2017-08-12

Task Condition and Pupillometry

Mina Shojaeizadeh
Worcester Polytechnic Institute

Et al.

Task Condition and Pupillometry

Full Papers

Mina Shojaeizadeh
Worcester Polytechnic Institute
minashojaei@wpi.edu
Ping Chen
UMASS Boston
ping.chen@umb.edu

Soussan Djamasbi
Worcester Polytechnic Institute
djamasbi@wpi.edu
John Rochford
UMASS Medical School
john.rochford@umassmed.edu

Abstract

While eye tracking is gaining popularity in IS research, pupillometry is relatively less explored in IS eye tracking studies. Research however suggests that pupillometry may serve as an excellent unobtrusive measure to study user information processing behavior. The Adaptive decision making theory asserts that task demand affects information processing behavior. Grounded in this theory, we argue that users' pupillary responses will be different under different task conditions (task demand). We tested our assertion via an eye tracking laboratory experiment. Our results show that pupillary responses were significantly different under different task conditions.

Keywords

Pupil dilation, pupil dilation variation, task conditions, eye-tracking, cognitive load

Introduction

Human eyes have a great deal of communicative power and as such examining eye movement data is gaining popularity in IS research (Djamasbi 2014). Pupillometry, however, is less explored in IS eye tracking studies. Research suggests pupillometry may serve as an excellent unobtrusive measure to study user information processing behavior. For example, studies show pupil dilation has a direct relation to mental activities (Bailey and Iqbal 2008, Hess and Polt 1964). Pupil dilation can be measured continuously during processing of a task. This would enable pupillary data to be a potentially robust measure of cognitive load (Beatty 1982, Iqbal et al. 2004). In addition to the pupil dilation, there is indication that variability in pupil dilation may also reveal users' cognitive load (Fehrenbacher and Djamasbi 2017). The relationship between cognitive load and pupil dilation variation (PDV) also has been supported by an IS study that shows a positive relationship between PDV and task difficulty (Chen S. et al 2011).

Grounded in the theory of adaptive decision making (Payne et al. 1993), which asserts that people choose an information processing behavior based on the demand placed on their cognitive resources, we argue that pupillary responses are likely to carry information about task condition and characteristics. We test this assertion via an eye tracking laboratory experiment.

Theoretical Background

Research shows pupillary data can serve as a reliable measure of cognitive effort. For example, Beatty and Kahneman (1966) observed an increase in pupil size as people completed harder tasks. Similarly, Chen et al. (2011) observed a positive relationship between pupil size and task difficulty (i.e., recalling the number of player positions in a basketball game). Klingner et al. (2011) measured pupil dilation during a mental multiplication; and found that easy-multiplication problems triggered the smallest pupil dilations and hard problems the largest. More recent IS scholars also suggest that pupil dilation is a reliable proxy of cognitive load (Klinger et al. 2008, King 2009, Piquado et al. 2010, Zhan et al. 2016).

Some recent IS studies have examined the relationship between cognitive load and pupil dilation variation (PDV). PD is defined as the size of pupil diameter (Kahneman and Beatty 1996) and PDV is defined as rate of change in pupil dilation measured as standard deviation of pupil dilation (Shojaeizadeh et al. 2015, Buettner et al. (2015)). For example, Buettner et al. (2015) showed that PDV has a positive relationship with performance, and argued that PDV is an appropriate measure of cognitive demand in IS research. Shojaeizadeh et al. (2015) measured average PD and PDV during a complex problem solving task, and suggested PDV may be a more sensitive measure of cognitive load, as compared to PD for complex problem solving tasks. They found a significant correlation between PDV and cognitive load, whereas they did not find a strong correlation between PD and cognitive load.

