Pupil responses to task requirement in goal-directed movements

by

Xianta Jiang

M.Sc., Zhejiang University, 1998

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Computing Science
Faculty of Science

SIMON FRASER UNIVERSITY

Fall 2014

Approval

Name: **Xianta Jiang**

Degree: **Doctor of Philosophy**

Title of Thesis: ***Pupil responses to task requirement in goal-directed movements***

Examining Committee: **Chair:** Zhigeng Pan
Professor

M. Stella Atkins
Senior Supervisor
Professor

Gencai Chen
Co-Senior Supervisor
Professor
Zhejiang University

Ze-Nian Li
Supervisor
Professor

Bin Zheng
Supervisor Assistant Professor
Endowed Research Chair
University of Alberta

Lyn Bartram
Internal Examiner
Associate Professor
School of Interactive Arts and
Technology, SFU

Roel Vertegaal
External Examiner
Professor
School of Computing
Queen's University

Date Defended/Approved: November 6, 2014

Partial Copyright Licence

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010

Abstract

Objectively measuring the operators' task workload in goal-directed motor tasks such as surgical operations, is important for performance and safety. This thesis presents an approach for objectively measuring task workload in goal-directed movements using an important eye response: the pupil diameter. We demonstrate how to capture movement-related pupil size changes during motor tasks, investigate how the pupil responds to task requirement, and show that the pupil diameter can be employed as an objective and quantitative indicator of task workload in motor tasks.

In particular, we studied tasks where a tool is used to manipulate a target. The challenges include how to quantitatively define the task requirement, accurately detect the movement-related pupil responses from other event-evoked pupil changes, and separate the pupil responses from consecutive movements. We quantify the task requirements using Fitts' index of difficulty. The movement-related pupil events are captured from a time window aligned at a specific moment of tool movement which is automatically detected from recorded task videos. To separate overlapping movement responses, each tool movement is divided into two phases: the Transport and Landing phases. Three experiments were conducted to verify correlations between pupil responses and task difficulty of goal-directed movements, including two discrete target-pointing tasks and a continuous target-pointing task. We also investigated the pupil responses to task difficulty in a realistic situation: that is, during performance of a simulated surgical task, where we found that the pupil responses to the subtasks were related to the task difficulty.

Overall, we found the pupil diameter can be employed to objectively measure task workload in goal-directed movements, by conducting three experiments of discrete and continuous target-point tasks and a study of real-life motor task. The findings constitute the foundation for developing methods to objectively and quantitatively evaluate task workload of motor tasks using pupil diameter, and have a variety of implications in enhancing psychophysiological interactions in human-centered HCI and evaluating mental workload in high skill-demanding domains such as driving, aviation, and surgery.

Keywords: Pupil diameter; movement-related pupil response; goal-directed movement; eye-tracking; HCI; human factors

Acknowledgements

Many people helped me throughout the journey of my Ph.D. study that I would like to cordially acknowledge, but I could only list those most significant ones here.

First and most I gratefully thank my senior supervisor Dr. M. Stella Atkins for her excellent guidance and endless support. Dr. Atkins seems like having a magic stick in her hand that guides students to success. I deeply feel this magic power comes from her excellent supervisory art, her great enthusiasm in academia, and her extremely hard working. I deeply appreciate every effort that Dr. Atkins made to build my skills and confidence in my research, her pervasive support in various ways, and the happy time we had.

I so much thank my co-senior supervisor Prof. Gencai Chen for his long time support. Prof. Chen was also my M.Sc. supervisor back to 1995 and we have worked together for over fifteen years at ZJU. I appreciate not only his guidance to my research, but also his help to my career. He foresaw my potentials in doing research and urged me to pursuit a Ph.D. degree. I really appreciate all the advice from Prof. Chen.

Dr. Bin Zheng contributed so much to my thesis research for the medical background knowledge, research methods, and many great ideas. Dr. Zheng always can draw precisely a research question to a right direction confidently. This is due to his strong research background and his sharp sense of needs for medical computer technology. Thanks Dr. Zheng for your great support. I look forward to the future cooperation.

Thank Dr. Ze-Nian Li for his very thoughtful and critical suggestions to my thesis research. Thanks also to my examiners Dr. Lyn Bartram and Dr. Roel Vertegaal for their valuable time and suggestions. Thank Dr. Roman Bednarik for the cooperation on my papers and suggestions. I would also like to thank Dr. Ling Chen for his always thoughtful ideas and advice for my research. I really appreciate his great support, especially during my first year's Ph.D. research.

Thank Geoffrey Tien for working together with me and the help. Thanks to my colleagues of Medical Image Analysis Lab and School of Computing Science for the good time we had in Simon Fraser University. Thank all people who helped me during Ph.D. study. Thanks all reviewers for their critical reviews for this thesis and all related submissions. Thank all the participants in the user studies for their valuable time.

This thesis is dedicated to my family, my daughter and my wife. Without their love and support, I would not able to finish my Ph.D. research.

Table of Contents

List of Tables

List of Figures

List of Acronyms

2D	2-Dimension
3D	3-Dimension
4-D	4-Dimension
ANOVA	Analysis of Variance
AVI	Audio Video Interleave
CD	Constant Distance
CS	Constant Size
EEG	Electroencephalography
EOG	Electro-OculoGraphy
HCI	Human-Computer Interaction
ICA	Index of Cognitive Activity
ID	Index of Difficulty
IP	Index of Performance
IR	Infrared Red
MT	Movement Time
RGB	Red Green Blue
RSME	Average Root Mean Square Error
TEPR	Task-Evoked Pupillary Response
TF/TT	Total Fixation time over Total execution Time
VOG	Video-OculoGraphy

Chapter 1.

Introduction

1.1. Background

Objectively measuring task workload of operators has a wide range of applications including driving and aviation (Benedetto et al., 2011; Haak, Bos, Panic, & Rothkrantz, 2009; Hancock, Wulf, Thom, & Fassnacht, 1990; Noguchi, Shimada, Ohsuga, Kamakura, & Inoue, 2009; Ohsuga et al., 2011; Smith, Shah, & Lobo, 2000; Tsai, Viirre, Strychacz, Chase, & Jung, 2007), human-computer interaction (Bailey & Iqbal, 2008; Bednarik, Vrzakova, & Hradis, 2012; Iqbal, Adamczyk, Zheng, & Bailey, 2005), and minimally invasive surgery (Berguer, Smith, & Chung, 2001; Richstone et al., 2010; B. Zheng, Cassera, Martinec, Spaun, & Swanstrom, 2010; Bin Zheng et al., 2012). For example, surgeons undergo higher mental workload in minimally invasive laparoscopic surgeries than in open procedures (Berguer, et al., 2001). Task requirements in a laparoscopic surgery are higher than in an open surgery due to the use of long-shafted tools, indirect mapping of visual field from the laparoscope, and the overall lack of natural vision on the surgical sites. It is important to measure and monitor surgeons' task workload in operating rooms regarding patients' safety, since overloaded surgeons may lose their ability to maintain stable performance in the operating room (Mishra, Catchpole, Dale, & McCulloch, 2008; Stefanidis, Scerbo, Korndorffer Jr, & Scott, 2007; B. Zheng, et al., 2010). They may lose vigilance to signs of life-threatening conditions (B. Zheng, et al., 2010) or make wrong decisions leading to undesirable consequence (Spaun, Zheng, & Swanström, 2009).

Task workload is generally induced by task difficulty, but generated from various sources such as perceptual load, cognitive load, and physical load (Wickens, 2002, 2008). These sources correspond to different stages of perception, cognition, and manual responding. Perceptual load is the requirement to perceive more items during a

visual searching task, cognitive load comes from the task demands in working memory in cognitive tasks such as mental arithmetic tasks, and physical load arises from physical demands typically in motor tasks (Backs, Ryan, & Wilson, 1994; S. Chen & Epps, 2014). A task may involve multiple sources of loads. For example, a mental arithmetic task may involve perceptual and cognitive loads: the subject has to take in the information of the question from either visual or acoustic channel (perceptual load), and then calculate from the items in the working memory (cognitive load). Chen et al. (2014) designed an experiment to separate perceptual load and cognitive load as two distinct sources of task difficulty, by manipulating five levels of difficulty of an arithmetic task performed in low and high perceptual load situations respectively.

In a visual-motor task such as target-pointing task, both perceptual load (visual) and physical load (manual responding) contribute to the task difficulty, but the latter usually dominates (Backs, et al., 1994). Backs et al.(1994) separated perceptual and physical loads in a manual tracking task. In the case of a target-pointing task, visual perception is involved at the beginning of the task to perceive the global visual field of the task setting, and before the hand/tool approaches the target to intake the specific target position information (Abrams, Meyer, & Kornblum, 1990; Elliott, Helsen, & Chua, 2001). Therefore, the task difficulty of an aiming task mainly comes from physical demands. In this thesis, the difficulty of the designed target-pointing task refers to that from the physical demands shaped by the target size and target distance.

There are three main categories of approaches to measure task workload, viz., using subjective rating scales, using performance measurements (usually by indirectly measuring the performance of a secondary task), and physiological response measurements. All of these measurements have significant limitations. The subjective rating methods have bias in the answers which are greatly affected by the working memory of the participant and can only usually be done at the end of the task; the secondary task measurement causes extra workload which may affect the primary task; and physiological methods usually require attaching sensors to the human body, which is intrusive.

With advances in eye-tracking technology, pupil diameter can be recorded remotely (unobtrusively) and continuously, and its subtle changes used to indicate the

cognitive task load (Klingner, Kumar, & Hanrahan, 2008). What is lacking is the evidence between the change of pupil dilation and the change of task requirements in goal-directed movements. Little work has been done 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 the true task requirements during goal-directed movements; the participants in this 1985 study simply flexed their fingers while looking at a blank screen for the purpose of pupil size recording.

In goal-directed interactive movements, such as aiming at a target and selecting a menu item, the tool or hand may move continuously with the coordination of eye movements, which may evoke a different pupil response pattern. The discovery of this knowledge 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.

1.2. Challenges

There are technical challenges in measuring task difficulty of motor tasks using pupil diameter. First, the movement-related pupil responses are mixed with other event-related pupil responses such as ambient lighting. Usually, signal averaging techniques can be employed to detect the movement-related pupil responses. By averaging many short epochs of the same task-evoked pupil responses, aligning at a specific common time point such as 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 (Beatty, 1982; Klingner, et al., 2008). The selection of a common time point for aligning the short pupil response epochs is critical to the signal to noise ratio (SNR) of the movement-related pupil response. However, goal-directed movement tasks are usually self-paced where the common time point for alignment is implicit. In these cases, the hand/tool positions and the critical moments of movement such as when the hand/tool starts to move carry important information for the

analysis of movement-related pupil responses. We developed a method for automatically extracting tooltip positions from the task videos (Appendix A). Based on the extracted tooltip positions, the tool movements and the common aligning time point were determined according to the kinematics of the tool movement.

Another challenge is how to quantitatively define the task difficulty. There is no common way of defining the task difficulty in movement-related pupil studies. For example, in the study performed by Richer and Beatty in 1985 (Richer & Beatty, 1985), the task difficulty was empirically defined by the increased number of fingers in the key-pressing task.

Fitts' law is a traditional model of human movement by analogy to the transmission of information (Paul M. 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 its movement time (MT), as shown in equation (1). To maintain a certain level of accuracy, the more difficult the task, the more movement time is needed. The index of difficulty (ID) is determined by the target distance (A) and the target width (W), as shown in equation (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 recently in the laparoscopic environment (H.-J. Chen & Lin, 2011; Prytz, Montano, & Scerbo, 2012).

$$\mathrm{IP} = \frac{ID}{MT} \qquad (1)$$

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

We quantitatively define task difficulty in target-pointing tasks (from Experiment 1 to Experiment 3) and the simulated surgical task using Fitts' index of difficulty (ID).

Yet another challenge is that the pupil response may be affected by neighbouring events in continuous movements due to the low frequency of pupil response (typically lower than 0.5Hz) (Moresi et al., 2011; Privitera, Renninger, Carney, Klein, & Aguilar, 2010; Richer & Beatty, 1985). To address this problem, we observed pupil responses in both discrete and continuous target-pointing tasks. We conducted two experiments

(Experiment 1 and Experiment 2) to explore the pupil response to task requirement in discrete target-pointing tasks, in which each movement was followed by a 10s waiting time to separate the pupil dilation from its previous movement and allow the pupil diameter to return to its baseline. On the basis of the basic pattern of pupil response to a discrete movement, we further explored pupil responses to task requirement in continuous movements i.e. continuous aiming movements (Experiment 3). For further analysis, we divided a movement into two phases e.g. the Transport and Landing phases according to the kinematics of the tool movement; this division provides a more detailed observation of the pupil responses in continuous movements. Specifically, the pupil response during the Landing phase of a movement positively correlates to the task difficulty and the pupil response during the Transport phase reflects the effect of the previous movement.

1.3. Overview of experiments

Three experiments were conducted to investigate the pupil changes during goal-directed tool movements.

In **Experiment 1**, we explored the pupil responses to the task requirement of a discrete target-pointing task, where the subjects were required to move a tool to point and touch pairs of target circles with different target sizes and distances following the definition of Fitts' ID and each movement was preceded by a 10s wait. We found a small but significant dilation starting about 1.5s before the tool starts to move, followed by a slight constriction, the "valley" in the pupil size profile. Before the tool touches the target, the pupil reaches its peak size. Both the pupil dilation and the duration from Valley-to-Peak size positively correlate with the increase of IDs. This evidence indicates that the change of pupil diameter is regulated by task requirement. However, the task requirement in that discrete task study was affected by the mixed effect of both target size changes and target distance changes.

To determine whether target size or target distance primarily contributes to the change of the pupil size, we conducted **Experiment 2** by asking subjects to perform two tasks: 1) aiming at different-sized targets located a constant distance apart, and 2)

aiming at targets of the same size at varying distances. The results demonstrated that the pupil dilates in response to both factors, either increase of target distance, or a reduction of the target size, in a similar pattern.

Both Experiment 1 and Experiment 2 confirmed the connection between pupillary response and task difficulty in discrete movement tasks. Examples of discrete visual-motor tasks in daily life include inserting a key into a lock, shooting a basketball, and mouse-clicking at a specific location in an editor, etc. However, in everyday interactive tasks, continuous movement tasks are much more common, such as steering the wheels of a vehicle, playing ping-pong, and selecting an item in a multiple-level cascade menu. In many cases, the movement frequency is higher than the pupil response frequency which is typically lower than 0.5Hz (Jiang, Atkins, Tien, Bednarik, & Zheng, 2014; Richer & Beatty, 1985), pupil response is inevitably affected by multiple movements. It is important to carefully examine pupil response and develop a method to distinguish if pupil response is in reaction to an upcoming movement or is just a residual effect from a previous movement. We needed to explore the pupil responses to the change of task requirements in a continuous aiming movement.

In **Experiment 3**, we employed a similar experiment setting to that in the discrete movement studies, but here the participants performed a continuous pointing task without any waiting time between movements. We hypothesized that pupil dilation will still respect Fitts' law in continuous movements, such that higher task difficulty will evoke higher peak pupil dilation. New challenges emerged during the analysis of pupil data in this continuous movement task as the pupil response was overlapped by the consecutive movements. By carefully choosing the baseline and the moment to align the movement windows and by dividing the movement into two phases—Transport and Landing, we addressed these challenges. The results supported our hypothesis.

In summary, we found how to employ pupil diameter as a quantitative measurement of task difficulty of visual-motor tasks, via a series of experiments based on various tasks including discrete Fitts' pointing tasks, continuous aiming task, and simulated surgical tasks. By carefully processing the pupil signal and tooltip position data, we found that pupil responses to task requirement respect Fitts' law in all these motor tasks, such that changes of pupil size positively correlate with the task difficulty.

These findings constitute the foundation of employing pupil diameter to measure task workload in motor tasks.

1.4. Contributions

This thesis has the following major contributions.

1) Verified correlations between pupil diameter and task difficulty of motor tasks.

- Innovatively set up the connection between pupil parameters e.g. peak pupil dilation to Fitts' index of difficulty. This enables us to quantitatively measure task workload using pupil diameter.
- Discovered pupil response patterns in both discrete and continuous target-pointing tasks, and confirmed the correlations between the changes of pupil diameter and the task difficulty defined by Fitts' index of difficulty. The findings constitute the foundation of employing pupil diameter as an objective, non-intrusive, and continuous measurement of task difficulty of visual-motor tasks.
- Applied pupil diameter to indicate task difficulties of real-life motor tasks. We found the pupil response patterns in a simulated surgical task, and the pupil diameter is able to indicate both the task difficulty and task execution order of complex motor tasks.

2) Developed key techniques to identify movement-related pupil responses. We employed signal averaging techniques to reduce data noise and preserve movement-related pupil responses, and determined the window length, the aligning time point, and the baseline which are critical to the validity of the results. Furthermore, we divided a movement or subtask into Transport and Landing phases to better understand the pupil responses to a movement.

- Employed windows to extract the movement-related pupil signal and aligned at a specific time point in the window. We found a shorter window length (4s) is proper for extracting pupil signals from the continuous target-pointing task, a longer window (7s) for discrete target-pointing task to preserve the pupil response to the preparation of the movement at around 1.5s before tool starts to move, and

a long window (8s) for the surgical task since it took much more longer time for grasping or releasing the object.

- Determined the valid time points for aligning the pupil signal windows. We found that by aligning at the timestamps around moments of the tool starting to move, and reaching the target, the movement-related pupil signal was effectively preserved. The moments of tool starts to move and reaching the target were automatically detected from the recorded task videos.
- Defined the baseline pupil diameter and applied baseline subtraction to each window of the movement and subtask. The baseline subtraction technique applied to each signal epoch greatly reduced the effects of pupil responses to factors such as unrelated events and changes of lighting and illumination of the screen. We found the short segment before the beginning of the movement (2s before tool starts to move) was proper for the discrete target-pointing task and the short segment at the moment when the tool reaches the target was suitable for continuous movement tasks (continuous target-pointing and peg transportation tasks).
- Decomposed a movement/subtask into Transport and Landing phases to better understand the pupil responses to a movement. We found the pupil responses in the Landing phase correlate to the task difficulty, and the pupil responses in the Transport phase reflect the task execution order in motor tasks.

3) Developed algorithms to identify tool movements from task videos of self-paced motor tasks.

- Developed video processing algorithms to automatically extract tooltip positions from recorded task videos for identifying tool movements and the analysis of pupil responses to tool movements.
- Developed algorithms to identify the moments when tool starts to move and reaches the target according to the kinemics of the tool movement. These import moments were employed to separate individual movements from the self-paced motor tasks and to divide the movement phases.

1.5. Organization of thesis

The thesis is organized as in Table 1.1.

Table 1.1 Road map of thesis organization

Chapter 1 Introduction
Chapter 2 Literature review
Chapter 3 Pupil responses in discrete movements: experiments 1 and 2
Chapter 4 Pupil responses in continuous movements: experiment 3
Chapter 5 Evaluation of pupil responses in a simulated surgical task
Chapter 6 Conclusion

Chapter 1 introduces the importance of developing an objective, non-invasive, and continuous method for assessing task workload and the challenges of employing the pupil diameter as indicator of task workload. Chapter 2 reviews the work in mental workload measurement, eye-tracking technologies on pupil measurement, and pupil responses to both mental workloads of cognitive and visual motor tasks. Chapter 3 examines the correlation between the changes of pupil diameter and the task difficulty of discrete goal-directed movements, and Chapter 4 in continuous target-pointing movements. Chapter 5 describes how we applied the technology to a real goal-directed movement task i.e. a peg transportation task under a laparoscopic environment. Chapter 6 summarizes the thesis.

Chapter 2.

Literature review

2.1. Mental workload, task difficulty, and measurement methods

Mental workload is a finite mental resource that one uses to perform a task under specific environmental and operational conditions (Cain, 2004; Cassenti & Kelley, 2006; Wickens, 2002); it reflects the mental demand of performing the tasks as well as the capacity of the operator to respond to those demands. For example, novice drivers may not be able to regularly check the side or back mirrors while driving, since the task of keeping driving forward costs most of the mental resources and leaves insufficient capacity for the safety checking; similarly, novices in a surgical task are not able to regularly check the vital life signals shown on the side screens (Geoffrey Tien, Atkins, Zheng, & Swindells, 2010; Bin Zheng et al., 2011).

The motivation for measuring mental workload is for the prediction of task performance by quantifying the mental cost, as increased task demands may lead to unacceptable performance. Human performance is a function of both individual processing capacity and task demands (May, Kennedy, Williams, Dunlap, & Brannan, 1990). Mental workload increases when the task becomes challenging, and when the task difficulty reaches or exceeds the processing capability of the operator, his performance may be jeopardized or the task may be failed. Therefore, the measurement of mental workload is very useful for measuring the possibility of task accomplishment in a complex task environment.

According to Wickens' 4-D multiple resource model (Wickens, 2002, 2008), mental workload is generally induced by task difficulty, generated from various sources such as perceptual load, cognitive load, and physical load. These sources correspond

to different stages of perception, cognition, and manual responding respectively. Perceptual load is the requirement such as to perceive more items during a visual searching task, cognitive load is related to the task demands on working memory in cognitive tasks such as mental arithmetic tasks, and physical load arises from physical demands typically in motor tasks (Backs, et al., 1994; S. Chen & Epps, 2014). A task may involve multiple sources of loads. For example, a mental arithmetic task may involve perceptual and cognitive loads: the subject has to take in the information of the question from either a visual or acoustic channel (perceptual load), and then calculate from the items in the working memory (cognitive load). Chen et al. (2014) designed an experiment to separate perceptual load and cognitive load as two distinct sources of task difficulty, by manipulating five levels of difficulty of an arithmetic task performed in low and high perceptual load situations respectively.

In a visual-motor task such as target-pointing task, both perceptual load (visual) and physical load (manual responding) contribute to the task difficulty, but the latter usually dominates (Backs, et al., 1994). Backs et al.(1994) separated perceptual and physical loads in a manual tracking task. In the case of a target-pointing task, visual perception is involved at the beginning of the task to perceive the global visual field of the task setting, before the hand/tool approaches the target to intake the specific target position information (Abrams, et al., 1990; Elliott, et al., 2001). After the visual search, the task difficulty of the aiming task mainly comes from physical demands. In the rest of this thesis, the difficulty of the designed target-pointing task refers to that from the physical demands shaped by the target size and target distance.

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 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. This empirical definition of task difficulty was not quantitative. The difficulty of two-finger flexion was not necessarily twice as hard as that of single-finger flexion and even two-finger flexion in one hand might not be easier than one-finger flexion in both hands. In goal-directed

movement tasks, the task difficulty is mostly from physical demands shaped by the target size and distance, and is governed by the law of speed-accuracy trade-off.

Fitts' law is a traditional model of human movement by analogy to the transmission of information (Paul M. Fitts, 1954), and serves as quantitative definition of difficulty in a variety of research areas, including kinematics, human factors, and human-computer interaction (HCI) (MacKenzie, 1992), and even recently in the surgical environment (Prytz, et al., 2012). Kourtis et al. (2012) explored the correlations between the electroencephalography (EEG) signals and the Fitts' index of difficulty of an aiming movement, and found that the index of difficulty of the planned movement correlated linearly with the amplitudes of the specific components of the EEG signals (N2 and P3b). In an eye-based interaction task where the participants were instructed to move their gaze from home position to target locations of various distance Fitts' law has been found be valid to saccadic parameters, that is, the peak saccadic velocity and the variability of the endpoint of saccades positively correlate to the gaze travel distance (Abrams, Meyer, & Kornblum, 1989; Al-Aidroos, Fischer, Adam, & Pratt, 2008).

Once a concrete and quantitative definition of the task difficulty is established, an analytic model such as linear regression between the difficulty and other physiological indicators such as EEG and pupil responses could be built (Backs, et al., 1994; Jiang, Atkins, Tien, Zheng, & Bednarik, 2014). The established mapping between pupil responses and task difficulty could be further applied to situations where the task difficulty cannot be measured by Fitts' law.

Measuring mental workload is not trivial. It needs to meet certain criteria for specific situations. In the case of measuring mental workload in high risky task environments such as surgery, an objective, non-intrusive, and continuous method is needed. 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, 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 as these measures cannot be taken continuously during tasks, and have a bias to the task difficulty (B. Zheng, et al., 2010).

The subjective shortcomings may be avoided by using a secondary task to the primary task where decreased secondary task performance implies increased mental workload caused by the primary task (Wickens, 2002). 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 bias and generally do not interfere with task performance, most of them present an intrusion, as they need sensors attached to the user.

2.2. Eye-tracking technology

2.2.1. Eye-tracking methods

There are several different way of tracking eye movements and measure pupil dilation (Duchowski, 2007), including Electro-OculoGraphy (EOG), scleral contact lens/search coil, Video-OculoGraphy (VOG), and video-based combined pupil and corneal reflection. In this thesis we mainly use remote video-based pupil and corneal reflection eye-tracking systems which usually contain illumination sources (often infrared red (IR)) and one or more eye-tracking cameras. To estimate gaze location, the point where the subject is looking needs to be inferred from the eye image recorded by the camera attached in the eye-tracker. That is, the *m*-dimensional image feature space must be mapped to the 3-D world coordinates by the parameter vector *c*: $\Phi_c: \mathbb{R}^m \rightarrow \mathbb{R}^3$. (Villanueva et al., 2009). The parameter vector *c* can be estimated by calibrating the gaze, which may differ in terms of system hardware and subject-specific variables. The process of gaze calibration requires the user to look at a number *N* of predefined points labeled as t_i on the screen, and calculate the corresponding eye features x_i. Then the mapping Φ_c should be derived from the tuple set $D\{(t_i, x_i)\}$.

According to Villanueva et al's (2009) simplified eye-tracking method (a single eye tracking camera, a single gaze target, and a 2D planar screen), there are three coordinate systems used in the geometrical modelling, the eyeball, the camera, and the

screen, assuming that the three components have their own coordinate systems (centered e_c, O_{cam}, and O_{mon}, respectively). Taking the screen as the real world coordinate system, the other two coordinate systems can be transformed to the real world coordinate system (Villanueva, et al., 2009). Consequently, the gaze point on the screen and the gaze direction can be derived from the relations between the visual axis, optical axis, and the screen, i.e., the intersection between visual axis and the screen (m_v) and the intersection between the optical axis and the screen (m_o).

In order to obtain the eye features, i.e., the pupil centre, the image captured by the eye tracking camera (centered O_{cam}) will undergo image processing (Villanueva, et al., 2009) to detect the pupil outline. Light sources are used to illuminate the eye area. The reflections of the light sources (IR) on the cornea are used to facilitate the image processing.

2.2.2. Pupil diameter output from eye-tracker

As described in section 2.2.1, the pupil outline is detected before the gaze location estimation in a remote video-based eye-tracking system, so that pupil size information is included in the output of most eye-trackers. Usually, the pupil size information includes the horizontal and/or vertical diameter in pixels or in mm, which could be utilized to derive other pupil parameters. For example, Benedetto et al. (2011) used the following formula to compute the pupil area.

$$APS = diaX / 2 * diaY / 2 * \pi$$

Where diaX and diaY are horizontal and vertical pupil diameters respectively and APS is the pupil area.

When extracting pupil size information from the output of a video-based eye-tracker, distortion caused by the view angle of the eye camera and eye movements may exist. Pomplun and Sunkara (2003) proposed a neural-network-based calibration method to eliminate this geometry-based distortion when using a video-based eye-tracker. The calibration interface was evaluated in an experiment and the results showed that it significantly improved the signal-to-noise ratio. Gagl et al. (2011) corrected the pupil distortion caused by the eye-camera angle by using an artificial eye

model. First the authors used an artificial eye model with a fixed pupil size to be captured by the eye-tracking system in different view directions. Then parameters derived from the captured data were filled into a physical function to correct the pupil size from a real eye.

Modern eye-tracking systems can monitor the eye movements remotely and provide pupil diameter information continuously and non-intrusively. However, the question remains whether the remote video-based eye-trackers are capable (e.g. have enough precision) of measuring task-evoked pupillary responses (TEPR). To answer this question, Klingner (2008) replicated two previous experiments i.e. mental multiplication (Ahern & Jackson, 1979) and short-term memory (Kahneman & Jackson, 1966), where he measured cognitive load using pupil diameter output from a remote video-based eye-tracker instead of using a television pupillometer (Watanabe, Ikeda, Suzuki, & Nakamura, 1990) or a customized pupillometer systems employed in these two previous studies. The results from both replicated experiments are consistent with those of the two previous experiments; this implies that remote eye-trackers can be used to measure cognitive workload.

2.3. Pupil responses to the changes of mental workload in cognitive tasks

The pioneer work exploring pupil size changes relating to mental workload 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 value 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 in different tasks, including mental arithmetic tasks (Bradshaw, 1967), recall or memory tasks (Beatty & Kahneman, 1966; Goldinger & Papesh, 2012; Libet, Gleason, Wright, & Pearl, 1983; Otero, Weekes, & Hutton, 2011; Peavler, 1974; van Rijn, Dalenberg, Borst, & Sprenger, 2012), and visual search tasks (Attar, Schneps, & Pomplun, 2013; Porter, Troscianko, & Gilchrist, 2007; Privitera, et al., 2010).

Pupil size also responds to a critical event such as mental arithmetic 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 as an approach to capture and evaluate the mental workload changes during a variety of tasks (Beatty, 1982; Fong, Sibley, Coyne, & Baldwin, 2011; Goldinger & Papesh, 2012; Karatekin, Couperus, & Marcus, 2004; Piquado, Isaacowitz, & Wingfield, 2010).

2.4. Pupil responses to task requirement in goal-directed movements

The pioneering work exploring pupil responses to movement was conducted by Richer and Beatty (Richer & Beatty, 1985), which examined the pupil responses to simple finger movements, the self-paced finger flexion. The participants were required to press their finger on the keyboard approximately every 5s-10s with their eyes fixating at a point at 2m distance while being recorded by a pupillometer. The task difficulty was evoked by the increasing number of fingers to press the specified keys in the keyboard. The authors found the typical pupil dilation pattern to a simple finger movement is that the pupil dilates about 1.5s before the finger movement and peaks after about 0.5s, and the peak pupil dilation positively correlates to the task difficulty i.e. the number of finger flexion. However, Richer and Beatty's finger flexion task is not a realistic goal-directed movement as the participants' eyes were looking at a fixed point during the finger movements, so the eyes and the hands were dissociated during the task. In goal-directed interactive movements, such as aiming at a target and selecting a menu item, the hand or tool may move continuously, which may evoke a different pupil response pattern. Our research group conducted a series of studies to explore pupil responses to goal-directed movements (Jiang, Atkins, Tien, Bednarik, et al., 2014; Jiang, Atkins, Tien, Zheng, et al., 2014; Jiang, Zheng, Bednarik, & Atkins, 2014), which will be described in detail in this thesis.

In the field of eye-hand coordination in health care training, 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, precise details of the pupil response to motor tasks were not reported.

Chapter 3.

Pupil responses in discrete movements

Two experiments were conducted to verify the correlations between the changes of pupil size and the task difficulty of discrete goal-directed movements—the discrete target-pointing tasks—with 10s wait before each movement. Experiment 1 investigated pupil responses to a mixed effect of the variations of both target size and distance and Experiment 2 explored the pupil responses to the effect of either target size or target distance variations. These two experiments employed the same experiment setting, apparatus, and were conducted at the same location in the same lighting and screen luminance conditions, but with different participants and different target settings. The same data analysis methods were applied to both sets of data.

