

A User Perspective of Quality of Service in m-Commerce

Gheorghita Ghinea and Marios C. Angelides

Department of Information Systems and Computing

Brunel University

Uxbridge, Middlesex

UB8 3 PH

UNITED KINGDOM

e-mail: {George.Ghinea@brunel.ac.uk, angelidesm@acm.org}

Abstract: In an m-commerce setting, the underlying communication system will have to provide a Quality of Service (QoS) in the presence of two competing factors - network bandwidth and, as the pressure to add value to the business-to-consumer (B2C) shopping experience by integrating multimedia applications grows, increasing data sizes. In this paper, developments in the area of QoS-dependent multimedia perceptual quality are reviewed and are integrated with recent work focusing on QoS for e-commerce. Based on previously identified user perceptual tolerance to varying multimedia QoS, we show that enhancing the m-commerce B2C user experience with multimedia, far from being an idealised scenario, is in fact feasible if perceptual considerations are employed.

Keywords: Quality of Service, User Perception, m-commerce

1. Introduction

Business to Consumer (B2C) applications represent a particularly high-growth area of electronic commerce. As these applications involve, by their very nature, users, it is they who will determine the ultimate success of the B2C sector. Whilst general user concerns in B2C environments relate to issues such as trust, privacy, and convenience [1][5][9][29], in a mobile communication context, considerations of user satisfaction are further hampered by matters relating to the relatively constrained input capabilities, as well as the limited size, processing capabilities and battery power of the devices themselves.

A central issue to the acceptance of B2C applications is the one of quality – and this shall be the main theme of the paper. As Figure 1 details, m-commerce B2C applications can be approached from three main perspectives, and, with each, one can associate a notion of quality. In this paper we shall, however, be concentrating primarily on the confluence between the user perspective and the data transmission viewpoint, showing how knowledge about perceived quality can be exploited from a data communication angle in a mobile B2C context. Accordingly, quality, in our work, has two main facets in an m-commerce environment: *of perception* and *of service*. The latter characterises the technical side of computer networking and represents the performance properties that the underlying network is able to provide. The former is a novel term coined by us and gives a more complete characterisation of the human side of the m-commerce experience.

That Quality of Service (QoS) impacts upon the success of m-commerce applications is without doubt, as it plays a pivotal role in attracting and retaining customers. Frustrated customers will leave a site if they perceive it as being slow (for instance), causing lost revenue. To compound the topic, m-commerce sites are likely to be compared to the user's actual shopping experiences in the physical world and their normal purchasing and browsing interactions in that environment. Whilst the advent of multimedia enhanced m-commerce applications offering, for instance, video footage of merchandise to be bought, will go some way towards alleviating this

factor, in a mobile communication environment such applications will inevitably have severe constraints put upon them by the limited bandwidth on offer, further highlighting the importance of QoS provision mechanisms in this context.

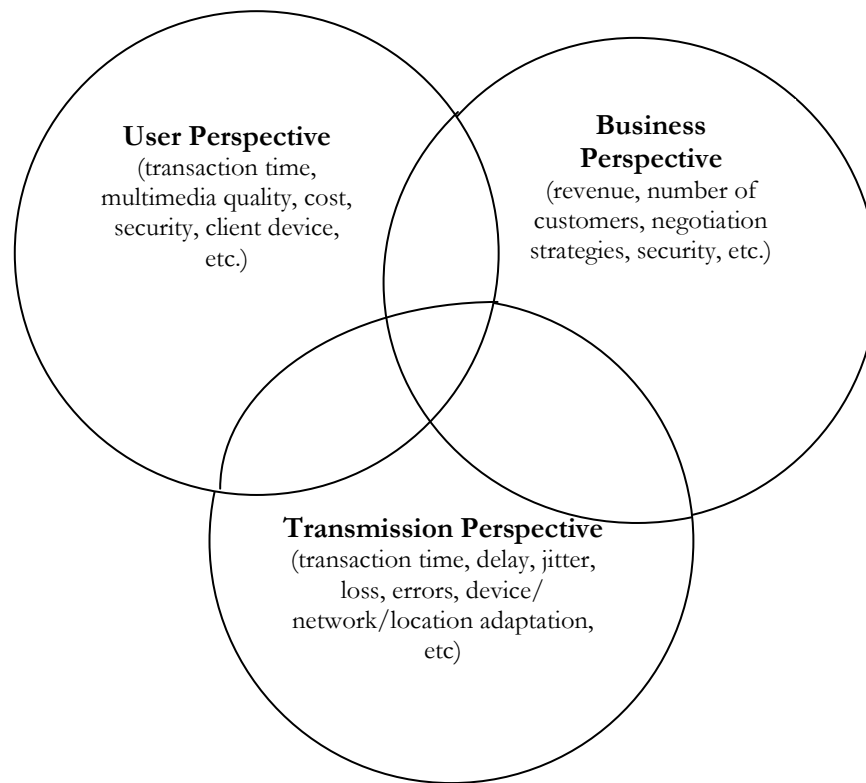


Figure 1. The different perspectives of multimedia enhanced B2C m-commerce

In this paper, we present an overview of QoS-related issues and how they impact on the user experience, with particular emphasis on m-commerce. Accordingly, the structure of the paper is as follows. Section 2 looks at how QoS issues in m-commerce from the prism of e-commerce. Whilst section 3 reviews research in the area of the perceptual impact of varying QoS and shows its applicability to wireless communications and m-commerce, section 4 brings together the main issues resulting from the previous two sections, integrating perceptual and QoS issues into a specifically-designed m-commerce communications architecture. Finally, conclusions are drawn in section 5.

2. The QoS Perspective

Traditional approaches of providing Quality of Service to distributed multimedia applications have focused on ways and modalities of ensuring and managing different technical parameters such as delay, jitter and packet loss over unreliable networks. Now, whilst issues raised by m-commerce systems do not necessarily preclude the use of such approaches, there are, nonetheless, a few-context specific characteristics which demand complementary methods of delivering QoS. Moreover, such methods of QoS provisioning have considerable overlap with those used by the more general category of e-commerce applications, and, as such, QoS considerations identified in the latter type of applications have direct applicability to the m-commerce arena.

The majority of these specific e-commerce QoS requirements stem from the fact that the e-commerce server has been identified as the main resource upon which rests the success of such applications. In a B2C session scenario, for instance, the user is, irrespective of the client device, interested in getting good quality search results and in the speed of the transactions, both of which are server dependent [8]. The longer the response delay, and the poorer the search results, the less inclined will the user be to shop from that specific e-commerce site, resulting in lost revenue.

Server response times have thus been identified as a key metric in e-commerce scenarios. It is therefore logical to for QoS considerations to concentrate on ways on how to manage such response times [4][7][8][11], and a primary focus has been on the identification of the specific characteristics of e-commerce servers' workloads. Such workloads have been found to be significantly different from traditional Web server workloads, and to be characterised by short and frequent requests, an abundance of dynamic data being generated, variable think times, and undefined session lengths [4][7][8][20]. Whilst there have been attempts to model e-commerce server traffic and requests using clustering techniques based on factors such as session-length,

request-class mix, and navigational behaviour [4][22], the frequency of such requests implies that QoS cannot be negotiated on a per-request basis, as is the case with distributed multimedia [7].

