Pupil dilations during target-pointing respect Fitts' Law

Xianta Jiang[*]
M. Stella Atkins[†]
Geoffrey Tien[‡]
Simon Fraser University

Bin Zheng[§]
University of Alberta

Roman Bednarik[¶]
University of Eastern Finland

Abstract

Pupil size is known to correlate with changes of cognitive task workloads, but the pupillary response to requirements of basic goal-directed motor tasks is not yet clear, although pointing with tools is a ubiquitous human task. This work describes a user study to investigate the pupil dilations during aiming in two tele-operation tasks with different target settings, one aiming at targets with different sizes located at constant distance apart, and the other aiming at targets varying in different distances. The task requirements in each task were defined by Fitts' index of difficulty (ID). The purpose of this work is to further explore how the changes in task requirements are reflected by the changes of pupil size, i.e., whether the pupil responds to either target size or target distance, or to both of them. Pupil responses to different task IDs were recorded in each task. The results showed that the pupil responds to the changes of ID, not just to the change of target size. This implies that pupil diameter can be employed as an indicator of task requirement in goal-directed movements, because higher task difficulty evoked higher peak pupil dilation which occurred with longer delay. These findings can be used for detailed understanding of eye-hand coordination mechanisms in interactive systems and contribute to the foundation for developing methods to objectively evaluate interactive task requirements using pupil parameters during goal-directed movements.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology;

Keywords: pupil diameter, movement-related pupil response, Fitts' law, task difficulty, minimally invasive surgery

1 Introduction

High workload experienced during a complex task may affect the performance and even cause failure of the task, since the human mental capacity is a finite resource [Cassenti and Kelley 2006]. This especially holds true in goal-directed movements with high task requirements. For example, surgeons experienced higher mental workloads in minimally invasive laparoscopic surgery under image-guided condition than in open procedures [Berguer et al. 2001]. Task requirements in laparoscopic surgery are higher than in open surgery due to the use of long-shaft tools, indirect mapping of visual field from the camera, and the overall lack of natural vision on the surgical sites. Similar difficulties are experienced in everyday interactions and in other skilled domains, too, where some targets we would like to interact with are further apart or are smaller than others.

The first step in the management of operators' workloads is to develop an objective, non-intrusive, and continuous measure of the mental workloads during a complex goal-directed motor task.

With advances in eye-tracking technology, pupil diameter can be recorded remotely and its subtle changes indicate the cognitive task load [Klingner et al. 2008]. What is lacking is the evidence between the change of pupil dilation and the change of task requirements in a goal-directed movement. Only a few works have been published in this area. One of the pioneering fundamental works was conducted by Richer and Beatty [1985], who found that the pupil dilated as a function of the complexity of finger movements. However, this was not testing of the true task requirements during goal-directed movement; the participants in this 1985 study simply flexed their fingers while looking at a blank screen for the purpose of pupil size recording.

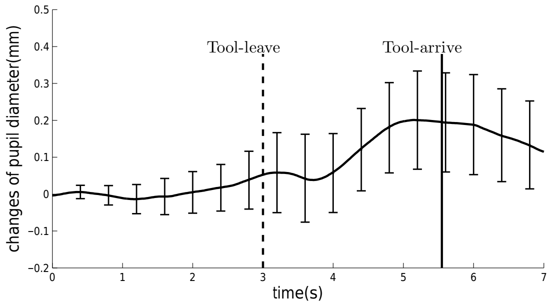
In our previous work [Jiang et al. 2014], we explored the pupil responses to the task requirement in a Fitts' pointing task, i.e. moving a tool to point and touch pairs of circles with target size and distance defined following Fitts' index of difficulty (ID). The detailed parameters of the target setting are shown in Figure 1(b). We found that the pupil responds to the task requirement following a pattern shown in Figure 1(a). Specifically, the subjects' pupil dilates significantly (about $0.05mm$) starting about $1.5s$ before the aiming movement, and then constricts (about $0.02mm$). Before the tool movement ended, the pupil peaks in size, ranging from $0.2mm$ to $0.25mm$ according to the different IDs. The peak pupil diameter and latency positively correlate to the increase of IDs. The evidence indicates that the changes of pupil diameter are regulated by task requirement. However, the task requirement in this previous study was affected by the mixture effect of both target size and target distance changes.

Previous work noticed that the pupil constricts during saccades when watching a fast moving object [Abrams et al. 1990; Elliott et al. 2001]. Because of this effect, we ask the question: would the pupil constricts more significantly in aiming at a target placed at longer distance than in aiming at a smaller target at the same distance? Specifically, we would like to know whether the two parameters used for calculating the ID, the target size and target distance, affect the pupil's size differently. We hypothesize that increasing target distance will not dilate the pupil as much as reducing the target size. In other words, the primary factor regulating the change of pupil size would come from the changes of target size rather than the changes of target distance.

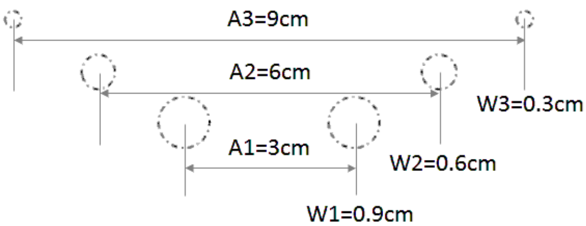
To test our hypothesis, we conducted the present study by asking subjects to perform two tasks: 1) aiming at targets with 4 different sizes located at constant distance apart, and 2) aiming at targets varying in 4 different distances. If our hypothesis is supported, the peak pupil size will increase as a function of the ID in the first setting with the smaller target sizes, but will maintain constant size in the second setting where only target distance is different. The findings from this study will further clarify the pupil response pattern

[*]e-mail: xiantaj@sfu.ca
[†]e-mail:stella@sfu.ca
[‡]e-mail:gtien@sfu.ca
[§]e-mail:bzheng1@ualberta.ca
[¶]e-mail:roman.bednarik@uef.fi

ETRA 2014, March 26 – 28, 2014, Safety Harbor, Florida, USA.



