Pupil Dilation as an Indicator of Cognitive Workload in Human-Computer Interaction

Marc Pomplun and Sindhura Sunkara

Department of Computer Science, University of Massachusetts at Boston
100 Morrissey Boulevard, Boston, MA 02125-3393, U.S.A.
Email: marc@cs.umb.edu, ssunkara@cs.umb.edu

Abstract

Pupil dilation is known to quickly respond to changes in the brightness in the visual field and a person's cognitive workload while performing a visual task. Pupil dilation is rarely analyzed in usability studies although it can be measured by most video-based eye-tracking systems and yields highly relevant workload information. This is mainly due to two problems: First, the variety of factors that can influence pupil dilation, and second, the distortion of pupil-size data by eye movements: The size of the pupil as seen by the eye-tracker camera depends on the person's gaze angle. In the present study, we developed and implemented a neural-network based calibration interface for eye-tracking systems, which is capable of almost completely eliminating the geometry-based distortion of pupil-size data for any human subject. Moreover, we compared the effects of cognitive workload and display brightness on pupil dilation and investigated the interaction of these two factors. The results of our study considerably facilitate the use of pupil dilation as a quick and reliable indicator of a person's cognitive workload.

1 Introduction

In the evaluation of human-computer interfaces, an increasing number of researchers conduct analyses of users' eye movements during task completion (e.g., Goldberg & Kotval, 1999). Gaze trajectories can indicate difficulties that users encounter with certain parts of the interface and point out inappropriate spatial arrangement of interface components. However, when performing such studies, scientists often neglect the analysis of another variable that they receive as a "byproduct" of video-based eye tracking, namely the size of the user's pupil.

It is well known from a variety of studies that participants' pupils dilate with increasing cognitive workload being imposed (see Kahneman, 1973). This effect has been demonstrated for tasks such as mental arithmetic (Hess, 1965), sentence comprehension (Just & Carpenter, 1993), and letter matching (Beatty & Wagoner, 1978). Besides cognitive workload, the intensity of ambient illumination is the other major factor determining the size of a person's pupil. Changes in illumination can therefore interfere with the use of pupil size as a measure of cognitive workload (Kramer, 1991). To reliably measure workload, we have to compensate for such changes in illumination (Nakayama, Yasuike & Shimizu, 1990; Porter, Troscianko & Gilchrist 2002). Furthermore, scientists face a technical problem: Since participants move their eyes during experiments, their pupils assume different angles and distances towards the monitoring camera of the eye tracker. This, in turn, means that the size of the pupil as measured by the system - the number of pixels that belong to the pupil in the camera image – varies with the participant's gaze angle. This effect is especially strong if the camera is located below the eye (see Figure 1).

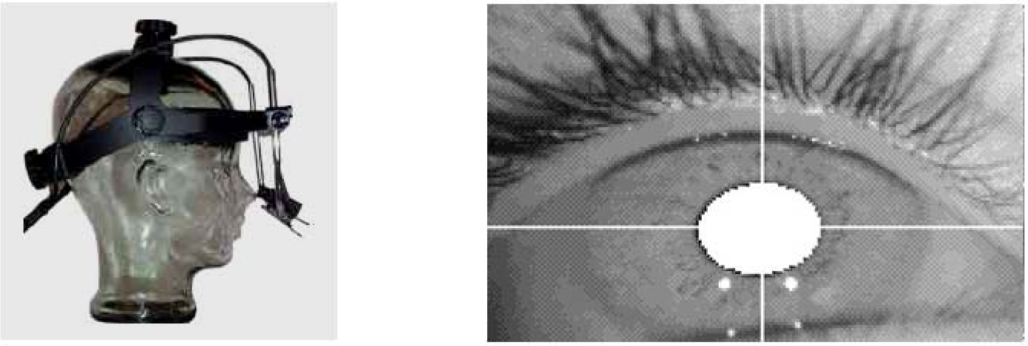

Figure 1: Left panel: The headset of the EyeLink-II system; Right panel: Camera image of a participant's left eye with the pupil area recognized by the system (white)

In order to reduce the noise in pupil size measurement caused by eye movements, we implemented a neural-network based calibration interface for video-based eye trackers and evaluated it empirically in Experiment 1. Using the increased precision achieved by the new interface, in Experiment 2 we investigated, from a practical perspective, in which way the brightness of the screen in a human-computer interaction task interferes with the measurement of cognitive workload as indicated by pupil size and whether this interference can be substantially reduced.