Even though e-commerce server workloads have been found to display time-of-day patterns (i.e. busiest during the day and least busy in early mornings), there are nonetheless substantial fluctuations to be found. Although the existence of ‘flash’ surges in server workload at times when promotions are running on the respective e-commerce site or during traditionally shopping-intensive periods, such as those preceding Valentine’s Day or Christmas Day, is to be expected, even when these factors are missing, observation of e-commerce server workload has found request rates to vary by a factor of nine [4]. However, it is precisely this relative unpredictability of server workload which make the application of traditional QoS management techniques especially opportune.

Admission control algorithms have thus been proposed in order to provide predictable e-commerce response times [10][11][18]. Such an algorithm would ensure user satisfaction with a B2C context from two points of view: not only would server delay latency be bounded, but denial of service attacks would also be prevented. The latter is especially an issue of concern in e-commerce systems, as server workload analysis has identified that a non-negligible percentage of requests are issued by robots, automatically collecting, for instance, price-information by ‘crawling’ through a site. Such robot-originated requests have a significant impact on server processing capacity, degrading the service provided to customers. Whilst work has been done on identifying particular characteristics of robot-sessions, the issue is further complicated by robots intentionally behaving like users [4] [23].

Admission control is inextricably linked to service differentiation provision, which is itself particularly suited to the for-profit transactions characteristic of e-commerce. For instance, personalisation, although benefiting the B2C shopping experience, has been identified as a potential threat to the performance and scalability of e-commerce systems [4]. The suggestion is

therefore to provide personalisation capabilities when server-usage is light and ban such capabilities at periods of high request; alternatively, users can have a personalised shopping experience, even at times of high system load, if they pay for the privilege.

Various server architectures have been put forward to enable faster response times. Accordingly, on starting an e-commerce session, the client first contacts a virtual server, known as a QoS broker, to whom initial QoS requests are sent. Based on overall system information and the requested QoS, the QoS broker can decide on whether or not to admit the session and then sends back an actual address of a 'real' server to the client, to whom subsequent requests are sent [7]. This server can itself be part of a multi-tiered server architecture and might only deal with requests for non-dynamic resources, forwarding requests for dynamic resources to application servers, which in turn, interact with database servers to provide support for personalisation, information retrieval and transaction management [4].

Whilst dynamically generated information greatly aids personalisation, it is a drain on server processing capacity and does add delays to server response times. Overall system performance could be improved, though, if information was cached on the application server side, and various mechanisms for dynamic content caching have been put forward. Thus, whilst [4] suggests the use of a cache large enough to store all of a site's responses, [26] proposes WebGraph, a graph-based representation of web pages, where nodes (called weblets) represent parts of a web page that either remain static or change simultaneously, and edges represent the inclusion relationship. When a dynamic web page is created, only those weblets that have changed need be refreshed, thus relieving some of the server processing burden. Moreover, this framework can be also be used for QoS maintenance, as edges in the graph have attributes representing QoS characteristics (such as delay-, throughput- and loss-sensitiveness) of the particular weblet to which it is pointing. Thus, if a particular weblet is indicated as being loss-sensitive (such as is the case with financial information) and if the environment is one in which losses have occurred recently, then

that edge is deleted and the respective weblet is not refreshed, thus economising on resource requests.

3. Perceptual Impact of Varying QoS

As the pressure to add value to customers' shopping experience increases, so will the usage of multimedia in m-commerce environments. However, the increased data sizes associated with even relatively low presentation quality multimedia (compared to that of voice and text traffic) as well as the relatively low bandwidth available to m-commerce applications will make the underlying communication system struggle to provide an optimum QoS, resulting in unwanted features such as congestion, as well as data loss and errors.

However, this does not necessarily imply that there is an incompatibility between m-commerce applications and their usage of multimedia. Thus, although a user might be slightly annoyed at the lack of synchronisation between audio and video streams in an advertising clip, say, it is highly unlikely that (s)he will notice, for instance, the loss of a video frame out of the 25 which could be transmitted during a second of footage, especially if the multimedia video in question is one in which the difference between successive frames is small. In fact there has been a rich body of work which has studied the perceptual impact of varying multimedia QoS, which shall now be presented, highlighting their relevance to mobile communications.

3.1 Media Synchronisation Perception

Media synchronisation refers to the temporal relationship between two or more kinds of media or separate data streams. In a multimedia context this definition can be extended such that synchronisation in multimedia systems comprises content, spatial and temporal relations between media objects.

The most comprehensive results on the perceptual impact of synchronisation skews between media were reported in [6] [30]. Firstly, we shall take a look at the synchronisation between the audio and video streams when a human speaks, also known as *lip synchronisation*.

In the lip synchronisation experiment [30], the test subjects viewed 30-second clips of a speaker in a TV news environment featured in head, shoulder and body shots. Such shots enabled the viewers not to be disturbed by background information and to concentrate their attention on the gesture, eyes and lip movement of the speaker. Moreover, the fact that the test scenes had high temporal redundancy make them ideal for transmission at low frame rates, with consequently relatively low bandwidth requirements, characteristic of wireless transmissions. In an m-commerce environment, such shots could be used, for example, in the cases of a tennis pro explaining the features and qualities of a tennis racquet that the customer is considering buying.

Skews were artificially introduced between the video and audio streams of the clip and were shown to the viewers, together with the original recording. Users who did notice something wrong with the synchronisation were asked to quantify their level of annoyance on a 3-point scale ('acceptable', 'annoying', 'not sure if acceptable or annoying'). The main result obtained was that the dependency between perceived quality and lip synchronisation skew displayed a U-shaped characteristic, whereby between -80ms and $+80\text{ms}$ lip synchronisation was deemed acceptable by most of the test subjects, with very few saying that if there was an error it affected the quality of presentation. Indeed, sometimes, some 'out of sync' clips were classified as being 'in sync'. Out of these bounds, however, synchronisation skew was increasingly detected and perceived as annoying, with the 'out of sync' portion comprising skews beyond -160ms and 160ms . At such skews, the annoyance that users felt with such presentations often had adverse effects, such as users becoming distracted by the 'out of sync' effect and not focusing on the content of the clip. In the 'transient' region of the characteristic, an interesting result that surfaced was that audio behind video could be tolerated much better than the opposite case.

Moreover, the closer the speaker and the better the resolution were, the easier it was for users to detect the synchronisation errors and the more likely for them to be described as disturbing.

A comparison using languages other than English (the language of the news-cast) revealed no difference in the results. Similarly, there were no variations between people with different habits regarding the amount of TV and films watched. Lastly, no difference was detected between the same person speaking in a fast, normal, or slow manner.

Synchronisation between *audio and animation* is also of potential importance to m-commerce applications, as can be exemplified by audio commentary of an animated representation of a product of interest to the user. Here, the perceptual tolerance limits identified were not as stringent as those in lip synchronisation, with a skew of $\pm 80\text{ms}$ being tolerable.

