Using Pupil Size as a Measure of Cognitive Workload in Video-Based Eye-Tracking Studies

Marc Pomplun*, Sindhura Sunkara, Alexander V. Fairley, and Mei Xiao

Department of Computer Science, University of Massachusetts Boston

100 Morrissey Boulevard, Boston, MA 02125-3393, USA

*Corresponding author

Phone: (617) 287-6485

Fax: (617) 287-6433

Email: marc@cs.umb.edu

Research Article

Abstract: 175 words

Text (excluding references and figure captions): 5526

References: 19 entries

Figures: 9

Tables: 0

Abstract

Pupil size is known to quickly adapt to changes in the luminance within the visual field and the observer's cognitive workload while performing a visual task. Pupil size is rarely analyzed in eye-movement studies although it is measured by most video-based eye-tracking systems. This is mainly due to two problems: First, the dependence of pupil size on luminance across the display and second, the distortion of pupil-size data by eye movements: The size of the pupil as measured by the eye-tracker camera depends on the subject's gaze angle. In the present study, we introduce and develop measures and heuristics to estimate luminance-based changes in pupil size. Moreover, we present a neural-network based pupil calibration interface for eye-tracking systems, which is capable of almost completely eliminating the geometry-based distortion of pupil-size data. Finally, we compare the effects of cognitive workload and display luminance on pupil dilation and investigate the interaction of these two factors. The results of our study facilitate the use of pupil dilation as a reliable and inexpensive indicator of a subject's cognitive workload.

Using Pupil Size as a Measure of Cognitive Workload in Video-Based Eye-Tracking Studies

Cognitive workload is an important concept for both the study of human cognition and the evaluation of human-machine interfaces such as head-up displays in cars or air-traffic controllers' workstations. There are several common methods for measuring cognitive workload: galvanic skin response, heart rate, and electroencephalography (e.g., Kramer, 1991; O'Donnell & Eggemeier, 1986; Wilson, 2001). Rather than taking cognitive workload measurements, many researchers evaluate interfaces by analyzing 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. Interestingly, almost all of these studies use video-based eye trackers, which means that they routinely disregard an indicator of cognitive workload that they receive as a "byproduct", namely the size of the user's pupil.

It is well known from a variety of studies that an observer's 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), letter matching (Beatty & Wagoner, 1978), and visual search (Porter, Troscianko & Gilchrist, 2007). Besides cognitive workload, also emotional factors - such as the emotional content of written words - influence pupil size (Võ et al., 2008). However, in typical laboratory tasks, emotional influence can easily be reduced so that it does not significantly interfere with cognitive workload measurement. Unfortunately, such control is more difficult to achieve for the third factor determining pupil size - the illumination of the observer's visual field (Reeves, 1920). If changes in illumination need to occur during experimental sessions, we can expect substantial interference with the use of pupil size as a measure of cognitive workload. To reliably measure workload, we have to account for such changes in illumination (Nakayama, Yasuike & Shimizu, 1990). Unfortunately, the only current approach to this problem is to control the illumination of the display (e.g., Porter, Troscianko & Gilchrist 2002; 2007), which is not always possible when evaluating interfaces.

Furthermore, scientists face a technical problem: Since participants usually move their eyes during experiments, their pupils assume different angles and distances toward 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 in the camera image covered by the pupil or an ellipse fitted to it – varies with the participant's gaze angle. This effect is especially strong if the camera is located below the eye, which is the case for most remote eye trackers (using desktop cameras) and head-mounted systems. Consequently, these systems report considerably larger average pupil size while subjects fixate targets at the bottom of the screen than when their gaze is near the top of the screen. This effect makes it impossible to measure pupil size consistently in tasks involving eye movements. Even when only the average or maximum pupil size during a trial is of interest, any systematic difference in the distribution of fixation positions across experimental conditions would invalidate the pupil size measurements.

To tackle these problems, the present study provides methods for both dissociating the pupil's responses to light versus cognitive workload, as well as for deriving a gaze-position invariant measure of pupil size. In Experiment 1, the basic pupillary response to different levels of luminance on a standard CRT screen was obtained. Based on the results, we propose the pupil constriction index as a suitable measure for the pupil's response to screen luminance. Experiment 2 examined the effect of stimulus color – its red, green, and blue components displayed by the monitor – and eccentricity on pupil size. The results are used to develop heuristics for estimating the pupil's light response to any given display. Moreover, we introduced a neural-network based pupil calibration interface and evaluated it empirically in Experiment 3. It is shown that this technique greatly reduces the distortion of pupil size measurement by gaze shifts and thereby provides a valid pupil size measure for tasks involving eye movements. In the final Experiment 4, the neural-network interface is used to analyze cognitive workload in a monitoring task. The findings suggest that illumination and cognitive workload control pupil size in distinct ways, which needs to be considered when dissociating these two factors.

Experiment 1: Pupillary Response to Changes in Luminance on a Computer Monitor

While the response of the pupil to changes in illumination has already been measured in previous studies (e.g., Reeves, 1920), the main purpose of Experiment 1 was to derive useful measures for estimating the pupillary response to changes in luminance on a standard CRT computer monitor. Another aim of this experiment was to determine the time course of this response.

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, 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 calibrated 19-inch Dell Trinitron CRT monitor with a refresh rate of 85 Hz and a screen resolution of 1024 by 768 pixels (CIE chromaticity values: red: $x = 0.625$, $y = 0.340$; green: $x = 0.275$, $y = 0.605$; blue: $x = 0.150$, $y = 0.065$; color temperature: 9300 K). The subjects were seated at an eye-screen distance of 50 cm.

Materials. The stimulus displays showed a plus sign (approximately 1° in diameter) centered on a gray disc (28° in diameter) on a black background (0.2 cd/m^2). The disc assumed one of 15 different luminance levels (0.2, 5.2, 10.2, …, 70.2 cd/m^2). The plus sign was white for disc luminance below 35 cd/m^2 and was black otherwise.

Procedure. Prior to the experiment, subjects performed a 9-point calibration procedure. The subsequent experiment consisted of 150 trials, which presented each disc luminance level ten times. These trials were presented in randomized order. The subjects started each trial by looking at a central drift correction marker – a white marker on a black background (0.2 cd/m^2) - and pressing a button on a game pad. Every trial lasted for 12 seconds or until the subjects shifted their gaze away from the plus marker by more than 1.5° of visual angle (gaze error). Trials with gaze error were presented again later in the experiment, and the data recorded during their first presentation were excluded from analysis.

Results

To obtain an estimate of the time it takes for the pupil to adjust its size to a change in screen luminance, we collapsed the data from trials across all subjects and all luminance levels except for the black disc condition (0.2 cd/m^2). All of these trials presented an increase in luminance from the black drift correction screen to the stimulus screen showing a disc of luminance $\geq$ 5.2 cd/m^2. Consequently, we expected a decrease in pupil diameter. We normalized the data by scaling them linearly in such a way that the average pupil diameter during the first ten and the last ten data samples (20 ms each) was 1 and 0, respectively. Figure 1 shows the average normalized pupil diameter as a function of time after the onset of the disc. We see that the diameter first undershoots the value 0 before asymptoting toward it. After approximately 4 s it reaches its final size.

----- insert Figure 1 about here -----

For the analysis of pupil diameter as a function of disc luminance, the relative pupil diameter was used as a measure. It was computed by dividing a given diameter by the average diameter measured for the same subject during the initial presentation of a black disc. Obviously, for the black disc condition (0.2 cd/m^2) we expected a relative pupil diameter close to 1, and decreasing diameter with greater luminance. The use of a relative measure instead of an absolute one was decided upon for practical reasons. For most eye trackers, the eye-camera distance is variable, and additional equipment is necessary to measure the absolute size of the pupil. Fortunately, the measurement of cognitive workload can be accomplished without knowledge of absolute values but through observation of relative changes in pupil diameter (e.g., Bailey & Iqbal, 2008).