(a) Pupil response to tool movement



(b) Target setting

Figure 1: *Findings and experimental setting of the authors' previous work. (a) shows the pattern of pupil response to tool movement; the mean pupil diameter changes over a 7-second window around the moment of aiming movement start. (b) shows the target setting of the task with combined variation of target size ($W1$ to $W3$) and distance ($A1$ to $A3$). The task was to move a tool to point between the circles in a Laparoscopic training box. The error bars showing 1 std. dev. are drawn every 400ms.*

in goal-directed movements, which will lead to constituting a foundation for construction of methods towards objectively measuring the task requirements during complex goal-directed motor tasks.

2 Related Works

2.1 Mental Workload, Task Difficulty, and Measurement Methods

Mental workload is usually considered as a finite mental resource that one uses to perform a task under specific environmental and operational conditions, which is induced mainly by task demands [Wickens 2002]. In order to evoke different levels of mental workload of a user in a study, the difficulty of the task has to be carefully manipulated. Most past studies empirically manipulated the task difficulty by changing related task factors such as the complexity of the task. For example, Richer and Beatty [1985] defined four levels of task difficulty by varying the complexity of finger movement: one-finger flexion, two-finger flexion of one hand, one-finger flexion of both hands, and three-finger flexion in one hand. Despite the wealth of research that we review next, there is currently no commonly accepted way to quantitatively define task difficulty in workload studies employing eye-tracking.

Fitts' law is a traditional model of human movement by analogy to the transmission of information [Fitts 1954]. Specifically, the information capacity of the human motor system, which is called index of performance (IP), is considered to be relatively stable and can be calculated by the ratio of the index of difficulty (ID) of a motor task and the movement time (MT).

$$IP = \frac{ID}{MT} \tag{1}$$

The index of difficulty (ID) is determined by the target distance (A) and the target width (W)

$$ID = \log_2 \frac{2A}{W} \tag{2}$$

Fitts' law has been widely adopted in a variety of research areas, including kinematics, human factors, and human-computer interaction (HCI) [MacKenzie 1992], and even recently in the Laparoscopic environment [Chen and Lin 2011; Prytz et al. 2012].

Past research introduced three main categories of mental workload measurement techniques: subjective rating scales (self-assessment), performance measures (including primary and secondary task measures), and psychophysiological measures [Gawron 2008]. The subjective rating methods are easy to perform, but the validity of the measurement is affected by the operator's working memory [Zheng et al. 2010]. The secondary task may cost extra workload to the primary task. Finally, although psychophysiological measures avoid subjective bias and generally do not interfere with task performance, most of them present an intrusion, as they need sensors attached to the user. One way to overcome the problem with intrusiveness is to apply remote sensors such as eye-tracking.

With advances in eye-tracking technology, pupil diameter can be remotely recorded and its subtle changes are able to indicate the cognitive task load [Klingner et al. 2008]. The pioneer work measuring pupil size changes relating to mental workload and task difficulty was done by Hess and Polt [1964] in a mental arithmetic experiment. The authors found that the pupil size of the subjects gradually dilated along with the time elapse of presentation of a multiplication problem and reached a peak immediately before the production was orally reported; then constricted rapidly back to the original size. The mean pupil dilation was also found to be a function of the level of difficulty of the problem. Following this work, extensive studies showed that the changes of pupil size reflect mental workload and the level of difficulty in different tasks, e.g., mental arithmetic tasks [Bradshaw 1967], recall or memory tasks [Otero et al. 2011; Goldinger and Papesh 2012], and visual search tasks [Privitera et al. 2010].

Pupil size also responds to the critical events during information processing, which is called Task-evoked Pupil Response (TEPR), appearing at the event onset with a short latency (averaging between $100ms$ and $200ms$), and terminating rapidly following the completion of the event [Beatty 1982]. The TEPR has been employed to capture and evaluate the mental workload changes during a variety of tasks [Beatty 1982].

2.2 Pupillary Responses to Movements of Hand and Tool

Richer and Beatty [1985] examined the pupil responses to simple hand movement, the self-paced finger flexion. The authors found the typical pupil dilation pattern to a simple hand movement is that pupil dilates at around $1.5s$ before the finger movement and peaks afterwards at around $0.5s$. This pupil dilation pattern was confirmed by other works [Privitera et al. 2010; Moresi et al. 2011]. For example, in a visual target search study, Privitera et al. [2010] found a quite similar pupil dilation pattern to that in Richer and Beatty's finding when the participants successfully detected a target and responded with a key-press.

However, in goal-directed interactive movements, such as aiming at a target and selecting a menu item, the tool may move continuously, which may evoke a different pupil response pattern. Such phenomena, if confirmed, would have important implications. For example, interactive environments could continuously adjust the presented information to accommodate the workload of the user in real-time. Design guidelines for fundamental movements of a tool could be objectively evaluated, and training and simulation systems could be developed with specific tasks toward the improvement of motor and aiming skills.

In the HCI domain, Iqbal et al. [2005] employed pupil diameter as an objective indicator of mental workload in interactive tasks including route planning and document editing, purporting to find proper moments for low cost interruption; they found the pupil size is relatively small during task boundaries which are suitable for interruption with lower mental workload.

