Perceptual Multimedia Quality:
Implications of an Empirical Study

S.R. Gulliver *G. Ghinea*

School of Information System, Computing and Mathematics.
Brunel University, United Kingdom.
{Stephen.Gulliver, George.Ghinea}@brunel.ac.uk

Abstract

If commercial multimedia development continues to ignore the user-perspective in preference of other factors, i.e. user fascination (i.e. the latest gimmick), then companies ultimately risk alienating the customer. Moreover, by ignoring the user-perspective, future distributed multimedia systems risk ignoring accessibility issues, by excluding access for users with abnormal perceptual requirements.

This paper presents an extensive examination of distributed multimedia quality. We define a model that considers multimedia quality from three distinct levels: the *network*, the *media*- and the *content-levels*; and two views: the *technical*- and the *user-perspective*. By manipulating both technical and user-perspective parameters, we examine the impact on quality perception at the three quality levels identified. Results show that: a significant reduction in frame rate does not proportionally reduce the user's understanding of the presentation, independent of technical parameters; the type of video clip significantly impacts user information assimilation, user level of enjoyment and user perception of quality; the display type impacts user information assimilation and user perception of quality. Finally, to ensure transfer of informational content, network parameter variation should be adapted; to maintain user enjoyment, video content variation should be adapted.

1. Introduction

Distributed multimedia quality, in our perspective, is deemed as having two main facets: *of perception* and *of service*. The former, Quality of Perception, considers the user perspective, measuring the infotainment impact of the presentation. The latter facet, Quality of Service, characterises the technical perspective and represents the performance properties provided by multimedia technology. To realize a truly extensive examination of distributed multimedia quality we propose a quality model, which integrates user and technical considerations at three abstraction levels and then examines the quality implications of parameter variations at each of the respective levels.

The structure of this document is as follows: In section 2 we introduce the reader to our model of distributed multimedia and define the research domain. In section 3, we consider the methodology, the video material and devices that was used in our work, whilst in section 4 we describe the process used in our experiments at the network-, media- and content- levels. Research findings are presented in section 5, and conclusions are drawn in section 6.

2. Modelling Distributed Multimedia Quality

Distributed multimedia quality is a multi-faceted concept that involves the integration of quality parameters at different levels of abstraction and from different perspectives. Integrating these considerations is non-trivial, but is necessary if an extensive picture of distributed multimedia quality is to be elaborated. To this end, we propose an extended version of a model initially suggested by Wikstrand (2003), in which quality is segregated into three discrete levels: the *network-level*, the *media-level* and *content-level* (Figure 1). The network-level concerns the transfer of data and all quality issues related to the flow of data around the network. The media-level concerns quality issues relating to the transference methods used to convert network data to perceptible media information, i.e. the video and audio media. The content-level concerns quality factors that influence how media information is perceived and understood by the end user.

In our work, and in addition to the model proposed by Wikstrand, we incorporated two distinct quality perspectives, which reflect the infotainment duality of multimedia: the user-perspective and the technical-perspective.

- ***User-Perspective*:** The user-perspective concerns quality issues that rely on user feedback or interaction. This can be varied and measured at the media- and content-levels. The network-level does not facilitate the user-perspective since user perception cannot be measured at this low level abstraction (see Figure 1).
- ***Technical-Perspective*:** The technical-perspective concerns quality issues that relate to the technological factors involved in distributed multimedia. Technical parameters can be varied and measured at all quality abstractions.

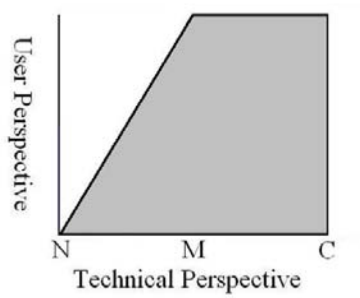

Figure 1: Quality Model, incorporates Network- (N), Media- (M) and Content-Level (C) abstractions and Technical- and User-Perspectives dimensions

At each quality abstraction defined in the quality model, parameters can be varied as demonstrated by the literature. Previous studies show that at the:

