A 3-Dimensional Approach to Assessing End-User Quality of Service

Anna Bouch, Gillian M. Wilson & M. Angela Sasse

University College London

Abstract: Due to the increasing popularity of applications such as media streaming and videoconferencing, multimedia traffic dominates the Internet today. So far, research into Quality of Service (QoS) for networked multimedia has focused on technical performance at the network level, rather than the quality experienced by the end-user. With a more service-oriented culture, however, the notion of user-centered QoS is gaining hold; this raises the question of how to measure quality experienced by end-users in meaningful terms. This paper introduces a 3-dimensional approach to assessing audio and video quality in networked multimedia applications: measuring task performance, user satisfaction and user cost (in terms of physiological impact). We argue that – for any given application – the relative importance of each dimension needs to be determined, and assessment techniques selected accordingly.

1 Introduction.

The number of Internet users is expected to triple between 1998 and 2002 [1], largely because of new applications Multimedia Conferencing (MMC) and new services (such as e-commerce). With increasing Internet traffic and moves towards charging for use, it is imperative that researchers establish the levels of Quality of Service (QoS) required by users. Human Computer Interaction (HCI) methods are the most powerful means of extracting user requirements. These methods are based on the recognition that human-computer systems consist of the user, the computer, underlying network functionality, and a context of use, or *task*.

Many different HCI methods exist for capturing users' QoS requirements [2]. What is often overlooked, however, is the need to fit those methods to the context of users' interaction. For example, a user that interacts with the network to retrieve a urgent document from a known Web location operates in a very different context than the user who browses the Web absentmindedly.

This paper aims to show that, for network MMC, it is necessary to address whether delivered quality is usable in a given task situation, where usability is defined in three different ways: subjective satisfaction, task performance and user-cost (Figure 1):

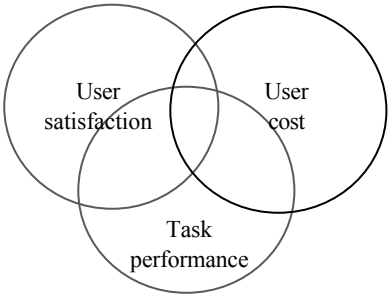

Figure 1: 3-D approach

2. The need for a 3-dimensional approach.

Different methods of capturing users' QoS requirements has led to a somewhat confusing array of prescribed audio and video quality thresholds that each claim to represent users' QoS requirements. More usefully - specific thresholds have been derived by using certain research methods in specific *contexts*. For example, if a user accesses a sports Web site the video should be streamed at no less than 4kbits/sec; the content of the video stream dictates the QoS requirement. It is our argument that, instead of trying to generalize users' QoS requirements, we must recognize the ability of each HCI method to answer certain questions about such requirements. Figure 2 shows the relation of each method described in the 3-D model to the context of interaction.

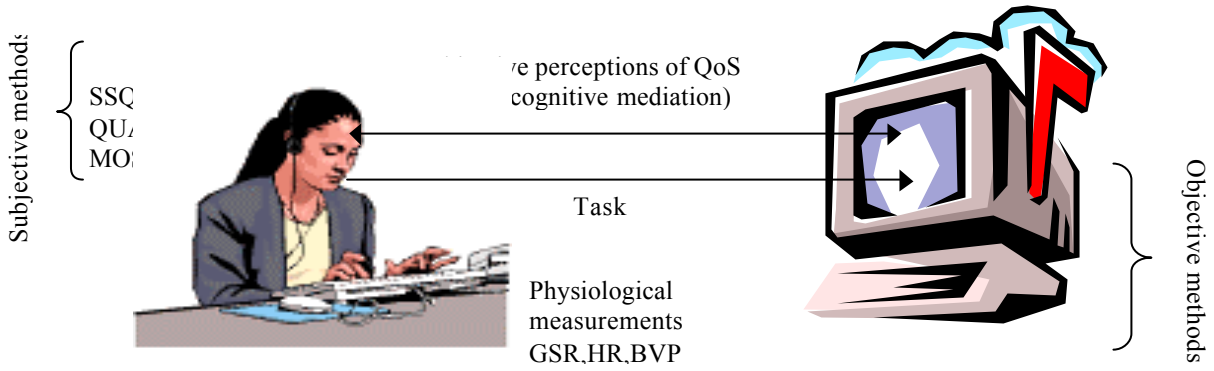

Figure 2: Relation of 3-D methods to context of interaction

3. Subjective methods.

In the subjective assessment of image quality single stimuli are rated using the 5-point quality or impairment scale resulting in the gathering of a Mean Opinion Score (MOS) (ITU-R BT. 500-8). These tests and scales are primarily concerned with establishing whether the effect of small degradations on *high-quality* video is noticeable. However, the fact that network conditions can change without warning, and long test segments are subject to memory effects [3] suggests that subjective measurement would be better served by turning the static MOS scale into a dynamic measurement scale.