Bednarik, Vrzakova, and Hradis [2012] proposed an approach to predict intent in gaze-base interaction purporting to avoid Midas touch by using the fusion of pupil diameter with other eye metrics. The authors found that the pupillary responses based model did not achieve significant discriminative performance between intent and non-intent events. The likely reason is that the pupil responses in the windows for feature extraction (two to three fixations around the occurrence of button-press events) may be an overlapped reflection of various events. This also reminds us it is important to separate the pupil responses to specific events when employing pupil diameter to measure event-related task workload, e.g. to measure the task requirement on discrete elementary tasks as done in the present work.

In the field of eye-hand coordination in health care training, Jiang et al. [2013] investigated the pupil changes in a simulated laparoscopic task, and found the pupil dilates faster during the execution of higher difficulty tasks, such as dropping a peg into a tiny cup, than during the execution of tasks with lower difficulty, such as touching a centrally located home position with the operating tool. However, the difficulty level for each of the tasks was not well defined in this study.

Marshall [2002] reported the Index of Cognitive Activity (ICA) that is capable of capturing subtle cognitive changes from pupil metrics, and was used to predict the expertise of surgeons, together with other eye metrics [Richstone et al. 2010]. However, details of the pupil response to motor tasks were not reported so far.

3 Methods

3.1 Experiment Design

The experiment was designed to explore whether the pupil size responds to both the changes of the target size and target distance apart in a goal-directed movement, such as to move the grasper inside the training box to point the circles printed on the paper. In order to clearly separate the effects of pupil responses to either the changes of target size or distance, either target size or distance apart is kept constant in the two settings, as shown in Figure 2. In setting 1, the target size changes between $1.1cm, 0.9cm, 0.6cm$, and $0.3cm$ with a constant distance of $6cm$, while in setting 2, the target size is constant ($0.6cm$) but the target distances vary from $3cm, 5cm, 7cm$, and $9cm$. The formula for calculating ID is as in Equation 2 [Fitts 1954]. Therefore, in this experiment, four IDs for setting 1 are $ID1 = 3.4, ID2 = 3.7, ID3 = 4.3$, and $ID4 = 5.3$ bits/response; and the four IDs for setting 2 are $ID1 = 3.3, ID2 = 4.1, ID3 = 4.5$, and $ID4 = 5.0$ bits/response. $ID1, ID2, ID3$, and $ID4$ are denoted as Easy, Middle, Hard, and Hardest respectively in this study.

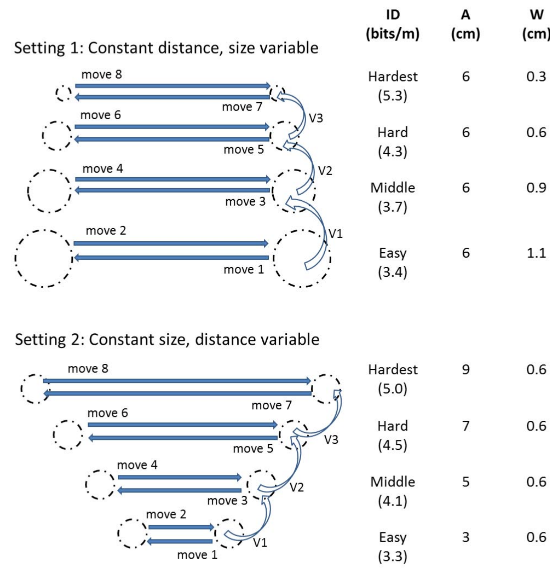


Figure 2: *Target settings and task execution sequences. The dash circles are the targets with distance (A) and size (W) shown in corresponding rows in the right table. The horizontal arrows represent the horizontal tool movement steps, labeled from move 1 to move 8. The vertical arrows labeled V1 to V3 represent the execution sequence between each pair of targets.*

Since the frequency of pupil response to a hand movement is typically lower than $0.5Hz$ [Richer and Beatty 1985; Privitera et al. 2010; Moresi et al. 2011] and the frequency of the Fitts' tool movement under laparoscopic environment is usually around $0.5Hz$ [Chen and Lin 2011; Prytz et al. 2012], in order to avoid overlap of the pupil dilation curve, we used discrete Fitts' pointing [Fitts and Peterson 1964], i.e., there was a $10s$ wait time before each tooltip move. Such discrete goal-directed movements and related eye-hand coordination are elementary parts of any human-computer interaction and arise in the vast majority of our interactions with surrounding objects.

3.2 Participants and Experiment Setting

Eight graduate students (four females and four males) participated in the study. All were right-hand users and had normal or corrected-to-normal vision. None of them were previously trained in any surgical procedures.

The equipment had three components: an eye-tracker, a surgical training box, and a PC system. The participants performed the task using the training box for laparoscopic procedures (Laparoscopic Trainer, 3-D Technical Services, Franklin, OH), holding a surgical grasper at a standing pose about $60cm$ from the eye-tracker, as shown in Figure 3. The tips of the grasper were black-taped together to keep its color consistent with the shaft, and make image processing easier. The scene of the work area inside the training box was illuminated and captured at $30Hz$ by a built-in camera and projected to a 17" display. The eye movements of the participants were recorded simultaneously using the remote eye-tracker (Tobii X50, Tobii Technology AB). A web camera was attached on the top center of the display frame recording the face expressions of the participants for the purpose of identifying eye blinks and lost data. The eye-tracker and the cameras were physically connected to the PC system and integrated using Tobii software, Clearview 2.7.0 in the PC system. The setting was physically isolated to keep the effect of ambient lighting relatively stable. The brightness and contrast of the display were constant and set to a moderate level to make the pupil to work at the center of the length-tension curve of the iris muscle, for the best pupil response to the task requirement [Privitera et al. 2010].

