

Performance Evaluation of the Gazepoint GP3 Eye Tracking Device Based on Pupil Dilation

Pujitha Mannaru¹(✉), Balakumar Balasingam¹, Krishna Pattipati¹,
Ciara Sibley², and Joseph T. Coyne²

¹ Department of Electrical and Computer Engineering,
University of Connecticut, Storrs, CT, USA

{pujitha.mannaru,balakumar.balasingam,krishna.pattipati}@uconn.edu

² Warfighter Human Systems Integration Lab, Naval Research Laboratory,
Washington DC, USA

{ciara.sibley,joseph.coyne}@nrl.navy.mil

Abstract. Eye tracking is considered one of the most salient methods to study the cognitive demands of humans in human computer interactive systems, due to the unobtrusiveness, flexibility and the development of inexpensive eye trackers. In this work, we evaluate the applicability of these low cost eyetrackers to study pupillary response to varying memory loads and luminance conditions. Specifically, we examine a low-cost eye tracker, the Gazepoint GP3, and objectively evaluate its ability to differentiate pupil dilation metrics under different cognitive loads and luminance conditions. The classification performance is computed in the form of a receiver operating characteristic (ROC) curve and the results indicate that Gazepoint provides a reliable eye tracker to human computer interaction applications requiring pupil dilation studies.

Keywords: Low-cost eye trackers · Eye tracker performance · Gazepoint · Pupil dilation · Memory load · TEPR · Power spectral density

1 Introduction

Eye tracking metrics are found to be useful indicators of visual attention and cognitive workload in numerous application areas, including reading and language comprehension [1], driving [2], individual differences [3], gaming devices [4], and medical applications [5]. Eye tracking devices (eye trackers) are used to collect measurements, such as pupil dilation, gaze locations and eye-closing patterns. Recent technical advances in video sensors and miniaturized computing power have resulted in cost-effective mass produced eye tracking devices; thus, several low-cost eye tracking devices have become available for researchers. However, the effectiveness of these low-cost devices to study human behavior remains an ongoing investigation [6–13] and is the objective of this paper. Specifically, we examine a low-cost eye tracker, the Gazepoint GP3 (cost \approx \$500), and objectively evaluate its ability to differentiate pupil dilation metrics under different

cognitive loads and luminance conditions. To our knowledge, this is one of the first studies reporting the effectiveness of Gazeport GP3 in capturing pupillary data.

Several pupillary metrics have been proposed in the past as useful indices of cognitive context [14–16]. Out of those, we employ two widely accepted metrics in this paper: one computed in the time domain and the other in the frequency domain. Using data collected by the Gazeport GP3 eye tracking device, a time domain measure, task evoked pupillary response (TEPR) [17], as well as a recently published frequency domain measure, pupillary power spectral density (PSD) [18], are computed and evaluated as indicators of mental workload under different luminance conditions. It has been well established that pupil diameter is impacted by both mental workload and luminance conditions [19–24]. Therefore, the objective of our experiment is to verify the potential use of Gazeport system to study the impact of these two factors on pupil diameter in studies involving cognitive context analysis.

Towards this end, we employed the digit span task [19] experiment under different luminance conditions, which is explained in Sect. 2. The rest of the paper is organized as follows: data collection and analysis methods are described in Sects. 2 and 3, respectively, the results of classification analysis are presented and discussed in Sect. 4, and the paper is concluded in Sect. 5.

2 Experiment

2.1 Subjects

Twenty participants ranging in age from 22 to 29 years ($M = 23.9$, $SD = 2.41$) voluntarily participated in the experiment conducted by researchers from the Naval Research laboratory (NRL) at the Naval Aerospace Medical Institute (NAMI).

2.2 Apparatus

All the eye tracking data were collected using the Gazeport GP3 system. The system was calibrated for each user according to the Gazeport Application Program Interface (API) manual [25]. GP3 collects the pupillary data, specifically, pupil size in pixels for each eye and their corresponding binary quality factor (valid/invalid) at 60 samples/s.