A slide show is the most obvious combination of *audio and images*. Here a skew of around 1s, equivalent to the time required to advance the projector, is tolerable. Synchronisation of *audio with text*, also known as audio annotation, was found to have a permissible skew closely related to the pronunciation duration of short words, which is about 500ms, leading to an experimentally verified skew of 240ms.

In the case of synchronisation between *video and text* or *video and images* two cases can be distinguished. In the *overlay* mode, the image or text offers additional information to the displayed video sequence, as is the practice of having subtitles placed close to the topic of discussion in a multimedia video. Irrespective of video content, a skew of around 240ms has been shown to be sufficient in this case. When *no overlay* occurs, skew is less serious. In this case one could imagine a drawing detailing assembly instructions of a product being displayed together with a low frame rate video detailing the product's appearance when assembled. Here, a synchronisation of around 500ms between the video and image or the video and text is deemed sufficient, which is half the value of the roughly 1s required for human perception of simple images.

3.2 Frame Rate Perception

The first recorded work on human perception of different frame rates was done by Apeteker *et al.*[2]. They coined the term ‘human receptivity’ to mean not only just how the human user perceives multimedia video shown at diverse frame rates, but also more distinct aspects of a user’s acceptance of a video message. These include clarity and acceptability of audio signals, continuity of visual messages, lip synchronisation during speech, and the general relationship between visual and auditory message components.

Human receptivity is expressed as a percentage measure, with 100% indicating complete user satisfaction with the multimedia data. The authors derived a set of categories of data resulting from a common *Video Classification Scheme* (VCS). These *VCS curves* were obtained on the basis of the temporal nature of the data (i.e. its dynamic nature) as well as the importance of the auditory and visual components of the video message. Various video clips were thus classified into eight main categories. A video from each of the eight categories was shown to users in a windowed multitasking environment. Each multimedia video clip in turn was presented at three different frame rates (15, 10, and 5 frames per second - fps), in a randomised order. The end-users in these experiments had, however, no knowledge of the particular frame rate at which they were watching the video. The users rated the quality of the multimedia videos on a 7-point scale. A total of 60 people were tested for the 24 types of clips.

The most relevant result to come out of their work was that the dependency between human receptivity and the required bandwidth of multimedia clips is non-linear. Consequently, for certain ranges of human receptivity, a small variation of it leads to a much larger relative variation of the required bandwidth – a feature referred to as the *asymptotic property* of the VCS curves.

Bearing in mind the low frame rate videos used in their experiments (5fps, with a bit rate well within the reach of 2.5 and 3G mobile technology), this perceptual property can be exploited in a

bandwidth-constrained mobile communication environment by sacrificing a small amount of human receptivity in return for the release of a much larger relative amount of bandwidth, which could then be potentially used by future m-commerce sessions.

Related work has been done by Fukuda *et al.*[12]. They take 4 video sequences and encode them using the MPEG-2 algorithm. They focus on three main QoS parameters:

1. The spatial resolution – 640*480, 320*240 and 160*120 pixels
2. The Signal to Noise Ratio – this is affected by the value of the quantiser used in the MPEG-2 algorithm. The quantiser is applied in the encoding algorithm to each block 16*16 pixels large; if a large quantiser is applied the quality of the decoded block becomes poorer, which in turn leads to a lower SNR value.
3. The number of frames per second – 10 and 30 fps.

The authors then go on to show that, in the case of the 4 MPEG-2 clips considered, there is a common mapping between these parameters and the required bandwidth, independent of video content. Through tests involving only 5 participants, the authors then gauge the human perception of the 4 clips in order to map users' preference onto bandwidth.

Each user gave a score ranging from 1 (Poor) to 5 (Excellent) to the video sequences; based on these ratings the overall video quality of each video is given by a *Mean Opinion Score* (MOS):

$$MOS = \sum_{i=1}^{\text{number of users}} i * \frac{\text{number of persons who gave score } i}{\text{number of users}} \quad (1)$$

Finally, based on the MOS vs. bandwidth curves thus obtained and their deduced mapping, the authors propose a QoS management mechanism which delivers a high perceptual score in spite of bandwidth constraints.

3.3 Media Loss Perception

In contrast to earlier work done by Apteker and Steinmetz which assumed that the underlying network communication system provided lossless multimedia streams, Wijesekera *et al.* have carried out a series of experiments which evaluate user perception in the presence of media losses[33][34]. Their work is of particular importance bearing in mind the noisy, bandwidth-constrained environment characteristic of mobile communications, and the QoS handoff issues which arise in this context.

One of their initial results was that missing a few media units will not be negatively perceived by a user, as long as too many such units are not missed consecutively and that this occurrence is infrequent. This is of especial importance in a noisy, bandwidth constrained environment. They also found out that media streams could drift in and out of synchronisation without considerable human annoyance.

They further evaluated human tolerance of transient continuity and synchronisation losses with respect to audio and video. Media loss was of two types in their study:

- Consecutive - this refers to the maximal number of consecutive media data units (e.g. frames or audio packets) the loss of which could be tolerated
- Aggregate - here, the number of consecutive media units loss was kept fixed, but was replicated at random intervals during the multimedia presentation

Rate variation was a parameter which was also studied in their experiments. It refers to the ideal rate of a media flow and the maximum permissible deviation from it and was examined both from an intra-stream as well as from an inter-stream point of view.

Only two video clips of 30s duration were used in their experiments. Both of them featured bust views of a speaker (different in each of the clips) explaining didactic and academic related matters. Users were asked to give their opinions on a 10-point Likert scale (where 1 was poor,

while 10 was excellent). Similar to [30], participants were also asked to categorise each clip as ‘Do not mind the effect if there is one’, ‘I dislike it and it’s annoying’, and ‘I’m not sure’. Their main results can be summarised by the following points [34]:

- The pattern of user sensitivity varies depending on the type of media defect.
- Viewer discontent for aggregate video losses increases gradually with the amount of losses, while for other types of losses and synchronisation defects there is an initial sharp rise in viewer annoyance which afterwards plateaus out.
- Video rate variations are tolerated much better than rate variations in audio.
- Because of the bursty nature of human speech (i.e. talk periods interspersed with intervals of silence), audio loss in this case is tolerated quite well by humans as it results merely in silence elimination (21% audio loss did not provoke user discontent).

3.4 Quality of Perception

A different approach to evaluating the perceptual impact of varying QoS was adopted in [13]. Recognising multimedia’s *infotainment* duality, the authors proposed to enhance the traditional view of QoS with a user-level defined *Quality of Perception* (QoP). This is a measure which encompasses not only a user’s satisfaction with multimedia clips, but also his/her ability to perceive, synthesise and analyse the informational content of such presentations. They subsequently investigated the interaction between QoP and QoS and its implications from both a user perspective as well as from a networking angle.

