The Eyes Never Lie: The Use of Eye Tracking Data in HCI Research

Daniel Bruneau, M. Angela Sasse, John McCarthy
Department of Computer Science
University College London
Gower Street
London WC1E 6BT
{d.bruneau, a.sasse, j.mccarthy}@cs.ucl.ac.uk

ABSTRACT

Recent technology developments have made eye-tracking systems more affordable and less cumbersome. Many members of the HCI community will therefore consider employing this method in usability research and assessment. This paper, provides an overview of the eye movement data that can be collected, what is involved in analysing them, and what it can add to current HCI research and usability assessment practice. We argue that, whilst much work is required to develop the use of eye-tracking as a practitioner's method, it can provide valuable objective data about the impact of visual design on human performance. To illustrate this point, some preliminary results of a study into web page usage are reported. The results are examined with regard to explaining task performance, user physiology as well as in the assessment of the difficulties and the potential behind eye-tracking.

Keywords
Eye tracking, physiological methods, usability evaluation

1 INTRODUCTION

Eye tracking is a technology that enables some feature(s) of the eye to be tracked visually by a camera or imaging system, delivering real-time coordinates of the user's gaze. Since eye movements are generally thought to be involuntary, eye tracking provides objective data of users' visual interaction with a system [3]. It has been used over the past two decades in a number of military and academic research institutes. Until recently, the equipment was prohibitively expensive, required that participants keep their head still (bite bars being the most common method to ensure this), or wear (often heavy) devices on their head. Recent developments have made the technology more affordable [3], and less intrusive for participants. LC Technology's Eyegaze system [6], for example, requires no equipment to be attached to the person. This means that an increasing number of HCI researchers and practitioners can consider employing eye- tracking. In this paper, we provide our view of the research to date: the eye movement data that can be collected, what is involved in analysing them, and what insights HCI researchers and practitioners should be able to derive from them.

2 EYE-TRACKING RESEARCH: AN OVERVIEW

The majority of studies on eye-tracking to date have been concerned with using eye-tracking as an input device (e.g. [14]), rather than as a data collection tool. The most commonly used eye movement data to assess the impact on participants are time and location measures of *fixations* and *saccades*. At a physiological level, changes in *arousal* can be detected in latency, number and frequency of saccades, and timing and frequency of fixation. *Fatigue* and *vigilance* can be examined in saccade speed, duration and precision, and the number of glissadic movements [13]. At the perceptual level, information processing activities can be assessed through a number of fixations and saccades, and accumulated duration of fixations.

There have been only a small number of studies of eye movements in evaluating user interfaces; [4] suggests that *fixation duration* and *scanpaths* are key indicators of users' interest in certain elements of the user interface. [13] lists a total of 31 eye tracking measures that could be employed to assess the distribution of information in the user interface, visual search processes, information processing, and to detect when the user is engaged in perceptual and cognitive activities.

This brief overview illustrates the first challenge HCI researchers face with eye tracking: to decide which measurements are reliable indicators for evaluating the impact of on users. Another challenge is how to summarise and analyse the vast amount of data that eye trackers can produce. A significant challenge at present is that there are currently only a few software tools that help researchers to analyse and present the data in a coherent manner [9,15]. Finally, less restrictive tracking equipment is also less reliable in recording, which means that the data of large numbers of subjects have to be discarded [15].

3 RATIONALE FOR APPLYING EYE TRACKING IN HCI

Despite these methodological challenges, we predict that eye tracking will become widely used in HCI research and practice, because it provides objective data on the physiological and perceptual impact of interaction, that have so far remained untapped. The need for objective data on user cost, for instance, has been highlighted [18].

3.1 Detecting Fatigue and Strain

Fatigue and *strain* are two well-known concepts within work psychology, but are rather neglected within HCI paradigms. Even though Shackel's [16] definition of usability includes *user cost*, this element has been rather neglected in usability assessments in the past 25 years. A more objective measure of user cost is required, as subjective methods in isolation have a drawback in that they are cognitively mediated. Physiological measures such as Heart Rate (HR), Blood Volume Pulse (BVP), and Galvanic Skin Response (GSR) have been successfully used in assessing the impact for video [9] and for a number of audio degradations in non-interactive tasks [19].

