Insights on Pupil Dilation, Interaction Technique and Effort

Abstract
Eye Tracking has been successfully applied in Human-Computer Interaction, be it as input mechanism or as measurement of quality e.g. for immersion [6]. In order to validate the quality of interaction modalities of a system, questionnaires or performance metrics e.g. task completion time can be used but these approaches only evaluate the overall quality of a system. Our experiment show that the mean Pupil Dilation increases upon performing a specific interaction technique with a public display ensemble, even more when an interaction does not lead towards the expected outcome – e.g. fails to perform the intended action. In turn, we have established a connection between Pupil Dilation, Interaction Technique and User Reported Effort Ratings in our experiment with a public display ensemble.

Author Keywords
Evaluation Techniques, Interaction, Usability, Pupil Dilation

ACM Classification Keywords
H.5.0 [Information interfaces and presentation (e.g., HCI)]: General

Acknowledgements
This work was supported by the 360Light project.
(FFG, Contract No. 838547)
Introduction
Eye Tracking has mainly been used to evaluate HCIs by analysing eye gaze behaviour such as fixations and saccades, see [13] for an elaborate overview. Yet, these evaluations focus on interface design and task performance and not on interaction techniques. The later is especially crucial for seldom used interaction modalities, e.g. found in a spontaneous interaction between public display and passerby. Public displays can be subject to display blindness – because their content is not attracting user attention [10] – or interaction blindness – because users are unaware that they can interact with the display [5]. While both problems have been studied in recent years and are already challenged, it is – then – crucial for a public display to choose interaction modalities which are effortless for an untrained passerby, otherwise the user will get frustrated and walk away. This effort can be monitored and evaluated using eye tracking prior to an actual deployment in the field as shown in this paper.

Early Psychophysiology work already revealed a connection between pupil dilation and cognitive load [3]. Eye tracking researchers are already using this connection for their studies [8, 11] to perform task and interface evaluation. Recent advances also show that pupil dilation changes can be retraced to a single events [16]. Previous research [14] revealed that pupil dilation signals surprise – to experience unexpected behaviour. In line with the simulation ‘hypothesis’ [4] it can be said, that if the users simulation of (expected) behaviour does not align with the later on perceived behaviour, the user has to interrupt his ongoing chain of behaviour [2] and invest cognitive resources to realign simulation and real world, an effect we experience as effort [7] and measurable in pupil dilation. To the best of our knowledge no connection between pupil dilation, a specific interaction techniques and its (un)expected behaviour and effort in HCI has been previously drawn.

In this paper we will investigate this connection with a conducted experiment. In an eye-tracking study, we observed participants performing hand gesture based interaction with our self-developed public display ensemble, in combination with an usability questionnaire in which the participants reported their perceived subject invested effort.

A Public Display Experiment
The public display ensemble – the 360Light System – used in the experiment consisted of five independent public displays. Each of which contains an independent computing system and an infrared depth sensor (Kinect v2.0), c.f. Figure 1. The visualization software was written in C++ using the high level open source 3D engine “Irrlicht”. Each displays has its own internal state which is then synchronised via networking and a self-developed software framework. Interactions between system and operator were implemented using dedicated mid-air hand gestures via modern, commercial infrared depth sensors and finally mapped to system-side reactions such as panning, zooming, selecting or navigating. The following fundamental hand gestures were used: (i) Push - A fast forward and backward movement of the hand; (ii) Move In/Out - the hand is extended towards the display or retract from it; (iii) Navigate - the hand which controls a cursor is positioned to the far left, right, top or bottom.

In order to evaluate our system an usability study with 22 voluntary study participants (age: $\mu=23.4$ $\sigma=3.7$; gender: 20 male, 2 female) was conducted and evaluated using an adapted version of Purdue Usability Testing Questionnaire [9]. On a 7-step Likert scale, study participants scored their prior experience with gesture-based interaction as $\mu=3$, $\sigma=1.8$ (1 signifying no prior experience) and their prior experience using public displays as $\mu=2.1$, $\sigma=1.36$. 2 out of
the 22 participants were unable to use the eye tracking glasses due to severe amblyopia. The experiment was conducted in a controlled environment to limit external influences on the setting. During the experiment, test subjects were asked to perform 9 usability tasks, designed for three specific content types typically found in such public display systems: panoramic & high-resolution images, 3D environments and a reading corpora. Each usability task consisted of a sequence of interactions that had to be completed by study participants while the experimenter guided and the eye tracking glasses recorded the procedure. The overall goal of this study was to test the psychophysiology theory of pupil dilation, expected and unexpected behaviour and effort [3, 14, 16, 4] and to see if their connection is also visible and usable in a much less constrained and controlled environment and hence applicable towards evaluating public displays or HCIs in general.

