35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

An Investigation of Pupil-based Cognitive Load Measurement with Low Cost Infrared Webcam under Light Reflex Interference

Siyuan Chen, Julien Epps, and Fang Chen

Abstract— **Using the task-evoked pupillary response (TEPR) to index cognitive load can contribute significantly to the assessment of memory function and cognitive skills in patients. However, the measurement of pupillary response is currently limited to a well-controlled lab environment due to light reflex and also relies heavily on expensive video-based eye trackers. Furthermore, commercial eye trackers are usually dedicated to gaze direction measurement, and their calibration procedure and computing resource are largely redundant for pupil-based cognitive load measurement (PCLM). In this study, we investigate the validity of cognitive load measurement with (i) pupil light reflex in a less controlled luminance background; (ii) a low-cost infrared (IR) webcam for the TEPR in a controlled luminance background. ANOVA results show that with an appropriate baseline selection and subtraction, the light reflex is significantly reduced, suggesting the possibility of less constrained practical applications of PCLM. Compared with the TEPR from a commercial remote eye tracker, a low-cost IR webcam achieved a similar TEPR pattern and no significant difference was found between the two devices in terms of cognitive load measurement across five induced load levels.**

I. Introduction

Cognitive load measurement aims to evaluate the working memory ability during a task [1], therefore it can be used to assess the memory function in cognitive tests to assist the diagnosis of cognitive impairment and some mental disorders [2], e.g. schizophrenia [3]. Meanwhile, cognitive skills are of the essence to daily living and social activities, thus cognitive load measurement can be a convenient means to screen discharge patients and evaluate their responses to treatment [3]. Such applications of cognitive technology call for a low cost, convenient and accurate measurement method. Among existing measures, task-evoked pupillary response (TEPR) has been found as a strong index of cognitive load [4]. Compared with using reaction time for detecting cognitive dysfunction in computerised cognitive tests [12], a pupil-based cognitive load measurement (PCLM) based e.g. on a smartphone can certainly provide benefit in terms of objectiveness, portability, and being a physiological measure. However, TEPR measurement to date has occurred only in well-controlled lab environments.

The main barriers for PCLM are the susceptibility to pupil light reflex (PLR) and the lack of a cheap and dedicated device for pupil measurement. As empirical work has demonstrated, light can cause the pupil to dilate and constrict in a range of around 3 mm [5] while changes in TEPR can only be as large as around 0.6 mm [2,4,6] under constant background. Therefore, the global luminance (i.e. due to the background change) and the local luminance (i.e. due to gaze shifting) of the visual field are required to be uniform across experimental conditions [4,6]. Other factors, such as near reflex (due to the distance to screen changes) and arousal state [2,4] can also influence pupillary response. Baseline subtraction and average pupillary response are common approaches to alleviate some of the above confounders [2,4,6], but the effectiveness of using these approaches without uniform luminance background in tasks has not been investigated. In other words, it is significant to know in practice whether PCLM can still be applied in a less controlled luminance background.

Ideally, if non-cognitive variability is the same during both the trial and the baseline, we can obtain the pupillary response resulting from cognition after baseline subtraction [4]. In the literature, baseline pupil size is usually collected before the experiment, in a non-task state, when participants are merely looking at a blank/pre-task screen without a task goal. Even in this situation, the pupillary response is not stable; therefore, an average pupil size over a few seconds is often used as a baseline. But the duration of the baseline region varies considerably between different studies, e.g. 10 s in [8], 0.4 s in [6] and 0.2 s in [9]. Furthermore, it is not easy to find out whether the baseline is reliable, since participants can think about something else in the absence of a driven task goal, and their arousal level can be different before and during the task-state. Thus, different baseline selections might result in different results.

