Detection of Changes in Surgical Difficulty: Evidence From Pupil Responses

Surgical Innovation 2015, Vol. 22(6) 629–635 © The Author(s) 2015 Reprints and permissions: sagepub.com/journalsPermissions.nav DOI: 10.1177/1553350615573582 sri.sagepub.com

Bin Zheng, MD, PhD¹, Xianta Jiang, PhD², and M. Stella Atkins, PhD²

Abstract

Background. Assessing the workload of surgeons requires technology to continuously monitor surgeons' behaviors without interfering with their performance. We investigated the feasibility of using eye-tracking to reveal surgeons' response to increasing task difficulty. *Methods.* A controlled study was conducted in a simulated operating room, where 14 subjects were required to perform a laparoscopic procedure that includes 9 subtasks. The subtasks could be divided into 3 types with different levels of task difficulty, calculated by the index of task difficulty (ID) proposed by Fitts in 1954. Pupillary responses of subjects in performing the procedure were recorded using Tobii eye-tracking equipment. Peak pupil dilation and movement time were compared between subtasks with different IDs as well as between fast moving and slow aiming phases within each subtask. *Results.* When the task difficulty was increased, task completion time increased. Meanwhile, the subjects' peak pupil size also increased. As the entire procedure was performed continuously, we found that pupil responses were not only affected by the ID in the current subtask but also influenced by subtasks before and after. *Discussion.* Decomposing a surgical procedure into meaningful subtasks and examining the surgeon's pupil response to each subtask enables us to identify the challenging steps within a continuous surgical procedure. Psychomotor evidence on surgeon's performance may lead to an innovation for designing a task-specific training curriculum.

Keywords

workload assessment, eye-tracking, psychomotor, Fitts' law, human factors in surgery, surgical simulation

Introduction

Quantifying the workload of surgeons has numerous implications for safe surgery.¹⁻³ It has been shown that overloaded surgeons may have difficulties in maintaining stable performance in the operating room.^{1,4} A surgeon may lose vigilance to signs of life-threatening conditions during the surgery⁵ and make wrong decisions leading to undesirable consequences.⁶ There is no question that multiple factors in the operating room can increase the workload of surgeons, ranging from patient variation, surgeon's competency, and resource availability. It is up to us to develop reliable methodologies to monitor when a surgeon's workload increases before we can explore what causes the workload to increase.^{3,7}

In the past a few years, we have used a paper instrument and the secondary task approach to investigate the workload of surgeons while they were performing laparoscopic surgeries.^{8,9} Each of these methodologies has its own limitations. Paper assessment is conducted at the end of a task performance. Results are acquired from the surgeon's selfreport and are heavily affected by the surgeon's perception of task difficulty.⁷ Recently, some researchers are making efforts to modify the paper instrument to better fit the surgical context.¹⁰ The secondary task approach can monitor the change of workload during the procedure.⁸ However, it can interfere with the primary surgical tasks in the operating room and pose risks to patient safety; as a result, the secondary task approach for assessing the workload is mainly used outside the operating room, such as in simulated surgical environments.^{8,9,11}

In this study, we decided to adopt a third approach to assess the workload of surgeons by examining the changes of their physiological signals. When workload increases, the surgeon's sympathetic system becomes excited and can be detected by examining the heart rate, sweating, and pupil responses.¹² By monitoring changes on these physiological signs, we have the capacity to detect the change

¹University of Alberta, Edmonton, Alberta, Canada ²Simon Fraser University, Burnaby, British Columbia, Canada

Corresponding Author:

Bin Zheng, Department of Surgery, Faculty of Medicine and Dentistry, University of Alberta, 3-002F Li Ka Shing Centre, 8440 112 St NW, Edmonton, Alberta, Canada T6G 2E1. Email: bin.zheng@ualberta.ca of workload of the surgeon without interruption.³ Among the multiple physiological signs that have been used to assess surgeons' workloads, we are particularly interested in examining surgeons' pupillary response.