3.1. Pupil responses to task requirement during discrete goal-directed movements (Experiment 1)

3.1.1. Experiment purpose and hypothesis

This experiment was designed to investigate the pupil responses to task difficulty of discrete goal-directed movements, where the participants performed a target-pointing task with 10s wait between each tool movement. To quantify the task difficulty, the target sizes and distances were defined using indices of difficulty calculated according to Fitts' law.

The hypothesis was that different indices of difficulty would result in distinguishable patterns of pupillary response.

3.1.2. Methods

Participants

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

Experimental setting and apparatus

The task was to move a long-shafted laparoscopic grasper horizontally to touch the circles printed on a paper inside a laparoscopic training box, 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 3.1*. Three pairs of target circles with combinations of variable sizes and distances between targets were designed with three indices of difficulty (ID). The detailed parameters of the target circles are illustrated in *Figure 3.2*. The formula for calculating ID is as in Equation (2) (Paul M. Fitts, 1954), where W is the diameter of the target circle and A is the distance between the pair of targets.

Figure 3.1 Experimental setting (Experiment 1 and 2).
Note. Tobii X50, training box, and web camera, showing the target circles on the Tobii display monitor.

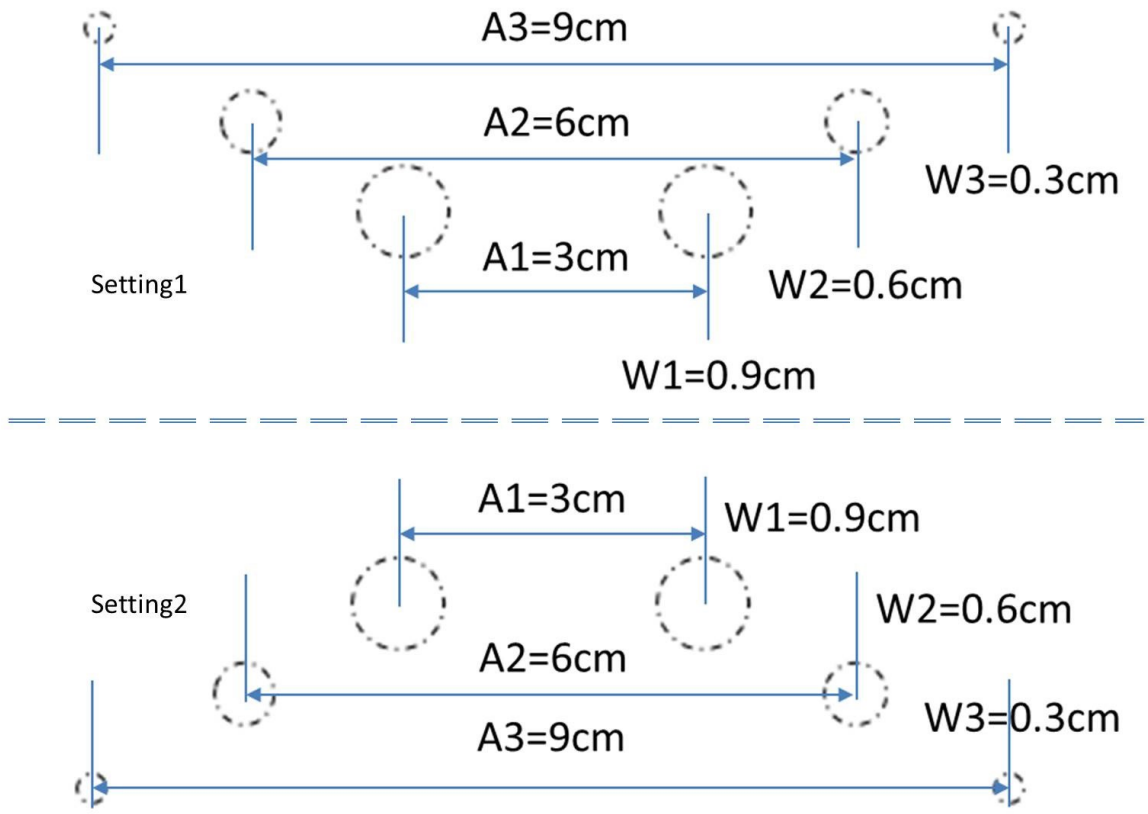


Figure 3.2 Illustration of target patterns (the pairs of target circles on the paper inside the training box showing on the display monitor as in Figure 3.1).

Note. 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. The following settings were employed: ID1=2.7 bit/response (W1=0.9cm, A1=3cm) denoted here as Easy, ID2=4.3 bit/response (W2=0.6cm, A2=6cm) denoted here as Medium, and ID3=5.9 bit/response (W3=0.3cm, A3=9cm) denoted as Hard.

Subjects held the surgical grasper at a standing pose about 60cm from the eye-tracker. The scene of the work area inside the training box was illuminated and captured at 30Hz by a built-in video camera and projected onto 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 to 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 (Privitera, et al., 2010). The tips of the grasper were black-taped to keep the tips' color consistent with the shaft, and make image processing easier.

Tasks and procedure

The task was to move the surgical tool to point to the circles printed on a piece of paper placed horizontally at the bottom of the training box. Since the frequency of pupil response to a movement is typically lower than 0.5Hz (Moresi, et al., 2011; Privitera, et al., 2010; Richer & Beatty, 1985) and the frequency of the tool movement of target-pointing under laparoscopic environment is usually around 0.5Hz (H.-J. Chen & Lin, 2011), in order to avoid overlap of the pupil dilation curve, we used discrete Fitts' pointing (P. M. Fitts & Peterson, 1964) by waiting 10s before each tool movement.

Specifically, each trial consisted of 16 discrete movement steps executed from the bottom to the top pairs of targets (phase 1, steps 1-8) and then from the top pairs to the bottom ones (phase 2, steps 9-16), each movement separated by 10s wait, with the execution sequence shown in Figure 3.3. 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 movements 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.

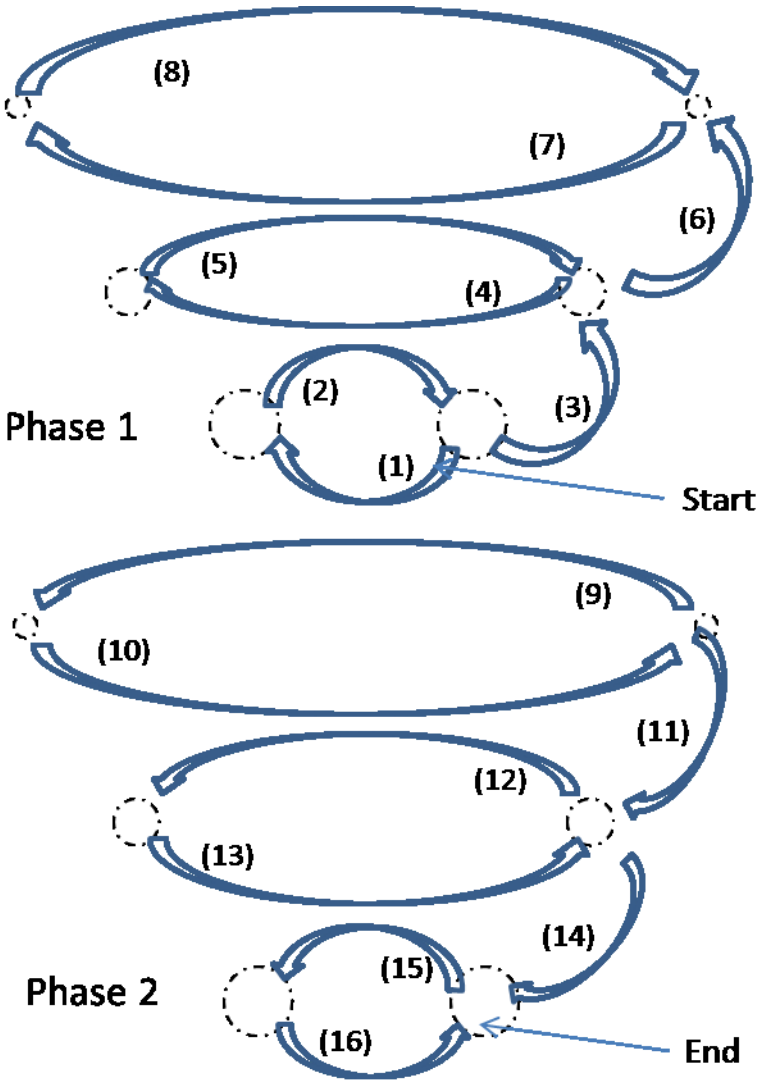


Figure 3.3 Execution sequence of a trial (target setting 1), steps shown in parentheses.

Note. The arrows represent tool movement from one circle to another; before each tool movement, the tool stops on the circle for 10s.

Each participant read and signed a 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 of tasks, each containing three trials with a 20s break between trials. To achieve counterbalance, each block used either task setting 1 (from easy to hard as shown in the top panel in *Figure 3.3*) or task setting 2 (from hard to easy as shown in the bottom panel in *Figure 3.3*). Half of the participants started with task setting 1 and the other started with task setting

2. We did not consider other ordering options like executing from medium 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.

Data analyses

Tooltip location

The task video was recorded by the camera inside the training box at 30Hz with a resolution of 352×288 pixels, and the text files (Combined Data file, CMD) containing eye movements (50Hz) were exported from Tobii Clearview for the offline analysis.

To analyze the pupil responses to tool movements, we needed the information of the tool movements that the pupil responded to. Therefore an algorithm was developed to automatically extract tooltip positions from the task videos. The detail of the algorithm for extracting tooltip positions is described in Appendix A.

The tooltip data were smoothed with a running-average-filter using equally weighted four samples window and then were synchronized to pupil signal in timeline as shown in *Figure 3.4*.

Tool movement division

A movement is defined from the tool starts to leave the current circle (Tool-leave), to when the tool reaches the vicinity above the target (Tool-reach), and eventually touches the target (Tool-touch). An algorithm was developed to detect the moments of Tool-leave and Tool-reach, as shown the dashed and solid vertical blue lines in *Figure 3.4*, respectively.

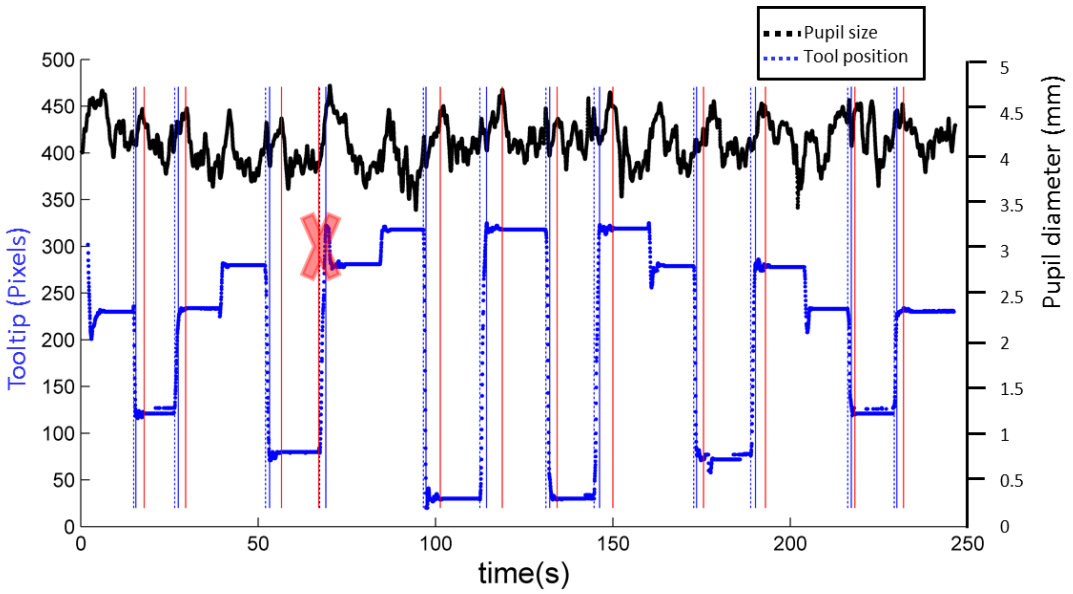


Figure 3.4 An example of tooltip movement and pupil size over time for a complete trial (subject 01, trial 01).

Note. The blue curve is the tooltip position in pixels and the black curve is the pupil diameter. The dashed and solid vertical blue lines and the solid vertical red lines represent the moments of Tool-leave, Tool-reach, and Tool-touch, respectively. The red cross at around 70s indicates a mis-operation (the tool points to a wrong target).

The algorithm finds the absolute tooltip movement peak velocity in a movement, and then searches backward and forward for the moments of Tool-leave and Tool-reach respectively, by checking whether the x-axis absolute velocity is lower than a threshold. The tool movement velocity was calculated by dividing the distance in x-axis direction (in pixels) for two consecutive movements to their time interval. The velocity thresholds for detecting the Tool-leave and Tool-reach moments were empirically determined, since there are abrupt changes of tool velocity at the moments of Tool-leave and Tool-reach. The Tool-leave and Tool-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 3.5*.

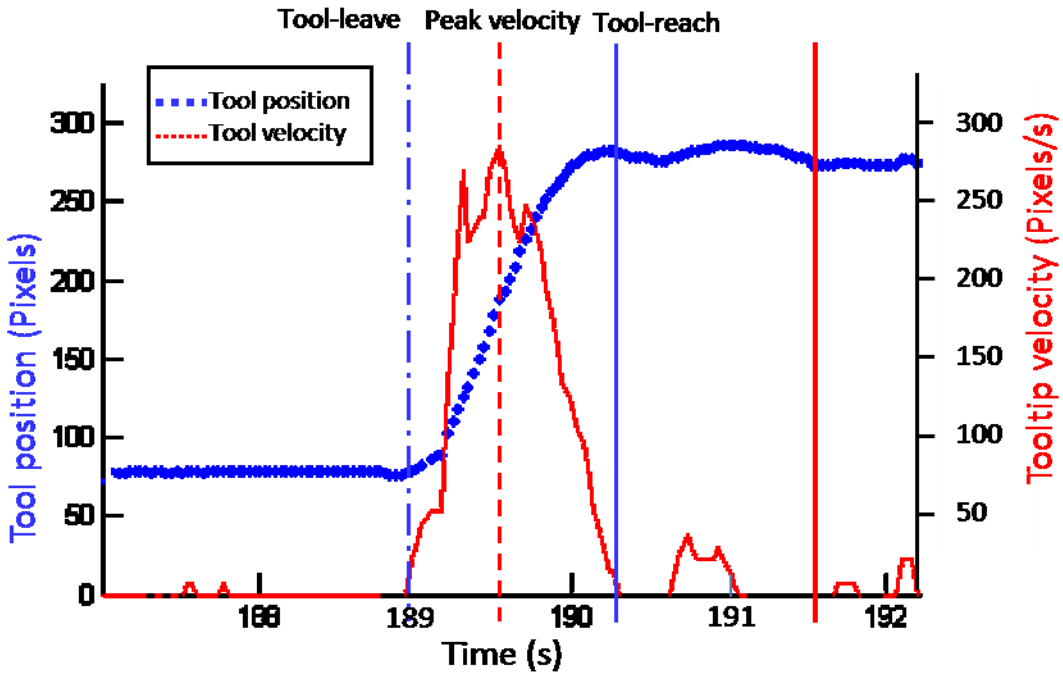


Figure 3.5 Kinematics of tool movement (subject 01, trial 01, from Experiment 1) during a typical horizontal movement between circles.

Note. 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 movement. The vertical dashed and solid blue lines represent the moments of Tool-leave and Tool-reach, and the vertical dashed and solid red lines represent the moments of peak velocity of a tool movement and the tool touches the target.

The tooltip usually leaves the current circle to the next target quickly in the horizontal direction once the tool starts to move, which makes it easy to accurately detect the Tool-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. Therefore the detected Tool-reach moment by this algorithm is not the moment that the tool touches the target circle. Therefore, we manually annotated the moments where the tool touches the target (called Tool-touch) by observing the surgical videos, as shown in the solid vertical red lines in *Figure 3.4* and in *Figure 3.5.*

Tool movement phases

A movement was further divided into Transport and Landing phases. The Transport phase starts when the tool leaves (Tool-leave) for the target circle and stops

when the tooltip reaches the vicinity above the target circle (Tool-reach). Starting from Tool-reach, the Landing phase slowly adjusts the tool till it touches the target circle (Tool-touch). This Transport and Landing phases division is consistent with the phase division (initial impulse and error correction phases) in rapid movement studies (Abrams, et al., 1990; Adam et al., 2000; Elliott, et al., 2001; Pelz, Hayhoe, & Loeber, 2001).

Movement-related pupil responses

The pupil diameter data was processed and synchronized with the trajectory of the tooltip for the analysis of the pupil responses during tool movements. The pupil diameter data was exported from Tobii Clearview to the text file (Combined Data file, CMD) and the tooltip positions were derived from the task video recorded by the built-in camera. The pupil data were in 50Hz and the task videos were in 30Hz with a resolution of 352×288 pixels. Segments of missed pupil data shorter than 100ms were interpolated, which might contain blinks and noise data. Then a Butterworth low-pass filter with a cut-off 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).

The pupil response data of a movement were extracted from a 7-second window starting 3 seconds before and 4 seconds after every Tool-leave, as shown in *Figure 3.6*. 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 (Moresi, et al., 2011; Privitera, et al., 2010; Richer & Beatty, 1985) and the pupil continuous to dilate afterwards to a peak value in a few seconds e.g. 3s-window in (Privitera, et al., 2010), 4s-window in (Richer & Beatty, 1985), and 5s-window in (Moresi, et al., 2011).

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. *Figure 3.7* shows a blow-up of a segment of the sample trial data from *Figure 3.6*.

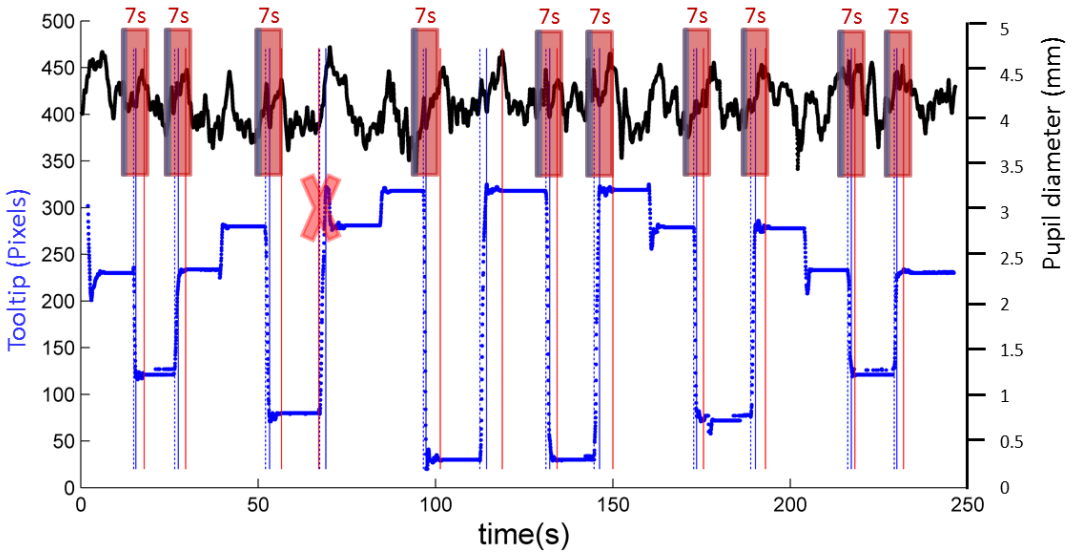


Figure 3.6 Illustration of the 7s windows extracting pupil responses over a complete trial (subject 01, trial 01).

Note. The blue curve is the tooltip position in pixels along the horizontal line between the circles. The dash and solid vertical blue lines and the solid vertical red lines represent the moments of Tool-start, Tool-reach, and Tool-touch, respectively. The red rectangles on the top represent the 7s-windows which start 3s before Tool-leave and end 4s afterword. The blue rectangle at the start of each 7s-window represents the baseline area for the window.

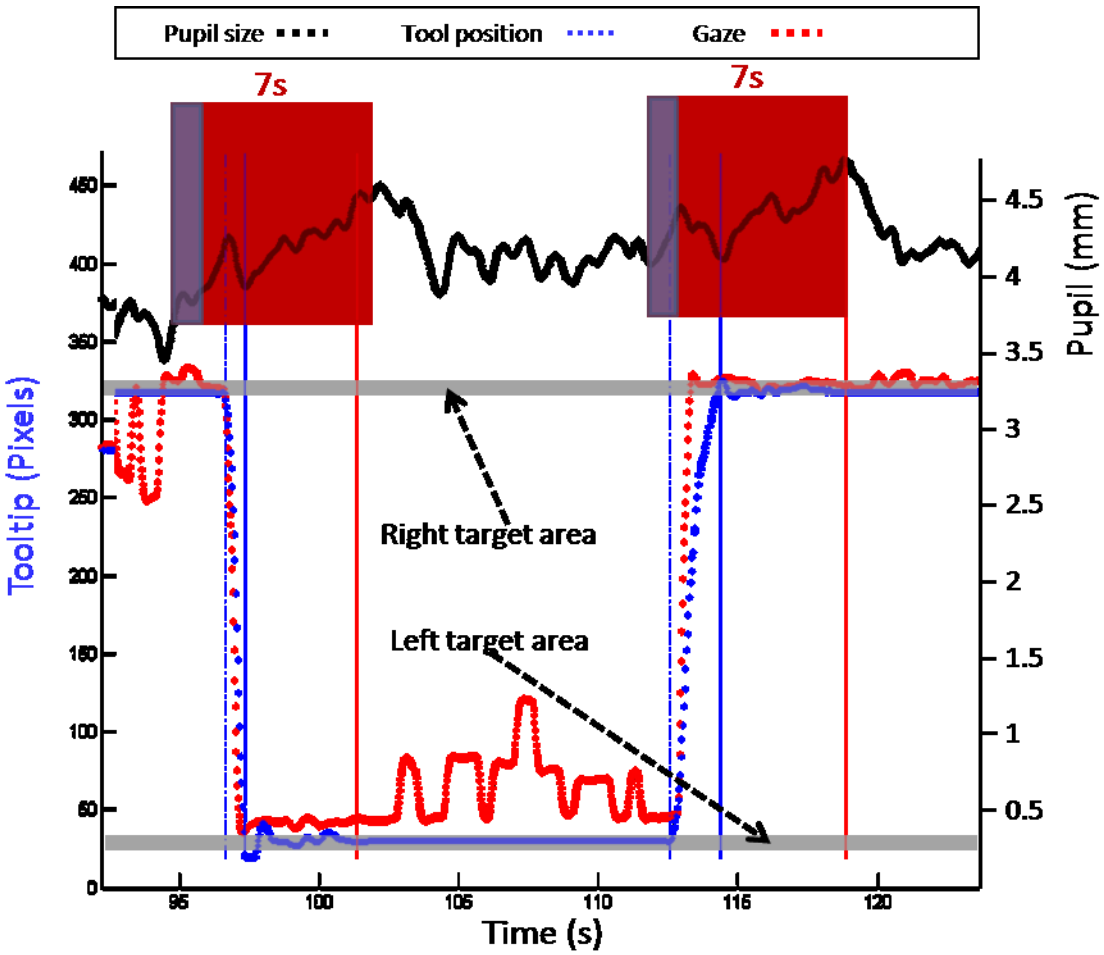


Figure 3.7 Blow up of a segment (95s-120s) from Figure 3.6, showing how the pupil size (black curve) increases when tool movements occur.

Note. The black, blue and red curves are the pupil diameter in mm, tooltip position and gaze location in pixels along the horizontal line between the circles. The dash and solid vertical blue lines and the solid vertical red lines represent the moments of Tool-start, Tool-reach, and Tool-touch, respectively. The red rectangles on the top represent the 7s-windows which start 3s before Tool-leave and end 4s afterword. The blue rectangle at the start of each 7s-window represents the baseline area for the window.

Since the pupil diameter baseline may drift during a task, usually baseline techniques (Bednarik, et al., 2012) are used to normalize the pupil size changes and eliminate the effect of drift of mean pupil size. The way that the baseline was chosen e.g. the position of the baseline may greatly affect the results of the data analysis (Bednarik, et al., 2012). The ideal baseline for event-related pupil diameter analysis should be derived from the period in the vicinity of the event where the pupil diameter is free from effects of the other events as possible.

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 (usually the moment of the stimulus onset), noise and other pupil size changes not correlated in time to the stimulus will be averaged to zero and the useful pupil size changes related to the task will be preserved (Klingner, et al., 2008). This is the signal averaging technique. Formally, assuming there are *m* movements from all trials of all subjects and each movement 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 $\overline{X} = \frac{\sum_{i=1}^{m} x_{ik}}{m}, k = 1...n$. 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, a paired *t*-test was applied 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).

For each 7-second window, the smallest and the biggest pupil sizes after tool-leave was searched as “valley” and “peak” pupil size respectively, and then the Valley-to-Peak pupil dilation (the difference between the peak and valley pupil sizes), the Valley-to-Peak pupil dilation duration (from the valley time point to the peak time point), and the peak pupil dilation duration (from the tool-leave time point to the peak time point) were calculated.

Experimental design

The experiment had an independent variable task difficulty with three levels of Easy, Medium, and Hard. The dependent variables were movement time, Valley-to-Peak pupil dilation, Valley-to-Peak pupil duration, and peak pupil dilation duration.

The total number of movements was 864 (12 participants, each performing 6 trials, each trial has 12 movements).

3.1.3. Results

Accuracy

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. However, we discarded 23 windows due to mis-operation such as when the participant moved the grasper to a wrong target. All the movements' endpoints (the tooltip positions when touching the target) were within the target circle. Therefore we had 805 valid movements.

Movement time

The mean movement time (MT) is the mean transportation time between Tool-leave and Tool-touch for all horizontal movements. The mean MT for all ID is 2.6±1.1s, and the MT changes as a function of ID, i.e., the mean MT increases when the difficulty level increases, as shown in the mustard bars of *Figure 3.8*. There is significant main effect between the three IDs in terms of mean MT ($F_{2,802}$ = 237.069, p < .0001). Post Hoc test (Tukey HSD) shows that the differences between pairs of the three IDs are significant (p < .0001), with mean MT of Easy, Medium, and Hard IDs being 1.8±0.6s, 2.6±0.8s, and 3.4±1.1s, respectively.

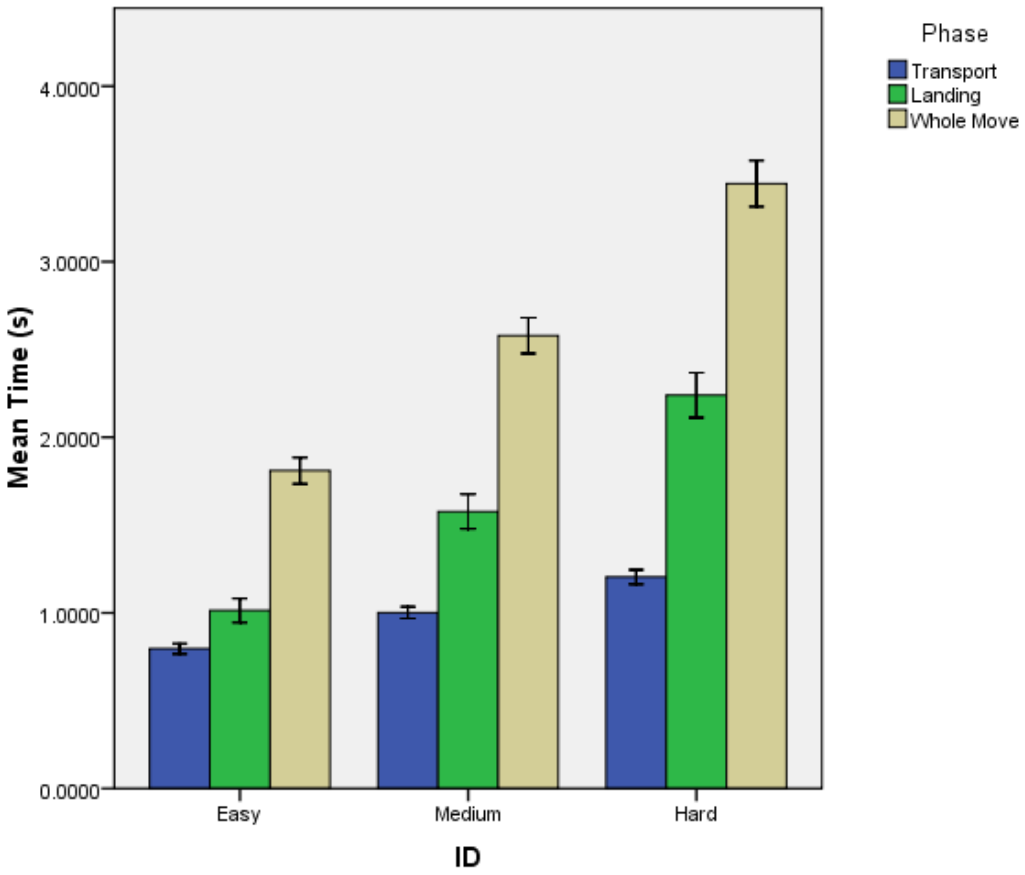


Figure 3.8 Mean time of movement phases and complete movement time for different IDs over all subjects.

Note. The blue bar is the movement time for the Transport phase, the green bar is the movement time for the Landing phase, and the mustard bar is the movement time for a complete move. The error bars are for 95% confidence intervals.

The mean MT of *Transport* phase of all IDs is 1.0±0.3s, and increases when the difficulty level increases, as shown in the blue bars of *Figure 3.8*. There is significant main effect between the three IDs in terms of mean Transport MT ($F_{2,802}$ = 134.150, p < .0001). Post Hoc test (Tukey HSD) shows that the differences between pairs of the three IDs are significant (p < .0001), with mean Transport MTs of Easy, Medium, and Hard IDs being 0.8±0.2s, 1.0±0.3s, and 1.2±0.3s, respectively.

The mean MT of *Landing* phase of all IDs is 1.6±1.0s, and increases when the difficulty level increases, as shown in the green bars of *Figure 3.8*. There is a significant main effect between the three IDs in terms of mean Landing MT ($F_{2,802}$ = 144.196, p <

.0001). Post Hoc test (Tukey HSD) shows that the differences between pairs of the three IDs are significant ($p < .0001$), with mean Landing MTs of Easy, Medium, and Hard IDs being 1.0±0.6s, 1.6±0.8s, and 2.2±1.1s, respectively.

Pupil responses to tool movements

Descriptive analysis

Figure 3.9 shows the mean changes of pupil diameter during horizontal tooltip movements over a 7-second window. The data is averaged across 805 valid moves of 69 trials from 12 subjects. For clarity, error bars are drawn every 1s rather than drawn for every time sample (at 50Hz).