The approach to evaluating QoP was mainly empirical, as is dictated by the fact that its primary focus is on the human-side of multimedia computing. Users from diverse backgrounds and ages (12-58 years old) were presented with a set of 12 short (30 - 45 seconds’ duration) multimedia clips. As detailed in Table 1, these were chosen to be as varied as possible, ranging from a relatively static news clip to a highly dynamic rugby football sequence. All of them depicted

excerpts from real world programmes and thus represent informational sources which an average user might encounter in everyday life. Each clip, selected snapshots of which are given in Figure 2, was shown with the same set of QoS parameters, unknown to the user. After each clip, the user was asked a series of questions (ranging from 10 to 12) based on what had just been seen and the experimenter duly noted the answers. Lastly, the user was asked to rate the quality of the clip that had just been seen on a scale of 1 - 6 (with scores of 1 and 6 representing the worst and, respectively, best perceived qualities possible).

VIDEO CATEGORY
1 - Action Movie
2 - Animated Movie
3 - Band
4 - Chorus
5 - Commercial
6 - Cooking
7 - Documentary
8 - News
9 - Pop Music
10 - Rugby
11 - Snooker
12 - Weather Forecast

Table 1. Video categories used in QoP experiments

Because of the relative importance of the audio stream in a multimedia presentation [6] as well as the fact that it takes up an extremely low amount of bandwidth compared to the video it was decided to transmit audio at full quality during the experiments. Parameters were, however, varied in the case of the video stream. These include both spatial parameters (such as colour depth) and temporal parameters (frame rate). Accordingly, two different colour depths were considered (8 and 24-bit), together with 3 different frame rates (5, 15 and 25 frames per second -



Figure 2 Snapshots of video clips used in QoP experiments

fps). A total of 12 users were tested for each (*frame_rate*, *colour_depth*) pair. In summary (see Figure 3), the results obtained in the QoP experiments showed that:

- A significant loss of frames (that is, reducing the frame rate) does not proportionally reduce the user's understanding and perception of the presentation. In fact, in some instances (s)he seemed to assimilate more information, thereby resulting in more correct answers to questions. This is because the user has more time to view a frame before the frame changes (at 25 fps, a frame is visible for only 0.04 sec, whereas at 5 fps a frame is visible for 0.2 sec), hence absorbing more information. This observation has implications on resource allocation.
- User assimilation of the informational content of clips is characterised by the wys<>wyg (what you see is not what you get) relation. What this means is that often users, whilst still absorbing information correctly, do not notice obvious cues in the clip. Instead the reasoning process by which they arrive at their conclusions is based a lot on intuition and past experience.
- Users have difficulty in absorbing audio, visual and textual information concurrently. Users tend to focus on one of these media at any one moment, although they may switch between the different media. This implies that critical and important messages in a

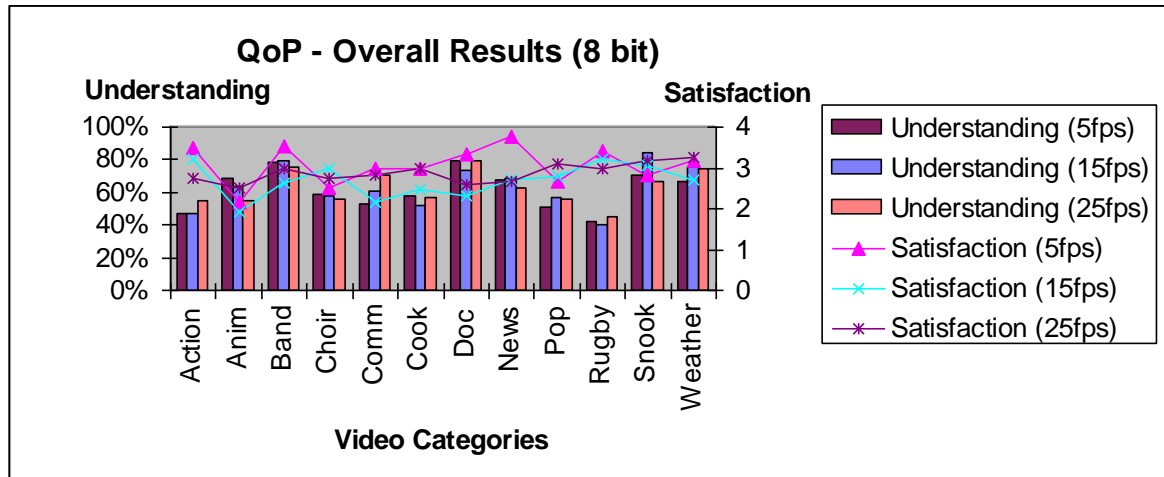


Figure 3 Experimental QoP results at 8 bit colour depth

multimedia presentation should be delivered in only one type of medium, or, if delivered concurrently, should be done so with maximal possible quality.

- The link between perception and understanding is a complex one; when the cause of the annoyance is visible (such as lip synchronisation), users will disregard it and focus on the audio message if that is considered to be contextually important.
- Highly dynamic scenes, although expensive in resources, have a negative impact on user understanding and information assimilation. Questions in this category obtained the least number of correct answers. However the entertainment value of such presentations seems to be consistent, irrespective of the frame rate at which they are shown. The link between entertainment and content understanding is therefore not direct and this was further confirmed by the second observation above.

All these results indicate that Quality of Service, typically specified in technical terms such as end-to-end delay, must also be specified in terms of perception, understanding and absorption of content - Quality of Perception in short - if multimedia presentations are to be truly effective.

3.5 Perceptual Impact of Audio Reconstruction Methods

A different angle on the perceptual impact of media loss was taken in [31] [32], which examined the effect on users of three different audio reconstruction methods. These were: silence substitution (here any missing audio packets are replaced with silence so as to preserve playback order), waveform substitution (a missing packet is replaced by a copy of the last correctly received packet) and Linear Predictive Coding (LPC - a synthetic quality speech coding algorithm which preserves about 60% of the informational content of the speech signal).

24 participants took part in the experiments and read out phonetically balanced sentences as given by the IEEE [15]. These essentially are made up of short syntactically varied sentences containing five key words each. Up to 40% loss was randomly generated in the list of words considered, while packet sizes were of 20, 40 or 80ms duration.

The results obtained showed that waveform substitution and LPC overall performed better than silence substitution irrespective of packet size. LPC outperformed waveform substitution for high loss rates and large packet sizes, since speech characteristics change with the lost packets of sound.

3.6 Perceptual Impact of Multimedia Presentations in Low Bandwidth Environments

Low bandwidth environments, such as the ones typical of mobile communications, need not necessarily imply their unsuitability for multimedia presentations, especially if perceptual considerations are taken into account. As was shown in the previous subsections, low video frame rates do not significantly impact upon the perceived quality of multimedia applications – this was true for both the human receptivity and QoP measures of perceptual multimedia quality. Moreover, media loss, as long as it infrequent and of short sizes, is well within the limits of perceptual tolerance.

The question arises, however, of what the cut-off rate is beyond which the quality of transmitted audio and video becomes unacceptable to human users. Whilst [6][30] did not explicitly consider, the impact of low frame rates on lip synchronisation, [19] showed that audio and video are not perceived as being synchronised for frame rates of less than 5 fps. This corroborates the findings of [31] [32], where it is reported that the lack of lip synchronisation at especially low frame rates (2-3 fps), although helpful to the user, has mainly a psychological benefit.

