

# Information Assimilation over the Internet

## An initial study

G.Ghinea

Dept. of Information Systems and Computing, Brunel University  
Uxbridge, Middlesex, United Kingdom

and

J.P.Thomas

School of Computer Science and Information Systems, Pace University  
Pleasantville, NY 10570, USA

### Abstract

Although the Internet holds the promise of long-distance education, multimedia entertainment etc., the quality of multimedia documents delivered by the Internet can vary enormously. In this paper we examine how varying quality of service affects a users' perception and understanding (and thereby learning) of multimedia presentations. Our results show that the quality of multimedia documents can be severely degraded without the user having to perceive any significant loss of informational content.

**Keywords:** Quality of Service, Quality of Perception, Multimedia

### 1. INTRODUCTION

The advent of the Internet promises 'live' or synchronous long-distance education, video-conferencing, multimedia presentations of sports, entertainment, documentaries and other educational programmes. Some of these applications such as synchronous long-distance education have advantages such as real-time interaction between students and the teacher. However, due to the enormous popularity of the Internet as well as the limitations of existing technologies and protocols, the quality of multimedia delivered by the Internet varies enormously. Thus it may not be possible for the Internet to consistently deliver multimedia of sufficiently high quality that is of benefit to the participants.

The aim of this research is to determine if the quality of the multimedia presentation has an impact on user perception and understanding. If the quality is such that the user gains very little from the presentation in terms of understanding and assimilating the

information contained in the presentation, the presentation serves little purpose from an educational point of view. The quality of the multimedia presentation is defined by a term call the Quality of Service (QoS). Multimedia QoS is typically measured using technical parameters such as end-to-end delay and jitter. One can envisage a situation where a professor or a student specifies the minimum required delay, packet loss and maximum jitter, before the class begins. To the average multimedia user, however, these parameters have little immediate meaning or impact. Moreover, even if a user understands what these terms mean, (s)he will most likely be unaware of the influence of these parameters on user perception or their impact on learning. Therefore, such technical parameters, although useful, disregard the users' perspective of the presentation. Instead, what the end user (such as student or professor) is interested in is more that (s)he enjoys the overall multimedia display while at the same time assimilating its informational content. The human element is an important part of the multimedia paradigm whose role, although appreciated, has often

been overlooked. This is because of the inherent difficulty and subjectivity associated with appreciating an individual's sense of multimedia perception and, consequently, precious little work has been done in this area. What work (Apteker 1995), (Fukuda 1997), (Steinmetz 1996) has been done, however, has indicated the existence of a threshold beyond which a user does not perceive an improvement in the QoS of multimedia applications, no matter the amount of resources allocated to them. Some research has also been done in order to establish the synchronisation limits between audio and video streams of a multimedia clip with which human observers are comfortable (Kawalek 1995). Although (s)he might be slightly annoyed at the lack of synchronisation between audio and video streams, 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.

This paper is therefore investigating the impact that multimedia QoS has not only on a user's satisfaction with the quality of the presentation itself, but also on his/her capacity to understand, analyse and synthesise the informational content of such presentations, i.e., the impact of the quality of multimedia presentations on education/learning. The motivation behind this approach is that in multimedia education it is not only the aesthetics that count, but also, also the effect of system and networking parameters on the user's potential to comprehend and assimilate the material and data presented in multimedia applications. This is especially important as multimedia databases become widespread and the technology is used in information-intensive domains such as education. The advent of the Internet where education over the Internet is becoming a reality, makes this an important and relevant issue. In other words, the QoS impact on user perception and understanding also has implications on Internet protocol design and resource allocation.

The focus of our research has been the enhancement of the traditional view of QoS with a user-level defined *Quality of Perception* (QoP). This is a measure that encompasses not only a user's satisfaction with multimedia clips, but also his/her ability to perceive, synthesize and analyze the informational content of such presentations. As such, we have investigated the interaction between QoP and QoS and its implications from both a user perspective as well as from a networking angle.