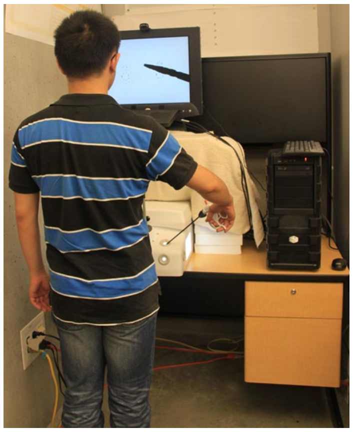


Figure 3: *Experimental setting.*

3.3 Tasks and Procedure

The task was to point to the circles printed on an A4 white paper pasted on a same size thin glass ($5mm$) at the bottom of the training box, using the surgical grasper. This was a discrete pointing task, i.e., participants had to wait $10s$ before each move to the next circle. Specifically, a trial consisted of 8 discrete horizontal tool movements (move1 to move 8) between the targets in pairs and 3 vertical transfer (V1 to V3), each separated by $10s$, as shown in Figure 2. The trial started by placing the tooltip on the right bottom circle for $10s$, then moving the tool to the left bottom circle (move 1), and ended by stopping the tooltip on the right bottom circle for $10s$ after move 8. Only the 8 horizontal tool movements were included in the analysis.

The participants were instructed to move the tool and hit the target as accurately and as fast as possible; once the target was hit, $10s$ were counted before moving to the next target. Each trial took about 2 minutes.

Each participant read and signed the consent form before entering the study, and then read the instructions. The participants practiced the task for a few minutes, until they were ready to begin.

Each participant performed two blocks using the two target settings shown in Figure 2 in a counterbalance way, i.e., half of the participants performed setting 1 in block 1 and setting 2 in block 2; another half executed setting 2 in block 1 and vice visa. Each block consists of 6 trials, with three trials starting from Easy to Hardest and the other three trials from Hardest to Easy (by flipping the target paper). There was $20s$ break between each trial. Between two blocks, the target setting was changed.

3.4 Data analyses

The pupil diameter data needed to be processed and synchronized with the trajectory of the tooltip for the analysis of the pupil responses during tool movements. The pupil diameter data was contained in the text file (Combined Data file, CMD) exported from Tobii Clearview and the tooltip positions were derived from the task video (recorded by the camera attached to the training box) using a customized algorithm. The pupil data were in $50Hz$ and the task videos were in $30Hz$ and resolution of 352×288 pixels.

3.4.1 Location of the tooltip and moments of tooltip-leave and tooltip-arrive

We developed a customized video processing algorithm written in Visual C++ (Microsoft Visual Studio, Microsoft Corporation) to derive the position of tooltip automatically. The algorithm involved three major steps. First, the RGB video was read in frame by frame and transferred to gray scale image format, and then binary-thresholded into black and white image format, where mostly only the tool was left in the image as shown in Figure 4A. Second, the biggest connected object was searched and identified as the tool, as shown within the red rectangle in Figure 4B.

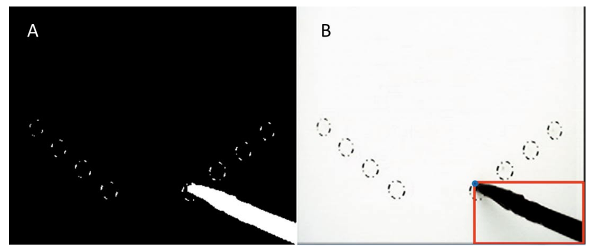


Figure 4: *Key steps in detecting the tooltip from surgical videos. An example frame from constant size setting.Panel A shows the binary-thresholded image with mostly the tool left. Panel B shows the recognized tool (in red rectangle) and tooltip (the blue dot on the left top corner of the red rectangle).*

Third, the coordinates of the left top corner of the tool rectangle was used as tooltip position, as the tool was always consistently north-west orientated. The determined tooltip position was the blue dot shown Figure 4B. The tooltip positions (x and y coordinates in pixels) together with timestamps were stored in a text file for further analysis. The tooltip data were smoothed with a running-average-filtered using equally weighted four samples window.

This algorithm is simple but effective, allowing detecting of the tooltip position in real-time.

Some critical moments of tooltip movements, i.e. the moments when the tooltip started to move (tooltip-leave) and reached (tooltip-arrive) the target, were detected by another customized algorithm, written as a Matlab script. The algorithm first finds the absolute tooltip movement peak velocity along the x-axis during a move, and then detects backward and forward respectively for the moments of tooltip-leave and tooltip-arrive, by checking whether the absolute velocity of the tool movement in x-axis is lower than a threshold. The velocity thresholds for detecting the tooltip-leave and tooltip-arrive moments were empirically determined, since there are abrupt changes of tool velocity at the moments of tooltip-leave and tooltip-arrive. The tooltip-leave and tooltip-arrive thresholds were both set to 30 pixels/s.

The tooltip usually leaves the current circle to the next target quickly in the horizontal direction once the tooltip starts to move, which makes it easy to accurately detect the tooltip-leave. However, the tooltip usually quickly arrives at a relative height position over the target circle and then slowly descends to touch the circle, which causes no change in the tooltip positions from the 2-D image during this descending period. This caused inaccuracy in detecting the tooltip-arrive moments by this algorithm. Therefore, we corrected moments of tooltip-arrive by observing the surgical videos and manually determining the moment of tooltip-arrive.

3.4.2 Pupil data processing and window alignment

Pupil data have to be carefully processed to eliminate noise without losing useful information. In this study, segments of missed pupil data shorter than $100ms$ were linearly interpolated. These missed data might be caused by many reasons such as short blinks and delay of eye-tracking recovery. Then a Butterworth low-pass filter with a cutoff frequency of $4Hz$ was applied to the pupil diameter data, since frequency above $2Hz$ of the pupil is considered as noise [Privitera et al. 2010].

