Pupil Responses during Discrete Goal-directed Movements

Xianta Jiang[1,2]   **M. Stella Atkins**[2]   **Geoffrey Tien**[2]   **Roman Bednarik**[3]   **Bin Zheng**[4]

[1]Zhejiang University, Hangzhou, China, jxt@zju.edu.cn
[2]Simon Fraser University, Burnaby, BC, {xiantaj, stella, gtien}@sfu.ca
[3]University of Eastern Finland, Joensuu, Finland, roman.bednarik@uef.fi
[4]University of Alberta, Edmonton, AB, bzheng1@ualberta.ca

ABSTRACT

Pupil size is known to correlate with the changes of cognitive task workloads, but how the pupil responds to requirements of basic goal-directed motor tasks involved in human-machine interactions is not yet clear. This work conducted a user study to investigate the pupil dilations during aiming in a tele-operation setting, with the purpose of better understanding how the changes in task requirements are reflected by the changes of pupil size. The task requirements, managed by Fitts' index of difficulty (ID), i.e. the size and distance apart of the targets, were varied between tasks, and pupil responses to different task IDs were recorded. The results showed that pupil diameter can be employed as an indicator of task requirements in goal-directed movements—higher task difficulty evoked higher valley to peak pupil dilation, and the peak pupil dilation occurred after a longer delay. These findings contribute to the foundation for developing methods to objectively evaluate interactive task requirements using pupil parameters during goal-directed movements in HCI.

Author Keywords

Pupil diameter; goal-directed movement; movement-evoked pupillary response; Fitts' Law.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces- *Evaluation/methodology*

INTRODUCTION

The extent of achievement of a complex task critically relies on the task requirements while performing the task, since the human mental capacity is a finite resource [5]. This especially holds true in goal-directed movements with high task requirements. For example, surgeons undergo higher mental workload in a minimally invasive

laparoscopic surgery than in open procedures [3], since the task requirements in laparoscopic surgery are higher than in open surgery, due to the long tool shaft and the lack of depth information when projecting the scene inside the body to a 2-D screen. Therefore, it is highly desirable to objectively and continuously measure the task requirements as a surrogate for mental workload of the users during goal-directed movements. Such a method would allow indirect evaluation of the performance during the procedure.

Pupil size has been well proven to correlate with the information processing load in a variety of cognitive tasks [2], but has not been thoroughly explored in the field of HCI [16]. In order to apply pupillary indicators to task workload in the field of HCI research, pupil size changes during goal-directed movements should be explored first. For example, how do the pupil size changes correlate with the level of the task difficulty in an eye-hand coordination movement? Only a few works have been published in this research area, and to a very limited extent. One of the pioneer fundamental works was done by Richer and Beatty [24], who found that the pupil dilated as a function of the complexity of finger movements. However, this is not a real situation of 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.

This work aims to explore the pupillary responses during a discrete goal-directed movement with the task difficulty level clearly defined by Fitts' law [8] (the size and distance between targets were varied corresponding to difficulty level) to quantify the pupil dilation during movements. The findings from this study constitute the foundation for constructing methods towards objectively measuring the task requirements during complex goal-directed movements. Such measurements can be used for adjusting proactive responses of user interfaces, for instance in medical educational simulations.

RELATED WORK

Mental Workload and Measurement Methods

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 [10]. The subjective rating methods are easy to perform, and they only require the

operator to answer survey questions about the mental stress associated with a task. However, the answers to the survey questions are affected by the operator's working memory and have a bias to the task difficulty [26].

The subjective shortcomings may be avoided by using a secondary task to the primary task where decreased secondary task performance means increased mental workload caused by the primary task. However, there may be extra workload caused by the secondary task that affects the performance of the primary task.

Although psychophysiological measures have the advantage of eliminating the possibility of subjective distortions and generally do not interfere with task performance, most of them present an intrusion, as they need sensors attached to the user. Modern eye-tracking systems can monitor the eye movements remotely. Pupil size data are a by-product of virtually all psychophysical experiments which use eye tracking and are thus available without adding any further technology to the experiment.

Pupil Dilation as an Indicator of Mental Workload

With advances in eye-tracking technology, pupil diameter can be recorded remotely and its subtle changes have been proven to indicate the cognitive task load [19].

The pioneer work exploring pupil size changes relating to mental workload and task difficulty was done by Hess and Polt [13] 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 have shown that the changes of pupil size reflect mental workload and the level of difficulty in different tasks, e.g., mental arithmetic tasks [4], recall or memory tasks [11, 22], and visual search tasks [23].

Pupil size also responds to the critical event 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 [2]. The TEPR has been employed as an approach to capture and evaluate the mental workload changes during a variety of tasks [2].

