

Pupillary Response Based Cognitive Workload Measurement under Luminance Changes^{*}

Jie Xu, Yang Wang, Fang Chen, and Eric Choi

National ICT Australia
University of New South Wales
jie.jackxu@gmail.com,
{yang.wang, fang.chen, eric.choi}@nicta.com.au

Abstract. Pupillary response has been widely accepted as a physiological index of cognitive workload. It can be reliably measured with remote eye trackers in a non-intrusive way. However, pupillometric measurement might fail to assess cognitive workload due to the variation of luminance conditions. To overcome this problem, we study the characteristics of pupillary responses at different stages of cognitive process when performing arithmetic tasks, and propose a fine-grained approach for cognitive workload measurement. Experimental results show that cognitive workload could be effectively measured even under luminance changes.

Keywords: Cognitive workload, eye tracker, luminance, pupillary response.

1 Introduction

Cognitive workload measurement plays an important role in various application areas involving human-computer interface, such as air traffic control, in-car safety and gaming [2]. By quantifying the mental efforts of a person when performing tasks, cognitive workload measurement helps predict or enhance the performance of the operator and system. Physiological measures are one class of workload measurement techniques, which attempts to interpret the cognitive processes through their effect on the operator's body state [5]. In the past, physiological measures usually entailed invasive equipment. With the advance of sensing technologies in recent years, the measuring techniques have become less intrusive, especially those through remote sensing. As a physiological index, eye activity has been considered as an effective indicator of cognitive workload assessment, as it is sensitive to changes of mental efforts. Eye activity based physiological measures [1] [3] [4], such as fixation and saccade, eye blink, and pupillary response, can be detected unobtrusively through remote sensing.

The fact that changes of pupillary response occur during mental activity has long been known in neurophysiology, and it has been utilized to investigate cognitive workload. In an early work, Beatty investigates the pupillary response through

^{*} NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council.

experiments that involve tasks of short-term memory, language processing, reasoning and perception [1]. Pupillary response is shown to serve as a reliable physiological measure of mental state in those tasks. Usually head-mounted eye trackers are used to measure pupillary response during the task. It is not until recently that remote eye tracking has become a popular approach for cognitive workload measurement [6] [9]. In comparison with the head-mounted eye trackers, remote eye tracker enables the non-intrusive measurement of cognitive workload without interfering with user's activity during the tasks. Moreover, remote video eye tracker is shown to be precise enough for measuring the pupillary response.

Though empirical evidence from the literature has demonstrated that eye activity based physiological measure is a useful indicator of mental efforts, it could be influenced by noise factors unrelated to the cognitive task. For example, it is reported that pupillary response is sensitive to illumination condition, fatigue, and emotional state [7] [8] [12]. These factors restrict the practical usage of pupillary response for cognitive workload measurement. In this paper we investigate the feasibility of measuring cognitive workload based on pupillary response even under luminance changes.

2 Related Work

As a non-intrusive means of measuring cognitive workload, remote eye tracker has been demonstrated to be precise enough for recording detailed information of pupillary response. Klingner *et al.* examine the pupil-measuring capability of video eye tracking in [6]. In their experiment, cognitive workload is measured using a remote video eye tracker during tasks of mental multiplication, short-term memory, and aural vigilance. It has been observed that the remote eye tracker can detect subtle changes in pupil size induced by cognitive workload variation. Similarly, Palinko *et al.* also use remote eye tracking to measure cognitive workload in their experiment [9]. In a simulated driving environment, pairs of subjects are involved in spoken dialogues and driving tasks. The driver's cognitive workload is estimated based on the pupillometric measurement acquired from the remote eye tracker. The pupillometric measurement and the driving performance exhibit significant correlation, which suggests the effectiveness of cognitive load measurement by remote eye tracker.

