An Eye Opener: Low Frame Rates do not Affect Fixations

S.R. Gulliver  
G. Ghinea

Department of Information Systems and Computing  
Brunel University, United Kingdom.  
{Stephen.Gulliver/George.Ghinea}@brunel.ac.uk

Abstract

Monitoring eye movement offers insights into visual perception, as well as the associated attention mechanisms and cognitive processes. This paper presents the results of an empirical study that investigated the impact of differing multimedia presentation frame rates on the user perceptual experience, as reflected through video eye path data. Results show that varying frame rates do not affect the median coordinate value of user eye gaze and therefore highlight that selection of visual cues, which defines a users region of interest, is not adversely affected by frame rate variation.

1. Introduction

Monitoring eye movements offers insights into visual perception, as well as the associated attention mechanisms and cognitive processes. Eye movement data can be interpreted based on the empirically validated assumption that the location of a person's gaze corresponds to the symbol currently being processed in working memory [6]. Moreover, the eye naturally focuses on informative areas [9] and allows data to be collected with a fine temporal grain without subjects needing instruction or training to produce informative data [12].

The study described in this paper uses eye tracking to study the impact that varied video frame rate has on user eye position. The underlying premise of our research is that degradation in the network QoS results in loss of quality of the multimedia presentation. By monitoring user eye movements, we have a more complete indication of how user information assimilation varies due to varying QoS. For example, drifting video eye paths could suggest a loss of concentration, possibly due to low QoS levels and user intolerance to them. Eye tracking thus serves to pinpoint the thresholds at which varying network QoS affects the user multimedia experience.

The structure of this paper is as follows. Section 2, considers the relationship between eye movement and user perception of multimedia. Section 3 describes the empirical study undertaken as part of our research, while Section 4 presents the main results obtained. Finally, in Section 5, conclusions are drawn and avenues for future research based on our findings are proposed.

2. Eye Tracking and User Perception of Multimedia

Eye tracking is increasingly being used as a tool for obtaining information about human perceptive and cognitive processes[7][12]. Thus, eye tracking is being employed in the design of user interfaces, as an efficient interface ensures, for instance, that commonly-used controls are located in areas where the eyes’ gaze is most likely to rest, and that eye movement between these controls is minimal. Moreover, eye-based interfaces also help users (especially disabled) to execute interface input actions, such as menu selection [1], eye-typing [13], and even mouse clicking, through the development of an ‘eye-mouse’ [8]. Web design guidelines based on results obtained using eye tracking technology have also been elaborated and are being used by commercial web designers to write more effective web pages [10]. Eye tracking is also currently being used in virtual reality-based education and training, ranging from such diverse topics as aircraft inspection [3] to driving [14]. However, in the context of this paper, we are interested in the relationship between eye movement and user perception of multimedia.

The relationship between eye movement and user perception of multimedia has been investigated in [4] and [11]. The former study explores both visual attention (given by eye tracking patterns) and information recall of subjects being presented with a single multimedia educational application, displayed with optimum QoS parameters. The authors then went on to propose a series of guidelines to be used in web animation based on ‘contact points’ (co-references between text and animation obtained from the initial eye tracking study) [4]. The first study, however, only focuses on one particular type of multimedia category with constant, optimum QoS parameters, which fails to reflect the multitude of multimedia applications and the variety of prevailing network conditions that exist in a real life distributed system. The second study investigated the effect that multi-resolution displays have on a user’s perception. The idea behind this piece of research was to reduce the use of resources by presenting a non-uniform level of visual detail across the whole display area, but rather to render a high level of visual detail only around the centre of the user’s gaze. Whilst a spatial QoS parameter was indeed varied, based on user eye gaze, this study does not consider the range of media content diversity that
exists in the real world environment. To the best of our knowledge no one has examined the relationship between eye location and varying QoS across a diverse range of multimedia content.

3. EXPERIMENTS

3.1 Participants

Our study involved 36 participants, who were evenly divided into six experimental groups. Participants were aged between 21 and 55 and were taken from a range of different nationalities and backgrounds. All participants spoke English as their first language or to a degree-level qualification, were computer literate, and were presented with a series of 12 windowed MPEG video clips with continuous audio, each between 31 and 45 seconds long (simulating video found on the internet).

3.2 Experimental Material

The multimedia video clips used in this experiment were chosen to cover a broad spectrum of infotainment subject matter. Multimedia video clips vary in nature from those that are informational in nature (such as a news/weather broadcast or a documentary) to ones that are usually viewed purely for entertainment purposes (such as an action sequence, a cartoon, a music clip or a sports event). Moreover, specific clips were chosen as a mixture of the two viewing goals, such as a cooking clip or a television commercial.

3.3 Experimental Set-up

In our experiment, only one QoS parameter – frame rate – was varied. We were particularly interested in frame rate, as the frame rate with which a multimedia presentation is shown is the one parameter that has the greatest bandwidth implications in today’s distributed multimedia systems – and bandwidth is arguably the most scarce networking resource in such environments. Accordingly, a within-subjects design was chosen, where participants viewed four video clips at 5 fps, four at 15 fps, and four at 25 fps. Moreover, in order to counteract any possible order effects, the video clips were shown in a number of order and frame-rate combinations.