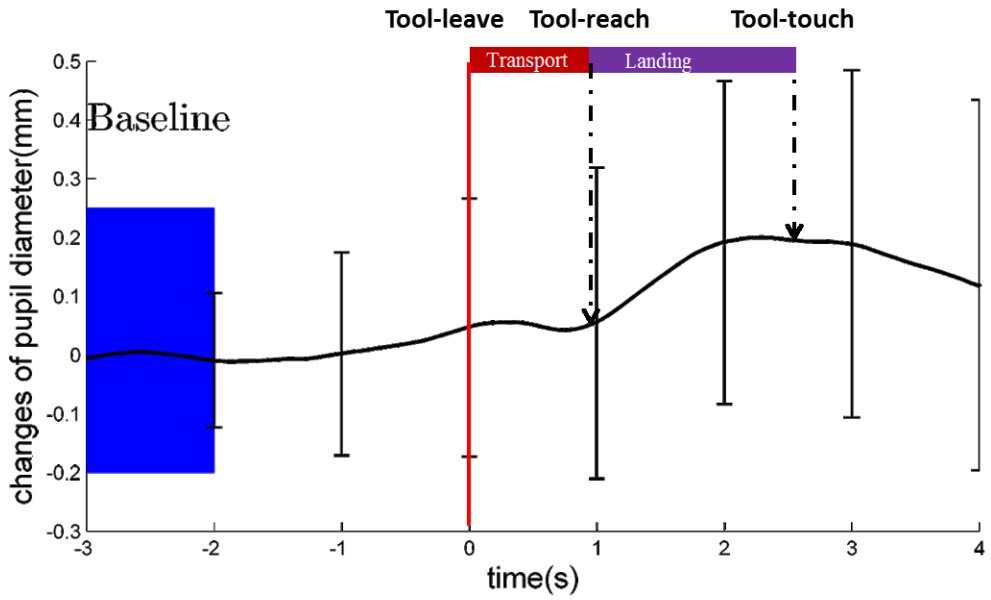


Figure 3.9 Mean pupil diameter changes for 805 valid moves of 69 trials from 12 subjects.

Note. Data were aligned over a 7 second window 3 seconds before the Tool-leave. 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 vertical red line is Tool-leave where all the data are aligned and the next two vertical dashed black arrows are the average Tool-reach time and Tool-touch time. The error bars for 1 std. dev. are drawn every 1s.

Figure 3.10 shows the means of pupil size changes in the 7s-window for the three IDs. The three ID curves share a common pupil change pattern as shown in *Figure 3.9*, i.e., all of them dilate in response to the movement preparation and execution in a relatively long period from about 1.5s before Tool-leave and peak 1-2s after Tool-

reach, corresponding to the movement preparation and execution (Privitera, et al., 2010; Richer & Beatty, 1985). There is also a small constriction commencing 200ms after the tooltip starts to move during Transport phase.

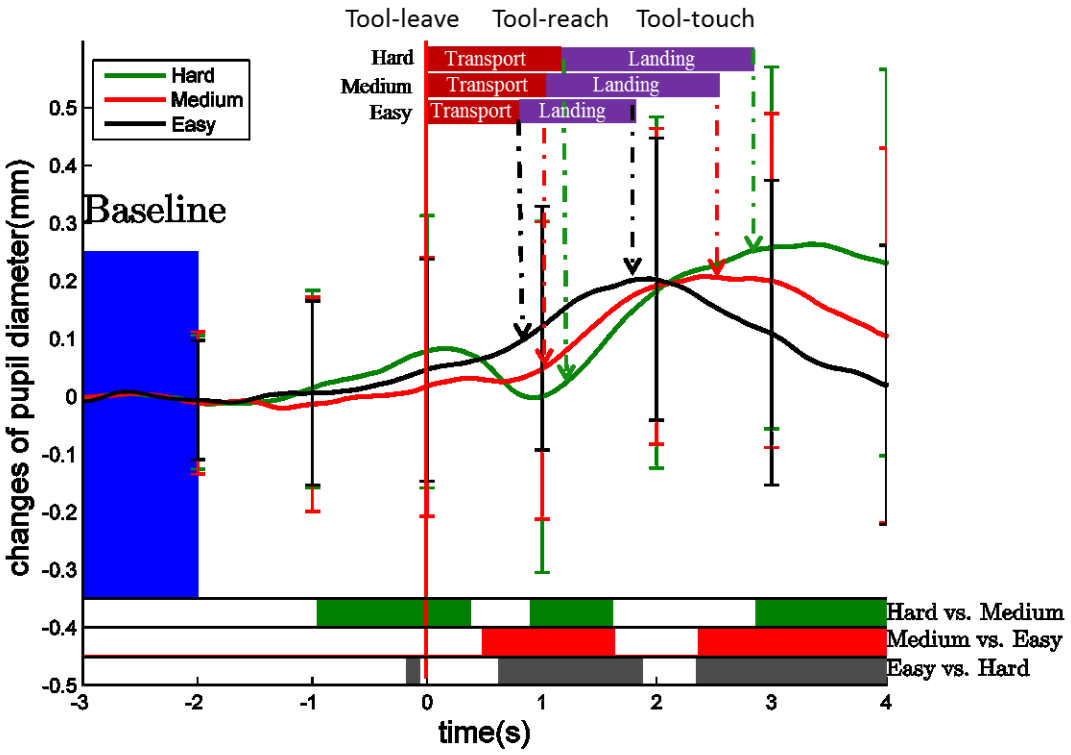


Figure 3.10 Mean pupil diameter changes against different IDs.

Note. Data are aligned over a 7-second window around Tool-leave. The vertical red line is the Tool-leave. The three colors of bars at the bottom indicate significant differences in pupil dilation between Easy, Medium and Hard ID with black representing Easy vs. Hard, red representing Easy vs. Medium, and green representing Medium vs. Hard. The error bars for 1 std. dev. are drawn every 1s.

However, major differences can be observed between the three IDs. First, during 2-3s in *Figure 3.10*, the pupil dilation for the Hard ID (green curve) is more than Easy (black) and Medium (red) ones. Second, during Transport phase, the pupil of the Hard ID constricts the most among the three IDs and the pupil of the Easy ID constricts the least. Third, the peak value and duration from Tool-reach 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 0s and 1s along the time window in *Figure 3.10*) 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 medium task (red 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

The moment-to-moment based difference was examined between the pupil size change curves for the three IDs, with the results as shown in the bottom color bars in *Figure 3.10*. 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 changes between Easy ID and Hard ID are shown in black bars (bottom row) in *Figure 3.10*. The first significant time period is located at 0.7s to 1.9s 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 2s) to the end of the window.

The red bars (middle row) in *Figure 3.10* show similar significant differences between Easy ID and Medium ID, only with the first bar starting a little bit earlier (0.5s to 1.6s in time axis).

The green bars (upper row) in *Figure 3.10* show significant differences between Medium ID and Hard ID. There are three major significant time periods: the first one (-1s to 0.4s along the time window) where the pupil dilates more during the hard ID than the Medium ID, the second one (0.9s to 1.6s along time window) where the pupil constricts more during the Hard ID than the Medium ID, and the third one starts after 2.8s along the time window.

According to the three colored bars at the bottom in *Figure 3.10*, significant differences between all three IDs can be found in two major time periods: one is around the end of the Transport phase 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.

Figure 3.11 shows the mean amount of the Valley-to-Peak pupil dilation for the three IDs. ANOVA shows there is significant difference in the pupil size increase from Valley-to-Peak between IDs ($F_{2,802}$ = 44.754, p < .0001). Post Hoc test (Tukey HSD) also shows that the mean increase of pupil size from valley to peak is significantly different (p < .0001) between Easy (0.3±0.2mm), Medium (0.4±0.2mm), and Hard ID (0.5±0.2mm).

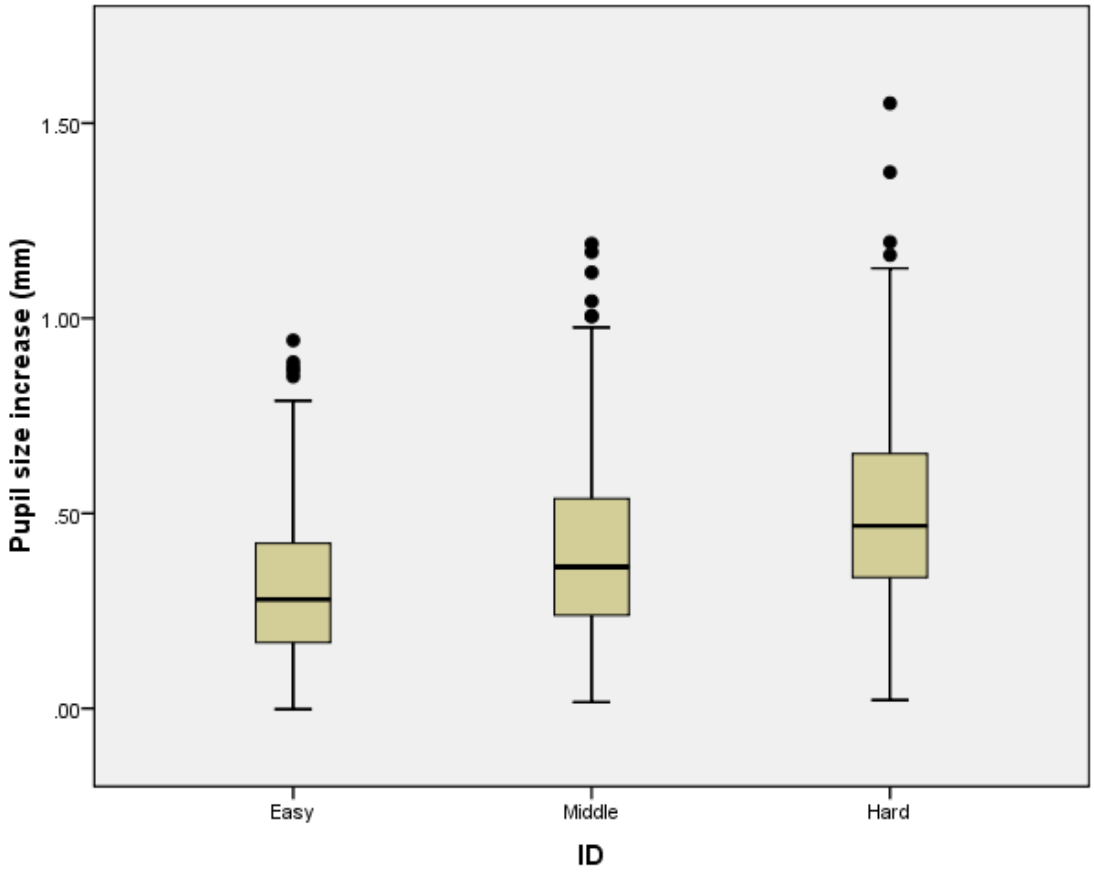


Figure 3.11 Box-whisker plot for Mean Valley-to-Peak pupil dilation for three difficulty IDs.

Figure 3.12 shows the mean Valley-to-Peak duration of pupil dilation for three IDs. ANOVA shows there is significant difference in the duration of the pupil dilation from its valley to peak size between IDs ($F_{2,802}$ = 91.938, p < .0001). Post Hoc test

(Tukey HSD) shows that the duration from valley to peak is significantly different (p < .0001) between Easy (1.6±0.7s), Medium (2.1±0.9s), and Hard ID (2.7±1.1s).

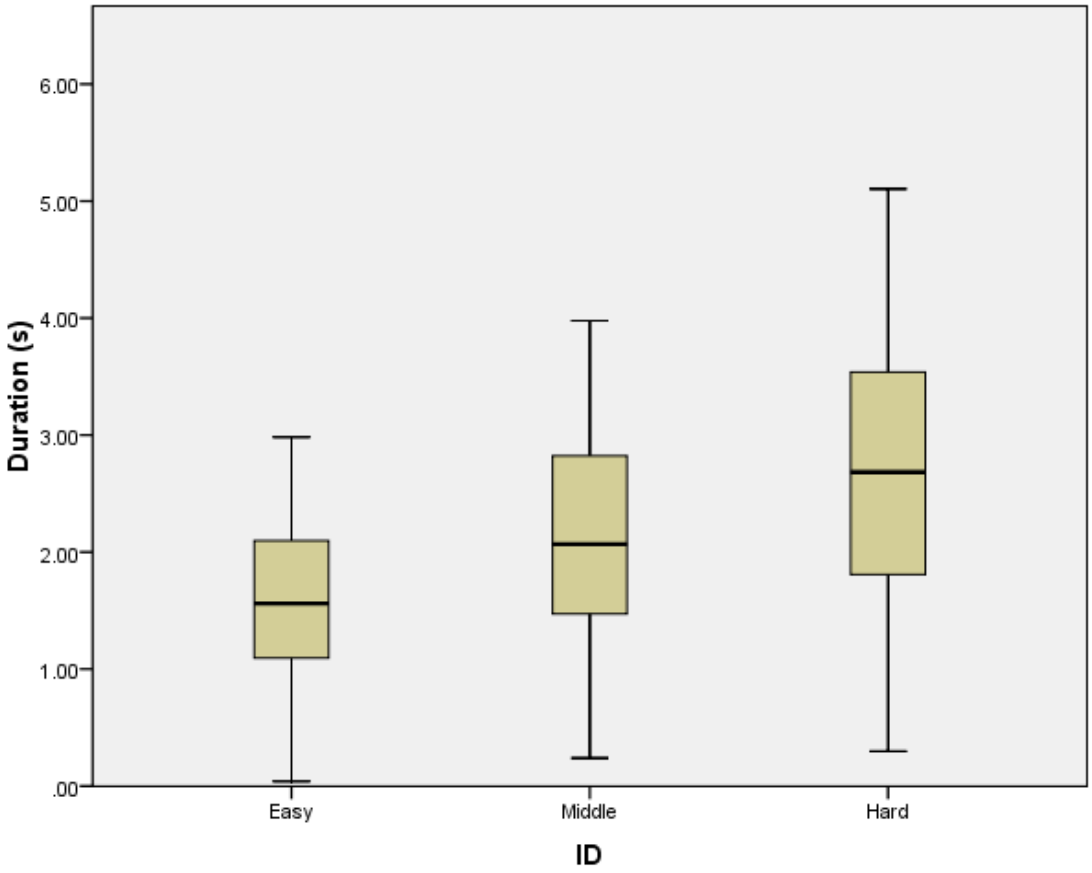


Figure 3.12 Box-whisker plot for mean Valley-to-Peak pupil dilation duration for three difficulty IDs.

Figure 3.13 shows the duration from Tool-reach to the moment where the pupil peaked in size for each ID, up to 4s after the tooltip reach. There is significant main effect between the three IDs in terms of this duration ($F_{2,802}$ = 30.558, p < .0001). Further Post Hoc test (Tukey HSD) shows that this duration for the Hard ID (2.2±1.0s) is significantly (p < .0001) longer than that of Easy (1.6±1.0s) and Medium (1.8±0.9s), and the Medium ID is significantly longer than that of Easy ID (p < .01).

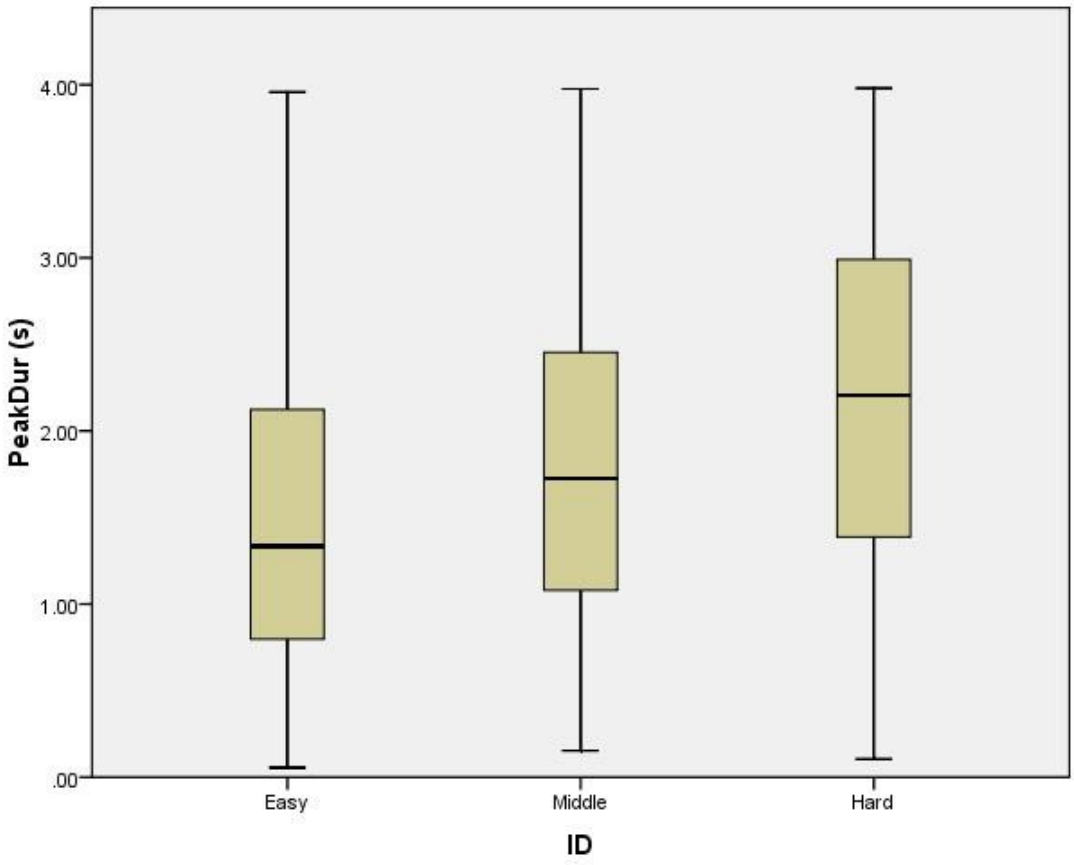


Figure 3.13 Box-whisker plot for Mean duration from Tool-reach to the moment where the pupil peaked in size for three difficulty IDs.

3.1.4. Discussion

The performance measures exhibited expected behavior—the MTs correlated perfectly to the increase of the ID, as shown in *Figure 3.8*. Therefore, Fitt's ID is a good measure for the task requirement in this study. Furthermore, both the mean MTs of Transport and Landing phases perfectly correlated to the increase of ID, where the timing of Transport phase positively correlated to the target distance and the timing of Landing phase negatively correlated to the 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 Tool-leave and continued 2s afterwards (*Figure 3.9*) and clearly reveals the task requirement during the preparation and execution of the movement. The preparation time for a movement (Abrams, et al., 1990; Elliott, et al., 2001) cannot be explicitly determined in this self-paced discrete aiming movement study, but we can roughly assume it is at the last minute of the 10s waiting time.

Our data is consistent with the finding of previous works (Moresi, et al., 2011; Privitera, et al., 2010; Richer & Beatty, 1985). The time that the pupil starts to dilate (1.5s before Tool-leave) in this study is same as that of Richer and Beatty's finding, while the peak duration (2s after Tool-leave) is longer than that of Richer and Beatty's (0.5s). The likely reason is that we have a tool movement time, while Richer and Beatty's is only a simple finger flexion.

The pupil constricts slightly during the Transport phase (*Figure 3.9*), which is the time when the tooltip moves speedily to a target. During the Transport phase of the aiming movement (corresponding to the initial impulse phase), the eye gaze usually saccades to the target ahead of tooltip for the visual guidance (Abrams, et al., 1990; Elliott, et al., 2001); the saccade causes the pupil constriction. Our results show that longer travel distances require longer saccades, causing deeper pupil constrictions, as shown in *Figure 3.10*.

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 *t*-testing shown in *Figure 3.10*, where the three pairs of IDs (Easy vs. Hard, Easy vs. Medium, and Medium vs. Hard) all show significant differences.

Although the peak pupil size of Easy and Medium ID shows similar amplitudes in *Figure 3.10*, 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 Tool-reach and peak pupil size, as shown in *Figure 3.13*) was correlated with the three IDs, i.e. Hard ID has the longest delay, and Medium ID has the medium one. This is consistent with the finding of previous work (Richer & Beatty, 1985).

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 ($p < .0001$) ability in distinguishing the three difficulty levels of task requirement.

The lighting condition and the screen illumination of this experiment were well-controlled. The ambient lighting in the case study room was kept constant throughout the entire experiment data collection. 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. The background of the screen was uniformly grey.

Besides lighting conditions, one may be concerned that the pupil dilation or constriction are affected by the viewing angle to the eye-tracker camera due to the changes of gaze location (Brisson et al., 2013; Gagl, et al., 2011) during target-pointing. In our case, the tool quickly moves from a target circle to another and the eyes mostly fixate on the targets, as shown in the red curve in *Figure 3.7*. To clarify the gaze angle problem, we only need to check whether the baseline pupil diameter of movements in two different directions (moving from right to left vs. from left to right) are significantly different. In this study, there is no significant difference between the baselines of left and right movements. Therefore, we are sure that gaze angle does not affect the measured pupil size in this study.

Similarly, concerns may also be raised that the pupil response could be affected by the variation of the illumination for different sizes of the dashed black target circles (e.g. Easy ID has the biggest circles) (Kun, Palinko, & Razumenić, 2012). However, there was no significant effect of the target size (task difficulty) on the pupil baseline.

For real-world application, several steps to capture the movement-related pupil responses should be carefully considered. First, critical moments of the movement (the

moments the tool/hand starts to move or reach) should be accurately detected as timing markers 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 the tool may need extra actions at the end of the 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.

3.2. Pupil responses to target size and distance during discrete goal-directed movements (Experiment 2)

In Experiment 1, we explored the pupil responses to the task requirement in a target-pointing task, i.e. moving a tool to point and touch the pairs of target circles with their size and distance defined following Fitts' index of difficulty (ID). We found that the pupil responds to the task requirement following a common pattern—the pupil dilates significantly (about 0.05mm) starting about 1.5s before the aiming movement, and then constricts slightly (about 0.02mm) 200ms after Tool-leave to a lower size (the valley). Before the tool movement ends, the pupil peaks in size, ranging from 0.2mm to 0.25mm according to the IDs. The Valley-to-Peak pupil dilation and duration positively correlate

to the increase of IDs. This 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 mixed effect of both target size and target distance changes. Previous studies have found that the pupil may constrict during saccades (Abrams, et al., 1990; Elliott, et al., 2001). If pupil size has reversed relationship to the speed of saccade, then we like to know if distance and the size of target can separately affect pupil response. If it does, then we like to ask to what degree of each target property affects the pupil. Specifically, we would like to know whether the two parameters used for calculating the ID, the target size and target distance, affect the pupil size differently.

Therefore, we decided to conduct Experiment 2 using a similar setting to that of Experiment 1, but asked the participants to perform two different tasks: 1) aiming at targets with 4 different sizes located at a constant distance apart, and 2) aiming at targets varying in 4 different distances.

3.2.1. Experiment purpose and hypothesis

Experiment 2 aimed to examine whether the variation of target size and distance has separate influence on the pupil size.

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. 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 changes.

3.2.2. Methods

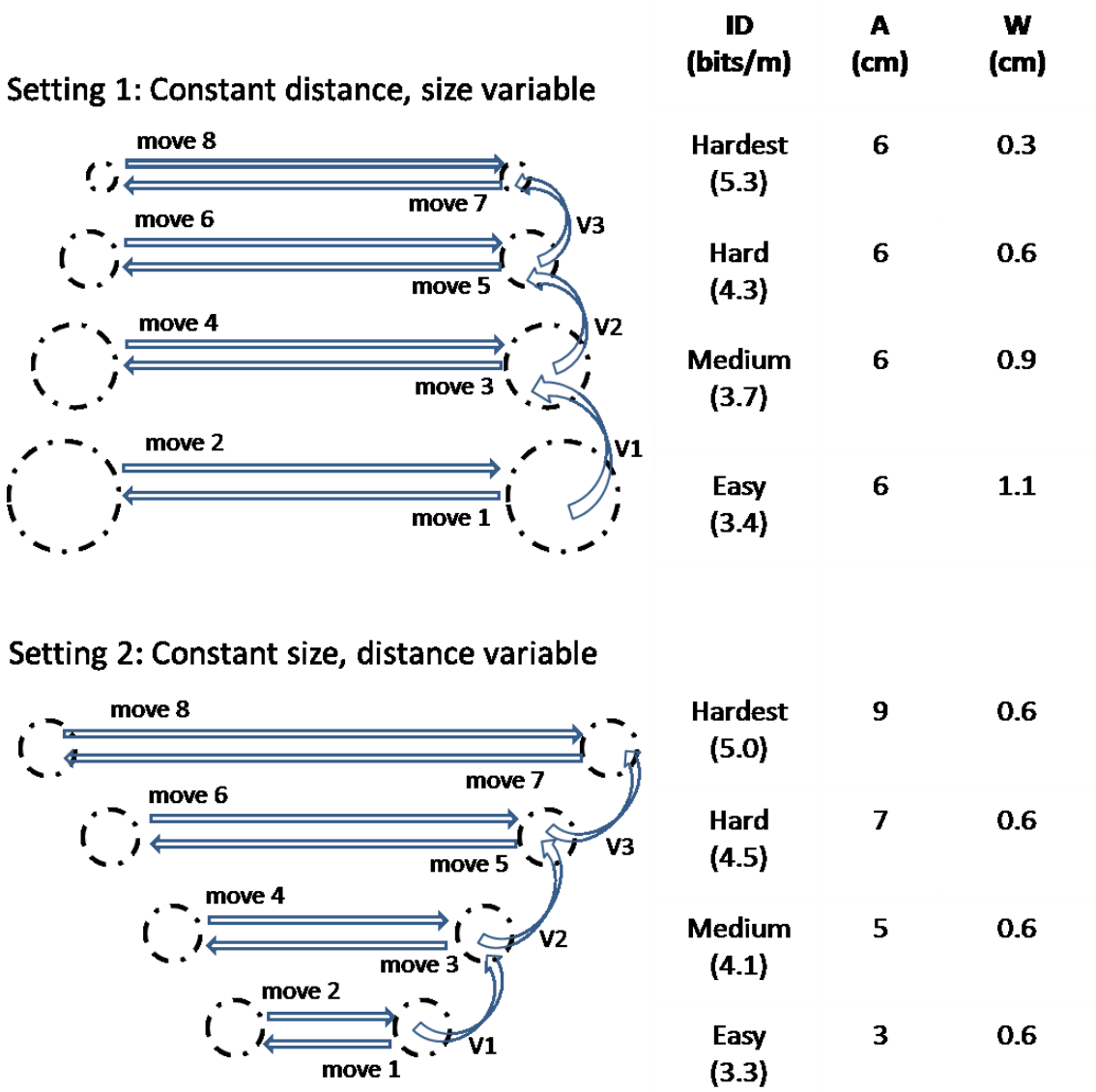
Participants

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

Experimental setting and apparatus

In order to clearly separate the effects of pupil responses to either the changes of target size or distance, one of target size and distance is kept constant in two different task settings, as shown in *Figure 3.14*. 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) (Paul M. 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 bit/response; and the four IDs for setting 2 are ID1=3.3, ID2=4.1, ID3=4.5, and ID4=5.0 bit/response. ID1, ID2, ID3, and ID4 are denoted as Easy, Medium, Hard, and Hardest respectively in this study.

The apparatus are same as that in Experiment 1 as shown in *Figure 3.1*.



	ID (bits/m)	A (cm)	W (cm)
Hardest	(5.3)	6	0.3
Hard	(4.3)	6	0.6
Medium	(3.7)	6	0.9
Easy	(3.4)	6	1.1
Hardest	(5.0)	9	0.6
Hard	(4.5)	7	0.6
Medium	(4.1)	5	0.6
Easy	(3.3)	3	0.6

Figure 3.14 Target settings and task execution sequences.

Note. 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.

Task and procedure

The task was to point and touch 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 a surgical grasper. This was a discrete target-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 move8) between the targets in pairs and 3 vertical transfer (V1 to V3), each separated by 10s, as shown in *Figure 3.14*. 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 felt ready to begin.

Each participant performed two blocks of tasks using the two target settings shown in *Figure 3.14* in a counterbalanced 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 versa. Each block consisted 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 trials. Between two blocks, the target setting was changed.

Data analysis

Tooltip location

The tooltip positions were automatically extracted using the algorithm described in Appendix A and were smoothed with a running-average-filtered using equally weighted four samples window. The moments when the tooltip started to move (Tool-leave) and reached (Tool-reach) the target, were detected by the algorithm described in Appendix A. Similar as Experiment 1, we manually annotated moments of Tool-touch by observing the surgical videos, and a movement was further divided into Transport (from Tool-leave to Tool-reach) and Landing (from Tool-reach to Tool-touch) phases.

Movement-related pupil responses

The pupil diameter data was exported from Tobii Clearview to the text file (Combined Data file, CMD). Pupil data segments of missed pupil data shorter than 100ms were linearly interpolated and a Butterworth low-pass filter with a cut-off 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).

Same as in Experiment 1, we were only interested in the pupil size changes in the window around the Tool-leave and Tool-reach, i.e. 3 seconds before Tool-leave and 4 seconds after Tool-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.

All the data in the windows were aligned at the Tool-leave moment (3 seconds into the window), and the mean pupil diameter changes were calculated for each time point in the window across all horizontal tooltip movements from all trials. Similarly to analysis of Experiment 1, 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, 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 Tool-leave, and the pupil peak duration (from Tool-leave) was recorded as well.

Experimental design

The experiment had an independent variable in each task setting which is the task difficulty with four levels of Easy, Medium, Hard, and Hardest. The dependent variables were movement time, peak pupil dilation, and peak pupil dilation duration.

The total number of movements was 768 (8 participants, each performing 6 trials in Constant Distance setting and 6 trials in Constant Size setting, each trial has 8 movements (each trial had 4 ID executions, with each ID execution having 2 moves—from right to left and left to right)).

3.2.3. Results

Accuracy

A total of 96 trials were recorded. Two trials in the CD setting and four trials in the 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 was window data for 720 horizontal tool movements available. However, we discarded 4 windows in the CD setting and 9 windows in the CS setting due to the mis-operation by the participants, e.g., when they moved the grasper to a wrong target. Therefore we had 361 correct windows in the CD setting and 342 correct windows in the CS setting.

Among the 361 movements in the CD setting, a total of 20 movements' endpoints (the tooltip positions when touching the target) were outside their target circles (13 for Hardest ID, 4 for Hard ID, 3 for Medium ID, and none for Easy ID). Most movements' end points are within 2.5mm on the target paper (corresponding to 10 mm displayed on the 17" screen) to their corresponding target edges (12 movements). The 8 movements with their endpoint distance greater than 2.5mm to their target circle edges were discarded. Usually, all error end points should be included or excluded in the analysis, but as the error rate is less than 4% and we did not compare between conditions, these partly-erroneous included end points would not affect the results. The remaining 353 valid movements were further analyzed. In the CS setting, all the 342 movements' endpoints were within their target circles.

Movement Time

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

Figure 3.15 shows the linear regression of mean MT of each ID to ID values in both target settings with R^2=0.970 and $p < .001$, which means MT is positively correlated with ID values.