Although the physiological measure based on remote eye tracker has exhibited its usage for cognitive workload measurement, its performance could be affected by various noise factors. Luminance condition is especially known as an important factor that influences the pupil size. Pomplun and Sunkara compare the effects of cognitive workload and display brightness on pupil dilation and investigate the interaction of both factors in [10]. They design a gaze-controlled human computer interaction task that involves three levels of task difficulty. In the experiments, each level of the difficulty is combined with two levels of background brightness (black and white), which results in six different trial types. The experiment results show that the pupil size is significantly influenced by both the task difficulty and the background brightness. There is a significant increase of pupil size when the workload demand becomes higher under both background conditions. However, the pupil size corresponds to the highest workload under white background is even smaller than that corresponds to the lowest workload under black background.

3 Experiment

3.1 Participants and Apparatus

Thirteen 24-to-46-year-old male subjects have been invited to participate in the experiment. All the subjects have normal or corrected-to-normal vision. Each subject receives a small-value reward for his participation.

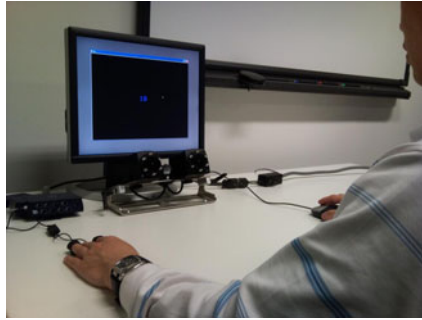


Fig. 1. Experiment setup

The pupillary response data of each subject is recorded with a remote eye tracker (faceLAB 4.5 of Seeing Machines Ltd), which operates at a sampling rate of 50 Hz and continuously measures the subject's pupil diameters. The skin conductance data is also recorded with a galvanic skin response (GSR) sensor (ProComp Infiniti of Thought Technology Ltd). However the analysis of the GSR data is out of the scope of this paper. Visual stimuli are presented on a 21-inch Dell monitor with a screen resolution of 1024 by 768 pixels. The experiment setup is demonstrated in Figure 1.

3.2 Experiment Design

Each subject is requested to perform arithmetic tasks under different luminance conditions. The arithmetic tasks have 4 levels of difficulty, and each level of task difficulty is combined with 4 levels of background brightness, which results in 16 different trial types in total.

For each arithmetic task, each subject is asked to sum up 4 different numbers sequentially displayed on the center of the screen, and then choose the correct answer on the screen through mouse input. The task difficulty depends on the range of numbers. For the first (lowest) difficulty level, each number is binary (0 or 1); for the second difficulty level, each number has 1 digit (1 to 9); for the third difficulty level, each number has 2 digits (10 to 99); for the fourth (highest) difficulty level, each number has 3 digits (100 to 999). Each number will be displayed for 3 seconds, and there is no time constraint for choosing the answer. Before the first number appears, different number of "X" will be displayed at the center of the screen for 3 seconds. The number of "X" corresponds to the number of digits for each arithmetic task.

During the experiment, the luminance condition varies when each subject performs arithmetic tasks. To produce different levels of luminance condition, luminance

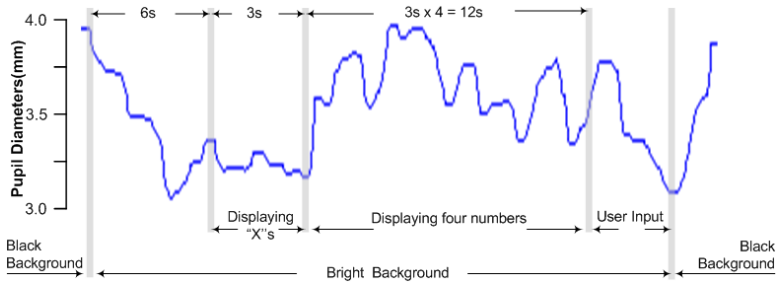


Fig. 2. Time setting of an arithmetic task

(grayscale value) of the background are set as 32, 96, 160 and 224 for the four levels of background brightness (L1, L2, L3, and L4), respectively. Black background will be displayed for 6 seconds before each arithmetic task. The time setting for each arithmetic task is depicted in Figure 2.