Based on the above data on temporal characteristics of pupil size changes, the average pupil diameter between 4.5 and 12 s after stimulus onset was used to compute the relative pupil size for each trial. Figure 2 shows the results across all 15 luminance levels. The data conform to the well-known (e.g., Reeves, 1920) functional relationship between luminance and relative pupil diameter, which obviously has to be non-linear since the relative pupil diameter can never reach zero or become negative. For the current purpose of measuring the effect of luminance and cognitive workload on the pupil, relative pupil diameter is not an ideal variable. As can be seen in Figure 2, the same increase in luminance has a larger effect on this variable when it starts at a low luminance level than when it starts at a high one.

----- insert Figure 2 about here -----

It would clearly be desirable to have a measure that enables researchers to estimate the strength of an influence on pupil size solely based on the change in that measure. Ideally, the amount of change in this measure should be independent of other, constant factors such as the absolute illumination of the visual field. Based on the current data, and evaluated in Experiments 2 to 4, we propose computing a pupil constriction index c as such a measure. This computation has the form

$$c = e^{-k\left(\frac{d}{d_{\max}}-1\right)}, \qquad (1)$$

where d is the pupil diameter, $d_{\max}$ is the maximum pupil diameter measured during the experiment (for example, during the presentation of a black calibration screen), and k is a constant. This equation yields $c = 1$ whenever the pupil is at its maximum diameter, and with a fitted value for k it increases approximately linearly with greater screen luminance that causes the pupillary response. For our experimental setting, we found the best fit with the empirical data for $k = 9.5$, and the same value was found for a different eye tracker, the EyeLink-2k. Using Equation (1) to compute the pupil constriction index and assuming this index to vary in proportion with the luminance of the gray disc, we can determine the best fit with the current data (see curved line in Figure 2). In the following experiments, we will employ the pupil constriction index for the analysis of pupil size data.

Experiment 2: The Effects of Color and Eccentricity on Pupil Size

The pupillary response to luminance is known to decrease with greater eccentricity in the observer's visual field (e.g., Reeves, 1920). The first aim of Experiment 2 was to measure this decrease for a subject sitting in front of a standard computer monitor. These data were used to devise a simple heuristic for estimating the eccentricity effect. Second, the individual and additive effects of the monitor's three color components (red, green, and blue), and their interaction with eccentricity was assessed. Finally, we tested whether the combined effect of luminance at different eccentricities on pupil size is simply the sum of the individual effects.

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. The apparatus was identical to Experiment 1.

Materials. Each stimulus display showed a central fixation marker identical to the one used in Experiment 1. The stimuli consisted of one or more concentric, non-overlapping rings (thickness 3.5°) being illuminated simultaneously (see Figure 3) in front of a black background (0.2 cd/m^2). In the multiple-ring condition, each ring was either black (0.2 cd/m^2) or white (71.5 cd/m^2). All possible black/white ring configurations were shown, resulting in 16 different types of stimulus for the multiple-ring condition. In the one-ring condition, only one ring was shown in each trial. Its color was chosen from a set of 64 different colors, which were composed of all combinations of four luminance levels of red (0.0, 2.0, 5.8, and 12.8 cd/m^2), green (0.0, 7.1, 22.2, and 50.5 cd/m^2), and blue (0.0, 1.4, 3.8, and 8.0 cd/m^2). The luminance of a given color was given by the luminance sum of the three RGB components plus the background luminance of 0.2 cd/m^2. The four levels for each of these three constituent colors were chosen in such a way that they corresponded to the RGB channel values 0, 85, 170, and 255 for the non-calibrated monitor. These values were selected to simplify the estimation of luminance effects on pupil size for non-calibrated monitors, which are more commonly used for human-computer interfaces than calibrated ones. The chosen luminance values span the entire range that a monitor can display, and they are approximately perceptually equidistant (see Pinoli, 1997). Figure 4 (top) illustrates the luminance values across channels and levels.

----- insert Figure 3 about here -----

Procedure. Prior to the experiment, subjects performed a 9-point calibration procedure. Subsequently, the 64 one-ring stimuli and 16 multiple-ring stimuli were each shown four times. The resulting 320 trials per subject were presented in random order. Prior to each trial, a drift correction procedure identical to Experiment 1 was performed. Based on the temporal pupillary response characteristics obtained in Experiment 1, in Experiment 2 each trial lasted for 4.7 seconds or until the subjects shifted their gaze away from the plus marker by more than 1.5° of visual angle (gaze error). Trials with gaze error were repeated at a later time in a manner identical to Experiment 1.

Results

As motivated by the pupillary response function (Figure 1), we measured pupil diameter in each trial as the average diameter during the time interval from 4.5 to 4.7 s after stimulus onset. For each stimulus type and each subject, the pupil constriction index (see Equation 1) was computed for the average pupil diameter across the four presentations of that stimulus type. The pupil constriction indices in the one-ring condition were entered into a four-way, repeated-measures Analysis of Variance (ANOVA) with the within-subject factors red channel, green channel, blue channel, and eccentricity (four levels each). As expected, each of the three RGB channels significantly influenced the pupil constriction index, all Fs(3; 7) > 6.69, ps < 0.05. As shown in Figure 4 (bottom), for all channels, greater luminance led to greater indices. Interestingly, the factor eccentricity had no significant effect, $F(3; 7) = 1.01$, $p > 0.4$. The illumination of each of the four rings induced an average increase in the pupil constriction index by approximately 9 to 10 units (see Figure 5). This finding suggests that the greater area covered by the more eccentric rings must have roughly compensated for the eccentricity effect in the pupillary response. No interactions between any of the four factors were found, indicating that the pupil constriction indices for the individual RGB channels simply add up for colors composed of multiple channels. Moreover, this result implies that the relative pupillary response across channels does not vary with eccentricity.

----- insert Figure 4 about here -----

----- insert Figure 5 about here -----

The pupil constriction indices measured in the multiple-ring condition were analyzed using a four-way, repeated measures ANOVA with the within-subject factors ring 1 to ring 4 (two levels each: black and white). Each ring had a significant effect on the pupil constriction index, all Fs(1; 9) > 15.06, ps < 0.005. No interactions between any factors revealed an effect, indicating that the pupillary responses to illumination at different eccentricities add when this illumination occurs simultaneously. This finding can be illustrated by computing the average indices for those conditions that present no rings (one condition), one ring (four conditions), two rings (six conditions), three rings (four conditions), or four rings (one condition) in white color at the same time, while disregarding the eccentricities of these rings. Figure 6 shows that each additional ring being presented adds a value of approximately 20 units to the pupil constriction index.

----- insert Figure 6 about here -----

Discussion

The results of Experiment 2 demonstrate that heuristics can be applied to predict – at least roughly - the pupillary response to a given display and gaze position. The absence of any significant interactions between RGB channels and eccentricity – as observed in Experiment 2 – allows a straightforward computation. First, we measure the observer's pupillary responses to a black screen and a white screen with gaze fixated on a central marker. For analysis purposes, during the experiment or user session the screen is divided into rings of equal thickness centered at the current fixation point, with neighboring rings touching but not overlapping. We then use Equation (1) to compute the pupil constriction index based on the average luminance in each ring. The sum of all these indices is the overall index. By using the pupillary response to the white screen as a reference, the value of k in Equation (1) can be determined with regard to the overall index. This algorithm can then be used to predict pupil size.