Pupillary Responses to Mental Workload in HCI

In the HCI domain, Iqbal et al.[15] 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 that the pupil diameter is relatively low during task boundaries, which are suitable for interruption with lower mental workload.

Richer and Beatty [24] examined the pupil responses to simple movement, the self-paced finger flexion. The authors found the typical pupil dilation pattern to a simple movement is that pupil dilates at around 1.5s before the movement and peaks afterwards at around 0.5s. This pupil dilation pattern was confirmed by other works [21, 23]. For example, in a visual target search study, Privitera et al.[23] 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 field of eye-hand coordination in health care training, Jiang et al. [17] investigated the pupil changes in a simulated laparoscopic task, and found that the pupil dilates faster during the execution of higher task difficulty subtasks, 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 [20] 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 [25]. However, details of the pupil response to motor tasks were not reported.

In summary, previous research confirmed that pupillary response reflects the internal cognitive processing. If we want to build interactive systems that accommodate various types of user's workload and respond to it, it is fundamental to investigate the relationships between the pupillary response and the varying task demands.

A handy conceptualization of task requirement is to apply Fitts' law's index of difficulty. Because no knowledge is available about the relationships between the motor goal-directed interactive tasks, the objective assessment of their difficulty, and the respective pupillary response, we conducted the study reported here.

Our hypothesis was that increased index of difficulty would result in distinguishable patterns of pupillary response. Confirmation of such a hypothesis would mean establishment of new measurements of the difficulty of fundamental motor tasks in HCI.

USER STUDY

To examine the pupil's responses to task difficulty, we designed a goal-directed aiming task using a laparoscopic grasper, in order to answer the question: do pupil size changes correlate with increases of task requirement during goal-directed movements?

Experiment Design

The task was to move the grasper horizontally to touch the circles printed on a paper inside a laparoscopy training box (see below for details), while the eyes of the user were looking at the projection of the inside of the box on a display screen, as shown in Figure 1. Three pairs of target circles with combinations of variable size and distance between targets were designed with three indices of difficulty (ID). The detailed parameters of the target circles are illustrated in Figure 2. The formula for calculating ID is as in Equation 1 [8]:

$$ID = -\log_2 \frac{W}{2A} \qquad (1)$$

where W is the diameter of the target circle and A is the distance between the pair of targets.

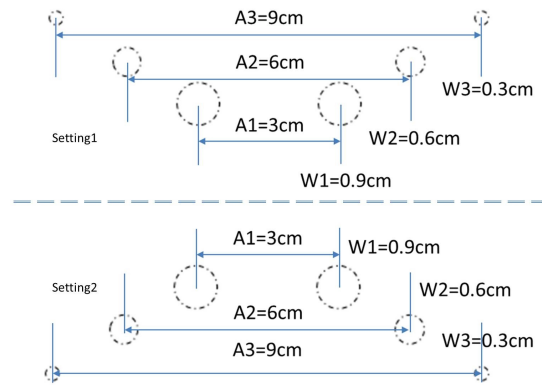**Figure 1. Experimental setting; Tobii X50, training box, and web camera.**

Figure 2. Illustration of target patterns (the pairs of target circles on the paper inside the training box showing on the screen as in Figure 1). One group of subjects executed bottom up from the easiest task to the hardest task and then back to the easiest task (shown in the upper panel); another group of subjects executed bottom up from hard to easy tasks and then back to the hardest task (lower panel). A1 to A3 represent the distances between targets and W1 to W3 represent the sizes of the targets of ID1 to ID3.

In our experiment, we employed the following settings: ID1=2.7 bits/response (W1=0.9cm, A1=3cm) denoted here as Easy, ID2=4.3 bits/response (W2=0.6cm, A2=6cm) denoted here as Middle, and ID3=5.9 bits/response (W3=0.3cm, A3=9cm) denoted as Hard.

Since the frequency of pupil response to a movement is typically lower than 0.5Hz [21, 23, 24] and the frequency of the Fitts' tool movement under laparoscopic environment is usually around 0.5Hz [6], in order to avoid overlap of the pupil dilation curve, we used discrete Fitts' pointing [9] by waiting 10s 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.

Participants and Apparatus

Twelve participants (four females) were recruited to the study, including ten graduate students, one undergraduate student, and one staff member from a local university. All were right-hand users and had normal or corrected-to-normal vision. None of them were previously trained in any surgical procedures.

The task was performed using a surgical training box for laparoscopic procedures (Laparoscopic Trainer, 3-D Technical Services, Franklin, OH). Subjects held a surgical grasper at a standing pose about 60cm from the eye-tracker. The tips of the grasper were black taped 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 a 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 integrated using Tobii software, Clearview 2.7.0. 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 [23].