- **Network-Level:** Technical-perspective network-level variation of bit error, segment loss, delay and jitter (Procter, Hartswood, McKinlay & Gallacher, 1999) has been used to simulate QoS deterioration. Technical-perspective network-level measurements of loss (Koodli & Krishna, 1998), delay and jitter (Claypool & Tanner, 1999; Wang, Claypool, & Zuo, 2001), as well as allocated bandwidth (Wang et al., 2001) have all been used to measure network level quality performance.
- **Media-Level:** Technical-perspective media-level variation of video and audio frame rate (Apteker, Fisher, Kisimov & Neishlos, 1995; Ghinea & Thomas, 1998; Masry, Hemami, Osberger & Rohaly, 2001; Wilson & Sasse, 2000b), captions (Gulliver & Ghinea, 2003), animation method (Wikstrand & Eriksson, 2002), inter-stream audio-video quality (Hollier & Voelcker, 1997), image resolution (Kies, Williges & Rosson, 1997), media stream skews (Steinmetz, 1996), synchronisation (Steinmetz, 1996) and video compression codecs (Masry et al., 2001) have been used to vary quality definition. Technical-perspective media-level measurement is generally based on linear and visual quality models (Teo & Heeger, 1994; van den Branden Lambrecht & Verscheure, 1996), with the exception of Wang et al. (2001) who use output frame rate as the quality criterion. User-perspective media-level variation requires user data feedback and is limited to attentive displays, which manipulate video quality around a user's point of gaze. User-perspective media-level measurement of quality has been used when measuring user 'watchability' (Apteker et al., 1995), assessing user rating of video quality (Ghinea & Thomas, 1998; Wijesekera, Srivastava, Nerode & Foresti, 1999), comparing streamed video against the non-degraded original video (Procter et al, 1999), as well as for continuous quality assessment (Wilson & Sasse, 2000a, 2000b) and gauging participant annoyance of synchronisation skews (Steinmetz, 1996).
- **Content-Level:** Technical-perspective content-level variation has been used to vary the content of experimental material (Ghinea & Thomas, 1998; Gulliver & Ghinea, 2003; Masry et al., 2001; Procter et al., 1999; Steinmetz, 1996), as well as the presentation language (Steinmetz, 1996). Technical-perspective content-level measurement has, to date, only included stress analysis ((Wilson & Sasse, 2000a, 2000b). User-perspective content-level variation has also been used to measure the impact of user demographics (Gulliver & Ghinea, 2003), as well as volume and type of microphone (Watson, & Sasse, 2000) on overall perception of multimedia quality. User-perspective content-level measurement has measured 'watchability' (Apteker et al., 1995), 'ease of understanding', 'recall', 'level of interest', 'level of comprehension'

(Procter et al., 1999), information assimilation (Ghinea & Thomas, 1998; Gulliver & Ghinea, 2003), predicted level of information assimilation (Gulliver & Ghinea, 2003) and enjoyment (Gulliver & Ghinea, 2003, Wijeskera et al, 1999).

To extensively consider distributed multimedia quality effectively it is essential that, where possible, both technical- and user-perspective parameter variations be made at all quality abstractions. Moreover, in order to effectively measure the infotainment duality of multimedia (information transfer and level of satisfaction), the user perspective must consider both the user's ability to assimilate and understand the informational content of the video, as well as the user's subjective satisfaction. None of the referenced studies achieved this set of criteria and it is on this that this paper shall focus its attention.

3. Experimental Design

Our research aims to extensively consider the user's perception of multimedia quality, by varying relevant technical- and user-perspective parameters at the three quality abstractions of our model. Due to the reduced bandwidth requirement and increased perceptual impact of corrupted audio, the audio stream was not manipulated in our research.

- **Objective 1: Measurement of the perceptual impact of network level parameter variation.** To consider network level technical parameter variation we measured the impact of delay and jitter on the users' perception of multimedia quality. Although other authors have considered the perceptual impact of delay and jitter (Procter et al., 1999), previous studies fail to consider both level of user understanding (information assimilation) and user satisfaction (both of the video quality of service and concerning the content of the video).

- **Objective 2: Measurement of the perceptual impact of media level parameter variation.** Attentive displays monitor and/or predict user gaze, in order to manipulate allocation of bandwidth, such that quality is improved around the point of gaze (Barnett, 1996). Attentive displays offer considerable potential for the reduction of network resources and facilitate media level quality variation with respect to both video content-based (technical-perspective) and user-based (user-perspective) data. In order to measure media level parameter variation, in respect of both technical- and user-perspectives, we measured the impact of a novel Region of Interest attentive display system, which was developed to produce both video content data and eye tracker data-dependent output video.

- **Objective 3: Measurement of the perceptual impact of content level parameter variation.** To consider user-perspective content level parameter variation, we measured the impact of various display types on user perception of multimedia quality. Technical-perspective content level parameter variation was achieved through use of diverse experimental video material.

In accordance with our quality model, and in order to explore the human side of the multimedia experience, three structured QoP experiments were used to achieve the defined objectives.