When applying this algorithm to situations of free viewing, it is also necessary to consider the pupil's temporal response characteristics, as shown in Figure 1. During periods of fast saccadic eye movements, the relatively slow pupil response makes it difficult to predict precisely the instantaneous pupil size. However, the current approach is useful to predict statistical changes in pupil size induced by varying luminance across experimental conditions and display positions.

As mentioned earlier, the spatial variation of the pupillary response can also be an artifact introduced by the video-based measurement of pupil size. Experiment 3 presents and evaluates an approach aimed at eliminating those systematic distortions of the pupil size measurement.

Experiment 3: Evaluation of a Neural-Network Based Pupil Calibration Interface

Due to the underlying geometry, measuring pupil size as the number of pixels in the image of an eye camera leads to significantly different results when the observer looks in different directions. 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. To tackle this problem, we introduced a pupil calibration procedure prior to the experiment to determine the relative size of the pupil as a function of the subject's gaze position. This procedure involves only nine calibration points to make it as quick and unobtrusive as possible. Subjects were asked to fixate sequentially on each point in a 3×3 array of black markers on a gray background (35.7 cd/m^2) for five seconds to collect pupil size data for these 3×3 gaze positions. The horizontal and vertical distances between the centers of neighboring markers were 18.5° and 13.6°, respectively. The mean pupil size during the final 500 ms at each point was measured.

Clearly, 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 Parameterized Self-Organizing Map (PSOM, see Ritter, 1993) has proven well-suited (Essig, Pomplun & Ritter, 2006; Pomplun, Velichkovsky & Ritter, 1994). PSOMs are a variant of the Self-Organizing Maps (SOMs, see Kohonen, 1990), but learn much more rapidly than the latter ones and are capable of representing continuous, highly non-linear functions. Because of these characteristics and its successful application to related tasks, we used a PSOM for the pupil size calibration and correction tasks. Other approaches such as polynomial interpolation may also be adequate for solving this problem. It is not the goal of the present work to demonstrate possible advantages of the PSOM method over other techniques, but simply to present one useful way of tackling the gaze-position problem in pupil-size measurement. Detailed descriptions of the PSOM network paradigm and its application to eye-tracking related interpolation tasks are given in Essig et al. (2006) and Velichkovsky et al. (1994) and are not repeated in the present context.

In the current application, a PSOM with nine neurons received as its input the measured average size of the pupil at the nine calibration points. During the subsequent experiment, the current gaze position – as measured by the eye tracker – was continually fed into the PSOM. By interpolating the calibration data, the PSOM estimated the factor by which the current 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 thus standardized. This correction was assumed to strongly reduce the eye-movement induced variance in pupil size data. We conducted Experiment 3 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 luminance.

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. The apparatus was identical to the one used in Experiments 1 and 2.

Materials. The stimulus displays showed small black cross markers (diameter approximately 0.5°), one at a time, arranged in a 4×4 array spanning almost the entire screen. The horizontal and vertical distances between the centers of neighboring markers were 12.3° and 9.2°, respectively. None of the 16 positions coincided with any of the nine target positions used for calibration. Two different displays were created that differed in the luminance of their background – dark gray (20.0 cd/m^2) versus light gray (30.0 cd/m^2). These values were thought to induce a small but clear difference in pupil size that is measurable for all gaze positions on the screen.

Procedure. Each subject performed in four trials, two of them showing a dark background and the other two showing a bright one. The order of presentation was counterbalanced across participants. In each trial, the 16 markers were presented in random order for five seconds each. The subjects' task was to keep their gaze fixated on the currently visible marker. We measured the average pupil size during the final 500 ms of each marker's presentation.

Results

All pupil size data, both the uncorrected and the corrected ones, were separated into 4×4 groups based on the position of the marker shown during their measurement. Figure 7 (top) presents the uncorrected, average data for both the dark and the bright background conditions. These uncorrected data were entered into a three-way, repeated measures ANOVA with the factors background luminance (dark vs. bright), horizontal (x) position of the marker (1 to 4), and vertical (y) position of the marker (1 to 4). It revealed significant effects by luminance, $F(1; 9) = 13.47$, $p < 0.01$, x-position, $F(3; 7) = 6.55$, $p < 0.05$, and y-position, $F(3; 7) = 12.98$, $p < 0.005$, as well as a significant interaction between luminance and y-position, $F(3; 7) = 5.22$, $p < 0.05$. The gaze-position and interaction effects demonstrate that, as predicted, the measured pupil area is systematically influenced by the participants' gaze position.

----- insert Figure 7 about here -----

The pupil size data were corrected using the PSOM-based interface (see Figure 7, bottom panel). An analogous ANOVA for the corrected pupil size data also revealed a significant effect by the factor luminance, $F(1; 9) = 13.28$, $p < 0.01$, but no significant effect by marker position or interactions between any factors, all Fs < 1.91, ps > 0.2. This result indicates that the calibration interface effectively reduced the systematic influence of the gaze position on the pupil size measurement.

Discussion

Experiment 3 provided evidence for the neural-network based calibration method to substantially reduce the noise in video-based pupil size measurement introduced by the observer's eye movements. This is especially beneficial to situations in which cognitive workload is to be measured as a function of gaze position. For example, if researchers intend to study the cognitive workload imposed by different parts of a given Web page, the uncorrected pupil size could not be used as a measure because of its dependence on gaze position. As a matter of fact, any experimental conditions that differ in the distribution of eye movements cannot be validly compared in terms of uncorrected pupil size data. The calibration interface presented here overcomes this restriction by providing pupil size measurements that are statistically invariant to changes in gaze position. Other approaches may be able to do this task just as well – in the present work, we only presented one possible solution and demonstrated its effectiveness. The following Experiment 4 will finally address the problem of measuring cognitive workload by examining the combined effect of luminance and workload in an interactive task.

Experiment 4: Luminance and Cognitive Workload Effects on Pupil Size

To investigate the effects of luminance and cognitive workload on pupil size, we devised a gaze-controlled human-computer interaction task that ran at three different speeds, thereby creating three different levels of task difficulty and, as we assumed, cognitive workload. The background luminance in the task displays was once again varied between two levels in order to study the pupillary response to variation in both cognitive workload and luminance.

Method

Participants. The same ten subjects from Experiment 3 also participated in Experiment 4.

Apparatus. The apparatus was the same as in Experiments 1 to 3.

Materials. The stimulus displays showed a grid of 4×3 square cells, each of them subtending a width and height of approximately 9.2° (see Figure 8). As in Experiment 3, the background luminance assumed one of two different levels – dark gray (20.0 cd/m^2) versus light gray (30.0 cd/m^2). At the beginning of a trial, all cells were empty. Then, in each cell, one of four possible items could be randomly chosen to appear: a red square, a red circle, a blue square, or a blue circle. These items then increased in size three times before they disappeared. The subjects' task was to prevent any blue circles from attaining their maximum size. To achieve this, the subjects could look at any growing blue circle and simultaneously press a designated game pad button to eliminate that item. Whenever a subject made a mistake, that is, let a blue circle reach its maximum size or eliminate a non-target object, a buzzer sound was played. In the "easy" condition, every 1000 ms 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.

----- insert Figure 8 about here -----

Procedure. Each of the three levels of task difficulty was combined with the two levels of background luminance, resulting in six different experimental conditions. Each subject performed in four trials of each condition. 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.

Results

To verify that background luminance did not significantly influence the subjects' task performance, the proportion of missed (fully grown) targets out of all targets shown within a trial was computed for each of the experimental conditions. A two-way ANOVA with the factors task difficulty and luminance revealed a significant influence of task difficulty on performance, $F(2; 8) = 10.79$, $p < 0.01$. Luminance had no effect, $F(1; 9) < 1$, and there was no significant interaction between the two factors, $F(2; 8) = 1.51$, $p > 0.2$. The mean proportion across difficulty levels (easy: 3.5%, medium: 3.4%, hard: 8.4%) indicates that performance at the hard level strongly deteriorated as compared to the easy and medium levels.