Instead of attaching uncomfortable sensors to the body of surgeons for measuring skin conductivity and heart rate, surgeons' eye motions can be monitored by a remote eyetracker that can be attached to the surgical screen without interference of the surgeon's performance.¹³ Pupillary response is very sensitive to the change of workloads of a human operator, regardless of whether the loads arises from cognitive judgment or task difficulty.¹⁴⁻¹⁶ Currently, studies of the linkage between pupillary response and workloads focus on cognitive loads, where subjects are asked to perform mental calculation, recitations, and visual judgment tasks. Studies of the direct linkage between pupillary responses to the change of task difficulty are sporadic.

Since 2013 our research group has examined pupil response to increasing task difficulty in a simple (discrete) aiming task.¹⁷⁻¹⁹ In aiming, increasing the target distance and reducing the target size, 2 key variables for calculating the index of difficulty (ID), shows exponential correlation to the performance time. This phenomenon has been described by Paul Fitts in 1954 and its mathematical form has been called Fitts' law.²⁰ Our recent findings suggest that the change of ID not only affects task performance but also pupillary response.^{17,19} In simple words, as the aiming ID increases, subjects' peak pupil size increases correspondingly from its baseline.

Needless to say that surgical tasks are more complicated than a simple aiming task, and more importantly, tasks are sequential rather than discrete. In sequential tasks, pupil response to a current action may mingle with the residual effect from the previous action as well as with the anticipatory effect for the upcoming action. Before we can confidently use the pupil response to interpret the change in workload during a surgery, we need to develop a strategy to dissect the varying impacts on pupil responses from different actions during a procedural task. In this study, we asked subjects to perform a multiplestep laparoscopic procedure in a simulated environment. The task included using a laparoscopic grasper to pick up an object, transport it, and then place it down at a designated location. The steps (action, subtask) can be clearly defined and the ID can be computed. We hypothesized that subjects' pupil response, measured by the peak amplitude of pupil size increase, will be more significantly observed in a laparoscopic task with large ID.

Methods

Environment and Apparatus

A simulated training environment for laparoscopic surgery was set up inside the Medical Image Analysis Lab at Simon Fraser University. The training system included a laparoscopic training box (Laparoscopic Trainer, 3-D Technical Services, Franklin, OH) and a pair of laparoscopic graspers (Ethicon Endo-Surgery, Cincinnati, OH). On the bottom of this training box, a custom-made wood plate was placed as the surgical site. Images were captured by a webcam and displayed on to the monitor of a Tobii 1750 eye-tracker (Tobii 1750, Tobii Technology, Danderyd, Sweden). When a subject was performing the task while watching the monitor, hidden infrared sensors below the Tobii monitor captured the pupils of the subjects. The experimental setting was controlled as much as possible for luminance and head position, which may have affected the apparent pupil diameter. Details of the experimental setting can be found in our previous report.²¹

Participants

Fourteen university students (9 males and 5 females; age = 20-36 years, mean = 28 years) with zero surgical experience participated in the study, as we intended to eliminate the influence of surgical expertise on the performance. All subjects were right-handed with normal or corrected to normal vision. Ethics approval was obtained from the *Research Ethics Board* of Simon Fraser University before the recruitment of human subjects. Written consent was obtained from each participant prior to entering the study.

Task and Procedure

The task was to transport a rubber object (4.5 mm \times 10.5 mm green cylinder) over 3 dishes (13 mm in diameter) in a predetermined order (Figure 1). The task can be divided into 9 steps of 3 types of subtask: reaching and grasping (RG, estimated ID = 4.0 bits/response), transporting and releasing (TR, estimated ID = 3.3 bits/response), and homing (H, estimated ID = 2.6 bits/response). Detailed description about the procedure and tasks can be found in our previous report.¹⁸ Basically, reaching for the small rubber object (target diameter of 4.5 mm) required a higher level of precision than bringing the object into a plastic dish (diameter of 13 mm), or transporting the unloaded grasper back to a home plate (diameter of 15 mm, with a shorter travelling distance). The ambient lighting and data recording conditions were maintained constant throughout the entire study.