We developed a software, dynamic version of the scale, where users could register their opinion of the delivered quality continuously, as and when it changed [4]. This tool is known as QUASS (QUality ASsessment Slider). Subjective methods are cognitively mediated. This means that users are required to consciously assess the level of QoS they receive. Subjective methods are therefore applicable when the subjective impression that user takes away from an interaction is important.

4. Task performance (TP).

In laboratory-based studies, many different types of task have been used, ranging from collaborative tasks such as decision making and problem solving tasks [5] to tasks with a more social content such as negotiating and bargaining [6]. This range and variety of tasks makes it difficult to draw comparisons and conclusions from the findings in general, and as Anderson et al. (1994) comment, "… *not only have different tasks been used in evaluation studies, but different methods of analysis have been adopted*" [5]. For the purposes of the 3-D approach, and this paper, our definition of a task is that which forms a (or the) main activity of a session.

In task performance measures, users' ability to complete a task is objectively measured under experimental conditions. Different levels of audio and/or video quality are applied to each condition. The resulting measurements are directly related to the ability of the application to support the particular task. For example, measures of task performance showed that, in a consonant identification task, task performance decreases drastically at an audio-visual skew of 160ms [7]. Tasks can be divided into *foreground* and *background* tasks [8]. Background tasks are likely to require lower quality since they are by their nature less critical (e.g. Dourish & Bly's, 1992 porthole system, for checking whether an office colleague is free to meet [9]). The advantage of this method is that users' assessments are not cognitively mediated - the subject doesn't have to be conscious of all the effects of QoS for the adequacy of that QoS to be assessed. This means that Task Performance measures are applicable where subjects are asked to perform a complex task.

Task performance measures and subjective assessments are appropriate where users are exposed to levels of QoS for relatively short time periods. Research suggests that physiological testing may be the only truly reliable predictor how intolerant users will become with levels of QoS in the long term. Also, as this method is unobtrusive, as a measure it is independent of task complexity.

5. User cost: Physiological indictors of stress.

One way of objectively measuring user cost is to monitor physiological signals that are indicative of stress and discomfort. When a user is presented with insufficient audio and video quality in a multimedia conference, he/she has to expend extra effort at the perceptual level. If they struggle to decode the information, this should induce a response of discomfort/stress, even if the user remains capable of performing his/her main task.

The following signals are measured: Heart Rate (HR), Blood Volume Pulse (BVP) and Galvanic Skin Response (GSR). For video, results showed that there was a significant increase in stress responses at 5fps from 25fps [10]. Interestingly, only 16% of participants noticed that the frame rate had changed.

We used 6 conditions to investigate audio degradations [11], 20% and 5% packet loss, audio that was loud & quiet, audio recorded using a 'bad' microphone and echo. Results showed that the most stressful condition, bad mike, was not subjectively rated as being poor. In addition, it was shown that out of the three most stressful and subjectively poor conditions, the network caused only one: 20% packet loss. Thus, the largely ignored problems due to the hardware set-up and user behavior affected users as much as, if not more so, than problems due to the network.

6. Discussion.

Our results show that different HCI methods can be used to answer specific questions concerning QoS requirements in different contexts. This is clearly shown by the fact that people can report no differences between, for example, two frame rates, yet the difference is registered physiologically (see Section 5). Each method in the 3-D model can be selectively emphasized (or weighted) so that the right questions can be answered. For example, if the context is entertainment, then the most important level is user satisfaction (Figure 3) - the opinion the user walks away with and the enjoyment they got from their interaction. If the context is distance learning the most important level is task performance; students must be able to learn from their interaction (Figure 4). If the user interviews daily my MMC, the most important level is user-cost, the longer term affects on health (Figure 5).

