Pupil Response to Precision in Surgical Task Execution

Xianta JIANG[1,2], M.Sc, Bin ZHENG[3], M.D Ph.D., Geoffrey TIEN[2], M.Sc, M. Stella ATKIN[2], Ph.D.

[1] College of Computer Science and Technology, Zhejiang University, *jxt@zju.edu.cn*

[2] School of Computing Science, Simon Fraser University, *stella@sfu.ca*

[3]Department of Surgery, University of Alberta, *bzheng1@ualberta.ca*

Abstract. Task-evoked pupil response (TEPR) has been extensively studied and well proven to be sensitive to mental workload changes. We aimed to explore how TEPR reflects mental workload changes in a surgical environment. We conducted a simulated surgical task that has 3 different subtasks with different levels of motor precision and different mental workload requirements. We found a significant effect among these different subtask groups by measuring pupil diameter change rate. This finding sheds light on improving patient safety in a real operating room by non-intrusively monitoring the surgeon's mental workload during performing a surgery using an eye-tracking system.

Introduction

The pupil has been well proven to react not only to the ambient light and psychological changes, but also to the intensity of mental effort demanded by the task performance [1]. Hess and Polt [2] found that the pupil gradually dilated when preparing the answer to a multiplication problem and reached a peak immediately before the answer was orally reported; then, it rapidly constricted back to the original size. They also found that the mean pupil dilation was a function of the level of difficulty of the problem.

The pupil size changes not only in response to the task difficulty overall, but also with respect to critical events during an information processing task, called the task-evoked pupillary response (TEPR) [3]. TEPR has been extensively studied and proven to be an efficient index of mental workload while participants are performing tasks. This has been tested in tasks of driving vehicles and airplanes [4, 5], interacting with computer interfaces [6], and performing surgery [7].

Objectively measuring mental workload in performing a laparoscopic operation has been proposed by several researchers [8-10]. Berguer et al. [10] found that performing laparoscopic surgery causes higher mental stress than open surgery by measuring physiological signals, i.e., skin conductance level (SCL) and blinks, collected from the participants when performing simulated tasks in both laparoscopic and open operating situations. Results have been confirmed by Zheng et al. using psychological assessment [8]. More recently, Zheng et al. [11] found that blink rate and mental workload of the participants were correlated during a simulated surgical task. Richstone et al. [7] used eye movement behaviors including pupil movements and blinks to predict surgical skill using linear discrimination analysis (LDA) and nonlinear neural network analyses (NNA), which distinguished expert and novice surgeons at 81.0% and 90.7% accuracy respectively in the live operating room setting. However, few studies directly report the relationship between pupillary change (TEPR) and mental workload of surgeons in performing a surgical task.

To examine the relationship between TEPR and surgical workload, we conducted an experiment in a surgical simulation lab where the participants were required to perform a simulated laparoscopic task with their pupillary movements recorded. The task includes 3 groups of subtasks, demanding different levels of intensity of mental workload. We hypothesized that the change rate of the pupil diameter would reflect the precision requirement of the subtask, i.e. the pupil diameter would increase faster during performing higher precision subtasks than lower precision subtasks. Similarly, the pupil would constrict rapidly while performing subtasks with low precision requirement.

Our research goal is to test the hypothesis that TEPR indicates the level of precision of subtask in a surgical task execution. Thus the pupil measurements can be used for monitoring the mental workload of the surgeon when performing a surgery. This will improve the patient safety in a real operating room.

Methods & Materials

Experiment setting & Apparatus

This study was conducted in a surgical simulation room at the Centre of Excellence for Surgical Education & Innovation, in Vancouver General Hospital. 12 subjects were recruited including surgeons and office staff. Each participant signed a consent form before the study.

As shown in Figure 1, there were three components in the experiment setting: a laparoscopic training box, a remote eye tracking system, and a web camera. The laparoscopic training box (Laparoscopic Trainer, 3-D Technical Services, Franklin, Ohio) was equipped with a single grasper in one of the four entrance ports. The remote eye-tracker (Tobii 1750, Tobii Technology, Sweden) has a built-in 17" LCD display, and records eye gaze points on the display at 50 Hz, if the operator stands 60-70 cm away and avoids large head movement. The web camera (C525 HD Webcam, Logitech, Fremont, CA) was placed below the display to record facial expressions of the operator for validation purpose, at 30 Hz.