2.3 Task

A visual *digit span task* (also known as memory span task), which is a common technique used for assessing working memory capacity, was employed to assess the pupillary response of the participants to mental workload. In this task, participants are presented with a series of numbers and are then asked to recall the digits in the order they saw them. Longer series of numbers present more of a challenge for working memory, while shorter series are expected to be easier.

A *luminance change task* was employed to assess the pupillary response of the participants to the screen luminance. While completing the digit span task, participants were fixating on a monitor which varied in the background luminance (black, gray, and white).

2.4 Procedure

As mentioned in the previous section, participants engaged in a digit span task. Each participant was given four sets of digits of sizes 3, 5, 7 and 9 under three different screen luminance conditions (black, gray and white). The experiment utilized a within subject design (i.e., repeated measures) in which each participant completed all digit span set sizes (3, 5, 7 and 9, randomly ordered and exhaustive) three times for each of the 3 different background colors (white, gray, and black). Thus, a total of $36 (= 4 \text{ set sizes} \times 3 \text{ colors} \times 3 \text{ times})$ trials were conducted. Participants were told to focus on a central fixation cross (a “+” sign ~ 50 pixels tall and wide) that was offset from the background color (80 brighter for the black and gray backgrounds, and 80 darker for the white background). The string of numbers was then sequentially presented ~ 1 s per number. Following each number set (e.g., “2, 6, 1, 8, 4”), a numeric keypad appeared on the screen and participants used the mouse to input the string of numbers (“2, 6, 1, 8, 4”) by clicking on the corresponding numbers in order. The keypad was used to ensure that participants continued to fixate on the screen, while they were making a response. When satisfied, the participants clicked the submit button. Participants were not given performance feedback on their response accuracy. Following each set of digits, there was a pause of ≈ 3 s before presenting the participant with a numeric keypad on the monitor to enter his/her response. The pupillary measures from this time segment, known as the *encoding phase* of the memory, are analyzed here. The total time to complete the digit span task varied from 10–15 min, depending on the participant’s response times.

3 Data Analysis

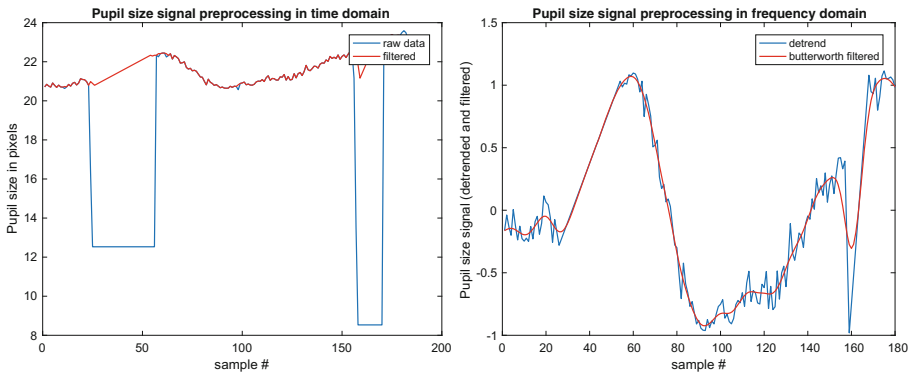
The Gazepoint GP3 collects the following pupillary data: pupil size in pixels for each eye and their corresponding binary quality factors (valid/invalid) at 60 samples/s, the scale factor of each eye pupil (unitless), whose value equals 1 at calibration depth, is less than 1 when the user is closer to the eye tracker and greater than 1 when the user is further away. Only data from the encoding time segment are analyzed in this work, as it has been established by the human factors researchers that the maximum pupil dilation occurs during the encoding of the stimulus materials for short term memory recall tasks [26, 27].