Moreover, the results presented in [19] also corroborate the finding [16] that the jitter associated with video frame presentation times (20ms) produces skew that is well within the limits of human perception.

The impact of low frame rates on speech intelligibility was also investigated in [3] where it is shown that when speakers can see each other on a low-frame-rate video screen, they articulate more clearly than the case where they cannot see each other and are communicating only over an audio link. This contrasts with the case when speakers can see each other face to face, when their speech is less clear.

So, whilst a frame rate of 5fps would seem to be a perceptual cut-off point beyond which quality is no longer acceptable, in practice this might actually be lower. The reason for this is that the perceptual impact of multimedia quality as experienced by users is task-dependent, as identified in [17], where although the general trend of video losses being much easily tolerated than audio ones irrespective of the task at hand is confirmed, the actual thresholds are heavily task dependent. In fact, a loss of 99% of video frames is still regarded as acceptable quality if the users engage in task-solving, a situation not unlike ones encountered in m-commerce applications. Thus, multimedia presentations, even when subjected to extremely high losses and presented at very low frame rates, need not necessarily impact negatively on the user multimedia-enhanced m-commerce experience.

QoS factors	Perceptual Effect
Media synchronisation	Lip synchronisation will be tolerated for frame rates of 5 fps. Apart from synchronisation between audio and animation which has more stringent requirements, synchronisation between other media will be generally well tolerated at low frame rates.
Frame rate	Low frame rates (such as 5fps) do not necessarily imply low perceptual satisfaction with presentations. Tolerance may extend to even lower frame rates if task solving is required.
Media losses	Media losses can be tolerated as long as they are infrequent and have small aggregate sizes. In the case of audio, silence substitution is a simple, perceptually acceptable, way of masking such losses
Quality of Perception	At low frame rates might even surpass that obtained at higher ones due to the persistence for longer on the screen of the particular frame.

Table 2. Perceptual considerations in low bandwidth environments

4. Perceptual considerations in m-commerce

Previous work on the user experience in e-commerce has focused on issues as diverse as privacy, trust, interface design, bargaining strategies and different cultural perceptions [1][5][9][21][24][27][29]. Whilst, with the exception of interface design, there is no reason to believe that conclusions from these studies would not be applicable in the m-commerce arena, all of this work has had the underlying assumption that the underlying communication system is capable of delivering an optimum QoS. As has been highlighted in the previous section, this will be difficult in an m-commerce setting due to two main competing factors – firstly, the increasing pressure to add multimedia content to the shopping experience in order to attract and retain

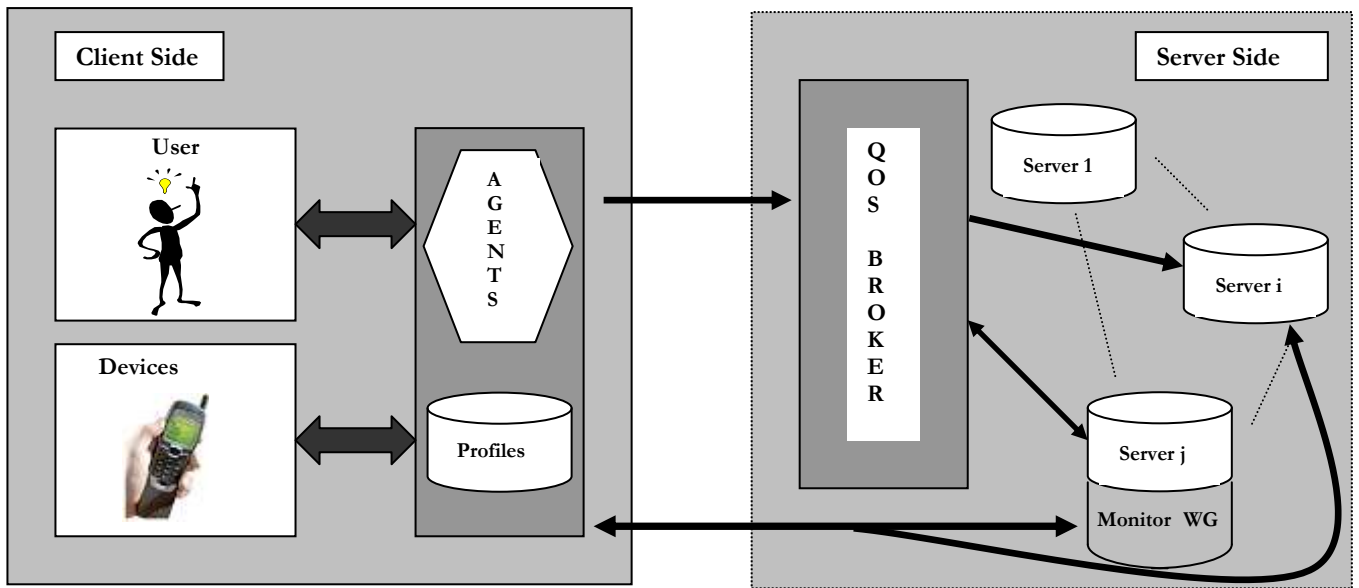


Figure 4 Integrating User Considerations in an m-Commerce Architecture

customers; secondly, the bandwidth on offer, even at 3G and 4G level, might not be sufficient for such applications, bearing in mind the potential increase in the number of users and the number and diversity of devices accessing m-commerce applications.

4.1 A User Oriented m-Commerce Architecture

Whereas server response time will still continue to be a main factor in determining user take-up of m-commerce, this does not preclude enriching the user m-commerce experience with multimedia content. This is primarily because, as has been pointed out in the preceding section, multimedia need not necessarily be sent with full-quality parameters (thus increasing the server burden), for users can be quite tolerant to low quality multimedia presentations.

A multimedia enabled B2C m-commerce architecture integrating user requirements is presented in Figure 3. Here, user profiles are created offline and store not only user shopping preferences but also multimedia quality preferences (such as ‘don’t mind a black and white versions of images’, ‘start transmitting video at 5 fps, and do not transmit video if the available bandwidth is not enough for a 3fps display’). Such profiles are stored together with device profiles in an

intelligent client interface, which may also, depending on the capabilities (processing power, memory size) of the client devices include agents geared to enhance the shopping experience.

The user and device profiles get sent, together with the initial client request to initiate the m-commerce session to a QoS broker, a virtual server which performs the call admission functions based on knowledge of workloads of servers comprising the server pool. If a session is admitted, the QoS broker passes on the client details both to an application server (server_i in Figure 4), in charge of commercial transaction management, as well as to a designated multimedia server (server_j in Figure 4). The latter accesses multimedia databases which store different quality versions of the same material, and dynamically retrieve into their caches the version corresponding to the requested quality. As mobile devices are unlikely to possess significant computational capabilities, any such multimedia material is decompressed on the server side and sent uncompressed to the requesting device. Whilst dedicated multimedia (and application) servers ensure that the processing burden does not fall on one server, the fact that full-sized multimedia data would have to be transmitted to the client implies that, in the initial stages at least, transmission might have to start off with a lower quality and then, depending on resource availability, build on to the requested quality. Moreover, such a scheme of quality adaptation has been shown to be perceptually preferable [17].