Eye tracking data provides an additional way of measuring user cost objectively, in applications such as web pages and video. We argue that fatigue and strain can be examined together using different eye movements. It is important to note that strain can be linked to instances of information overload with the resultant effect of the user having to work harder to accomplish a task. Thus the user may become more irritable and more long-term physical stress can set in. Indeed it has been found that pupil diameter decreases with fatigue [5,7]. Thus pupil constriction reflects fatigue in a work environment and could be related to strain and the impact of differing quality levels in a video conferencing session or on a complex web page layout.

Another eye movement that can be utilised is *eye blinks*. Eye blinks are closely related to psychological factors including mood state and task demands and reflect the viewer's attention and tension. While it can be logically argued that the causes of changes in pupil size and blink can be difficult to specify, since all components within the act of viewing can realistically have an influence on the user, eye-movement and changes in pupil size and blink can show where the user is gazing within a temporally and spatially orientated domain [9].

Differences in re-visits of the eye or backtracking to certain locations on the screen may also occur as a result of fatigue and strain and may be an indicator of the user having difficulty in accomplishing a task [17].

3.2 Affective values

A key concept in HCI usability is the role that emotion plays. For instance, the emotional impact of different user interfaces to enhance levels of motivation in the workplace has been assessed [8]. Furthermore, research at MIT [12] has demonstrated that physiological signals can be used in computer recognition of affective information. If we are to assume that the context of affect in HCI usability evaluation is to induce and maintain a positive mood, then there are a number of different eye movements that can be used to examine this. For example, pupil diameter changes are related to positive affect (pupil dilation) and negative affect (pupil constriction) [11]. The controversial nature of the relationship between affective processing and pupil size has been highlighted by [17] who suggest that systematic studies on affective pupillary responses seem to be small in number, a view that is keenly supported by this paper. In a study examining emotionally provocative sound stimuli, the results suggest that there are differences in pupil size both during and after different kinds of emotional sound stimuli when they are played [17]. The results are discussed in the context of possibly utilising pupillometry as a computer input signal for affective computing. A difference in re-visits of the eye or backtracking to certain locations on the screen has been illustrated [4]. What needs to be examined within the context of HCI and affective values is whether or not these re-visits may also occur as affective mood changes. Thus these different measures may be implemented to assess the impact of different media setups within HCI.

3.3 Explaining Task performance

A key eye movement measure using the eye tracker system is the examination of *scanpaths*. Scanpaths explicate the path of movement that the eye makes on a screen or on the page of a book. With regard to task performance, it is proposed that number and duration of fixations, and particular patterns of scanpaths could be related to differences in efficiency in task performance. One of the few studies in this area [4] was limited to evaluating screen layouts. There is also a lack of research examining task performance within a wider usability context where the real impact on users is examined as task complexity varies.

4 Experiment: Examining Task Performance- A Web based Study

To investigate this issue further we examined 4 web sites within the context of the most common uses of the Internet. Eye tracking measures provide objective measures of where attentional resources are being allocated. We also examined whether differences in eye movement would be paralleled in physiological changes. The initial results of the experiment are discussed below.

We investigated user interaction with four web sites – the home pages of the 4 largest Internet Service providers (ISPs) in the UK. They have a web base of over 7 million users. Figure 1 illustrates the four web pages with specific *Regions of Interest* (ROI) identified. ROI's are classified by function based on Nielsen's classification [10] and are simply highlighted regions on the web page considered by the experimenter to hold some salient value. The tasks were based on a recent survey of web users and the activities they typically perform when using the worldwide web [2]. The most common uses of the Internet are:

1) To search for something
2) To check for news/sports headlines
3) To shop
4) To write/receive personal e-mail

Figure 1


1
2
3
4
Freeserve
no problem
TISCALI
FOR £20 ON SELECTED BOOKS
egg:card
Egg lets you be...
AOL.CO.UK
AOL Members' Email Login
FREE reasons to try AOL
CLICK HERE TO SIGN ON TODAY
Search the Web
People & Chat
Shop@AOL
AOL.co.uk Today
Web Centres
BTopenworld
Now's the time to apply.
Click Here Now
Mega Deals
excite

4.2 Results

Below are some preliminary results from the present study.