3.1 Quality of Perception: Experimental Methodology

Ghinea and Thomas (1998) first used Quality of Perception (QoP) to measure level of information assimilation and satisfaction, where multimedia video clips were shown at varied frame rates. Quality of perception is based on the idea that the technical-perspective alone is incapable of defining the perceived quality of multimedia video, especially at the content-level (Bouch, Wilson, & Sasse, 2001; Ghinea & Thomas, 1998; Watson & Sasse, 2000). Quality of Perception uses level of 'information transfer' (QoP-IA) and user 'satisfaction' (QOP-S) to determine the perceived level of multimedia quality. To this end, quality of perception is a term used in our work to encompass not only a user's satisfaction with the quality of multimedia presentations ('Satisfaction' - S), but also his/her understanding, that is an ability to analyse, synthesise and assimilate the informational content of multimedia content ('Information Assimilation' – IA). In our study QoP-S is subjective in nature and consists of two component parts: QoP–LoQ (the user's judgement concerning the objective Level of Quality assigned to the multimedia content being visualised) and QoP–LoE (the user's Level of Enjoyment whilst viewing multimedia content), targeting objective perceptual quality at both media- and content-levels respectively. QoP-S therefore considers the user-perspective at both abstractions defined in our model.

3.1.1. Measuring Information Assimilation (QoP-IA)

QoP-IA allows us to measure a user's ability to assimilate the content of the video clip (content-level). Thus, after watching a particular multimedia clip, the user is asked a number of questions that examined the information being assimilated from certain information sources. QoP-IA is expressed as a percentage representing the proportion of correctly answered questions. For each feedback question, the source of the answer was determined as having originated from one or more of the following information sources: V (Video-based information that comes from the video window), A (Audio-based information that is presented in the audio stream), T (Textual-based information that is contained in the video window). Since QoP-IA is calculated as the percentage of correctly assimilated information, all QoP-IA questions are designed so that specific information must be assimilated in order to correctly answer each question. Although the majority of questions can trace their answer to a single information source, a number of specific questions do however relate to multiple information sources. The following example (pop video clip) shows how questions were used to test the user's assimilation of V, A and T information sources (the source of data is contained in brackets and the answer is underlined):

- What was the bald man doing in the video? (V) <u>Moving a chair / furniture.</u>
- Name two features of the clip that relate to the Orient? (V) <u>She is wearing a t-shirt that has a dragon logo</u>, (T) <u>She performed in a Japanese video commercial</u>
- According to the lyrics of the song, is the male character on time? (A) <u>He is late.</u>
- As all QoP-IA questions have unambiguous answers it is possible to calculate the percentage of correctly assimilated information, facilitating examination of user information assimilation, as a result of quality parameter variation.

3.1.2. Measuring Subjective Perception (QoP-S)

In our study subjective perception consists of two component parts: User perceived level of quality and user level of enjoyment. Measuring user perceived level of quality (the user's subjective Level of Quality) ensures that user satisfaction includes measurement at the media-level. Measuring the user's level of enjoyment ensures that user satisfaction includes measurement at the content-level. In order to measure user perceived level of quality and level of enjoyment, users were asked to respectively indicate, how they judged the presentation quality (independent of the subject matter) and how much they enjoyed the multimedia presentation. A five point scale was used, with scores of 0 and 5 representing "no" and, respectively, "absolute" user satisfaction. Accordingly, subjective perception incorporates the media-and content-level user-perspective of our model.

3.2 Experimental Material

The set of video clips used in our experiments consists of a series of 12 windowed MPEG video clips. The multimedia video clips, of duration between 26 and 45 seconds, were specifically chosen to cover a broad spectrum of infotainment. Moreover, the clips were chosen to present the majority of individuals with no peak in personal interest, whilst limiting the number of individuals watching the clip with previous knowledge and experience. The multimedia video clips used in our experiments varied from those that are informational in nature (such as a news or weather broadcast) to ones that are usually viewed purely for entertainment purposes (such as an action sequence, a cartoon or a sports event). Specific clips, such as the cooking clip, were chosen as a mixture of the two viewing goals.