Data Organization and Analyses

Pupil responses were uninterruptedly recorded by Tobii 1750 at 50 Hz, except for the moments of eye blinks and large head movements. The pupil data were divided based on 3 subtasks (RG, TR, and H). The tool tip was identified automatically in every frame from the task video recordings, using our newly developed algorithm.²²



Figure 1. Illustration laparoscopic tasks performed in the study. The participant moves the surgical grasper from the home plate (the rectangle in the center) to reach and grasp (RG) the object (a green plastic cylinder) in the red dish (at top), to transport and release (TR) the cylinder to the next dish (low left), then bring the grasper back to the home plate (homing, H). These 3 subtasks were repeated 3 times, moving the object first to the green (low left), then to the blue dish (low right), and back to the reddish at top.

Within each subtask, movement can be further divided into 2 phases based on movement speed. In the first phase of movement (fast moving, FM), subjects accelerated to cover a large amount of travel distance. Once near the target, they decelerated to make adjustments to the movement (slow aiming, SA), either for picking up or releasing the objects. Practically, the FM phase started from the moment the grasper breaking contact with its current location (either a dish or a home plate) to the moment when the grasper reached the vicinity of the next target. In the recorded video images, tool-reaching was identified when the image of the tool tip reached over a dish or the home plate. The SA phase was defined from the moment when the tool reached the vicinity location above the current target (the end of FM) to the moment when the tool started to move for the next location. Actions performed in the SA phase differed among the 3 types of subtasks: grasping the cylinder from a dish in the RG subtask, releasing the cylinder into a dish in the TR subtask, and pointing down the unloaded grasper onto the home plate in the H subtask. Movement time and pupillary response were recorded between these 2 phases for each type of subtask.

The moment of tool reaching above the target (a dish or a home plate), that is, the end of the FM phase, was set as time zero. Pupillary response on each subtask was observed over a period of 6 seconds, with 1 second before and 5 seconds after the tool reaching above the target. The 6-second window was chosen for analyzing pupil response, because the average duration of the FM was less than 1 second (0.7 \pm 0.3 seconds) and the average duration of the SA phase was less than 5 seconds $(3.5 \pm$ 3.0 seconds). The baseline pupil size was computed on the pupil diameter over a 400-ms period before the start of each trial (RG1). Actual peak size over the performance was reduced by this baseline data to acquire an adjusted pupil size for each subtask. As the entire procedure was performed continuously, the observational phase may overlap between subtasks, that is, pupillary response to transporting may start before the completion of reaching and grasping. This gives us a chance to examine the combined influence of 2 consecutive actions on the pupil dilation.

Statistics

Movement time (divided into FM and SA) and peak pupil size will be subjected to a 3 task type (RG vs TR vs H) within-subject ANOVA.

Post hoc test (Tukey HSD) was further applied to compare between pairs of RG, TR, and H subtasks. To examine segments where significant differences occurred in the 6-second window between the 3 subtasks, the graphical significance testing approach^{22,23} 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 waveforms, we considered a series of more than 4 consecutive samples (80 ms) with *P* values <.05 as significantly different.²² The results are reported in this article as mean \pm standard deviation unless stated otherwise.

Results

A total of 70 trials were recorded (14 subjects, each performed 5 trials). Nineteen trials were excluded from analysis due to low ratio of total fixation time over total execution time (TF/TT). The TF/TT ratio has been known as a goal-keeping marker for eye-tracking data. When the TF/TT ratio is lower than a certain value (ie, 70% in our studies), the eye-tracking data are not valid for interpretation as a normal human eye-movement.¹³ From the 51 valid trials, there would be a total 459 subtasks (each trial has 9 subtasks). We also excluded data from 16 subtasks where the cylinder was dropped outside the field of vision. We also excluded the first subtask (RG1) and last



Figure 2. Mean movement time for 2 different moving phases and a complete subtask of all subjects. Data are displaced in means with 95% confidence intervals (error bars).

subtask (H3) from each trial because we decided to examine pupil response under the influence from both preceeding and following actions. To keep the sample size equal over 3 different types of subtask, we also excluded the TR3. As a result, a total of 290 subtasks remained presenting data from TR1, H1, RG2, TR2, H2, and RG3 for further analysis.