A QoS monitor on the multimedia server periodically updates the QoS broker with QoS information relating to the transmitted multimedia content. Moreover, each such server adapts the WebGraph framework [26] to provide scaleable, dynamic presentations of multimedia, with each weblet potentially representing a different quality version of a multimedia clip (a 5 fps clip would have a weblet with a refresh rate of 0.2s, whilst one transmitted at 2 fps would have an associated weblet with a refresh rate of 0.5s).

There are no barriers to implementing the proposed architecture using current technologies, such as ones based on WAP (the Wireless Application Protocol) [36] and Bluetooth [35].

Accordingly, if one assumes the usage of WAP-enabled Bluetooth devices which actively listen for a WAP-enabled QoS broker, then a multimedia enhanced B2C session starts when the client finds itself in the piconet of such a QoS broker and discovers it using the Bluetooth Service Discovery Protocol (SDP). A connection to the server-QoS broker can then be done automatically or at the client's specific request. Alternatively, the server can also check periodically for client devices and, if such WAP-enabled devices are found, start pushing data to the client (which the client has the option to refuse).

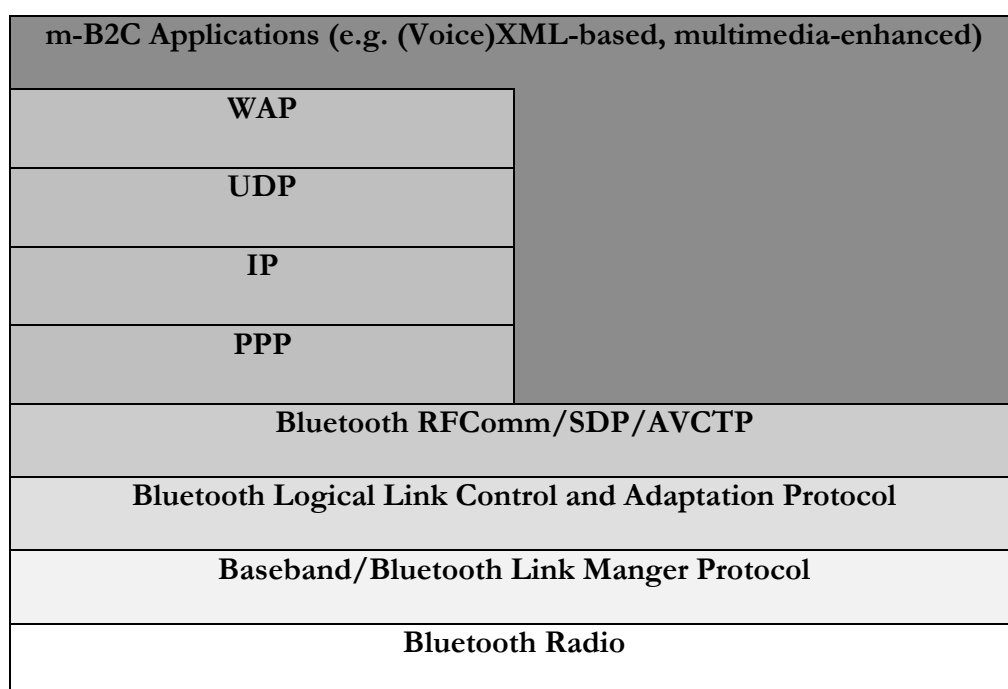


Figure 5 Multimedia-enhanced B2C protocol stack in a WAP- Bluetooth environment

In this example, it is assumed that WAP services use the Point-to-Point Protocol (PPP), although other protocol stack configurations can be supported and must be indicated by the Bluetooth SDP. Accordingly, if the connection is requested, then, as a result of the Service Discovery Protocol, the server-QoS broker is in a position to determine the client's capabilities, and instructs the application and multimedia servers to initiate the m-commerce session. The latter can transmit the multimedia content of the application using the Bluetooth Audio/Video Control Transport Protocol (AVTCTP), as detailed in Figure 5 (here Baseband, the Link Manager Protocol, and the Logical Link Control and Adaptation Protocol correspond to the OSI layer 1

and 2 Bluetooth protocols [35], while the RFComm protocol emulates the signalling through which serial ports send and receive data). The m-commerce transaction then continues till conclusion, or, if the client leaves the piconet coverage of the server, discovery information could be cached so that the transaction could proceed at a later stage.

4.2 Maximising the Perceptual Experience

As the previous section has shown, there are strong drivers for including perceptual considerations in resource allocation. The issue of how to allocate such resources for the best possible user benefit therefore arises. If we denote by QoP_{mB2C} the Quality of Perception user experience in a mobile B2C environment then, analogously to QoS, QoP_{mB2C} is a function of different perceptual parameters:

$$QoP_{mB2C} = QoP_{mB2C}(qop_1, qop_2, \dots, qop_n) \quad (2)$$

Here the qop_i , $1 \leq i \leq n$, are parameters that impact upon the mobile commerce perceptual experience. Such parameters could include the quality of the video, audio and textual components of a multimedia presentation, the type of client device, and the transaction response time. For instance in [14] it was shown that, if the definition of QoP is taken to be a user's satisfaction with multimedia clips, together with his/her ability to perceive, synthesise and analyse the informational content of such presentation, then, for a general multimedia presentation, the latter component is given by:

$$QoP_{IT} \cong \frac{V * QoS_V + A * QoS_A + T * QoS_T}{V + A + T} \quad (3)$$

Here, V , A and T denote the relative importance of the video, audio and textual components as conveyors of information, whilst QoS_V , QoS_A and QoS_T are parameterised values representing the delivered quality of video, audio and text respectively in the context of the multimedia clip being visualised. Relation (3) thus highlights the dependence of inverse proportionality between a person's QoP and the informational complexity of the clip (represented by $V+A+T$). It

moreover also shows that QoP is directly proportional to the quality with which the main conveyor of information is delivered in the context of the clip.