The experiment starts with a practice trial of which the data is not analyzed. Subsequently a one-minute resting data with black background is recorded before the test trials start. There are two tasks for each trial type, which results in 32 arithmetic tasks for each subject in the experiment. The tasks are presented randomly during the experiment. Once the subject finishes all the tasks, another one-minute resting data is also recorded. The whole experiment lasts about 25 minutes for each subject.

4 Analysis

In this section we analyze the correlation of the pupillary response and cognitive workload under different luminance conditions from the experimental data.

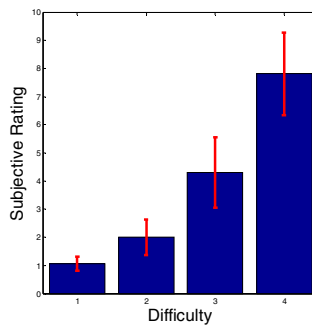


Fig. 3. Subjective rating of task difficulty

Figure 3 shows average subjective rating scores for the four levels of task difficulty from all the subjects. The scores range from 1 to 9, which correspond to the easiest and hardest tasks respectively. It can be known that there are significant differences between the subjective ratings of different task difficulty levels ($F=108.63$, $p<0.05$ in

ANOVA test), which indicates the overall effectiveness of cognitive workload manipulation in the experiment.

4.1 Coarse-Grained Analysis

For each subject, the pupillary response data of every arithmetic task during the experiment is examined. As a coarse-grained analysis, the average pupil diameter from the whole task period is used to characterize the cognitive workload. Figure 4 shows the average pupil diameters from that period under different levels of task difficulty and background brightness. It can be seen from the figure that the pupil diameter is influenced by the background brightness, in the sense that a smaller pupil diameter is usually observed under brighter background. On the other hand, the pupil diameter is also influenced by cognitive workload. For each background brightness level, the pupil diameter often increases when the task difficulty level becomes high. Together background brightness and cognitive workload could affect the pupil diameter. It can be observed that the pupil diameter at the highest task difficulty with highest background brightness is, in fact, smaller than that at the lowest task difficulty with lowest background brightness. This observation is consistent with previous empirical study that, luminance conditions take priority over cognitive demands in pupil diameter changes. Thus it is difficult to directly use the average pupil size or dilation to measure cognitive workload in the experiment.

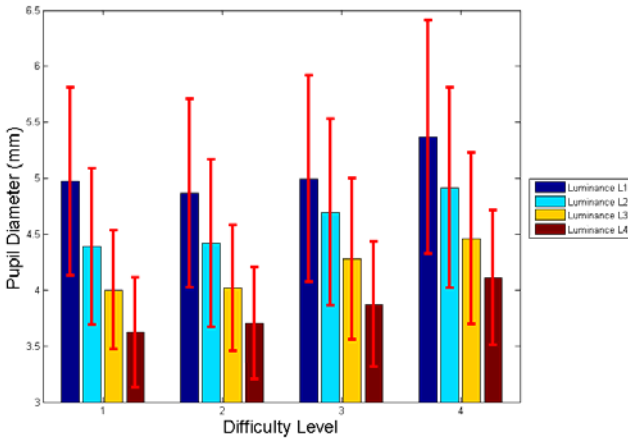


Fig. 4. Pupil diameter under different task difficulty levels and background brightness conditions

The above analysis shows that the coarse-grained measures of pupillary response could not effectively measure cognitive workload under luminance changes. To overcome this problem, we propose a fine-grained analysis of pupillary response in the following section. It is expected that the dynamic characteristics of cognitive process could be reflected by the fine-grained measures of pupillary response, which will improve cognitive workload measurement under complex environments.

