

# Defining the Users Perception of Distributed Multimedia Quality

S.R. Gulliver

G. Ghinea

*Department of Information Systems and Computing*

*Brunel University, United Kingdom.*

*{Stephen.Gulliver, George.Ghinea}@brunel.ac.uk*

## Abstract

In our study, we explore the human side of the multimedia experience. The authors propose a model that assesses quality variation from three distinct levels: the *network-*, the *media-* and the *content-levels*; and from two views: the *technical-* and the *user-perspective*. By facilitating parameter variation at each of the quality levels and from each of the perspectives, we were able to examine their impact on user quality perception. 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.

## Keywords

Quality of Perception, Distributed Multimedia, Quality, User Perspective

## 1. Defining Multimedia Quality

Distributed multimedia quality is not defined by a “single monotone dimension”; it is judged instead using numerous factors, which have been shown to influence user criteria concerning presentation excellence, e.g. delay or loss of frames, audio clarity, lip synchronisation during speech, as well as the general relationship between visual auditory components [2]. As a result, considerable work has been done looking at different aspects of distributed multimedia video quality at many different levels. Due to these multiple influences, the comparable examination of perceived quality becomes complex. To aid this comparison this paper extends a quality definition model first used by Wikstrand [33] that segregates quality into three discrete levels: the *network-level*, the *media-level* and *content-level*. Wikstrand showed that all factors that influence distributed multimedia quality can be categorised by assessing the information abstraction. 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.

- The network-level is concerned with how data is communicated over the network and includes variation and measurement of parameters including: bandwidth, delay, jitter and loss.
- The media-level is concerned with how the media is coded for the transport of information over the network and / or whether the user perceives the video as being of good or bad quality. Media-level parameters include: frame rate, bit rate, screen resolution, colour depth and compression techniques.
- The content-level is concerned with the transfer of information and level of satisfaction between the video media and the user, i.e. level of enjoyment, ability to perform a defined task, or the user's assimilate critical information from a multimedia presentation.

At each quality abstraction defined in Wikstrand's model, quality parameters can be varied, e.g. jitter at the network-level, frame rate at the media-level and finally display-type at the content-level. Similarly, at each level of the model, quality can be measured, e.g. percentage of loss at the network-level, user mean opinion score (MOS) at the media-level, and task performance at the content-level.

As well as possessing three distinct information abstractions, distributed multimedia covers a range of applications, which reflects the symbiotic *infotainment* duality of multimedia, i.e. the ability to transfer information to the user, yet also provide the user with a level of subjective satisfaction in respect of its perceived quality. Consequently, the user perspective concerning multimedia quality should consider both how a multimedia presentation is understood by the user, yet also examine the user's level of satisfaction (both satisfaction with the perceived Quality of Service setting and level of enjoyment concerning the video material). As multimedia applications are ultimately produced for the education and / or enjoyment of human viewers, the user's perspective concerning the presentation quality is surely of considerable importance. Accordingly, distributed multimedia quality, in our perspective, is deemed as having two main facets: *Quality of Service (QoS)* and *Quality of Perception (QoP)*. The former (QoS) characterises the *technical perspective* and represents the performance properties provided by multimedia technology. The latter facet (QoP) considers the *user perspective*, measuring the infotainment impact of the presentation. Accordingly, and

in addition to the model defined by Wikstrand, we incorporate in our studies both user- and the technical-perspectives.

- **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 can not 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.

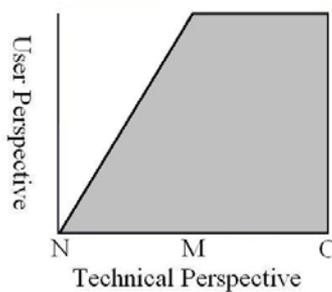


Figure 1: Quality Model, incorporates Network- (N), Media- (M) and Content-Level (C) abstractions and Technical- and User-Perspectives dimensions.

Since three quality abstractions have been defined (network, media and content levels), and two perspectives (technical and user), an extensive examination concerning the impact of multimedia quality variation must examine the perceived quality implications of parameter variations at each of the respective levels and from each of the defined perspectives.

The structure of this document is as follows: in section 2 we introduce the reader further to the research domain, we consider previous studies that involve quality variation and measurement at the defined three levels (both technical and user perspectives), and subsequently justify the need for our work. In section 3, we describe the research methodology and the experimental material that was used in our work, whilst in section 4 we describe how technical- and user-perspective parameter variation was achieved at network-, media- and content- levels. Research findings are presented in section 5, with conclusions being drawn in section 6.

## 2. Assessing Multimedia Quality

In this section we consider previous studies, which involve multimedia quality variation and measurement at the three levels of quality abstraction identified. Special attention has been made to differentiate the two distinct quality perspectives (the technical- or user-perspective).