As relation (3) shows, each qop_i parameter is related to the underlying QoS provided by the communication system, which comes with a corresponding consumption of resources:

$$qop_i = qop_i(R_1, R_2, \dots, R_m) \quad (4)$$

Here, the R_j , $1 \leq j \leq m$, represent the networking resources needed by parameter qop_i . Without loss of generality it can be assumed that the same set of resources $\{R_j\}$ are used by each qop_i parameter. The problem of maximising the user perceptual experience in B2C m-commerce can therefore be stated as:

$$\text{Maximise } QoP_{mB2C} \text{ such that } \sum R_j \leq R \quad (5)$$

By R we have denoted the available networking resources. The constrained optimisation problem represented by (5) can be converted to an unconstrained optimisation problem by considering the associated Lagrangian:

$$L(R, \lambda) = QoP_{mB2C} + \lambda (R - \sum R_j) \quad (6)$$

and solving the equation:

$$dL = 0 \quad (7)$$

This boils down to:

$$\sum_{i=1}^n \frac{\partial L}{\partial qop_i} \frac{\partial qop_i}{\partial R_j} = \lambda \quad 1 \leq j \leq m \quad (8)$$

Whilst solving the solving the system of equations given by (8) is a non-trivial problem due to the difficulties in mapping functional requirements of perceptual considerations to associated resources and QoS parameters, in practice there have been approaches to this end and are inevitably based on empirically obtained results. Thus, for instance in [14] a simple but effective

mapping linking QoP parameters to QoS is given by a matrix representation of Table 3. Thus, if importance is attached to the textual component of a multimedia application (such as those encountered in collaborative environments) then, although individual bit errors may be tolerated, loss of contiguous frames will not. The human brain's ability to filter noise and reconstruct missing information means that even if bit errors do occur, perception and understanding will not suffer proportionally. However the loss of a series of frames will result in a volume of text being displayed for a much shorter length of time than intended or not at all. In either case, this would entail the loss, rather than corruption, of textual information whose inference, even in context, might be extremely difficult.

QoP To QoS MAPPING		QoP		
		VIDEO	AUDIO	TEXT
Q o S	Bit error rate (BER)	low	low	low
	Segment loss (SL)	low	high	high
	Segment order (SO)	high	medium	medium
	Delay (DEL)	medium	medium	low
	Jitter (JIT)	medium	low	medium

Table 3 Relative importance of QoS elements for QoP

In order for a point of maximum to occur, besides (8) a further condition must occur, namely that $d^2L < 0$. However, even if this condition does not hold, [25] has shown that, as long as the functional relation between QoP_{mB2C} and resource usage is clarified, then greedy allocation algorithms, where resources are preferentially managed in order of their perceived utility, are optimal in this case.

4.3 Further Considerations

Future m-commerce activities will place considerable processing pressure on servers. Whilst our proposed architecture tries to relieve some of these constraints by having designated application

and multimedia servers, other considerations need to be addressed as well. Thus, the variety of devices that might engage in m-commerce activities, means that servers must provide suitably tailored content for each of these devices. The XML-FO (eXtensible Mark-up Language Formatting Objects) standards can be used in such scenarios. Moreover, since current wireless devices can handle presentation aspects when specified with XML-FO, this need not necessarily be done on the server-side.

From a different perspective, current network conditions (such as static vs. moving, battery powered vs. live power) will further increase the pressure for servers to provide scaleable multimedia content, but will also furnish the impetus to provide dynamic switching capabilities dependent on contextual characteristics.

In [18] the need for novel ways of interaction between mobile device users and servers in m-commerce scenarios is identified, particularly a voice recognition-based natural language query mechanism which would simplify transactions, particularly from a user perspective, if one bears in mind the relatively limited interfaces characteristic of wireless devices. In our architecture, such an interaction mechanism can be integrated in the intelligent client interface. Again, the XML family of languages can be used here, through VoiceXML [28], particularly in the synthesis of the elaborated query responses.

Last but certainly not least, another important issue which must be taken into account when integrating multimedia perceptual considerations in m-commerce applications is the fact that they are task-dependent, and more work needs to be done in order to explore the perceptual impact of multimedia clips presented with differing spatial and temporal parameters, when done so in this specific context. Indeed, if previous results stemming from work exploring the perceptual impact of varying multimedia quality, which have shown that user tolerance increases when tasks are involved (as opposed to task-independent settings) [17], are an indication of the influence of the task at hand on perceived quality, then chances are that resource requirements needed to maintain an acceptable quality of user experience will not increase.

5. Conclusions

User considerations are of paramount importance for the proliferation and increased acceptance of m-commerce applications. Although multimedia applications have enhanced the user computing experience, contributing to the success of the World-Wide Web, little work has been done exploring how multimedia can improve and contribute to the take-up of m-commerce. Whilst this might partly stem from concerns rising from the narrow bandwidth available and the limited processing capabilities of m-commerce client devices, our paper has shown that, if user perceptual considerations are taken into account, previously identified human tolerance levels certainly do not exclude incorporation of multimedia applications in m-commerce. On the contrary, multimedia applications can be transmitted with perceptually tolerable parameters, augmenting the user m-commerce experience. Furthermore, if one bears in mind that technological improvements will lead to the increase of mobile devices' processing power, incorporating multimedia in m-commerce applications becomes less of a challenge and more of a necessity.

Exploiting these results, we have consequently proposed an architecture for multimedia-enhanced m-commerce, which integrates both user-perceptual requirements with technical transmission considerations. Key to this architecture is the issue of mapping perceptual requirements to underlying resources. Whilst obtaining such a map is a non-trivial task and must inevitably be based on extensive user tests, previous work has underlined the feasibility of such an undertaking, which will represent the main thrust of our future work.

In concluding, we remark that multimedia, even of relatively low technical quality, holds great promise for future m-commerce applications, for not only do limited resources not imply an automatic degradation of the user experience, but might, if integrated taking into account perceptual considerations, in fact lead to its enhancement.

References

- [1] Ackerman, M.S., Cranor, L.F., Reagle, J. "Privacy in e-commerce: examining user scenarios and privacy preferences", *Proceedings of the 1st ACM conference on Electronic Commerce*, pp. 1 - 8, Denver, Colorado, 1999.
- [2] Apteker, R.T., Fisher, J.A., Kisimov, V.S., and Neishlos, H. "Video Acceptability and Frame Rate", *IEEE Multimedia*, 2(3), pp. 32 - 40, 1995.
- [3] Anderson, A. and Blockland, A. "Intelligibility of Speech Mediated by Low Frame-Rate Video", *Proceedings of the Audio-Visual Speech Processing Conference*, Rhodes, Greece 1997.
- [4] Arlitt, M., Krishnamurthy, D., and Rolia, J. "Characterizing the Scalability of a Large Web-Based Shopping System", *ACM Transactions on Internet Technology*, 1(1), pp.44-69, 2001.
- [5] Basso, A., Goldberg, D., Greenspan, S., and Weimer, D. "First impressions: emotional and cognitive factors underlying judgments of trust e-commerce", *Proceedings of the 3rd ACM conference on Electronic Commerce*, pp. 137 – 143, Tampa, Florida, 2001.
- [6] Blakowski, G. and Steinmetz, R. "A Media Synchronisation Survey: Reference Model, Specification, and Case Studies", *IEEE Journal on Selected Areas in Communications*, 14(1), pp. 5 - 35, 1996.
- [7] Bochmann, G., Kerhervé, B., and Salem, M. "Service management issues in electronic commerce applications", in *Electronic Commerce Technology Trends: Challenges and Opportunities*, W. Kou and Y. Yesha (eds), pp. 227 - 238, IBM Press, 2000.
- [8] Bochmann, G., Kerhervé, B., Lutfiyya, H., Salem, M., and Ye, H. "Introducing QoS into electronic commerce applications", *Proceedings of 2nd International Symposium on Electronic*

Commerce, pp. 138-147, April 2001, Hong Kong, China, published as “Electronic Commerce Technologies”, LNCS 2004, Springer Verlag.

