Measuring the Task-Evoked Pupillary Response with a Remote Eye Tracker

Jeff Klingner, Rakshit Kumar, and Pat Hanrahan
Stanford University
353 Serra Mall, Stanford, CA 94305, USA (klingner, rakshit, hanrahan)@stanford.edu

Abstract

The pupil-measuring capability of video eye trackers can detect the task-evoked pupillary response: subtle changes in pupil size which indicate cognitive load. We performed several experiments to measure cognitive load using a remote video eye tracker, which demonstrate two extensions to current research in this area. First, we show that cognitive pupillometry can be extended from head-mounted to remote eye tracking systems. Second, we demonstrate the feasibility of a more fine-grained approach to analyzing pupil size data gathered with an eye tracker, which provides more detail about the timing and magnitude of changes in cognitive load.

Keywords: eye tracking, remote eye tracking, pupillometry, pupil, cognitive load, task-evoked pupillary response

CCS: B.4.2 Input/Output Devices, H.5.2 User Interfaces

1 Introduction

Video eye trackers determine a user's gaze direction by observing the eye(s) with a high-resolution camera and measuring the relative positions of the pupil center and one or more corneal reflections [Duchowski 2003]. The high-resolution image of the eye necessary for gaze tracking also enables a high-precision measurement of the size of the pupil. This additional measurement is useful, because small, short-term changes in pupil size can indicate cognitive load.

1.1 The Task-evoked Pupillary Response

Whenever somebody recalls something from memory, pays close attention, parses a complicated sentence, or otherwise thinks hard, her pupils dilate slightly and then return to their previous size within a few seconds of completing the mental work [Beatty 1982, Beatty & Lucero-Wagoner 2000]. This reaction, called the task-evoked pupillary response (TEPR), is small (usually less than 0.5 mm dilation), involuntary, and reliably associated with a broad set of cognitive processes that are characterized as cognitive load. These include short and long-term memory access, mental arithmetic, sentence comprehension, vigilance, and visual and auditory perception tasks. (See Beatty & Lucero-Wagoner [2000], pp. 142-162 for a summary of this work.)

ETRA 2008, Savannah, Georgia, March 26–28, 2008.
© 2008 ACM 978-1-59593-982-1/08/0003 $5.00

1.2 Prior TEPR research using eye trackers

In the fields of cognitive science and psychophysiology, custom pupillometry systems are used to measure the TEPR [Alexandris et al. 1991]. Recently though, several human-computer interaction research groups have begun to use the pupillometric capability of head-mounted video eye trackers to measure cognitive load. Marshall [2002] applied a wavelet decomposition to the pupil size signal in order to estimate the average number of abrupt discontinuities in pupil size per second, and used this measure as a general index of cognitive activity. Pomplun and Sankara [2003] described how to correct bias in observed pupil size based on gaze direction. Maloney et al. [2006] used differences in pupil responses to distinguish older, visually impaired subjects from younger, visually healthy control groups performing a drag-and-drop task. Iqbal et al. [2005] applied eye tracker pupillometry to show that mental workload drops at task boundaries in a hierarchical task model and can be used as an indicator of interruptability.

Contributions: We show that existing work on measuring cognitive load with eye trackers can be extended in two ways. First, we demonstrate that remote video eye trackers can be used for pupillometric measurement of cognitive load. Second, we show that eye trackers can be used with an experimental technique that focuses on the detailed timing and magnitude of short-term pupillary responses, rather than a simple aggregated measurement over a long period of time. We demonstrate the feasibility of both extensions by replicating prior studies which used specialized pupillometers.

1.3 TEPR measurement with remote eye trackers

All existing work utilizing eye trackers to estimate cognitive load has used head-mounted cameras, which provide high precision but can be cumbersome and annoying to users. Marshall reported that some of her experimental subjects were bothered by wearing a head-mounted eye tracker, and that this may have distorted some of her results [Marshall 2002]. In addition, care must be taken with head mounted cameras to keep the head band from slipping.

Remote eye trackers, which employ desktop or display-mounted cameras, eliminate the need for distracting head-mounted equipment or chin rests. This enables unobtrusive measurement of cognitive load during normal computer use with a computer system that resembles standard desktop models. Although remote systems typically have less precision and are subject to more measurement noise than head-mounted systems, we found that our remote eye tracker can be used for measuring cognitive load. Section 3 describes some of the issues we encountered when adapting a remote eye tracker for this purpose.

