Look into my Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation

Thomas Kosch, Mariam Hassib,
Daniel Buschek, Albrecht Schmidt
LMU Munich, Munich, Germany
{firstname.lastname}@ifi.lmu.de

Figure 1: Left: User solving math tasks which induces cognitive workload. **Right:** Measured pupil dilation used for displaying suitable math task complexities

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188643

Abstract
Cognition-aware systems acquire physiological data to derive implications about physical and mental states. Pupil dilation has recently attracted attention in the HCI community as an indicator for mental workload. The impact of mental workload on pupillary behavior has been extensively examined. However, systems making use of these measurements to alleviate mental workload have been scarcely evaluated. Our work investigates the expediency of task complexity adaption based on pupillary data in real-time. By conducting math tasks with different complexities, we calibrate a complexity adjustment system. In a pilot study (N=6), we evaluate the feasibility of changing task complexity using two different complexities. Our findings show less perceived mental workload during task complexity adaptation compared to presenting high task complexities only. We show the potential of pupil dilation as a valid metric for assessing mental workload as a modality for cognition-aware user interfaces.

Author Keywords
Cognition-Aware Interfaces; Workload-Aware Computing; Pupil Dilation; Eye Tracking

ACM Classification Keywords
H.1.2 [Models and Principles]: User/Machine Systems — Human Information Processing

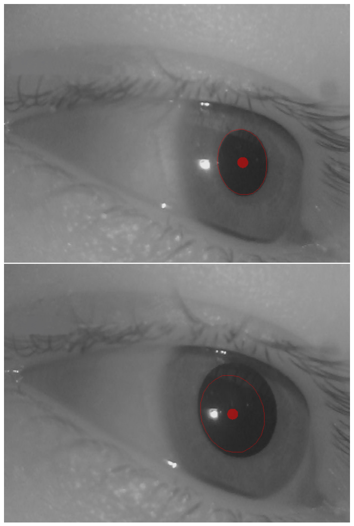

Figure 2: Exemplary impact of cognitive workload on pupil dilation. **Top:** Measured pupil diameter under low cognitive workload. **Bottom:** Measured pupil diameter under high cognitive workload.

Introduction and Background

Context-aware computing has grown in recent years, proliferating devices and day-to-day life. Adapting system behavior according to location, physical activity, or social interactions is accepted and often expected by users. Sensing emotional states has attracted attention as a modality for context-aware computing and has been researched thoroughly in the domain of affective computing [4, 15]. Additionally, the analysis of cognitive processes has great potential for providing adapted content from a mental point of view [5]. Such cognition-aware systems analyze cognitive processes, derive implications from psycho-physiological states, and provide tailored content based on the user's individual cognitive state. Various experiments in Human-Computer Interaction (HCI) depend on measures for quantifying cognitive effort. To specify the complexity of experimental conditions, questionnaires such as the NASA-TLX [7, 8] or the Driver Activity Load Index [13], are used to derive cognitive implications which depend on subjective perceptions of participants. However, these measures of mental task complexities are often generalized for all participants, leading to a lack of individual user dependent cognition-awareness for later designed systems.

Recently, pupil dilation has attracted attention in the HCI community as a measure for cognitive workload (see Figure 1). Demanding the short-term memory frequently over a timespan causes cognitive effort [1, 2, 9, 10, 11], a mental process which extends the pupil diameter (see Figure 2). Eye trackers are used to retrieve eye gaze positions which can be used at the same time to evaluate pupillary data in real-time. Using such a metric as an assessment of cognitive workload can be utilized by applications to provide help and assistance for users when high cognitive workload is detected. For example, cognition-aware systems can provide assistance in real-time during a cognitively demanding task. Furthermore, the usability of interfaces can be evaluated by conjunctively assessing pupil diameter and eye gaze during user interaction. Evaluating pupil dilation to infer cognitive workload has been examined by various researchers before. Benedetto et al. [3] explored the impact of cognitive workload on pupil dilation and eye blink duration. Since pupillary measurements are error prone to lighting conditions, Pfleging et al. [14] proposed a model for classifying cognitive workload of pupil dilation under different lighting conditions. Kiefer et al. [12] evaluated the pupil diameter under different task difficulties to assess the perceived task complexity. Gollan et al. [6] examined the assessment of pupil dilation under cognitive workload in real-time. Hess et al. [10] investigated the impact of simple mental processes, such as solving one and two-digit multiplications, on the pupil diameter. Zbrodoff and Logan [16] found a correlation between higher pupil extensions and increasing math exercise complexity. By using multiplication tasks, a higher pupil diameter was measured when multiplying two-digit numbers than one-digit numbers.