The same type of ANOVA was then computed for the pupil constriction index based on the pupil size data corrected by the calibration interface. It showed that this index was significantly influenced by the factor task difficulty, $F(2; 8) = 24.26$, $p < 0.001$, and by the factor background luminance, $F(1; 9) = 47.19$, $p < 0.001$. There was also a significant interaction between these factors, $F(2; 8) = 15.93$, $p < 0.005$. Figure 9 (top) illustrates both the gradual effect of task difficulty and the substantial offset caused by the difference in background luminance. Moreover, it can clearly be seen that the interaction is due to a steeper decrease of the pupil constriction index for the bright background than for the dark background. This result suggests that task difficulty had a stronger effect on pupil size for the bright background than for the dark background. However, there is no evidence of task difficulty and cognitive workload changing with background luminance – for example, through different visibility of objects – since luminance did not affect performance (see above). A possible explanation of this outcome is that luminance and cognitive workload affect pupil size through distinct mechanisms that require different measures to estimate the strength of the causal factor. To illustrate this idea, Figure 9 (bottom) shows pupil size, measured as the number of pixels in the eye-camera image covered by the pupil, for the same six conditions as before. For this measure, as opposed to the pupil constriction index, the difference between dark and bright backgrounds remains approximately constant across task difficulties (359, 354, and 340 pixels). A two-way ANOVA revealed that both the factors task difficulty, $F(2; 8) = 22.74$, $p < 0.005$, and luminance, $F(1; 9) = 41.08$, $p < 0.001$, influence pupil size, while there was no interaction, $F(2; 8) < 0.1$.

----- insert Figure 9 about here -----

<u>Discussion</u>

The results obtained in Experiment 4 indicate two things: First, the pupil calibration method introduced in Experiment 3 enabled us to distinguish and measure the separate influences by task difficulty and background luminance on pupil size. Second, cognitive workload seems to differ from luminance in the way it affects pupil size. While the pupil constriction index we introduced above provides a useful framework for estimating the pupillary response to changes in illumination of the visual field (and vice versa), it may not be appropriate for calculating quantitative changes in pupil size triggered by variation in cognitive workload. The current results suggest that such a workload change leads to a particular change in the absolute size (area) of the pupil, which is almost independent of the initial size. However, further research under systematically varied conditions is necessary to quantify this effect. The present results at best propose that differences between these two types of influence on pupil size may exist.

Conclusions

This study has provided some tools and data that may be useful and convenient for analyzing pupil size as a measure of cognitive workload in video-based eye-tracking experiments. First, we have introduced a pupil constriction index as a practical measure for pupillary response to changes in illumination. Second, we have developed some heuristics of estimating the effect of illumination of different intensity, chromaticity, and foveal eccentricity on pupil size. These heuristics can help to estimate pupillary responses in situations when display luminance and gaze position cannot be held constant. Deviations from such estimates can then serve as indicators of changes in cognitive workload. Third, we have presented a calibration technique for substantially reducing the eye-movement induced variance in video-based pupil dilation measurement. Our proposed calibration procedure takes approximately 50 seconds and thus does not add considerable overhead to experimental sessions. While we have presented and evaluated a neural-network approach, other approaches may yield similar results. Finally, we have measured the combined effects of changes in display luminance and cognitive workload on pupil size. The pupil size data, corrected by the calibration interface, allowed a clear identification of both individual effects. However, the current data suggest that workload affects pupil size in a different way than does luminance. More research is necessary to examine this difference and develop an appropriate measure that better quantifies the workload effect on pupil size.

The present study can be considered a small but significant advance in using pupil dilation for the analysis of cognitive workload in video-based eye tracking experiments. In such experiments, researchers receive pupil size data “for free”, without any further hard- or software requirements. By adding a brief pupil calibration procedure to their experimental sessions and using the techniques, measures and heuristics developed in the present study, these raw data can be turned into a valid and sensitive measure of cognitive workload. This measure can be used by itself or to complement other variables such as galvanic skin response, heart rate, and electroencephalographic data. In either case, the proposed methods can expand and enhance data acquisition without adding significant cost to the experiment.

References

Bailey, B.P. & Iqbal, S.T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction*, 14, 1-28.

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

Essig, K., Pomplun, M. & Ritter, H. (2006). A neural network for 3D gaze recording with binocular eye trackers. *International Journal of Parallel, Emergent, and Distributed Systems*, 21, 79-95.

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.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. *Handbook of Perception and Human Performance,* 2, 1-49.

Pinoli, J.C. (1997). The Logarithmic Image Processing Model: Connections with Human Brightness Perception and Contrast Estimators. *Journal of Mathematical Imaging and Vision*, 7, 341–358.

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.

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

Porter, G., Troscianko, T. & Gilchrist, I.D. (2007). Effort during visual search and counting: insights from pupillometry. *Quarterly Journal of Experimental Psychology, 60,* 211-229.

Reeves, P. (1920). The response of the average pupil to various intensities of light. Journal of the Optical Society of America, 4, 35–43.

Ritter, H. (1993). Parametrized self-organizing maps. *ICANN93-Proceedings* (Berlin: Springer), pp. 568–577.

Wilson, G.F. (2001). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *International Journal of Aviation Psychology*, 12, 3-18.

Võ, M.L.H., Jacobs, A.M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A. & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology*, 45, 130-140.

Figure Captions

Figure 1. Time course of pupil diameter after the onset of a bright stimulus. The pupil diameter is scaled in such a way that 1 corresponds to its initial value before stimulus onset and 0 corresponds to its final state after complete adaptation.

Figure 2. Relative pupil size after adaptation to stimuli of different luminance with value 1 indicating pupil size for a completely black screen. Error bars in all figures show the standard error of the mean.

Figure 3. Concentric rings 1 to 4 used as stimuli in Experiment 2.

Figure 4. Luminance of the stimulus (top panel) and resulting mean pupil constriction index (bottom panel) for different values of the red, green, and blue channels of a standard RGB monitor.

Figure 5. Mean pupil constriction index for the pupillary response to the illumination of the four stimulus rings.

Figure 6. Mean pupil constriction index as a function of the number of simultaneously illuminated stimulus rings.

Figure 7. Mean pupil constriction indices while fixating on markers on a 4×4 grid, measured before correction by the neural-network calibration interface (top panel) and afterwards (bottom panel). The columns of the grid are indicated by x = 1, …, 4 (left to right), and the rows are indicated by y = 1, …, 4 (bottom to top). Measurements were taken for bright and dark backgrounds, as illustrated by hollow and solid markers, respectively.

Figure 8. Screenshot of the interactive task used in Experiment 3. The objects were red and blue, and a green background indicated the cell that the subject was currently looking at. The objects grew over time, and the subjects' task was to eliminate the blue circles by looking at them and pressing a button before they reached their maximum size.

Figure 9. The effects of task difficulty and background luminance on pupil size, as indicated by the pupil constriction index (top panel) and the pupil area in the eye-camera image (bottom panel).