Task & Procedure

The task was to transport the green rubber peg between three dishes using the grasper as shown in Figure 1. Each participant was given a brief oral description of the task and practiced a few minutes to become familiar with the task before starting to perform the task. Each participant performed five trials, with a pause between each trial. At the beginning and end of each trial, a camera flash was given for synchronization purpose. The ambient light was constant and controlled for all trials.

A trial has 9 subtasks as shown in Figure 2, which can be grouped into three basic movements with different precision requirements: reaching and grasping the object (RG), transporting and releasing the object (TR), and bringing the instrument to the home position in the white central square (H). Taking the first three subtasks as an example, starting from the home position, the grasper was moved to the red dish (6 mm) and picked up the peg (2 mm) (RG). The peg was then transported from the red to the green dish using the grasper, which was opened to release the peg into the green dish (TR). After releasing the peg, the empty grasper was moved to the home position (H). Each basic movement was repeated three times until the peg was brought back to the original place and the grasper back to the home position.

When the participants were performing reaching and grasping (RG), greater mental effort was demanded since the operator had to finish a complex process where speed and motor need to be well controlled. First, the operator had to control the grasper moving perfectly, i.e., the operator had to slow down and open the grasper when it approached the target after a relatively fast move from the home position. Second, the operator had to decide and locate a proper position for the tool tip of the grasper to stop for the peg. Third, the operator had to perform the grasping action very carefully to avoid dropping the peg; even after successfully grasping the peg, he had to lift up the peg carefully and in the right way, avoiding touching the edge of the dish. Transporting the object to a cup (TR) might be also demanding as it required the subject to carefully place the peg into the 6 mm dish. In contrast, bringing the empty grasper back to the home position (H) was less demanding.

Data Analysis

Surgical videos were captured from the display on the Tobii 17” monitor while participants were performing the task, and were manually annotated by recording the start time of each subtask in milliseconds. The captured surgical video contains the whole trajectory of the tool tip of the grasper and peg movement. The criteria for judging the start of a subtask are as follows:

- **The start of the RG movement:** the first frame that the tool starts to move towards the target dish after being in the home position.
- **The start of the TR movement:** the first frame that the tool moves after successfully grasping the peg and lifting it up above the dish.
- **The start of the H movement:** the first frame that the empty grasper moves towards the home position after successfully releasing the peg into the dish.

Since we are interested in the pupillary changes to the task requirement, we need to adjust the pupil diameter (ranging from 3.18 mm to 5.60 mm among the 12 subjects) by an appropriate baseline diameter. The baseline for each trial is calculated based on the average pupil diameter of the samples in the 400 ms period around the start of RG1 (200 ms before and 200 ms after the start of RG1), as shown in the pink background in Figure 3. Blinks and artifacts data were discarded prior to the baseline adjustment.

Results

Theoretically, we should record a total 540 subtasks produced by the 12 participants each performing 5 trials (12 subjects × 5 trials × 9 subtasks); however 6 subtasks were excluded because of peg dropping. As a result, we had data for 534 valid subtasks for analysis. For each subtask, the rate of pupil diameter change was calculated as the slope of the adjusted pupil diameter over time using a simple linear regression.

We performed a one-way ANOVA analysis on the 534 valid subtask data to find pupillary response differences among the three groups of subtasks, i.e., RG, TR, and H, based on adjusted pupil diameter and the slope of rate of change of pupil diameter. The output of the analysis is shown in Table 1.

Adjusted pupil size did not show significant differences among the three types of movement (Row 1 in Table 1. $F(2,33) = 1.553$, $p = 0.227$). The largest pupil size was recorded in performing the TR movement, rather than in RG movement, which seemed contradictory to our expectation. Possible reasons for this are given in the discussion session.