## 2. APPROACH

Our approach has been mainly empirical. Users were presented with a set of 12 short (30 - 45 seconds'

duration) multimedia clips in MPEG-1 format. The multimedia clips selected could have been limited to a typical education scenarios such as a classroom scene where a professor can be seen teaching or to a slide presentation containing text and diagrams. The research was not limited to such traditional education sessions for three reasons: firstly, multimedia opens up the range of types of presentations that can be applied to the education domain. For example, a documentary type of presentation or even a commercial style of presentation may be used to get a point across to a student. Secondly, education covers a wide range of subjects and fields. For example, a music class may wish to view a video clip about a chorus or a pop group. Thirdly, we also wished to investigate if the type of multimedia clip (or the contents of the presentation) has an impact on user perception and understanding/assimilation. For example, would a documentary style of clip be educationally more beneficial than a 'pop' style of multimedia presentation containing rich visual effects, dynamic scenes and a wide range of sounds. The multimedia clips were therefore chosen to be as varied as possible, ranging from a relatively static news clip to a highly dynamic rugby football sequence. Each clip was shown with the same set of QoS parameters. 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).

A classification of the clips from a QoP viewpoint is given in Table 1. As has been mentioned, an average multimedia user will find it difficult to classify clips according to the impact that the various QoS parameters have on its transmission. It is much more likely that (s)he will be able to classify such clips according to the relative importance of the video, audio and textual components in the context of a clip, as well as how dynamic the clip is, and this is the approach that we have adopted. We have thus assigned a score of 0, 1, or 2 according to the relevance (with 0 being least relevant and 2 being most) of the three identified components, and have thus obtained the above classification.

The clips themselves were chosen to cover a broad spectrum of subject matter in which the following factors were specifically taken into account:

- spatial parameters (intraframe)
- temporal parameters (interframe)
- importance of audio information in the context of the clip
- importance of the video information in the context of the clip

- importance of textual information in the context of the clip

Multimedia QoS parameters can be broadly classified into temporal and spatial parameters. In the case of audio, temporal parameters would include the sample rate, while spatial parameters might refer to mono sound. Because of the relative importance of the audio stream in a multimedia presentation (Kawalek 1995) as well as the fact that it takes up an extremely low amount of bandwidth compared to the video component, it was decided to transmit audio at full quality during the experiments. Moreover, bandwidth is the main resource we are interested in using more efficiently. As the audio stream occupies a very small bandwidth of the multimedia clip, compression at this level will not result in major gains in bandwidth. 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

VIDEO CATEGORY	Dynamic	Audio	Video	Text
1 - Action Movie	2	1	2	0
2 - Animated Movie	1	1	2	0
3 - Band	1	2	1	0
4 - Chorus	0	2	1	0
5 - Commercial	1	2	2	1
6 - Cooking	0	2	2	0
7 - Documentary	1	2	2	0
8 - News	0	2	2	1
9 - Pop Music	1	2	2	2
10 - Rugby	2	1	2	1
11 - Snooker	0	1	1	2
12 - Weather Forecast	0	2	2	2