A low cost and convenient device can certainly proliferate the applications of PCLM. However, to date, in the literature, TEPR determination is heavily dependent on commercial eye trackers. Although some commercial eye trackers can offer an output of pupil size, they are expensive. More importantly, they are dedicated to gaze direction measurement since the aim of most eye tracker is to obtain a precise center of the pupil rather than a precise pupil size. For the purpose of the former, a few opposite points on the pupil boundary can determine the center of the pupil through a fitted circle or as the center of the longest line by scanning [11]. For commercial eye trackers, the implementation details of pupil measurement are often not available. Moreover, a large computing resource is used to transform the distance between the centers of the pupil and the glint to an absolute eye position by the parameters obtained in the calibration procedure [10]. Thus, for TEPR measurement,

S. Chen and J. Epps are with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia, and ATP Research Laboratory, National ICT Australia. (Phone: +61 2 9385 4803; fax: +61 2 9385 5993; e-mail: siyuan.chen@student.unsw.edu.au, j.epps@unsw.edu.au).

F. Chen is with ATP Research Laboratory, National ICT Australia, Sydney, NSW 2015, Australia. (e-mail: fang.chen@nicta.com.au).

the tedious calibration procedure and computing resources required for gaze calculation are redundant.

In this study, we carefully select the baseline for subtraction and averaging to reduce the PLR and measurement noise, and compare TEPR obtained by a remote eye tracker under a gray background with that under image backgrounds to investigate the efficacy of PCLM in a less controlled luminance background. We also used low-cost head-mounted IR webcams to record eye activity without any stereo calibration and compared the TEPR obtained by an IR webcam with that obtained by a remote eye tracker to demonstrate the validity of using a low-cost device for PCLM. In a previous study [6], TEPR obtained by a remote eye tracker was found able to index cognitive load, although head motion might cause variations in the estimated pupil size.

II. Method

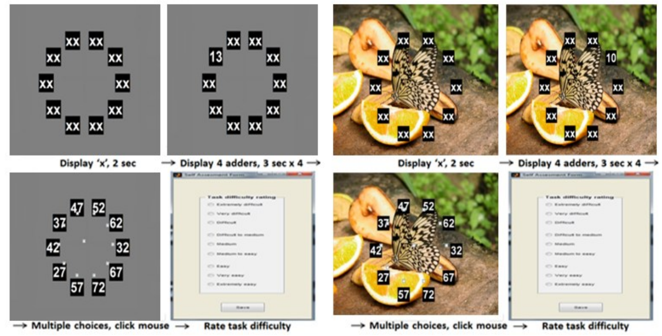
A. Experiment setup

A mental arithmetic task is often employed in cognitive load studies and TEPR has been successfully found to have a distinct pattern in a two-number multiplication task with three levels of difficulty, as pupil size was found increasing in a more difficult task [4,6]. However, to avoid gaze-shift affected PLR, the luminance of their task background was carefully controlled to be constant. We also employed a mental arithmetic task but in three different conditions, (i) constant gray background with FaceLab4 - a remote eye tracker ('gray_eye_tracker'), where there is little PLR due to global luminance and minimum PLR due to gaze shift; (ii) image background with the remote eye tracker ('image_eye_tracker'), where both the PLR due to different global luminance and the PLR due to gaze shift occur; and (iii) constant gray background with IR webcams ('gray_webcam') for pupillary response recording.

In the arithmetic task, participants were required to add four numbers that were sequentially displayed every 3 s on the screen (step 2 in Figure 1 (a) and (b)), use the mouse to select the correct answer from 10 displayed choices (step 3 in Figure 1 (a) and (b)), and then rate the current task difficulty (step 4 in Figure 1 (a) and (b)). Before the beginning of task stimulus display, there was a 2-s period, showing 10 placeholders in the background (step 1 in Figure 1 (a) and (b)) in order to let the eye adapt to the background luminance. The PLR due to global luminance can be expected to reduce to a minimum after around 2 s [9]. Task difficulty levels were regulated by the number of digits for addition and carries produced by addition. For each trial, the four integers were randomly selected from (i) {0,1}, (ii) {1,…,5}, (iii) {5,…,9}, (iv) {10,…,19}, (v) {84,…,93}.

