Applied Ergonomics 75 (2019) 99–107

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Pupillary measurement during an assembly task

Per Bækgaard[a,*], Shahram Jalaliniya[b], John Paulin Hansen[c]

[a] DTU Compute, Technical University of Denmark, Richard Petersens Plads, Building 321, DK-2800 Kgs, Lyngby, Denmark
[b] Department of Computer Science, IoTaP Research Center, Malmö University, Sweden
[c] DTU Management Engineering, Technical University of Denmark, DK-2800 Kgs, Lyngby, Denmark

ARTICLE INFO

Keywords:
Assembly manual
Eye tracking
Pupil dilation

ABSTRACT

We conducted an empirical study of 57 children using a printed Booklet and a digital Tablet instruction for LEGO® construction while they wore a head-mounted gaze tracker. Booklets caused a particularly strong pupil dilation when encountered as the first media. Subjective responses confirmed the booklet to be more difficult to use. The children who were least productive and asked for assistance more often had a significantly different pupil pattern than the rest. Our findings suggest that it is possible to collect pupil size data in unconstrained work scenarios, providing insight to task effort and difficulties.

1. Introduction

Mental workloads are commonly measured by subjective ratings (Reid and Nygren, 1988), which risk being rationalizations made in hindsight, or using dual-task paradigms (Wickens, 1991) that interfere with execution of the primary task. EEG provides another source of data for studying continuous cognitive load (Klimesch, 1999), but placing electrodes on a participant's head can be a challenge, and subjects wearing them may feel awkward. Likewise, Galvanic Skin Response (GSR) and heart rate measurements require sensors to be attached to the subject, which may hinder mobility.

Pupillometry holds potential as an unobtrusive way to measure the cognitive effort associated with a given task. Gaze tracking is becoming low-cost and is likely to be integrated with future displays. However, in real task situations, one of the main challenges is to distinguish the rather large pupil reactions caused by unknown and uncontrollable changes in the ambient light from the minuscule dilations that reflect changes in cognitive effort. Another challenge is the large individual differences in pupil size and in people's reactions to task difficulties. This paper seeks to clarify whether pupillary measures can be used to analyse user experiences in-the-wild. The main research question addressed is whether noisy pupil data can provide information about the user experience and task engagement for a given population.

2. Theory

2.1. Pictorial assembly instructions

The preferred medium for step-by-step instructions has been print on paper. In recent years, new digital forms of instruction material have emerged. Apps with interactive building instructions have been provided by LEGO® for some years. When downloaded to a PC, Tablet or smartphone they offer instructions on how to build a model. Sequential navigation is done by touching a forward or a backward button located in the corners. Details can be examined by zoom and rotation; animations show how to place a component and a forward/rewind slider helps find a particular event in the construction sequence (cf Fig. 1) (see Fig. 2).

We expect digital instructions to become more common in the future because they offer easy updating, extended explanations (e.g. more steps and animations) and may be less costly to produce as a large amount of products are routinely modelled in 3D for design, production, and marketing.

2.2. Previous work

A number of human factor studies of building instructions have been reported (Table 1). Common interest was the impact of different media and display forms (Tang et al., 2003; Henderson and Feiner, 2011; Alexander, 2013; Wille et al., 2014; Funk et al., 2015) and how to best present the pictorial instructions (Pillay, 1998; Rodriguez, 2002; Richardson et al., 2004; Martin and Smith-Jackson, 2008). A few

* Corresponding author.
E-mail addresses: pgba@dtu.dk (P. Bækgaard), shahram.jalaliniya@mah.se (S. Jalaliniya), jpha@dtu.dk (J.P. Hansen).

https://doi.org/10.1016/j.apergo.2018.09.004
Received 1 November 2017; Received in revised form 10 August 2018; Accepted 7 September 2018

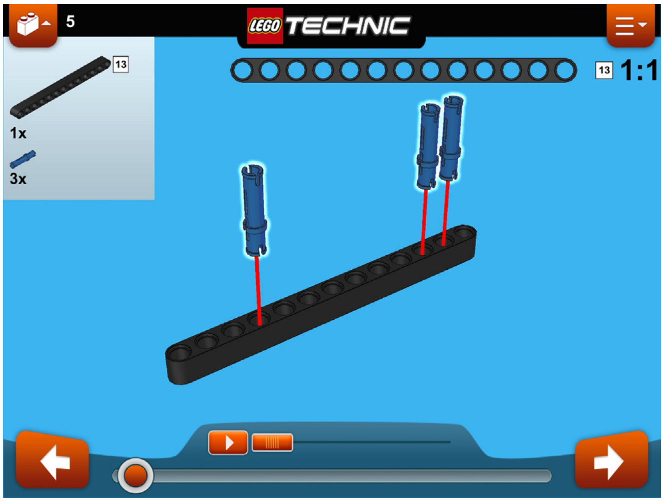

Fig. 1. LEGO® TECHNIC digital building instruction. Arrows in the corners step backward (left corner) and forward (right corner). The play button (middle section) offers a short animation of where to place the two pieces. The slider next to the play button allows the user to control the animation frame-by-frame. The bottom slider with the circle serves as an interactive progress bar. Two-finger pinch on the touch-screen will zoom, and moving the finger on the Tablet will rotate the model.

studies (Martin and Smith-Jackson, 2008; Gupta et al., 2012) looked at the influence of individual factors like age, gender, education and building experience. The measurements most often taken were task time and errors; while five studies (Tang et al., 2003; Richardson et al., 2004; Martin and Smith-Jackson, 2008; Alexander, 2013; Wille et al., 2014) included subjective ratings of experienced difficulty, mental effort, fatigue and/or satisfaction and two (Henderson and Feiner, 2011; Alexander, 2013) asked for the participants' preferences. Interestingly, three studies (Pillay, 1998; Richardson et al., 2004; Martin and Smith-Jackson, 2008) included a manual record of visual behaviour (in terms of "number of looks", "study time", "viewing time" and "gaze time") but without the use of gaze tracking equipment.