In this study, we were only interested in the pupil size changes in the window around the tooltip-leave and tooltip-arrive, i.e. 3 seconds before tooltip-leave and 4 seconds after tooltip-leave. Relative pupil diameter changes in the window were derived by subtracting each sample from a baseline pupil diameter which was the mean of the pupil size during the $400ms$ from start of the window.

The magnitude of workload-related pupil dilations (less than $0.5mm$) is usually smaller than the magnitude of other simultaneously ongoing pupil changes caused by light reflex, respirations, and other brain activities [Klingner et al. 2008], which causes difficulty in detecting the movement-related pupil dilation. By averaging many repetitions of short epochs of the same task aligned at a specific common time point (the moment of stimulus onset), noise and other pupil size changes not correlated in time to the stimulus will be averaged to zero and the useful pupil changes related to the task will be preserved [Klingner et al. 2008]. In this study, all the data in the windows were aligned at the tooltip-leave (3 seconds into the window), and the mean pupil diameter changes were calculated for each time point in the window across all horizontal tooltip moves from all trials. Similarly, the mean pupil diameter changes were calculated across all moves from all trials for each ID. The mean pupil diameter changes in the 7-second window were drawn in a graph for visual analysis.

To examine which parts of the pupil size changes in the 7-second window have significant differences between the four IDs, the graphical significance testing approach [Guthrie and Buchwald 1991; Privitera et al. 2010] was employed. This method applies a paired t-test to the same time point sample and examines all the p-values along the time axis to determine which segments of the curves are significantly different. Due to the temporal autocorrelation of pupil waveform, we considered a series of more than 4 consecutive samples ($80ms$) with p-values $< .05$ as significantly different [Privitera et al. 2010].

Peak pupil dilation was searched within 4 seconds after tooltip-leave, and the pupil peak duration (from tooltip-leave) was recorded as well.

4 Results

A total of 96 trials were recorded (8 participants, each performed 6 trials in Constant Distance (CD) setting and 6 trials in Constant Size (CS) setting). Two trials in CD setting and four trials in CS setting were excluded from analysis due to low ratio of total fixation time over total execution time (TF/TT), since we have observed that the quality of the eye movement data cannot be guaranteed when TF/TT is lower than a certain value (less than about 70%). From the 90 valid trials, there were window data for 720 horizontal tool movements available (each trial had 4 ID executions, with each ID execution having 2 moves — from right to left and left to right). However, we discarded 4 windows in CD setting and 9 windows in CS setting due to the mis-operation, e.g., when the participants moved the grasper to a wrong target. Therefore we had 364 valid windows in CD setting and 343 valid windows in CS setting.

4.1 Tool Movement Time

The mean tool movement time (MT) is the mean transportation time between tooltip-leave and tooltip-arrive for all horizontal movements.

Figure 5 shows the linear regression of mean MT of each ID to ID values in both target settings with $R^2 = 0.970$ and $p < .001$; the mean MT is positively correlated with ID values.

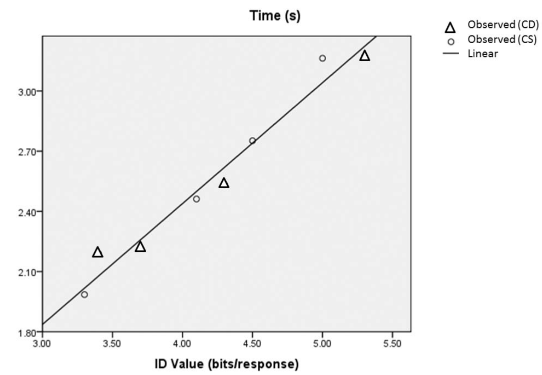


Figure 5: *Linear regression of mean movement time (MT) of each ID to Fitts' ID value ($R^2 = 0.970$ and $p < .001$).*

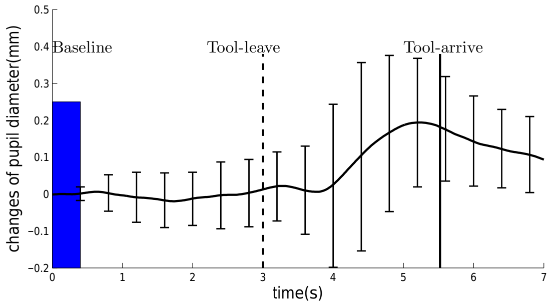
In the constant distance (CD) setting, the mean MT for all IDs is $2.5 \pm 1.0s$. There is significant main effect between four IDs in terms of mean $MT(F(3, 360) = 20.428, p < .001)$. Post Hoc test (Tukey HSD) shows that the mean MT of Hardest ID ($3.2 \pm 1.1s$) is significantly longer ($p < .001$) than other three (Easy, Middle, and Hard IDs being $2.2 \pm 0.9s$, $2.2 \pm 0.9s$, and $2.5 \pm 0.9s$ respectively), and there is no significant difference between other pairs of ID, but the MT of Hard ID is marginally longer than that of the Easy ID ($p = .076$).

In the constant size setting (CS), the mean MT for all IDs is $2.5 \pm 1.0s$. There is significant main effect between four IDs in terms of mean $MT(F(3, 339) = 18.389, p < .001)$. Post Hoc test (Tukey HSD) shows there are significant differences ($p < .05$) between all the pairs of four IDs except that the Hardest ($3.2 \pm 1.4s$) is marginally longer than that of Hard ID ($2.8 \pm 1.1s, p = .056$) and there is no significant difference between Hard and Middle ID ($2.5 \pm 1.0s$).