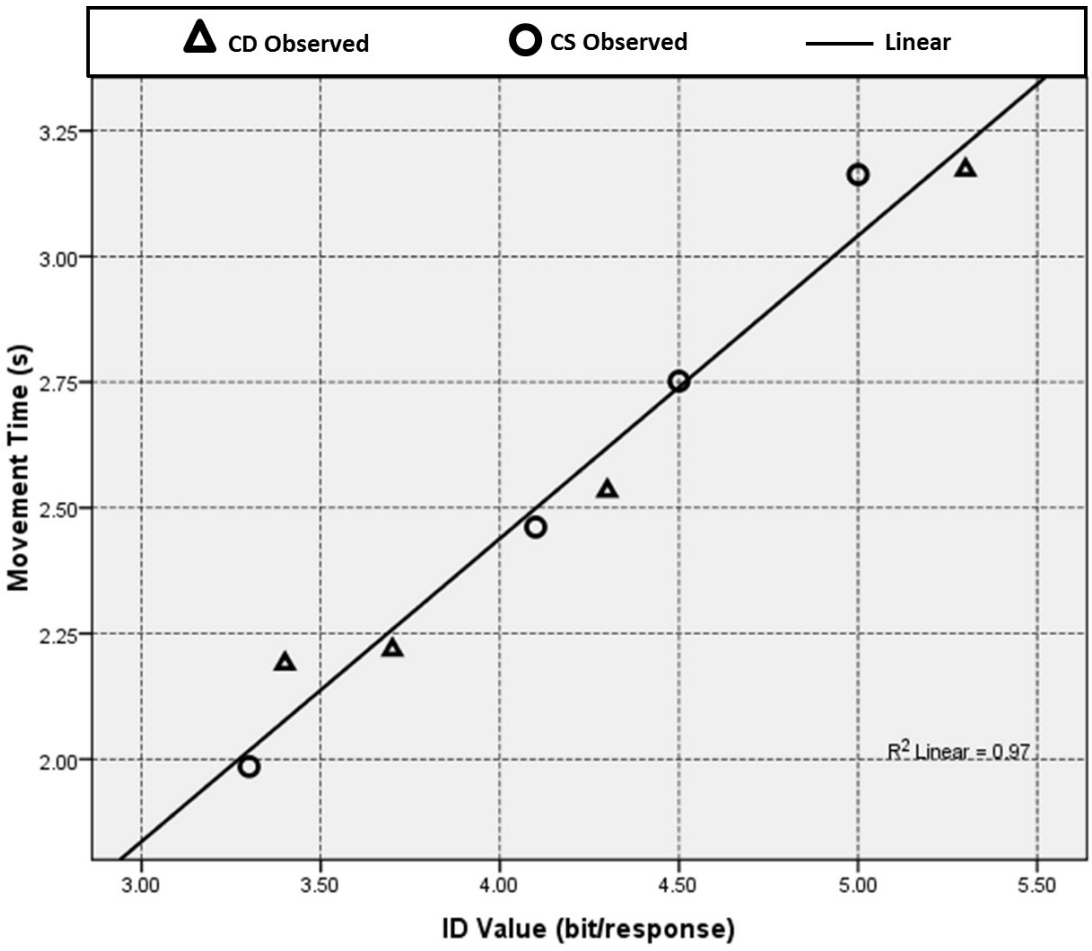


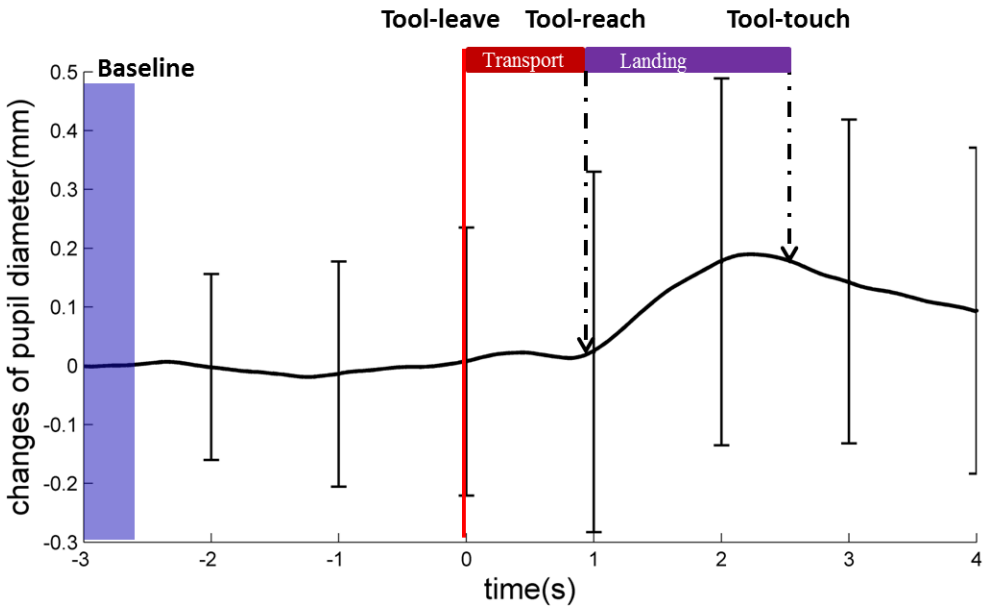
Figure 3.15 Linear regression of mean movement time (MT) of Each ID to Fitts' ID value (R^2=0.970 and $p < .001$).

In the constant distance (CD) setting, the mean MT for all IDs is 2.5±1.0s. There is significant main effect between four IDs in terms of mean MT ($F_{3,349}$ = 20.428, $p < .0001$). Post Hoc test (Tukey HSD) shows that the mean MT of Hardest ID (3.2±1.1s) is significantly longer ($p < .0001$) than other three (Easy, Medium, and Hard IDs being 2.2±0.9s, 2.2±0.9s, and 2.5±0.9s respectively), and there is no significant difference between other pairs of ID, but the Hard ID is marginally longer than that of the Easy ID ($p = .076$).

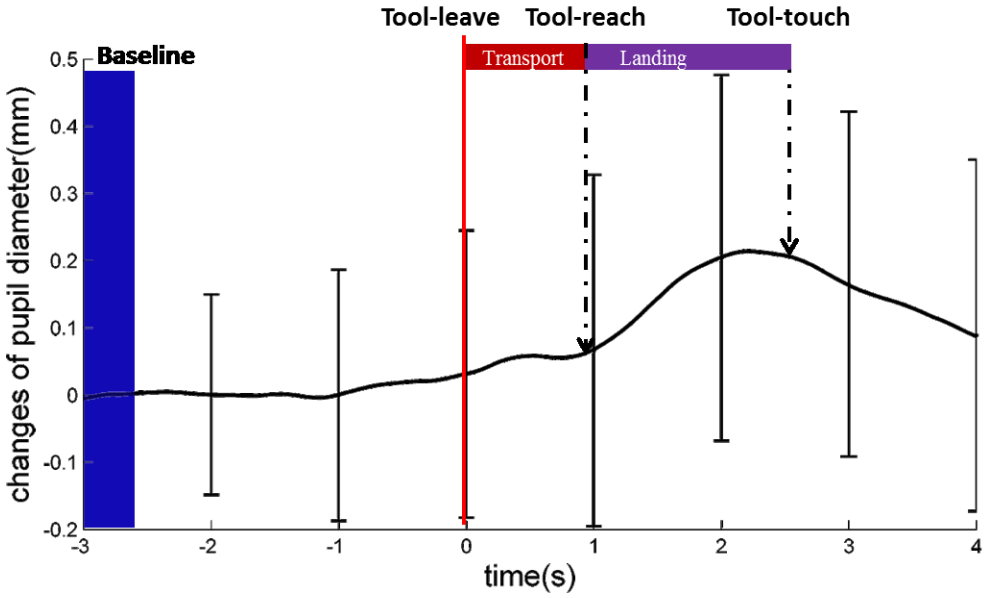
In the constant size setting (CS), the mean MT for all IDs is 2.5±1.0s. There is significant main effect between four IDs in terms of mean MT ($F_{3,338}$ = 18.389, p < .0001). Post Hoc test (Tukey HSD) shows there are significant differences (p < .05) between all the pairs of four IDs except that the Hardest (3.2±1.4s) is marginally longer than that of Hard ID (2.8±1.1s, p = .056) and there is no significant difference between Hard and Medium ID (2.5±1.0s).

Pupil responses to tool movements

Figure 3.16 shows the mean changes of pupil diameter during horizontal tooltip movements over a window in both the CD and CS settings. Both *Figure 3.16*A and *Figure 3.16*B show a very similar pupil diameter change pattern; the pupil starts to dilate slightly at 1.2s to 1.5s before Tool-leave (increasing less than 0.05mm), and then peaks (0.2mm compared to baseline) right before Tool-touch.



(A) Constant Distance (CD)

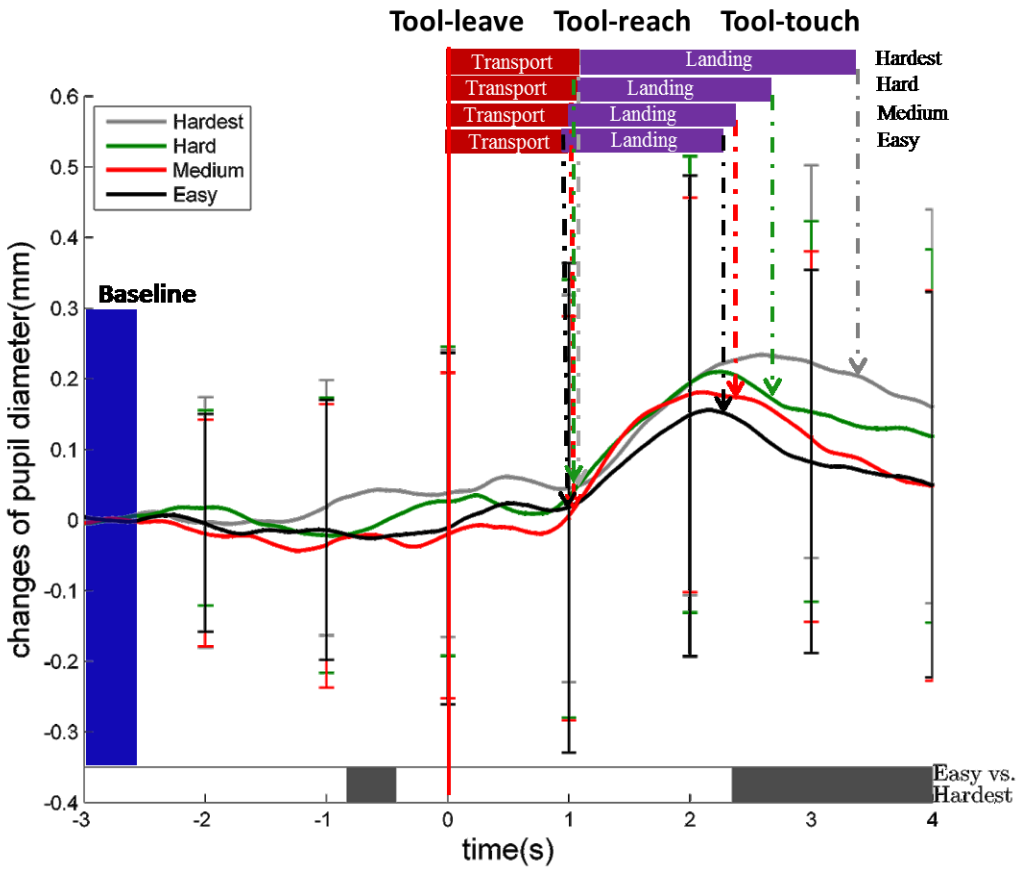


(B) Constant Size (CS)

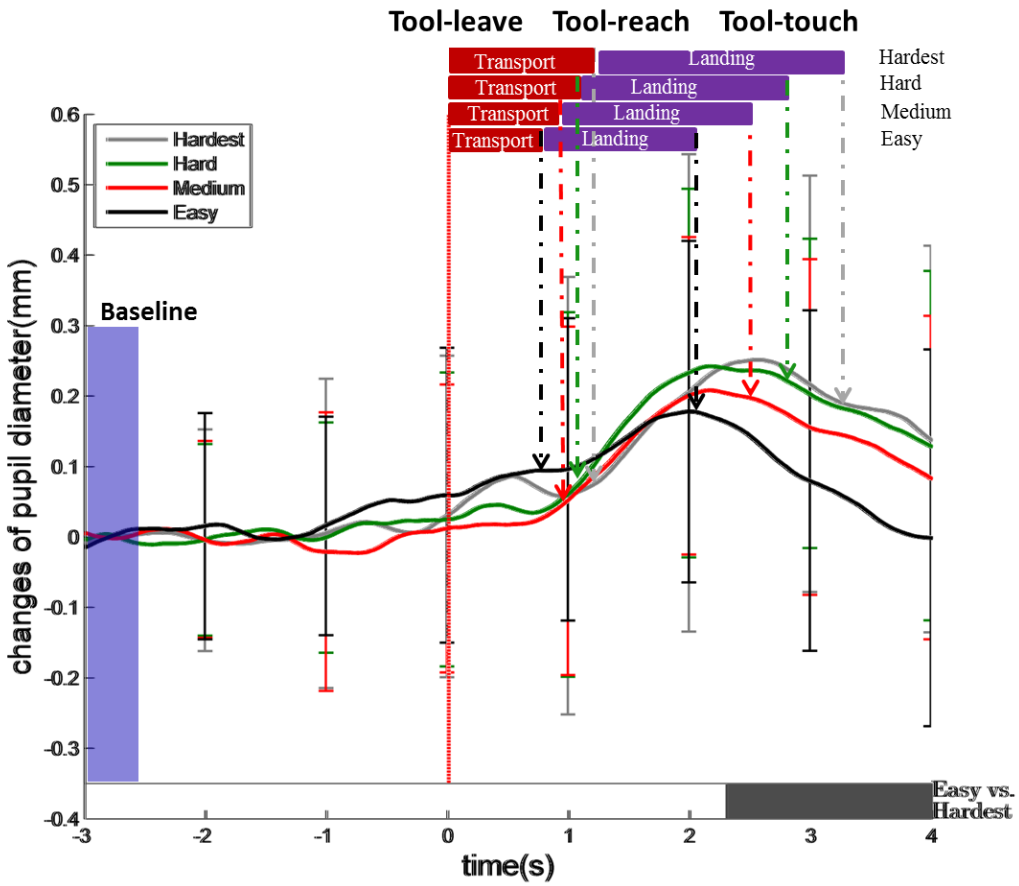
Figure 3.16 Mean pupil diameter changes in both target settings.

Note. (a) mean pupil diameter changes for 353 moves of 46 trials from 8 subjects in the CD setting; (b) mean pupil diameter changes for 342 moves of 44 trials from 8 subjects in the CS setting. Data were aligned over a 7-second window 3 seconds before the Tool-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 red line is the moment of Tool-leave where all the data are aligned and the rest two vertical dashed arrows are the mean moments the Tool-reach and Tool-touch. The error bars showing 1 std. dev. are drawn every 1s.

Figure 3.17 shows the mean changes of pupil diameter in a window for four IDs in both the CD and CS settings. Besides sharing a common pupil change pattern as shown in *Figure 3.16*, the four ID curves in both *Figure 3.17A* and *Figure 3.17B* are actually distinguishable. First, as shown in *Figure 3.17A*, the moments when the pupil starts to dilate are different for the four IDs; the Hardest ID (the grey solid curve) starts the earliest at around 1.5s before Tool-leave, the Easy ID (the black solid curve) starts the latest at around 0.5s before Tool-leave, and the Medium and Hard ones are in the medium.



(A) Constant Distance (CD)



(B) Constant Size (CS)

Figure 3.17 Mean pupil diameter changes against different IDs in two target settings.

Note. Data are aligned over a 7-second window around Tool-leave. The black, red, green, and grey curves are means of pupil changes cross all trials for Easy ID, Medium ID, Hard ID, and Hardest ID respectively. The read vertical line is the Tool-leave and other four color vertical arrows represent the Tool-reach and Tool-touch 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 1s.

Second, as shown in *Figure 3.17B*, the amplitude of the constrictions of the four curves in Transport phase 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 (grey solid curve) has the deepest constriction, and the other two IDs have the medium level of constrictions. Third, as shown in both *Figure 3.17A* and *Figure 3.17B*,

the peak pupil dilation value and duration positively correlates with the level of difficulty of the task, i.e., the harder ID has higher peak pupil dilation value and longer duration.

Moment-to-moment *t*-testing was applied between the curves for the Easy and Hardest IDs, with the results as shown in the bottom gray bars in *Figure 3.17A* and *Figure 3.17B*. For each time point, the paired *t*-test was performed for each time point for two IDs each containing samples from the all trials respectively. The horizontal black bars in both *Figure 3.17A* and *Figure 3.17B* represent the significant segment of the graphical significant testing between Easy and Hardest curves; there are significant differences between Easy and Hardest ID after 2.5s in both target settings.

Figure 3.18 shows the linear regression of mean peak pupil dilation for each ID to ID value, with $R^2 = 0.849$ and $p < .005$.

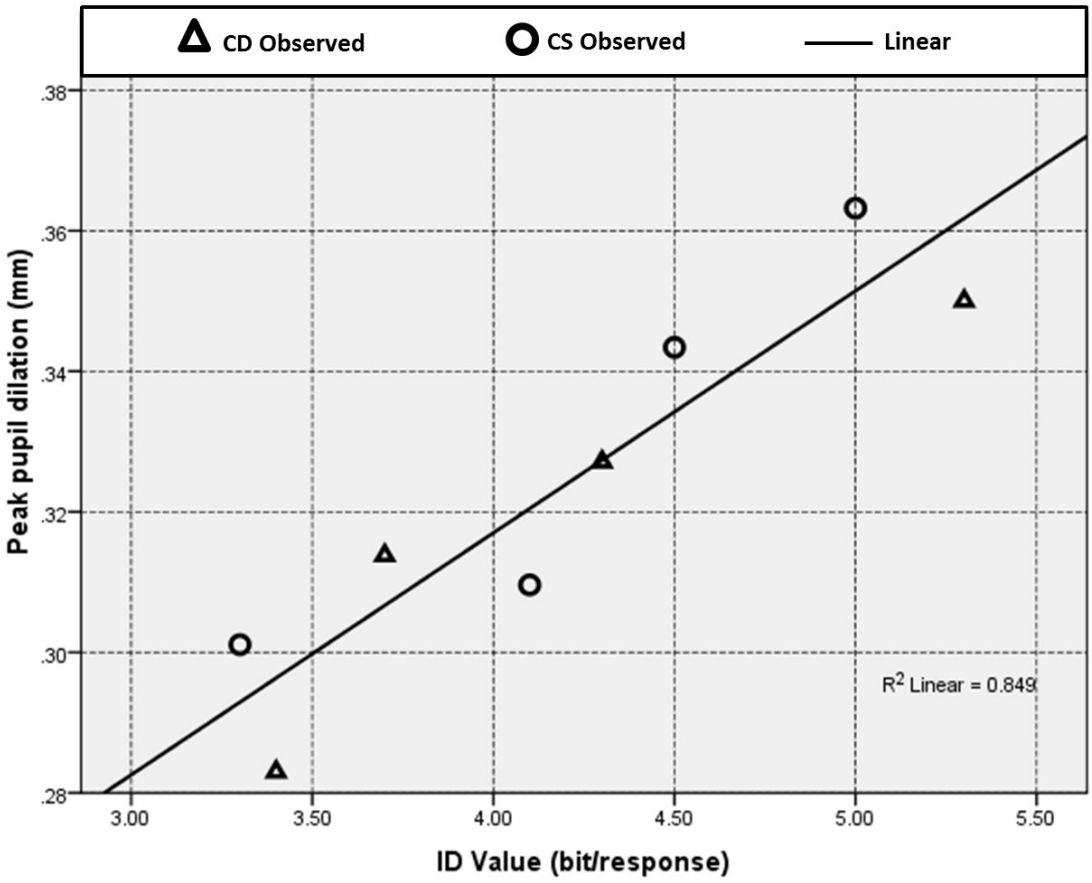


Figure 3.18 Linear regression of mean peak pupil dilation of Each ID to Fitts' ID value ($R^2 = 0.849$ and $p < .005$).

Figure 3.19 shows the linear regression of mean peak pupil dilation duration (from Tool-leave) for each ID to ID value, with $R^2 = 0.662$ and $p < .05$.

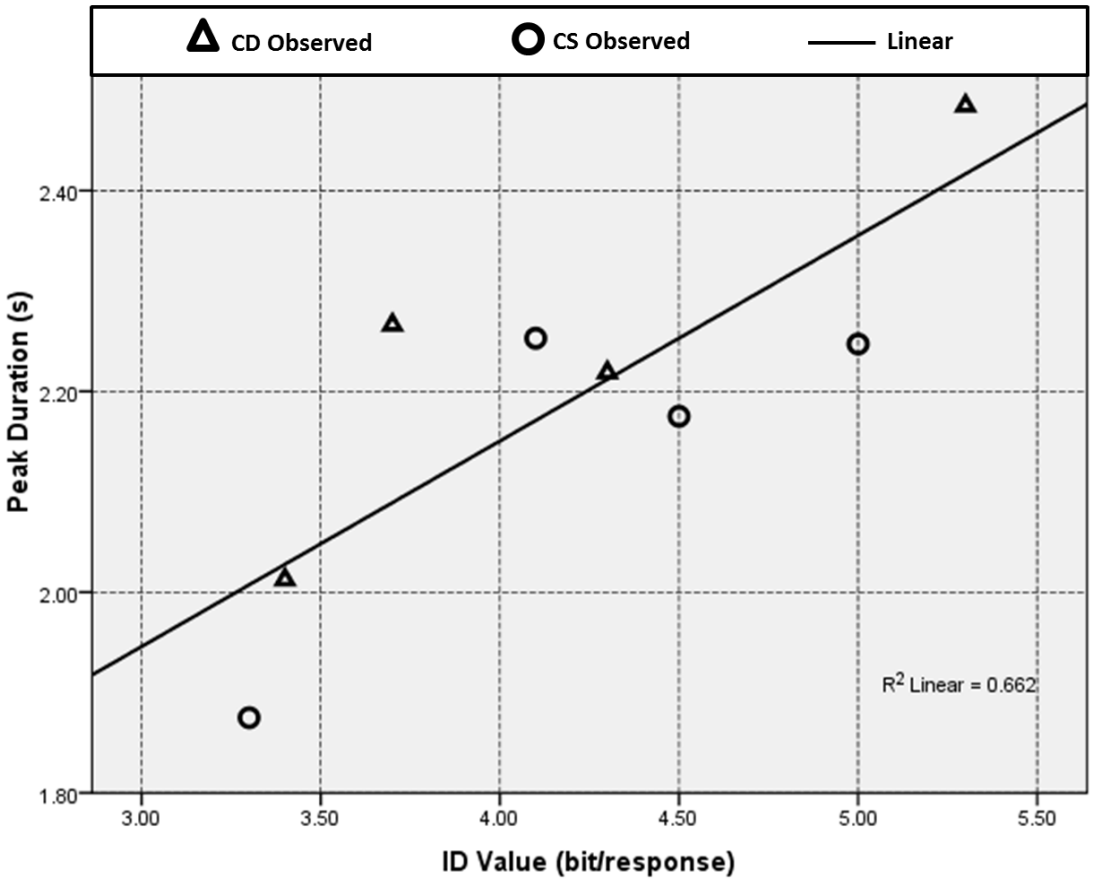


Figure 3.19 Linear regression of mean peak pupil dilation duration (from Tool-leave) of Each ID to Fitts' ID value ($R^2 = 0.662$ and $p < .05$).

3.2.4. Discussion

The hypothesis of this experiment was that the increasing target distance will not dilate the pupil as much as reducing the target size. However, the results 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 3.16*; it starts to dilate at around 1.2s to 1.5s before Tool-leave, following a constriction right after Tool-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 ID of task will be perceived by human operators as a single parameter and responded in the pupil size changes with

the same pattern. Specifically, the pupil reacts uniformly during the preparation and execution of the aiming movement. Our findings are consistent with those in previous work (Moresi, et al., 2011; Privitera, et al., 2010; Richer & Beatty, 1985). The uniform pupil response to the change of target distance and the target size has an important implication—that pupil dilation can be employed as a valid indicator for the overall 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 3.17*. After the pupil reached its peak size, the pupil maintained a larger size in harder IDs than in easier IDs.

Second, after tool moving, as shown in *Figure 3.17B*, 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 3.17A*, the extent of pupil constriction for different IDs during the Transport phase 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 3.17A*, the pupil started to react to tool movement at about 2s before Tool-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 & Beatty, 1985); the smaller target size requires an earlier preparation.

Figure 3.18 shows a strong correlation between peak pupil dilation with task difficulty, and *Figure 3.19* 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.

3.3. Summary

This Chapter verified the correlations between the pupil diameter and the task difficulty in discrete target-pointing movements using two experiments, where the participants were asked to use a surgical tool (grasper) to point and touch the target circles with a 10s waiting time before each movement. The results of Experiment 1 showed a typical pupil response pattern that the pupil dilates 1.5s in preparation before the tool starts to move, followed by a slight constriction, and then peaks in its size before the tool touches the target. The extent of the Valley-to-Peak pupil dilation positively correlates to the task difficulty defined according to Fitts' law. The pupil diameter was further proved to respond to both the variation of target size and to the target distance in the second experiment, where the participants performed two discrete movement tasks, with only target size or distance changing in each task target setting.

Our work is the first in the 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.

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 of motor tasks using pupil parameters.

The results of Experiment 1 were orally presented at the SIGCHI conference 2014, published in the Conference Proceedings of CHI 2014, and included in the ACM

Digital Library (Jiang, Atkins, Tien, Bednarik, et al., 2014)[1]. This paper was selected to be a SIGCHI Best of CHI Honorable Mention Award (identified by the CHI Associated Chairs as being the top 5% of all submissions to CHI 2014).

The results from Experiment 2 were orally presented at the conference of Eye Tracking Research and Applications (ETRA) 2014, published in the Conference Proceedings, and included in ACM Digital Library (Jiang, Atkins, Tien, Zheng, et al., 2014)[2].

1 **X.T. Jiang**, M.S. Atkins, G. Tien, R. Bednarik, and B. Zheng, "Pupil Responses during Discrete Goal-directed Movements", Proc. CHI 2014, 2075-2084 (2014) **(Best of CHI Paper Honorable Mention Award)**.

2 **X.T Jiang**, B. Zheng, G. Tien, R. Bednarik, & M.S. Atkins, "Pupil Dilations during Target-pointing Respect Fitts' Law". In the Proceedings of the Symposium on Eye Tracking Research and Applications, pp 175-182 (2014).

Chapter 4.

Pupil responses in continuous movements (Experiment 3)

The above two studies (Experiment 1 and Experiment 2) found the connection between pupillary response and task difficulty in discrete tasks. The subjects were instructed to wait 10s before the next aiming movement to ensure that the recorded pupil response would not be affected by the previous movement, and the pupil had time to return to its baseline size. However, continuous movement tasks are much more common in the reality of everyday life, such as steering a vehicle, playing ping-pong, and selecting an item in a multiple-level cascade menu. In many cases, the continuous movement frequency is higher than the pupil response frequency which is typically lower than 0.5Hz (Jiang, Atkins, Tien, Bednarik, et al., 2014; Richer & Beatty, 1985), so pupil response is inevitably affected by multiple movements. It is important to confirm whether the pupil size can still serve as an indicator of task difficulty, via developing a method to distinguish if pupil response is in reaction to an upcoming movement or is just a residual effect from a previous movement. We therefore explore the pupil responses to the change of task requirements in a continuous movement such as continuous aiming tasks, with the following research questions in mind. First, is there a difference between the patterns of pupil size responses to discrete and continuous visual-motor tasks? Second, is the change of pupil size still able to distinguish task difficulty in continuous visual-motor tasks?

We conducted the present study using a similar experimental setting as that in the discrete movement study (Jiang, Atkins, Tien, Bednarik, et al., 2014) but here the participants performed a continuous pointing task without any waiting time between movements. We hypothesized that pupil dilation will respect Fitts' law in continuous movements, such that a higher task difficulty evokes a higher peak pupil dilation. If the hypothesis holds, it may be possible to measure the task difficulty in continuous visually-

guided motor tasks. Such measurements can be used for continuously adjusting proactive responses of user interfaces, for example in medical educational simulations involving visual-motor tasks.

4.1. Experiment purpose and hypothesis

In this experiment, we explored pupil responses to the change of task requirements in a continuous aiming movement. This study used a similar experimental setting to that in the discrete movement studies (experiments 1 and 2), but here the participants performed a continuous pointing task without any waiting time between movements.

We hypothesized that pupil dilation will respect Fitts' law in continuous movements, such that a higher task difficulty evokes s higher peak pupil dilation and a longer dilation duration.

4.2. Methods

4.2.1. Participants

Fourteen participants (three females) were recruited to the study, including three graduate students, seven undergraduate students, and four staff members from the University of Alberta. All were right-handed and had normal or corrected-to-normal vision. None were previously trained in surgical procedures.

4.2.2. Experimental setting and apparatus

The task was to move the grasper horizontally inside the training box to point at the horizontal circles printed on the paper, while viewing the scene inside the box which was displayed on a monitor, as shown in *Figure 3.1*. The parameters of the target (the circles) setting, i.e., the target size and the distance between the targets, are the same as those in Experiment 1, shown in *Figure 3.2*. Three difficulty levels were designed in this experiment according to the Fitt's index of difficulty (ID) (Paul M. Fitts, 1954), i.e.

Easy (ID1=2.7 bit/response, W1=0.9cm, A1=3cm), Medium (ID2=4.3 bit/response, W2=0.6cm, A2=6cm), and Hard (ID3=5.9 bit/response, W3=0.3cm, A3=9cm). The equation for calculating ID1, ID2, and ID3 is in Equation (2).

The three difficulty levels of targets were organized in two target settings by their execution sequences. Target setting 1 is shown in *Figure 3.2*, executed from Easy task (bottom circles) to Hard task (top circles). Target setting 2 is similar to target setting 1, except the Hard target was on the bottom, so the task was executed from Hard to Easy.

Since the purpose of this study was to explore the pupil responses in continuous aiming movements, the pointing task was designed to execute consecutively between the pair of circles for a certain number of times (10 times back and forth movements).

The apparatus are the same as that in Experiment 1 but was located in a well-isolated surgical simulation room in the University of Alberta where the ambient lighting was kept relatively stable.

4.2.3. Tasks and procedure

The task was to move the surgical tool to point to and touch the circles printed on a piece of paper placed horizontally at the bottom inside the training box, as for Experiment 1. However, unlike Experiment 1, in this experiment each ID (each step) was executed continuously for 20 movements. A trial consisted of 6 steps executing from bottom to top pairs of targets (phase 1, steps 1-3) and then from top pairs to the bottom ones (phase 2, steps 4-6), as shown in *Figure 4.1*. The trial started by placing the tooltip on the right bottom circle for 20s for the recording of baseline for the whole trial. The tool was then continuously moved to the left bottom circle and back to the right circle repetitively for ten times (20 movements) without any waiting time in-between movements (step (1)). After step 1 the tool was moved to the medium right circle after staying 10s on the bottom right circle, then similarly step 2 and step 3 were performed (*Figure 4.1*, Phase 1). Right after step 3, on the same target setting, step 4 started with a 10s pause on the top right circle, then steps 4, 5 and 6 were completed with a 10s pause between each step (*Figure 4.1*, phase 2). The trial was ended by stopping the

tooltip on the right bottom circle for 10s after step (6). Only the 120 horizontal moves were used for analysis.