In summary:

- **Network-Level:** Technical-perspective network-level variation of bit error, segment loss, segment order [8], delay and jitter [5] [8] [17] has been used to simulate QoS deterioration. Technical-perspective network-level measurements of loss [8] [13], delay and jitter [26], as well as allocated bandwidth [26] have all been used to measure network level quality performance.
- **Media-Level:** Technical-perspective media-level variation of video and audio frame rate [2] [7] [11] [12] [16] [31] [34] [35], captions [9], animation method [32], inter-stream audio-video quality [10], image resolution [12], media stream skews [20] [31], synchronisation [20] and video compression codecs [16] [36] have been used to vary quality definition. 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. Technical-perspective media-level measurement is generally based on linear and visual quality models [1] [14] [18] [21] [22] [23] [24] [25] [29] [36], with the exception of [26] who uses output frame rate as the quality criterion. User-perspective media-level measurement of quality has been used when measuring user 'watchability' (receptivity) [2], assessing user rating of video quality [7] [32], comparing streamed video against the non-degraded original video [17] [36], as well as for continuous quality assessment [34] [35] and gauging participant annoyance of synchronisation skews [20].
- **Content-Level:** Technical-perspective content-level variation was used to vary the content of experimental material [7] [9] [16] [17] [19] [20] as well as the presentation language [20]. User-perspective content-level variation has been used to measure the impact of user demographics [9], as well as volume and type of microphone [28] on overall perception of multimedia quality. Technical-perspective content-level measurement has, to date, only included stress analysis [34] [35]. User-perspective content-level measurement has measured 'watchability' (receptivity) [2], 'ease of understanding', 'recall', 'level of interest', 'level of comprehension' [17], information assimilation [7] [9], predicted level of information assimilation [9] and enjoyment [9] [32]. These results are summarised in Table 1.

Table 1: Comparison of User Perceptual Studies varied

(N = Network level, M = Media level, C = Content level).

Study	Participants	Varied Parameters	Measured Output
Aptker et al. [2]	60 students	<ul style="list-style-type: none"> <li>• Frame rate (M)</li> <li>• Video Content (C)</li> </ul>	<ul style="list-style-type: none"> <li>• Watchability (M)(C)</li> </ul>
Ghinea and Thomas [7]	30 participants	<ul style="list-style-type: none"> <li>• Frame rate (M)</li> <li>• Video Content (C)</li> </ul>	<ul style="list-style-type: none"> <li>• Information Assimilation (C)</li> <li>• Satisfaction (M)</li> </ul>
Gulliver and Ghinea [9]	50 participants (30 hearing / 20 deaf)	<ul style="list-style-type: none"> <li>• FrameRate (M)</li> <li>• Captions (M)</li> <li>• Video Content (C)</li> <li>• Demographics (C)</li> </ul>	<ul style="list-style-type: none"> <li>• Information Assimilation (C)</li> <li>• Satisfaction (C)</li> <li>• Self perceived ability (C)</li> </ul>
Procter et al. [17]	24 participants	<ul style="list-style-type: none"> <li>• Network Load (N)</li> <li>• Video Content (C)</li> </ul>	<ul style="list-style-type: none"> <li>• Comprehension (C)</li> <li>• Uptake of non-verbal information (C)</li> <li>• Satisfaction (M)</li> </ul>
Wilson and Sasse [34] [35]	24 participants	<ul style="list-style-type: none"> <li>• Frame Rate (M)</li> </ul>	<ul style="list-style-type: none"> <li>• Galvanic Skin Resistance (C)</li> <li>• Heart Rate (C)</li> <li>• Blood Volume Pulse (C)</li> <li>• QUASS (M)</li> </ul>

To extensively consider distributed multimedia quality effectively from a user-perspective it is essential that, where possible, both technical- and user-perspective parameter variations be made at all quality abstractions of our model, i.e. network-level (technical-perspective), media-level (technical- and user-perspective) and content-level (technical- and user-perspective) parameter variation – see Figure 1. Moreover, in order to effectively measure the infotainment duality of multimedia, i.e. 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, thus assessing the content-level user-perspective.
- The user's subjective satisfaction, measuring both the user's perception of objective QoS settings, yet also user enjoyment.

Interestingly, none of the afore-mentioned studies achieved this set of criteria and it is on this that our research shall focus its attention.

### 3. Research Methodology

In our study, we intend to explore the human side of the multimedia experience. In accordance with their proposed quality model, we used three structured laboratory-based experiments to investigate the user quality perspective at network-, media- and content-levels respectively, incorporating the QoP concept in order to explore the human side of the multimedia experience.

#### 3.1. Quality of Perception: An Adaptable Approach

Ghinea and Thomas [7] initially used QoP to measure level of information assimilation and satisfaction, where multimedia video clips were shown at varied frame rates. QoP 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 [4] [7] [27]. 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, QoP is a term used in our work that encompasses 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).

Originally, Ghinea and Thomas defined QoP-S using a 7-point Likert scale to measure the user's satisfaction with the quality of the video presentation. Although determining the user perception of video QoS (at a media-level), in our work variation to the original methodology was required, in order to measure user satisfaction at both media- and content-level quality abstractions. Previously [9], QoP was used to measure the impact of hearing level on a user's level of enjoyment (QoP-LoE) and self-predicted level of information assimilation (QoP-PIA). Interestingly, both QoP-LoE and QoP-PIA are measured at the content-level, which demonstrates that QoP-S facilitates the effective use of content-level user feedback.