When using the slope of rate of change of pupil diameter for one-way ANOVA analysis, we found a significant effect among RG, TR, and H groups on value ($F(2,33) = 41.837$, $p < 0.001$), as shown in row 2 in Table 1. Post-hoc analysis using Tukey HSD test shows that the average slope of H subtask was less than that of TR subtask ($p < 0.001$) and RG subtask ($p < 0.001$), and the average slope of the TR subtask is less than that of RG subtask ($p = 0.05$). The results show that performing RG subtasks demands the highest mental workload in the three groups which causes rapid pupil dilation (mean slope = 1.629); the H subtask is easy to do which causes rapid pupil relaxing (mean slope = -2.987); the TR subtask is in the middle requirement of mental workload which tends to slightly enlarge the pupil size (mean slope = 0.348). The slope of pupil diameter changes over different subtasks is illustrated in Figure 3 and Figure 4.

Discussion:

Our hypothesis was partially supported. Recall that the largest pupil size was not recorded during the most challenging movement (RG), but in the TR movement. We think it was caused by the order of the 3 types of subtasks. As shown in Figure 4, the pupil diameter in RG starts to increase from a rest state with a small pupil size to an active movement with a larger pupil size. On average pupil diameter increased from the RG baseline (0.427 mm) and continued to enlarge slightly in performing the TR task (0.603 mm). The pupil diameter in H subtasks drops down rapidly towards the baseline, resulting in a size close to that in RG. Pupil size failed to show significant difference among the three types of movement, suggesting that average pupil diameter does not correlate directly with the task difficulty.

However, pupil size changes over time (calculated by the slope) supports our hypothesis well. Pupil size increased significantly during RG movements and increased slightly during TR movements; while in H movements, the pupil started to shrink rapidly. There are two reasons why the pupil enlarged less during the TR movement than in RG. First, the mental workload demanded by the TR is lower than the RG subtask since releasing an object is relatively easier than grasping an object (the cup size is larger than the size of peg). The second reason is that the TR subtask follows the RG subtask where the pupil diameter nearly reached its maximum and has to decrease. As shown in Figure 4, the pupil diameter in most of the TR subtasks first underwent a decrease and then increased again in a V-shape. Thus the average pupil diameter and the slope during the TR subtask did not reflect the real difficulty of the task. In contrast, it was relatively easier for the participants to perform the H subtask since there was no peg to carry or to pick up, and the target (the white central square) is relatively big and easy to touch.

Results show that we must use caution when using pupil size per se to interpret task workload. The size of the pupil may be affected by the previous task before the current task. However, the rate of change of pupil size matches well to the task requirement, and can serve as a better behavioral indicator for assessing mental workload of a surgeon.

Conclusions

Although TEPR has been extensively studied and well proven to be sensitive to mental workload changes, there are few works examining how TEPR correlates with mental workload intensity during surgical task execution. Our preliminary study shows that the rate of change of pupil diameter matches well to the change of precision requirement of a surgical task, better than the pupil diameter. This finding sheds light to non-intrusively monitoring surgeon’s mental workload during performing a surgery in a real operating room using a remote eye-tracking system.

Future work will conduct more case studies and explore individual differences among surgeons using TEPR.

Acknowledgments

We thank the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Royal College of Physicians and Surgeons in Canada (RCPSC) for funding this project.

References

[1] Goldwater, B.C., Psychological Significance of Pupillary Movements. *Psychological Bulletin*, 1972. 77(5): p. 340-355.

[2] Hess, E.H. and J.M. Polt, Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, 1964. 143(3611): p. 1190-1192.

[3] Beatty, J., Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, 1982. 91(2): p. 276.

[4] Recarte, M.Á., et al., Mental Workload and Visual Impairment: Differences between Pupil, Blink, and Subjective Rating. *The Spanish Journal of Psychology*, 2008. 11(2): p. 374-385.

[5] Ahlstrom, U. and F.J. Friedman-Berg, Using Eye Movement Activity as a Correlate of Cognitive Workload. *International Journal of Industrial Ergonomics*, 2006. 36(7): p. 623-636.

[6] Iqbal, S.T., et al., Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2005, ACM: Portland, Oregon, USA. p. 311-320.

[7] Richstone, L., et al., Eye Metrics as an Objective Assessment of Surgical Skill. *Annals Surgery*, 2010. 252(1): p. 177-82.