3.3 Experimental Devices

To achieve our objectives, we used five different devices in our experiments to measure the perceptual impact of network-, media- and content-level factor variation on a user's ability to assimilate information from multimedia video. Devices were chosen to simulate a varying level of mobility. Each participant that was used in our experiments had never participated in a quality of perception experiment before, thus minimising the existence of participant pre-knowledge. Participants used in our experiments were taken from a range of different nationalities and backgrounds, however all participants spoke English as their first language, or to a degree-level qualification, and were computer literate. Accordingly:

- 126 participants (Objectives 1 and 3) viewed the video clips using a standard computer monitor, whilst simultaneously interacting with a Mac Arrington ViewPoint EyeTracker, used in combination with QuickClamp Hardware, which limits head mobility. An eye-tracker was used as the eye naturally selects / fixates on areas that are likely to be most informative (Kaufman & Richards, 1969). Eye tracking data was collected in order to aid adaptation of multimedia in specific regions of interest, assuming spontaneous looking.
- 54 participants (Objective 2) viewed the video clips using a standard computer laptop monitor.
- to act as a control, 18 participants (Objective 3) viewed video clips using a normal 15 inch SVGA generic computer monitor enabled with a Matrox Rainbow Runner Video Card.
- to consider greater autonomy of movement, than is available using a generic computer monitor, 18 participants viewed the multimedia video clips using an Olympus Eye-Trek FMD 200 Head-Mounted Display (HMD), which uses two head-mounted liquid crystal displays. Each one of the displays contains 180,000 pixels and the viewing angle is 30.0° horizontal, 27.0° vertical. It supports PAL (Phase Alternating Line) format and has a display weight of 85g.
- to consider full mobility, 18 participants (Objective 3) used a Hewlett-Packard iPAQ 5450 personal digital assistant with 16-bit touch sensitive TFT liquid crystal display that supports 65,536 colours. The display pixel pitch of the device is 0.24 mm and its viewable image size is 2.26 inch wide and 3.02 inch tall. The PDA ran the Microsoft Windows for Pocket personal computer 2002 operating system on an Intel 400 Mhz XSCALE processor and allows the user complete mobility. By default, it contains 64MB standard memory (RAM) and 48MB internal flash read-only memory (ROM). In order to complete this experiment a 128 MB secure digital memory card was used for multimedia video storage purposes.

4. Experimental Process

To avoid audio and visual distraction, a dedicated, uncluttered room was used throughout all experiments. All participants were asked a number of short questions concerning their sight. Also a basic eye-test was undertaken to ensure that user were able to view menu text on the screen. This was especially important for those using the eye-tracking device, as participants were not able to wear corrective spectacles for the duration of the experiment. Participants were informed that after each video clip they would be required to stop and answer a number of questions that related to the video clip that had just been presented to them. To ensure that participants did not feel that their intelligence was being tested it was clearly explained that they should not be concerned if they were unable to answer any of the quality of perception information assimilation questions. In the case of the participants using the eye-tracker, time was taken to adjust the chin-rest, infrared red capture camera and software settings to ensure that pupil fix was maintained throughout the user's entire visual field. When appropriate calibration was complete, the participant was asked to get into a comfortable position and, in the case of the eye-tracker, place his/her chin on the chin-rest.

After showing each clip, the video window was closed and the participant was asked a number of questions. The participant was asked all questions aurally and the answers to all questions were noted at the time of asking. Once a user had answered all questions relating to a specific video clip, and all responses had been noted, participants were presented with the next video clip. This was done for all videos, independent of the display device.

4.1 Objective 1: Network-Level Parameter Variation (Delay and Jitter)

Three experimental variables were manipulated in this study: network-level error type (control, delay and jitter), multimedia video frame rate and multimedia content. Accordingly, original, delay and jitter video conditions were considered in our experiment, and three multimedia video frame rates: 5, 15 and 25 frames per second (fps). 108 participants were evenly divided into three groups, which related to the perceptual impact of control, jitter and delay videos respectively. Participants in each group (36 participants in total) were subdivided into three groups, each containing 12 participants. Sub-groups were used to distinguish the specific participants viewing order and frame rate. In each experimental sub-group, a within-subjects design was used, where participants viewed each of the 12 video clips in turn at one of three pre-recorded frame rates (5, 15 or 25 fps). Thus, each participant viewed four video clips at 5 fps, four at 15 fps, and four at 25 fps.

4.2 Objective 2: Media-Level Parameter Variation (RoI Display)

When viewing a multimedia presentation, a user can only be attending to a relatively small part of the video display at any one point in time (Kaufman & Richards, 1969). Accordingly, by shifting allocation of bandwidth from peripheral regions of the screen to *Regions-of-Interest* (RoI), attentive displays can be produced. Two main approaches have been developed to implement adaptive attentive displays: *Gaze-contingent Display* (GCD) and *Region of Interest Display* (RoID) systems. Attentive gaze-contingent displays select the region of interest by actively tracking the viewer's eyes and in real time adapting video with a high level of detail at the point of gaze. On the other hand, Region of Interest Display use region of interest coordinates, determined from either previously obtained eye-tracking data or from analysis of video content characteristics (e.g. colour, movement, edges) and adapts the displayed video quality such that resource allocation is biased to region of interest areas.