Figure 1

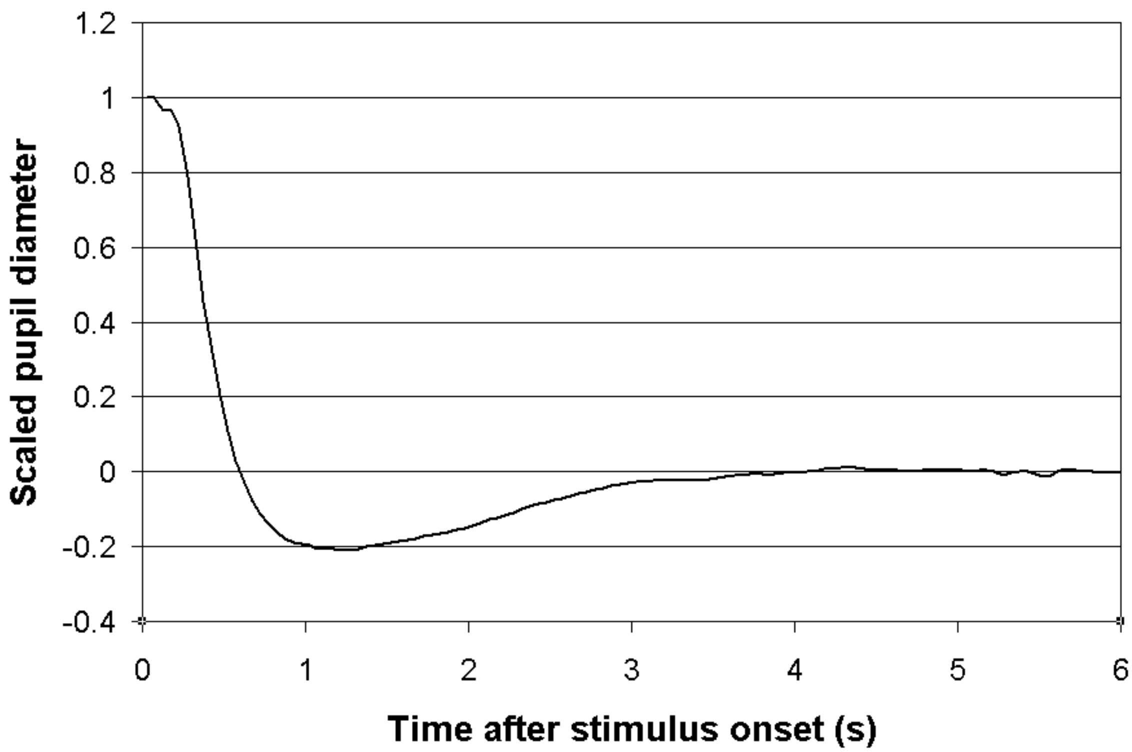

1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Scaled pupil diameter
0
1
2
3
4
5
6
Time after stimulus onset (s)

Figure 2

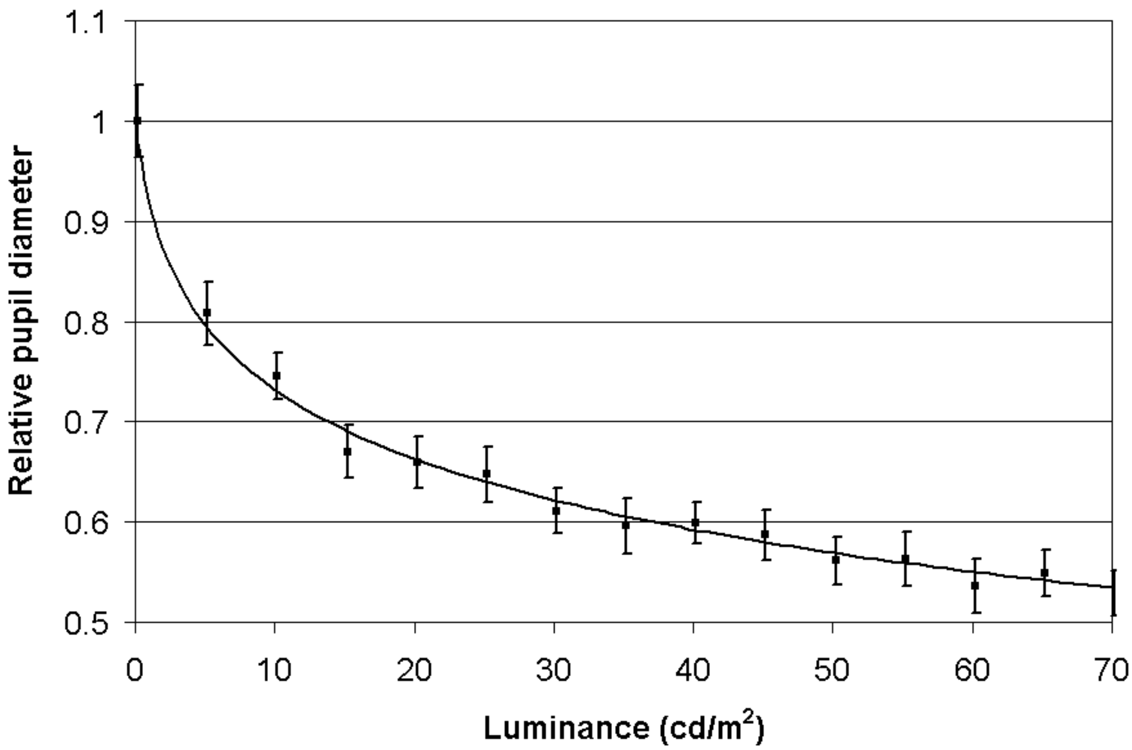

1.1
1
0.9
0.8
0.7
0.6
0.5
Relative pupil diameter
0
10
20
30
40
50
60
70
Luminance (cd/m2)