1.4 Measuring the details of TEPR timing, duration, and shape

The task-evoked pupillary response does not occur reliably for any one single episode of mental effort. In order to measure it, observations of several pupillary responses must be combined [Beatty 1982, Beatty & Lucero-Wagoner 2000]. Current HCI researchers employing eye trackers do this by aggregating pupil diameter measurements over a long period of time, usually simplifying them into a single number and ignoring the details of timing, size, or shape of particular responses. For example, the index of cognitive activity [Marshall 2002] summarizes cognitive load as the frequency of sudden discontinuities in pupil size during an ongoing task. Similarly, Pomplun and Sunkara [2003] measure cognitive load as the average pupil size over a long task involving many visual searches. This time-aggregated style of data processing leads to a very robust but coarse measurement of cognitive load.

In contrast, research done in the field of cognition, which typically uses specialized pupillometers, deals with the unreliability of single pupil responses by aligning and averaging data over several identical short trials [Beatty 1982, Beatty & Lucero-Wagoner 2000]. This experimental method can only deal with short, simple task components, but it is able to measure the details of the onset timing, shape, magnitude, and duration of the pupillary response. The work of Iqbal et al. [2005] takes steps in this direction, measuring pupil diameter at specific points in time corresponding to sub-task boundaries.

The trial-aggregated fine-grained analysis method used in cognition research complements the time-aggregated coarse measurement of cognitive load currently used in HCI studies. We think that both approaches are valuable, and the experiments we ran demonstrate the feasibility of applying the more subtle trial-aggregated method to data gathered with eye trackers.

2 Experiments

To determine whether remote eye trackers are precise enough to measure the TEPR at all or to enable a trial-aggregated style of data analysis, we conducted several experiments, three of which we report here. The first two replicate prior cognitive pupillometry studies, to determine whether we could observe the same details of pupil response measured in those studies with custom pupillometry systems. The third experiment is original.

For all our experiments, we used a Tobii 1750 eye tracker, which has a sampling rate of 50 Hz [Tobii 2007]. The eye tracking precision of the Tobii is variable and depends on the distance from the camera to the eyes. We recruited eight subjects from an introductory undergraduate design course. Subjects with strong vision correction (more than 5 diopters) or astigmatism were excluded. We covered all the room's windows to maintain a constant ambient luminance. Stimuli for the three experiments reported here were all auditory, and we used a fixation target to keep subjects' gaze on the center of the screen while they listened. Each subject repeated each task 10-15 times. We encouraged genuine effort by giving subjects a monetary incentive proportional to their task performance.

2.1 Experiment One: Mental Multiplication

We first replicated an experiment performed by Ahern and Beatty on the TEPR for mental arithmetic [1979]. In each trial, subjects listened to two spoken numbers between 5 and 19. Five seconds after the second number was spoken, subjects were prompted to type the numbers' product into an on-screen keypad using the mouse. The use of an on-screen keypad kept users' eyes on the screen, in order to maintain pupil visibility and to avoid pupillary light reflexes caused by looking away from the screen. We observed a similarly-shaped pupil response to that reported by Ahern and Beatty (Figure 1).