A recent study provided evidence that task condition has an impact on pupillary response, and suggested the adaptive decision making theory may serve as a suitable theoretical framework for IS decision making eye tracking studies (Fehrenbacher and Djamasbi 2017). According to adaptive decision making theory (Payne et al. 1993), people choose an information processing behavior based on the demand placed on their cognitive resources. Grounded in this theory, we argue different task demands impact cognitive resources in different ways. This, in turn, is likely to impact how people manage their cognitive loads and thus is likely to impact their pupillary responses. Therefore, we hypothesize:

H1) Pupil dilation (PD) will be different in different task conditions.

H2) Pupil dilation variation (PDV) will be different in different task conditions.

Methodology

We recruited 54 volunteers to participate in our eye-tracking study. Participants were randomly selected among college students (between the age of 18 and 24) in a northeastern university with technical disciplines. Before starting the task, each participants went through a short calibration process. The task required participants to read a text passage and answer two questions about the passage they just read. The text passage and the questions were displayed on the same screen. Participants were randomly assigned to one of the two text conditions (original or simplified). The passages used in this study were adopted from a previous study (Djamasbi et al. 2016-b), where the authors indicated that participants who read the simplified version of the passage experienced less cognitive load and performed significantly better in reading comprehension. The original passage in Djamasbi et al. (2016-b) study was simplified using plain language standards from their previous study (Djamasbi et al. 2016-a). One of the questions about the passage was literal and the other was inferential. To avoid order effect, the order in which the questions were displayed on the screen was randomized. Cognitive load was manipulated as task demand or task condition in two ways: 1) reading an original passage, 2) answering to inferential question.

To capture eye-movements we used Tobii x300 and Tobii studio 3.2.3. For each participants, two video segments were created: one capturing user eye movement activity when completing the inferential question and one when completing the literal question.

Data Analysis

Because the task was not timed task duration was different among participants, which resulted in different PD and PDV data points for each participant. To standardize the number of PD and PDV values we used a cubic spline interpolation (McKinley and Levine 1998). We then computed the average PD time-series of 26 subjects who were given the original version of the passage, and then 28 subjects who performed the task on a simplified version. The same procedure was also applied to PDV data. This process allowed us to obtain a similar length of data for every participant who completed the task in the original or simplified text conditions.

Results

Hypotheses testing

To test our hypotheses, we investigated pupillary responses to 1) text type (original/simplified), and 2) to question type (inferential/literal). We performed two mixed model ANOVAs, one for PD and one for PDV. The results (Table 1 and Table 2) indicate that PD values for people in the simplified text condition were significantly different from the PD values for people in the original text condition ($F(1,477)=16.92$ and p-value=0.00). Additionally, the results show that when answering a literal question compared to when answering an inferential question, PD values were significantly different (F(1,477)=65.71, p-value=0.00). Further, the results show that PD values were impacted significantly by the interaction between the type of text (simplified vs. original) and the type of question (inferential vs. literal) (F(1,477)=22.29, p-value=0.00). The pairwise comparisons show that in the simplified text condition PD during responding to an inferential question was significantly different from PD during answering a literal question (PD-inferential=2.992±.003 vs. PD-literal=3.007±.004, F(1,477)=47.72, p-value=0.00). However, we did not observe the same in the original text condition (PD-inferential=2.994±.004 vs. PD-literal=2.998±.004, F(1,477)=1.18, p-value=0.28) (Table 2).

PDV values for people in the simplified text condition were also significantly different from the PDV values for people in the original text condition (F(1,524)=85.17, p-value=0.00). Similarly, PDV values were also significantly different between literal and inferential questions (F(1,524)=11.12, p-value=0.00). The results, however, did not show a significant interaction effect (F(1,524)=0.43, p-value=0. 52). In other words, regardless of the type of passage the participants read (original or simplified) their PDV was significantly affected by the type of questions they answered (Table 3). The results of pairwise comparisons in Table 4 show that PDV in the original text condition was significantly different between inferential and literal questions (PDV-inferential=0.0060±0.0002 vs. PDV-literal=0.0066±0.0002, F(1,524)=58.80, p-value=0.00). PDV in the simplified text conditions was also significantly different between inferential and literal questions (PD-inferential=0.0041 ± 0.0001 vs. PD-literal=0.0049 ± 0.0002, F(1,524)=31.97, p-value=0.00).