In the gray_eye_tracker experiment, the task was conducted in a gray background as shown in Figure 1 (a) and pupil size was collected from 15 participants (7 females and 8 males; age M=26.8, SD=7.2) by a remote eye tracker (FaceLab4, shown in Figure 2 (a)), running at 60 Hz. Each participant completed 10 trials (5 levels × 2 repetitions).

In the image_eye_tracker experiment, the task settings were the same as in the gray_eye_tracker experiment, except different images were used as the background in each trial, as shown in Figure 1 (b). Each participant completed 60 trials (5 levels × 12 repetitions).



(a) (b)

Figure 1. An arithmetic task with a constant gray background (a) in the gray_eye_tracker and gray_webcam experiments and an image background (b) in the image_eye_tracker experiment to investigate the validity of cognitive load measurement under light reflex interference. There were four steps, indicated by arrows, in the arithmetic task and the 12 s interval in step 2 is of most interest for comparison. The rating form in step 4 was a 9-point scale from extreme easy to extreme difficult.

(a) remote (b) near-field

Figure 2. A remote eye tracker (a) and head-mounted IR webcams (b) were used to obtain the pupillary response for cognitive load measurement. (a) was used in the experiment of gray_eye_tracker and image_eye_tracker and (b) was used in the experiment of gray_webcam.

In the gray_webcam experiment, the task had a gray background, identical to that in the gray_eye_tracker experiment, but was completed by a different 22 participants (9 females and 13 males, age: M=26.8, SD=4.0). They were required to wear a pair of glasses frames with two IR webcams mounted and tracking each pupil, as shown in Figure 2 (b). Video was recorded at 30 Hz. Each participant completed 150 trials (5 levels × 30 repetitions). As experiments were conducted in different time and improved by having more participants and trials, all data were used.

B. Data processing for pupil size

In the gray_eye_tacker and image_eye_tracker experiment, pupil size was directly measured by the eye tracker and was linearly interpolated during zero pupil size intervals and then low-pass filtered at around 4 Hz cut off frequency [2]. In the gray_webcam experiment, pupil size was recorded as the length (in pixels) of the major axis of a fitted ellipse in a binary image, produced by thresholding each frame of the video sequences [11]. Pupil size was linearly interpolated during blinks, which were detected by a dual-ellipse algorithm [11] and manually checked by superimposing the ellipse on the video. The pupil size was then passed through a median filter with a length of 3 frames to remove the noise caused by rapid eye movements. Pupil size was finally converted to millimeters by the ratio of the true eye length and the eye length in the video.

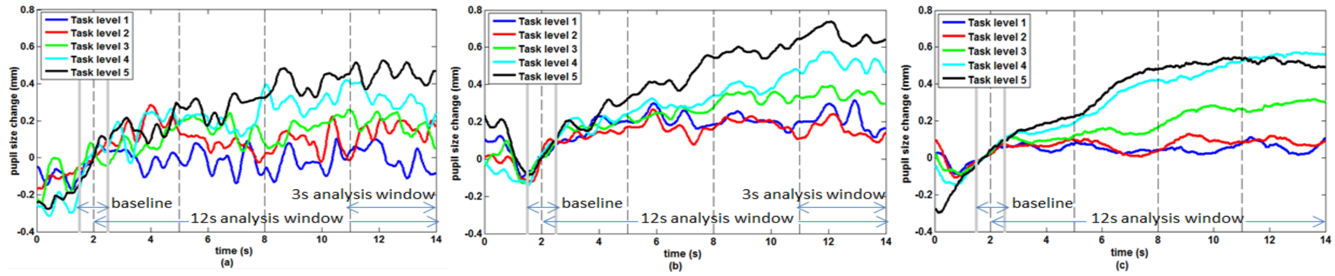


Figure 3. (a) Average TEPR of five cognitive load levels obtained by a remote eye tracker over the 14 s with gray background and (b) with image background. (c) Average TEPR of five cognitive load levels obtained by an IR webcam over the 14 s with gray background. The average pupil size over the 1 s duration between the two gray vertical lines was used as a baseline for subtraction. This region was chosen because the eye was almost adapted to the global luminance of the task background, and it was at the instant of task beginning, close to task state but involving minimum cognitive load. Therefore, at least 2 s of task background presence was required before task beginning in practice for a baseline region and we need to know when a stimulus begins in order to define one.