[8] Zheng, B., et al., Measuring Mental Workload during the Performance of Advanced Laparoscopic Tasks. *Surgical Endoscopy*, 2010. 24(1): p. 45-50.

[9] Carswell, C.M., D. Clarke, and W.B. Seales, Assessing mental workload during laparoscopic surgery. *Surgical Innovation*, 2005. 12(1): p. 80-90.

[10] Berguer, R., W.D. Smith, and Y.H. Chung, Performing Laparoscopic Surgery is Significantly more Stressful for the Surgeon than Open Surgery. *Surgical Endoscopy*, 2001. 15(10): p. 1204-1207.

[11] Zheng, B., et al., Workload Assement of Surgeons: Correlation between NASA TLX and Blinks. *Surgical Endoscopy*, 2012: p. In press.

Illustrations

	RG	TR	H	P value
Adjusted pupil size (mm)	0.427 ±0.202	0.603 ±0.288	0.473 ±0.228	0.227
Slope (100 µm/sec)	1.629 ±1.081	0.348 ±0.325	-2.987 ±1.790	0.000

Table 1. The output of the single ANOVA analysis on the adjusted pupil diameter and rate of change of diameter (slope) over three types of subtasks.

Figure 1. The experimental setting.

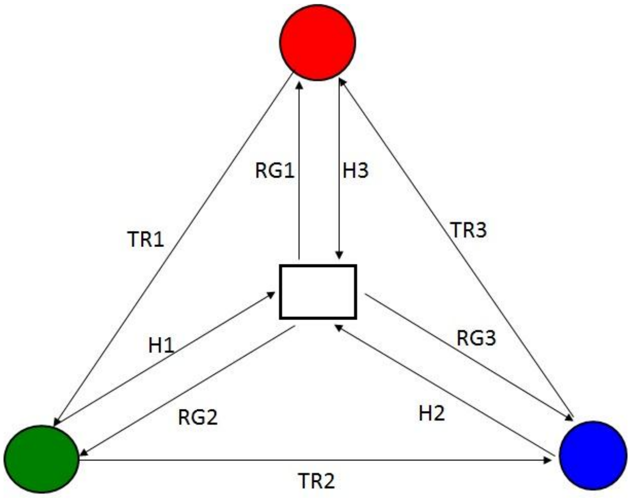


Figure 2. Illustration of the task. The subject is required to move a surgical grasper from the home plate (the rectangle in the centre) to grasp a peg (small green ball) sitting on a cup (reaching and grasping, RG); transport and release the peg to the next cup (TR), then bring the grasper back to the home plate (homing task, H). This is repeated three times, moving the peg to the green and then the blue cup, then back to the red starting cup.