To create effective region of interest displays, we produced multimedia videos that had an adaptive non-uniform distribution of resource allocation. To achieve this we used eye tracker- and video content-dependent data, which facilitated the variation of frame rate in particular regions of the screen. Whilst eye tracker-dependent data related the location of participant gaze during the original control experiment, video content-dependent data related to significantly important visual primitives, i.e. edges, colour distribution, contrast and movement. Thus region of interest areas, herewith referred to as foreground areas, were refreshed at a relatively higher frame rate than that of the non-region of interest areas (background areas) – Figure 2. Our study considered nine *video quality variations* as part of our experiment: control videos at 5 fps (C5), 15fps (C15), and 25fps (C25); eye data based region of interest display video with 5 / 15 (E5_15), 5 / 25 (E5_25) and 15 / 25 (E15_25) frames per second background / foreground combinations; and video content-dependent region of interest display video with 5 / 15 (V5_15), 5 / 25 (V5_25) and 15 / 25 (V15_25) frames per second background / foreground combinations.

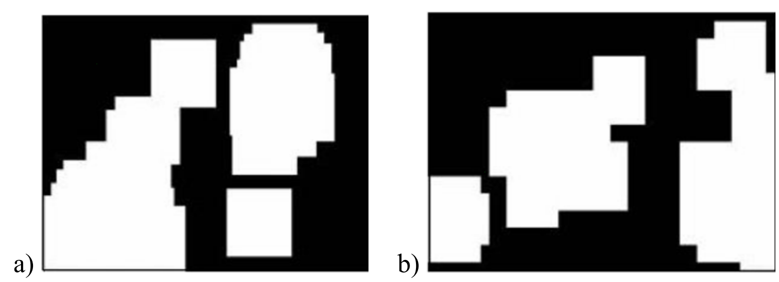

Figure 2: a) Eye-based RoI areas, b) Content-dependent RoI areas - (white areas represent foreground)

Three experimental variables were manipulated in this experiment: region of interest display presentation technique (i.e. control, eye-tracker based and video content-dependent data), multimedia video frame rate combinations, and multimedia content. Consequently, both eye- and content-based region of interest display video were considered in our experiments. To ensure experimental consistency a within-subjects design was used to ensure that participants view all nine video quality variation types across the 12 videos. 54 participants were evenly divided into nine experimental groups.

4.3 Content-Level Parameter Variation: Display Type

Three experimental variables were manipulated in this experiment: type of device, multimedia video frame rate and multimedia content (video clip type). To allow the perceptual comparison of different display equipment on a user's ability to assimilate information from multimedia video, 72 participants were evenly allocated to four different experimental groups. Within each group, users were presented the video clips using certain display equipment. For more detail concerning the experimental display equipment, see section 3.3. Group 1 acted as a control group (standard mobility) and was therefore shown the video clips using a generic computer monitor. Group 2 viewed the video using a computer monitor, however, the participants were simultaneously interacting with a Eye-tracker that provides limited head mobility. Group 3 viewed the multimedia video clips using an Olympus Eye-Trek FMD 200 head-mounted display. Group 4 viewed the video clips using a Hewlett-Packard iPAQ 5450 personal digital assistant.

In addition to different display devices, participants viewed video clips using one of three configurations. Thus, each participant viewed four video clips at 5 frames per second, four video clips at 15 frames per second, and four video clips at 25 frames per second. To ensure technical-perspective content-level quality parameter variation and experiment consistency we used the same video clips, as employed in the previous experiments.

5. Research Findings

Quality of Perception (QoP) was used in our study to extensively characterise the user's perception of multimedia quality at the three levels of our multimedia quality model. This involved three experiments which measured QoP-IA (the user's ability to assimilate information) and user QoP-S (the user's satisfaction), as a result of relevant technical- and user-perspective parameter variation, made at the network-level (technical-perspective), the media-level (both technical- and user-perspectives), and the content-level (both technical- and user-perspectives), respectively. In addition to abstraction-level quality parameter variation, we also measured the impact of video frame rate and video clip type at each level of our quality model. The findings of our work (Table 1) highlights a number of important issues relating to the effective provision of user-centric quality multimedia that will now be discussed.