C. *Baseline pupil size selection*

Different to the baseline selected before the experiment with a blank background, where the associated factors such as arousal state, luminance to the eye and cognitive activity during resting might not be same as these during tasks, we considered averaging pupil size over 1 s around the onset of the first addend in each task as the baseline, as shown in Figure 3. The reason for this baseline region is that (i) the eye has almost been adapted to the global luminance of the background since it has been presented for nearly 2 s; (ii) it is the closest moment to the arousal state throughout the task since it is at the beginning of the task; (iii) it is relatively stable because participants paid attention to the task stimulus, involving minimum cognitive load. The TEPR is an average phasic pupil size (*PS*) change [4] (average to minimize the PLR effect due to gaze shift), obtained by

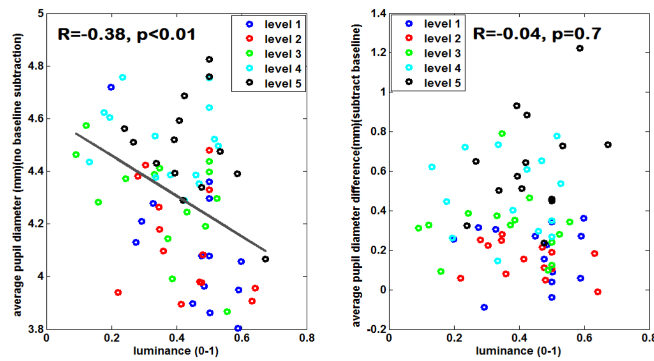
$$TEPR = \sum_{t=t_1}^{t=t_2} \Big(PS - \underbrace{\sum_{t=1.5}^{t=2.5} PS}_{baseline} \Big) \qquad (1)$$

where the pupil size during the 12 s, t_1=2 and t_2=14 s, and during the 3 s, t_1=11 and t_2=14 s, as shown in Figure 3, are of interest and calculated for later analysis.

D. *ANOVA Analysis*

To examine whether the obtained TEPR is still valid for PCLM, that is, whether PLR still has a great effect after baseline subtraction in the image_eye_tracker experiment compared with that in the gray_eye_tacker experiment, we conducted two repeated two-way ANOVA tests to see the significant effects on cognitive load, average luminance of each image, and their interaction under two conditions: with and without baseline subtraction. We set 0.05 as the critical *p* value. For those within-subject tests that violated the assumption of sphericity, the degrees of freedom were corrected by Greenhouse-Geisser, as indicated by fractions.

We used the average TEPR over 3 s for the repeated ANOVA analysis. The last 3 s is the most difficult part of the task, where cognitive load has the strongest effect on pupil size, therefore the pupil size change due to arousal state resulted by the image was minimal [2] to see the PLR effect. To calculate the luminance of an image, we firstly converted an RGB image to an intensity image by 0.2989 * Red + 0.5870 * Green + 0.1140 * Blue to obtain the value of each pixel in Matlab, and averaged them to provide the average luminance of each image. Then the luminance values were grouped into six 0.1-interval bins centred at 0.15, 0.25, 0.35, 0.45, 0.55, 0.65. All the five workload levels shared the 4 levels of luminance, 0.25, 0.35, 0.45, 0.55, therefore 4 levels of luminance were used.



(a) **(b)**

Figure 4. Average pupil size over the 11th to 14th s vs. average luminance for each of 14 different backgrounds across 15 participants in five cognitive load levels without baseline subtraction (a) and with baseline subtraction (b). Irrespective of induced cognitive load level, the pupil size in (a) has a general trend (indicated by the gray line using robust regression analysis) of being a function of luminance, that is, the pupil size is larger when the luminance is lower. However, the significant correlation is lost after baseline subtraction in (b). This demonstrates that most variability of pupil size is now due to cognitive load. Therefore, after baseline subtraction, the cognitive load can be more accurately classified using TEPR.