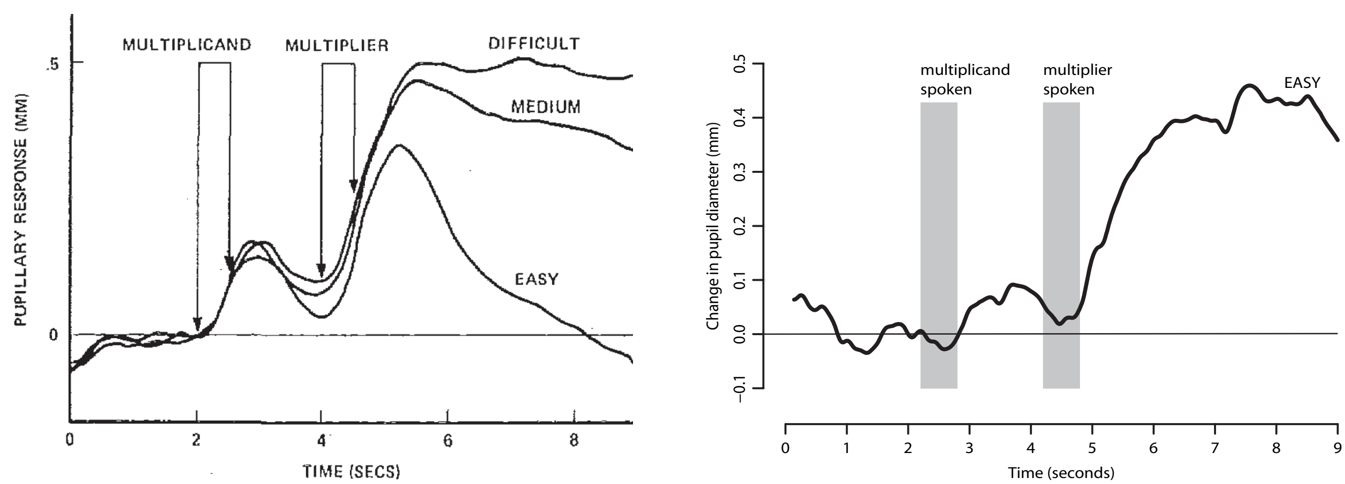

Figure 1: Pupillary response during the mental multiplication task. There is a small (0.1 mm) increase in pupil size as the multiplicand is committed to short term memory and a larger, longer-lasting increase after the subjects hear the multiplier and begin computing the product. The graph on the left is from Ahern and Beatty [1979], reprinted with permission from AAAS. The graph on the right shows the results from our replication of their experiment. The two graphs are aligned and plotted at the same scale. Although we gave problems at all three difficulties, the easy level was the only one for which we collected sufficient correct responses for analysis. The pupillary response we observed for these easy problems resembles the prior result for medium and difficult problems. We speculate that students in 1979 had more practice with mental arithmetic.

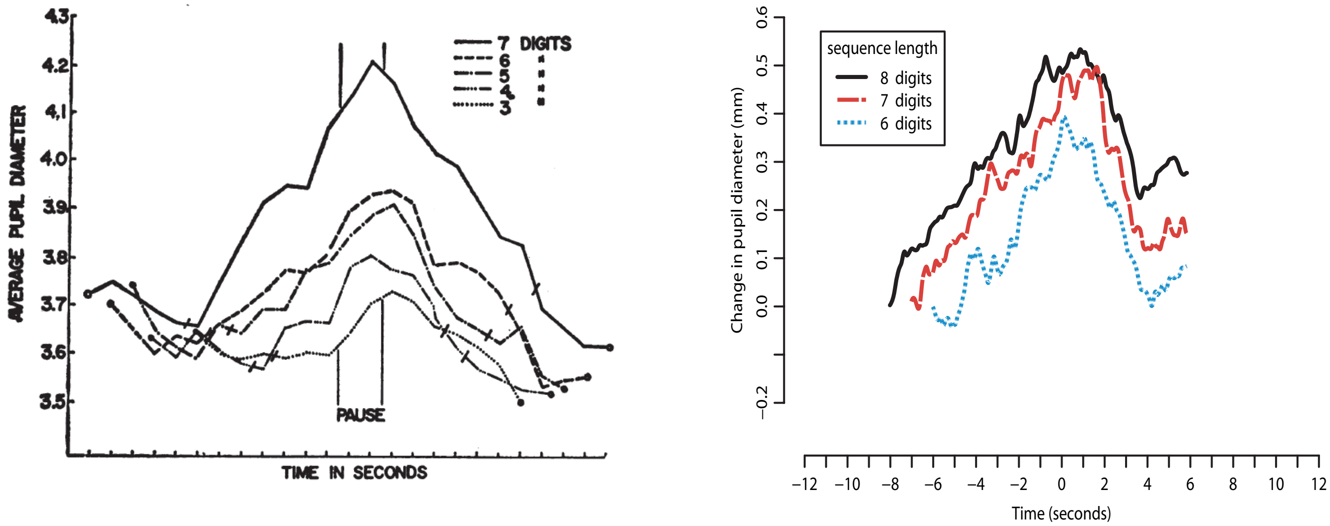

Figure 2: Pupillary response during the short-term memory task. The graph on the left is from Kahneman and Beatty [1966], reprinted with permission from AAAS; the graph on the right shows our data. The two graphs are aligned and plotted at the same scale. In both experiments, pupil diameter increases as the digits to be memorized are heard and encoded, peaks during the pause while they are retained, and declines as the subjects report them back. The magnitude of the response increases monotonically with the length of the memorized sequence. In the 1966 study, subjects repeated the sequence aloud, one digit per second, while in our study, the response was entered using an on-screen numeric keypad. This enabled a faster response and resulted in the observed steeper decline in pupil diameter in our results.