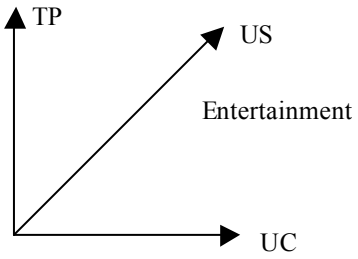

Figure 3: Method weighting (entertainment)

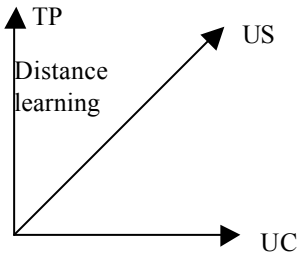

Figure 4: Method weighting (distance learning)

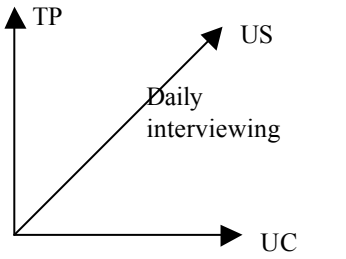

Figure 5: Method weighting (Daily interviewing)

The 3-D approach employed in our research has provided a more complete and integrated framework from which to conduct valid assessments of perceived QoS. The approach enables the assessment of users' dynamic, task-driven and unconscious QoS requirements. Through developing this framework we have:

- Attempted to widen the perspective on user QoS requirements capture, and show that this wider scope is indeed essential, as users' requirements should be described at a level of granularity that is appropriate to the context of their interaction.
- As a guide, indicated in which contexts each method of the 3-D approach should be used, thus providing researchers with a basic framework from which to conduct further exploration.
- Investigated the impact and relative importance of problems caused by the network, such as audio packet loss, video frame-rate, and hardware issues [12]. This allows network designers to allocate resources with users requirements specified, thus improving applications for end users, and to input provably meaningful variables into objective models.

In the literature to date, only a 3-D approach recognises the complexity of human response to – and requirements for – different levels of QoS. Passive rating will tell us little that is meaningful about how audio and video quality is perceived in a lengthy task, over a real-world network, nor will it help us to establish what quality

users *require* for different tasks to be completed successfully. Conversely, physiological methods should not be applied in investigating users' short-term behavior in, for example, an e-commerce interaction. This paper has argued that only by recognizing the relative benefits of a number of methodologies when defining QoS requirements, is it possible to produce truly meaningful results.

7. References.

[1] Cullinane, P. (1998). Ready, set, crash. Telephony 3, pp. 3-13.

[2] Bouch, A., Sasse, M., & DeMeer, H. G. (2000). Of packets and people: A user-centered approach to Quality of Service. Proceedings of IWQoS'00, June 5-7, 2000.

[3] Aldridge, R.P., Hands, D.S., Pearson, D.E. and Lodge, N.K. (1995): Continuous assessment of digitally-coded television pictures. IEE Proceedings - Vision, Image and Signal Processing, 145 (2), 116-123.

[4] Bouch, A. Watson, A., Sasse, M.A. (1998): QUASS – A tool for measuring the subjective quality of real-time multimedia audio and video. Proceedings of HCI '98, 1-4 September, Sheffield, UK.

[5] Anderson, A. H., Mullin, J., Newlands, A., Doherty-Sneddon, G. & Fleming, A. (1994). Video-mediated communication in CSCW: Effects on communication and collaboration. Presented at workshop on VMC at CSCW '94, Oct. 20-25, Chapel Hill, NC.

[6] Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. London. Wiley.

[7] Knoche, H., De Meer, H.G., & Kirsh, D. (1999. Utility curves: Mean opinion scores considered biased. Proceedings of IWQoS'97, pp. 12-14.

[8] Buxton, W. (1995). Integrating the Periphery and Context: A New Model of Telematics. Proceedings of GI '95, pp. 239-246.

[9] Dourish, P., & Bly, S. (1992). Portholes: Supporting Awareness in a Distributed Work Group. Proceedings CHI'92.

[10] Wilson, G. & Sasse, M.A. (2000a) Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods. Proceedings of OZCHI 2000, pp. 135-142.

[11] Wilson, G. & Sasse, M.A. (2000b) Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods. In C. Paris, N. Ozkan, S. Howard & S. Lu (eds.) Proceedings of OZCHI 2000: Interfacing Reality in the New Millenium, pp. 135-142.

[12] Bouch, A., & Sasse, M.A. (2001). Why value is everything: A user-centred approach to network Quality of Service and pricing. Proceedings of 9th International conference on Quality of Service (IWQoS'01), Germany, June 6-8, 2001.