To evaluate the comparative performance of IR webcams and the remote eye tracker for PCLM, we conducted a two-way ANOVA test to examine whether there is significant difference in TEPR obtained by the two devices for the five levels of cognitive load, and whether the cognitive load effect interacts with different devices.

III. Results and Discussion

Figure 3 and Table I present the TEPR patterns of the five levels of cognitive load in the three experiments. They demonstrate that the average TEPR with baseline subtraction achieved similar patterns, that is, the pupil size increased further during a more difficult addition task, in spite of the PLR effect with a non-constant background or the differences between the two pupil size measurement devices. Meanwhile, the TEPR across the five cognitive load levels in image_eye_tracker was larger than that in the other two experiments, as shown in Table 1. It is likely due to the arousal effect induced by image backgrounds throughout task execution [2]. Although there was likely to have been non-task-induced variability, we can see that the cognitive load is the dominating factor since the pupil size is a function of cognitive load, increasing with higher load levels, as shown in Figure 3 and Table I.

TABLE I. THE MEAN AND STANDARD DEVIATION OF TEPR ACROSS ALL TRIALS AND ALL PARTICIPANTS OVER 12 S IN THE THREE EXPERIMENTS.

TEPR (mm)	gray_eye_tacker	image_eye_tracker	gray_web cam
Level 1	-0.01 (0.16)	0.19 (0.11)	0.03 (0.10)
Level 2	0.11 (0.18)	0.16 (0.12)	0.05 (0.14)
Level 3	0.14 (0.21)	0.27 (0.15)	0.16 (0.17)
Level 4	0.24 (0.26)	0.34 (0.13)	0.33 (0.27)
Level 5	0.32 (0.26)	0.46 (0.18)	0.36 (0.28)

TABLE II. TWO-WAY (LUMINANCE LEVELS (4) × COGNITIVE LOAD LEVELS (5)) REPEATED ANOVA TEST OF THE AVERAGE PUPIL SIZE OVER THE LAST 3 S DURING TASKS FOR THE GRAY_EYE_TRACKER AND IMAGE_EYE_TRACKER CONDITIONS.

TEPR without baseline subtraction	TEPR with baseline subtraction
Luminance levels: $F(1.9,26.2)=9.33$, $p<0.01$. Cognitive load levels: $F(2.4, 33.3)=44.46$, $p<0.01$. Luminance × cognitive load: $F(4.5,62.4)=6.58$, $p<0.01$.	Luminance levels: $F(2.0,27.5)=1.27$, $p=0.3$. Cognitive load levels: $F(2.8, 38.9)=28.59$, $p<0.01$. Luminance × cognitive load: $F(5.5,76.4)=2.13$, $p=0.06$.

TABLE III. TWO-WAY (DEVICE (2) × COGNITIVE LOAD LEVELS (5)) ANOVA TEST OF AVERAGE PUPIL SIZE OVER THE 12 S TASK DURATION FOR THE GRAY_EYE_TRACKER AND GRAY_WEBCAM CONDITIONS.

TEPR with baseline subtraction
Device: $F(1,175)=0.56$, $p=0.45$. Cognitive load levels: $F(4, 175)=15.53$, $p<0.01$. Device × cognitive load: $F(4,175)=0.63$, $p=0.64$.

Specifically, Figure 4 and Table II show that with the baseline selection and average methods, the PLR effect was significantly reduced. In Figure 4, in each cognitive load level, the 14 points represent the average pupil size over 3 s in the 14 trials across 15 participants. Before baseline subtraction, pupil size was affected by the average luminance of images, as shown in Figure 4 (a). From Table II, there are significant effects of pupil size on cognitive load, average luminance of images and the interaction between them. However, after baseline subtraction, as shown by Figure 4 (b) and Table II, the average luminance of images effect was not significant but the cognitive load effect was still significant and the interaction between them was not significant any more. These results suggest that the proposed baseline method and average phasic pupil size can effectively reduce PLR for PCLM.