3.1 Data Preprocessing

For *time-domain analysis* (TEPR), the poor quality samples (quality factor = 0) of the pupil size signals were marked as missing values (or NaN in

MATLAB[®] [28]). Pupil size data of the eye with fewer missing observations [29] were utilized for analysis. A “clean-up” function was employed to remove all the data below 4th percentile and above 98th percentile, in order to remove any sudden dips/peaks in the pupil size signal. Then, a hamper filter (of order 6) [30] was applied to remove outliers and a linear interpolator was used to recover missing values. Figure 1a shows an example of raw data and filtered data signals.

For *frequency-domain analysis* (PSD), the linear trend in the above pre-processed signals was removed using the *detrend* function in MATLAB[®] and the resulting signals were passed through a zero-phase lowpass butterworth filter with a cutoff frequency $f_c = 4$ Hz using the *filtfilt* function, since most of the pupillary activity falls in the frequency range of 0–4 Hz [31]. Figure 1b shows an example of detrended data and filtered data signals.



(a) Preprocessing for Time domain analysis (b) Preprocessing for frequency-domain analysis

Fig. 1. Pupil size signal preprocessing

3.2 Data Analysis

Task Evoked Pupillary Response (TEPR): To evaluate the ability of the eye tracker in capturing the changes in pupil diameter caused by mental workload changes, we analyzed the data of set sizes 3 (labeled as EASY), 5 (labeled as MEDIUM) and 7 (labeled as HARD) only. The set size 9 was excluded from the analysis since recall performance dropped to 65% (i.e., only remembering 65% of the 9 numbers) and there was increased variability between participants, suggesting it was either too difficult for some participants or that some participants gave up. For classification purposes, the median values of the pupil size in the encoding phase (TEPR), for each person, for each set size, each background color, and for each trial, (e.g., pupil size of person 13, set size 3 in a black background for the first trial) were computed over a sliding window of size

30 samples with an overlap of 25 samples ($\approx 80\%$ overlap). A simple cut-point grouping into binary classes was implemented for pairs of set sizes 3 (EASY) vs. 7 (HARD), 3 (EASY) vs 5 (MEDIUM) and 5 (MEDIUM) vs. 7 (HARD) for the corresponding pairs of the moving-median filtered signals. The Receiver Operating Characteristic (ROC) curves [32] were drawn by varying the cut-points from the minimum of the two signals, in steps of 0.01 pixels, to the maximum value of the two signals.

Power Spectral Density (PSD): PSD of the pupil diameter signals was computed for each person using the Welch’s method with segments of 50 samples with 50% overlap [18]. Each segment was windowed with a Hamming window. Only the ‘encoding’ phase was considered when computing PSD under the memory tasks of set size 3 (EASY) vs set size 5 (MEDIUM) vs. set size 7 (HARD). PSD presented here is the average PSD over 20 participants * 3 trials; thus averaged over a total of 60 trials for each background luminance color.

4 Results and Discussion

At the preprocessing stage, an average of 37% data was missing due to poor quality recordings. Figure 2 shows the boxplots for average pupil diameters across different background luminance conditions and workload conditions. It is evident that the average pupil diameter in a black background is higher than that of the grey background which, in turn, is greater than that of the white background; this pattern agrees with earlier pupillary light reflex studies, thereby assuring the GP3’s capability to capture light-sensitive pupillary readings. Figure 2 also shows the differences in average pupil diameter for different workload tasks within the same background conditions and it can be seen that the average pupil diameter for set size 3 is lower than that of set size 7 under all 3 luminance conditions. However, the pupil diameters of set size 5 is not clearly greater than (or lesser than) for set size 3 (or for set size 7) under black and grey background luminance conditions.