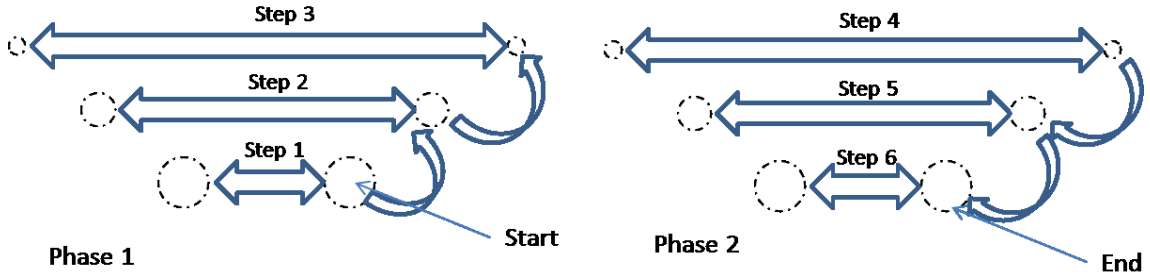


Figure 4.1 The execution sequence of target setting 1, including phase 1 (Step 1 to Step 3) and phase 2 (Step 4 to Step 6).

Note. Each step includes moving the tooltip from right to left and then left to right 10 times. The horizontal arrows represent 20 tool movements from one circle to another within a step, and the vertical arrows represent Transporting to the next step. Before moving to the next step, the tool stops on the right side circle for 10s for baseline recording.

The participants were instructed to move the tool and touch the target as accurately and as fast as possible. Each trial took about 7 minutes.

Each participant read and signed the consent form before entering the study, and then read the instructions. The ethical consent was approved by the Research Ethics office of the University of Alberta. The participants practiced the task for a few minutes, until they were ready to begin. Each participant performed one trial either on task setting 1 (from easy to hard and then from hard to easy as shown in *Figure 4.1*) or task setting 2 (by flipping the paper, same execution sequence as in *Figure 4.1* but starting with the hard targets on the bottom and easy targets on the top). For counterbalance, half of the participants started the target setting 1 first and then performed the task setting 2, and the other half of the participants did the reverse.

4.2.4. Data analysis

Tooltip location

The tooltip positions were automatically extracted from the task videos using the algorithm described in Appendix A and were smoothed with a running-average-filtered using equally weighted four samples window. The moments when the tooltip started to

move (Tool-leave) and reached (Tool-reach) the target, were detected by the algorithm described in section 3.1.2. An example trial of pupil signal synchronized with tooltip positions in timeline is shown in *Figure 4.2*.

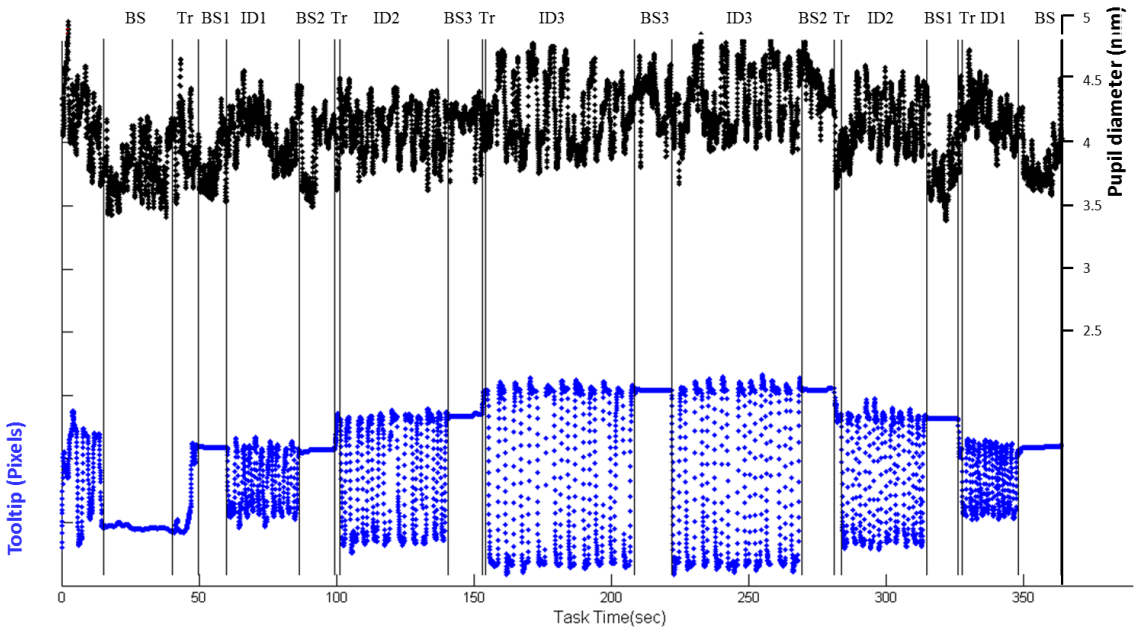


Figure 4.2 An example of tooltip movement and pupil size over time for a complete trial (subject 01, trial 01).

Note. The blue curve is the tooltip position in pixels along the horizontal line between the target circles over time. The black curve is the pupil diameter (mm) synchronized with the tooltip positions in the timeline. The vertical lines represent the phases of procedure, with labels above the lines: BS is the baseline recording, Tr is transferring to a new ID execution, and ID is the period of movements between targets.

Tool movement phases

A movement time for this continuous aiming movement is calculated from when the tool starts to leave, to when next Tool-leave moment occurs, as shown in Figure 4.3. This definition of movement time follows that in Fitts' study (1954), which includes the dwell time (Adam & Paas, 1996) which is from when the tooltip touches the target to the start of the next movement.

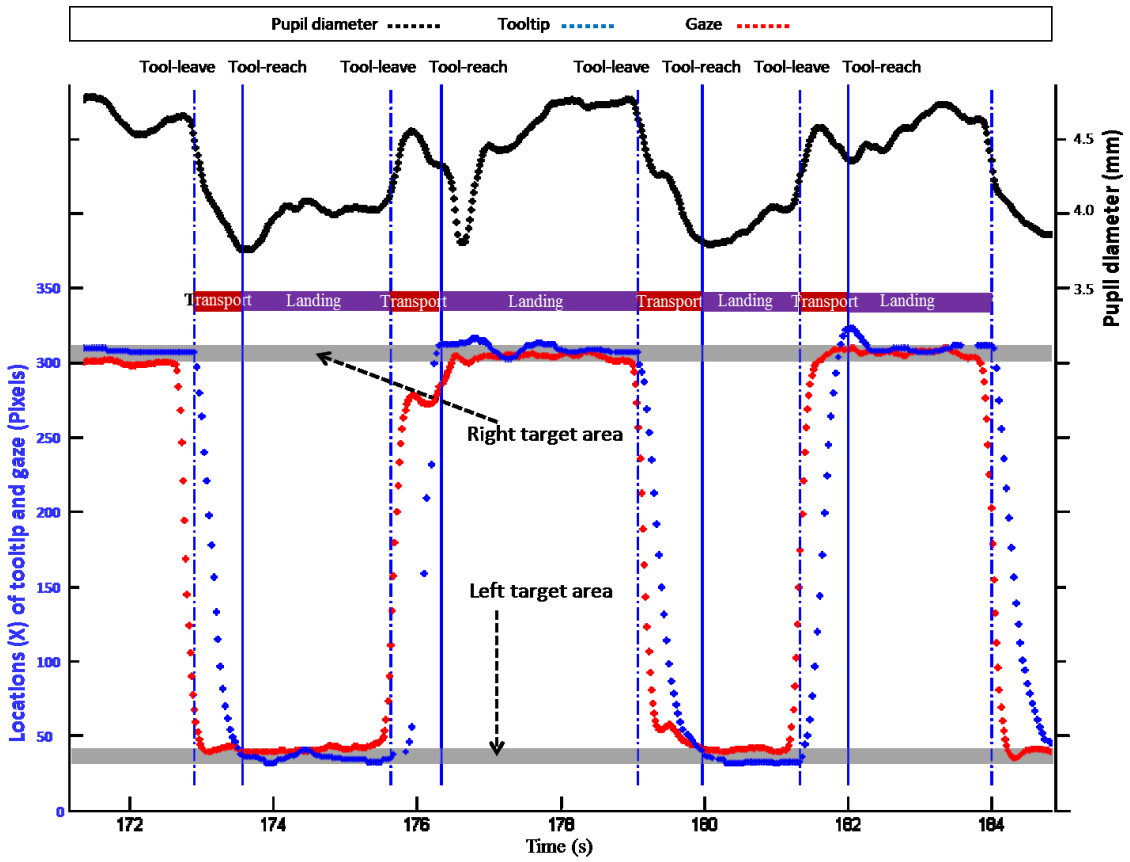


Figure 4.3 A blow-up plot showing four consecutive movements of the first round of hard ID execution (ID3, during 173-182s in Figure 4.2).

Note. The black curve on the top is pupil diameter, the blue curve is the tooltip position (X coordinate) in pixels along the horizontal line between the target circles, and the red curve is the gaze locations (X coordinate). The vertical dashed and solid blue lines represent the moments of start of a movement and the tool reaching the vicinity above the target.

As shown in *Figure 4.3*, a movement is further divided into *Transport* and *Landing* phases (Abrams, et al., 1990; Adam, et al., 2000; Elliott, et al., 2001; Pelz, et al., 2001) to better distinguish the impact on the pupil response from a continuous movement. The Transport phase starts when the tool leaves (Tool-leave) for the target circle and stops when the tooltip arrived at the vicinity above the target circle (Tool-reach). In the Landing phase the tool is slowly adjusted to touch the target circle before the immediate start of the next movement (next Tool-leave). The moments of Tool-leave and Tool-reach are detected using the algorithm described in section 3.1.2.

Movement-related pupil responses

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

Segments of missed pupil data shorter than 100ms were linearly interpolated. These missed data were caused by several reasons such as short blinks and delay in eye-tracking recovery. Then a Butterworth low-pass filter with a cut-off 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).

A signal averaging technique was employed to capture the movement-related pupil responses, as described in section 3.1.2. In signal averaging, properly choosing the time point for signal alignment and the size of window is critical for the detection of movement-related pupil changes. In our previous study on discrete tasks (Experiment 1 and 2), the repetitive 7s-windows of pupil responses were aligned at the moment of Tool-leave (the tart of Transport phase) to preserve the pupil dilation 1-2s before the movement which is considered to be a reaction to the preparation of a movement. However, in the present continuous movement study, the preparation period for a movement is overlaid with the Landing phase. Therefore, we decided to align the task repetitions at the moment of Tool-reach (the end of Transport phase) to preserve the pupil diameter changes in both Transport and Landing phases. Also, a shorter time window i.e. 4s-window was employed to extract and show pupil diameter data in this continuous movement study, since the longest movement time of the hardest task was 3.2s.

In our previous studies (Experiment 1 and 2), the baseline was chosen from the period at 2s before Tool-leave, since this period was the end of the 10s waiting time and occurred right before the pupil started to dilate for the upcoming movement. In the present continuous aiming task, since there was no waiting time between movements, we instead considered the 400ms period at the beginning of the Landing phase, when

the eyes start to focus on the new target and the pupil starts to dilate. Therefore, we used the 400ms period around the Tool-reach moment for the baseline pupil diameter, which is the beginning of the Landing phase.

Relative pupil diameter changes in the 4s-window were derived by subtracting each sample from the baseline pupil diameter which was the mean of the pupil size during the above determined baseline period. All the data in the windows were aligned at the moments of Tool-reach (2 seconds into the 4s window), and the mean pupil diameter change was calculated for each time point in the window across all horizontal tooltip movements from all trials from all subjects. Similarly, the mean pupil diameter changes were calculated across all moves from all trials for each ID. The mean pupil diameter changes in the 4-second window were drawn in a graph for visual analysis.

To examine which parts of the pupil size changes in the 4-second window have significant differences between the three IDs, the moment-to-moment based significance testing was applied in the 4s-window as in Experiment 1 and 2.

Experimental design

The experiment had an independent variable task difficulty with three levels of Easy, Medium, and Hard. The dependent variables were movement time, peak-to-valley pupil constriction, peak-to-valley pupil constriction duration, Valley-to-Peak pupil dilation, and Valley-to-Peak pupil dilation duration.

The total number of movements was 1680 (14 participants, each performing 6 trials, each trial has 6 ID executions (each ID was executed twice), with each ID execution having 20 moves—10 times from right to left and from left to right, for a total of 120 movements for each subject).

4.3. Results

4.3.1. Accuracy

From the 1680 movements recorded by the 14 participants, we excluded the first 2 moves and last 2 moves from each ID execution (96 moves remained for each

subject), and also discarded a total of 19 movements due to mis-operation, for example when the subject moved the pointing tool to a wrong target. Therefore we obtained 1325 correctly performed movement recordings.

Among the 1325 movements, a total of 73 movements' endpoints (the tooltip positions when touching the target) were outside the target circle (63 for Hard ID, 10 for Medium ID, and none for Easy ID). *Figure 4.4* shows how inaccurate these 73 movements were. Most movements' (46 movements) end points are within 10 pixels to their corresponding target edges in the recorded videos frame (corresponding to 2.5mm on the target paper which is 10 mm displayed on the 17" screen). The 27 movements with their endpoint distance greater than 2.5mm to their target circle edges were discarded. The remaining 1298 valid movements were further analyzed. Usually, all error end points should be included or excluded in the analysis, but as the error rate in this study is less than 4% and we did not compare between conditions, these partly-erroneous included error end points would not affect the results.

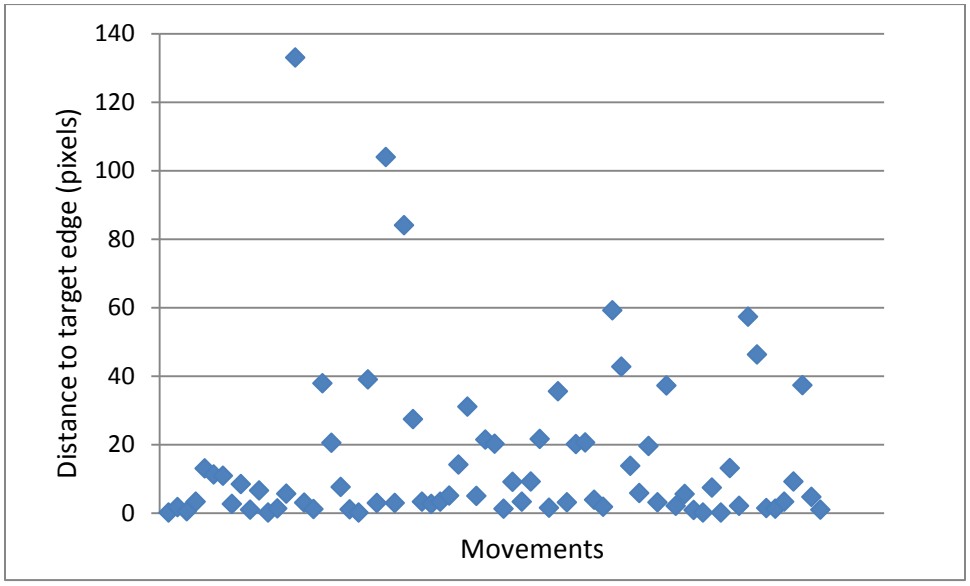


Figure 4.4 Scatter plot of the distribution of errors for the 73 inaccurate movements.

4.3.2. Movement time

The movement time (MT) for a complete movement starts from when the tool leaves the current target, and ends when the tool leaves that target for the next

movement (the next Tool-leave), such that the MT is the interval between two consecutive Tool-leave moments. The mean MT for all movements is 2.0±0.8s, and the MT increases when the difficulty level increases (*Figure 4.5*). There is a significant main effect of ID in terms of mean MT ($F_{2,1295}$ = 412.258, p < .0001). Post Hoc test (Tukey HSD) shows significant differences (p < .0001) between pairs of the three IDs, with mean MT of Easy, Medium, and Hard IDs being 1.3±0.5s, 1.9±0.6s, and 2.6±0.8s respectively.

To distinguish the impact on the pupil response from a continuous movement, we divided the entire movement into *Transport* and *Landing* phases (Abrams, et al., 1990; Adam, et al., 2000; Elliott, et al., 2001; Pelz, et al., 2001). The phase separation was defined by the kinematics of the tool. The *Transport* phase starts from the moment of Tool-leave to Tool-reach (when the tool quickly moves to the vicinity above the target), and the *Landing* phase occurs between the Tool-reach and the next Tool-leave (when the tool descends and touches the target).

Figure 4.5 shows the mean MT over all subjects for the Transport and Landing phases, and for the complete movement. The mean MT of both Transport and Landing phases are significantly different (p < .0001) between the three IDs.

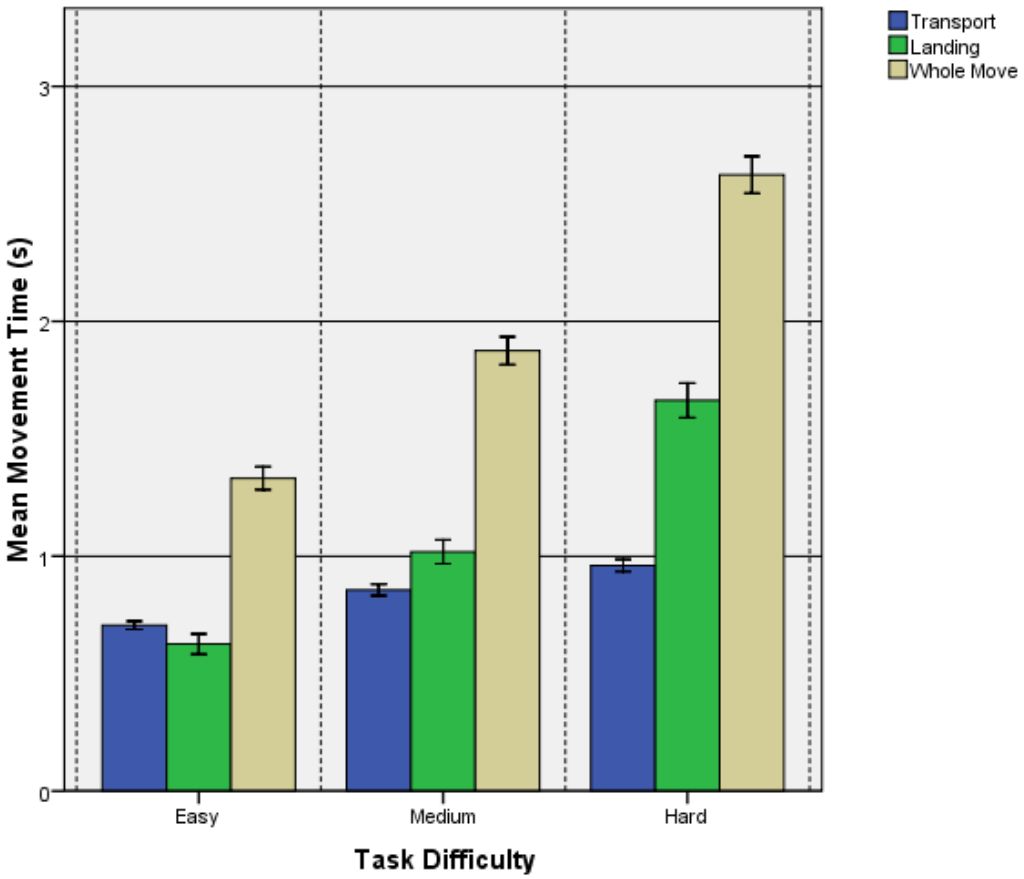


Figure 4.5 Mean time of movement phases and complete movement time for different IDs over all subjects.

Note. The blue bar is the movement time for the Transport phase, the green bar is the movement time for the Landing phase, and the mustard bar is the movement time for a complete move. The error bars are for 95% confidence intervals.

4.3.3. Pupil responses to continuous tool movements

Figure 4.6 shows the mean changes of pupil diameter averaged for all IDs for all subjects. Data are aligned over a 4-second window starting 2 seconds before the Tool-reach. The mean pupil diameter change is shown from the baseline over time. The baseline pupil diameter is the mean pupil diameter over a 400ms period at the Tool-reach moment in each window, as shown in the orange rectangle. The vertical red line indicates the moment of Tool-reach where all the data are aligned. The red and purple bars represent Transport and Landing phases respectively. The pupil starts to constrict about 150ms into the Transport phase and dilates in the Landing phase with an

approximate 100ms delay; this is the typical pattern of pupil size changes found in a single aiming movement found in this study.

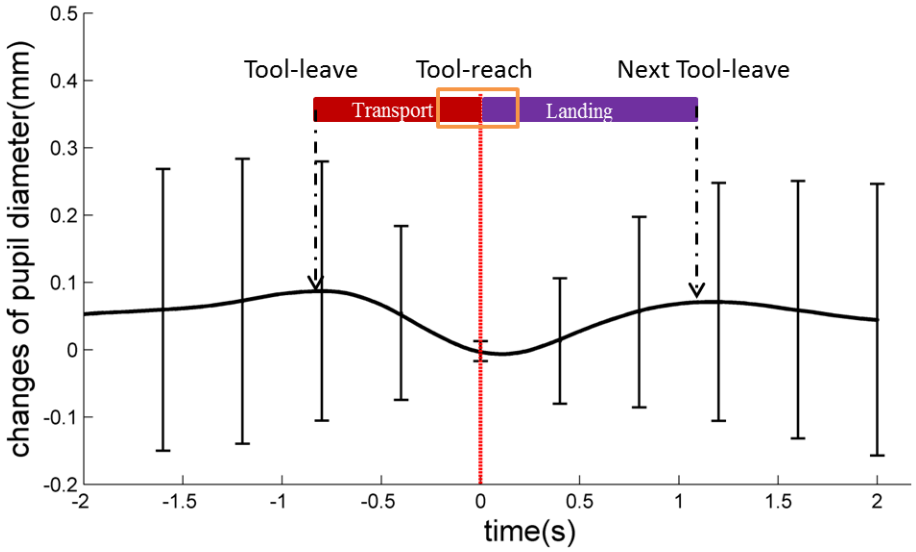


Figure 4.6 Mean pupil diameter changes for 1298 valid movements from all IDs from all 14 subjects.

Note. Data were aligned over a 4-second window starting 2 seconds before the Tool-reach, shown by the vertical red line. The solid black curve is the mean pupil diameter change from the baseline over time. The pupil baseline diameter is the mean pupil diameter over a 400ms period at the Tool-reach moment, as shown in the orange rectangle. The error bars for 1 std. dev. are drawn every 400ms.

Figure 4.7 shows the mean changes of pupil diameter in a 4-second window aligned at the tool reach moment, for each of the three IDs. The curves of the pupil response to the three IDs are clearly separated in both Transport and Landing phases; the hard ID constricts from a higher peak pupil size to the baseline in the Transport phase and dilates from the baseline to a higher peak pupil size in the Landing phase.

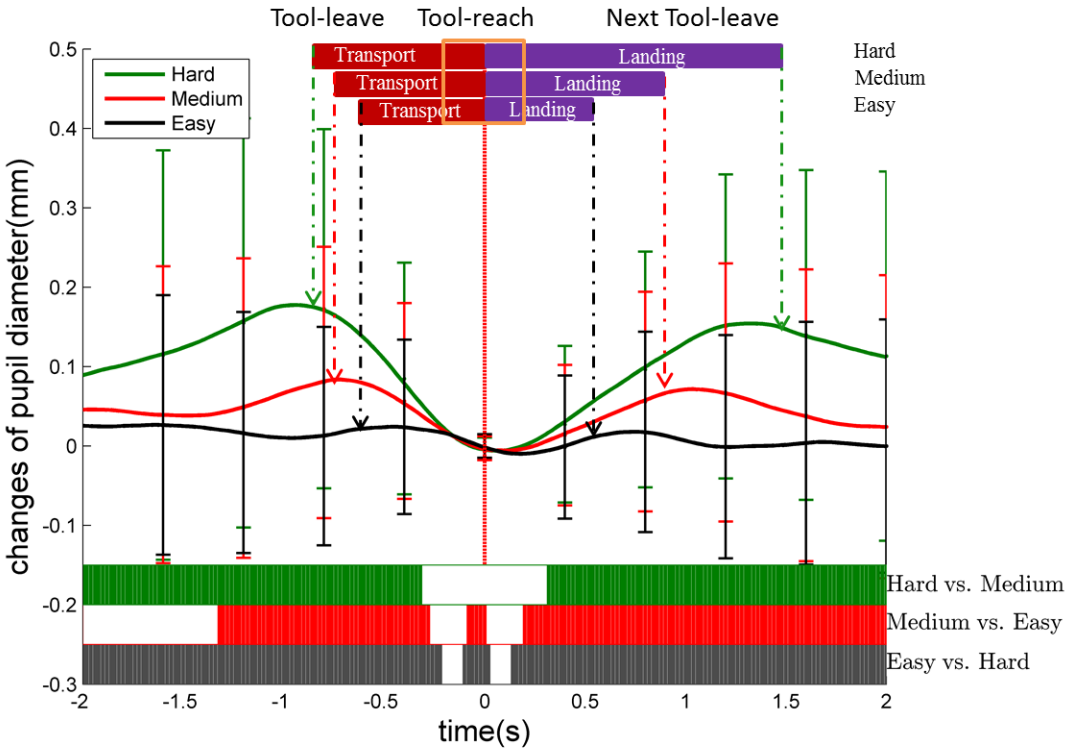


Figure 4.7 Mean pupil diameter changes (subtracted by baseline) for 3 different IDs from all subjects.

Note. Data are aligned over a 4-second window 2 seconds before Tool-reach, shown by the vertical red line. The three colored curves—black, red, and green—are the mean pupil diameter changes for Easy, Medium, and Hard ID respectively. The pupil baseline diameter is the mean pupil diameter over a 400ms period at the Tool-reach moment, as shown in the orange rectangle. The three rows of bars at the top represent Transport (red) and Landing (purple) phases for the three IDs Hard, Medium, and Easy, respectively. The three colors of bars at the bottom indicate significant differences in pupil dilation between Easy, Medium and Hard ID with black representing Easy vs. Hard, red representing Easy vs. Medium, and green representing Medium vs. Hard. The error bars for 1 std. dev. are drawn every 400ms.

Table 4.1 shows the extent and duration of the pupil diameter changes between the three IDs in Transport and Landing phases. ANOVA results show that there are significant main effects in terms of Peak-to-Valley pupil constriction ($F_{2,1295}$ = 157.463, p < .0001) and Peak-to-Valley duration ($F_{2,1295}$ = 144.223, p < .0001) in the Transport phase and Valley-to-Peak pupil dilation ($F_{2,1295}$ = 157.052, p < .0001) and Valley-to-Peak duration ($F_{2,1295}$ = 408.863, p < .0001) in the Landing phase between the three IDs. The values of the above four pupil diameter parameters (Peak-to-Valley pupil constriction,

Peak-to-Valley duration, Valley-to-Peak pupil dilation, and Valley-to-Peak duration) positively correlate with IDs ($p < .0001$).

Table 4.1 The mean pupil diameter changes and duration over all subjects for Transport and Landing phases, for different IDs

	Easy	Medium	Hard
Transport phase			
Peak-to-Valley Constriction (mm)	0.09±0.10mm	0.16±0.15mm	0.28±0.21mm
Peak-to-Valley Duration (seconds)	0.52±0.33s	0.68±0.35s	0.97±0.48s
Landing phase			
Valley-to-Peak Dilation (mm)	0.08±0.09mm	0.15±0.14mm	0.26±0.20mm
Valley-to-Peak Duration (seconds)	0.53±0.35s	0.91±0.46s	1.44±0.58s

A paired *t*-test was performed for each time point in each pair of IDs containing samples from the 14 subjects, with the results as shown in the bottom colored bars in *Figure 4.7*. The colored areas of the bars at the bottom of *Figure 4.7* indicate significant differences ($p < .05$), with the black comparing Easy and Hard, the red comparing Easy and Medium, and the green comparing Medium and Hard. Almost all moments along the 4s window are significantly different between pairs of the three IDs except in the area of the Tool-reach moment (baseline).

4.4. Discussion

The results support our hypothesis, that the pupil responses to a movement in this continuous aiming task respect Fitts' law. Our data revealed a pattern of pupil response to a single goal-directed movement in the continuous aiming task, where the pupil constricts during the Transport phase and dilates during the Landing phase (*Figure 4.6*); this holds for different difficulty levels of movements, where harder task IDs elicit a higher magnitude of pupil response and longer pupil dilation/constriction duration (*Figure 4.7*). Both ANOVA on the pupil diameter features (Peak-to-Valley pupil constriction and duration in Transport phase and Valley-to-Peak pupil dilation and duration in Landing

phase), and moment-to-moment based *t*-test analysis on the changes of pupil size during a movement (the horizontal color bars on the bottom of *Figure 4.7*) show significant differences between the three IDs. The pupil size of Easy ID did not show a significant peak (*Figure 4.7*). We believe the reason to be that in the continuous aiming task, the frequency of tool movement for Easy ID is higher (0.7Hz) than that of pupil's response (typically lower than 0.5Hz) (Jiang, Atkins, Tien, Bednarik, et al., 2014; Richer & Beatty, 1985).

New challenges emerged during the analysis of pupil data in continuous movement tasks due to the interference between two consecutive movements without waiting time. The challenges included how to choose the moment for the windows to align for the pupil signal averaging, and how to determine the period for baseline. The end of the Transport phase was found to be a good timestamp for aligning windows, and the short period around this moment was suitable for the pupil baseline in this continuous movement task, where the pupil starts to dilate in response to the Landing phase.

Our solution of aligning tasks at the end of tool transportation resolved the above-mentioned two challenges and also found new results which were not identified in the prior discrete aiming movement studies (Jiang, Atkins, Tien, Bednarik, et al., 2014; Jiang, Atkins, Tien, Zheng, et al., 2014). First, two peak pupil dilations occurred, at around both the start of Transport and at the end of Landing phases (*Figure 4.6*), compared to only one peak at 2.0s after the tool started to move in the previous discrete movement studies. This is because, in the continuous aiming movement, the 1-2s period before Tool-leave is also a Landing phase for the previous target, and pupil dilation is a combined response to the preparation for the Tool-leave and the execution of touching the target.

Second, pupil response around the Tool-leave differs most significantly between the three IDs (*Figure 4.7*), where the end of the movement is followed immediately by a new movement. The effect of the pupil dilation after the end of movement is overlapped by the new movement which yields a significant pupil response. In contrast, in the discrete task (Experiment 1), the most significantly distinguishable phase is around the

end of each movement, since the pupil diameter develops to its peak followed by the 10s waiting time.

Therefore, by aligning the windows at the end of Transport phase and subtracting the baseline, the pupil response patterns are clearly shown for the Transport and Landing phases. Once the end of tool transportation is detected, the pupil diameter features including Peak-to-Valley constriction and duration and Valley-to-Peak pupil dilation and duration can be employed to distinguish the task difficulty of the movements.

The lighting condition and screen illumination of this experiment were well controlled. The ambient lighting in the case study room (a windowless room) was kept constant throughout the entire experiment data collection. The luminance of monitor that the subjects looked at during performing the task was kept constant as well. The background of the screen was uniformly grey. Furthermore, the baseline subtraction performed for each window and the averaging of the windows would cancel most of the possible unrelated pupil responses.