Table 1: A summary of our QoP finding. (✗ - no significant difference; ✓- significant difference)

<table>
<tr><th></th><th></th><th>QoP-IA</th><th>QoP-LoQ</th><th>QoP-LoE</th></tr>
<tr><td rowspan="4">Network Level</td><td>Delay</td><td>✗</td><td rowspan="2">✓
F(1,2) = 8.547
$p<0.001$
Jitter p=0.001
Delay p=0.002</td><td rowspan="2">✓
F(1,2) = 3.954
p=0.019
Jitter p=0.037
Delay p=0.019</td></tr>
<tr><td>Jitter</td><td>✗</td></tr>
<tr><td>Video Variation Type (Frame Rate)</td><td>✗</td><td>✓
F(1,8) = 7.706
$p<0.001$</td><td>✓
F(1,8) = 2.221
p=0.024</td></tr>
<tr><td>Video Clip</td><td>✓
F(1,11) = 12.700
$p<0.001$</td><td>✓
F(1,11) = 7.085
$p<0.001$</td><td>✓
F(1,11) =8.322,
$p<0.001$</td></tr>
<tr><td rowspan="3">Media Level</td><td>Attentive Display</td><td rowspan="2">✗</td><td rowspan="2">✓
F(1,8) = 19.462
$p<0.001$</td><td rowspan="2">✗</td></tr>
<tr><td>Frame Rate</td></tr>
<tr><td>Video Clip</td><td>✓
F(1,11) = 8.696
$p<0.001$</td><td>✓
F(1,11) = 6.772
$p<0.001$</td><td>✓
F(1,11) = 10.317
$p<0.001$</td></tr>
<tr><td rowspan="3">Content Level</td><td>Device Type</td><td>✓
F(1,3) = 3.048,
p=0.028</td><td>✓
χ^2(3, N = 576) =
11.578,
p= .009</td><td>✗</td></tr>
<tr><td>Frame Rate</td><td>✗</td><td>✓
F(1,2) = 4.766,
p=0.009</td><td>✗</td></tr>
<tr><td>Video Clip</td><td>✓
F(1,11) = 10.769
$p<0.001$</td><td>✗</td><td>✓
F(1,11) = 9.676,
$p<0.005$</td></tr>
</table>

A significant loss of frames (that is, a reduction in frame rate) does not proportionally reduce the user's understanding of the presentation (Table 1). This finding supports the conclusions of Ghinea and Thomas (1998) and justifies the reduction in bandwidth allocation, if and only-if user information assimilation is the primary aim of the multimedia presentation. However, the use of frame rates below 15 fps was found to significant impact user Perceived level of quality (Figure 3a and Table 1). This supports the work of Wijesekera et al. (1999), who showed that video frame-rate should be maintained above 12 fps if user perception of multimedia quality is to be maintained. Interestingly, this result also raises considerable concerns regarding the usability of frame rate-based attention display systems, since our findings show no positive bandwidth benefits were associated to such displays.

Video clip type impacts user information assimilation, which means that variation in user information assimilation varies across the range of experimental video material, independent of technical video quality. Additionally, variation in user level of enjoyment shows that certain videos were perceived as being overall more enjoyable. This finding is of interest, especially in the fields of advertising and education, as it implies that the type of video is more important to the users' level of enjoyment than quality parameter variation, e.g. varied device type. Indeed, further work is required to fully understand the relationship between video content and user enjoyment, yet this aim lies outside the scope of our study.

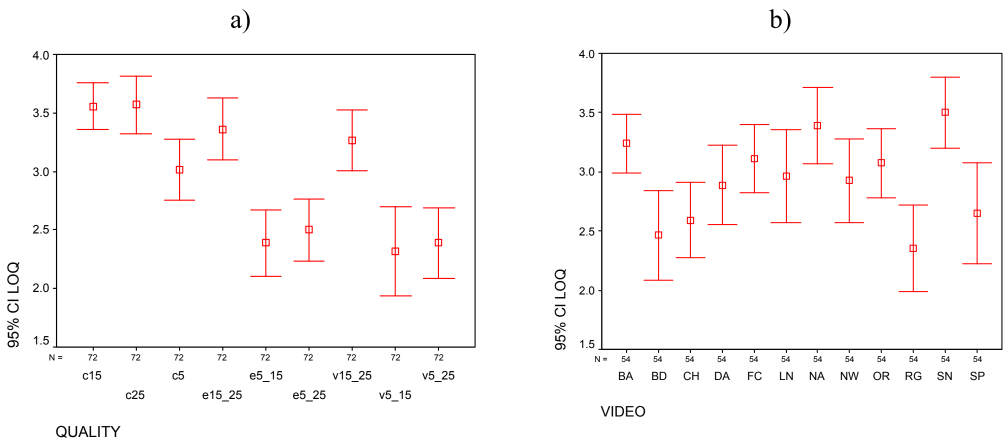

Figure 3: User perception of quality, depending on quality (a) and video (b) type