Previous work has explored the influence of changing pupil diameters during different task complexities and possibilities to classify cognitive workload based on pupil extensions. However, the presented theoretical concepts have not been evaluated in the context of a real-world task, where the complexity of presented content is adapted according to the mental resources of the user. In our work, we investigate the impact of adjusting the complexity of a cognitive inducing task in real-time using pupillary data. Task complexity is set to a suitable difficulty to keep users engaged, thus becoming more difficult or easier depending on current mental workload measurements. This is complemented by presenting and evaluating a proof-of-concept application, which sets its task complexity based on pupillary data to prevent mental underload or overload.

Prototype Application

To examine the usefulness of adapting task complexity based on pupillary data, we constructed a prototype capable of evaluating pupil dilation measurements in real-time.

Recording and Feedback Setup

A mobile eye tracker from Pupil Labs[1] is used to receive pupillary data. Data was updated at 30 Hertz and processed on the attached computer. An external screen is used to display stimuli. To avoid distractions during the experiment, the setup was divided into two areas using separators, respectively for the experimenter and participant. The experiment was conducted in a room without windows and constant lighting conditions. The stimulus monitor had a resolution of 1920×1080 and a screen size of 23 inches.

Classifier

Two baseline trials are conducted in the beginning to obtain ground truth data about pupil dilation during an *easy* and a *difficult* math task, each lasting three minutes. A single complexity was assigned to every trial (*easy*/*difficult*). The trained classifier is then used to adaptively control the complexity of the displayed multiplications in a third trial. The classifier aims to estimate low and high cognitive workload based on the data retrieved from the two baseline tasks. The prototype simulates a cognition-aware system by setting task complexity to *easy* if high mental workload is measured. In contrast, task complexity is set to *difficult* when low cognitive workload is classified from the previous baseline measurements.

To assess cognitive workload from the users' pupil dilation, we trained individual support vector machines (SVMs) with a linear kernel to infer required changes of task difficulty in real-time. We used the individual pupil dilation as the only feature. We defined two classes, *easy* and *difficult*. The classifier was trained individually for every participant after the two baseline trials. In the last trial, the previously trained classifier is used to predict cognitive workload. The task difficulty is set to the opposite task complexity to force cognitive alleviation or effort. The person-dependent classifier accuracies ranged between 64$\%$ and 99$\%$ ($M = 79\%$, $SD = 0.16\%$) resulting from a k-fold cross-validation with $k = 5$.

[1] www.pupil-labs.com - last access 2018-02-22

Study

We conducted a pilot study to evaluate the feasibility of changing task complexity based on pupil diameter measurements as an indicator for cognitive workload. The study configuration conforms the previously described prototype.

Cognitive Task

To induce cognitive workload, we use a multiplication math task with two different complexity levels including one-digit and two-digit multiplications. Both complexities lead to different pupil dilations [1, 2, 16]. To add a time constraint, the multiplication math tasks were moving centered from the top to bottom of the screen within five seconds. During the time limit, participants were asked to type the correct solution on a keyboard number pad. When the multiplication reached the bottom of the screen, it disappeared and a new multiplication is displayed at the top of the screen. If the user fails to enter a solution during the given time frame, an error is counted. If the solution entered during the time is correct, the multiplication disappears and a new multiplication task is displayed at the top of the screen. The complexity is divided into *easy* and *difficult*. The *easy* condition uses numbers ranging from $0 - 9$, randomly selecting two numbers which have to be multiplied together. In the *difficult* condition, two numbers are randomly chosen, where the first number ranges from $10 - 19$ and the second from

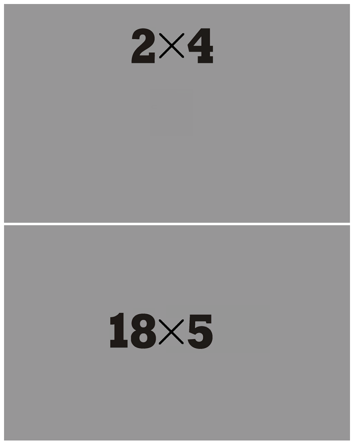

Figure 3: Multiplication math task with equations moving from top to bottom comprising two different complexities. **Left:** *Easy* multiplication containing one-digit numbers. **Right:** *Difficult* multiplication comprising a two-digit number and a one-digit number.

$0-9$ (see Figure 3). Numbers are constantly displayed in a black font on a gray background (RGB: $[150, 150, 150]$).