Besides lighting conditions, one may be concerned that the pupil dilation or constriction are affected by the viewing angle to the eye-tracker camera due to the changes of gaze location (Brisson, et al., 2013; Gagl, et al., 2011) during target-pointing. In our case, the tool quickly moves from a target circle to another and the eyes mostly fixate on the targets, as shown in the red curve in *Figure 4.3*. To clarify the gaze angle problem, we only need to check whether the baseline pupil diameter of movements in two different directions (moving from right to left vs. from left to right) are significantly different. In our two discrete studies (Experiment 1 and 2), there is no significant difference between the baselines of left and right movements. Therefore, we are sure that gaze angle does not affect the measured pupil size in this study.

Similarly, concerns may also be raised that the pupil response could be affected by the variation of illumination for different sizes of the dashed black target circles (e.g. Easy ID has the biggest circles) (Kun, et al., 2012). However, there was no significant effect of the target size (task difficulty) on the pupil baseline ($F_{2,1295} = 2.476$, $p > 0.05$).

There are some threats to the validity of employing the pupil diameter to measure task workload in continuous movements. First the pupil response may be much slower

than the continuous tool movement time, so that the pupil response pattern could not reflect the task workload changes. Second, the learning effect may affect the accuracy of the pupil diameter measurement to task workloads. That is, after many repeated tool movements for the same ID, the participants may get used to the task and the task workload would decrease. Third, individual variations also should be considered as a violation to the validity of the pupil measurement. The pupil diameter, as a bio-signal, differs between individuals, by varying in its profile including amplitude, rate of change, and response latency to stimuli.

4.5. Summary

The previous discrete movement studies showed that the pupil dilation positively correlates to task requirements modeled by Fitts' index of difficulty. The present study shows that pupil responses still respect Fitts' law in continuous aiming tasks, which are more complex and common than discrete aiming movements. This implies that pupil size changes can provide indicators for measuring task requirements in goal-directed movements.

The detailed results have been written in a manuscript under review (Jiang, Zheng, et al., 2014)[3].

[3] **X.T. Jiang**, B. Zheng, R. Bednarik, and M.S. Atkins, "Pupil Responses to Continuous Aiming Movements", Under review (2014).

Chapter 5.

Evaluation of pupil responses in simulated surgical tasks

In Experiment 3 (Chapter 4), we explored the changes of pupil diameter in response to the task difficulty in continuous target-pointing movements. The pupil response was affected by consecutive movements in continuous motor tasks, by aligning and averaging the movements at the end of tool transportation, we discovered a typical pattern of pupil responses to continuous aiming movement that the pupil size drops in the Transport phase and dilates in the Landing phase, and the amplitude of the pupil constriction (in Transport phase) and dilation (in Landing phase) positively correlate to the task difficulty, which supported Fitts' law.

However, the above findings were obtained from a relatively simple and well-controlled continuous task, where the movement only involves a simple pointing and touching action with the task difficulty clearly defined by the target size and movement distance. In a real-life motor task, the action may be more complex rather than only pointing and touching a target, for example reaching and grasping an object or transporting an object to another place. Furthermore, the object of the target location could be an irregular size and Fitts' index of difficulty is not readily applicable. In these cases of real-life motor tasks, would the changes of pupil diameter be a valid indicator of task difficulty? With this question in mind, we decided to explore the pupil responses in real-life motor tasks to determine whether the pupil diameter can be employed as an indicator of task difficulty.

Our research team had previously conducted a study of a simulated surgical task, the peg transportation task performed in a surgical training box, for the purpose of exploring differences in gaze behaviors for different difficulty subtasks. However, there

was no analysis of the corresponding tool position, and no analysis of pupil responses to the task difficulty of the subtasks.

The task was to transport a peg between three dishes, and was divided into 3 different subtasks: picking up a peg, transporting it to another location and releasing the peg into a container, and touching a home location with the tip of the grasper. These 3 subtasks have different difficulties.

5.1. Study description

5.1.1. Participants

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

5.1.2. Experimental setting and apparatus

The simulated training environment for laparoscopic surgery was set up inside the Medical Image Analysis Lab at SFU. The training system including a laparoscopic training box (Laparoscopic Trainer, 3-D Technical Services, Franklin, Ohio) and a laparoscopic grasper (Ethicon Endo-Surgery, Cincinnati, Ohio), as shown in *Figure 5.1*. On the bottom of this training box, a custom-made wood plate was placed as the simulated surgical site. The scene of the inside of the training box was captured by a webcam and displayed on to the monitor of Tobii 1750 eye-tracker (Tobii 1750, Tobii Technology, Danderyd, Sweden). The web camera recorded facial expressions of the operator for validation purpose.

Figure 5.1 The experimental setting includes 3 components: Tobii eye-tracker, laparoscopic training box, and a webcam.

Note. The screen shows the display scene projected from inside of the training box, consisting of the peg board with 3 colored dishes of 13 mm diameter, a white homing square (12 mm in side length), and a green peg of 4.5 mm diameter.

5.1.3. Task and procedure

The task was to transport the rubber peg (a 4.5 mm × 10.5 mm green cylindrical peg) between three dishes using the grasper—from the red dish on the top to the green dish on the left, then to the blue dish on the right, and back to the red dish, touching the center white square (home plate) with grasper tip every time after the object was placed into a dish.

A trial has 9 subtasks as shown in *Figure 5.2*, which can be grouped into three basic categories with different task 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). The peg was initially placed in the top dish. Taking the first three subtasks as an example, starting from the home position, the

grasper had to be moved to the top dish and the peg then had to be picked up out of the dish (RG). The peg then had to be transported from the top to the left bottom dish using the grasper, which was then opened to release the peg into the left bottom dish (TR). After releasing the peg, the empty grasper had to be moved to the home position (H). These 3 subtasks were repeated three times until the peg was brought back to the original dish and the grasper back to the home position, as shown in *Figure 5.2*.

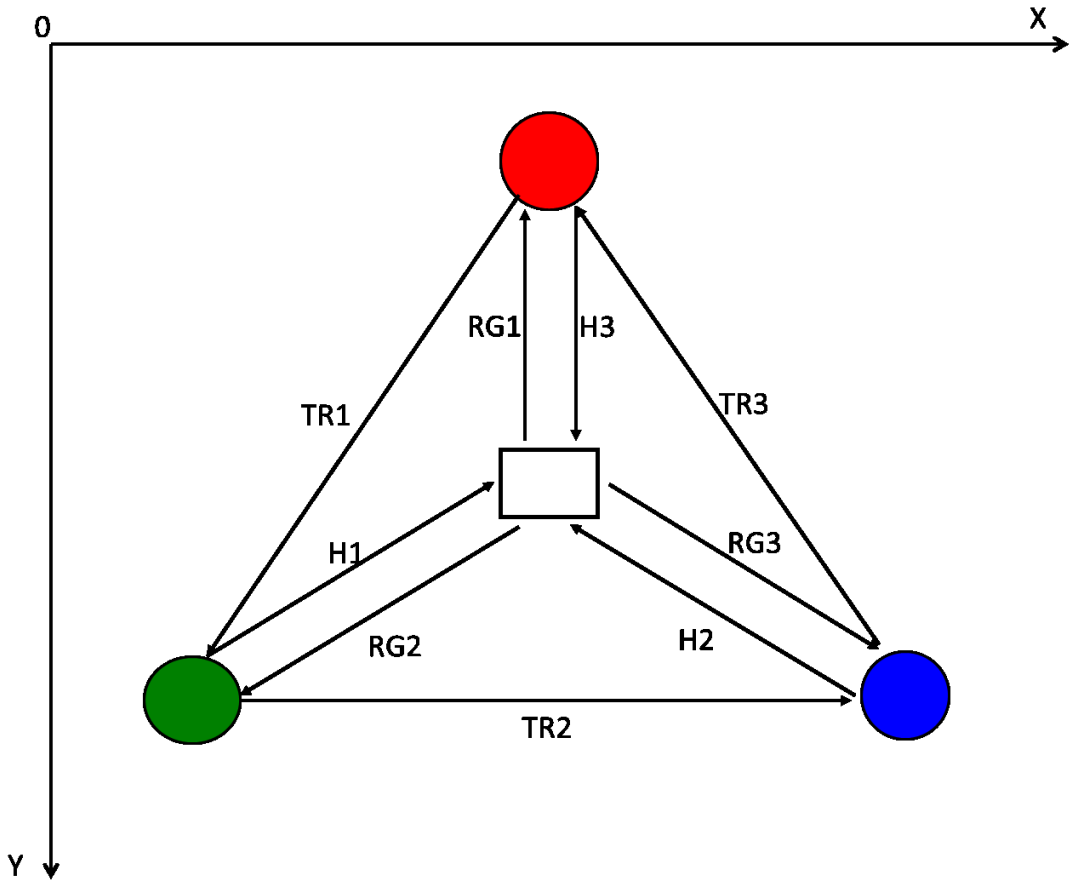


Figure 5.2 Illustration of the task and subtasks.

Note. In each group of subtasks, the participant moves the surgical grasper from the home plate (the rectangle in the centre) to grasp a peg (small green cylinder, 4.5mm × 10.5mm) in a dish (reaching and grasping, RG); transport and release the peg to the next dish (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 dish, then back to the red starting dish. The origin of the image of the target setting is at the left top corner and horizontal is x-axis and vertical is y-axis. The tool positions of subtasks RG1 and H3 change in y-axis, the tool positions of subtask TR2 change in x-axis, and the tool positions of the rest subtasks change in both x and y axis.

Each participant was given a brief oral description of the task and practiced a few minutes 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 lighting and data recording condition were maintained constant throughout entire study.

5.2. Data analysis

5.2.1. Estimation of difficulty of subtasks

The task requirements of the three subtasks cannot be simply defined according to the target size and distance using Fitts' index of difficulty (ID), since there is an extra action (grasping or releasing) added to the end of the tool movement in the RG and TR tasks, instead of just touching the target. Table 5.1 shows the estimated ID of the three subtasks according to their target size and distance using Fitts' ID according to equation (2).

Table 5.1. The estimated task difficulty of the three subtasks using Fitts' index of difficulty (ID).

Subtask	ID (bit/response)	Size (mm)	Distance (mm)
Reaching and Grasping	4.0	4.5	37
Transport and Releasing	3.3	13	65
Homing	2.6	12	37

The TR task requires the longest distance to move (see Table 5.1), and the Homing and RG tasks the same, shorter, distance, so based on distance alone we would expect the Transport phase time of TR to be the greatest, and the RG and H Transport times to be equivalent. However, the RG may be the hardest task, as the RG task requires picking up a small peg, whereas the TR task requires releasing the peg into a much larger dish (see Table 5.1 for actual size details). Using Fitts' Law naively, we have estimated the difficulty of the three subtasks, shown in Table 5.1. Here, the RG task clearly is the most difficult, and the Homing task is the easiest with only a Transport

phase movement to home position and a small extra action to touch the Home plate (ID=2.6bits/response).

5.2.2. Tooltip location

Surgical videos were captured from the scene camera attached to the training box while the participants were performing the task, at 30Hz with resolution of 352×288 pixels. The tooltip positions were automatically extracted from the surgical videos using the algorithm described in Appendix B and were smoothed with a running-average-filtered using equally weighted four samples window. Since the scene of the task videos from this study is more complex than that of the videos from Fitts' pointing study, an algorithm based on background subtraction was applied before searching for the tool object (Appendix B).

5.2.3. Subtask separation and movement phases

The subtasks needed to be separated sequentially for the analysis of pupil responses to different subtasks. Two important tool movement moments for the separation of subtasks and movement phases are when the tool starts to leave the current target to next target (Tool-leave) and when the tool reaches the vicinity above the target (Tool-reach) detected by the following algorithm.

The algorithm of searching the moments of Tool-leave and Tool-reach is similar to that in section 3.1.2 but according to the velocity of tool movement in different directions (x-axis, y-axis, or both directions) of the subtask; first the peak velocity of tool movement in the subtask's specified coordinate is found and then the Tool-leave and Tool-reach can be determined by searching backwards and forward respectively to when the tool velocity is lower than a threshold.

As shown in *Figure 5.3*, the detectable changes of the tool locations in coordinates differ between subtasks: the tool location of subtasks RG1, TR1, and H3 change significantly in the y-axis, subtasks H1, RG2, and TR2 change in the x-axis, subtasks H2, RG3, and TR3 change in both x and y-axis.

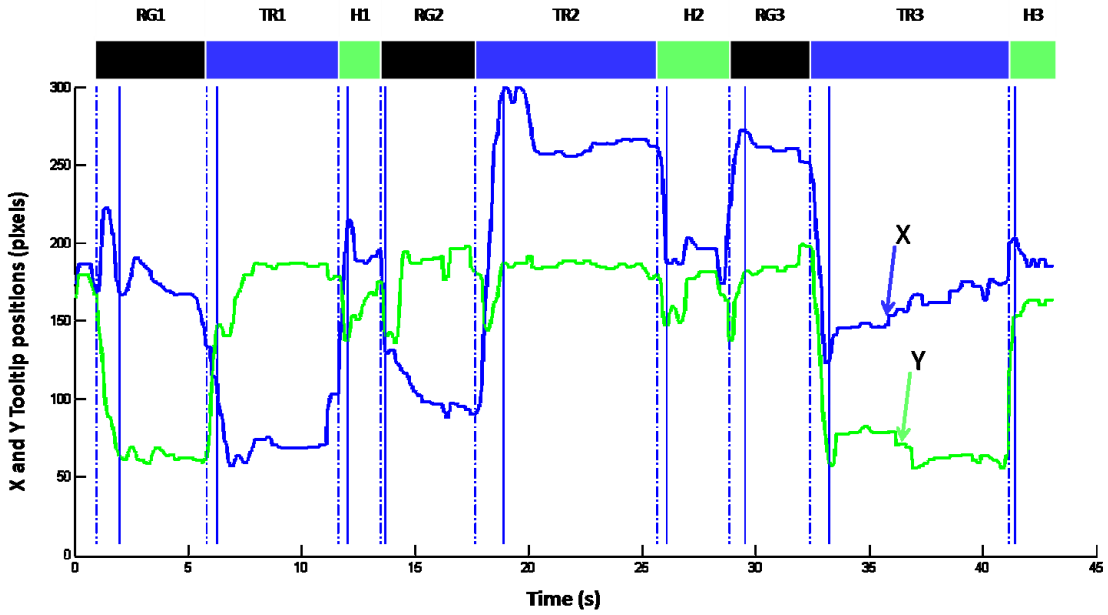


Figure 5.3 An example (from the 3rd trial of subject 1) of tooltip positions (X and Y) curves over time and the subtasks.

Note. The blue and green curves are the tooltip positions in X and Y coordinates, respectively. The dashed and solid vertical blue lines represent the moments of tool starts to leave the current target and the tool reaches the vicinity above the next target, respectively. RG1, TR1, and H3 change significantly in Y coordinates, H1, RG2, and TR2 in X coordinates, and H2, RG2, and TR3 in both X and Y coordinates.

The subtasks are then separated from the continuous tool moving by the moments of Tool-leave. As shown in *Figure 5.2*, the RG subtask starts from the tool leaving home position (Tool-leave) and lasts until when the tool has successfully grasped the peg and starts to leave the dish to Transport the peg to the next dish (Next Tool-leave). The TR subtask starts from the end of RG (defined above as when the grasper starts accelerating away from the dish) and lasts until the tool has released the peg into the dish and leaves for the home position. The homing subtask starts from the end of TR (defined above as when the grasper starts accelerating away from the dish) and ends at the start of RG (defined above as when the tool leaves the homing position).

To distinguish the impact on the pupil response from a continuous task execution, we divided a subtask into Transport and Landing phases (Abrams, et al., 1990; Adam, et al., 2000; Bootsma, Marteniuk, MacKenzie, & Zaal, 1994; Elliott, et al., 2001; Girgenrath, Bock, & Jüngling, 2004; Pelz, et al., 2001). The phase separation was defined according to the kinematics of the tool. The Transport phase starts from the

moment when the tool starts to leave (Tool-leave) to when the tool reaches the vicinity above the target (Tool-reach), and the Landing phase occurs between the Tool-reach and when the tool leaves for the next subtask (next Tool-leave). The action of the Landing phase differs for each type of subtask: grasping the peg in RG subtasks, releasing the peg into the dish in TR subtasks, and resting at the home position in H subtasks.

5.2.4. Movement-related pupil responses

Pupil diameter data was continuously recorded by a Tobii 17/50 eye-tracker at 50Hz, except for the moment of eye blinks and mis-capturing due to the large head movement. Segments of missed pupil data shorter than 100ms were linearly interpolated and a Butterworth low-pass filter with a cut-off 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).

Since the peg transportation task is a continuous movement task, the pupil response to an individual subtask has to be separated according to the timestamp when tool starts to move. Similar to the continuous aiming task, the time at the end of the Transport phase (the Tool-reach moment) was chosen as the alignment timestamp to preserve the pupil responses in both Transport and Landing phases, and the baseline pupil diameter was derived from the 400ms period at the end of Transport phase, when the eyes start to focus on the new target and the pupil starts to dilate.

A 8s-window with 2s before Tool-reach and 6s afterwards was employed to extract pupillary responses for each subtask, since the mean Transport phase time is less than 1 second and the Landing phase is around 2-6 seconds. Relative pupil diameter changes in the 8s-window were derived by subtracting each sample from the baseline pupil diameter. All the data in the windows were aligned at the moments of Tool-reach (2 seconds into the 8s window), and the mean pupil diameter change was calculated for each time point in the window across all horizontal tooltip moves from all trials from all subjects. Similarly, the mean pupil diameter changes were calculated across all moves from all trials for each subtask. The mean pupil diameter changes in the 8-second window were drawn in a graph for visual analysis.

The pupil dilation or constriction may be affected by the viewing angle to the eye-tracker camera due to the changes of gaze location (Brisson, et al., 2013; Gagl, et al., 2011). *Figure 5.4* shows an example frame of the task video where the physical targets were projected. Since the scene camera was not perpendicular to the plan of the peg board when recording, the red dish was at a relatively higher y-axis position in the image with a wider viewing angle to the eye-tracker camera. That could cause a significant pupil distortion, whereas the home plate, green and blue dishes were at a similar horizontal level.

Figure 5.4 An example frame of the task video showing the target setting.

Figure 5.5 shows the mean baseline pupil diameter of each subtask of all trials from all subjects. The baseline pupil sizes of RG1 and TR3 are the smallest among all the subtasks due to their wider viewing angle to the eye-tracker camera as discussed above. RG1 is far smaller than TR3 because it is the first subtask of a trial and the pupil

starts to dilate from a relatively small size. Therefore, we excluded the first subtask (RG1) and the last transport and release subtask (TR3). The subtask (H3) was also excluded as it was the last action, with no following actions. As a result, we retained data from TR1, H1, RG2, TR2, H2, and RG3 for further analysis.

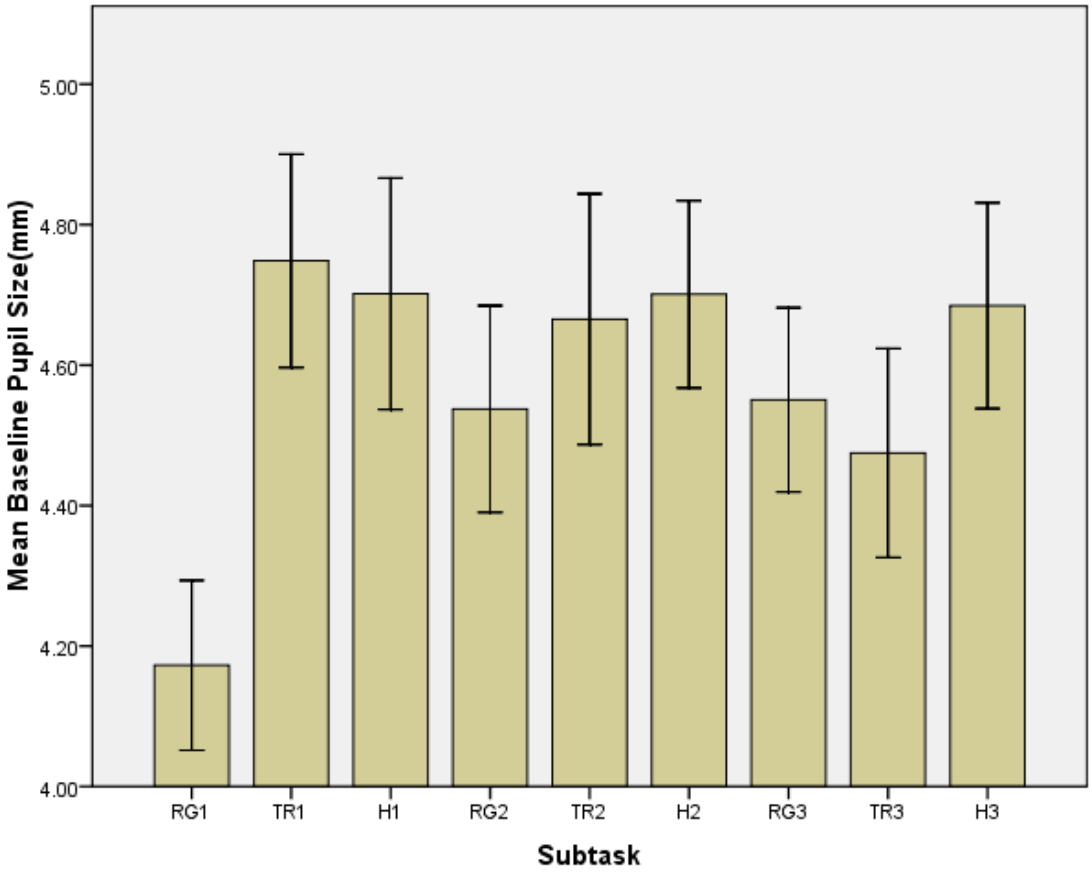


Figure 5.5 Mean of baseline pupil diameter of each subtask for all trials of all subtasks when performing.

Note. The baseline is the mean pupil size of the 400ms period at the end of Transport phase.

To examine which parts of the pupil diameter in the 8s-window had significant differences between the three types of subtasks, the moment-to-moment based *t*-testing approach described in experiments 1-3 was applied.

5.3. Results

5.3.1. Accuracy

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), since we have observed that the quality of the eye movement data cannot be guaranteed when TF/TT rate is lower than a certain value (less than about 70%). From the 51 valid trials, there would be a total of 459 subtasks (each trial has 9 subtasks). However, we excluded the first subtask (RG1) and last two subtasks (TR3 and H3) of each trial (153 subtasks) and subtasks with peg-dropping (16 subtasks). As a result, 159 subtasks were excluded and 290 valid subtasks were available for further analysis.

5.3.2. Movement time

An execution time for a subtask starts from when the tool leaves the current target, and ends when the tool leaves that target for the next subtask execution, such that that the execution time of a subtask is the interval between two consecutive Tool-leave moments. *Figure 5.6* shows the mean execution time over all subjects for the Transport and Landing phases and for the complete subtask, for the three types of subtasks. The mean execution time for all subtasks is 4.3±3.1s, and differs between the three types of subtasks. There is a significant main effect of subtask types in terms of mean subtask execution time ($F_{2,287} = 25.730$, $p < .0001$). Post Hoc test (Tukey HSD) shows that the mean execution time of the H subtask (2.7±1.5s) is significantly shorter ($p < .0001$) than those of the TR and RG subtasks (4.9±1.7s and 5.4±4.3s respectively), and there is no significant difference between TR and RG.

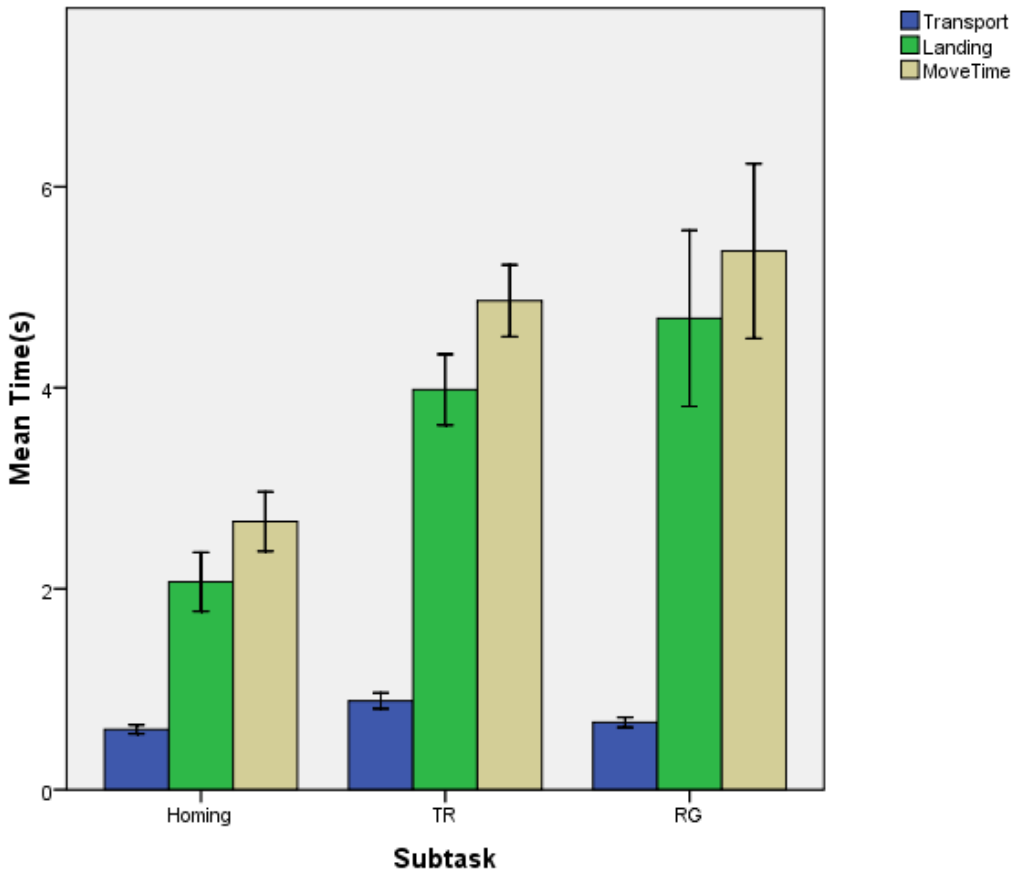


Figure 5.6 Mean execution time for movement phases and a complete subtask of different subtasks for all subjects.

Note. The blue column is the execution time for the Transport phase, the green column is the execution time for the Landing phase, and the mustard column is the execution time for the complete subtask. The error bars are for 95% confidence intervals.

The mean execution time of Transport phases for all subtasks is 0.7±0.3s, and differs significantly between the three types of subtasks ($F_{2,287}$ = 25.174, p < .0001). Post Hoc test (Tukey HSD) shows that the mean Transport time of the TR subtask (0.9±0.4s) is significantly longer (p < .0001) those of RG and H subtasks (0.7±0.2s and 0.6±0.2s 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 two tasks (see Table 5.1).

The mean execution time of Landing phases for all subtasks is 3.5±3.0s, and differs significantly between the three types of subtasks ($F_{2,287}$ = 22.934, p < .0001).

Post Hoc test (Tukey HSD) shows that the mean time of the Landing phase of the H subtask (2.1±1.5s) is significantly shorter ($p < .0001$) those of TR and RG subtasks (4.0±1.7s and 4.7±4.3s respectively), and there is no significant difference between the Landing phases of the RG and TR subtasks.

5.3.3. Pupil responses to task difficulty in performing surgical tasks

Figure 5.7 shows the mean changes of pupil diameter in an 8-second window for each of the three subtasks from all subjects when performing the task, aligned at the end of the Transport phase. The curves of the pupil response to the three subtasks clearly exhibit different shapes.

During the RG subtask (green curve) the pupil dilates in both Transport and Landing phases and develops to its peak pupil size at about 3.9s in the Landing phase.

During the H subtask (black curve) the pupil behaves oppositely to that of the RG—the pupil mostly constricts during both Transport and Landing phases.

During the TR subtask (red curve) the pupil diameter shows a V-shape—the pupil starts from its peak size at the beginning of the Transport phase, drops to its smallest (valley) size when the tool reached, and then dilates to its peak size in the middle of the Landing phase (about 2.7s after the start of landing).

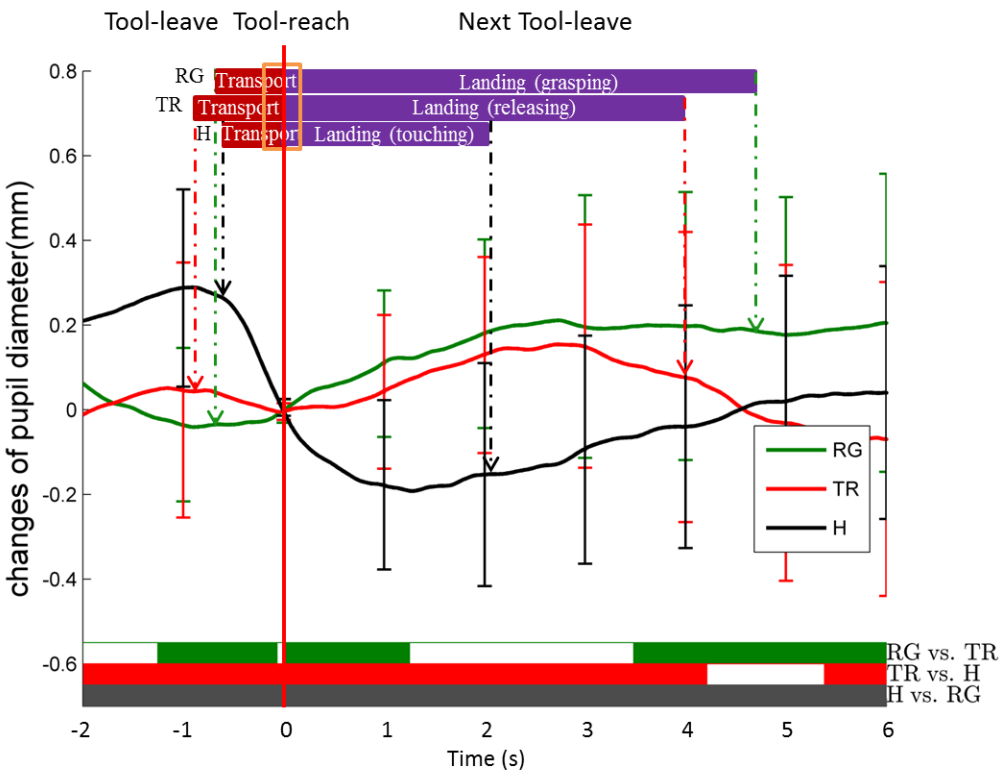


Figure 5.7 Mean pupil diameter changes (subtracted by baseline) for 3 different subtasks from all subjects when performing the task.

Note. Data are aligned over an 8-second window 2 seconds before Tool-reach, shown by the vertical red line. The three colored curves—black, red, and green—are the mean pupil diameter changes for H, TR, and RG subtasks respectively. The pupil baseline diameter is the mean pupil diameter over a 400ms period at the Tool-reach moment, as shown in the red rectangle. The three rows of bars at the top represent Transport (red) and Landing (purple) phases for the three subtasks RG, TR, and H, respectively. The three colors of bars at the bottom indicate significant differences in pupil dilation between H, TR and RG with black representing H vs. RG, red representing H vs. TR, and green representing TR vs. RG. The error bars showing 1 std. dev. are drawn every 1s.