User information assimilation is affected by variation in device type (content-level parameter variation), yet is not affected by network-level and media quality parameter variation. A significant difference was measured between the HMD device and eye-tracking device, which were identified as respectively the best and worst devices for user information assimilation. We believe that the reason for the difference in information assimilation is due to the level of immersion, as high-immersion devices, i.e. a head-mounted display appear to facilitate a greater level of information assimilation. Interestingly, those using the HMD rated it as having the lowest overall user perceived level of video quality, despite enabling the greatest level of video information transfer. We suggest that this reduction in user perceived level of quality is due to pixel distortion as a result of a higher field of view and highlights the information / satisfaction compromise of display systems, i.e. for consistent video clips, a higher field of view provides a higher level of information assimilation, yet provides a lower user perceived level of quality (and visa-versa). This conclusion has possible implications on the future of fully immersive head-mounted display devices, as the authors believe that any device that is perceived to deliver low quality, despite its ability to improve the transfer video information, will rarely be commercially accepted by the user.

User Perceived level of quality is affected by all quality parameter variation, which shows that participants can effectively distinguish between a video presentation with and without error. This supports (Wijesekera & Srivastava, 1996), who showed that the presence of even low amounts of error results in a severe degradation in perceptual quality. Consequently, it is essential to identify the purpose of the multimedia when defining Quality of Service (QoS) provision, e.g. applications relying on user perception of multimedia quality should be given priority over and above purely educational applications. User Perceived level of quality is also affected by video clip type at the network- and media-level, yet user perceived level of quality is not affected by video clip type at the content-level. This result is believed to be as a consequence of network- and media-level video content variation (i.e. delay, jitter and Region of Interest display manipulation). This finding suggests that variation of video content is more easily identified by users in certain video clips. Consequently, this disparity in level of perceived quality, as a result of video clip type, reflects the ability of specific video to mask network- and media-level video variation errors, e.g. the bath advert (BA), the pop video (NA) and the snooker clip (SN) appear to effectively mask video variation errors (see Figure 3b); yet the band (BD) and rugby (RG) clip (both of which are highly dynamic videos) do not effective hide network- and media-level video variation errors. Video variation was not made at the content level. Accordingly no significant impact was measured on user perceived level of quality, thus supporting the previous finding that participants can effectively distinguish between a video presentation with and without error.

User perceived level of enjoyment is affected by network-level quality parameter variation (jitter and delay), yet is not affected by media-level and content-level quality parameter variation (attentive region of interest display manipulation and display type). This findings support Procter et al. (1999), who observed that degradation of network level QoS has a greater influence on a subjects' uptake of emotive / affective content than on their uptake of factual content. This result has serious implications on the effective provision of user-centric quality multimedia, implying that if one wished to ensure user information assimilation, then network level quality parameter variation should be used, however, if one wishes to maintain user perceived level of enjoyment, then content-level quality parameter variation should be used (see Table 5).

6. Conclusion

In this paper, we proposed a multimedia quality model which incorporates both user and technical perspectives in its composition. Our work has shown that user perception of distributed multimedia quality cannot be achieved by purely technical-perspective QoS parameter variation. If commercial multimedia development effectively considered the user-perspective in combination with QoS quality parameters, then multimedia provision would aspire to facilitate appropriate multimedia, in context of the perceptual, hardware and network criteria of a specific user, thus maximising the user's perception of quality. Furthermore, the development of user-perspective personalisation and adaptive media streaming offers the promise of providing the customer with truly user-defined, accessible multimedia that allows users to directly interact with multimedia systems on their own perceptual terms.

References

Apteker, R.T., Fisher, J.A., Kisimov, V.S., & Neishlos H. (1995). Video Acceptability and Frame Rate. *IEEE Multimedia*, 2(3), 32 - 40.

Barnett, B. S. (1996) Motion Compensated visual pattern image sequence coding for full motion multi-session video conferencing on multimedia workstation. *Journal of Electronic Imaging,* 5, 129-143.

Bouch, A., Wilson, G. & Sasse M. A. (2001). A 3-Dimensional Approach to Assessing End-User Quality of Service. *Proc. of the London Communications Symposium,* pp.47-50.