Tasks and Procedure

The task was to move the surgical tool to point to the circles printed on a piece of paper at the bottom of the training box. This was a discrete pointing task, so that participants had to wait 10 seconds before each move to the next circle. Specifically, a trial consisted of 16 discrete movement steps, each separated by 10s, with the execution sequence shown in Figure 3. The trial started by placing the tooltip on the right bottom circle for 10s, then moving the tool to the left bottom circle (step (1)), and ended by stopping the tooltip on the right bottom circle for 10s after step (16). Only the 12 horizontal moves were used for 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 180s.

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, each containing three trials with 20s break between trials. To achieve counterbalance, each block used either task setting 1 (from easy to hard as shown in top panel in Figure 2) or task setting 2 (from hard to easy as shown in bottom panel in Figure 2). Half of the participants started with setting 1 and the other started with task setting 2. We did not consider other ordering options like executing from middle difficulty to hardest to easiest, because the ANOVA results showed there was no significant difference in movement time between the two groups, and no interaction effect of ID and group order.
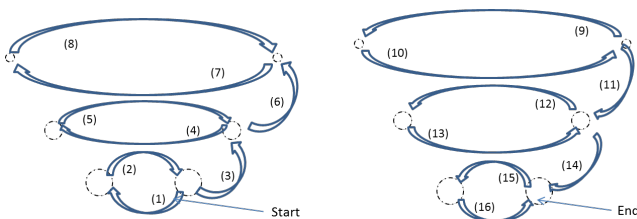

Figure 3. Execution sequence of a trial (target setting 1), steps shown in parentheses. The arrows represent tool movement from one circle to another; before each tool movement, the tool stops on the circle for 10s.
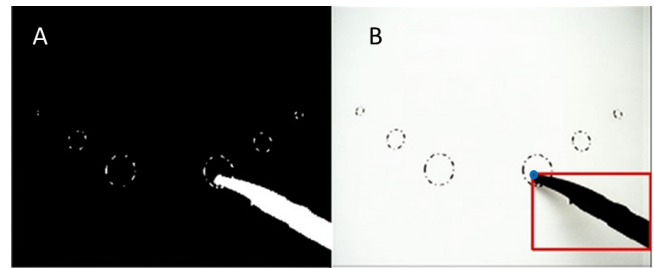

Figure 4. Key steps in detecting tooltip from surgical videos. Panel A shows the binary-thresholded image with mostly the tool left. Panel B shows the recognized tool (in red rectangle) and tooltip (the blue dot).

Data analyses

The task video recorded at 30Hz and resolution of 352×288 pixels by the camera attached to the training box, and the text files containing eye movements (50Hz) were exported from Tobii Clearview for the offline analysis. The tooltip position and eye movements were calculated and then synchronized for subsequent analysis.

Location of the tooltip

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 while 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.

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 in 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.

Identification of moments of tooltip leaving and arriving

Some critical moments of tooltip movements—the moments when the tooltip started to move (tooltip-start) and reached (tooltip-reach) 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-start and tooltip-reach, by checking whether the x-axis absolute velocity is lower than a threshold. The velocity thresholds for detecting the tooltip-start and tooltip-reach moments were empirically determined, since
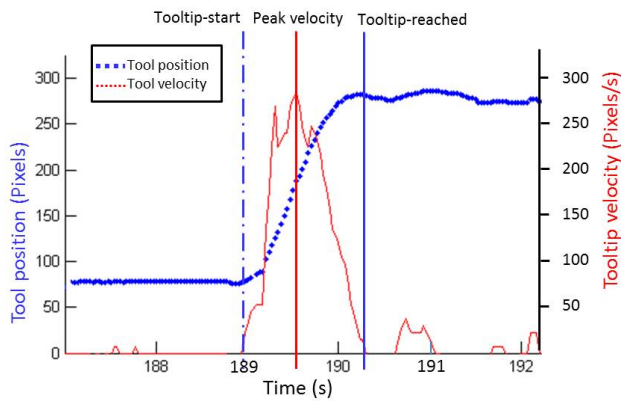
Figure 5. Kinematics of tool movement (subject 01, trial 01) during a typical horizontal movement between circles. The blue curve is the tooltip position in pixels along the horizontal line between the circles, and the red curve is the velocity of tooltip.

there are abrupt changes of tool velocity at the moments of tooltip-start and tooltip-reach. The tooltip-start and tooltip-reach thresholds were both set to 30pixels/s. The moment of tool-reach detected by the algorithm is the moment that the tool quickly arrives at a position above (but not necessarily touching) the target circle. An example of kinematics of tool movement is shown in Figure 5.

Pupil data processing and segment extraction
Segments of missed pupil data shorter than 440ms were interpolated, since they might have contained blinks. 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 [24].