(1) Fixations per region of interest

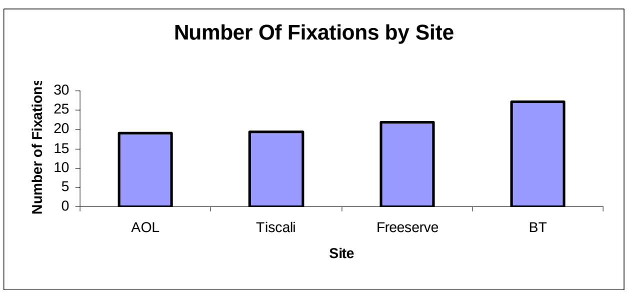
The table below shows the number of fixations for the most salient region of interests for each website for Figure 1

ROI	Number of fixations for region	Total number of fixations for site
1	535	2745
2	783	2278
3	1056	2267
4	995	2773

(2) Fixations Correlate with Site Usability Reports

In a post-experimental interview it was found that the majority of participants rated AOL as being the most usable site. An objective examination of the four web sites using eye-tracking data parallels this finding. AOL had the least number of fixations across the four sites – indicating less processing and leading to faster task completion times. (See Figure 2)

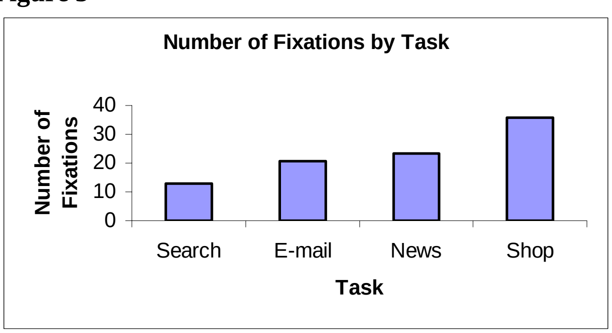
Figure 2



(3) Fixations Correlate with Task Difficulty Reports

The post-experimental interview also highlighted the fact that participants found the shopping task to be the most difficult. By examining the number of fixations for the different tasks, it was found (see Figure 3) that the shopping task (asking participants to search for a digital camera) was significantly different from the other three sites in that fixations significantly increased.

Figure 3



(4) Is Attention Independent of Layout?

The four sites examined vary greatly in their design and layout. To quantify these differences we classified "screen real-estate" into three functional areas based on [10]:

Search/Navigation, Content and Advertising/ Promotional.

Figure 4 illustrates the proportion of screen area devoted to each function across the four sites.

Figure 4

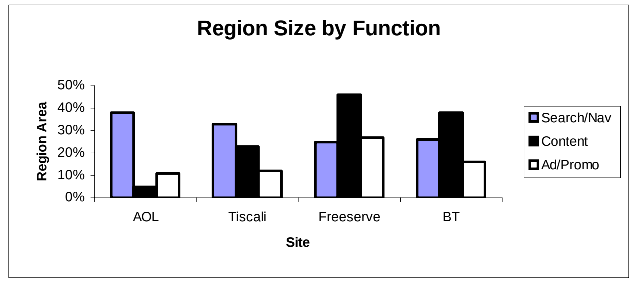
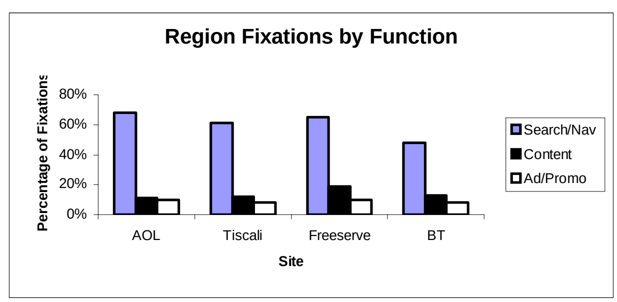


Figure 5 shows the distribution of fixations across these functional areas. The proportion of fixations allocated to each functional region is remarkably similar despite clear differences in spatial layout (Figure 4). In other words, the attention focused on a particular function seems to be largely independent of the screen space devoted to that function.