Table 5.2 shows the extent and duration of the pupil diameter changes between the three subtasks in Transport and Landing phases while performing the task. Both the peak pupil size of RG in the Transport phase and the peak pupil size of Homing in the Landing phase are negative since the pupil constricts in these two specific phases of the subtasks. ANOVA results show that there are significant main effects in terms of peak pupil size ($F_{2,287}$ = 215.777, p < .0001) in the Transport phase and peak pupil size ($F_{2, 287}$ = 358.024, p < .0001) and peak duration ($F_{2, 287}$ = 115.894, p < .0001) in the Landing

phase between the three subtasks. Post Hoc test (Tukey HSD) analysis shows that the values of the two pupil diameter parameters in the Landing phase (peak pupil size, and peak duration) positively correlate with the type of subtask ($p < .0001$), i.e., RG subtask (the hardest) has the highest peak pupil size and longest peak duration, TR subtask (the medium hard) has the medium peak pupil size and peak duration, and Homing subtask (the easiest) has the lowest peak pupil size (negative to the baseline) and the shortest peak pupil dilation. In the Transport phase, the peak pupil size of Homing is significantly more than that of TR ($p < .05$) and RG ($p < .0001$), and TR is significantly more than RG ($p < .0001$).

Table 5.2 The mean pupil diameter changes and duration over all subjects for Transport and Landing phases, for different subtasks

	H	TR	RG
Transport phase			
Peak pupil size (constriction)	0.38±0.21mm	0.22±0.22mm	-0.16±0.13mm
Peak duration (constriction)	0.95±0.48s	1.01±0.78s	0.87±0.59s
Landing phase			
Peak pupil size (dilation)	-0.29±0.18mm	0.34±0.22mm	0.46±0.23mm
Peak duration (dilation)	1.24±0.61s	2.71±1.15s	3.86±1.67s

Moment-to-moment base significant testing was applied between the curves of the three subtasks, with the results as shown in the bottom colored bars in *Figure 5.7*. A paired *t*-test was performed for each time point in each pair of subtasks containing samples from the 14 subjects; significance is indicated by ($p < .05$). The colored areas of the bars at the bottom of *Figure 5.7* indicate significant differences, with the black comparing Homing and RG, the red comparing Homing and TR, and the green comparing TR and RG. Almost all moments along the 8s window are significantly different between pairs of three subtasks except around the Tool-reach moment.

5.4. Discussion

Figure 5.6 shows that the mean execution time of a subtask increased as the difficulty of the subtask increased. However, there is no significant difference in the mean execution time between RG and TR, although RG is estimated to be more difficult than TR (Table 5.1). The probable reason is that the naïve application of Fitts' law defining the target sizes to be the peg size in the RG task and the dish size in the TR task does not totally reflect the Landing difficulty. The transport time of TR subtask is longer than those of RG and Homing due to its longer tool travelling distance (65mm of TR vs. 37mm of RG and Homing).

The changes of pupil diameter are able to distinguish the difficulty of the subtasks in the simulated surgical task, such that the peak pupil dilation in the Landing phase positively correlates to the difficulty of the subtask: RG has more peak pupil dilation than TR and Homing, as shown in *Figure 5.7*.

However, the pupil response patterns in the Transport phase did not reflect the ID value of the subtask; instead they followed the execution order of subtasks. This means the pupil response was not just regulated by the current action in the hand, but was also influenced by the previous action, and sometimes, by an action in planning. Starting from the end of the Homing subtask with a relative low task requirement, the RG subtask evokes an increase of pupil diameter through the Transport phase into the Landing phase. In contrast, both TR and Homing subtasks have a decrease of pupil diameter in the Transport phase since they start from completion of a relatively high requirement task (RG and TR).

On the basis of the observation, it can be concluded that the extent of pupil dilation in the Landing phase positively correlates to the task requirement, whereas for the Transport phase the pattern of pupil dilation reveals the execution order of the subtasks which is affected by the previous subtask. Therefore, the pupil diameter may be employed to indicate the task difficulty of the simulated surgical task.

Some threats to validity have to be considered in this study. First, the difficulty of subtasks cannot be quantitatively defined due to the complex action added at the end of the tool movement and the irregular size of the target objects. Second, the execution

order of the subtasks is fixed, i.e. from RG to TR to Homing, and then back to RG etc. We have not tested other execution orders of the subtasks due to the characteristic of the peg transportation task. More types of real-life motor tasks are needed to be tested. Third, mental activities such as anxiety concerning the performance during the execution of the task would affect the physical load.

5.5. Summary

This Chapter investigated the pupil responses during performing a relatively complex motor task, i.e. a simulated surgical task containing different subtasks executed sequentially in a mixed order. We automatically separated each subtask employing video processing techniques and identified the critical tool movement moments for the signal averaging of pupil diameter to capture the movement-related pupil responses.

We found the pupil responses in performing the surgical task exhibit a pattern that the changes of pupil diameter in the Landing phase correlate with task requirement. This finding suggests that the pupil diameter could be employed to measure task workload in performing simulated surgical tasks.

The results of this study have been written up in a manuscript under review for publication (Bin Zheng, Jiang, & Atkins, 2014)[4].

[4] B. Zheng, **X.T. Jiang**, M. S. Atkins, "Detection of Changes of Surgical Difficulty: Evidence from Pupil Responses", Under review (2014).

Chapter 6.

Conclusion and future work

6.1. Conclusions

> While the work presented here follows in the footsteps of decades of work on using pupil diameter to estimate the cognitive load of various tasks, it is the first to tie Fitts' index of difficulty to pupil diameter changes. This in turn might be a significant step in standardizing our evaluations of the demands of manual-visual tasks. These evaluations can even be used in a closed-loop system which would adjust the HMI to keep the user's cognitive load under some threshold. The paper describes a well thought-out and carefully conducted experiment. The ideas are presented clearly and relevant work is adequately discussed. This is exciting work that CHI 2014 attendees would enjoy hearing about.
>
> (one of the reviewers to our submission to CHI 2014)

This thesis presented a method of objectively measuring task workload of motor tasks employing pupil diameter and investigated how the pupil responds to a tool movement and how pupil diameter can be employed as an objective indicator of task workload in goal-directed motor tasks. We conducted a series of experiments, starting with simple discrete target-pointing tasks, followed by a more complex continuous aiming task, and finally a complex surgical simulation task, to reveal the basic pupil response patterns during tool movements and to correlate the changes of pupil diameter to the changes to task workload of motor tasks.

In Experiment 1, we found a common pattern of pupil responses to the task requirement of a discrete movement. The pupil dilates 1.5s before the tool starts to move, followed by a slight constriction 200ms after the tool starts to move. The pupil develops to its peak dilation around 2s after the tool leaves the target. The Valley-to-Peak pupil dilation and duration positively correlate to the task difficulty which is

quantitatively defined by Fitts' index of difficulty. These findings indicate that the changes of pupil size are regulated by task requirement, such that the pupil responses to task workload in goal-directed movements respect Fitts' law. The results of Experiment 1 were orally presented and published in the Proceedings of Conference of CHI 2014[5]. This paper won Best of CHI Honorable Mention Award.

These findings were further confirmed by Experiment 2 using a similar simple discrete pointing task to that of Experiment 1, but with different size targets at a constant distance apart, and the same size targets at different distances apart. It was found that the pupil size changes in response to both different target sizes and different distances between targets. Furthermore, the extent of pupil dilation positively correlates to the task workload evoked by both decrease of target size and increase of target distance. The results of Experiment 2 were orally presented and published in the Proceedings of Conference of ETRA 2014 (Jiang, Atkins, Tien, Zheng, et al., 2014)[6].

In Experiment 3, we further investigated the pupil responses to task workload in more complex continuous aiming movements, which are more common than discrete movements in real life. We found the pupil responses still respect Fitts' law but with a slightly different pattern due to the overlap in pupil size changes between arriving at a target and the preparation for the next movement; the pupil constricts during the Transport phase and dilates during the Landing phase, and this holds for different difficulty levels of movements, where harder task IDs elicit a higher magnitude of pupil response and longer pupil dilation/constriction duration. These results have been written in a manuscript for publication (Jiang, Zheng, et al., 2014)[7].

On the basis of the findings from experiments 1-3, we analyzed the pupil data from a previous experiment to examine whether the pupil responses can be employed to indicate task workload in a complex surgical simulation task. We found that the pupil

[5] **X.T. Jiang**, M.S. Atkins, G. Tien, R. Bednarik, and B. Zheng, "Pupil Responses during Discrete Goal-directed Movements", Proc. CHI 2014, 2075-2084 (2014) **(Best of CHI Paper Honorable Mention Award)**.

[6] **X.T. Jiang**, B. Zheng, G. Tien, R. Bednarik, and M.S. Atkins, "Pupil Dilations during Target-pointing Respect Fitts' Law", Proc. ETRA 2014, 175-182 (2014).

[7] **X.T. Jiang**, B. Zheng, R. Bednarik, and M.S. Atkins, "Pupil Responses to Continuous Aiming Movements", Under review (2014).

diameter changes in the Landing phase reflect the task difficulty but the pupil diameter changes in the Transport phase reflect the execution order of the subtasks. The results have been written in a manuscript and submitted for publication (Bin Zheng, Jiang, & Atkins, 2014)[8].

We tackled the challenges of capturing pupil responses to self-paced movements. Numerous factors cause variations of pupil diameter including changes of lighting, luminance, emotion, and mental workload. The signal averaging technique is effective to capture the movement-related pupil responses by aligning and averaging many repetitive movements, whereas here the common time point for alignment is implicit. We extracted the pupil responses to movements using a time window defined according to kinematics of tool movement, and found the moments when the tool starts to move (Tool-leave) and when the tool reaches the vicinity above the target (Tool-reach) are effective for the signal alignment in goal-directed movements. Tooltip positions were automatically extracted from the recorded task videos without adding any motion tracking equipment.

To perform the data analysis, the tooltip positions were automatically extracted from the task videos using customized algorithms. For the task videos of Experiment 1 to 3, a very simple but effective algorithm employing a biggest connected object searching method was developed (Appendix A). A relatively complex video processing algorithm was developed to extract tooltip positions from the peg transportation task videos using background subtraction technique (Appendix B). The output of this algorithm was compared with manually annotated ground truth data and very good results were achieved, with an average overlap rate (within less than 0.75˚ viewing angle, around 28 pixels on the 17″ screen when the participants were 60 cm away the eye-tracker, corresponding to 7.4 mm in physical distance on the display) between computed tooltip positions and the ground truth of 98.4%. These tooltip extraction algorithms could be applied to similar environments e.g. simulated laparoscopic surgery

[8] B. Zheng, **X.T. Jiang**, and M. S. Atkins, "Rapid Detection on the Change of Surgical Task Load: Evidence from Pupil Responses", Under review (2014).

tasks. The details of this algorithm have been written into a paper for publication in Surgical Innovation (Jiang, Atkins, & Zheng, 2014)[9].

We quantified the task difficulty of goal-directed movement by innovatively employing Fitts' index of difficulty. The size and distance of the targets were manipulated to generate different task difficulties.

To observe the effect of pupil responses of continuous movements, we divided a movement into Transport and Landing phases according to the kinematics of the tool movement. We found that the peak pupil dilation in the Landing phase positively correlated to the task difficulty and the changes of pupil in the Transport phase reflected the influence of the previous movement.

Overall, this thesis lies in the area of interdisciplinary research of Computing Science, Human Factors, Experimental Research, Psychophysiology, and Surgical Simulation, with a focus on exploring how the user responds to the task workload of motor tasks in human-machine interaction. We developed techniques to objectively and quantitatively measure mental workload in goal-directed tasks using pupil diameter, and found the pupil diameter can serve as an objective measure of task workload in goal-directed movements, by conducting three experiments of discrete and continuous target-point tasks and a study of real-life motor task. The findings constitute the foundation for developing methods to objectively and quantitatively evaluate task workload of motor tasks using pupil diameter, and have a variety of implications in enhancing psychophysiological interactions in human-centered HCI and evaluating mental workload in high skill-demanding domains such as driving, aviation, and surgery.

6.2. Threats to validity

Individual variation exists in pupil diameter changes between subjects even in response to a same difficulty task, in terms of amplitude, rate of changing size, and response latency to stimuli. The distinguishable pupil response patterns to different

[9] **X.T. Jiang,** B. Zheng, and M. S. Atkins, "Video Processing to Locate the Tooltip Position in Surgical Eye-hand Coordination Tasks", Surgical Innovation, In press (2014).

difficulty tasks in the studies of this thesis were derived by averaging many repeated windows of pupil signal over several individuals. This signal averaging technique may not be applicable to a real-time analysis in terms of individual performance without some kind of individual calibration using a pre-defined task.

The learning effect may affect the validity of the pupil diameter as an indicator of task workloads throughout continuous movements and peg transportation task. That is, after many repeated tool movements for the same ID and the same task, the participants may get used to the task and the task workload would decrease.

Pupil response may be much slower than the continuous tool movement time, so that the pupil response pattern could not reflect the task workload changes in continuous motor tasks such as a repetitive peg transportation task.

In most real-life motor tasks, such as the peg transportation task, the difficulty of subtasks cannot be quantitatively defined due to the complex action added at the end of the tool movement and the irregular size of the target objects. Also in the peg task, the execution order of the subtasks is fixed, i.e. from RG to TR to Homing, and then back to RG etc. We could not test other execution orders of the subtasks due to the characteristic of the peg transportation task. More types of real-life motor tasks need to be tested. Lastly, mental activities such as anxiety about the performance during the execution of the task would affect the physical load.

6.3. Future work

In future, we plan to monitor and model the pupil over short intervals, so we can make accurate predictions about the difficulties a user experiences. This will contribute to building more intelligent interfaces both for general HCI and for surgery procedures, for example for automatically evaluating and monitoring the task workload continuously in a very fine resolution during image-guided procedures.

More work needs to be done before we can bring this pupil measurement technique into real-world applications. The approach of averaging many repetitive task epochs to show the pattern of pupil size change may not be able to measure task

requirements in real-world applications. We need to employ machine learning to train pupil response-to-difficulty models using pupil diameter features such as Valley-to-Peak pupil dilation and duration, and then use the trained model to classify each of new movements or subtasks in real-time. Furthermore, the learning effect should be considered during the training and classifying process, for example, by discarding tasks according to the learning curve of the performance monitored.

Another important work in the future is to build Fitts' law model by conducting more experiments with more IDs in different experimental settings, for example in open operation settings instead of tele-manipulation laparoscopic settings. Once the Fitts' model and its parameters are accurately estimated in terms of pupil diameter changes, the pupil measurement could be further applied to complex motor tasks where Fitts' law is not readily applicable since the targets are in irregular size.

In a real-life application, such as monitoring mental workloads of surgeons in the Operating Room or drivers in the car, many factors are needed to be considered affecting the validity of the pupil measurement. First, multiple sources of the workload should be separated and measured by employing multiple modalities such as heart rate, skin conductivity, and respiration rate rather than merely pupil diameter. These physiological signals would provide a different aspect of the measurement of mental workload. For example, cardiac measures are able to distinguish the overall task complexity (Gao, et al., 2013), which could be a supplementary to the event-based pupil measurement method as proposed in this thesis. It is also possible to resolve the individual variation problem of the pupil measurement by calibrating individual profiles employing physiological signals or other eye metrics such as blinks. The key question becomes how to find out a correlation between a physiological signal and the individual profile of pupil diameter changes, which remains a subject for future research.

References

Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 529-543. doi: 10.1037/0096-1523.15.3.529

Abrams, R. A., Meyer, D. E., & 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. doi: 10.1037/0096-1523.16.2.248

Adam, J. J., Nieuwenstein, J. H., Huys, R., Paas, F. G. W. C., Kingma, H., Willems, P., & Werry, M. (2000). Control of rapid aimed hand movements: The one-target advantage. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 295-312. doi: 10.1037/0096-1523.26.1.295

Adam, J. J., & Paas, F. G. W. C. (1996). Dwell time in reciprocal aiming tasks. *Human Movement Science, 15*(1), 1-24. doi: http://dx.doi.org/10.1016/0167-9457(95)00041-0

Aggarwal, R., Dosis, A., Bello, F., & Darzi, A. (2007). Motion tracking systems for assessment of surgical skill. *Surgical Endoscopy, 21*(2), 339-339. doi: 10.1007/s00464-005-0561-3

Ahern, S., & Jackson, B. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science, 205*(4412), 1289-1292.

Al-Aidroos, N., Fischer, M. H., Adam, J. J., & Pratt, J. (2008). Structured Perceptual Arrays and the Modulation of Fitts's Law: Examining Saccadic Eye Movements. *Journal of Motor Behavior, 40*(2), 155-164. doi: 10.3200/jmbr.40.2.155-164

Atkins, M. S., Jiang, X., Tien, G., & Zheng, B. (2012). *Saccadic Delays on Targets while Watching Videos*. Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications, Santa Barbara, California.

Attar, N., Schneps, M., & Pomplun, M. (2013). Pupil size as a measure of working memory load during a complex visual search task. *Journal of Vision, 13*(9), 160. doi: 10.1167/13.9.160

Backs, R. W., Ryan, A. M., & Wilson, G. F. (1994). Psychophysiological Measures of Workload during Continuous Manual Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 36*(3), 514-531. doi: 10.1177/001872089403600308

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI), 14*(4), 1-28. doi: 10.1145/1314683.1314689

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

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science, 5*(10), 371-372.

Bednarik, R., Vrzakova, H., & Hradis, M. (2012). *What do you want to do next: a novel approach for intent prediction in gaze-based interaction.* Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications, Santa Barbara.

Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(3), 199-208. doi: 10.1016/j.trf.2010.12.001

Berguer, R., Smith, W. D., & Chung, Y. H. (2001). Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical Endoscopy, 15*(10), 1204-1207. doi: 10.1007/s004640080030

Bootsma, R., Marteniuk, R., MacKenzie, C., & Zaal, F. J. M. (1994). The speed-accuracy trade-off in manual prehension: effects of movement amplitude, object size and object width on kinematic characteristics. *Experimental Brain Research, 98*(3), 535-541. doi: 10.1007/bf00233990

Bradshaw, J. (1967). Pupil size as a measure of arousal during information processing. [10.1038/216515a0]. *Nature, 216*(5114), 515-516.

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods, 45*(4), 1322-1331. doi: 10.3758/s13428-013-0327-0

Brown, D. (2013). Tracker Video Analysis and Modeling Tool Retrieved Feb. 06, 2013, from http://www.cabrillo.edu/~dbrown/tracker/

Cain, B. (2004). A Review of the Mental Workload Literature. *English, 0*(1998), 1-34.

Cassenti, D. N., & Kelley, T. D. (2006). Towards the shape of mental workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(11), 1147-1151. doi: 10.1177/154193120605001107

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

Chen, S., & Epps, J. (2014). Using Task-Induced Pupil Diameter and Blink Rate to Infer Cognitive Load. *Human–Computer Interaction, 29*(4), 390-413. doi: 10.1080/07370024.2014.892428

Colque, R. M., & Cámara-Chávez, G. (2011). *Progressive Background Image Generation of Surveillance Traffic Videos Based on a Temporal Histogram Ruled by a Reward/Penalty Function.* Paper presented at the 24th SIBGRAPI Conference on Graphics, Patterns and Images.

Cristani, M., Farenzena, M., Bloisi, D., & Murino, V. (2010). Background Subtraction for Automated Multisensor Surveillance: a Comprehensive Review. *EURASIP Journal on Advances in Signal Processing - Special issue on advanced image processing for defense and security applications, 2010*, 1-24. doi: 10.1155/2010/343057

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice*: Springer-Verlag New York, Inc.

Elliott, D., Helsen, W. F., & Chua, R. (2001). A century later: Woodworth's (1899) two-component model of goal-directed aiming. *Psychological Bulletin, 127*(3), 342-357. doi: 10.1037/0033-2909.127.3.342

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. doi: 10.1037/h0055392

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

Fong, A., Sibley, C., Coyne, J., & Baldwin, C. (2011). Method for Characterizing and Identifying Task Evoked Pupillary Responses During Varying Workload Levels. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*(1), 207-211. doi: 10.1177/1071181311551043

Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: analysis and correction. *Behavior Research Methods, 43*(4), 1171-1181. doi: 10.3758/s13428-011-0109-5

Gao, Q., Wang, Y., Song, F., Li, Z., & Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics, 56*(7), 1070-1085. doi: 10.1080/00140139.2013.790483

Gawron, V. J. (2008). Human Workload *Human Performance, Workload, and Situational Awareness Measures Handbook, Second Edition* (pp. 87-230): CRC Press.

Girgenrath, M., Bock, O., & Jüngling, S. (2004). Validity of the speed-accuracy tradeoff for prehension movements. *Experimental Brain Research, 158*(4), 415-420. doi: 10.1007/s00221-004-1915-0

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21*(2), 90-95. doi: 10.1177/0963721412436811

Haak, M., Bos, S., Panic, S., & Rothkrantz, L. J. M. (2009). *Detecting Stress Using Eye Blinks and Brain Activity From Eeg Signals*. Paper presented at the Driver Car Interaction & Interface 2009, Prague.

Hancock, P. A., Wulf, G., Thom, D., & Fassnacht, P. (1990). Driver workload during differing driving maneuvers. *Accident Analysis & Prevention, 22*(3), 281-290.

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

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). *Towards an index of opportunity: understanding changes in mental workload during task execution.* Paper presented at the Proc. CHI 2005, Portland, Oregon, USA.

Jiang, X., Atkins, M. S., Tien, G., Bednarik, R., & Zheng, B. (2014). *Pupil responses during discrete goal-directed movements.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto.

Jiang, X., Atkins, M. S., Tien, G., Zheng, B., & Bednarik, R. (2014). *Pupil dilations during target-pointing respect Fitts' law.* Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications, Safety Harbor, Florida.

Jiang, X., Atkins, M. S., & Zheng, B. (2014). Video processing to locate the tooltip position in surgical eye-hand coordination tasks. *Surgical Innovation*, In press.

Jiang, X., Tien, G., Huang, D., Zheng, B., & Atkins, M. S. (2013). Capturing and evaluating blinks from video-based eyetrackers. *Behavior Research Methods, 45*(3), 656-663. doi: 10.3758/s13428-012-0294-x

Jiang, X., Zheng, B., Bednarik, R., & Atkins, M. S. (2014). Pupil responses to continuous aiming movements *Under review*.

Jiang, X., Zheng, B., Tien, G., & Atkins, M. S. (2013). Pupil response to precision in surgical task execution. *Studies in Health Technology and Informatics, 184*, 210-214. doi: doi:10.3233/978-1-61499-209-7-210

Kahneman, D., & Jackson, B. (1966). Pupil diameter and load on memory. *Science, 154*(3756), 1583-1585.

Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology, 41*(2), 175-185.

Klingner, J., Kumar, R., & Hanrahan, P. (2008). *Measuring the task-evoked pupillary response with a remote eye tracker.* Paper presented at the Proceedings of the 2008 symposium on Eye tracking research & applications, Savannah, Georgia.

Kourtis, D., Sebanz, N., & Knoblich, G. (2012). EEG correlates of Fitts's law during preparation for action. *Psychological Research, 76*(4), 514-524. doi: 10.1007/s00426-012-0418-z

Kun, A. L., Palinko, O., & Razumenić. (2012). *Exploring the effects of size and luminance of visual targets on the pupillary light reflex*. Paper presented at the Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, New Hampshire.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain, 106*(3), 623-642. doi: 10.1093/brain/106.3.623

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human Computer Interaction, 7*(1), 91-139. doi: 10.1207/s15327051hci0701_3

Marshall, S. P. (2002). *The Index of Cognitive Activity: measuring cognitive workload.* Paper presented at the Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants Scottsdale Arizona.

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. (1990). Eye Movement Indices of Mental Workload. *Acta Psychologica, 75*(1), 75-89. doi: 10.1016/0001-6918(90)90067-p

Mishra, A., Catchpole, K., Dale, T., & McCulloch, P. (2008). The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surgical Endoscopy, 22*(1), 68-73. doi: 10.1007/s00464-007-9346-1

Moresi, S., Adam, J. J., Rijcken, J., Kuipers, H., Severens, M., & Van Gerven, P. W. M. (2011). Response preparation with adjacent versus overlapped hands: A pupillometric study. *International Journal of Psychophysiology, 79*(2), 280-286. doi: 10.1016/j.ijpsycho.2010.11.005

Noguchi, Y., Shimada, K., Ohsuga, M., Kamakura, Y., & Inoue, Y. (2009). *The Assessment of Driver's Arousal States from the Classification of Eye-Blink Patterns*. Paper presented at the Proceedings of the 8th International Conference on Engineering Psychology and Cognitive Ergonomics: Held as Part of HCI International 2009, San Diego, CA.

Ohsuga, M., Kamakura, Y., Inoue, Y., Noguchi, Y., Shimada, K., & Mishiro, M. (2011). Estimation of Driver's Arousal State Using Multi-dimensional Physiological Indices. *Engineering Psychology and Cognitive Ergonomics, Lecture Notes in Computer Science, 6781*, 176-185. doi: 10.1007/978-3-642-21741-8_20

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*(10), 1346-1353. doi: 10.1111/j.1469-8986.2011.01217.x

Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology, 11*(5), 559-566.

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research, 139*(3), 266-277.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*(3), 560-569. doi: 10.1111/j.1469-8986.2009.00947.x

Pomplun, M., & Sunkara, S. (2003). *Pupil dilation as an indicator of cognitive workload in human-computer interaction.* Paper presented at the Proceedings of the 10th International Conference on Human-Computer Interaction (HCII), Greece.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during Visual Search and Counting: Insights from Pupillometry. [Article]. *Quarterly Journal of Experimental Psychology, 60*(2), 211-229. doi: 10.1080/17470210600673818

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

Prytz, E., Montano, M., & 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. doi: 10.1177/1071181312561396

Richer, F., & 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., & Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Annals Surgery, 252*(1), 177-182.

Smith, P., Shah, M., & Lobo, N. D. V. (2000). *Monitoring Head/Eye Motion for Driver Alertness with One Camera*. Paper presented at the Proceedings of the International Conference on Pattern Recognition (ICPR'00), Session P4.3A.

Spaun, G., Zheng, B., & Swanström, L. (2009). A multitasking platform for natural orifice translumenal endoscopic surgery (NOTES): a benchtop comparison of a new device for flexible endoscopic surgery and a standard dual-channel endoscope. *Surgical Endoscopy, 23*(12), 2720-2727. doi: 10.1007/s00464-009-0476-5

Stefanidis, D., Scerbo, M. W., Korndorffer Jr, J. R., & Scott, D. J. (2007). Redefining simulator proficiency using automaticity theory. *The American Journal of Surgery, 193*(4), 502-506. doi: http://dx.doi.org/10.1016/j.amjsurg.2006.11.010

Tien, G., Atkins, M. S., Jiang, X., Khan, R. S., & Zheng, B. (2013). Identifying eye gaze mismatch during laparoscopic surgery. *Studies in Health Technology and Informatics, 184*, 453-457.

Tien, G., Atkins, M. S., Jiang, X., Zheng, B., & Bednarik, R. (2014). *Verbal gaze instruction matches visual gaze guidance in laparoscopic skills training*. Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications, Safety Harbor, Florida.

Tien, G., Atkins, M. S., Zheng, B., & Swindells, C. (2010). *Measuring Situation Awareness of Surgeons in Laparoscopic Training*. Paper presented at the Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, Austin, Texas.

Tsai, Y.-F., Viirre, E., Strychacz, C., Chase, B., & Jung, T.-P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine, 78*(5 Suppl), B176-185.

van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil Dilation Co-Varies with Memory Strength of Individual Traces in a Delayed Response Paired-Associate Task. *PLoS ONE, 7*(12), e51134. doi: 10.1371/journal.pone.0051134

Villanueva, A., Daunys, G., Hansen, D., Böhme, M., Cabeza, R., Meyer, A., & Barth, E. (2009). A geometric approach to remote eye tracking. *Universal Access in the Information Society, 8*(4), 241-257. doi: 10.1007/s10209-009-0149-0

Watanabe, T., Ikeda, M., Suzuki, T., & Nakamura, F. (1990). Infrared television pupillometer revised: Bright‐pupil illumination and computer automation. *Review of Scientific Instruments, 61*(1), 36-41. doi: http://dx.doi.org/10.1063/1.1141297

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177. doi: 10.1080/14639220210123806

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 449-455. doi: 10.1518/001872008x288394

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

Zheng, B., Jiang, X., & Atkins, M. S. (2014). Rapid Detection on the Change of Surgical Difficulty: Evidence from Pupil Responses. Under revision.

Zheng, B., Jiang, X., Tien, G., Bednarik, R., & Atkins, M. S. (2014). Gaze Characteristics in Video Watching: Loss of Visual Guidance. *PLoS One*, Under review. doi: 10.1007/s00464-012-2268-6

Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., & Atkins, M. S. (2012). Workload Assessment of Surgeons: Correlation between NASA TLX and Blinks. *Surgical Endoscopy, 26*(10), 2746-2750. doi: 10.1007/s00464-012-2268-6

Zheng, B., Tien, G., Atkins, S. M., Swindells, C., Tanin, H., Meneghetti, A., . . . Panton, M. (2011). Surgeon's Vigilance in the Operating Room. *American Journal of Surgery, 201*(5), 667-671.

Appendix A.

Tool detection from Fitts' pointing task videos

This algorithm was designed to detect tooltip positions from the task videos recorded from Experiment 1-3, in which the participants were asked to move and touch the circle printed on the paper inside the training box using a long shafted grasper with their eye motions recorded by a Tobii X50 eye-tracker. The scene inside the training box was illuminated and recorded by the endoscopic camera attached to the training box, in the format of Audio Video Interleave (AVI) at 30 fps/s.

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 A.1A* Second, the biggest connected object was searched and identified as the tool, as shown within the red rectangle in *Figure A.1*B. To simplify tool detection, the targets in the background of the image were designed in black dash-dotted circles in Experiment 1-3, in case of the target circle being the biggest object in the image when the tool is only a small part in the image. Furthermore, the tip of the grasper was black taped to increase the contrast between the tool and the back ground.

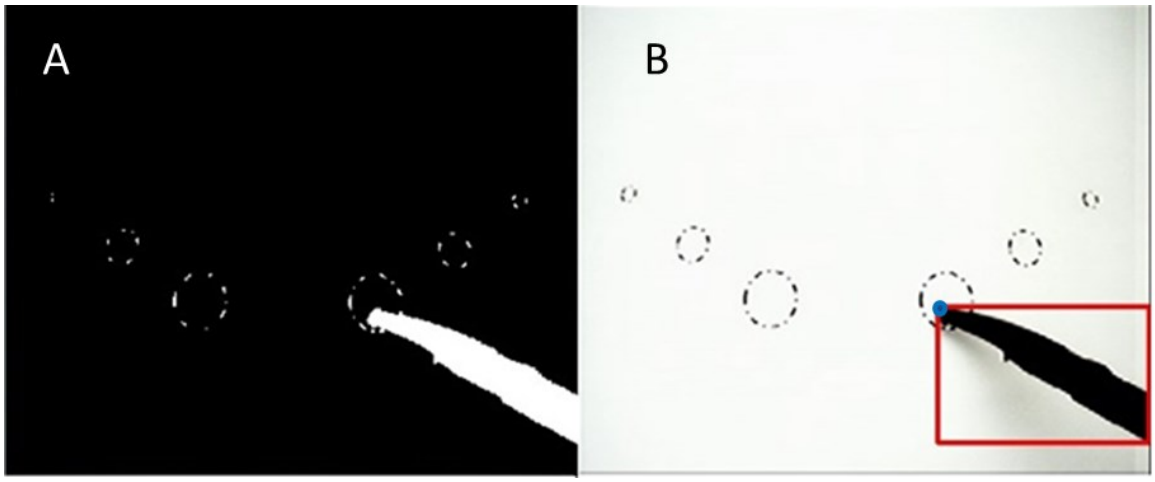


Figure A.1. Key steps in detecting tooltip from surgical videos. Panel A shows the binary-thresholded image with mostly the tool left. Panel B shows the recognized tool (in red rectangle) and tooltip (the blue dot).