Table 1 Video categories

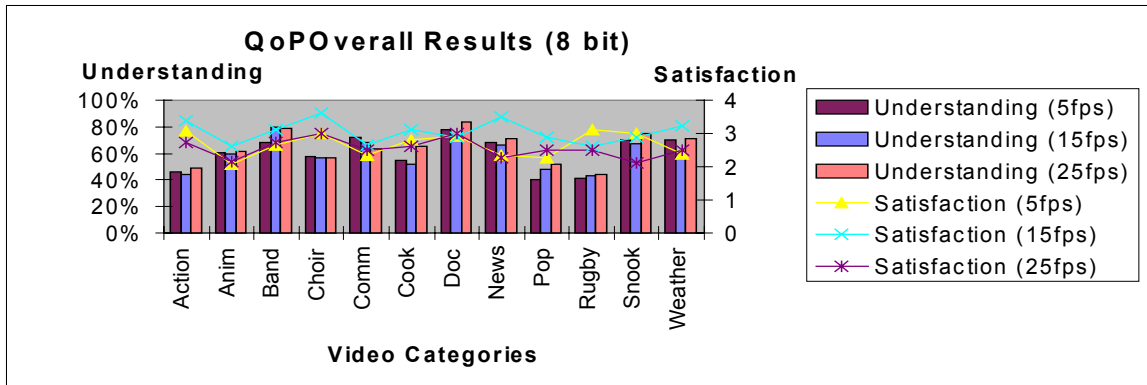


Figure 1 Quality of Perception at 8 bits

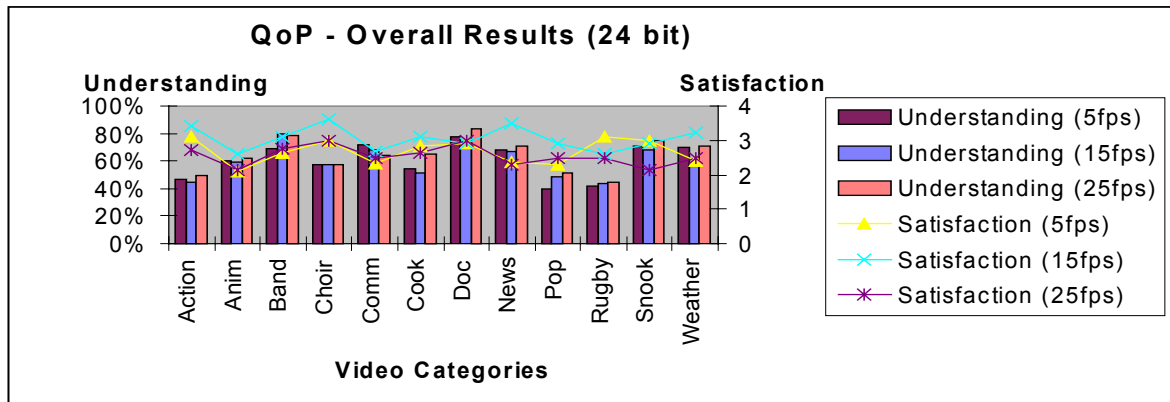
second - fps). A total of 10 users have been tested for each (*frame\_rate*, *colour\_depth*) pair.

For each clip, the questions were chosen to encompass all aspects of the information - audio, visual or textual - presented in the clips. In addition to this, some questions could only have been answered if the user had grasped pieces of both visual and audio pieces of information from the clip. Other questions, as will be shown later, were also chosen to see what was perceived as being the feel or the atmosphere of the clip. In order to be confident that the results were based purely on variations in the frame rate, questions were asked immediately after each clip so that the information contained was still fresh in the memory of the participants. Lastly, although there were no 'trick' questions as such, quite a few of them could not be answered by observation of the video alone, but by the user

making inferences and deductions from the information that had just been presented.

### 3. RESULTS

A few interesting remarks must be made about how the users answered some of the questions. The commercial video clip, for instance, is of a washing liquid for bathrooms and depicts a couple extolling its qualities. One of the questions which were asked was what the user thought the relationship between the couple was. The clue here was that there was a shot of the man's hand cleaning the bathroom in which a wedding ring could clearly be distinguished. What was, however, interesting was that 96% of the tested users said that the couple in the ad must be married. In all such cases bar one, this answer was given not because the participants had seen the wedding ring, but for a plethora of other reasons:



**Figure 2 Quality of Perception at a colour depth of 24 bits**

- this was the way in which they had perceived the situation to be
- this was the target market they felt was being addressed, or, lastly,
- some respondents gave the correct answer because the couple had appeared in previous commercials advertising the product, and in those commercials it had been made clear that the couple was married.

Another observation concerns the manner in which users answered questions regarding the cooking clip. Here, respondents were asked if they thought that the meal being cooked was going to be spicy hot. The clue was that the participants in the program had mentioned that they were going to use Tabasco, so therefore one would reasonably expect that the meal being prepared was going to be hot. A sizeable proportion (36%) of participants got the answer right, but for a rather different reason - they had seen a red sauce being prepared during the course of the clip, and they associated such sauces with chilli food.

Our results show that there is no significant difference between the percentage of correct answers given by respondents at different video frame rates or color depths. In the former case, the results would seem to indicate that severe frame dropping does not have a proportional impact on users' capacity to assimilate video clip material. Indeed, in some cases, the percentage of correct responses was marginally higher at lower frame rates. This could be explained by the fact that the complementary process to frame dropping is one of frame replication. Due to the latter, information that might had been lost had the clip been played with its designated frame rate, would now appear for a longer period of time (3 or even 5 times longer in the case of our experiments) on the screen. This would therefore increase the chance of the user noticing the respective information.

The fact that colour depth has no impact on the QoP leads to the conclusion that human users quickly ignore any annoying or distracting factors associated with a reduction in the quality of spatial parameters. Users then go on to focus on the application at hand as if it were transmitted with normal QoS.

As expected, the lowest percentages of correct answers were given in action clips with rapidly varying scenes - the action movie, the rugby clip - or those boasting a rich diversity of informational content such as the pop clip. In the action movie, for instance, one of the main events was an explosion in a communications centre. Although the cause of the explosion - a grenade - is clearly discernible, many people simply said that they were engulfed by the clip and hadn't actually noticed a grenade being thrown. When clip scenes are varying rapidly, it is of course difficult to get any sort of visual information, the most one can do is abstract the message of the clip. The fact that frame dropping has little impact here should not, therefore, surprise. To illustrate this, in the case of the rugby clip, subjects were asked which team had won ball possession from a line-out. Many of the people interviewed answered correctly, but not because they had actually remembered that particular team winning the ball. Rather, the respective team had scored a try soon afterwards, so it was natural to assume that it was they who had won the line out.