In this study, around 50-60% of fixations are distributed on Search and Navigation screen areas (see Figure 5). This is a reflection of the task given to users, which is primarily a search task of the form –
"Where would you go on this site to find X?"

Figure 5



With less directed instructions we might expect more attention to the content and advertising regions of the page.

User Physiology and Mood

To analyze the GSR data we examined the extreme high and low GSR's to see if there was any association with task and site. This investigation revealed that one specific task - site combination (i.e. task = search for e-mail and site = AOL) gave the highest GSR responses with an unusually high frequency. The probability that this was due to chance alone is < 0.01.

Finally, concerning the analysis of the Profile of Mood Scale questionnaire administered pre- and post test, of the seven different ratings for mood, only the hostile-agreeable ratings scale proved to be significant.

5 Discussion

5.1 substantive results

The close relationship between subjective reports and eye tracking data gives some substance to our initial argument that there must be a mutual dependence upon both subjective and objective measures within HCI evaluation techniques.

An intriguing question arises which concerns the general distribution of attention to screen regions and why this did not differ although there were clear layout differences between the web sites. The simplest explanation would be task based. Since the participants were asked to search for something, it is logical to assume that fixations would be focused in the search and navigational areas of the web page as was seen in Figure 1.4

This may apply when tasks are straightforward and follow pre-existing schemas of what usually exists on a web page (e.g. these pre-existing schemas may include the idea that ISP web pages usually contain elements of e-mail, news and so forth). For less straightforward tasks such as shopping for a digital camera the number of fixations increases dramatically. Shopping for a digital camera was ambiguous in nature because none of the web sites had a specific link to a camera section. However, since participants were asked to look for a camera, it can be assumed that their goal was to specifically find an icon or word for a digital camera. Once they realized that an icon or word for such an item did not exist, then reliance upon existing schemas most likely came into play. However, even at this stage, a number of possibilities and the opportunity to switch between schemas were offered to the participant: they could either click on the shopping icon or go to the search bar. Both types of behaviour where observed in participants.

Concerning the GSR scores, where the highest GSR scores are associated with the E-mail task on AOL, we have the first indication of a relationship between GSR and eye movements. For all sites except AOL the Navigation bars received the most fixations across tasks. However, on AOL the E-mail login region received the most fixations (see Figure 1). The extensive processing of this region gives rise to large GSR spikes when participants are asked to locate the region. The affective interpretation of this finding is unclear but the finding is completely opposite to our expectation that elevated GSR would be associated with poor usability. AOL was the fastest of the four sites on the e-mail task.

.

5.2 methodological issues

Certain methodological constraints may have led to the increased "hostility" of the participants. To control for learning effects, we required participants to complete complex sequences of task goals, based on a Latin squares design. This constant goal switching caused confusion among some participants to the extent that a few forgot what they were supposed to be looking for. An objective of future studies is to try to evaluate learning effects in the context of a more naturalistic task rather then attempt to eliminate or control for them.

5.3 Future Work

To deal with the large volume of data from the Eye-Tracker the screen area was divided into a distinct set of regions. This provides an indication of the scan path taken by the eye and simplifies the analysis of raw traces. Of particular interest is the temporal progression of the scanpath. Important components that can be measured include backtracking or re-visits where empirical evidence is provided for the number of times the eye goes back to a certain region on a screen. The potential for this type of measure is logical; more re-visits to a particular area would clearly signify some cognitive interest in that particular area of the screen. However, if re-visits occur in an area other than that specified by the task, then there is clearly some aspect of the overall design that is not fulfilling the user's requirements.