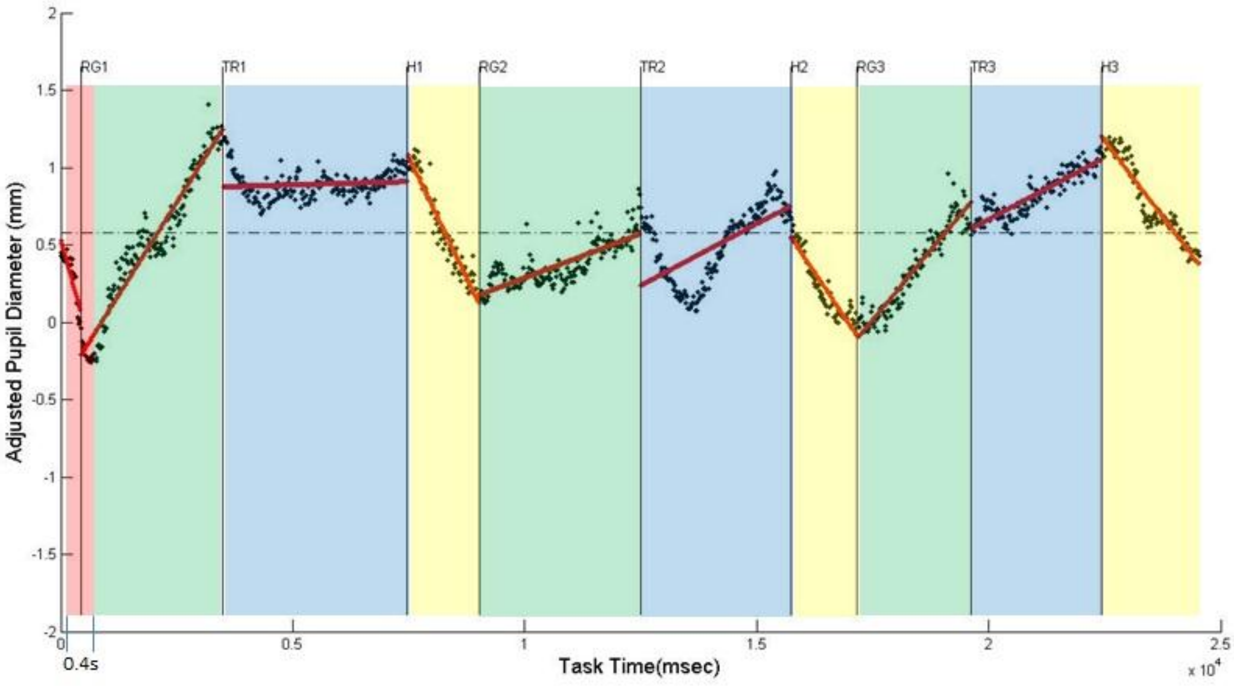
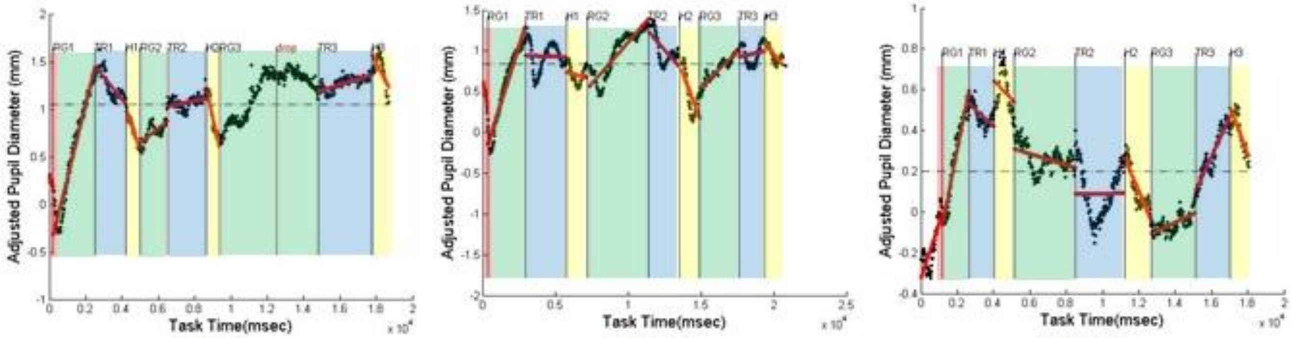
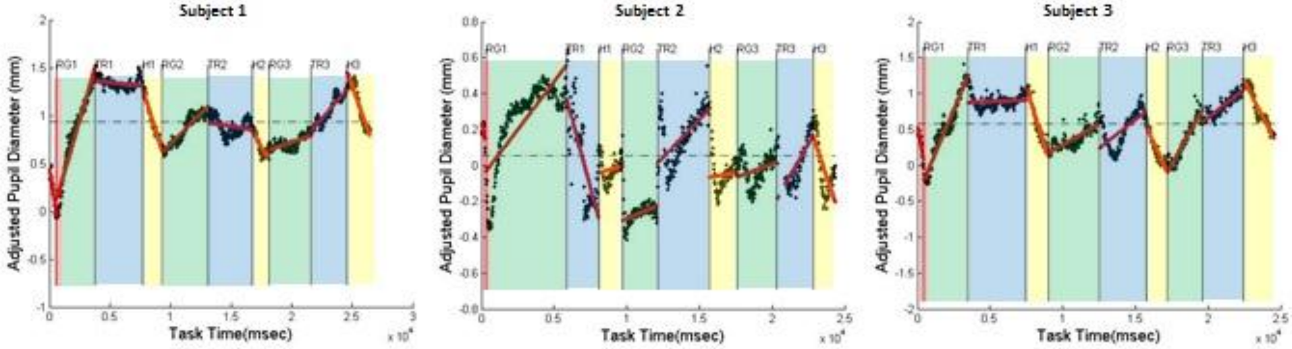