In our study QoP-S is subjective in nature and consists of two component parts: QoP-LoQ (the user's judgement concerning the objective QoS) and QoP-LoE (the user's Level of Enjoyment), thus targeting perceptual quality at both media- and content-levels respectively. Accordingly, QoP-S successfully considers the user-perspective from both user-perspective quality paradigms.

##### 3.1.1. Measuring QoP

###### 3.1.1.1. Measuring Information Assimilation / Understanding (QoP-IA)

QoP-IA implements content query and allows us to measure a user's ability to understand / assimilate the content of the video clip (content-level). Thus, after watching a particular

multimedia clip, the user was asked a number of questions that examined the information being assimilated from certain information sources. QoP-IA was then 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, which does not contain text. Originally Ghinea and Thomas [7] defined (V) and dynamic-based (D) information separately. However, as user feedback suggested that the distinction between these variables were confusing, these information sources were combined in our study.
- A : Audio-based information that is presented in the audio stream.
- T : Textual-based information that is contained in the video window, e.g. the newscaster's name in a caption window.

Since QoP-IA is calculated as being 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 shows how questions were used to test the user's assimilation and understanding of V, A and T information sources (the source of the data is contained in brackets and the answer is underlined) in a pop video clip:

- What was the bald man doing in the video? (V) Moving a chair / furniture.
- Name two features of the clip that relate to the Orient? (V) She is wearing a t-shirt that has a dragon logo, (T) She performed in a Japanese video commercial

As all questions gauging QoP-IA have unambiguous answers it is possible to calculate the percentage of correctly assimilated information, facilitating examination of user information assimilation / understanding, as a result of quality parameter variation.

### 3.1.1.2. Measuring Subjective Level of Quality (QoP-LoQ)

To ensure that user satisfaction includes measurement at the media-level we have used QoP-LoQ (the users subjective perception of QoS provision), the first component part of QoP-S in our approach. In order to measure QoP-LoQ, users were asked to indicate, on a scale of 0 - 5, how they judged, independent of the subject matter, the presentation quality of a particular piece of multimedia content they had just seen (with scores of 0 and 5 representing “no” and, respectively, “absolute” user satisfaction with the multimedia presentation quality). Accordingly, QoP-S incorporates the media-level user-perspective of our model.

### 3.1.1.3. Measuring Subjective Level of Enjoyment (QoP-LoE)

To ensure that user satisfaction includes measurement at the content-level we have used QoP-LoE (the subjective Level of Enjoyment), which is the second and final component part of QoP-S in our study. In order to measure QoP-LoE, the user was asked to express, on a scale of 0 - 5, how much they enjoyed the a multimedia presentation (with scores of 0 and 5 representing “no” and, respectively, “absolute” user satisfaction with the multimedia video presentation). Accordingly, QoP-S also incorporates the content-level user-perspective, in addition to the media-level user-perspective.

## 3.2. Experimental Material

The set of video clips used in our experiments consists of a series of 12 windowed MPEG video clips, with duration of video clips was between 26 and 45 seconds. The multimedia video clips 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 varied from those that are informational in nature (such as a news / weather broadcast) to ones that are usually viewed purely for entertainment purposes (such as an action sequence, a cartoon, a music clip or a sports event) – see Figure 2. Specific clips, such as the cooking clip, were chosen as a mixture of the two viewing goals. These videos are described in Figure 2.

**COMMERCIAL****(BA)**

**Commercial Clip (BA)** - an advertisement for a bathroom cleaner is being presented. The qualities of the product are praised in four ways - by the narrator, both audio and visually by the couple being shown in the commercial, and textually, through a slogan display.

**BAND****(BD)**

**Band clip (BD)** - this shows a high school band playing a jazz tune against a background of multicoloured and changing lights.

**CHORUS****(CH)**

**Chorus clip (CH)** - this clip presents a chorus comprising 11 members performing mediaeval Latin music. A digital watermark bearing the name of the TV channel is subtly embedded in the image throughout the recording.

**ANIMATION****(DA)**

**Animation clip (DA)** - this clip features a disagreement between two main characters. Although dynamically limited, there are several subtle nuances in the clip, for example: the correspondence between the stormy weather and the argument.



**WEATHER FORECAST  
(FC)**

**Weather clip (FC)** - this is a clip about forthcoming weather in Europe and the United Kingdom. This information is presented through the three main channels possible: visually (through the use of weather maps), textually (information regarding envisaged temperatures, visibility in foggy areas) and orally (by the presentation of the forecaster).



**INDIAN LIONS  
(LN)**

**Documentary clip (LN)** - a feature on lions in India. Both audio and video streams are important, although there is no textual information present.



**NATALIE'S POP MUSIC  
(NA)**

**Pop clip (NA)** - is characterised by the unusual importance of the textual component, which details facts about the singer's life. From a visual viewpoint it is characterised by the fact that the clip was shot from a single camera position.



**NEWS  
(NW)**

**News clip (NW)** - contains two main stories. One of them is presented purely by verbal means, while the other has some supporting video footage. Rudimentary textual information (channel name, newscaster's name) is also displayed at various stages.