Task Performance

The mean movement time for all subtasks is 4.3 ± 3.1 seconds, and differs between 3 types of subtasks (Figure 2; $F_{(2,287)} = 25.730$, P < .0001). Post hoc test (Tukey HSD) shows that the mean movement time of the H subtask (2.7 \pm 1.5 seconds) is significantly shorter (P < .0001) than those of the TR and RG subtasks, being 4.9 ± 1.7 seconds and 5.4 ± 4.3 seconds respectively; there is no significant difference between TR and RG.

When examining movement times in the FM and SA phases, we also found significant difference among the 3 different subtasks (FM: $F_{(2.287)} = 25.174$, P < .0001; SA:

 $F_{(2,287)} = 22.934$, P < .0001). Post hoc test (Tukey HSD) revealed that the fast moving time in the TR subtask (0.9 \pm 0.4 seconds) is significantly longer (P < .0001) than that in the RG (0.7 ± 0.2 seconds) and H subtasks ($0.6 \pm$ 0.2 seconds), and there is no significant difference between RG and H. The slow aiming period in the H subtask (2.1 ± 1.5 seconds) is significantly shorter (P <.0001) than those in the RG (4.7 ± 4.3 seconds), and TR (4.0 ± 1.7 seconds), and there is no significant difference between RG and TR subtasks.

The mean execution time of FM phases for all subtasks is 0.7 ± 0.3 seconds, and differs significantly between the 3 types of subtasks ($F_{(2,287)} = 25.174$, P < .001). Post hoc test (Tukey HSD) shows that the mean FM time of the TR subtask (0.9 ± 0.4 seconds) is significantly longer (P < .001) compared those of RG and H subtasks (0.7 ± 0.2 seconds and 0.6 ± 0.2 seconds, respectively), and there is no significant difference between RG and H. This result reflects that the distance to move in the TR task is much larger than the other 2 tasks.



Figure 3. Mean pupil diameter changes (subtracted by baseline) over 3 different subtasks (RG: bold solid line in green; TR: solid line in red, and H: thin solid line in black). Data are aligned at time of tool-reach (bold black vertical dash line), which separates the entire movement into the the fast moving and slow aiming phases. The pair comparison results from the graphic significant test are displayed on the lower bar (solid bar: significant difference; hollow bar: insignificant difference).

The mean execution time of SA phases for all subtasks is 3.5 ± 3.0 seconds and differs significantly between the 3 types of subtasks ($F_{(2,287)} = 22.934$, P < .001). Post hoc test (Tukey HSD) shows that the mean time of the SA phase of the H subtask (2.1 ± 1.5 seconds) is significantly shorter (P < .001) those of TR and RG subtasks (4.0 ± 1.7 seconds and 4.7 ± 4.3 seconds, respectively), and there is no significant difference between the SA phases of the RG and TR subtasks.

Pupil Response

Figure 3 shows pupil response in a 6-second window for each of the 3 subtasks; they all displayed differences. In the RG subtask, the pupil dilation started from the fast moving phase, lasting to the slow aiming phase, and developed the highest peak pupil size at the end of the aiming phase (3.9 seconds after tool reaching). In performing the TR subtask, we recorded a small pupil constriction during the fast moving phase, followed by moderate pupil dilation during the slow aiming phase. Peak pupil size occurred 2.7 seconds after reaching for the target. In the H subtask, the pupil also constricted in the fast moving phase and the pupil size continued to reduce until the end of the movement. The fact that the pupil dilated over the entire RG subtask whereas the pupil contracted over the H subtask generally indicates the pupil responses are mapping well to the change of task difficulty.