Claypool, M., & Tanner, J. (1999). The Effects of Jitter on the Perceptual Quality of Video, *ACM Multimedia'99 (Part 2),* Orlando, FL., pp. 115-118.

Ghinea, G., & Thomas, J.P. (1998). QoS Impact on User Perception and Understanding of multimedia Video Clips, *Proc. of ACM Multimedia '98*, Bristol UK, pp. 49- 54.

Gulliver S.R. and Ghinea G. (2003). How Level and Type of Deafness Affects User Perception of Multimedia Video Clips, *Universal Access in the Information Society*, *2*(4), 374-386.

Hollier, M. P. & Voelcker, R. M. (1997). Towards a multimodal perceptual model, *BT technological Journal,* 15(4), 162-171.

Kies J. K., Williges, R.C. & Rosson, M. B. (1997). Evaluating desktop video conferencing for distance learning. *Computers and Education,* 28, 79-91.

Koodli, R., & Krishna, C.M. (1998). A loss model for sorting QoS in multimedia applications. *Proc.. of ISCA CATA-98: Thirteenth International Conference on Computers and Their Applications*, ISCA, Cary, NC, USA, pp. 234-237.

Kaufman, L., & Richards, W. (1969) Spontaneous fixation tendencies for visual forms. *Perception and Psychophysics,* 5, 85-88.

Masry, M., Hemami, S.S, Osberger, W.M. & Rohaly, A.M. (2001). Subjective quality evaluation of low-bit-rate video. *Human vision and electronic imaging VI – Proc. of the SPIE*, Rogowitz, B.E. and Pappas, T.N. (Eds.), SPIE, Bellingham, WA, USA, pp. 102-113.

Procter, R., Hartswood, M., McKinlay, A. & Gallacher, S. (1999). An investigation of the influence of network quality of service on the effectiveness of multimedia communication. *Proc. of the international ACM SIGGROUP conference on supporting group work*. ACM. New York, NY, USA, pp. 160-168.

Steinmetz, R. (1996). Human Perception of Jitter and Media Synchronisation. *IEEE Journal on Selected Areas in Communications*, 14 (1), 61-72.

Teo, P. C., & Heeger, D. J. (1994). Perceptual Image Distortion. *Human Vision, Visual Processing and Digital Display V, IS&T / SPIE's Syposium on Electronic Imaging: Science & Technology, Vol. 2179*, San Jose, CA, pp. 127-141.

van den Branden Lambrecht, C. J., & Verscheure, O. (1996). Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System. *Proc. of the SPIE, Vol. 2668,* San Jose, CA, pp. 450-461.

Wang, Y., Claypool, M., & Zuo, Z. (2001). An empirical study of RealVideo performance across the internet. *Proc. of the First ACM SIGCOMM Workshop on Internet Measurement.* ACM Press, New York, NY, USA, pp. 295-309.

Watson, A., & Sasse, M.A. (1997). Multimedia conferencing via multicasting: Determining the quality of service required by the end user. *Proc. of AVSPN '97*, Aberdeen, Scotland, pp. 189-194.

Watson, A., & Sasse, M.A. (2000). The Good, the Bad, and the Muffled: The Impact of Different Degradations on Internet Speech. *Proc. of the 8th ACM International Conference on Multimedia*, Marina Del Rey, CA, pp. 269-302.

Wijesekera, D., & Srivastava, J. (1996). Quality of Service (QoS) Metrics for Continuous Media. *Multimedia Tools Applications,* 3(1), 127-166.

Wijesekera, D., Srivastava, J., Nerode, A., & Foresti M. (1999). Experimental Evaluation of loss perception in continuous media, *Multimedia Systems,* 7, 486-499.

Wikstrand, G., & Eriksson, S. (2002). Football Animation for Mobile Phones, *Proc. of NordiCHI,* pp. 255-258.

Wikstrand, G. (2003). Improving user comprehension and entertainment in wireless streaming media, *Introducing Cognitive Quality of Service,* Department of Computer Science, Umeå University, Umeå, Sweden.

Wilson, G.M., & Sasse, M. A. (2000a). Do Users Always Know What's Good For Them? Utilising Physiological Responses to Assess Media Quality. *Proc. of HCI 2000: People and Computers XIV - Usability or Else! Springer,* McDonald S., Waern Y. and Cockton G. (Ed.), pp. 327-339, Sunderland, UK.

Wilson, G.M., & Sasse, M.A. (2000b). Listen to your heart rate: counting the cost of media quality. *Affective Interactions Towards a New Generation of Computer Interfaces*. Paiva A.M. (Ed.), Springer, Berlin, Germany, pp. 9-20.