2 Experiment 1: A Pupil Calibration Interface and its Evaluation

Since the setup of the eye tracker - that is, the camera position and orientation relative to the participant's eye – is different for every experimental session, it is not feasible to use a fixed geometric calculation for correcting the measured pupil size. Instead, we introduced a pupil calibration procedure prior to the experiment to determine the relative size of the pupil as a function of the participant's gaze position. Participants were asked to fixate on each point in a 3×3 array four times to collect pupil size data for these 3×3 gaze positions. We chose to use only nine calibration points to make the calibration procedure as quick and little disruptive as possible. Given the continuous small changes in pupil size and the resulting variance, additional calibration points would not have led to a substantial improvement of the calibration.

Obviously, interpolation is necessary to estimate, based on the calibration data, the change in the measured pupil size as a function of the current gaze position. For such interpolation tasks, a type of artificial neural network called Parametrized Self-Organizing Map (PSOM) has proven to be well-suited (Pomplun, Velichkovsky & Ritter, 1994). PSOMs are a variant of the Self-Organizing Maps (Kohonen, 1990), but learn much more rapidly than the latter ones and are capable of representing continuous, highly non-linear functions. In the present context we used a PSOM with nine neurons and fed it with the calibration data, that is, the measured average size of the pupil at the nine calibration points, divided by the pupil size measured while looking at the center of the screen. During the subsequent experiment, by interpolating the calibration data, the PSOM estimated the factor by which the measured pupil size differed from the one that would have been measured if the subject had looked at the center of the screen. Then the currently measured pupil size was divided by the PSOM's output and thereby standardized, which we assumed to strongly reduce the variance in pupil size data that is due to eye movements. We conducted Experiment 1 in order to test the effectiveness of our calibration interface at improving the signal-to-noise ratio when measuring the effect on pupil size exerted by changes in display brightness.

2.1 Method

Participants. Ten students from the University of Massachusetts at Boston were tested individually. All participants had normal or corrected-to-normal vision. They were naïve with respect to the purpose of the study and were paid for their participation.

Apparatus. Eye movements were recorded with the SR Research Ltd. EyeLink-II system (see Figure 1), which operates at a sampling rate of 500 Hz and measures a participant's gaze position with an average error of less than 0.5 degrees of visual angle. Stimuli were presented on a 21-inch Dell Trinitron monitor with a refresh rate of 85 Hz and a screen resolution of 1152 by 864 pixels.

Materials. The stimulus displays showed the numbers from one to 16 arranged in a 4×4 array spanning almost the entire screen. None of the 16 positions coincided with any of the nine target positions used for calibration. Two different displays were created: One showing white numbers on a black background (luminance < 1 cd/m^2) and another one presenting black numbers on a white background (luminance 82.4 cd/m^2).

Procedure. Each participant was sequentially presented with the two stimulus displays. The order of presentation was counterbalanced across participants. They were asked to find the numbers in ascending order and read them out loud. Subsequently, participants were asked to repeat the task, but this time in descending order.

2.2 Results

All pupil size data, both the uncorrected and the corrected ones, were separated into 16 groups based on the participant's gaze position during their measurement. For this purpose, the screen area was divided into four by four equally large rectangular parts. A two-way analysis of variance (ANOVA) with the factors background color (two levels: black and white) and gaze position (16 levels) revealed significant effects by background color, $F(1; 9) = 32.82$, $p < 0.001$, and gaze position, $F(15; 135) = 22.30$, $p < 0.001$, as well as a significant interaction between the two factors, $F(15; 135) = 2.00$, $p < 0.05$. The gaze-position and interaction effects demonstrate that, as predicted, the measured pupil area is systematically influenced by the participant's gaze position. Figure 2 (left) illustrates this finding.