Third, the coordinates of the left top corner of the tool rectangle (red rectangle) was used as tooltip position, as the tool was always consistently north-west orientated. The determined tooltip position was shown as the blue dot in Figure A.1 B. The tool rectangle is determined by the left-upmost and the right-bottommost coordinates of the pixel blob of the detected biggest object.

Appendix B.

Tool detection from peg transportation task videos

This algorithm was designed to detect tooltip positions from the task videos recorded from peg transportation tasks (Chapter 5), in which the participants were asked to transport a green peg between three dishes using a long shafted grasper with their eye motions recorded by a Tobii 17/50 eye-tracker. The scenery inside the training box was illuminated and recorded by the endoscopic camera attached to the training box, in the format of Audio Video Interleave (AVI) at 30 fps/s.

Since the task videos from the peg transportation task is relatively complex than those of target-pointing tasks, we validated the algorithm by comparing the algorithm captured tooltip positions to manually annotated tooltip position from the same video.

B.1. Tool tracking algorithms

The video first underwent a background subtraction, and then the tool was located by searching the biggest connected object from the thresholded foreground binary image. Lastly, the tooltip position was determined according to the status of the grasper, i.e. opened or closed.

B.1.1. Background subtraction

The background subtraction algorithm has two steps including selecting an initial background image and updating the background image (see Figure B.1a and Figure B.1b for examples of typical task video frames with the tool absent and present). A frame without the tool in the image (which usually can be found at the first frame of the video) was used as the initial background image for the background subtraction, as shown in Figure B.1a. Only one frame was needed for all trial videos of a subject if the camera setting (the distance and the direction to the scene) did not change over the trials for a particular subject. However, different background images had to be chosen for different subjects, since they performed the task on different days and the camera setting might have been changed. The initial background frames for the subjects were manually chosen from the videos, and the frame IDs were recorded in a list for the system to automatically read in.

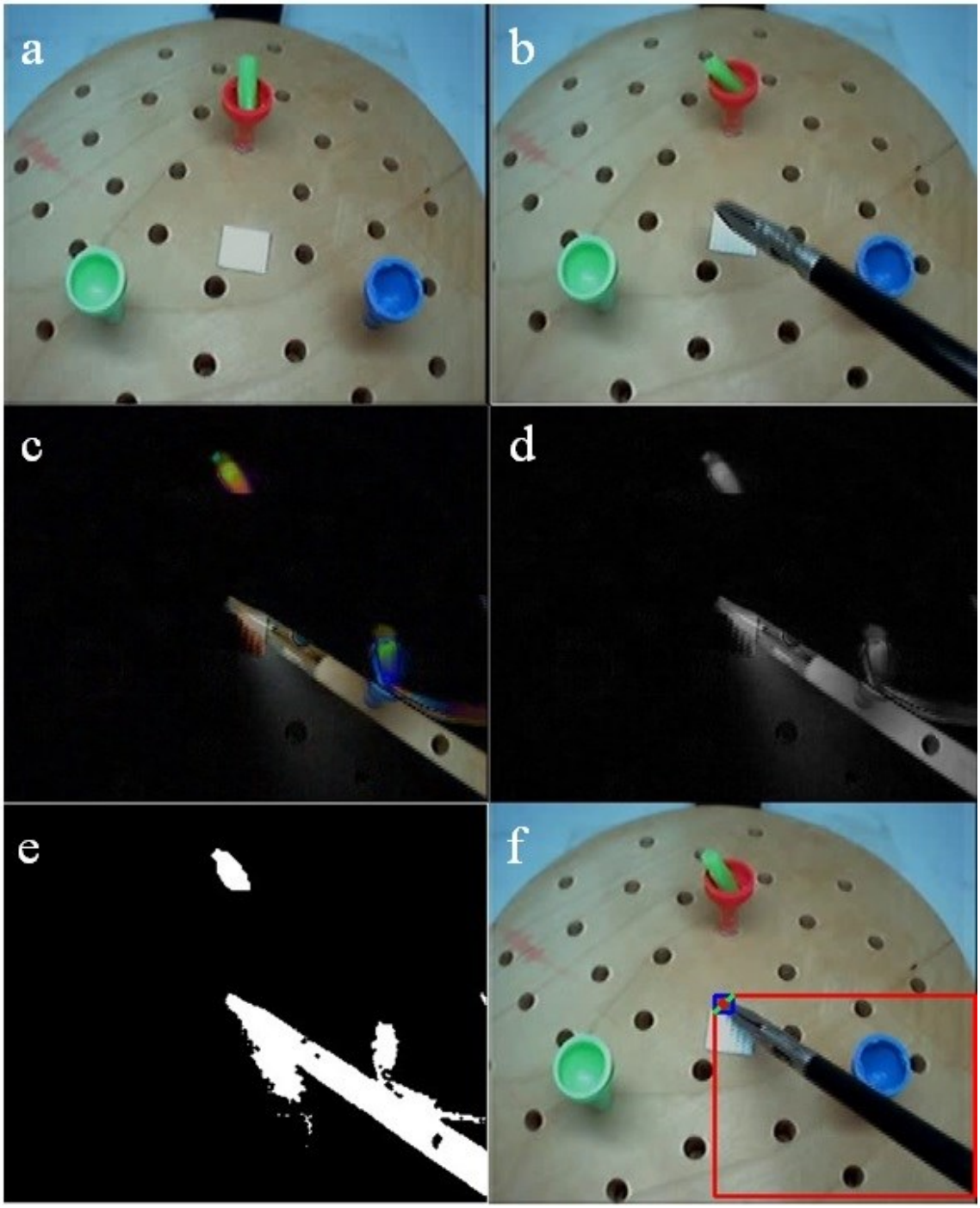


Figure B.1 **An example of the processing of tooltip location. (a) a background image, (b) a target image for tool tracking, (c) the foreground image in RGB color mode, (d) the foreground image in grayscale color mode, (e) the thresholded binary image from (d), and (f) the located tool position as is shown in the red rectangle, in which the red dot on the green line segment is the tooltip position.**

A partial background updating strategy (Colque & Cámara-Chávez, 2011) was employed to update the background image. Only the part of the image outside the rectangle of the recognized tool (see the tooltip location section below) was updated to the background image, since there usually was some noise in the tool rectangle such as shadows. The criterion whether the background image needed to be updated was defined based on the detected object numbers and the number of consecutive noisy frames, where the number of separate objects detected in the frame was greater than a value. For example, when three consecutive noisy frames happened, the background was updated. This criterion was set up to avoid some recoverable image changes, e.g., the effect of a sudden strong flash of light or an incidental touch to the dishes or plate by the grasper, which usually caused 1-2 frames of noisy images and then recovery to clean images. When there was no valid tooltip detected in the image and the image was noisy, the current background was replaced by the initial background image.

B.1.2. Tool location

Each frame was subtracted by the background image to get the foreground image, mostly with only the tool left in the foreground image, as shown in Figure B.1c. Then the foreground images were transformed to grayscale images, and were thresholded to binary images, as shown in Figure B.1e A width-first pixel-wise search strategy was employed to search the connected objects in the binary images. Any adjacent pixels in the binary image were included in an object.

To locate the tool, a horizontal rectangle was derived to surround the found tool, shown as a red rectangle in Figure B.1f. The rectangle was determined by the upper-left and lower-right corners of the tool, i.e., the x-axis value of the left-most point and the y-axis value of the upper-most point in the tool object were used as the x and y axis values of the upper-left corner of the rectangle. Similarly, the x-axis value of the right-most point and the y-axis value of the lower-most point in the tool object were used as the x and y axis values of the lower-right corner of the rectangle, as shown in Figure B.2a.

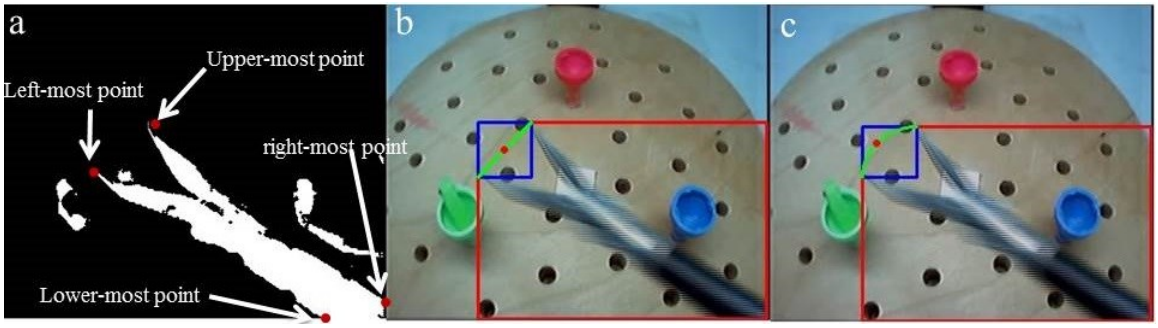


Figure B.2 Illustration of tool rectangle and opened tooltip location (a) shows the four points to determine the tool rectangle, (b) shows the approximated tooltip position using the middle point of the line segment between the two tooltips (the red point on the green line segment), and (c) shows the approximated tooltip position using the middle point of the arc between the two tooltips.

B.1.3. Tooltip location

The calculations for the locations of the closed and opened tooltip are different. For most surgical tasks performed in a laparoscopic environment, the direction of the tool is consistently oriented, e.g., right-handed tool is north-west oriented. When the grasper is closed, the position of the tooltip is derived from the position of the north-west corner of the red rectangle, as is shown in Figure B.1f, and when the grasper is opened, the position of the tooltip is approximated from the middle point of the two opened tooltips, which is the middle point of the line segment between left-most and upper-most points, as shown in Figure B.2b. As an alternative, the algorithm also provides the tooltip approximation from the middle point of the arc of the two opened tooltips, as shown in Figure B.2c.

Several situations might arise which could provide erroneous location of the tool.

Disconnection

To avoid unnecessary disconnected parts belonging to the tool caused by the thresholding problem, as is shown in Figure B.3, pixels with gap less than a certain value were grouped to the same object. Usually the biggest object was selected as the tool from the connected objects according to the total number of pixels in the object.

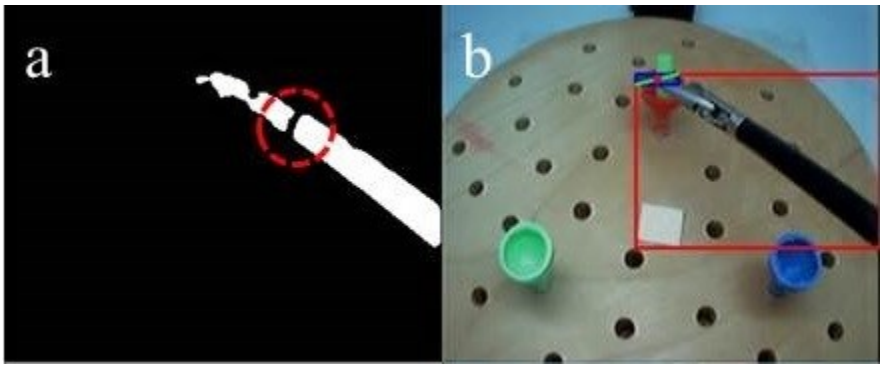


Figure B.3 Example of the disconnection of tool parts. (a) the disconnected part (in the red circle) of the tool, and (b) the recognition of the whole part of the object as a tool.

Tool as largest object

Criteria needed to be set up for judging a valid tool object, since the biggest object in the image was not necessarily always the tool. For example, when the grasper was moved nearly out of the image or inserted into the working field (especially in parallel to the optical axis of the camera) with only a very small part visible in the image, the tool was smaller than some other objects in the image, as shown in Figure B.4. The valid tool judgment criteria were defined on the basis of the characteristics of the tool in the task videos. For example, the lower-right end of the tool should always be connected with the bottom or right edge of the image, as is shown in Figure B.1b. Also the open amount of the tooltip should be less than a certain value, e.g. 80 pixels in the image (which is roughly twice the diameter of a dish in the image).

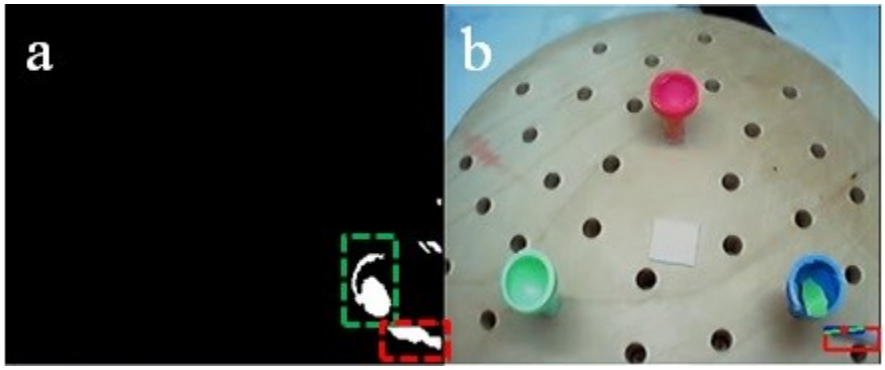


Figure B.4 Example of taking the second largest object as the tool: (a) the largest object (in the green dash-line rectangle) caused by the sleeping foreground object (the green peg), and (b) the recognized tool (in the red rectangle).

Peg in the grasper

When the green peg is picked up by the grasper, it may be misrecognized as a part of the tool, as shown in Figure B.5a. We take this as tool opened situation as shown in Figure B.5b.

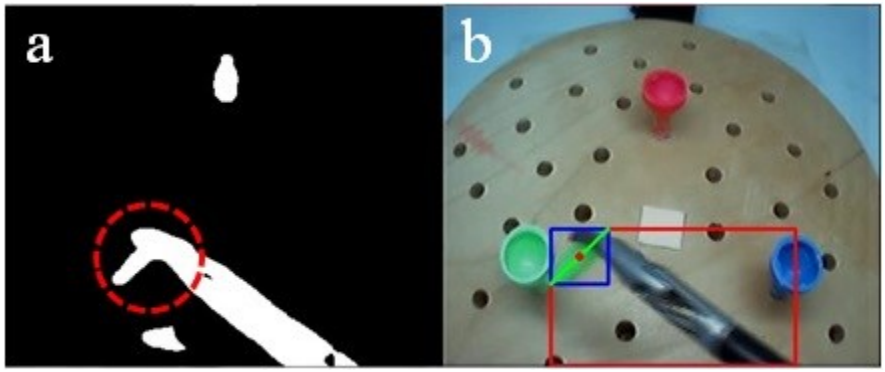


Figure B.5 Example of peg in the grasper: (a) the binary image with the connected peg and grasper (in the red circle), and (b) the recognition result.

B.2. Algorithm validation

We manually annotated the tooltip positions from the task videos as the ground truth data for the evaluation of the algorithm. The task videos were examined frame by frame using an open source video analysis software (Brown, 2013), in which the x and y coordinates of the tooltip were automatically recorded into a spreadsheet by mouse-clicking at the observed tooltip position in the image.

During the manual tooltip annotation, the tooltip was defined as the end point of the grasper when the grasper was closed and the middle point of the two opened tooltips when the grasper was open.

Two parameters were used to indicate the performance of the algorithm: the percentage of the pairs of true tooltip position and computed tooltip position with the distance within a certain value, and the average distance between the pairs of true tooltip and computed tooltip positions.

B.3. Results

A total of 12 task videos (the first trial of each participant) from previous study (Jiang, Zheng, Tien, & Atkins, 2013) were processed by the algorithm to output the tooltip positions, and these task videos were manually annotated to record the true tooltip position in each frame as the ground truth data.

The average root mean square error (RMSE) of the true tooltip positions to the calculated tooltip positions on the captured video was 9.2 ± 2.1 pixels (about 2.4 mm in physical distance on the 17" display). The average overlay rate between the true tooltip and the calculated tooltip from 12 task videos was 98.4% ± 1.7%. An overlay was determined when the distance between the true tooltip and the calculated tooltip was within a 0.75˚ viewing angle (the definition for the radius of a fixation, about 28 pixels in the video frames, corresponding to 7.4 mm in physical distance on the 17˝ display when the participants were 60 cm away the eye-tracker.

Two examples of the computer output tooltip position overlaid with the true tool positions over time are shown in Figure B.6. In the example shown in Figure B.6a, the computer-captured tooltip matched the true tooltip positions very well with the tooltip within a 0.75˚ viewing angle for 99.5% of the video frames. Figure B.6b shows an example with slight displacement between true and recognized tool positions, where the green peg was recognized as a part of the grasper when the participant was holding the green peg in a special direction as shown in Figure B.5. Even in the second example, most of the distances between the computer-recognized tooltip and the true tooltip were within 0.75˚ viewing angle (the overlay rate for this trial was 98.9%).

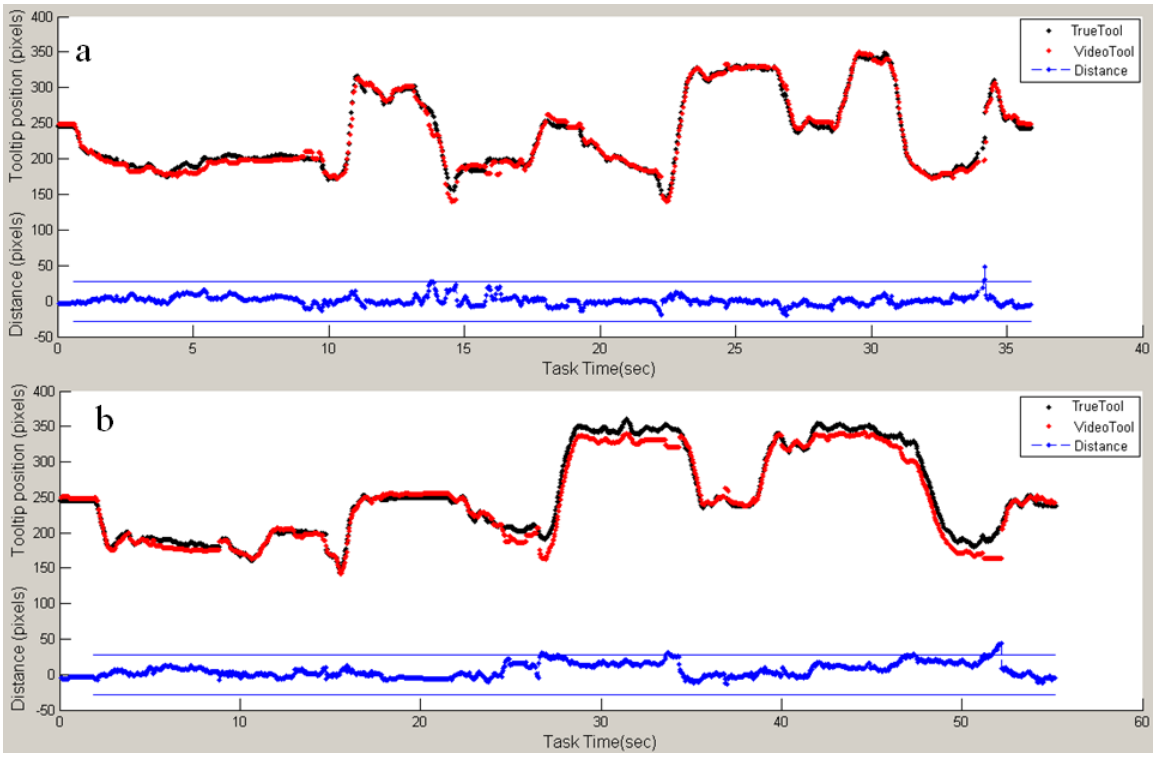


Figure B.6 **Examples of the overlay between the computer output tooltip positions and the true tool positions over time. The black curve and red curve in the figure are the manually annotated true tooltip and the computer output tooltip (video tooltip) positions (Euclidian distance to the origin (top-left corner of the image)), respectively. The blue dots are the distances between the true tooltip and video tooltip positions. The blue horizontal lines are overlay threshold (0.75˚ viewing angle, 28 pixels in the video frames, corresponding to 7.4 mm on the physical display). The panel a shows an example (subject 6) with well-matched computer output tooltip to the true tooltip positions. The panel b shows an example (subject 9) with slight displacement between the true tooltip position and recognized tooltip position during some segments, since the green peg was recognized as a part of the grasper as is shown in Figure B..**

Several hard problems when using background subtraction such as the effects of background oscillating, shadows, and sleeping foreground (Cristani, Farenzena, Bloisi, & Murino, 2010) were successfully avoided. Figure B.7a shows an example of background oscillation, which arose when the tool inadvertently touched the dishes; in this situation the tooltip was correctly located; Figure B.7b shows the correctly recognized result. Figure B.8a shows the sleeping foreground object from the green peg and the shadow of the tool in the binary image; Figure B.8b shows the correctly recognized result.

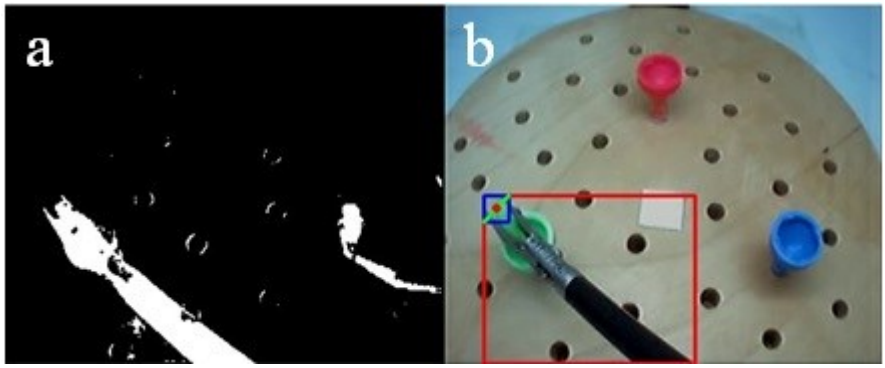


Figure B.7 **Example of the effect of oscillation: (a) the binary image having small objects caused by the oscillation, and (b) the recognition result.**

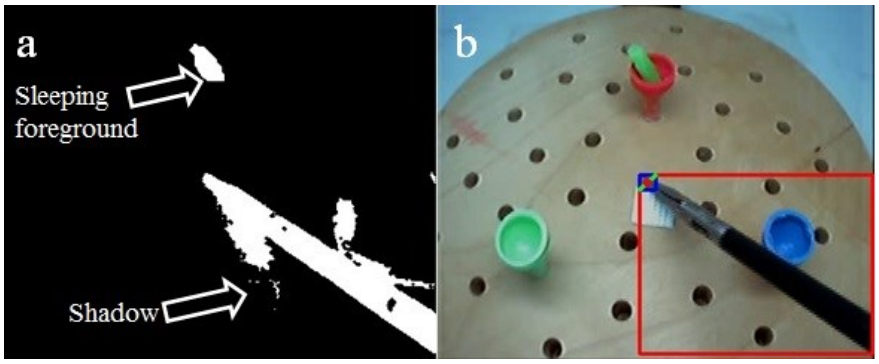


Figure B.8 **Example of the sleeping foreground and shadow: (a) the binary image having sleeping foreground and shadow of the tool, and (b) the recognition result.**

B.4. Discussion

The performance is good with RMSE of about 2.4 mm which is far less than the accuracy of the eye-tracker (0.5˚ viewing angle, corresponding to around 5.2 mm at 60 cm standing distance).

The present tool tracking method is simple but effective, and does not require attaching any extra sensors or color markers to the instruments, which makes it likely to succeed in terms of market dissemination. Some hard problems encountered in moving object detection using background subtraction, e.g., the effects of background oscillation, sleeping foreground objects, and shadows (Cristani, et al., 2010), were overcome in this case by employing object searching and the tooltip point determining strategy.

The background oscillation did not cause problems because the tool was still the biggest connected object in the binary foreground image in most of time (see Figure B.7). In some cases, even when the tool was not the biggest object in the foreground image, the tool location algorithm can successfully distinguish the valid tool according to the criteria described in the method section; similarly, the false alarm object from the sleeping foreground objects (e.g., caused when the green peg was Transported to and stayed in another dish) was mostly smaller than the tool object in the foreground images too, as is shown in Figure B.8a, and also could be avoided when it was bigger than the tool object,

by using the tool validation criteria, as is shown in Figure B.8b. The sleeping object (see Figure B.8) in the foreground also was updated to the background after several frames (according to the parameters in the background updating strategy). The shadow (see Figure B.8) by the illumination from the upper-front side of the tool which mostly lay on the downside of the tool, also did not affect the accuracy of the tooltip position, as the tooltip always was in the upper-left corner of the rectangle of the tool object, as illustrated in Figure B.1f, and only the part outside of tool rectangle is updated to the background image.

The initial background image was manually selected from the task videos instead of being dynamically generated in this study.

The peg would be recognized as the part of the tooltip when the grasper carried the peg in such a position as shown in Figure B.5. By using the middle point between the touching points on the edges (the left-most and upper-most points), the estimated tooltip was pretty close to the actual tooltip. A similar situation happened when the tool was open—the estimated tooltip position was more reasonable than taking the top-left corner of the rectangle as the tooltip position, as shown in Figure B.2b and Figure B.2c.

Although only 2-Dimensional tooltip locations were derived from the eye-tracker output task videos instead of the actual 3-Dimensional scene in the training box, the tooltip locations were still sufficiently accurate for the eye-hand coordination analysis in laparoscopic tasks, since the performers looked at the same 2-Dimensional images from the eye-tracker display screen.

The results were very good. There was nearly no mis-recognition of the tool. There was a slight displacement between the true tooltip and the recognized tooltip, as is shown in Figure B.6b, which was mainly contributed from two factors, one was that the peg in the grasper caused a problem as discussed before, and the other was the poor image quality when the tool was fast moving.

The algorithm could easily be extended to detect more than one moving tools in the training box from the task videos by enhancing the background subtraction and tool detection algorithms, e.g., by enabling a multiple objects detection during the tool object searching.

This method was designed for the analysis of eye-hand coordination under simulated laparoscopy environments. In the future, however, the algorithms could be enhanced for applications to laparoscopic surgery for tracking tool trajectories from real surgical videos recorded in the operating room. Challenges would arise because the videos recorded from real patients would present much more noise, and the algorithm would need to employ dynamic background calculations to work in a changing view. However, the scope view often remains unchanged when the tools are working on a site, so we still could take advantage of the stable background for tracking important actions. Surgeons believe tool trajectories are associated tightly with their surgical skills (Aggarwal, Dosis, Bello, & Darzi, 2007). Therefore, we would expect results from such a study would allow us to further assess surgeons' skills in the operating room, as the characteristic orientations of the tools are similar.

B.5. Contributions

A method was presented for tracking the tooltip surgical videos using background subtraction and object searching techniques. This method is perfectly adequate for the analysis of eye-hand coordination in tasks employing stable backgrounds such as laparoscopic tasks using a training box, by taking advantage of the characteristics of the videos, i.e., the stable background and the orientation of the tool in the video. This work achieves the first step towards tool tracking in a real laparoscopic operating room and training environment, where the surgical videos are more complex and more sophisticated video processing methods are needed.

The results of this study have been written in a manuscript and submitted to Surgical Innovation (Jiang, Atkins, & Zheng, 2014)[10].

[10] **X.T. Jiang,** B. Zheng, and M.S. Atkins, "Video Processing to Locate the Tooltip Position in Surgical Eye-hand Coordination Tasks", Surgical Innovation, In press (2014)

Appendix C.

Other work and contributions

The work described in this thesis is a part of an ongoing research project of eye-tracking applications in surgical simulation and training. The author is also involved in other parts of the project and contributed as the first author or coauthor in the following publications. Our research group conducted a vigilance study in operating rooms (Geoffrey Tien, et al., 2010; Bin Zheng, et al., 2011)[11,12]. On the basis of this vigilance study, I developed an algorithm to detect eye blinks from the eye videos recorded from a head-mounted eye-tracker. This algorithm and the evaluation results were published in Journal of Behaviour Research Methods (Jiang, Tien, Huang, Zheng, & Atkins, 2013)[13]. This blink detecting algorithm was applied in another study to classify surgeons' mental workload during operating, and the results were published in Surgical Endoscopy (Bin Zheng, et al., 2012)[14]. I was also involved in analyzing gaze behaviours in performing surgical tasks and watching surgical videos for training purpose. Findings were published in Proceedings of the Symposium on Eye Tracking Research and Applications (Atkins, Jiang, Tien, & Zheng, 2012; Geoffrey Tien, Atkins, Jiang, Zheng, & Bednarik, 2014)15,16 and Studies in Health Technology and Informatics (G. Tien, Atkins, Jiang, Khan, & Zheng, 2013)[17], and another paper is under review for PLoS One (Bin Zheng, Jiang, Tien, Bednarik, & Atkins, 2014)[18].

[11] Tien, G., M. S. Atkins, et al. (2010). Measuring Situation Awareness of Surgeons in Laparoscopic Training. Eye Tracking Research and Applications, Austin, TX, ETRA.

[12] Zheng, B., G. Tien, et al. (2011). "Surgeon's Vigilance in the Operating Room." American Journal of Surgery **201**(5): 667-671.

[13] **X.T. Jiang,** G. Tien, D. Huang, B. Zheng, and M. S. Atkins, "Capturing and Evaluating Blinks from Video-based Eyetrackers", Behavior Research Methods, 1-8 (2013).

[14] B. Zheng, **X.T. Jiang,** G. Tien, A. Meneghetti, O. N. M. Panton, M. S. Atkins, "Workload Assessment of Surgeons: Correlation between NASA TLX and Blinks", Surgical Endoscopy 26, 2746-2750 (2012).

[15] M.S. Atkins, **X. T. Jiang,** G. Tien, and B. Zheng, "Saccadic Delays on Targets while Watching Videos", Proceedings of the Symposium on Eye Tracking Research and Applications, 405-408 (2012).

[16] G. Tien, M.S. Atkins, **X.T. Jiang,** B. Zheng, and R. Bednarik, "Verbal Gaze Instruction Matches Visual Gaze Guidance in Laparoscopic Skills Training", Proc. ETRA 2014, 331-334 (2014).

[17] G. Tien, M.S. Atkins, **X.T. Jiang,** R. S.A. Khan, and B. Zheng, "Identifying Eye Gaze Mismatch During Laparoscopic Surgery", Studies in Health Technology and Informatics 184, 453 – 457 (2013).

[18] B. Zheng, **X.T. Jiang,** R. Bednarik, and M. S. Atkins, "Gaze Characteristics in Video Watching: Loss of Visual Guidance", Under revision (2014).