Graphical significance testing^{23,24} was applied between the curves of the 3 subtasks, with the results as shown in the bottom colored bars in Figure 3. Specifically, t tests were performed between paired subtasks on each time frame. If the mean pupil sizes were significant between 2 conditions at a particular time, this time moment was highlighted. Results showed that pupil responses were significantly different between 3 subtasks during the fast moving phase, but not between RG and TR during the slow aiming phase. In slow aiming, the shrinking pupil during homing differed to the dilating pupil in RG and TR (Figure 3, bottom bars).

Discussion

Our hypotheses were supported by the data. The overall movement time was prolonged as task difficulty

increased. This means that in the sequential surgical tasks, Fitts' law still holds true once we decompose the entire surgical procedure into meaningful surgical steps. Cao et al²⁵ conducted the pioneer study where a complicated laparoscopic procedure was decomposed into several basic human movements. They argued that task decomposition will help us understand surgical behaviors on a deeper level.²⁵ Results from this study add new evidence to this argument. Once we employ an appropriate method to examine surgical procedures, a common law for understanding human behaviors can be applied to reveal interesting insights.

Besides movement time, we are more interested in examining surgeons' pupil response to the change of task difficulty. In our previous study,^{17,19} we found that human's pupil dilation increased as the ID of the aiming task increased. Results presented in this study show that this phenomenon also exists in sequential surgical tasks. Basically, peak pupil size increased significantly in the subtasks with a high ID, such as reaching and grasping.

When examining the pupil response in a detailed manner, we found pupil response was not just regulated by the current action, but was also influenced by the previous action, and sometimes, by a subsequent action. For example, the subject's pupil constricted slightly when transporting the rubber cylinder to a dish (TR, fast moving phase in Figure 3). This action was following the reaching and grasping subtask, where the pupil reached its peak size in grasping the little cylinder from a dish. Subjects might feel relief as the transporting is relatively easier. The pupil size increased again when subjects tried to load the object into a dish; this requires precise manipulation. Once they successfully loaded the cylinder into a dish, the pupil size decreased over the entire homing phase, as bringing the unloaded grasper to the home plate was much easier than transporting and releasing the object.

Establishing a meaningful connection between pupillary response and the change of task difficulty provides an opportunity for us to monitor the change of task workload of surgeons in performing surgical procedures. In this study, we used a remote eye-tracker to capture subjects' eye pupil data continuously without interfering with their performance. This offers a potential to objectively label the difficult tasks within a surgical procedure once the procedure is video-recorded with surgeon's eyetracking. This may further lead to the innovation of designing task-specific simulation training for improving training outcomes in a shorter time.

Despite the encouraging results presented in this study obtained from a controlled simulation environment, we need to point out that task difficulty of a true surgical procedure is driven by numerous factors beyond simply the target size and movement distance. One limitation is that

the non-surgeon subjects included in this study will limit our findings to surgical scenarios. Further studies are needed using eye-tracking in the real surgical setting with expert surgeons before valid connections between pupil response and surgical task loads can be confirmed. Another limitation is that the surgical tasks included in this controlled study did not require bimanual coordination of the operators, which undoubtedly is a limitation for applying our results to a surgical context. Surgical procedures require bimanual coordination and are often more complicated than simulated tasks. In the future, since we have established the linkage between pupil dilation and the change of task difficulty, we plan to carry out more studies to investigate surgeons' gaze behaviors in more complex surgical procedures and provide further psychomotor evidences to improve surgical training.

Author Contributions

BZ, XJ, MSA contributed to concept and design. Data collection, analysis and interpretation was carried on by XJ, BZ, MSA. Manuscript preparation was done by XJ, BZ and critical comments on manuscript was given by MSA.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Arora S, Sevdalis N, Nestel D, Woloshynowych M, Darzi A, Kneebone R. The impact of stress on surgical performance: a systematic review of the literature. *Surgery*. 2010;147:318-330.
- Hull L, Arora S, Kassab E, Kneebone R, Sevdalis N. Assessment of stress and teamwork in the operating room: an exploratory study. *Am J Surg.* 2011;201:24-30.
- Satava RM. Mental workload: a new parameter for objective assessment? Surg Innov. 2005;12:79.
- Arora S, Sevdalis N, Aggarwal R, Sirimanna P, Darzi A, Kneebone R. Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surg Endosc.* 2010;24: 2588-2593.
- 5. Zheng B, Tien G, Atkins SM, et al. Surgeon's vigilance in the operating room. *Am J Surg*. 2011;201:673-677.
- Hull L, Arora S, Aggarwal R, Darzi A, Vincent C, Sevdalis N. The impact of nontechnical skills on technical performance in surgery: a systematic review. *J Am Coll Surg.* 2012;214:214-230.
- Carswell CM, Clarke D, Seales WB. Assessing mental workload during laparoscopic surgery. *Surg Innov.* 2005;12:80-90.