2.2 Experiment Two: Short-term memory

We also replicated a study by Kahneman and Beatty [1966], in which subjects were required to memorize and repeat back a sequence of digits spoken aloud, one per second. Our data appear somewhat noisier than the original study's, but we replicated the details in shape, magnitude, and timing of the short-term memory task TEPR (Figure 2).

2.3 Experiment Three: Aural vigilance

In a third experiment, we measured the TEPR for an aural vigilance task. Subjects listened to somebody counting up from 1 to 19 and were told that the counter might make a mistake by saying a number out of sequence instead of the correct number at 6, 12, or 18. Such mistakes were inserted randomly, and subjects were required to click the mouse as soon as they noticed a mistake. We observed sharp spikes in pupil diameter with consistent magnitude, onset timing, duration, and shape following all three mistake points.

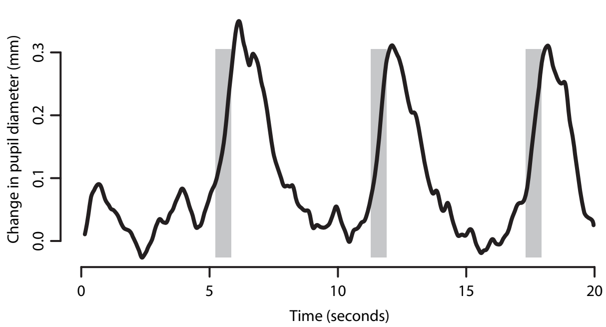

Figure 3: Pupillary response to an aural vigilance task. The grey bars mark moments when subjects needed to listen carefully and react quickly to mistakes in a spoken sequence of numbers.

3 Video pupillometry data processing issues

3.1 Smoothing

The pupil measurements made by eye trackers are rather noisy, and remote eye trackers are especially bad, because the freedom of head motion poses two additional problems: First, unless the eye-tracker's camera is actively pointed at the eyes, it must maintain a wider field of view, in which fewer pixels can be devoted to observing the pupil. Second, pupil measurements must be corrected for foreshortening by dividing the raw pixel-based pupil size by the distance from the camera to each eye. Drift, tremors, and non-spherical eye shape introduce noise into this distance measure [Duchowski 2003], which in turn causes a noisy pupil size signal.

Because this instrument noise is high-frequency, and pupils are known to dilate and constrict at low frequencies [McLaren et al. 1992], we smoothed the pupil size signal with a low-pass filter. To determine the appropriate cutoff frequency for this filter, we analyzed the correlation between the pupil size signal from the left and right eyes at different frequencies. Since the instrument noise is independent for the two eyes, we expected the noisy frequency components of the pupil signal to be uncorrelated. In contrast, the frequency components containing the true pupil size signal should be correlated, because they are driven by general cognitive activation, which affects both eyes [Beatty & Lucero-Wagoner 2000].

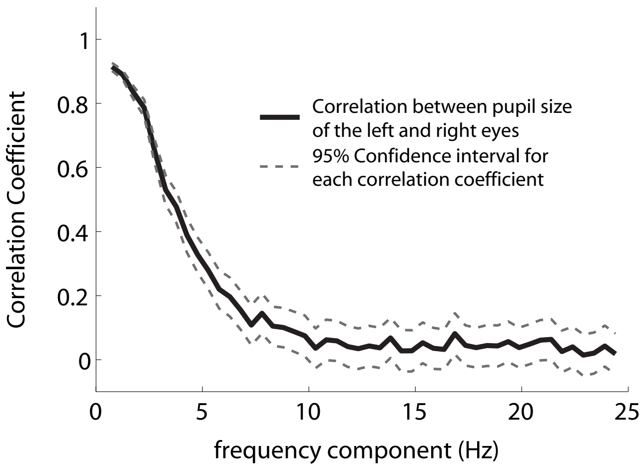

Figure 4: Correlation between the measured pupil size of the left and right eyes, broken down by frequency component. Pupil size signal components above about 10 Hz are uncorrelated. We therefore considered this part of the pupil signal to be noise and removed it with a low-pass filter.