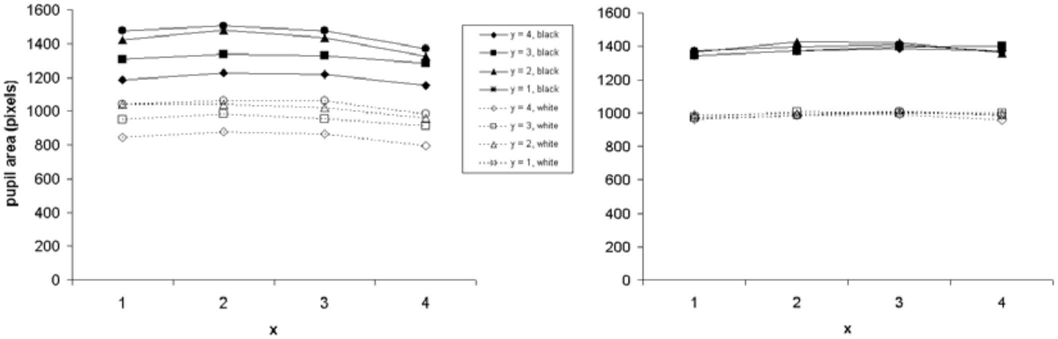

Figure 2: Measured average pupil size by gaze position ($x = 1, \ldots, 4$; $y = 1, \ldots, 4$) and background color before correction by the PSOM (left panel) and afterwards (right panel).

An analogous ANOVA for the corrected pupil size data also showed a significant effect by the factor background color, $F(1; 9) = 33.21$, $p < 0.001$, but no significant effect by gaze position, F(15; 135) < 1, and no interaction effect, F(15; 135) < 1. This indicates that our calibration interface greatly reduced the systematic influence of the gaze position on the pupil size measurement (see Figure 2, right).

3 Experiment 2: Workload and Brightness Effects on Pupil Dilation

To investigate brightness and cognitive workload effects on pupil size in human-computer interaction, we devised a gaze-controlled human-computer interaction task that ran in three different speeds, thereby creating three different levels of task difficulty and, assumedly, cognitive workload.

3.1 Method

Participants. The same ten participants from Experiment 1 also participated in Experiment 2.

Apparatus. The apparatus was the same as in Experiment 1.

Materials. The stimulus displays showed a grid of 4×3 cells (see Figure 3, left). At the beginning of a trial, all cells were empty. Then, in each cell, one of four possible items could appear: a red square, a red circle, a blue square, or a blue circle. These items then increased in size twice before they disappeared. The participants' task was to avoid any blue circles from attaining their maximum size. To achieve this, they could look at any growing blue circle and press a designated button at the same time to eliminate that item. Any failure caused a loud buzzer sound to be played. In the "easy" condition, every second one cell was randomly chosen to be updated, that is, if it contained an item, this item would grow (or disappear if already fully-grown), otherwise a new, small item of random type would be placed in the cell. In the "medium" and "hard" conditions, the updating interval was reduced to 200 and 75 milliseconds, respectively.

Procedure. Each of the three levels of task difficulty was combined with two levels of background brightness (black and white, as in Experiment 1), resulting in six different trial types. Each type was presented to each participant four times. Before the experiment, participants were instructed not to let any blue circle reach its full size. The experiment started with an easy practice trial whose data were not analyzed, followed by the 24 experimental trials in random order. Each trial lasted 30 seconds.