The pupil diameter data sequences were extracted from a 7-second window starting 3 seconds before and 4 seconds after every tooltip-start. The 7-second window is motivated by considering the extra movement time (about 1s-2s) before the tool leaves in our study compared with others who noted pupil increases start 1.5s before the stimulus [21, 23, 24] and the pupil continuous to dilate afterwards to a peak value in a few seconds e.g. 3s-window in [23] 4s-window in [24] and 5s-window in [21]. 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 first second

of the window. All the data in the windows were aligned at the tooltip-start (3 seconds into the window), and the mean pupil diameter change was calculated for each time point in the window across all horizontal tooltip moves from all trials. Assuming there are m moves from all trials of all subjects and each move has n samples in a 7-second window, $X = \{x_{ik}, i = 1...m, k = 1...n\}$ is the matrix of pupil diameter values in the 7-second windows across all tooltip moves of all trials of all subjects. The mean pupil diameter change of each time point in the window is in a

vector $\overline{X} = \dfrac{\sum\limits_{i=1}^{m} x_{ik}}{m}, k = 1...n$. 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 segments where significant differences occurred in the 7-second window between the three IDs, the graphical significance testing approach [12, 23] 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 [23].

RESULTS
A total of 72 trials were recorded (12 participants, each performed 6 trials). Three trials 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 69 valid trials, there was window data for 828 movements available (each trial has 6 ID executions as each ID has been executed twice, with each ID execution having 2 moves—from right to left and left to right). However, we discarded 20 windows due to mis-operation such as when the participant moved the grasper to a wrong target. Therefore we have 808 valid movements.
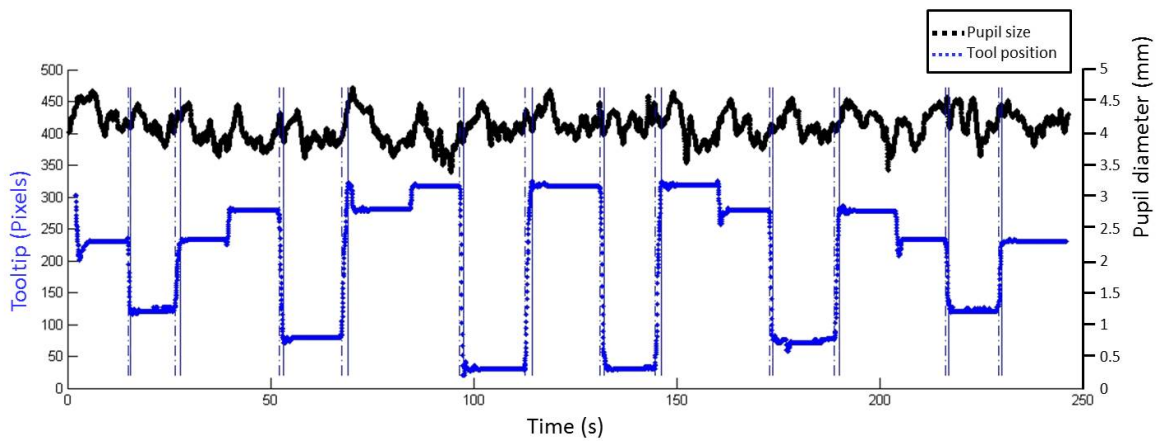
Figure 6: Example of tooltip movement (blue) and pupil size (black) over time for a complete trial (subject 01, trial 01). The blue curve is the tooltip position in pixels along the horizontal line between the circles. The dash and solid vertical lines represent the moments of tooltip-start and tooltip-reach.
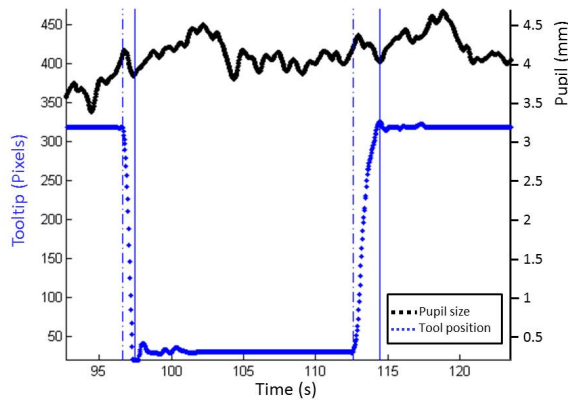

Figure 7. Blow up of a segment (95s-120s) from Figure 6, showing how the pupil size (black curve) increases when tool movements occur. The blue curve is the tooltip position in pixels along the horizontal line between the circles. The dash and solid vertical lines represent the moments of tooltip-start and tooltip-reach.

Figure 6 shows an example of pupil size aligned with horizontal tooltip movement over time for a complete trial. The pupil responses to the tooltip movement are actually recognizable in the graph, although there are several more fluctuations which might relate to pupil unrest. Figure 7 shows a blow-up of a segment of the trial data from Figure 6; there are significant dilations when the tool moves.