	F(1,477)	p-value
Original_Simplified	16.92	0.00
Inferential_Literal	65.71	0.00
Interaction	22.29	0.00

Table 1. Results of mixed model ANOVA comparing the means of PD among literal and inferential questions

Task Condition	Question type	Mean ± SD	F(1,477)	p-value
Original Text	Inferential	2.994 ±.004	1.18	0.28
	Literal	2.998 ±.004		
Simplified Text	Inferential	2.992 ±.003	47.72	0.00
	Literal	3.007 ±.004		

Table 2. Descriptive statistics and pairwise comparison between means of PD in different task conditions

These results together support our hypothesis that different task demands significantly affect PD and PDV. The results also support the assertion that question type also affects PD and PDV values. The results show an interaction effect between text type and question type on PDV, but not on PD.

	F (1,524)	p-value
Original_Simplified	85.17	0.00
Inferential_Literal	11.12	0.00
Interaction	0.43	0.52

Table 3. Results of mixed model ANOVA comparing the means of PDV among literal and inferential questions

Task Condition	Question type	Mean ± SD	F(1,524)	p-value
Original Text	Inferential	.0060± .0002	58.80	0.00
	Literal	.0066 ± .0002		
Simplified Text	Inferential	.0041 ± .0001	31.97	0.00
	Literal	.0049 ± .0002		

Table 4. Descriptive statistics and pairwise comparison between means of PDV of different task conditions

Exploratory Analysis

A recent study shows that examining pupil data in various time intervals is useful (Fehrenbacher and Djamasbi 2017). The purpose of this exploratory study is to understand how users experience cognitive load during different periods of making decision. This is because Fehrenbacher and Djamasbi (2017) showed that the way people distribute their effort over the decision time in order to complete a cognitive task varies within different time intervals. Here the objective was to see whether we can see the same user behavior pattern when it comes to making decisions in answering questions about an original passage and its simplified version. Therefore, we divided the interpolated time into three equal size portions as it was suggested by Fehrenbacher and Djamasbi (2017): beginning, middle and end, and investigated the differences in the means of PD and PDV during these time intervals. As mentioned earlier, because the task was not timed task duration was different among the participants, which resulted in different PD and PDV data points for each participant. To standardize the number of PD and PDV values we used a cubic spline interpolation (McKinley and Levine 1998). Therefore, the number of data points in each time interval is equivalent in the both task conditions (literal vs. inferential questions). Next, we performed two separate two-way ANOVAs. The first ANOVA tested whether PD during answering the inferential question (PD-inferential) was significantly different between different text conditions during different time intervals (PD-inferential); and the second ANOVA tested the same effects on PD but this time during answering the literal questions (PD-literal).

The results displayed in Table 5 show that, when answering inferential and literal questions, PD values for participants in the simplified text condition was not significantly different from PD values for participants in the original text condition during both inferential and literal questions. In other words, the results do not show a main effect for text type on PD-inferential ($F(1,950)=0.13$, p-value=0.72) as well as on PD-literal ($F(1,1042)=3.42$, p-value=0.07). The results however show that PD values were significantly different during the three time intervals ($F(2,950) = 23.14$, p-value=0.00). Furthermore, the results show that PD values were significantly affected by the interaction between the text type condition and time interval divisions ($F(2,950)=7.59$, p-value=0.00) during the inferential question. Results also show that PD values were not significantly affected by the interaction between the text type condition and time interval divisions during the literal question ($F(2,1042)=0.58$, p-value=0.56). Figure 3a shows the pairwise comparison between PD of the original text condition and PD of the simplified text condition during the three different time intervals and during inferential question. Results show that PD value during the simplified text condition is significantly different from PD value during the original text condition during the beginning and middle time intervals but not during the end time interval. An upward trend for PD values can be observed in both text conditions suggesting more intense cognitive activity at the end, when users were making decisions (selecting answers for an inferential question). This figure also shows opposite trends, in different text conditions, in the beginning and the middle time intervals.