Participants
Six participants (4 male, 2 female) took part in the pilot study, ranging from an age between 22 and 34 years ($M = 29.17$, $SD = 4.41$). All participants were computer science students or researchers. Participants had normal or corrected-to-normal vision. The overall duration of the study was approximately 15 minutes. Before the start of the experiment, participant signed an informed consent form and provided their demographic data.

Procedure
First, we explained to the participants the study procedure and familiarized them with the setting. The mobile eye tracker was then calibrated. The study consisted of three trials; two baseline multiplication trials (*easy*/*difficult*) and one *adaptive* multiplication trial. The duration of each task was three minutes. The order of the first two baseline trials was counterbalanced according to the balanced Latin square. The pupillary data from the baseline tasks was used to derive an SVM classifier capable to distinguish between cognitive workload induced by a *easy* or *difficult* task. During the *adaptive* trial, the trained user-dependent classifier is used to adaptively set the complexity of the multiplication task. The complexity is set by classifying measured pupil diameter data in five-second intervals. If the classifier estimates a high diameter difference, the task complexity is set to *easy* to avoid mental overload. In contrast, the task complexity is set to *difficult* if the classifier estimates a low pupil diameter to avoid mental underload. If the classifier determines a complexity change during a multiplication task, the new task complexity is adjusted to the next appearing multiplication. After every trial, participants filled a NASA-TLX [8] questionnaire to obtain subjectively perceived workload during the trials.

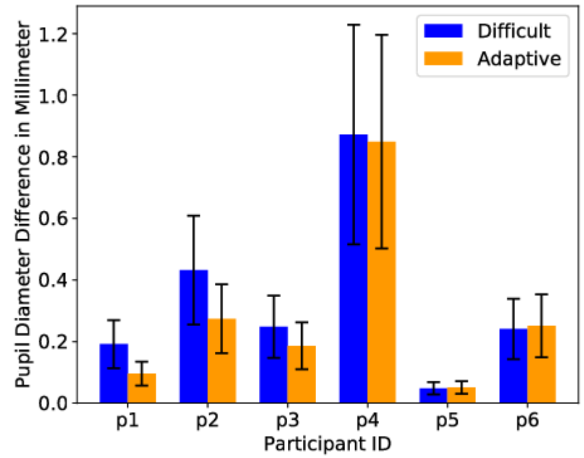

Figure 4: Averaged differences of pupil diameter per participant between *easy* and *difficult* task complexities as well as *easy* and *adaptive* task complexities. The difference between the *easy* and *difficult* task complexity is higher compared to the *adaptive* task complexity except for two participants (p5, p6). The whisker lines depict the standard error.

Exploratory Results
Overall, the *easy* multiplication condition consisted of 618 displayed calculations, the *difficult* condition 420 displayed multiplications, and 466 displayed calculations for the *adaptive* condition. Least errors were measured during *easy* multiplication, comprising 571 correct (92%) and 47 wrong (8%) answers. Most errors were made during *difficult* trials with 255 correct (61%) and 165 wrong (39%) answers. The *adaptive* condition placed itself in between with 313 correct (67%) and 153 wrong (33%) answers. We investigate individual differences in the pupil dilation measures between the different task complexities. The differences in the *easy*

and *difficult* task complexity ranged between 0.05 and 0.87 millimeters. The difference between the *easy* and *adaptive* complexity ranged between 0.05 and 0.85 millimeters. Figure 4 shows the difference between the averaged pupillary measurements per participant. The mean number of complexity switches during the *adaptive* condition results in 2.6 complexity changes for a total of 466 multiplication tasks during the *adaptive* condition. The mean NASA-TLX score (see Figure 5) reveals lowest subjectively perceived workload during the *easy* condition ($M = 9.75, SD = 1.45$) and highest workload measurements during the *difficult* complexity ($M = 12.05, SD = 2.15$). The *adaptive* condition was rated between both baseline trials ($M = 11.30, SD = 2.14$).

Discussion

In this study, we explored the impact of adapting the complexity of a cognitive workload inducing task based on the measurement of pupillary changes.

Adaptive User Interfaces

Our findings support the use of pupil dilation to detect mental effort. Assessing mental workload in real-time to change the task complexity has shown its feasibility when participants performed adapted math tasks. The adaptive trial shows how computing systems can make use of the pupil diameter to create engaging user interfaces while avoiding the perception of high or low mental workload.

Creating User Specific Models

The study shows the viability of individually trained classifiers. Since physiological measurements differ among users, person-dependent measurements cannot be avoided. Limited to pupillary data, encouraging results were achieved by allocating short training times to create a user-specific model. This is supported by averaged pupil diameter measurements and subjective feedback provided through NASA-TLX questionnaires.