4.2 Fine-Grained Analysis

For a fine-grained analysis of pupillary response, the 12-second task period is divided into smaller-size intervals. As shown in Figure 5, we examine five 3-second intervals corresponding to different stages of the cognitive process when performing the task. We denote X as the interval for the “X” displaying interval, and N1, N2, N3 and N4 as the four 3-second number displaying intervals respectively. Additionally, we also examine two 6-second intervals based on N1, N2, N3 and N4. Let M1 be the first 6-second of number displaying interval, and M2 the second 6-second interval. The setting of task intervals can be found in Figure 5. On the basis of these interval definitions, there are 6 intervals for each arithmetic task. We measure the average pupil diameters from these 6 intervals.

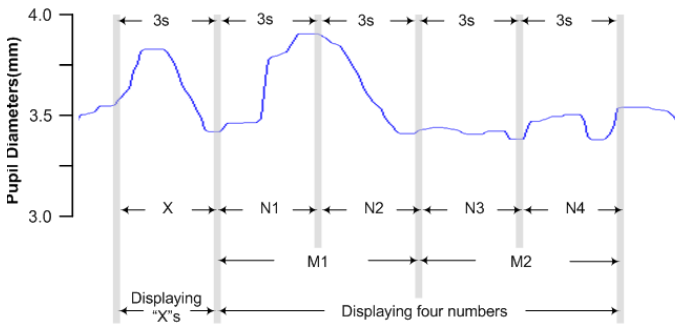


Fig. 5. The setting of task intervals for fine-grained analysis

For each task, average pupil diameter is measured for all the intervals. To reduce the influence of luminance condition, we normalize the measurement values for interval N1, N2, N3, N4, M1, and M2 using average pupil diameter d_X from the X interval, as there is no cognitive workload involved in that interval. Let d_N be one pupil size measurement from one task interval, its normalized measurement value is defined as $d_n = \frac{(d_N - d_X)}{d_X}$.

Figure 6 demonstrates the distributions of measurement values from M1 and M2 under different task difficulty levels. The figure demonstrates the characteristics of pupillary response at different time intervals. Even under the influence of luminance changes, the measurement values increase as the task difficulty increases. Such trend is more significant in the measurement values from M2 ($F=3.93, p<0.05$ in ANOVA test). We further examine interval M2 by studying the measurement values from N3 and N4, which are shown in Figure 7. As shown in Figure 7, the trend of the increase in the measurement values with increasing task difficulty is more significant in N4 ($F=3.43, p<0.05$).

In addition to the above analysis, cognitive workload classification has also been investigated using measurements from different intervals. We employ a decision tree-based classification scheme [11] to classify the cognitive workload. Specifically, given the measurement values from different classes, a threshold is estimated such

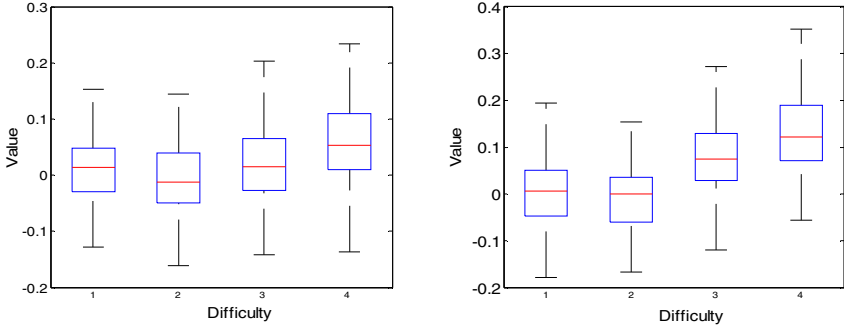


Fig. 6. Box plots of measurement values (sample minimum, lower quartile, median, upper quartile, and maximum) under different task difficulty levels: (left) M1, (right) M2