Performance

The mean movement time (MT) is the mean transportation time between tooltip-start and tooltip-reach for all horizontal movements. The mean MT for all IDs is 1.0±0.3s, and the MT changes as a function of ID, i.e., the mean MT increases when the difficulty level increases (Figure 8). There is significant main effect between three IDs in terms of mean MT ($F_{(2,805)} = 209.868$, $p < .001$). Post Hoc test (Tukey HSD) shows that the differences between
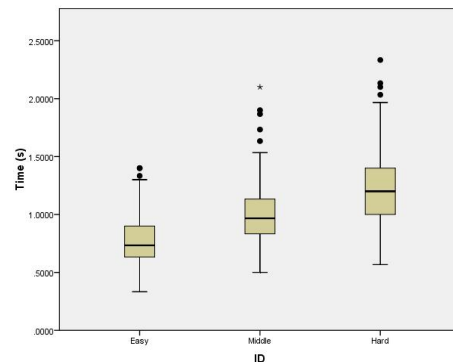

Figure 8. Box-whisker plot for Mean movement time for each ID.

pairs of the three IDs are significant ($p < .001$), with mean MT of Easy, Middle, and Hard IDs being 0.8±0.2s, 1.0±0.2s, and 1.2±0.3s respectively.

Pupil Diameter Changes

Descriptive analysis

Figure 9 shows the mean changes of pupil diameter during horizontal tooltip movements over a 7-second window. The data is averaged across 808 valid moves of 69 trials from 12 subjects. For clarity, error bars are drawn every 400ms rather than drawn for every time sample (at 50Hz).

Figure 10 shows the means of pupil size changes in the 7-second window for three IDs. The three ID curves share a common pupil change pattern as shown in Figure 9, i.e., all of them dilate in response to the movement preparation and execution in a relatively long period from about 1.5s before tooltip-start and peak about 1s after the tool is reached, corresponding to movement preparation and execution [23, 24]. There is also a small constriction 200ms after the tooltip starts to move, between the moments of tooltip-start and tooltip-reach.
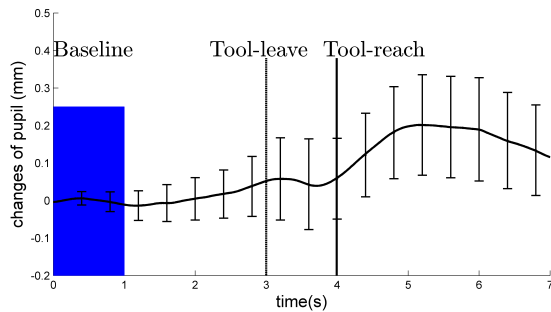
Figure 9. Mean pupil diameter changes for 808 valid moves of 69 trials from 12 subjects. Data were aligned over a 7 second window 3 seconds before the tooltip-start. The baseline is defined as the mean diameter of the pupil over the first second of the window, and the solid black curve is the mean pupil diameter change from the baseline over time. The black vertical dashed line is tooltip-start where all the data are aligned and the vertical solid black line is the average tool-reach time. The error bars for 1 std. dev. are drawn every 400ms.
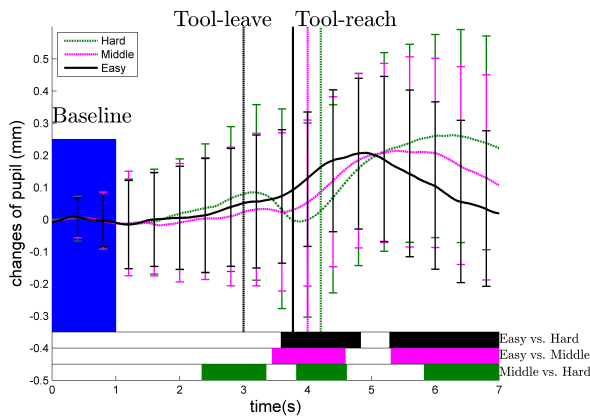

Figure 10. Mean pupil diameter changes against different IDs; data are aligned over a 7-second window around tooltip-start. The vertical dash black line is the tooltip-start and other three solid color vertical lines represent the tooltip-reach moments of three IDs respectively. The three colors of bars at the bottom indicate significant differences in pupil dilation between Easy, Middle and Hard ID with black representing Easy vs. Hard, pink representing Easy vs. Middle, and green representing Middle vs. Hard. The error bars for 1 std. dev. are drawn every 400ms.

However, major differences can be observed between the three IDs. First, during 2-3s in Figure 10, the pupil dilation for the Hard ID (green curve) is more than Easy (black) and Middle (pink) ones. Second, during tool moving, the pupil of the Hard ID constricts the most among three IDs and the pupil of the Easy ID constricts the least. Third, the peak value and duration from valley to peak pupil size of the three ID curves are different: the Hard ID has the highest peak and longest duration and the Easy ID has the smallest peak and the shortest duration.