	Inferential		Literal	
	F	p-value	F	p-value
Original_Simplified	0.13	0.72	3.42	0.07
Intervals	23.14	0.00	28.30	0.00
Interaction	7.59	0.00	0.58	0.56

Table 5. Results of ANOVA for PD within different time intervals during inferential and literal questions

(a) Inferential Question (b) Literal Question

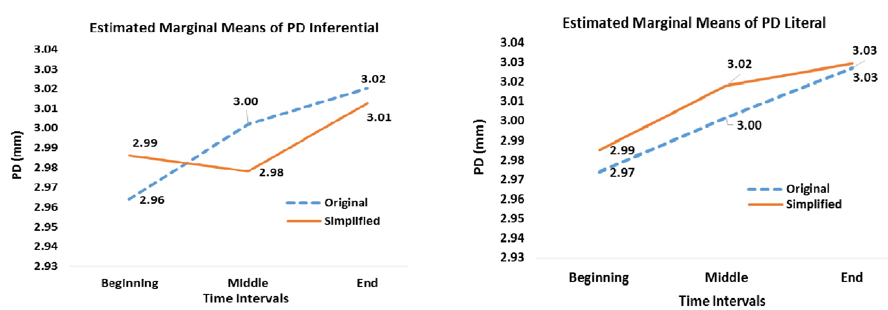

Figure 3. Average values of PD during three different time intervals (beginning, middle, and end) and two text conditions (original vs. simplified), and two different questions (inferential vs. literal)

Figure 3b shows the pairwise comparisons between PD-literal values in the original text condition, and PD-literal values in the simplified text condition during different time intervals. Results show that PD was not significantly different during any of the three intervals between the simplified and the original text conditions ($P_{Beginning}$, P_{Middle}, $P_{End} > 0.05$).

Next, we investigated the differences in the means of PDV during the three time intervals: beginning, middle, and end. As before, we conducted two separate two-way ANOVAs, one for PDV values captured during answering an inferential question (PD-inferential), and one for those collected during answering a literal questions (PD-literal).

The results of ANOVA and pairwise comparisons are indicated in Tables 6 and Figure 4 for PDV during inferential and literal questions. As shown by the results, PDV is significantly different between simplified and original text conditions. PDV values are also significanlty different during different time intervals when answering an inferential question. There is also a significant interaction effect between text type and time intervals during inferential questions.

The results of ANOVA for literal questions were different from those obtained for the inferential questions. The results in Table 6 show that PDV values were significantly different between two text conditions, but they were not significantly different during the three time intervals. There was also no significant interaction effect (F(2,1044)=1.20, p-value=0.30) between text condition and time intervals. This suggests the significant effect of text condition on PDV did not depend on the time interval when answering literal questions. Results of the pairwise comparison in Figure 4a indicate that, only during the last time interval.

	Inferential		Literal	
	F	p-value	F	p-value
Original Simplified	59.03	0.00	32.91	0.00
Intervals	26.25	0.00	0.83	0.44
Interaction	8.89	0.00	1.20	0.30

Table 6. Results of ANOVA for PDV within different time intervals during inferential and literal questions

PDV is not significantly different between original and simplified text conditions; and that PDV is significantly different during the first and second time intervals for the beginning interval and for the middle interval. A downward trend for PDV can be observed in both text conditions during answering inferential questions. The downward trend is more pronounced in the original text condition. At the end, when participants were making decisions, PDV values were quite similar. A downward trend for PDV can be observed in both text conditions during answering literal questions (Figure 4b). There was a slight increase in PDV values in the middle of the original text condition and a slight decrease in PDV in the middle of the simplified text condition.