Figure 3

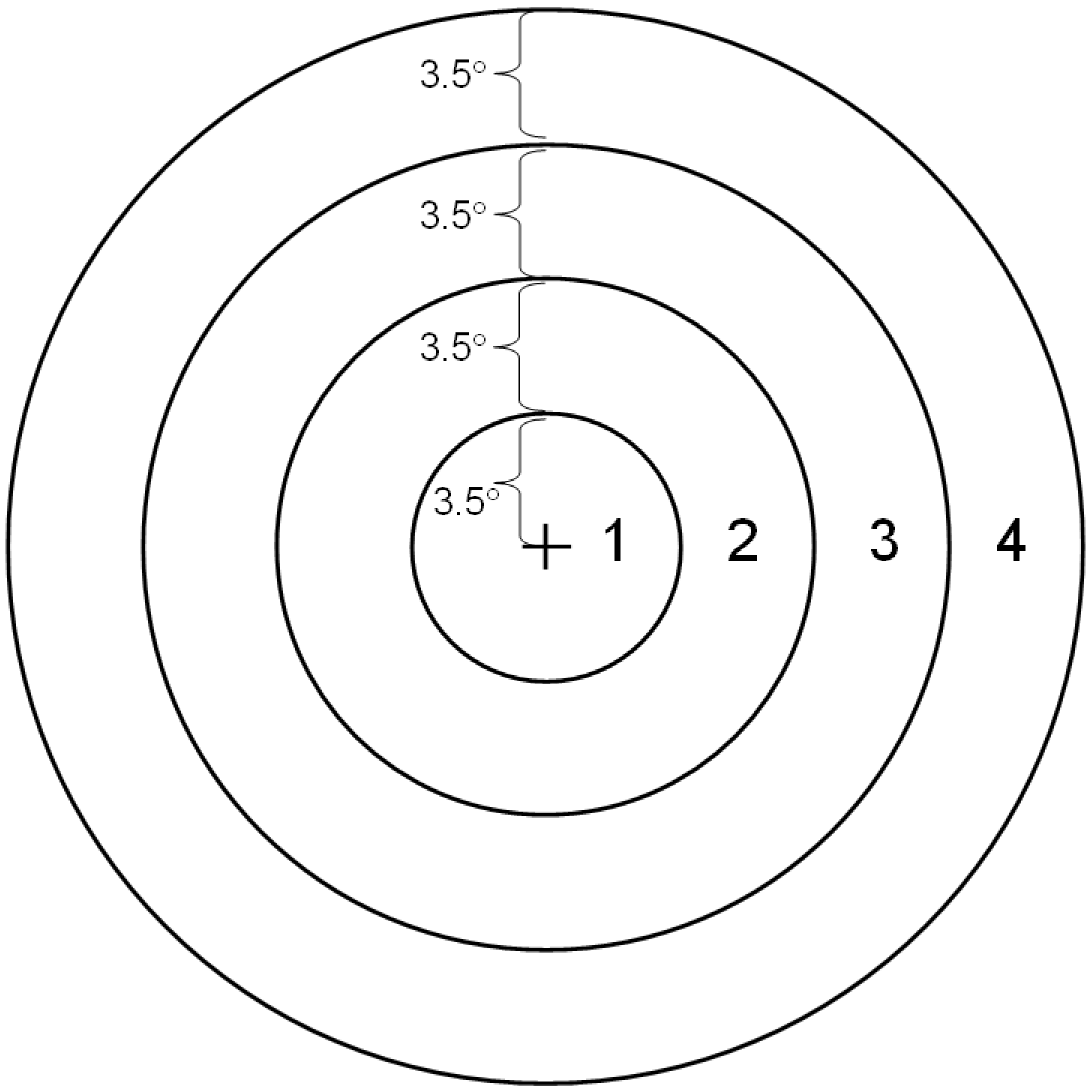
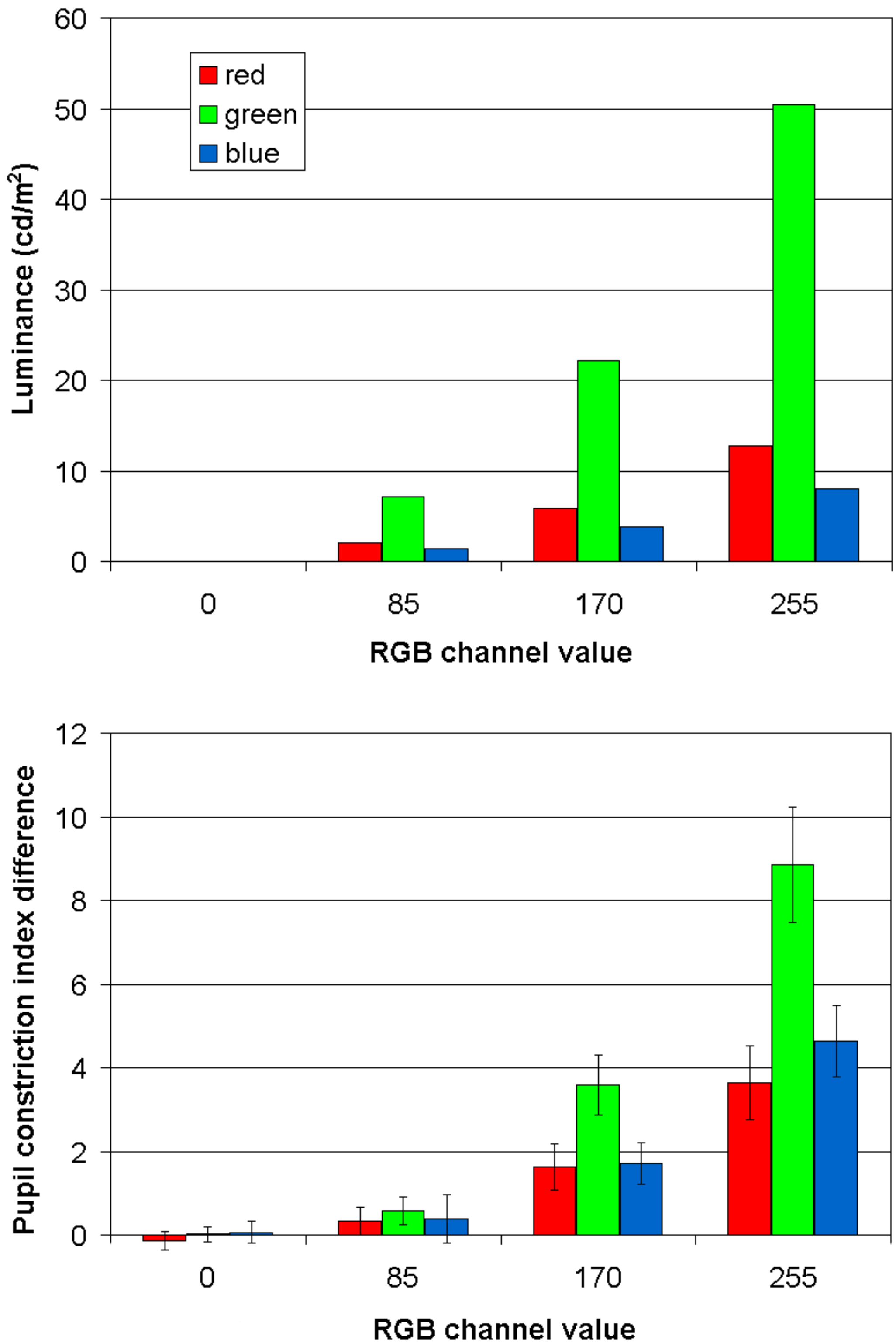


Figure 4



60
50
40
30
20
10
0
red
green
blue
Luminance (cd/m2)
0
85
170
255
RGB channel value
12
10
8
6
4
2
0
Pupil constriction index difference
0
85
170
255
RGB channel value