Evaluation and Discussion
While the system showed an overall satisfactory performance on a 7 Step Likert Scale in the questionnaire ( "Would you like to use the system again?": μ=5.9, σ=1.15 (Higher is Better)), we were intrigued by interaction errors and performance, which were most obvious – to the experimenter – during push gestures, c.f. Table 3, and likely influenced the effort stated in the questionnaire by the users (μ=4.68, σ=1.46 (Lower is Better)). While push gestures are easily detectable by a human observer – not only by watching the execution of the gesture, but also by looking at secondary clues like eye gaze [1] or by knowing the next work-flow step – they are hard to detect for computers [12]. We define successful pushes as pushes which were intended by the user and recognised by the system (Succeeded Push). However most push interactions errors observed were, on the one hand, due to the fact, that the user did not perform the push with enough force or at an unusual angle, hence the system did not react as expected (Failed Push). On the other hand, the system sometimes recognised a push without the user actually intending one (System Push), e.g. because the user turned around its own axis with the hand raised.

An illustrative behaviour of a (normalised) pupil response towards expected and unexpected push events (Succeeded and System Push) can be seen in Figure 2. Since pupil dilation is affected by different light settings [15], the field of view camera on the eye tracker was used to estimate the perceived luminance (RGB to LUM formula from the W3C http://www.w3.org/TR/AERT#color-contrast) for each videoframe and the corresponding pupil dilation event. Luminance invariant pupil dilation value was derived by dividing pupil dilation with luminance.

Pupil diameters with less then 2mm were discarded [15] and normalised to a [0, 1] range using feature scaling ($x' = \frac{x - \min(x)}{\max(x) - \min(x)}$). We aggregated the mean pupil dilation over a 4 second window after a "push" event occurred, which will contain the maximum peak dilation according to [16]. Additionally we derived a ground truth by calculating the mean pupil dilation for all other data recorded during the experiment. The collected and derived data is shown in Table 1. If less than 2 events of a type occurred for a user the mean value of that event was discarded and is not displayed. The upper half of the table contains the normalised mean pupil dilation for each event type (Succeeded, Failed & System Push and Ground Truth). The lower half contains the difference in % from ground truth value for a given study participants, whereas Ground Truth’s mean pupil dilation is defined as 100%.

For 18, out of the 19 study participants who were
eyetracked, the pupil dilation increased after execution of a Succeeded Push. Only one study participant had a lower mean value for pushes than the ground truth (difference of 0.011). For all of the 15 study participants who experienced a Failed Push their mean pupil dilation was indeed higher than the ground truth and also higher than for Succeeded Pushes. For all of the 2 participants who experienced System Pushes the mean pupil dilation was also higher than for all other event types or the ground truth. This establishes the fact that pupil dilation reacts to push events and reveals if the perceived – system – is also expected behaviour (Succeeded Pushes lower mean value) and not unexpected (Failed and System Pushes higher mean value) – from a user’s perspective – and confirms that the theoretical foundations [14, 16, 4] are useable in a public display and HCI evaluations. When looking at user reported effort ratings in the questionnaire, c.f Table 1, it can be seen that high differences between ground truth and Succeeded Push’s mean pupil dilation correlates. For the study participants where this is not the case a high rating of either Failed Push or System Push events seems to be the cause of the high effort ratings. The measured effect on pupil dilation, can be linked to the cognitive resources invested by users to realign their simulation of the world with perceived events – their effort. The data also shows that the more unexpected the behaviour – System Pushes vs Failed Pushes mean value – is, the higher the measured effect in pupil dilation is. This in turn confirms the applicability of the psychophysiological theory stated in [14, 3, 16, 4] in HCI and public display research.

**Conclusion**

While the psychophysiology theory of pupil dilation, expected and unexpected behaviour and effort [3, 14, 16, 4] has already been established. This connection has never – to the best of our knowledge – been used for evaluation of HCIs. Our experiments confirmed the applicability of the theory and established that pupil dilation (i) responds to push events indicating cognitive selection process, (ii) reveals that the perceived behaviour of a system is also the expected behaviour by a human indicating cognitive resources spend on realignment between simulated and real world and (iii) it can be used to derive effort of an interaction technique. Future work should investigate how this manual process can be adapted using machine learning towards a fully automated interaction evaluation.
References