(a) Inferential Question (b) Literal Question

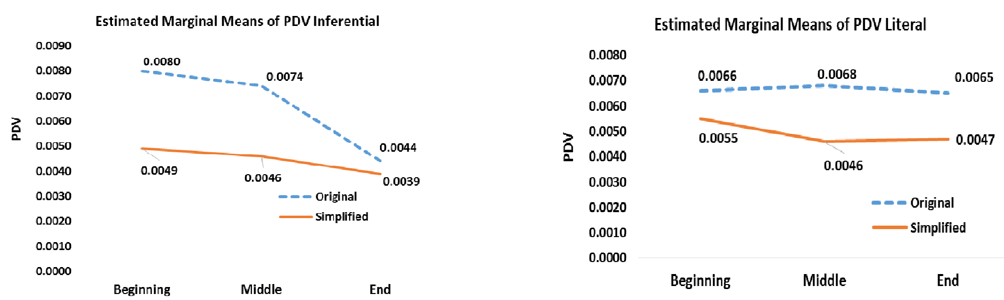

Figure 4. Average values of PDV during three time intervals (beginning, middle and end), two different text conditions (original vs. simplified), and two different questions (inferential vs. literal)

Discussion and Conclusion

The main objective of this paper was to test whether differences in task demand could be detected via pupil data. Because pupil dilation and variation have been associated with cognitive activity, we hypothesized that pupillary responses were likely to reflect task demand in our study (simplified vs. original text conditions; and answering inferential vs. literal questions). The results of ANOVA tests showed there was an overall significant difference in PD and PDV between the two text conditions and the two types of questions. While the results showed an interaction effect between text conditions and question type for PD, no interaction effect was observed for PDV. These results support our hypotheses. We also conducted exploratory ANOVA tests to refine our understanding of the results. We divided the task time into three intervals, and investigated PD and PDV values in each interval. Prior research indicate that studying PD and PDV during different time intervals could provide interesting insights for future pupillometry studies (Fehrenbacher et al., 2017). Supporting the results of previous research, our data showed larger pupil dilation at end of decision period. Our results also supported the finding of this prior research, which showed higher PD values were associated with lower PDV values.

Our results have important implications. First, supporting the results of a recent exploratory study examining the impact of four different task demand on pupillary responses (D. Fehrenbacher et al., 2017), our results show that PD and PDV can carry information about task condition. It is important to note that both the task and task conditions in the previous study were different from the task and task conditions in our study. Hence, our results provide evidence for the robustness of pupillometry in IS research.

The results also show that examining pupil data in various time intervals can provide valuable insight. For example, our results showed an upward trend for PD values during the decision time. This trend is consistent with prior research (Fehrenbacher et al., 2017). The upward trend of pupil dilation during the last parts of decision period is consistent with observations of pupil dilation trends in the study by Einhauser et al. (2010). The upward trend in pupil dilation for both inferential and literal questions at the beginning of decision periods suggests that participants were experiencing increased cognitive load at the beginning of the decision task (especially in the original text condition). The results also showed that during inferential question the difference in PDV values between text condition depended on the time interval of the decision task, they were signficatly different at the beginning and in the middle of the task but not signficantly different at the end. This suggests that people in the two different text conditions may have used different information processing strategies at the begining and in the middle of the task but used the same strategy at the end. Our results suggest that PD and PDV can provide continuous measurement of cognitive load in a decision task. Future research is needed to examine these possibilities. Future studies can benefit from our findings which suggest calculating PDV as well as PD within different time intervals is likely to provide a more comprehensive understanding of cognitve load in cognitive tasks. In this study we looked at three different time intervals, however, future studies can explore whether breaking the total time into smaller intervals (more time intervals) can improve understanding of user cognitve load.

As in any laboratory experiment, the results of our study are limited to the setting and the task used. We used a single text passage. More text passages with varying level of complexity are needed to replicate our results. The participants in our study were college students. Different participant groups can help to test whether the results extend to other populations.