COOKING CLIP

(OR)

**Cooking clip (OR)** - although largely static, there is a wealth of culinary information being passed to the viewer. This is done both through the dialogue being pursued and visually, through the presentation of ingredients being used in cooking of the meal.



RUGBY

(RG)

**Rugby clip (RG)** - presents a test match between England and New Zealand. Textual information (the score) is displayed in the upper left corner of the screen. The main event in the clip is the scoring of a try. The clip is characterised by great dynamism.



SNOOKER

(SN)

**Snooker clip (SN)** - the lack of dynamism is in stark contrast to the Rugby clip. Textual information (the score and the names of the two players involved) is clearly displayed on the screen.



SPACE

(SP)

**Space clip (SP)** - this was an action scene from a popular science fiction series. As is common in such sequences it involves rapid scene changes, with accompanying visual effects (explosions).

Figure 2: Video Frame 500, for the 12 video clips used in our experiment.

## 4. Parameter Variation

Our research aimed 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. Accordingly our study incorporated three major research objectives:

- **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 user perception of multimedia quality. Although other authors have considered the perceptual impact of delay and jitter, previous studies fail to consider both level of user understanding (information assimilation) and user satisfaction (both of the video perceived QoS and concerning the user level of enjoyment).
- **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 [3]. 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 (RoI) attentive display system, which was developed to produce both video content- and user- 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.

We now proceed to describe the experimental methodology associated with each of these studies.

### 4.1. General Experimental Process

All experiments used in our work followed a similar consistent 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, which was followed by a basic eye-test to ensure that they 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 QoP-IA questions.

After introducing the participant to the experiment, the appropriate experimental software and video order were configured. 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. The correct video order was loaded and the first video was displayed.

The content of the videos used in our experimental presentations was manipulated to simulate specific quality parameter variation. Due to the reduced bandwidth requirement and increased perceptual impact of corrupted audio, the audio stream will not be manipulated in our research. By purely manipulating video content we minimise the number of variables that impact the user's perception of quality. After showing each video clip, the video window was closed and the participant was asked a number of QoP questions relating to the video that they had just been shown. QoP questions were used to encompass both QoP-IA and QoP-S (QoP-LOE and QoP-LOQ) aspects of the information being presented to the user. 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 12 videos, independent of the display device.

## 4.2. Experimental Participants

Participant numbers were determined by two factors: the number of variable factors in each experiment and the practical availability of subjects. Each participant that was used in our experiments had never participated in a QoP 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 – students, clerical and academic staff, white collar workers, as well as a number of retired persons. All participants, however, spoke English as their first language, or to a degree-level qualification, and were computer literate.

In previous studies, Ghinea and Thomas [7] used 30 participants to measure the impact of both frame-rate and video content on user perception. Procter et al. [17] used 24 participants to measure the impact of both network load and video content on user perception. For each of the experiments in our study, we matched the participant numbers used in previous perceptual studies.

## 4.2. 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 fps.

To simulate delay and jitter we artificially manipulated skew between audio and video media streams. We manipulated video so that the number of delay and jitter errors equalled 2% the number of video frames, which corresponds to one video error every two seconds (the minimum time taken to identify perceptually relevant regions in a visual stimuli [6] [15] [37]). Consequently, to simulate accumulated video delay, after every 50 video frames a single video frame was repeated, i.e. for 50 original frames, 51 were shown. At no point was the audio manipulated. As a consequence of duplicated video frames, the manipulated delay video was 2% longer than the audio stream. To simulate video jitter - the variation in delay - a number of jitter points were simulated that was equal to 2% the number of video frames, e.g. for a 918 frame video (at 25 frames per second), 18 separate jitter points were simulated. The location of jitter points was randomly defined. The direction (+/-) and amplitude of each video skew (0 - 4 frames) was also randomly defined, however, minute adjustments were made to ensure that the net delay was equal to zero, i.e. the first and last video frame synchronised with the audio stream. Randomly-sized video skew (0 - 4 frames) was used to ensure variation in jitter, ranging from 0ms to 160ms, which represents a maximum skew equal to two times the minimal noticeable synchronisation error between video and audio media [20]. By duplicating frames, videos were produced with the perception of 5, 15 and 25 fps, which allowed users perception to be measured as a result of both quality variation and frame-rate variation. Video variation therefore includes: 5, 15 and 25 fps video containing no error (control), delay and jitter.

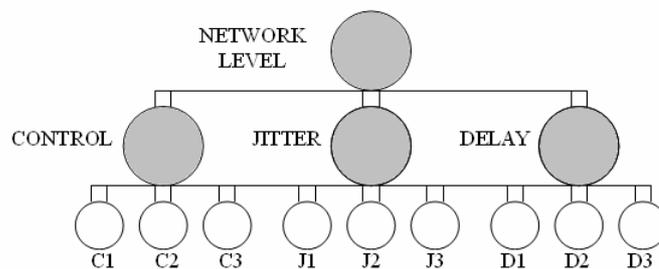


Figure 3: Participant distribution in order to measure impact of network quality parameter variation (Delay and Jitter).