The potential benefit of using eye activity to measure cognitive workload in tasks has been addressed in several previous studies. For instance, Van Orden et al. (2001) presented a display with target density as the workload variable. Blink frequencies, fixation frequency and pupil diameter showed strong correlation with the density of targets. The changes in pupil size from 1 to 9 targets were highly significant, even though the average change in actual size (calculated as moving estimates of means over a 2 s window) was less than 1 mm. Ahlstrom and Friedman-Berg (2006) found no significant effects of task conditions on subjective workload ratings, but significant effects on blink durations, which became shorter when conditions were difficult. The mean pupil diameter increased from 2.4 mm to 3.9 mm for the least responsive subject, while the most responsive had an increase from 2.62 mm to 4.39 mm. The correlation between number of objects (i.e. aircraft) to be supervised (in a simulated air traffic controller operation) and the pupil dilations was rather high (i.e. $r^2 = 0.7$). They concluded that measurements of eye activity provide a more sensitive measure of workload over task time than subjective ratings, and suggest this to be particularly relevant when trying to identify display components that cause workload changes which are not reported by operators themselves. In a study (Bhavsar et al., 2015) the variations of operators' pupil size were found to be a reliable indicator of the perceived mental workload during a simulated plant emergency; subjects with low, moderate and high task loads showed distinguishably different dilation patterns across the incident.

Kiefer et al. (2016) found significant differences in mean pupil diameter between map tasks conducted by subjects. The authors interpret the results as an indication of low cognitive load when exploring maps freely, while route planning and focused search entailed a high cognitive load.

Dehais et al. (2008) conducted a study with six pilots flying under low-light (nightfall) conditions with a remote gaze tracker mounted in the cockpit. They observed a large dilation after a simulated engine failure compared to the pilots' pupils during normal flight. In a study presented by Palinko et al. (2010) performance data recorded in a driving simulator (i.e. variances of steering wheel angle and lane position) showed a high correspondence to changes in pupil size, even without explicit control of the lighting conditions.

Orlandi and Brooks (2018) found that marine pilots had higher pupil scores for difficult berthings compared to easy berthings and a light correlations between a self-assessment Likert scale and pupil dilation (r = 0.243).

Recent studies by Čegovnik et al. (2018) and Hansen et al. (2018) has found that even a low-cost 30 Hz remote gaze tracker may provide reliable pupil measures that corresponds well with changes in task load during simulated driving and assembling of construction toy, respectively.

2.3. *The pupil as an indicator of cognitive effort*

The size of the pupil varies from 2 mm to 8 mm across different subjects and light conditions (Walker et al., 1990). Large variations (3 mm) within subjects are caused by changes in light levels controlled by the pupillary light reflex, as reported by e.g. Ellis (1981).

However, Hess and Polt (1964) in a now famous experiment reported that pupil dilation could also be used as an index of mental activity during multiplication problems. This was based on earlier work (Hess and Polt, 1960) relating pupil size to emotional interest in presented material. Kahneman and Beatty (1966) subsequently confirmed this finding in a separate study, further suggesting that pupil responses are indicative of memory and processing load. This effectively (re)introduced pupillometry as a discipline (Beatty and Lucero-Wagoner, 2000; Stanners et al., 1979; Laeng et al., 2012).

Pupillary dilations in response to cognitive processes can be small (Beatty and Lucero-Wagoner, 2000) in less than extreme test conditions, often only around 0.5 mm, which in typical conditions may amount to around 15% of the baseline pupil size. It is therefore common to record several responses from repeated presentations of the same stimuli and conditions, and calculate the average of these in order to filter out noise from other processes. These are referred to as *Task-Evoked Pupillary Responses* (TEPR), a term probably coined by Beatty (1982) and Ahern and Beatty (1979).

Marshall (2002, 2007) has proposed a (patented) technology, the Index of Cognitive Activity, based on filtering abrupt pupil dilations using wavelet transforms to index and identify cognitively induced phasic responses from constrictions and dilations caused by the pupillary light reflex. In the present paper, we instead focus on the simpler concept of averaged means of the pupil size; a concept which reflects not only phasic responses but rather a combination of the current tonic level (arousal) and any phasic activations that take place. It may therefore serve as an index into a combined *level of cognitive effort* similar to what Hyönä et al. (1995) refers to as *global processing load* or the *pupillometric estimate of mental load*. Similar ideas have also been proposed elsewhere, e.g. Iqbal et al. (2004) *percentage change in pupil size* or Palinko et al. (2010) *mean pupil diameter change*.

3. Materials and methods

3.1. *Participants*

57 children (54 boys and 3 girls) between 8 and 10 years-old (average 8.3 years) were recruited for the experiment at a LEGO® exhibition. Two were Swedish, the remaining 55 were Danish. Only one participant used glasses; none used contact lenses. All except five were