To further analyze the differences in TEPRs corresponding to the different set sizes, we plotted the ROC curves from classification as described in Sect. 3. An example set of ROC curves for one person are shown in Figs. 3, 4 and 5. For this particular example, Fig. 3 shows a 100% accuracy in classifying pupil size signals of set size 3 vs. 7 for all three background conditions, whereas a 68% accuracy in classifying pupil size signals of set size 3 vs. 5 in grey background conditions and a 78% accuracy in classifying pupil size signals of set size 5 vs. 7 in white background conditions. Table 1 gives the average classification accuracy values over all participants and over all 3 repeated trials. Therefore, the minimum average classification accuracy is approximately 80%, which is considered a significant value by psychologists in detecting human cognitive context.

Figure 6 shows the results of PSD analysis, where Figs. 6(a–c) correspond to black, grey and white background conditions, respectively. The results agree with earlier studies only in the average power spectral densities of set size 3 vs.

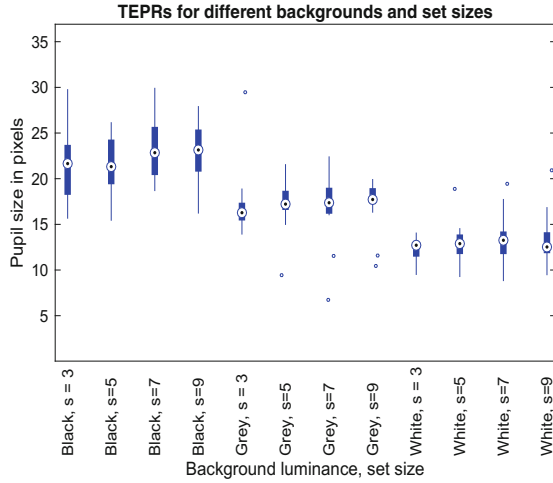


Fig. 2. Boxplot of average pupil diameters under different backgrounds and mental workloads

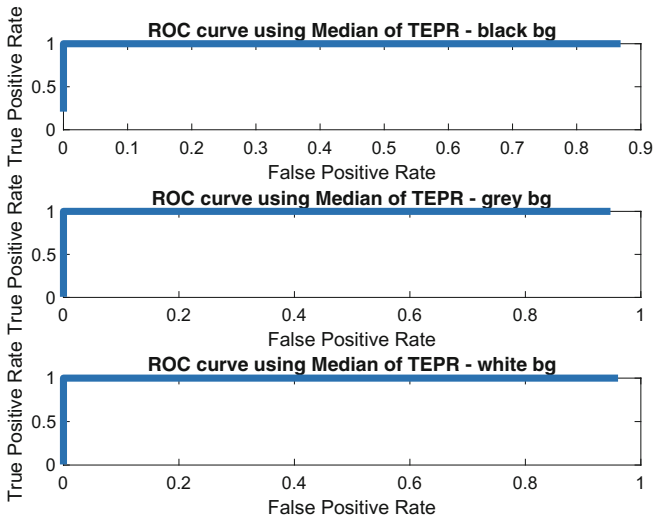


Fig. 3. ROC curves from classification of TEPRs between set size 3 and 7

set size 5 or 7. However, the results we obtained do not conform to the finding that average PSD increases in the frequency range of 0.1–0.5 Hz and 1.6–3.5 Hz with increase in cognitive workload as the average PSD in set size 5 is seen to be greater than that of set size 7. This could be due to the recovery of lost data points by using a linear interpolator or due to similar spectral behavior of pupils during set sizes 5 and 7. Also, to our knowledge, there is no detailed mechanism

