista a complementá-lo devido a algumas limitações encontradas, foi desenvolvido um método baseado em correlação.

A ferramenta foi utilizada em cenário de produção durante a realização de testes subjetivos no IT-IUL. Nestes testes participaram 15 observadores de ambos os géneros avaliando 56 sequências de vídeo sujeitas a perdas de transmissão num cenário IPTV. A metodologia aplicada durante os testes foi o ACR-HR, em que o participante atribui uma pontuação de qualidade a cada sequência de vídeo. A realização dos testes permite extrair os valores de MOS que são utilizados no desenvolvimento e análise de desempenho de algoritmos de avaliação objetiva da qualidade de vídeo.

O desenvolvimento da aplicação foi feito usando a linguagem de programação *java* e, para permitir a visualização das sequências de vídeo, utilizou-se a biblioteca *Java Media Framework* (JMF).

6.2 Trabalho futuro

Sendo o objetivo principal deste trabalho o desenvolvimento de uma aplicação de *software*, será de esperar que posteriormente surjam novas ideias com vista a melhorar ou acrescentar novas funcionalidades à aplicação desenvolvida.

De entre as várias possibilidades de implementação que visam enriquecer o *software* desenvolvido no âmbito deste trabalho, sugerem-se as seguintes:

- Implementação de métodos para avaliação subjetiva de imagens fixas – facilmente se poderia implementar, visto que a preparação e configuração dos testes seria muito semelhante.

- Implementação de novos métodos de avaliação subjetiva de vídeo – para a ferramenta ficar mais completa e abrangente, por exemplo novos métodos de avaliação contínua.
- Integração com um codificador de sequências de vídeo – visto ser necessário um programa que codifique as sequências, tendo como base uma sequência original (referencia), para que seja possível fazer as avaliações subjetivas da qualidade de vídeo. Ao ser implementado nesta ferramenta não era necessário recorrer a um programa externo para gerar sequências codificadas. Utilizava-se a mesma ferramenta para efetuar todas estas operações.
- Utilizar uma outra biblioteca para implementar os procedimentos envolvidos na visualização de vídeo, que seja mais atual do que a biblioteca JMF. Esta deverá permitir ler formatos de vídeo mais recentes e ter suporte para máquinas com sistema operativo de 64bits.
- Fazer os testes de acuidade visual e despiste de daltonismo a partir da própria ferramenta.

Após a finalização do plano de trabalhos (e concluída a dissertação), a ferramenta, respetiva documentação, resultados de avaliações subjetivas e sequências de vídeo feitas com a ferramenta serão disponibilizados e disseminados pela comunidade científica.

Referências

- [1] ITU-T, Recommendation H.264, “Advanced video coding for generic audiovisual services”, 2003–2012.
- [2] ITU-T, Recommendation H.262, “Information technology - Generic coding of moving pictures and associated audio information: Video”, 1995–2012.
- [3] ITU-T, Recommendation P.910, “Subjective video quality assessment methods for multimedia applications”, 2008.
- [4] ITU-R, Recommendation BT.500-13, “Methodology for the subjective assessment of the quality of television pictures”, 2012.
- [5] S. Kanumuri, P. C. Cosman, A. R. Reibman, V. A. Vaishampayan, “Modeling packet-loss visibility in MPEG-2 video”, *IEEE Transactions on Multimedia*, Vol. 8, n. 2, pp. 341-355, 2006.
- [6] T. L. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. Cosman, A. R. Reibman, “A versatile model for packet loss visibility and its application in packet prioritization” *IEEE Transactions on Image Processing*, Vol. 19, n. 3, pp. 722-735, 2010.
- [7] M. Nezveda, S. Buchinger, W. Robitza, E. Hotop, P. Hummelbrunner, H. Hlavacs, “Test persons for subjective video quality testing: experts or non-experts?”, *Proc. European Conference on Interactive TV (EuroITV)*, Tampere, Finlândia, 2010.
- [8] F.Kozamernik, P. Sunna, E. Wyckens, D. Pettersen, “Subjective quality of internet Videocodecs – Phase 2 evaluations using SAMVIQ”, *EBU Technical Review*, European Broadcast Union (EBU), 2005.
- [9] *Applied Statistics and Probability for Engineers*, 5th Edition, Douglas C. Montgomery e George C. Runger, Wiley, 2010.
- [10] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, T. Ebrahimi, “Subjective assessment of H264/AVC video sequences transmitted over a noisy channel”, *Proc. International Workshop on Quality of Multimedia Experience (QoMEX)*, San Diego, California, E.U.A., 2009.
- [11] F. Simone, M. Tagliasacchi, M. Naccari, S. Tubaro, T. Ebrahimi, “H.264/AVC video database for the evaluation of quality metrics”, 35th International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, E.U.A., 2010.
- [12] T. Brandão, M.P. Queluz, “No-reference perceptual quality metric for H.264/AVC encoded video”, *Proc. International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Scottsdale, E.U.A., 2010.

- [13] T. Brandão e M.P. Queluz,; “No-reference quality assessment of H.264/AVC encoded video”, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 20, No. 11, pp. 1437 - 1447, 2010.
- [14] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, “Study of Subjective and Objective Quality Assessment of Video”, *IEEE Transactions on Image Processing*, Vol.19, n. 6, pp. 1427-1441, 2010.
- [15] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, “A Subjective Study to Evaluate Video Quality Assessment Algorithms”, *Proc. SPIE Human Vision and Electronic Imaging*, S. Jose, E.U.A., 2010.
- [16] P. Le Callet e F. Autrusseau, “Subjective quality assessment IRCCYN/IVC database”, Relatório Técnico disponível online em: <http://www.irccyn.ec-nantes.fr/ivcdb> , Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCyN), 2005.
- [17] Stéphane Péchard, Romuald Pépion, Patrick Le Callet, "Suitable methodology in subjective video quality assessment: a resolution dependent paradigm", *International Workshop on Image Media Quality and its Applications*, Kyoto, Japão, 2008.
- [18] MSU Perceptual Video Quality tool, *Lomonosov Moscow State University – Graphics and Media Laboratory*, disponível online em http://compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html.
- [19] ACR (Absolute Category Rating) Test Interface, *AcceptTV*, disponível online http://www.acceptv.com/page/products_acr_test_interface.
- [20] Java Media Framework API, *Oracle*, disponível online em <http://www.oracle.com/technetwork/java/javase/tech/index-jsp-140239.html>.
- [21] Projeto Open Source Computer Vision (Opencv), disponível online em <http://opencv.willowgarage.com/wiki/>.
- [22] Software de referência JM H.264 encoder, *HHI*, disponível online em <http://iphome.hhi.de/suehring/tml/>.
- [23] ITU-T, Recommendation G.1050, “Network model for evaluating multimedia transmission performance over the Internet Protocol”, 2011.