Figure 5

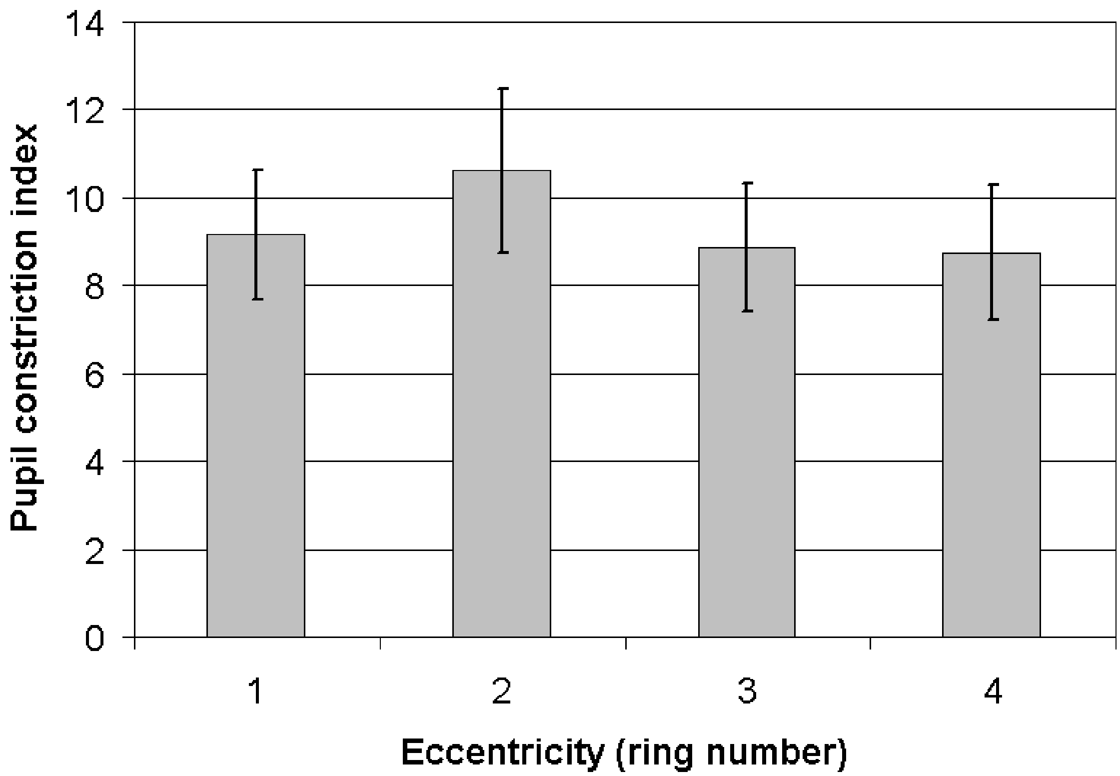
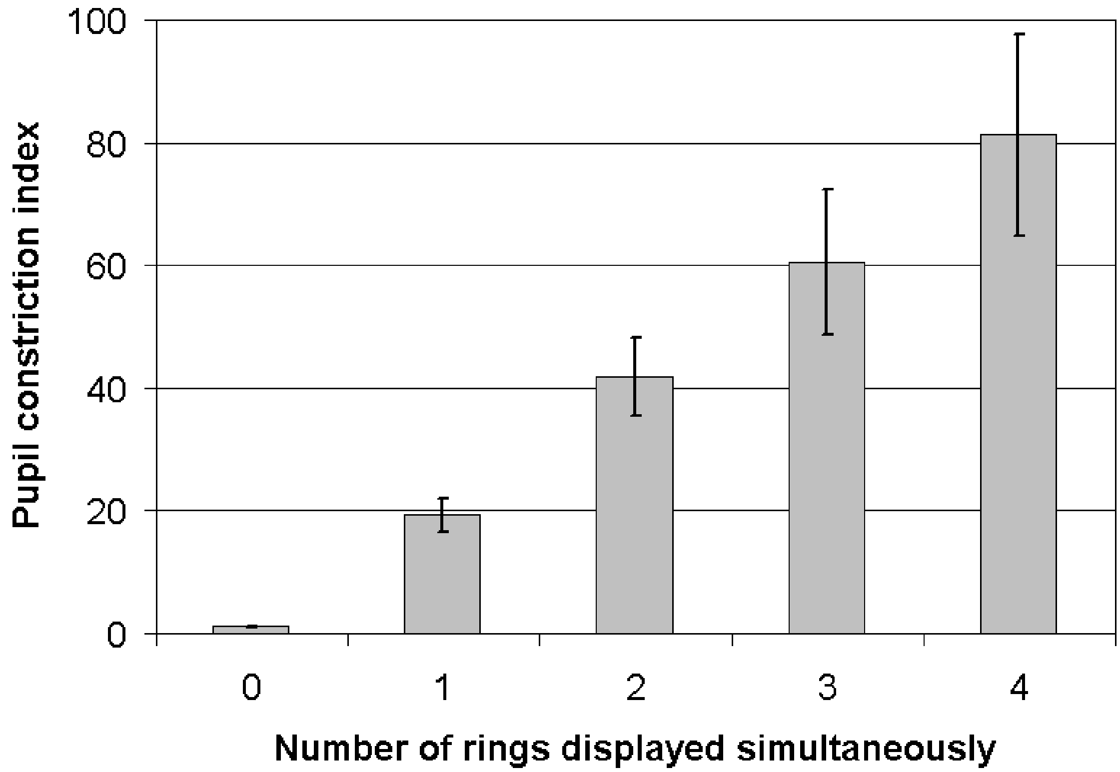

14
12
10
8
6
4
2
0
Pupil constriction index
1
2
3
4
Eccentricity (ring number)

Figure 6



100
80
60
40
20
0
Pupil constriction index
0
1
2
3
4
Number of rings displayed simultaneously

Figure 7

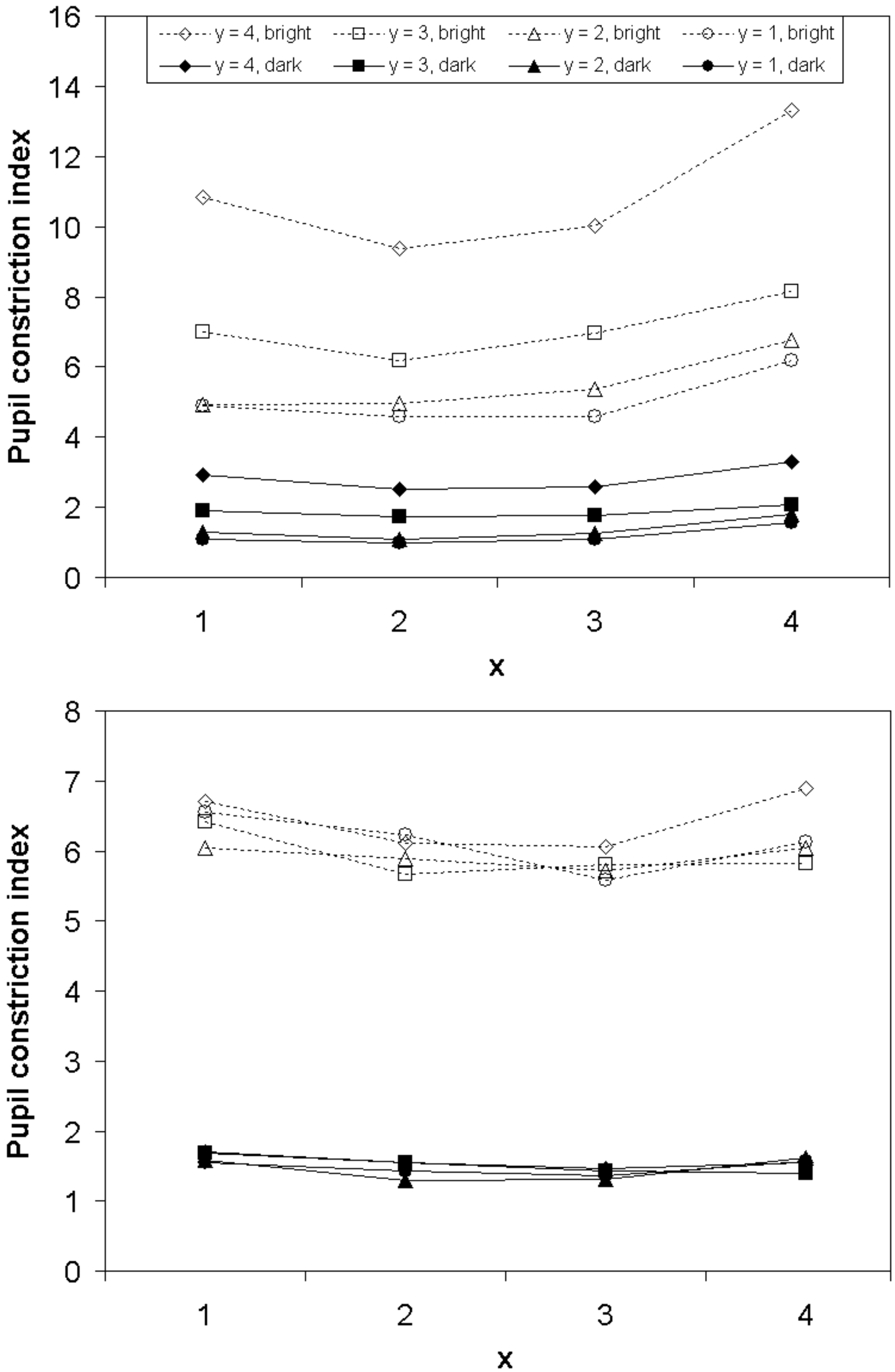

y = 4, bright
y = 3, bright
y = 2, bright
y = 1, bright
y = 4, dark
y = 3, dark
y = 2, dark
y = 1, dark
Pupil constriction index
x
Pupil constriction index
x