To ensure that experimental conditions remained constant throughout, consistent environmental conditions were used for all participants. An Arrington Research Power Mac G3 (9.2) based ViewPoint EyeTracker was used in combination with QuickClamp Hardware - see Fig. 1. To avoid audio and visual distraction a dedicated, uncluttered room was used throughout all experiments. To limit physical constraints, except from those imposed by the QuickClamp hardware, tabletop multimedia speakers were used instead of headphone speakers. A consistent audio level (70dB) was used for all participants. State transition scripts were developed and implemented in the ViewPoint software. Transition scripts allow movement through a number of defined states and are dependent on participant key presses. They allow each experiment to proceed at a flexible rate, marking relevant experimental points, such as the start of the video or a key-press, on the stored eye-tracking data file.

Figure 1: Use of Power Mac G3 Eye Tracker

3.4 Experimental Process

Each participant was asked a number of short questions concerning their sight, which was followed by a basic eye-test to ensure that all participants would be able to view the eye-tracker screen without spectacles. Participants wearing contact lenses were not asked to remove lenses, however, special note was made and extra time was given when configuring the eye tracker to ensure that pupil fix was maintained throughout the entire visual field. Once configuration set-up was complete, automatic calibration was made using a full screen stimulus window. However, point re-calibration was also used if an error, such as head movement, caused a non-smooth pupil mapping in eye-space window.

When calibration was complete, eye-space settings were stored. Before viewing each multimedia video the participant was asked to get into a comfortable position and place his/her chin on the chin-rest. Errors, caused due to slight shifts in head position, were corrected and the participant was reminded to keep his/her head still for the duration of the video. No defined task was given to the participants, thus allowing spontaneous looking.

3.5 Extracting Frame Based Eye-Tracking Coordinates

Data captured in each eye-tracking data sample included: frame number, x coordinate (range: 0-10000) and y coordinate (range: 0-10000). Eye-tracking data
was sampled at 25Hz, which corresponds to the maximum frame rate displayed.

Frame based eye-tracking coordinates for each of the participants was saved in a file for each of the video clips presented. Data relating to specific videos, at defined frame-rates, were combined so that x and y coordinates for participant relating video frames could be analysed (Fig 2).

4. RESULTS

To allow comparison of eye-position between frame rates (5, 15 and 25 fps) over the duration of the video clip (between 650-1000 frames, depending on the video clip), only three coordinate points were required for each frame, each one relating to a specific participant group (5, 15 and 25 fps). As we are not aware of any previous eye-tracking data analysis across multiple frames, no precedent for summarising participant group data has been defined.

Therefore, to avoid inclusion of extreme outlying points whilst removing unwanted data, such as error coordinates as a result of participant blinking, we determined that the coordinate points for each video frame should be the median value of the data within each of the participant groups (5, 15 and 25 fps). By mapping these x and y median coordinate values in time we were able to calculate the median eye-path through each multimedia video clip (video eye-path), for each of the available frame-rates (Fig. 3).

Statistical correlations (Kendall’s tau-b and Spearman’s 2-tailed nonparametric tests) between median coordinate values, for frame rates of 5, 15 and 25 fps (i.e. 5fps compared to 15 fps, 5 fps compared to 25 fps, and 15 fps compared to 25 fps), were then carried out for all of the 12 multimedia video clips (72 tests in total). This allows us to determine whether x and y coordinate values, from specific video clips shown at varied frame rates, statistically correlate, i.e. similar median trends of eye movement occur for groups of people shown the same video clip at different frame rates. All 72 correlation tests showed a significance value of \( p < 0.001 \) between the video eye-paths across the different frame rates, which implies a strong correlation between the median eye-position of participants, independent of the frame rate.

This result shows that, for median coordinate values mapped across time, eye movement significantly correlates independent of the video frame rate viewed. With such strong correlation between participants and the fact that strong correlation exists for each of the diverse multimedia video clips, we can conclude that frame rate does not significantly impact median video eye-path. At first sight, it seems to be counterintuitive, as one would expect that less smooth multimedia video presentations would be perceptually more annoying and therefore lead to bigger variability in video eye paths. However, our results suggest that lower frame rate video does not adversely impact the flow of information assimilation. This result reinforces the conclusions of other studies \([1][5]\) (which, however, did not examine video eye paths) and is of particular interest, for it suggests that selection of visual cues, which defines a users region of interest is not adversely affected by frame rate variation. Moreover, it also has significant implications in bandwidth-constrained environments, for it highlights the fact that multimedia video presentations need not necessarily be transmitted with full quality (in our case, best possible frame rate) in order for users to follow the same input stimuli.

5. CONCLUSIONS

In this paper, we have used eye-tracking data to investigate the impact of varying multimedia presentation frame rates on the user perceptual experience, by analysing user gaze location.

Our results show that the median video eye paths are not significantly affected by varying frame rates, for video 30-45 seconds in length. Further study is required to identify whether a match across eye paths degrades over longer durations. Bearing in mind that the location of eye gaze can be considered to indicate the object currently being processed in working memory, this finding indicates that users process information similarly, irrespective of the underlying video frame rate. This suggests that the location of a user’s focus of attention does not significantly change if (s)he is presented with what is technically recognised as a lower quality multimedia presentation (i.e. a multimedia presentation with a reduced frame rate). This finding support previous work \([5]\) and suggests that selection of visual cues, which defines a users region of interest, is not adversely affected by frame rate variation.
Figure 3: Rugby median Y-coordinate eye-path (5, 15 and 25 fps).

6. REFERENCES