In this experiment, 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 viewing order and frame rate that participants were ultimately going to view multimedia video clips (Figure 3). In each experimental sub-group (e.g. C1, C2, etc.), a within-subjects design was used. Thus, each participant viewed four video clips at 5 fps, four at 15 fps, and four at 25 fps. In order to counteract order effects, the video clips were shown in a number of order and frame-rate combinations, defined by the experimental sub-group name, e.g. participants in C3, J3 and D3 sub-groups (see Figure 3) viewed videos with frame-rates as defined by column ‘Order 3’ (see Table 2).

Table 2: Frame-rate order for Control, Jitter and Delay sub-groups.

Video	Code	Order 1	Order 2	Order 3
Commercial	BA	5	15	25
Band	BD	25	5	15
Chorus	CH	15	5	25
Animation	DA	25	15	5
Weather	FC	5	25	15
Documentary	LN	5	15	25
Pop	NA	15	25	5
News	NW	5	25	15
Cooking	OR	15	25	5
Rugby	RG	25	5	15
Snooker	SN	15	5	25
Space	SP	25	15	5

### 4.3. Media-Level Parameter Variation: Region of Interest Display

To create effective Region of Interest Displays (RoIDs), we produced multimedia videos that had an adaptive non-uniform distribution of resource allocation. To achieve this we used output data from an eye tracker and information about the content of the video, 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, content-dependent data related to significantly important visual primitives, e.g. edges, colour distribution, contrast and movement. Thus RoI (Region of Interest) areas, herewith referred to as foreground areas, were refreshed at a relatively higher frame rate than that of the non-RoI areas (background areas). Considerable effort was taken to make sure that each RoI foreground square covered at least  $4^\circ$  of the visual field ( $\pm 2^\circ$  around the point of gaze), thus ensuring that the high acuity area of the fovea is contained within the foreground area.

Software was developed, using the Java Media Framework, which takes the original video (at 25fps) and a RoI information (either containing eye-tracker or content-dependent RoI data) and, using a 5 frame count, produces a playable multi-frame rate MPEG video that presents using foreground and background regions at different frame rate combinations. At playback, this video can be considered as a RoID, since it displays a higher level of quality in significant RoI. Moreover, as the system adapts video based on both eye-tracker (user-perspective) and video content (technical perspective) data the RoID fulfils the defined objective 2.

To identify how varied foreground and background frame rate impacts user perception, our study considered three possible foreground and background combinations. Accordingly, nine *video quality variations* were considered as part of our experiment: control 25fps (c25), control 15fps (c15), control 5fps (c5), eye-based and content-dependent 25fps foreground / 15fps background video (e25\_15, v25\_15); eye-based and content-dependent 25fps foreground / 5fps background video (e25\_5, v25\_5) and, finally, eye-based and content-dependent 15fps foreground / 5 fps background video (e15\_5, v15\_5).

Three experimental variables were manipulated in this experiment: RoID 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 RoID video was considered as part of our experiments.

To ensure experimental consistency a within-subjects design was again used to ensure that participants view all nine video quality variation types (c25, c15, c5, e25\_15, e25\_5, e15\_5, v25\_15, v25\_5, v5\_5) across the 12 videos. Accordingly, nine experimental groups were required, with video quality shown as described in Table 3.

Table 3: Order of Video Quality Variations in Media-level Perceptual Experiments.

	<b>Order 1</b>	<b>Order 2</b>	<b>Order 3</b>	<b>Order 4</b>	<b>Order 5</b>	<b>Order 6</b>	<b>Order 7</b>	<b>Order 8</b>	<b>Order 9</b>
<b>BA</b>	C5	C15	C25	E25_15	E_25_5	E15_5	V25_15	V25_5	V15_5
<b>BD</b>	V15_5	C5	C15	C25	E25_15	E_25_5	E15_5	V25_15	V25_5
<b>CH</b>	V25_5	V15_5	C5	C15	C25	E25_15	E_25_5	E15_5	V25_15
<b>DA</b>	V25_15	V25_5	V15_5	C5	C15	C25	E25_15	E_25_5	E15_5
<b>FC</b>	E15_5	V25_15	V25_5	V15_5	C5	C15	C25	E25_15	E_25_5
<b>LN</b>	E_25_5	E15_5	V25_15	V25_5	V15_5	C5	C15	C25	E25_15
<b>NA</b>	E25_15	E_25_5	E15_5	V25_15	V25_5	V15_5	C5	C15	C25
<b>NW</b>	C25	E25_15	E_25_5	E15_5	V25_15	V25_5	V15_5	C5	C15
<b>OR</b>	C15	C25	E25_15	E_25_5	E15_5	V25_15	V25_5	V15_5	C5
<b>RG</b>	C5	C15	C25	E25_15	E_25_5	E15_5	V25_15	V25_5	V15_5
<b>SN</b>	V15_5	C5	C15	C25	E25_15	E_25_5	E15_5	V25_15	V25_5
<b>SP</b>	V25_5	V15_5	C5	C15	C25	E25_15	E_25_5	E15_5	V25_15

54 participants were evenly divided into nine experimental groups. Participants were aged between 21 and 59 and were taken from a range of different nationalities and backgrounds.