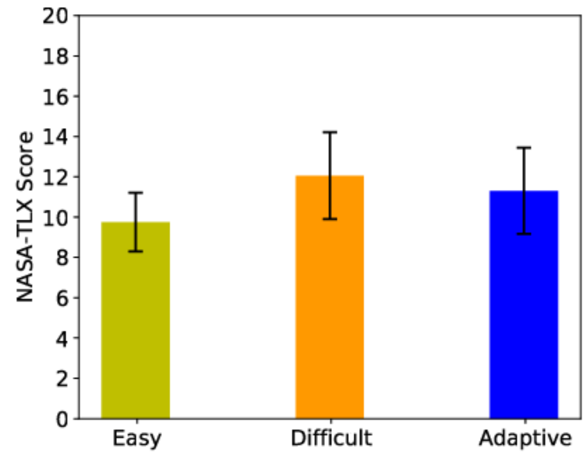

Figure 5: Averaged NASA-TLX scores for the different task difficulties. The whisker lines depict the standard error.

Conclusion and Future Work

This work investigates the assessment of cognitive workload based on pupillary data to change the complexity of a task in real-time. In a preliminary study, we trained person-dependent classifiers through math tasks with different complexities. During an *adaptive* trial, task complexity was changed depending on the classification of pupillary data. The results show the practicability of using pupillary data in controlled environments to evaluate user interface adaptation mechanics and user interface assessment. In future work, we plan to compensate noise introduced by varying illuminations through in situ light measurements. Furthermore, we investigate the suitability of distinguishing multiple task complexities in a second study. This includes a large scale sample size to derive a general classifier which can be dynamically deployed to evaluate user interfaces.

References

[1] Mark H Ashcraft and Elizabeth P Kirk. 2001. The relationships among working memory, math anxiety, and performance. *Journal of experimental psychology: General* 130, 2 (2001), 224. DOI: http://dx.doi.org/10.1037/0096-3445.130.2.224

[2] Mark H Ashcraft and Jeremy A Krause. 2007. Working memory, math performance, and math anxiety. *Psychonomic bulletin & review* 14, 2 (2007), 243–248. DOI: http://dx.doi.org/10.3758/BF03194059

[3] Simone Benedetto, Marco Pedrotti, Luca Minin, Thierry Baccino, Alessandra Re, and Roberto Montanari. 2011. Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour* 14, 3 (2011), 199–208. DOI: http://dx.doi.org/10.1016/j.trf.2010.12.001

[4] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *CSUR* 46, 3 (2014), 33. DOI: http://dx.doi.org/10.1145/2499621

[5] Andreas Bulling and Thorsten O Zander. 2014. Cognition-aware computing. *IEEE Pervasive Computing* 13, 3 (2014), 80–83. DOI: http://dx.doi.org/10.1109/MPRV.2014.42

[6] Benedikt Gollan, Michael Haslgrübler, and Alois Ferscha. 2016. Demonstrator for Extracting Cognitive Load from Pupil Dilation for Attention Management Services. In *UbiComp'16*. ACM, New York, NY, USA, 1566–1571. DOI: http://dx.doi.org/10.1145/2968219.2968550

[7] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage Publications, 904–908.

[8] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9

[9] Richard P Heitz, Josef C Schrock, Tabitha W Payne, and Randall W Engle. 2008. Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology* 45, 1 (2008), 119–129. DOI: http://dx.doi.org/10.1111/j.1469-8986.2007.00605.x

[10] Eckhard H Hess and James M Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 3611 (1964), 1190–1192. DOI: http://dx.doi.org/10.1126/science.143.3611.1190

[11] Daniel Kahneman and Jackson Beatty. 1966. Pupil diameter and load on memory. *Science* 154, 3756 (1966).

[12] Peter Kiefer, Ioannis Giannopoulos, Andrew Duchowski, and Martin Raubal. 2016. Measuring cognitive load for map tasks through pupil diameter. In *International Conference on Geographic Information Science*. Springer, 323–337. DOI: http://dx.doi.org/10.1007/978-3-319-45738-3_21

[13] Annie Pauzié. 2008. A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems* 2, 4 (2008), 315–322. DOI: http://dx.doi.org/10.1049/iet-its:20080023

[14] Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. 2016. A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. In *CHI'16*. ACM, 5776–5788. DOI: http://dx.doi.org/10.1145/2858036.2858117

[15] Jianhua Tao and Tieniu Tan. 2005. Affective computing: A review. In *International Conference on Affective Computing and Intelligent Interaction*. Springer. DOI: http://dx.doi.org/10.1007/11573548_125

[16] N Jane Zbrodoff and Gordon D Logan. 2005. What everyone finds: The problem-size effect. (2005).