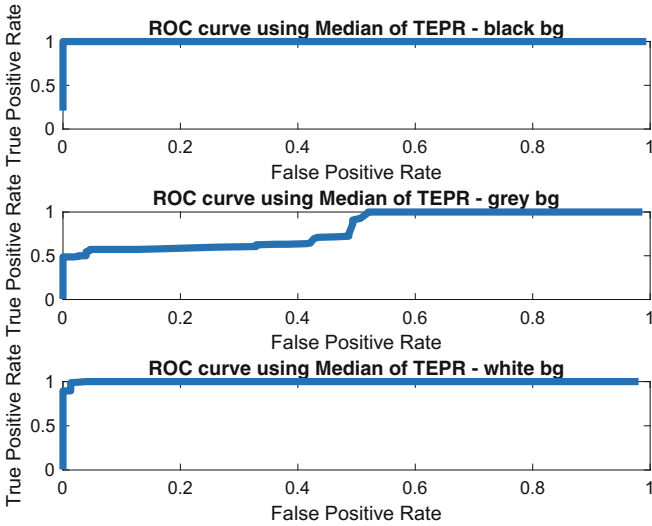


Fig. 4. ROC curves from classification of TEPRs between set size 3 and 5

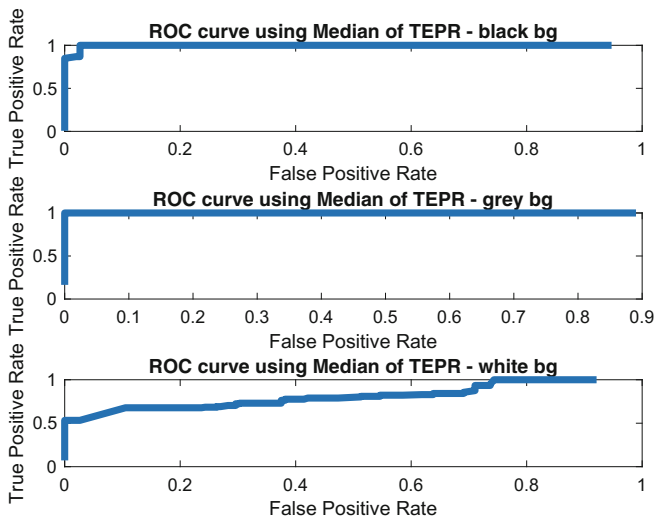
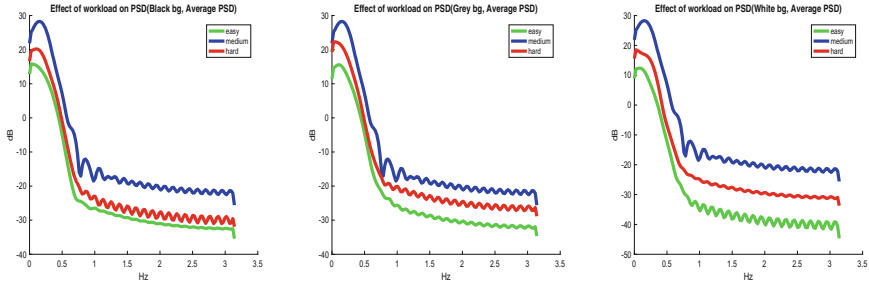


Fig. 5. ROC curves from classification of TEPRs between set size 5 and 7

for this phenomena of pupil control and PSD, yet. Future research will integrate the PSD metrics in classification studies to attempt to validate the findings of Peysakhovich et al. [18] and Nakayama and Shimizu [31].

Table 1. Average accuracies in TEPR classification

Average accuracies in %			
Background luminance conditions	Set size 3 vs. 7	Set size 3 vs. 5	Set size 5 vs. 7
Black	90.73	79.83	83.28
Grey	87.53	80.92	81.77
White	86.89	81.45	79.68



(a) PSD in black luminance conditions (b) PSD in grey luminance conditions (c) PSD in white luminance conditions

Fig. 6. Power spectral density under different workload conditions

5 Summary and Conclusion

In this paper, we evaluated the performance of Gazeport GP3, a low-cost eye tracker, by using pupillary metrics that are already tested and used by human factors researchers: TEPRs and PSD. We collected pupil size data from 20 volunteers engaged in a visual digit span task. First, a preprocessing routine was employed to filter out outliers from the data for time domain analysis, and low pass filtering was performed prior to frequency domain analysis. Then, TEPRs and PSDs were computed and studied for different digit set sizes. The classification performance is computed in the form of a receiver operating characteristic (ROC) curve and the results show the applicability and limitations of low-cost eye tracking devices by cognitive workload researchers.