Furthermore, the increase of pupil size from the valley (between 3s and 4s along the time window in Figure 10) to its peak correlates to IDs: the easy task (black curve) increased about 0.13mm (0.07mm at valley to 0.2mm at peak), the middle task (pink curve) increased about 0.18mm (0.02 to 0.2mm), and the hardest task (green) had the largest increase of about 0.27mm (-0.01 to 0.26mm). Also, the valley to peak duration can be observed to be different between IDs.

Statistical Analysis

Graphical significance testing was applied between the pupil size change curves for the three IDs, with the results as shown in the bottom bars in Figure 10. For each time point, the paired *t*-test was performed in two IDs each containing samples from the 69 trials.

Significant differences in pupil size between Easy ID and Hard ID are shown in black bars (upper row) in Figure 10. The first significant time period is located at 3.7s to 4.9s along the time window, where the Hard ID pupil curve (green) is at its local valley while the Easy ID curve (black) has already recovered from a short period of constriction. The second significant time period starts at the intersection of Hard ID curve and Easy ID curve (at around 5s) to the end of the window.

The pink bars (middle row) in Figure 10 show similar significant differences between Easy ID and Middle ID, only with the first bar starting a little bit earlier (3.5s to 4.6s in time axis).

The green bars (bottom row) in Figure 10 show significant differences between Middle ID and Hard ID. There are three major significant time periods: the first one (2.3s to 3.4s along the time window) where the pupil dilates more during the hard ID than the Middle ID, the second one (3.8s to 4.6s along time window) where the pupil constricts more during the Hard ID than the Middle ID, and the third one starts after 5.8s along the time window.

According to three color bars in Figure 10, significant differences between all three IDs can be found in two major time periods: one is around tool-reach where the hardest ID has deeper pupil constriction than easier ones, and another area is after the peak, where the hardest ID has bigger pupil diameter than the easier IDs.

ANOVA shows there is significant difference in the pupil size increase from valley to peak between IDs ($F_{(2,805)} = 44.754$, $p < .001$). Post Hoc test (Tukey HSD) also shows that the mean increase of pupil size from valley to peak is significantly different ($p < .001$) between Easy (0.3±0.2mm), Middle (0.4±0.2mm), and Hard ID (0.5±0.2mm).

ANOVA shows there is significant difference in the duration of the pupil dilation from its valley to peak size between IDs ($F_{(2,805)} = 91.938$, $p < .001$). Post Hoc test (Tukey HSD) shows that the duration from valley to peak is

significantly different ($p < .001$) between Easy (1.6±0.7s), Middle (2.1±0.9s), and Hard ID (2.7±1.1s).
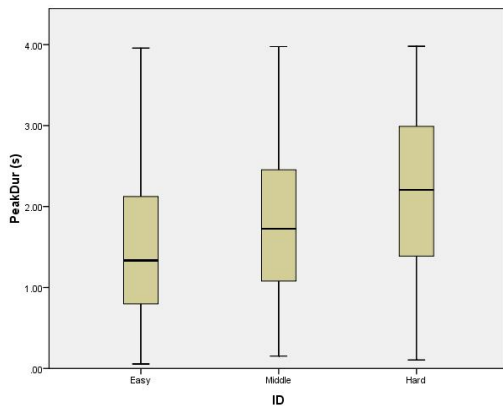

Figure 11. Box-whisker plot for Mean duration from tooltip-reach to the moment where the pupil peaked in size for three difficulty IDs.

Figure 11 shows the duration from tooltip-reach to the moment where the pupil peaked in size for each ID, up to 4s after the tool-tip reach. There is significant main effect between three IDs in terms of this duration ($F_{(2,805)} = 30.558$, $p < .001$). Further Post Hoc test (Tukey HSD) shows that this duration for the Hard ID (2.2±1.0s) is significantly ($p < .001$) longer than that of Easy (1.6±1.0s) and Middle (1.8±0.9s), and the Middle ID is significantly longer than that of Easy ID ($p < .01$).

DISCUSSION

Our work is the first one in the HCI literature that relates pupillary response and task difficulty measured by Fitts' Law. Previous research investigated either movement time (MT) and task difficulty, or pupillary response and mental workload. The present study bridges these two directions and uncovers new knowledge about the relationships between tool-movement difficulty and respective pupillary response. Previous work [15] conducted investigations for cognitive tasks; this study however, examines and takes the results to the motor domain of HCI.