#### **4.4. 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, aged between 18 and 56, were evenly allocated in to four different experimental groups. Within each group, users were shown the video clips using certain display equipment. Group 1 acted as a control group (standard mobility) and was shown the video clips using a 15 inch SVGA generic computer monitor enabled with a Matrox Rainbow Runner Video Card. Group 2 also viewed the video clips full screen using a computer monitor, however, the participants were simultaneously interacting with a Power Mac G3 (9.2) powered Arrington ViewPoint EyeTracker, used in combination with QuickClamp Hardware, which provides limited head mobility. Group 3 viewed the multimedia video clips using an Olympus Eye-Trek FMD 200 head-mounted display, which uses two liquid crystal displays and allows a greater autonomy of movement than a generic computer monitor. 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. Group 4 viewed the video clips using 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 using 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.

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, with the order as defined in Table 2. Accordingly, four types of display devices were considered in our experiments (representing varying levels of user mobility), and in keeping with previous experiments three multimedia video frame rates: 5, 15 and 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 two experimental studies. A pilot test of two participants was used to check and validate the output of each display device (8

participants in total). During this pilot, participants using the PDA commented that environmental noises interfered with the audio output. As we hoped to provide participants with a consistent audio level, headphones were used for all devices to limit interference from the surrounding environment.

## 5. Research Findings

QoP was used in our study to extensively characterise the user's perception of multimedia quality at all three levels of our 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 (see Table 4) highlight a number of important issues relating to the effective provision of user-centric quality multimedia. These issues will now be discussed.

A significant loss of frames (that is, a reduction in frame rate) does not proportionally reduce the user's understanding of the presentation (see Table 4). This finding supports the conclusions of Ghinea and Thomas [7] and justifies a reduction in bandwidth allocation, if and only-if user QoP-IA (information assimilation / understanding) is the primary aim of the multimedia presentation.

The use of frame rates below 15 fps was found to significant impact user QoP-LoQ (see Figure 4a and Table 4). This finding supports the work of Wijesekera et al. [31], who showed that frame-rate should be maintained at or above 12 fps if the user perception of multimedia quality is to be maintained. Interestingly, this finding also raises considerable concerns regarding the usability of frame rate-based attention display systems, since our findings show no positive benefits associated to such display techniques.

Table 4: A summary of our QoP finding. (\* - no significant difference; ✓ - significant difference).

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

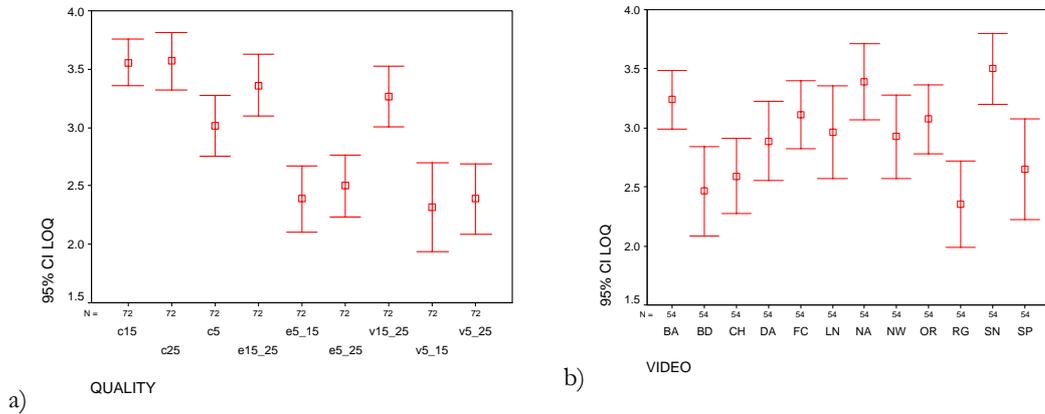


Figure 4: QoP-LoQ, dependent on quality (a) and video (b) type.

Video clip type significantly impacts user QoP-IA (Information Assimilation). Variation in user QoP-IA shows that the level of information assimilated varies significantly across the range of experimental video material. As the informational content of video determines the use of QoP-IA questions, and ultimately the reliability of QoP-IA, this finding supports the use of QoP-IA for each of our experiments.

Video clip type significantly impacts user QoP-LoE (Level of Enjoyment). Variation in user QoP-LoE shows that certain videos (NA, LN, and DA in our study – see Figure 5) were perceived as being overall more enjoyable, some (FC, RG) were perceived as generally less enjoyable. This finding is of interest, especially in the fields of advertising and education, as it implies that the type of video is more significantly important to the users’ level of enjoyment than implementing certain quality parameter variation, e.g. variation in the device type. Further work is required to fully understand the relationship between video content and user enjoyment, yet this aim lies outside the scope of this study.