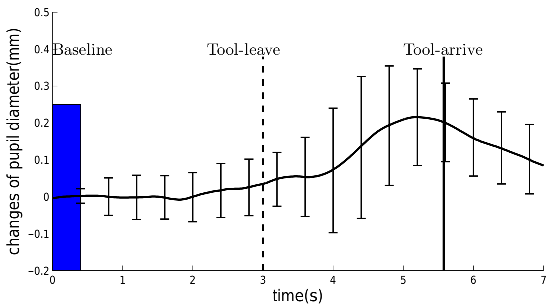
4.2 Pupil Responses to Tool Movements

Figure 6 shows the mean changes of pupil diameter during horizontal tooltip movements over a window in both CD and CS settings. Both Figure 6(a) and Figure 6(b) show a very similar pattern of pupil diameter change—the pupil starts to dilate slightly at $1.2s$ to $1.5s$ before tooltip-leave (increasing less than $0.05mm$), and then peaks ($0.2mm$ compared to baseline) right before tooltip-arrive—except a significant constriction (decreasing around $0.02mm$) at about $200ms$ right after tooltip-leave in CD setting (Figure 6(a)).

Figure 7 shows the mean changes of pupil diameter in a window for four IDs in both CD and CS settings. Besides sharing a common pupil change pattern as shown in Figure 6, the four ID curves in both Figure 7(a) and Figure 7(b) are actually distinguishable. First, as shown in Figure 7(a), the moments when the pupil starts to dilate are different for the four IDs: the Hardest ID (the pink solid curve) starts the earliest at around 1.5s before tooltip-leave, the Easy ID (the black solid curve) starts the latest at around $0.5s$ before tooltip-leave, and the Middle and Hard ones are in the middle.

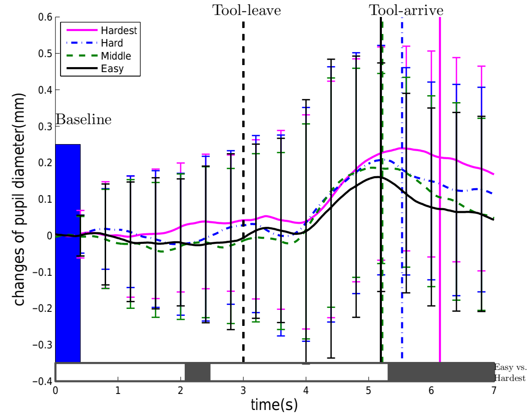


(a) Constant Distance (CD)

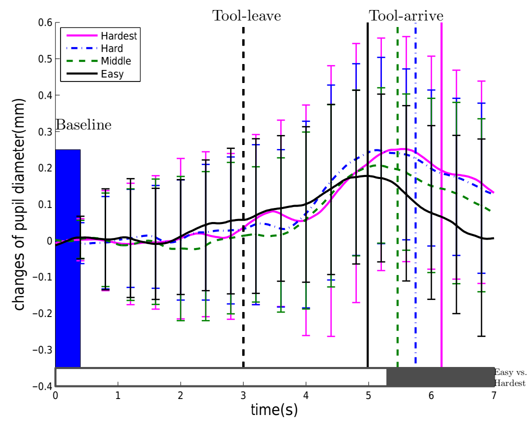


(b) Constant Size (CS)

Figure 6: *(a) mean pupil diameter changes for 364 moves of 46 trials from 8 subjects in CD setting; (b) mean pupil diameter changes for 343 moves of 44 trials from 8 subjects in CS setting. Data were aligned over a 7-second window 3 seconds before the tooltip-leave. The baseline is defined as the mean diameter of the pupil over the 400ms at the beginning of the window, and the solid black curve is the mean pupil diameter change from the baseline over time. The vertical dashed line is the moment of tooltip-leave where all the data are aligned and the vertical solid line is the mean moment the tooltip-arrive. The error bars showing 1 std. dev. are drawn every 400ms.*



(a) Constant Distance (CD)



(b) Constant Size (CS)

Figure 7: *Mean pupil diameter changes against different IDs in two target settings; data are aligned over a 7-second window around tooltip-leave. The black vertical dashed line is the tooltip-leave and other four color vertical lines represent the tooltip-arrive moments of four IDs respectively. The black bar at the bottom of each figure indicates the period having significant differences in pupil dilation between Easy and Hardest ID. The error bars showing 1 std. dev. are drawn every 400ms.*

Second, as shown in Figure 7(b), the amplitude of the constrictions of the four curves right after tooltip-leave is different: the Easy ID one (the black solid curve) nearly does not constrict due to the very short travel distance between the targets, the Hardest ID (pink solid curve) has the deepest constriction, and the other two IDs have the middle level of constrictions. Third, as shown in both Figure 7(a) and Figure 7(b), the peak pupil dilation value and delay positively correlates with the level of difficulty of the task— the harder ID has higher peak pupil dilation value and longer delay.

Graphical significance testing was applied between the curves for the Easy and Hardest IDs, with the results as shown in the bottom black bars in Figure 7(a) and Figure 7(b). For each time point, the paired t-test was performed for each time point for two ID each containing samples from the all trials respectively. The horizontal gray bars in both Figure 7(a) and Figure 7(b) represent the significant segment of the graphical significant testing between Easy and Hardest curves. There are significant differences between Easy and Hardest ID after $5.3s$ in both target settings, and also there is short a segment of signicant difference area ($2.1s$ to $2.4s$) in CD setting.

Figure 8 shows the linear regression of mean peak pupil dilation for each ID to ID value, with $R^2 = 0.849$ and $p < .005$.

Figure 9 shows the linear regression of mean peak pupil dilation duration (from tooltip-leave) for each ID to ID value, with $R^2 = 0.662$ and $p < .05$.