- [9] Callahan, E. and Koenemann, J. “A comparative usability evaluation of user interfaces for online product catalog”, *Proceedings of the 2nd ACM conference on Electronic Commerce*, pp. 197 – 206, Minneapolis, Minnesota, 2000.
- [10]Chen, H. and Mohapatra, P. “Session-Based Overload Control in QoS-Aware Web Servers”, to appear in *IEEE Infocom*, 2002.
- [11]Chen, X., Mohapatra, P., and Chen, H. “An Admission Control Scheme for Predictable Server Response Time for Web Accesses”, *Proceedings of the 10th World Wide Web Conference*, pp. 545 – 554, Hong Kong, China, 2001.
- [12]Fukuda, K., Wakamiya, N., Murata, M., and Miyahara, H., “QoS Mapping between User’s Preference and Bandwidth Control for Video Transport”, *Proceedings of the 5th International Workshop on QoS (IWQoS)*, pp. 291 – 301, New York, USA , 1997.
- [13]Ghinea, G. and Thomas, J.P., “QoS Impact on User Perception and Understanding of Multimedia Video Clips”, *Proceedings of ACM Multimedia ‘98*, pp. 49 - 54, Bristol, U.K., 1998.
- [14]Ghinea, G. and Thomas, J.P., “Crossing the Man-Machine Divide: A Mapping Based on Empirical Results”, *Journal of VLSI Signal Processing*, 29(1/2), pp. 139-147, 2001.
- [15]“IEEE Recommended Practice for Speech Quality Measurements”, *IEEE Transactions on Audio and Electroacoustics*, AU-17(3), pp. 225-246, 1969.
- [16]Jardetzky, P.W., Sreenan, C.J., and Needham, R.M., “Storage and synchronisation for distributed continuous media”, *Multimedia Systems*, 3(3), pp. 151 - 161, September 1995.

- [17]Kawalek, J. "A User Perspective for QoS Management", *Proceedings of the QoS Workshop aligned with the 3rd International Conference on Intelligence in Broadband Services and Network (IS&N 95)*, Crete, Greece, September 1995.
- [18]Kant, K. and Mohapatra, P. "Scalable Internet Servers: Issues and Challenges", *ACM SIGMETRICS Performance Evaluation Review*, 28(2), pp. 5-8, 2000.
- [19]Kouvelas, I., Hardman, V., and Watson, A., "Lip Synchronisation for use over the Internet: Analysis and Implementation", *Proceedings of IEEE Globecom '96*, London, U.K., November 1996.
- [20]Krishnamurthy, D. and Rolia, J. "Predicting the QoS of an Electronic Commerce Server: Those Mean Percentiles", *ACM SIGMETRICS Performance Evaluation Review*, 26(3), pp. 16-22, 2000.
- [21]Liang, T. and Doong, H. "Effect of Bargaining in Electronic Commerce", *International Journal of Electronic Commerce*, 4(3), pp. 23-43, 2000.
- [22]Menascé, D.A., Almeida, V.A.F., Fonseca, R., and Mendes, M.A. "A methodology for workload characterization of E-commerce sites", *Proceedings of the 1st ACM conference on Electronic Commerce*, pp. 119 – 128, Denver, Colorado, 1999.
- [23]Menascé, D.A., Almeida, V.A.F., Riedi, R., Ribeiro, F., Fonseca, R., and Meria, W. "In search of invariants for e-business workloads", *Proceedings of the 2nd ACM conference on Electronic Commerce*, pp. 56 - 65, Minneapolis, Minnesota, 2000.
- [24]Miles, G.E. and Howes, A. "A framework for understanding human factors in web-based electronic commerce", *International Journal of Human Computer Studies*, 52(1), pp. 131-163, 2000.

- [25]Mohan, R., Smith, J.R., and Li, C-S., “Adapting Multimedia Internet Content for Universal Access”, *IEEE Transactions on Multimedia*, 1(1), pp.104-114, 1999.
- [26]Mohapatra, P. and Chen, H. “WebGraph: A Framework for Managing and Improving Performance of Dynamic Web Content”, to appear *IEEE Journal on Selected Areas in Communications*, 2002.
- [27]Moore, D.A., Kurtzberg, T.R., and Thompson, L.L. “Long and Short Routes to Success in Electronically Mediated Negotiations: Group Affiliations and Good Vibrations”, *Organizational Behavior and Human Decision Processes*, 77(1), 23-43, 1999.
- [28]Sharma, C. and Kunins, J. *Voice XML*, New York, Wiley, 2002.
- [29] Spiekermann, S., Grossklags, J., and Berendt, B. “E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior”, *Proceedings of the 3rd ACM conference on Electronic Commerce*, pp. 38 – 47, Tampa, Florida, 2001.
- [30]Steinmetz, R. “Human Perception of Jitter and Media Synchronisation”, *IEEE Journal on Selected Areas in Communications*, 14(1), pp. 61 –72, 1996.
- [31]Watson, A. and Sasse, M.A. “Evaluating Audio and Video Quality in Low-Cost Multimedia Conferencing Systems”, *Interacting with Computers*, 8(3), pp. 255-275, 1996.
- [32]Watson, A. and Sasse, M.A. “Multimedia Conferencing via Multicast: Determining the Quality of Service Required by the End User” *Proceedings of AVSPN '97 - International Workshop on Audio-Visual Services over Packet Networks*, pp 189-194, Aberdeen, U.K., 1997.
- [33]Wijesekera, D. and Srivastava, J. “Quality of Service (QoS) Metrics for Continuous Media”, *Multimedia Tools and Applications*, 3(1), pp. 127-136, September 1996.

- [34] Wijesekera, D., Srivastava, J, Nerode, A. and Foresti, M. “Experimental Evaluation of Loss Perception in Continuous Media”, *Multimedia Systems*, 7(6), pp. 486-499, 1999.
- [35] www.bluetooth.com, “Bluetooth Specification Version 1.1”, 2001.
- [36] www.wapforum.org, “Wireless Application Protocol”, 2000



Gheorghita Ghinea is Lecturer in the Department of Information Systems and Computing at Brunel University. He obtained the BSc (Honours) and MSc degrees in Computer Science from the University of the Witwatersrand, Johannesburg, South Africa, and the PhD degree, also in Computer Science, from the University of Reading in the United Kingdom. His research interests span multimedia (especially perpetual aspects), Quality of Service, as well as computer networking and security issues. He is a member of the IEEE. His home page is at <http://www.brunel.ac.uk/~csstggg2>



Marios C. Angelides is Professor of Computing in the Department of Information Systems and Computing at Brunel University. He holds a BSc in Computing and a PhD in Information Systems both from The London School of Economics and Political Science where he began his academic career as a Lecturer in Information Systems. He has over 10 years of research experience in multimedia information systems where he has published extensively in journal and book format. He is the author of Multimedia Information Systems which was published by Kluwer. He is a member of the ACM, the IEEE Computer Society, and the British Computer Society. His home page is at <http://www.brunel.ac.uk/~csstmca>