Figure 8

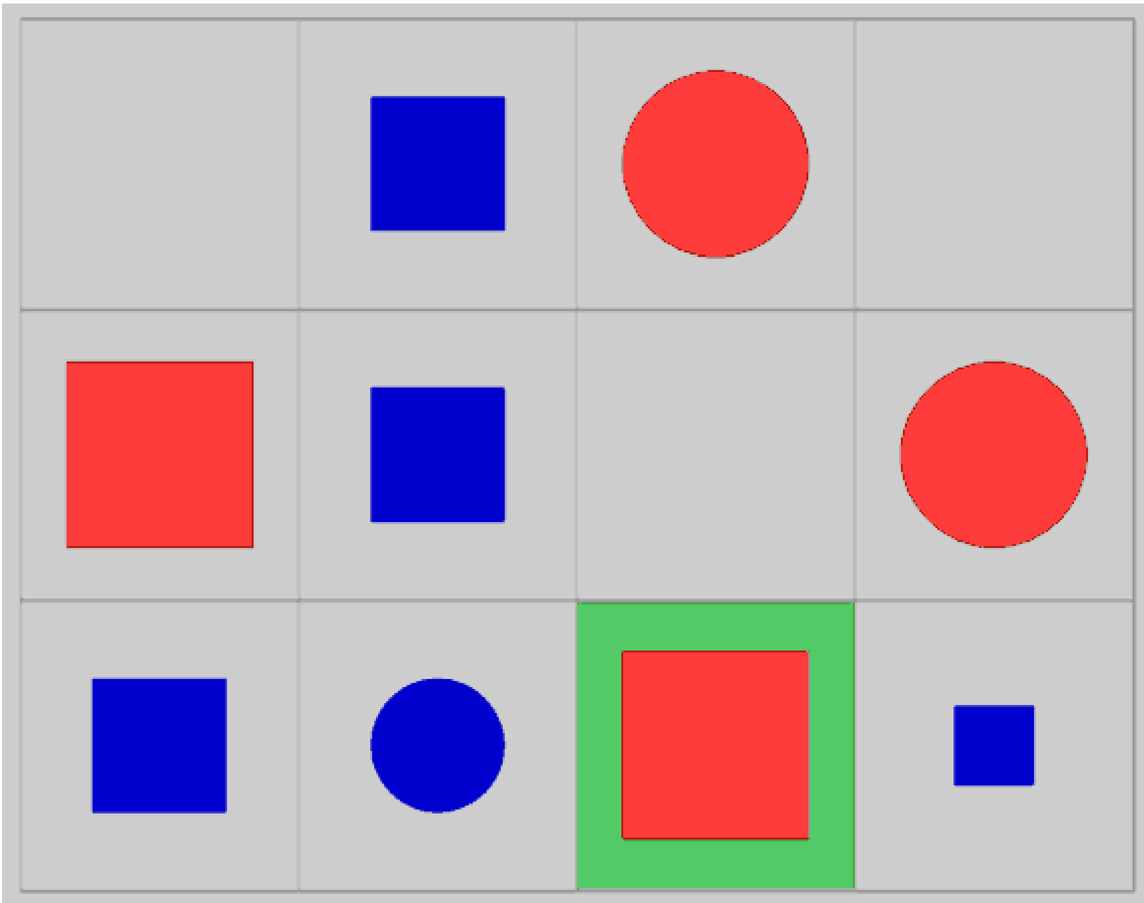
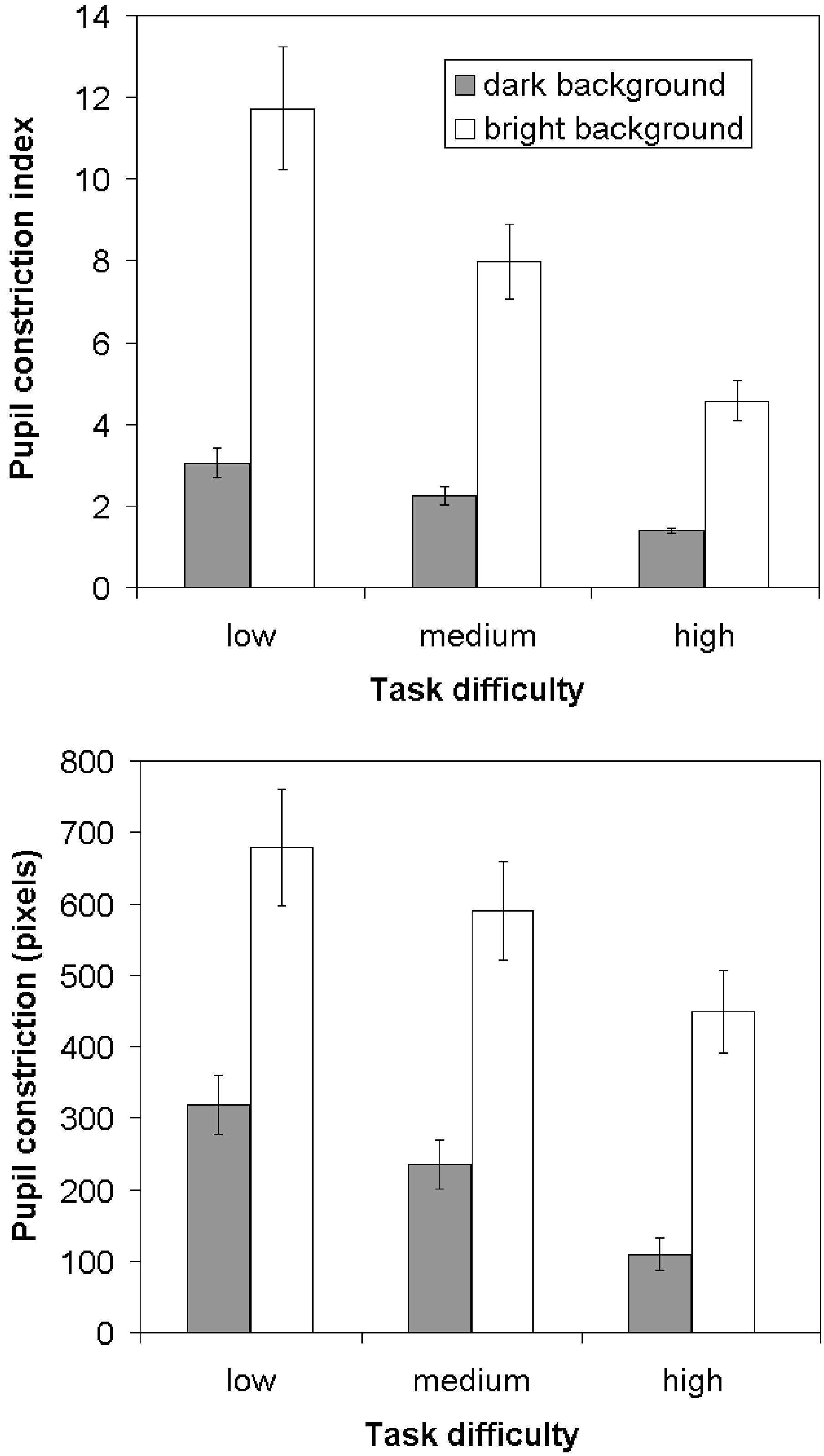


Figure 9



14
12
10
8
6
4
2
0
Pupil constriction index
dark background
bright background
low
medium
high
Task difficulty
800
700
600
500
400
300
200
100
0
Pupil constriction (pixels)
low
medium
high
Task difficulty