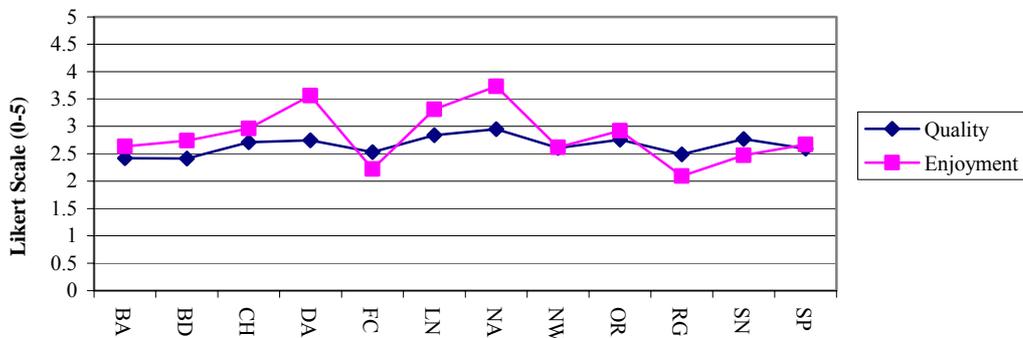


Figure 5: Average perceived level of Quality and Enjoyment.

User QoP-IA is significantly affected by variation in content-level parameter variation (device type), yet is not significantly affected by network-level or media quality parameter variation. Results show that the display device, used to watch a distributed multimedia video, significantly affects a user level of understanding. Moreover, a significant difference was measured between the head-mounted display (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 user QoP-IA is due to the level of immersion, with high-immersion devices (i.e. the HMD) facilitating a greater level of information assimilation.

Although variation in device type does not significantly impact user level of enjoyment, HMDs were found to significantly lower overall user perceived level of video quality (QoP-LoQ), despite enabling the greatest level of video information transfer. We suggest that this reduction in QoP-LoQ is due to pixel distortion as a result of a higher field of view and highlights a information / satisfaction compromise, i.e. for consistent video clips, a higher field of view provides a higher QoP-IA, yet provides a lower QoP-LoQ (and visa-versa). Unless high special detail can be achieved, 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 QoP-LoQ is significantly affected by network-, media-, and content-level quality parameter variation, i.e.: delay, jitter, attentive display RoI manipulation, and device type. This finding shows that participants are able to effectively distinguish between a video presentation with and without error. This supports [30], 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 appropriate QoS provision, e.g. applications relying on user perception of multimedia quality should be given priority over and above purely educational applications.

User QoP-LoQ is significantly affected by video clip type at the network- and media-level, yet QoP-LoQ is not significantly 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 attentive display RoI manipulation). This finding suggests that variation of video content is more easily identified by users in certain video clips. Consequently, this disparity in QoP-LoQ, 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) and snooker clip (SN) appear to effectively mask video variation errors (see Figure 4b); yet the band (BD) and rugby (RG) clip (both highly dynamic videos) do not effective hide network- and media-

level video variation errors (see figure 4b). Video variation was not made at the content level, which explains why no significant impact was measured on user QoP-LoQ.

User QoP-LoE is significantly affected by network-level quality parameter variation (jitter and delay), yet is not significantly affected by media-level and content-level quality parameter variation (attentive display RoI manipulation and display type). This findings support Procter et al. [17], 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 QoP-IA, then network level quality parameter variation should be used; however, if one wishes to maintain user QoP-LoE, then content-level quality parameter variation should be used (see Table 4).

## 6. Conclusion

In this paper, we have 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 means of purely technical-perspective QoS parameter variation. Accordingly, the future of multimedia research contains both promise and danger for user-perspective concerns.

We believe that a user will not continue paying for a multimedia system or device that they perceive to be of low quality, irrespective of its intrinsic appeal. Consequently, if commercial multimedia development continues to ignore the user-perspective in preference to other factors, i.e. user fascination (i.e. the latest gimmick), then companies risk ultimately alienating the customer. Moreover, by ignoring the user-perspective, future multimedia systems also risk ignoring accessibility issues, by excluding access for users with abnormal perceptual requirements, e.g. the deaf [9].

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.

By providing a extensive study of the distributed multimedia quality, our work shows that the user-perspective is as critically important to distributed multimedia quality definition, as QoS considerations. In conclusion, although multimedia applications are produced for the

education and / or enjoyment of human viewers, effective integration and consideration of the user-perspective in multimedia systems still has a long way to go....