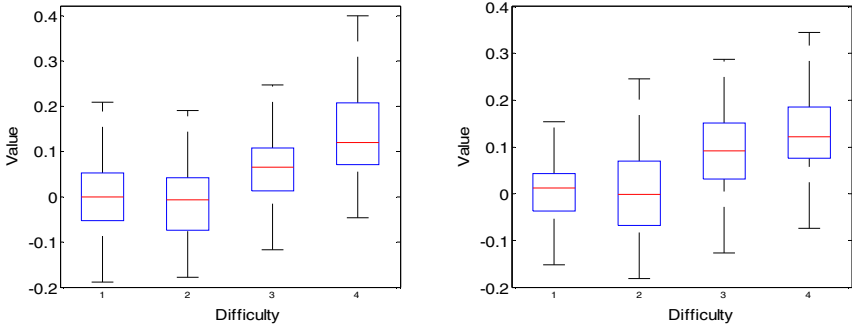


Fig. 7. Box plots of measurement values (sample minimum, lower quartile, median, upper quartile, and maximum) under different task difficulty levels: (left) N3, (right) N4

that maximum information gain can be achieved by splitting the data using that threshold. One threshold is needed for two-class classification while three thresholds are required for the four-class classification. We conduct both two-class classification (task difficulty 1, 2 vs. task difficulty 3, 4) and four-class classification of cognitive workload. The classification results are shown in Table 1. As shown in Table 1, M2 outperforms M1 for both two-class and four-class classification. N4 achieves the highest performance in both tasks. The measurements from different intervals reveal the dynamic characteristics of pupillary response at different stages of cognitive process, which can be utilized to improve the performance of cognitive workload assessment under complex environments.

Table 1. The classification results of different pupillary response measurements

Pupillary Measurements	M1	M2	N3	N4
Two-class Classification	59.3%	71.6%	68.9%	72.7%
Four-class Classification	36.6%	41.7%	43.0%	43.9%

6 Conclusion

This work investigates the measurement of cognitive workload through remote eye tracking under the influence of luminance condition. We study the characteristics of pupillary response, by analyzing the measurements acquired from different stages of cognitive process. The experimental results demonstrate the feasibility of cognitive workload measurement under complex environments using the proposed fine-grained analysis. Our future work will be applying machine learning techniques to improve fine-grained analysis for cognitive workload measurement.

References

1. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91, 276–292 (1982)
2. Cain, B.: A review of the mental workload literature. Technical Report, Defence Research and Development Canada Toronto (2007)
3. Chen, S., Epps, J., Ruiz, N., Chen, F.: Eye activity as a measure of human mental effort in HCI. In: *International Conference on Intelligent User Interfaces*, pp. 315–318 (2011)
4. Fogarty, C., Stern, J.: Eye movements and blinks: Their relationship to higher cognitive processes. *International Journal of Psychophysiology* 8, 35–42 (1989)
5. Grootjen, M., Neerinx, M., Weert, J.: Task-based interpretation of operator state information for adaptive support. In: Schmorrow, D., Stanney, M., Reeves, M. (eds.) *Foundations of Augmented Cognition*, 2nd edn., pp. 236–242 (2006)
6. Klingner, J., Kumar, R., Hanrahan, P.: Measuring the task-evoked pupillary response with a remote eye tracker. In: *Eye Tracking Research and Applications Symposium*, pp. 69–72 (2008)
7. Kramer, A.: Physiological metrics of mental workload: A review of recent progress. In: Damos, D. (ed.) *Multiple-Task Performance*, pp. 279–328. Taylor and Francis, Abington (1991)
8. Marshall, S.: The index of cognitive activity: Measuring cognitive workload. In: *IEEE Human Factors Meeting*, pp. 7-5–7-9 (2002)
9. Palinko, O., Kun, A., Shyrovkov, A., Heeman, P.: Estimating cognitive load using remote eye tracking in a driving simulator. In: *Eye Tracking Research and Applications Symposium*, pp. 141–144 (2010)
10. Pomplun, M., Sunkara, S.: Pupil dilation as an indicator of cognitive workload in human-computer interaction. In: *International Conference on Human-Computer Interaction*, pp. 542–546 (2003)
11. Quinlan, J.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Francisco (1993)
12. Xu, J., Wang, Y., Chen, F., Choi, H., Li, G., Chen, S., Hussain, S.: Pupillary response based cognitive workload index under luminance and emotional changes. In: *Annual Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1627–1632 (2011)