The results indicate that the Gazeport GP3 is an easy and inexpensive tool that can be utilized in psychological studies involving pupil diameter data. The classification results indicate that the eye tracker does a good job in classifying mental workloads under different background luminance conditions; however, it is not a robust tool for frequency domain analysis which could be attributable to linear interpolation of poor quality readings. Researchers, with budget constraints, who are interested in incorporating pupillary measures of cognitive workload now have access to a reliable inexpensive eye tracker. However, they should keep in mind the GP3 is limited to collecting pupil diameter data for tasks

which use a single screen and is vulnerable to loss of chunks of data. Finally, we believe that the low cost eyetrackers are of great value to researchers from all disciplines trying to incorporate human factors aspects in their systems.

Acknowledgements. The authors would like to thank Dr. Jeffrey Morrison and the Command Decision Making (CDM) program at the U.S. Office of Naval Research and Department of Defense High Performance Computing Modernization Program for supporting this work. In addition, the authors would like to thank the symposium organizers for their encouragement of this work. This research was funded by the U.S. Office of Naval Research and the Department of Defense under contracts #N00014-12-1-0238, #N00014-16-1-2036 and #HPCM034125HQU.

References

1. Just, M.A., Carpenter, P.A.: The intensity dimension of thought: pupillometric indices of sentence processing. *Can. J. Exp. Psychol./Revue canadienne de psychologie expérimentale* **47**(2), 310 (1993)
2. Palinko, O., Kun, A.L., Shyrokov, A., Heeman, P.: Estimating cognitive load using remote eye tracking in a driving simulator. In: *Proceedings of 2010 Symposium on Eye-Tracking Research and Applications*, pp. 141–144. ACM (2010)
3. Odenheimer, G., Funkenstein, H., Beckett, L., Chown, M., Pilgrim, D., Evans, D., Albert, M.: Comparison of neurologic changes in 'successfully aging' persons vs the total aging population. *Arch. Neurol.* **51**(6), 573–580 (1994)
4. Ekman, I.M., Poikola, A.W., Mäkäräinen, M.K.: Invisible eni: using gaze and pupil size to control a game. In: *CHI 2008 Extended Abstracts on Human Factors in Computing Systems*, pp. 3135–3140. ACM (2008)
5. Ren, P., Barreto, A., Gao, Y., Adjouadi, M.: Affective assessment of computer users based on processing the pupil diameter signal. In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 2594–2597. IEEE (2011)
6. Zugal, S., Pinggera, J.: Low-cost eye-trackers: useful for information systems research? In: Iliadis, L., Papazoglou, M., Pohl, K. (eds.) *CAiSE 2014. LNBP*, vol. 178, pp. 159–170. Springer, Cham (2014). doi:[10.1007/978-3-319-07869-4_14](https://doi.org/10.1007/978-3-319-07869-4_14)
7. Dalmaijer, E.: Is the low-cost eye tribe eye tracker any good for research? Technical report, PeerJ PrePrints (2014)
8. Ooms, K., Dupont, L., Lapon, L., Popelka, S.: Accuracy and precision of fixation locations recorded with the low-cost eye tribe tracker in different experimental setups. *J. Eye Mov. Res.* **8**(1) (2015)
9. Ferhat, O., Vilariño, F.: Low cost eye tracking. *Comput. Intell. Neurosci.* **2016**, 17 (2016)
10. Coyne, J., Sibley, C.: Investigating the use of two low cost eye tracking systems for detecting pupillary response to changes in mental workload. In: *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, vol. 60, pp. 37–41. SAGE Publications (2016)
11. Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., Menke, L.: Which eye tracker is right for your research? Performance evaluation of several cost variant eye trackers. In: *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, vol. 60, pp. 1240–1244. SAGE Publications (2016)