5 Discussion

The study yielded results that did not support our hypotheses. The pupil size increases in response to the tool movements in both target settings in a very similar pattern, as shown in Figure 6; it starts to dilate at around $1.2s$ to $1.5s$ before tool-leave, following a constriction right after tooltip-leave, and then peaking at around $2.5s$ after tool-leave. This evidence indicated that, once the task requirement changed (either by target size or distance), the human operators perceived the difficulty as a single parameter and responded uniformly during the preparation and execution of the goal-directed tool movement. Our findings are consistent with those in previous work [Richer and Beatty 1985; Privitera et al. 2010; Moresi et al. 2011]. The uniform response between pupil dilation and the change of task requirement has an important implication—that pupil dila-

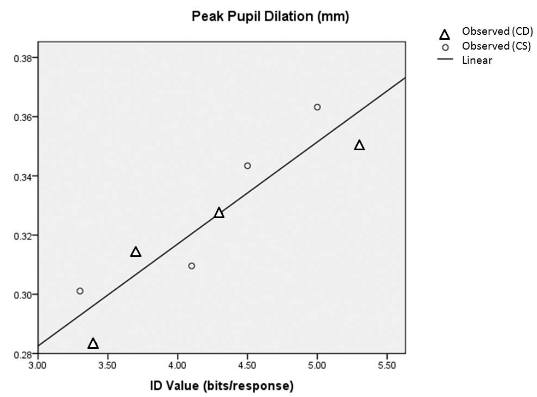


Figure 8: *Linear regression of mean peak pupil dilation of each ID to Fitts' ID value ($R^2 = 0.849$ and $p < .005$).*

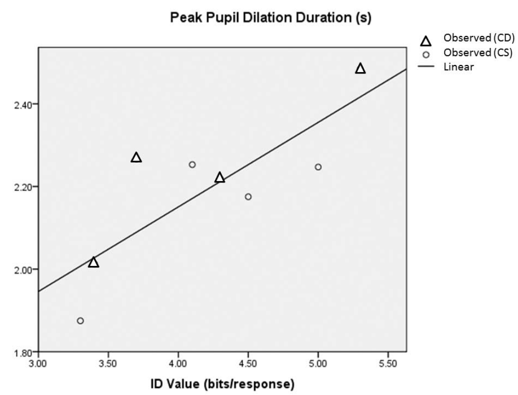


Figure 9: *Linear regression of mean peak pupil dilation duration (from tooltip-leave) of each ID to Fitts' ID value ($R^2 = 0.662$ and $p < .05$).*

tion can be employed as a valid indicator for task requirement in goal-directed movements.

This is a significant extension of the previous knowledge of pupil as an indicator of cognitive task workloads, to eye-hand coordination involving motor tasks. As the pupil dilation can be monitored uninterruptedly, it can be used for measuring the change of task loads to a human operator during a continuous performance. In fact, Gao et al. [2013] have used pupil responses to monitor workload changes on operators working at a nuclear power plant. Our results can be easily extended to the study of workloads of surgeons in the operating room setting.

Except for above main point, we also noticed some interesting findings from this study based on observation on the pupil behaviors. First, the peak pupil dilation occurred earlier with easier IDs than harder IDs, as shown in Figure 7. After the pupil reached its peak size, the pupil maintained a larger size in harder IDs than in easier IDs. It seemed that the participants maintained a relatively higher level of effort to collect visual input on the target during the harder pointing task, either on a smaller target or one with a longer travelling distance.

Second, after tool moving, as shown in Figure 7(b), the extent of pupil constriction was different among four IDs in the constant target size setting. It seemed that the extent of pupil constriction was associated with the tool travel distance: the longer the movement distance, the longer saccade, and the deeper pupil constriction. To support this, in the constant distance setting as shown in Figure 7(a), the extent of pupil constriction for different IDs during tool-leave and tooltip-arrive was maintained constant as target distance was kept the same in this experimental setting.

The third interesting observation came from the moment when the pupil started to react to the coming movement. As shown in Figure 7(a), the pupil started to react to tool movement at about $2s$ before tooltip-leave, and this varied among different IDs, i.e. for the harder ID the pupil started to dilate earlier than for the easier IDs. This may be a reflection of the level of mental preparation for an ongoing movement with different task requirement [Richer and Beatty 1985]; the smaller target size requires an earlier preparation.

Figure 8 shows a strong correlation between peak pupil dilation with task difficulty, and Figure 9 shows a correlation between the duration of the pupil dilation from the moment of tool leaving to peak value. For real-world application, some features such as peak pupil size and pupil dilation duration from tool leaving derived from the vicinity of the moment of movement start can be employed to classify the task difficulty levels.

Movement-related pupil response is easily mixed with pupil responses evoked by other events such as emotion and illumination changes. The effictive solution is the window averaging approach presented here, where many repetitions of short epochs of the same task are aligned at a specific common time point and averaged to remove the effect of unrelated events and preserve usful information.

The method by which we measured the task requirements using pupil diameter in a discrete elementary task, that is the ID-to-pupil mapping in this motor-coordination task, could be further applied to complex tasks which can be decomposed into several elementary tasks [Hoeks and Levelt 1993]. This opens the door towards objectively and continuously measuring the task workload during complex motor tasks using pupil diameter.

Furthermore, if we monitor and model pupil on short intervals, such as the 7-second intervals as was done here, we could make accurate predictions about the difficulties user experience. This would contribute to building more intelligent interfaces both for general interactive systems and for surgery procedures, for example for automatically evaluating and monitoring the task workload in a very fine resolution during image-guided procedures. Compared to other physiological signals such as Galvanic Skin Response (GSR), pupil size processing changes have low latency (around $100 - 200ms$ [Beatty 1982]), which allows for near real-time evaluation.

Knowledge gained from this study contributes to the understanding of how pupil responses indicate the changes of task requirement during a goal-directed movement. The research opens an opportunity to develop valid methods of measuring task load using pupil parameters available in eye-tracking technologies.