In the pop clip, in addition to the audio (of primary importance in this case) and video streams (where the body language and demeanour of the singer also tries to convey a message), textual information about the singer was regularly displayed on the screen. In the case of informationally rich clips, what usually happens is that users cannot distribute their attention. For example, frequent remarks in the case of this clip were that "I was enjoying the music and wasn't interested in the text" or, alternatively, that "the text was enormously

distracting”, which would explain why respondents got such low percentages of correct answers in this case.

Since the audio stream is unaffected by frame dropping, one would initially expect that participants would score much better when asked about the audio content of the clip. This is, generally speaking, the case. However, there are some exceptions. For instance, in the pop clip, when asked questions pertaining to the lyrics of the song, many people said that they hadn't paid attention to the lyrics themselves, as they were enjoying the melody in general. This happened especially when the clip was run at the full 25 fps and 24-bit colour depth - probably people were enjoying the overall quality of the clip, without giving regard to specifics.

As far as the satisfaction associated with media clips is concerned, a few observations need to be made. The first is that, even though users were instructed not to let their personal bias towards the subject matter being shown influence their decision, many of them gave better quality ratings to types of multimedia clips which they confessed to liking. Conversely, clips which the users thought uninteresting were given low marks, even though they might have been transmitted at the maximum possible quality. Generally speaking however, the lower the frame rate is, the lower the user's satisfaction with it, although the variation is not linear. Users seem to have enjoyed the animated clip at the expense of assimilating the data; a similar remark can be made about the rugby clip. As far as users' perception of dynamism goes, this latter clip and the action movie received similar across the board satisfaction ratings. Lastly, as concerns the news clip, users were annoyed at the newscaster's visible lack of lip synchronisation and thus, even though it was a static clip, only gave it average values as far as satisfaction is concerned. Users then essentially treated the bulletin as an audio broadcast, proof being the consistently high percentages of correctly answered questions related to the audio stream of this particular clip.

#### 4. CONCLUSIONS

This paper defines Quality of Perception as a novel term comprising a user's perception of multimedia presentations together with the benefit of such presentations from a user's angle in terms of content assimilation and understanding. The main conclusions drawn from this work may be summarized as follows:

- A significant loss of frames (that is, reducing the frame rate) does not proportionally reduce the user's understanding and perception (and thereby learning) 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.

- 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 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 synchronization), 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 seem 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 is 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. The above results hold implications on education over the Internet. For example, the results indicate that degradation in quality of the multimedia teaching material does not result in a corresponding reduction in the knowledge gained. The results also show that teaching material rich in multimedia content or in dynamicity may not aid the learning process. The first result in particular showing that a significant loss of frames does not proportionally reduce the user's understanding and perception of the presentation is also relevant to education over mobile devices where bandwidth limitation is a serious problem. Future work will focus on determining the limits below which user perception and information assimilation become unacceptable. Work will also concentrate on the impact of content in multimedia clips on user perception and understanding. The impact of different media on user perception/assimilation is another areas for further research. The research will also be extended to measure

the effects of Internet protocols such as TCP/IP on user perception and understanding. Would other protocols yield better results for perception (and thereby education) than existing protocols such as TCP/IP? This research also serves to show students the importance of networking and system parameters on the quality of a multimedia display and ultimately user perception/learning. The above research is based on real users rather than simulations or a theoretical model. Moreover, the research has focused on a wide range of multimedia documents rather than being limited to traditional teaching modes. This makes the research generic as no assumptions are made about the types of multimedia clips that may be applied in education.

## 5. REFERENCES

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

Fukuda, K., Wakamiya, N., Murata, M., and Miyahara, H., 1997, "QoS Mapping between User's Preference and Bandwidth Control for Video Transport", in *Proceedings of the 5<sup>th</sup> International Workshop on QoS (IWQoS)*, New York, USA, May 21-23, pp. 291 – 301

Kawalek, J., 1995, "A User Perspective for QoS Management", in *Proceedings of the QoS Workshop aligned with the 3<sup>rd</sup> International Conference on Intelligence in Broadband Services and Network (IS&N 95)*, Crete, Greece.

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