12. Gibaldi, A., Vanegas, M., Bex, P.J., Maiello, G.: Evaluation of the Tobii EyeX eye tracking controller and Matlab toolkit for research. *Behav. Res. Methods* 1–24 (2016)
13. Janthanasub, V., Meesad, P.: Evaluation of a low-cost eye tracking system for computer input. *King Mongkuts Univ. Technol. North Bangk. Int. J. Appl. Sci. Technol.* **8**(3), 185–196 (2015)
14. Causse, M., Sénard, J.-M., Démonet, J.F., Pastor, J.: Monitoring cognitive and emotional processes through pupil and cardiac response during dynamic versus logical task. *Appl. Psychophysiol. Biofeedback* **35**(2), 115–123 (2010)
15. Mannaru, P., Balasingam, B., Pattipati, K., Sibley, C., Coyne, J.: Cognitive context detection in UAS operators using pupillary measurements. In: *SPIE Defense+ Security*, p. 98510Q. International Society for Optics and Photonics (2016)
16. Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., Causse, M.: Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biol. Psychol.* **121**, 62–73 (2016)
17. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* **91**(2), 276 (1982)
18. Peysakhovich, V., Causse, M., Scannella, S., Dehais, F.: Frequency analysis of a task-evoked pupillary response: luminance-independent measure of mental effort. *Int. J. Psychophysiol.* **97**(1), 30–37 (2015)
19. Kahneman, D., Beatty, J.: Pupil diameter and load on memory. *Science* **154**(3756), 1583–1585 (1966)
20. Tryon, W.W.: Pupillometry: a survey of sources of variation. *Psychophysiology* **12**(1), 90–93 (1975)
21. Taptagaporn, S., Saito, S.: How display polarity and lighting conditions affect the pupil size of VDT operators. *Ergonomics* **33**(2), 201–208 (1990)
22. Goldinger, S.D., Papesh, M.H.: Pupil dilation reflects the creation and retrieval of memories. *Curr. Dir. Psychol. Sci.* **21**(2), 90–95 (2012)
23. Winn, B., Whitaker, D., Elliott, D.B., Phillips, N.J.: Factors affecting light-adapted pupil size in normal human subjects. *Investig. Ophthalmol. Vis. Sci.* **35**(3), 1132–1137 (1994)
24. Peysakhovich, V., Vachon, F., Dehais, F.: The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *Int. J. Psychophysiol.* **112**, 40–45 (2017)
25. Gazept API. http://www.gazept.com/dl/Gazept_API_v2.0.pdf
26. Beatty, J., Lucero-Wagoner, B.: The pupillary system. *Handb. Psychophysiol.* **2**, 142–162 (2000)
27. Gardner, R.M., Beltramo, J.S., Krinsky, R.: Pupillary changes during encoding, storage, and retrieval of information. *Percept. Mot. Skills* **41**(3), 951–955 (1975)
28. MATLAB: R2016a. The MathWorks Inc., Natick (2016)
29. Papesh, M.H., Goldinger, S.D., Hout, M.C.: Memory strength and specificity revealed by pupillometry. *Int. J. Psychophysiol.* **83**(1), 56–64 (2012)
30. Pearson, R.K.: Outliers in process modeling and identification. *IEEE Trans. Control Syst. Technol.* **10**(1), 55–63 (2002)
31. Nakayama, M., Shimizu, Y.: Frequency analysis of task evoked pupillary response and eye-movement. In: *Proceedings of 2004 Symposium on Eye Tracking Research Applications*, pp. 71–76. ACM (2004)
32. Fawcett, T.: An introduction to ROC analysis. *Pattern Recogn. Lett.* **27**(8), 861–874 (2006)