Finally, integration of pupil diameter changes and eye blink measurements also need to be taken into account for future studies. The challenge is to integrate these measures with minimal obstruction to the participants, so that they remain as unaware as possible of the different tools being used to measure numerous different types of eye movements. This would enhance ecological validity of the experiment and follows from our earlier suggestion that a greater emphasis on more naturalistic tasks is needed.

6 Summary

The type of eye movement metric that is used to assess a particular aspect of a web site study depends on the question being asked. It is clear from the initial results that there is still a long way to go in establishing eye tracking as a concrete objective measurement tool within HCI, however, this does not hide the fact that the potential behind its usage is enormous and the insights gained from this first study provide strong evidence for this, particularly concerning the methodological issues. The continued purpose is not to use eye-tracking measures in isolation, but together with qualitative methods, and more established physiological measures.

Acknowledgements

Many thanks to Gillian Wilson for her useful comments. This research is funded by BT.

REFERENCES

1. Andreassi, J.L. (2000). Psychophysiology*: Human behavior and physiological response* (4th edition). Lawrence Erlbaum Associates, Publishers: Mahwah, New Jersey.
2. Angus Reid Group (2000) http://www.ipsos-reid.com.
3. Duchowski, A.T and Vertegaal, R. (2000) Eye- based interaction in Graphical systems: theory and Practice. *Siggraph 2000 Course notes.*
4. Goldberg, H, and Kotval (1999) Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics* 24, 631-645
5. Hess, E.H. (1972). Pupillometrics. In Greenfield, N.S., and Sternbach R.A. (Eds.), *Handbook of psychophysiology* (pp. 491-5310. New York: Holt, Rinehart and Winston.
6. L C Technologies, inc. (2001) www.eyegaze.com
7. Lowenstein, O., and Loewenfeld, I.E. (1964). The sleep-waking cycle and pupillary activity. *Annals of the New York Academy of Sciences, 117*, 142-156
8. Millard, N., Hole, L and Crowle, S (1999). Smiling through: motivation at the user interface. *Proceedings of HCI International*, Munich, Germany, August 22-26
9. Narayanan, N.H., and Schrimpsher, D.J.(2000). Extending eye tracking to analyse interactions with multimedia information presentations. *Proceedings of HCI 2000 on People and Computer XIV – usability or else, p271-286*
10. Nielsen and Tahir (2001). *Homepage Usability – 50 Websites Deconstructed.* New Riders Publishing, Indianapolis
11. Partal,, T., Maria, J., and Surakka, V. (2000). Pupillary responses to emotionally provocative stimuli. *Proceedings of ETRA 2000*, Florida, USA, November 6-8
12. Picard, R (1999). Affective Computing, *The MIT Press*
13. Rötting, M.: Parametersystematik der Augen-und Blickbewegungen für arbeitswissenschaftliche Untersuchungen. Shaker Verlag.
14. Salvucci and Anderson (2000). Salvucci, D.D. and Anderson, J.R. (2000) Intelligent gaze-added interfaces. *Proceeding of CHI 2000, CHI Letters 2*, 1, Addison-Wesley, 273-280.
15. Schnipke. S.K, and Todd, M.W. (2000.) Trials and Tribulations of Using and Eye-tracking system*. Proceedings of CHI 2000extended abstracts, p273-274*
16. Shackel, B. (1975). *Applied ergonomics handbook*. IPC Science and Technology Press Ltd.
17. Takahasi, K., Nakayama, M., and Shimizu, Y. (2000). The response of eye-movement and pupil size to audio instruction while viewing a moving target.*Proceedings of ETRA 2000*, Florida, USA, November 6-8
18. Wilson, G.M. and Sasse, M.A. (2000). Do Users always know what's good for them? Utilising Physiological responses to assess media quality. *HCI 2000*, Sunderland, UK, Springer.
19. Wilson, G. & Sasse, M.A. (2000) Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods. In C. Paris, N. Ozkan, S. Howard & S. Lu (eds.) *Proceedings of OZCHI 2000: Interfacing Reality in the New Millennium, pp135-142, December 4th - 8th, Sydney, Australia.*