6 Conclusion

We showed that pupil diameter can be used as a reliable indicator of the task requirement in motor tasks, responding to both changes of target size, and also to changes in target distance. This is a piece of new evidence that validates Fitts' law in accounting for movement behavior in eyes, as well as in hands.

With low-cost eye-tracking sensors, it will be possible to embed such pupil size estimation in a variety of user interactive tasks and have intelligent systems to monitor the second-to-second workload of a user.

The results of this study have important implications to the applications of eye-tracking technologies in eye-hand coordination involved systems, e.g. interactive system and training and simulation environments.

Acknowledgements

We thank the Natural Sciences Research Council of Canada (NSERC), Royal College of Physicians and Surgeons of Canada (RCPSC) Medical Education Research Grant, and all the volunteers in the case study.

References

ABRAMS, R. A., MEYER, D. E., AND KORNBLUM, S. 1990. Eye-hand coordination: oculomotor control in rapid aimed limb movements. *Journal of Experimental Psychology: Human Perception and Performance 16*, 2, 248–267.

BEATTY, J. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin 91*, 2, 276–292.

BEDNARIK, R., VRZAKOVA, H., AND HRADIS, M. 2012. What do you want to do next: a novel approach for intent prediction in gaze-based interaction. In *Proc. ETRA 2012*, ACM, 83–90.

BERGUER, R., SMITH, W. D., AND CHUNG, Y. H. 2001. Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical Endoscopy 15*, 10, 1204–1207.

BRADSHAW, J. 1967. Pupil size as a measure of arousal during information processing. *Nature 216*, 5114, 515–516.

CASSENTI, D. N., AND KELLEY, T. D. 2006. Towards the shape of mental workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50*, 11, 1147–1151.

CHEN, H.-J., AND LIN, C. J. 2011. The investigation of laparoscopic instrument movement control. *Lecture Notes in Engineering and Computer Science 2198*, 1, 1309–1313.

ELLIOTT, D., HELSEN, W. F., AND CHUA, R. 2001. A century later: Woodworth's (1899) two-component model of goal-directed aiming. *Psychological Bulletin 127*, 3, 342–357.

FITTS, P. M., AND PETERSON, J. R. 1964. Information capacity of discrete motor responses. *Journal of Experimental Psychology 67*, 103–112.

FITTS, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology 47*, 6, 381–391.

GAO, Q., WANG, Y., SONG, F., LI, Z., AND DONG, X. 2013. Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics 56*, 7, 1070–1085.

GAWRON, V. J. 2008. *Human Workload*. CRC Press, 87–230.

GOLDINGER, S. D., AND PAPESH, M. H. 2012. Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science 21*, 2, 90–95.

GUTHRIE, D., AND BUCHWALD, J. S. 1991. Significance testing of difference potentials. *Psychophysiology 28*, 2, 240–244.

HESS, E. H., AND POLT, J. M. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science 143*, 3611, 1190–1192.

HOEKS, B., AND LEVELT, W. 1993. Pupillary dilation as a measure of attention: a quantitative system analysis. *Behavior Research Methods 25*, 1, 16–26.

IQBAL, S. T., ADAMCZYK, P. D., ZHENG, X. S., AND BAILEY, B. P. 2005. Towards an index of opportunity: understanding changes in mental workload during task execution. In *Proc. CHI 2005*, ACM, 311–320.

JIANG, X., ZHENG, B., TIEN, G., AND ATKINS, M. S. 2013. Pupil response to precision in surgical task execution. *Studies in Health Technology and Informatics 184*, 210–214.

JIANG, X., ATKINS, M. S., TIEN, G., BEDNARIK, R., AND ZHENG, B. 2014. Pupil responses during discrete goal-directed movements. In *Proc. CHI 2014*, ACM, In press.

KLINGNER, J., KUMAR, R., AND HANRAHAN, P. 2008. Measuring the task-evoked pupillary response with a remote eye tracker. In *Proc. ETRA 2008*, ACM, 69–72.

MACKENZIE, I. S. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction 7*, 1, 91–139.

MARSHALL, S. P. 2002. The index of cognitive activity: measuring cognitive workload. In *Proc. the 2002 IEEE 7th Conference on Human Factors and Power Plants*, IEEE, 7–5 – 7–9.

MORESI, S., ADAM, J. J., RIJCKEN, J., KUIPERS, H., SEVERENS, M., AND VAN GERVEN, P. W. M. 2011. Response preparation with adjacent versus overlapped hands: A pupillometric study. *International Journal of Psychophysiology 79*, 2, 280–286.

OTERO, S. C., WEEKES, B. S., AND HUTTON, S. B. 2011. Pupil size changes during recognition memory. *Psychophysiology 48*, 10, 1346–1353.

PRIVITERA, C. M., RENNINGER, L. W., CARNEY, T., KLEIN, S., AND AGUILAR, M. 2010. Pupil dilation during visual target detection. *Journal of Vision 10*, 10, 1–14.

PRYTZ, E., MONTANO, M., AND SCERBO, M. W. 2012. Using fitts law for a 3d pointing task on a 2d display: Effects of depth and vantage point. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 56*, 1, 1391–1395.

RICHER, F., AND BEATTY, J. 1985. Pupillary dilations in movement preparation and execution. *Psychophysiology 22*, 2, 204–207.

RICHSTONE, L., SCHWARTZ, M. J., SEIDEMAN, C., CADEDDU, J., MARSHALL, S., AND KAVOUSSI, L. R. 2010. Eye metrics as an objective assessment of surgical skill. *Annals Surgery 252*, 1, 177–182.

WICKENS, C. D. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science 3*, 2, 159–177.

ZHENG, B., CASSERA, M. A., MARTINEC, D. V., SPAUN, G. O., AND SWANSTROM, L. L. 2010. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surgical Endoscopy 24*, 1, 45–50.