To find the boundary between these two parts of the signal, we applied a band-pass filter with a bandwidth of 0.5 Hz and varying central frequencies to the pupil size signal of each eye separately to isolate their individual frequency components and computed the correlation between the left and right eyes at different frequencies (Figure 4).

3.2 Pixel-counting vs. ellipse-fitting

Pomplun and Sunkara [2003] reported a systematic dependence of pupil size on gaze direction. We replicated the ascending numeral visual search task they used to check for this bias but did not find it in our pupil measurements. We believe this is because the system used by Pomplun and Sunkara measured pupil size as the number of pixels encompassed by the pupil image, and optical perspective causes this size to vary with gaze direction. The Tobii 1750 we used instead measures pupil size as the length of the major axis of an ellipse fitted to the pupil image. This method is not affected by perspective distortion, though it is still subject to small errors caused by non-circular pupil shape. We recommend either using the ellipse-fitting method or calibrating for the bias as per Pomplun and Sunkara.

3.3 Controlling Light

For the experiments reported above, we also conducted trials while changing the brightness of the screen in front of the subjects, to see how TEPRs mixed with the pupillary light reflex. We found that the two influences on pupil size were not additive. Although the TEPRs could still sometimes be seen, the light reflex was much larger and also tended to suppress and distort the TEPR. We therefore echo the advice of prior investigators of pupillary responses to carefully control lighting during TEPR measurement [Beatty & Lucero-Wagoner 2000].

4 Conclusion and ongoing work

We found that remote video eye trackers have enough precision as pupillometers to be used for detailed measurements of task-evoked pupillary responses. We are excited by the potential of this capability, especially when combined with simultaneous measurement of gaze direction. We are proceeding with investigations into the relationship between cognitive load and gaze direction for a variety of visual tasks, including visual search, mental rotation, and shape discrimination.

Funding & Permissions

This work was funded by the Stanford Regional Visual Analytics Center, through the U.S. Department of Energy's Pacific Northwest National Laboratory. Our eye tracker was funded by the Stanford MediaX project and the Stanford School of Engineering.

References

AHERN, S. AND BEATTY, J. 1979. Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*. 205(4412): p. 1289-1292.

ALEXANDRIDIS, J.A., LEENDERTZ, J.A., AND BARBUR, J.L. 1991. Methods of studying the behaviour of the pupil. *Journal of psychophysiology*. 5: p. 223-239.

BEATTY, J. 1982. Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*. 91(2): p. 276-92.

BEATTY, J. AND LUCERO-WAGONER, B. Pupillary System. Chapter 6 in CACIOPPO, J.T., TASSINARY, L.G., AND BERNTSON, G. *Handbook of Psychophysiology*. 2000. Cambridge University Press.

DUCHOWSKI, A.T. *Eye Tracking Methodology: Theory and Practice*. 2003. Springer.

IQBAL, S.T., ADAMCZYK, P.D., ZHENG, X.S., AND BAILEY, B.P., Towards an index of opportunity: understanding changes in mental workload during task execution, in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2005. ACM: Portland, Oregon, USA. p. 311-320.

KAHNEMAN, D. AND BEATTY, J. 1966. Pupil Diameter and Load on Memory. *Science*. 154(3756): p. 1583-1585.

MARSHALL, S.P., 2002. The Index of Cognitive Activity: measuring cognitive workload, in *Proceedings of the IEEE Conference on Human Factors and Power Plants*. p. 7-5--7-9.

MCLAREN, J.W., ERIE, J.C., AND BRUBAKER, R.F. 1992. Computerized analysis of pupillograms in studies of alertness. *Investigative Ophthalmology & Visual Sciences* 33(3): p. 671-676.

MOLONEY, K.P., JACKO, J.A., VIDAKOVIC, B., SAINFORT, F., LEONARD, V.K., AND SHI, B. 2006 Leveraging data complexity: Pupillary behavior of older adults with visual impairment during HCI. *ACM Trans. Comput.-Hum. Interact*. 13(3): p. 376-402.

POMPLUN, M. AND SUNKARA, S., 2003. Pupil Dilation as an Indicator of Cognitive Workload in Human-Computer Interaction, in *Proceedings of the International Conference on HCI*.

TOBII TECHNOLOGY, INC.. 2007. Tobii 1750. http://www.tobii.com