- Zheng B, Cassera MA, Martinec DV, Spaun GO, Swanström LL. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surg Endosc.* 2010;24:45-50.
- Zheng B, Rieder E, Cassera MA, et al. Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures. *Surg Endosc.* 2012;26:1352-1358.
- Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RS. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). World J Surg. 2011;35:1961-1969.
- Meneghetti AT, Pachev G, Zheng B, Panton ON, Qayumi K. Objective assessment of laparoscopic skills: dual-task approach. *Surg Innov*. 2012;19:452-459.
- Rohleder N, Wolf JM, Maldonado EF, Kirschbaum C. The psychosocial stress-induced increase in salivary alpha-amylase is independent of saliva flow rate. *Psychophysiology*. 2006;43:645-652.
- Atkins MS, Tien G, Khan RS, Meneghetti A, Zheng B. What do surgeons see: capturing and synchronizing eye gaze for surgery applications. *Surg Innov*. 2013;20:243-250.
- 14. Jainta S, Vernet M, Yang Q, Kapoula Z. The pupil reflects motor preparation for saccades: even before the eye starts to move. *Front Hum Neurosci*. 2011;5:10.
- 15. Porter G, Troscianko T, Gilchrist LD. Effort during visual search and counting: insights from pupillometry. *Q J Exp Psychol (Hove)*. 2007;60:211-229.
- Reinhard G, Lachnit H. The effect of stimulus probability on pupillary response as an indicator of cognitive processing in human learning and categorization. *Biol Psychol.* 2002;60:199-215.

- Jiang X, Atkins MS, Tien G, Bednarik R, Zheng B. Pupil responses during discrete goal-directed movement. Paper presented at: CHI 2014; April 4, 2014; Toronto, Ontario, Canada.
- Jiang X, Zheng B, Tien G, Atkins MS. Pupil response to precision in surgical task execution. *Stud Health Technol Inform.* 2013;184:210-214.
- Jiang X, Zheng B, Tien G, Zheng B, Bednarik R. Pupil dilations during target-pointing respect Fitts' law. In: *Proceedings of the Symposium on Eye Tracking Research and Applications* (ETRA '14). New York, NY: ACM; 2014:175-182. doi:10.1145/2578153.2578178.
- Accot J, Zhai S. Beyond Fitts' law: models for trajectory-based HCI tasks. In: *Proceedings of the ACM SIGCHI Conference on Human factors in Computing Systems* (CHI '97). New York, NY: ACM; 1997:295-302. doi:10.1145/258549.258760.
- Atkins MS, Jiang X, Tien G, Zheng B. Saccadic delays on targets while watching videos. Paper presented at: ETRA 2012; March 28, 2012; Santa Barbara, CA.
- Jiang X, Zheng B, Atkins MS. Video processing to locate the tooltip position in surgical eye-hand coordination tasks [published online July 21, 2014]. Surg Innov. doi:10.1177/1553350614541859.
- Privitera CM, Renninger LW, Carney T, Klein S, Aguilar M. Pupil dilation during visual target detection. J Vis. 2010;10(10):3.
- Guthrie D, Buchwald JS. Significance testing of difference potentials. *Psychophysiology*. 1991;28:240-244.
- Cao CG, MacKenzie CL, Ibbotson JA, Turner LJ, Blair NP, Nagy AG. Hierarchical decomposition of laparoscopic procedures. *Stud Health Technol Inform.* 1999;62:83-89.