Furthermore, cognitive load can be successfully measured with a low-cost IR webcam. As Table III shows, there is no significant difference between the TEPR obtained from the commercial remote eye tracker and the IR webcam. Also the cognitive load effect does not depend on the devices as there is no significant effect on the interaction. Moreover, the significant effect on cognitive load demonstrates the promise of using a low-cost and dedicated processing system to proliferate the applications of PCLM.

IV. CONCLUSION

In this study, we demonstrated that in a less controlled luminance background, with appropriate baseline selection and average method, the pupillary light reflex effect can be significantly reduced. Pupillary response obtained from a low-cost head-mounted infrared webcam is also suitable for cognitive load measurement, compared with a commercial eye tracker, since no significant difference was found between the two devices. Obtaining pupillary response from a cheap device and in less controlled luminance conditions allows this physiological measure to be applied in a more realistic environment, which is a small but significant step for the pupil-based measurement. Typical applications include assessing working memory function in cognitive tests or evaluating treatment responses associated with cognition. However, at least 2s eye adaptation time is needed after any significant luminance change and the timestamp of the beginning of the task is needed for the baseline region. Future work will focus on automatic methods to detect the changes in luminance and task endpoints.

ACKNOWLEDGMENT

We thank all the volunteers. This research was supported by the Asian Office of Aerospace Research & Development, through Grant No. FA2386-12-1-4049.

REFERENCES

[1] F. Paas, J. E. Tuovinen et al., “Cognitive load measurement as means to advance cognitive load theory” in *Educational Psychologist*, vol. 38, no.1, 2003, pp. 63-71.

[2] S. Chen, J. Epps, “Automatic Classification of eye Activity for cognitive load measurement with emotion interference”, *Computer Methods and Programs in Biomedicine*, vol. 110, 2013, pp. 111-124.

[3] W. D. Spaulding, S. K. Fleming et al., “Cognitive functioning in schizophrenia: Implications for psychiatric rehabilitation”, *Schizophrenia Bulletin*, vol. 25, 1999, pp. 275-289.

[4] J. Beatty, “Task-evoked pupillary response, processing load, and the structure of processing resources”, *Psychological Bulletin*, vol. 91, 1982, pp. 276-292.

[5] B. Winn, D. Whitaker et al., “Factors Affecting Light-Adapted Pupil Size in Normal Human Subjects”, *Investigative Ophthalmology & Visual Science*, vol. 35, no. 2, 1994, pp. 1132-1137.

[6] J. Klingner, B. Tversky, P. Hanrahan, “Effects of visual and verbal presentation on workload in vigilance, memory and arithmetic tasks”, *Psychophysiology*, vol. 48, no. 3, 2011, pp. 323-332.

[7] A. F. Kramer, “Physiological metrics of mental workload: A review of recent progress”, *In:* Damos, D. L. (ed.) *Multiple-task Performance.* London: Taylor & Francis Ltd, 1991, pp. 279-317.

[8] S. T. Iqbal, X. S. Zheng, B. P. Bailey, “Task-Evoked Pupillary Response to Mental Workload in Human-Computer Interaction”, *Ext. Abs CHI 2004*, pp. 1477-1480.

[9] S. R. Steinhauer, R. Condray, A. Kasparek, “Cognitive modulation of midbrain function: task-induced reduction of the pupillary light reflex”, *Int’ Journal of Psychophysiology*, vol. 39, 2000, pp. 21-30.

[10] D. W. Hansen, Q. Ji, “In the eye of the beholder: a survey of models for eyes and gaze”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 32, no. 3, 2010, pp. 478-500.

[11] S. Chen, J. Epps, “An efficient and robust method for pupil size estimation and blink detection from near-field video sequences,” *ICIP2013*, submitted for publication.

[12] A. Collie, D. Darby, P. Maruff, “Computerised cognitive assessment of athletes with sports related head injury”, *British Journal of Sports Medicine*, vol. 35, 2001, pp. 297–302.