The performance measures exhibited expected behavior—the MT correlated perfectly to the increase of the ID, as shown in Figure 8. Therefore, Fitt's ID is a good measure for the task requirement in this study. Although it would be interesting to see if the differences hold if the movement time is normalized by distance and then compared across the three difficulty levels in this study, we did not do that since our other research [18] showed the pupil dilations and the movement time in goal-directed movements still respect Fitts' Law, by only changing the target distance or target size.

The pupil response to simple goal-directed movement tasks was shown by our data, and furthermore, the different task IDs elicited different pupil responses. Thus our hypothesis, that increased index of difficulty would result in

distinguishable patterns of pupillary response, was confirmed.

The pupil dilation started 1.5s before tooltip-start and continued 2s afterwards (Figure 9) and clearly reveals the task requirement during the preparation and execution of the movement. Our data is consistent with the finding of previous works [21, 23, 24]. The time that the pupil starts to dilate (1.5s before tooltip-start) in this study is same as that of Richer and Beatty's finding, while the peak duration (2s after tooltip-start) is longer than that of Richer and Beatty's (0.5s). The likely reason is that we have a tool movement time around 1s, while Richer and Beatty's is only a simple finger flexion.

The pupil constricts slightly between tooltip-start and tooltip-reach (Figure 9), which is the time when the tooltip moves speedily to a target. During the initial phase of the aiming movement, the eye gaze usually saccades to the target ahead of tooltip for the visual guidance [1, 7]; the saccade causes the pupil constriction. Our results show that longer travel distances require longer saccades, causing deeper pupil constriction, as shown in Figure 10.

The two significantly different time periods for pupil size (the pupil constriction around the tool-leave movement, and the peak pupil size after the tool-reach) were proven by the graphical significance testing shown in Figure 10, where the three pairs of IDs (Easy vs. Hard, Easy vs. Middle, and Middle vs. Hard) all show significant differences.

Although the peak pupil size of Easy and Middle ID shows similar amplitudes in Figure 10, the increase of pupil size from valley to peak is significantly different between all three IDs.

Furthermore, the delay of peak pupil dilation (represented by the duration between tooltip-reach and peak pupil size, as shown in Figure 11) was correlated with the three IDs, i.e. the Hard ID has the longest delay, and the Middle ID has the middle one. This is consistent with the finding of previous work [24].

The 7s window approach of averaging many repetitive task epochs to show the pattern of pupil size change may not be able to measure task requirement in real-time. However other features introduced here, such as the valley to peak pupil size dilation, the valley to peak duration, and the duration between the moments of tool-reach and peak pupil size, can be employed to classify the task difficulty of each targeting step alone using techniques such as machine learning. The ANOVA results showed that all these features have strong ($p < .001$) ability in distinguishing the three difficulty levels of task requirement.

For real-world application, several steps should be carefully considered. First, the critical moment of the movement (the moment the tool/hand starts to move or reach) should be accurately detected as a timing marker of the movement event that the pupil responses to. Usually tool-leave is more

promising as a movement event marker than tool-reach since in a complex movement such as surgical tasks [17] the tool may need extra actions at the end of movements such as grasping or releasing an object. One effective and robust way to detect tool-leave and tool-reach is to first find the moment of peak velocity of the tool movement and then search backward and forward to get tool start and reached moments, as presented here.

Second, the features of pupil size change should be extracted within a window in the vicinity of the movement start. The window size should be properly defined according to the property of the application, since if the window is too small or too wide, useful pupil information may be excluded, or false pupil information introduced.

Third, the baseline of pupil diameter used for deriving each relative pupil diameter change in the window should be carefully chosen. In the present discrete movement study, the selection of the position of the baseline is quite straightforward, using the period between 2s to 3s before the tool-leave moment, which is near the end of the 10s waiting time and is right before the pupil starts to dilate. However in a real-world motor task where the movements may occur consecutively without a clear waiting time between them it may be a challenge to select the position of the baseline.

The luminance of the screen was well controlled to induce a mean pupil diameter of 4.1mm (±0.8mm) which falls in the center of length-tension curve of the iris muscle, enabling the pupil to capture the cognitive changes well. ANOVA results show that there was no significant difference between three IDs in terms of mean pupil size in baselines of the moves in each ID.

The results have important implications for the design of interactive systems and training and simulation environments. First, we showed that elementary task demands can reliably be measured in real-time. With the availability of low-cost eye-tracking sensors, it will be possible to embed such pupil size estimations in a variety of user interactive tasks and have intelligent systems to monitor the second-to-second workload of a user. Therefore, not only internal processing load can be estimated, but also the motor-level component of workload can now be incorporated into the interaction design.

Second, 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 [14], and is a subject for future research.

Third, if we monitor and model the pupil over short intervals, such as the 7-second intervals as done here, we can make accurate predictions about the difficulties user experience. This contributes to building more intelligent interfaces both for general HCI and for surgery procedures,

for example for automatically evaluating and monitoring the task workload in a very fine resolution during image-guided procedures.