## References

- [1] Ardito, M., Barbero, M., Stroppiana, M., and Visca, M., (1994). Compression and Quality. *Proceedings of the International Workshop on HDTV '94*, Chiariglione, L., (Ed.), Torino, Italy, October 26 - 28, 1994, Springer Verlag, pp. B-8-2.
- [2] Apteker, R.T., Fisher, J.A., Kisimov, V.S., and Neishlos H., (1995). Video Acceptability and Frame Rate. *IEEE Multimedia*, Vol. 2, No. 3, Fall, pp. 32 - 40.
- [3] 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*, Vol. 5, pp. 129-143.
- [4] Bouch, A., Wilson, G. and Sasse M. A., (2001) A 3-Dimensional Approach to Assessing End-User Quality of Service. *Proc. of the London Communications Symposium*, pp.47-50.
- [5] Claypool, M., and Tanner, J., (1999). The Effects of Jitter on the Perceptual Quality of Video, *ACM Multimedia '99 (Part 2)*, Orlando, FL, pp. 115-118.
- [6] De Groot, A. D., (1966). Perception and memory versus thought: Some old ideas and recent findings. *Problem solving: Research, method, and theory*, Klinmuntz, B. (Ed.), New York: John Wiley.
- [7] Ghinea, G., and 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.
- [8] Ghinea, G. and Thomas, J.P. (2000). Impact of Protocol Stacks on Quality of Perception, *Proc. IEEE International Conference on Multimedia and Expo*, Vol. 2, New York, pp. 847 -850.
- [9] 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*, Vol. 2, (4), pp. 374-386.
- [10] Hollier, M. P. and Voelcker, R. M. (1997) Towards a multimodal perceptual model, *BT technological Journal*, Vol. 15 (4), pp. 162-171.
- [11] Kawalek, J. A. (1995) User perspective for QoS Management. *Proc. of the QoS Workshop aligned with the 3<sup>rd</sup> Internations Conference on Intelligence in Broadband Services and Network (IS&N 95)*, Crete, Greece.
- [12] Kies J. K., Williges, R.C. and Rosson, M. B. (1997) Evaluating desktop video conferencing for distance learning. *Computers and Education*, Vol. 28, pp. 79-91.

- [13] Koodli, R., and 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.
- [14] Lindh, P. and van den Branden Lambrecht, C. J., (1996). Efficient Spatio-Temporal Decomposition for Perceptual Processing of Video Sequences. *Proc. of the International Conference on Image Processing, Vol. 3*, Lausanne, Switzerland, pp. 331-334.
- [15] Mackworth, J. F., and Morandi, A. J., (1967). The gaze selects informative details within pictures. *Perception and Psychophysics, Vol. 2*, pp. 547-552.
- [16] Masry, M., Hemami, S.S, Osberger, W.M. and 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.
- [17] Procter, R., Hartswood, M., McKinlay, A. and 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.
- [18] Quaglia D., and De Martin J. C., (2002). Delivery of MPEG Video Streams with Constant Perceptual Quality of Service. *Proc. of IEEE International Conference on Multimedia and Expo (ICME), Vol. 2*, Lausanne, Switzerland, pp. 85-88.
- [19] Rimmell, A. M., and Hollier M. P., (1999). The significance of cross-modal interaction in audio-visual quality perception, *Multimedia Signal Processing*, IEEE Signal Processing Society 1999 Workshop on Multimedia Signal Processing, pp. 509-514.
- [20] Steinmetz, R., (1996). Human Perception of Jitter and Media Synchronisation. *IEEE Journal on Selected Areas in Communications, Vol.. 14 (1)*, pp. 61-72.
- [21] Teo, P. C. and Heeger, D. J., (1994). Perceptual Image Distortion. *Human Vision, Visual Processing and Digital Display V, IS&T / SPIE's Symposium on Electronic Imaging: Science & Technology, Vol. 2179*, San Jose, CA, pp. 127-141.
- [22] van den Branden Lambrecht, C. J., Farrell, J. E., (1996). Perceptual quality metric for digitally coded color images. *Proc. of the VIII European Signal Processing Conference EUSIPCO*, Trieste, Italy, pp. 1175-1178.
- [23] van den Bradnden Lambrecht, C. J., (1996). Colour moving pictures quality metric. *Proc. ICIP, Vol.. 1*, Lausanne, Switzerland, pp 885-888.
- [24] van den Branden Lambrecht, C. J., and 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.

- [25] Verscheure, O., and Hubaux, J. P., (1996). Perceptual Video Quality and Activity Metrics: Optimization of Video Services based on MPEG-2 Encoding, *COST 237 Workshop on Multimedia Telecommunications and Applications*, Barcelona.
- [26] Wang, Y., Claypool, M., and 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.
- [27] Watson, A., and 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
- [28] Watson, A., and 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.
- [29] Watson, A. B., (1998). Toward a perceptual video metric." *Proc. SPIE, Vol. 3299*, San Jose, CA, pp. 139-147.
- [30] Wijesekera, D., and Stivastava, J., (1996). Quality of Service (QoS) Metrics for Continuous Media. *Multimedia Tools Applications, Vol. 3* (1), pp 127-166.
- [31] Wijesekera, D., Srivastava, J., Nerode, A., and Foresti M, (1999). Experimental Evaluation of loss perception in continuous media, *Multimedia Systems, Vol. 7*, pp. 486-499.
- [32] Wikstrand, G., and Eriksson, S., (2002). Football Animation for Mobile Phones, *Proc. of NordiCHI*, pp. 255-258.
- [33] 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, 2003.
- [34] Wilson, G.M., and Sasse, M.A., (2000). 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, DE, pp. 9-20.
- [35] Wilson, G.M., and Sasse, M. A., (2000). 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
- [36] Winkler, S., (2001). Visual fidelity and perceived quality: toward extensive metrics. *Human vision and electronic imaging VI – Proc. of SPIE*, Rogowitz B. E. and Pappas T.N. (Eds.), Bellingham, WA, USA. pp. 114-125.
- [37] Yarbus, A. L., (1967). Eye movement and vision, trans. B. Haigh. Plenum Press, New York.