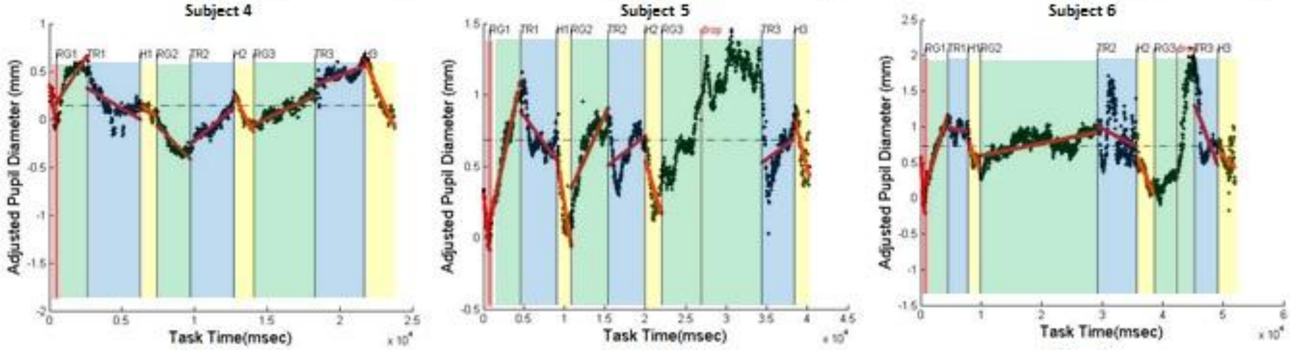
Figure 3 An example (from the 3rd trial of subject 6) of pupil diameter curve over time, aligned by subtasks. The area with the pink background is used for the baseline. The areas with green, blue, and yellow backgrounds are RG, TR, and H subtasks respectively.



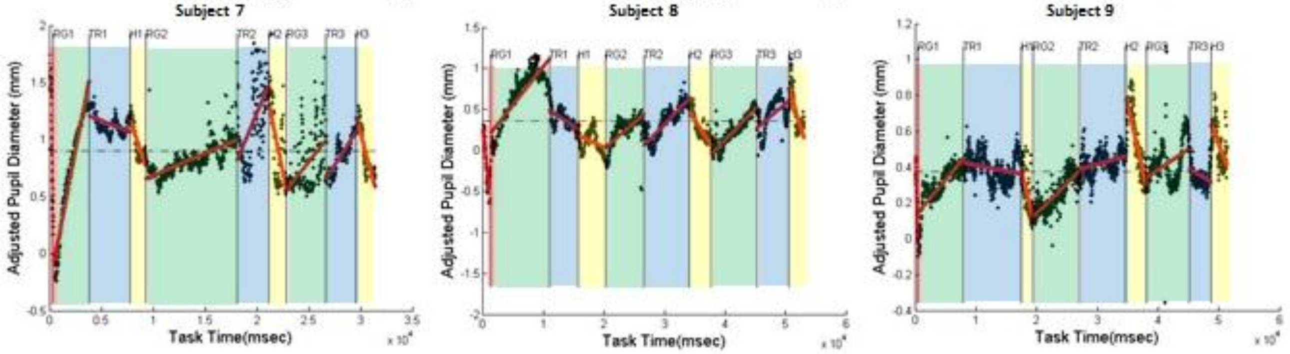
Subject 1 Subject 2 Subject 3



Subject 4 Subject 5 Subject 6



Subject 7 Subject 8 Subject 9



Subject 10 Subject 11 Subject 12

Figure 4. The raw data of pupil changes in trial 3 of the 12 subjects. The area with the pink background is used for the baseline. The areas with green, blue, and yellow backgrounds are RG, TR, and H subtasks respectively.