CONCLUSION

The results of this study answered positively the proposed question that the pupil response pattern is distinguishable for different levels of task difficulty. The pupil diameter increase for harder task IDs has a higher valley to peak value and longer delay.

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 for us to develop valid methods of measuring task load using pupil parameters in HCI.

REFERENCES

1. Abrams, R.A., Meyer, D.E., and Kornblum, S. Eye-hand coordination: oculomotor control in rapid aimed limb movements. *Journal of Experimental Psychology: Human Perception and Performance 16*, 2 (1990), 248-267.

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

3. Berguer, R., Smith, W.D., and Chung, Y.H. Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical Endoscopy 15*, 10 (2001), 1204-1207.

4. Bradshaw, J. Pupil size as a measure of arousal during information processing. *Nature 216*, 5114 (1967), 515-516.

5. Cassenti, D.N. and Kelley, T.D. Towards the shape of mental workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50*, 11 (2006), 1147-1151.

6. Chen, H.-J. and Lin, C.J. The investigation of laparoscopic instrument movement control. *Lecture Notes in Engineering and Computer Science 2198*, 1 (2011), 1309-1313.

7. Elliott, D., Helsen, W.F., and Chua, R. A Century Later: Woodworth's (1899) Two-component Model of

Goal-directed Aiming. *Psychological Bulletin 127*, 3 (2001), 342-357.

8. Fitts, P.M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology 47*, 6 (1954), 381-391.

9. Fitts, P.M. and Peterson, J.R. Information capacity of discrete motor responses. *Journal of Experimental Psychology 67*, (1964), 103-112.

10. Gawron, V.J., *Human Workload*, in *Human Performance, Workload, and Situational Awareness Measures Handbook, Second Edition*. 2008, CRC Press. 87-230.

11. Goldinger, S.D. and Papesh, M.H. Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science 21*, 2 (2012), 90-95.

12. Guthrie, D. and Buchwald, J.S. Significance testing of difference potentials. *Psychophysiology 28*, 2 (1991), 240-244.

13. Hess, E.H. and Polt, J.M. Pupil size in relation to mental activity during simple problem-solving. *Science 143*, 3611 (1964), 1190-1192.

14. Hoeks, B. and Levelt, W. Pupillary dilation as a measure of attention: a quantitative system analysis. *Behavior Research Methods 25*, 1 (1993), 16-26.

15. Iqbal, S.T., Adamczyk, P.D., Zheng, X.S., and Bailey, B.P. Towards an index of opportunity: understanding changes in mental workload during task execution. In *Proc. CHI 2005*, ACM Press (2005), 311-320.

16. Iqbal, S.T., Zheng, X.S., and Bailey, B.P. Task-evoked pupillary response to mental workload in human-computer interaction. *Ext. Abstracts CHI 2004*, ACM Press (2004), 1477-1480.

17. Jiang, X., Zheng, B., Tien, G., and Atkins, M.S. Pupil response to precision in surgical task execution. *Studies in Health Technology and Informatics 184*, (2013), 210-214.

18. Jiang, X., Zheng, B., Tien, G., Bednarik, R., and Atkins, M.S. Pupil dilations during target-pointing respect Fitts' Law. In *Proc. ETRA 2014*, ACM Press (2014), In press.

19. Klingner, J., Kumar, R., and Hanrahan, P. Measuring the task-evoked pupillary response with a remote eye tracker. In *Proc. ETRA 2008*, ACM Press (2008), 69-72.

20. Marshall, S.P., Method and Apparatus for Eye Tracking and Monitoring Pupil Dilation to Evaluate Cognitive Activity. 2000: United States.

21. Moresi, S., et al. Response preparation with adjacent versus overlapped hands: A pupillometric study. *International Journal of Psychophysiology 79*, 2 (2011), 280-286.

22. Otero, S.C., Weekes, B.S., and Hutton, S.B. Pupil size changes during recognition memory. *Psychophysiology 48*, 10 (2011), 1346-1353.

23. Privitera, C.M., Renninger, L.W., Carney, T., Klein, S., and Aguilar, M. Pupil dilation during visual target detection. *Journal of Vision 10*, 10 (2010), 1-14.

24. Richer, F. and Beatty, J. Pupillary dilations in movement preparation and execution. *Psychophysiology 22*, 2 (1985), 204-207.

25. Richstone, L., et al. Eye metrics as an objective assessment of surgical skill. *Annals Surgery 252*, 1 (2010), 177-182.

26. Zheng, B., Cassera, M.A., Martinec, D.V., Spaun, G.O., and Swanstrom, L.L. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surgical Endoscopy 24*, 1 (2010), 45-50.