3.2 Results and General Discussion

A two-way ANOVA revealed that the (corrected) pupil size was significantly influenced by the factor task difficulty (levels easy, medium, and hard), F(2;18) = 35.13, $p < 0.001$, and the factor background color (levels black and white), F(1; 9) = 41.08, $p < 0.001$, while there was no interaction between the two factors, F (2; 18) < 1. Figure 3 (right) illustrates how the increase in pupil area induced by higher task demands was almost identical for black backgrounds (1231, 1315, and 1441 pixels) and white backgrounds (872, 961, and 1102 pixels).

This finding suggests a method for accurate cognitive workload measurement even in situations where the display brightness cannot be kept constant. The idea is to perform an additional calibration procedure in which the display brightness - in the same display that will be used in the subsequent experiment - is systematically varied to determine the participant's pupil size as a function of brightness. During the following experiment, by subtracting the calibration value for the current display brightness from the currently measured pupil size, the amount of pupil dilation induced by cognitive workload can be computed.

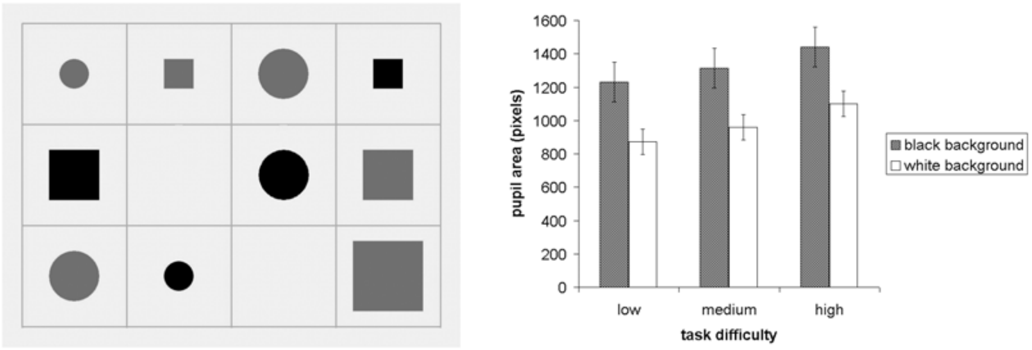

Figure 3: Left panel: Screenshot of Experiment 2 - red objects are shown in light gray, blue ones in dark gray. Right panel: Results of Experiment 2.

In summary, we have presented a technique for substantially reducing the eye-movement induced variance in video-based pupil dilation measurement. Our proposed calibration procedure takes only about 30 seconds and strongly and reliably improves measurement precision. Moreover, we have pointed out how to separate brightness effects from workload effects on pupil size. All in all, the present study can be considered a small but significant advance in using pupil dilation for the analysis of cognitive workload in human-computer interaction.

Acknowledgements. The authors would like to thank Naing Naing Maw for her valuable help with programming the experiments for the present study.

4 References

Beatty, J. & Wagoner, B.L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science, 199*, 1216-1218.

Goldberg, J.H. & Kotval, X.P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24*, 631-645.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American, 212*, 46-54.

Just, M.A. & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*, 310-339.

Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice Hall.

Kohonen, T. (1990). The self-organizing map. *Proceedings of IEEE, 78,* 1464-1480.

Kramer, A.F. (1991). Physiological metrics of mental workload: A review of recent progress. In D.L. Damos (Ed.), *Multiple-task performance*. (pp. 279-328). London: Taylor & Francis.

Nakayama, M., Yasuike, I. & Shimizu, Y. (1990). Pupil size changing by pattern brightness and pattern contents. *The Journal of the Institute of Television Engineers of Japan, 44*, 288-293.

Pomplun, M., Velichkovsky, B.M. & Ritter, H. (1994). An artificial neural network for high precision eye movement tracking. In Nebel, B. & Dreschler-Fischer, L. (Eds.), *Lecture notes in artificial intelligence: AI-94 Proceedings* (pp. 63-69). Berlin: Springer Verlag.

Porter, G., Troscianko, T. & Gilchrist, I.D. (2002). Pupil size as a measure of task difficulty in vision. *Perception 31S,* 170.