Contribution

Grounded in the adaptive decision making theory and the results of a previous exploratory research we hypothesized that pupillary responses carry information about task conditions. The results of our study, which support our assertion, suggests that pupillometry may serve as an unobtrusive method to detect a user experience of task conditions. Our study provides a framework for examining information processing behavior continuously via eye tracking. Future studies can benefit from the results of this research, which suggest that in addition to PD, PDV can provide valuable insight from user's cognitive load in a cognitive task.

REFERENCES

Bailey B. P. and Iqbal S. T. "Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management," *ACM Trans. on Computer-Human Interaction*, Vol. 14(4), pp. 1–28, 2008.

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

Beatty J., Kahneman D. 1966. "Pupillary changes in two memory tasks, "*Psycho. Sci.* 5 (10) 371–372.

Chen S., Epps J., Ruiz N., and Chen F. 2011. "Eye Activity As a Measure of Human Mental Effort in HCI," *In Proceedings of the 16th International Conference on Intelligent User Interfaces* (IUI '11). ACM, New York, NY, USA, 315–318.

Buettner R., Sauer S., Maier C., Eckhardt A., 2015. "Towards ex ante prediction of user performance: a novel NeuroIS methodology based on real-time measurement of mental effort," *In: 48th Hawaii International Conference on System Sciences (HICSS), IEEE*, pp. 533–542

Djamasbi, S. 2014. "Eye Tracking and Web Experience," *AIS Transactions on Human-Computer Interaction* (6) 2, pp. 37-54

Djamasbi S., Shojaeizadeh, Mina; Chen, Ping; and Rochford, John, "Text Simplification and Generation Y: An Eye Tracking Study" (2016). SIGHCI 2016 Proceedings. Paper 12. http://aisel.aisnet.org/sighci2016/12

Djamasbi, S., Rochford, J., DaBoll-Lavoie, A., Greff, T.,Lally, J., & Kayla McAvoy, "Text Simplification and User Experience", *In International Conference on Augmented Cognition* (pp. 285-295). Springer International Publishing.

Einhauser W., Koch C., Carter O., Pupil dilation betrays the timing of decisions Front. *Hum. Neurosci*, 4 (18) (2010), pp. 1–9.

Fehrenbacher D. D., Djamasbi S., 2017. "Information systems and task demand: An exploratory pupillometry study of computerized decision making," *Decision Support Systems*, Available online 20 February 2017, ISSN 0167-9236.

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., X.S. Zheng and B.P. Bailey. 2004. "Task-Evoked Pupillary Response to Mental Workload in Human-Computer Interaction," *CHI*, 1477-1480.

King L. 2009. "Visual navigation patterns and cognitive load, in: D. Schmorrow, I. Estabrooke, M. Grootjen (Eds.)," *Foundations of Augmented Cognition Neuroergonomics and Operational Neuroscience Springer, Berlin*, pp. 254–259.

Klingner J., Tversky B., Hanrahan P. 2011. "Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks," *Psychophysiology* 48(3):323–332

Klingner J., Kumar R., and Hanrahan P. 2008. "Measuring the Task-evoked Pupillary Response with a Remote Eye Tracker," *In Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 69–72.

McKinley S. and Levine M., 1998. "Cubic spline interpolation," *College of the Redwoods*, vol. 45, no. 1, pp. 1049–1060.

Payne, J. W., Bettman, J. R., & Johnson, E. J. 1993. "The adaptive decision maker," *Cambridge University Press*.

Piquado T., Isaacowitz D., Wingfield A. 2010. "Pupillometry as a measure of cognitive effort in younger and older adults," *Psychophysiology* 47 (3) 560–569

Shojaeizadeh M., Djamasbi S., Trapp A. 2015. "Does pupillary data differ during fixations and saccades? Does it carry information about task demand?," *In: Proceedings of the Fourteenth Annual Workshop on HCI Research in MIS, Fort Worth, Texas*

Zhan, Z., Zhang, L., Mei, H., Fong, P.S.W., 2016. "Online Learners' Reading Ability Detection Based on Eye-Tracking Sensors," *Sensors*, 16, 1457