

Identifying Web Usability Problems from Eye-Tracking Data

Claudia Ehmke
Centre for HCI Design
City University
London EC1V 0HB
+44 20 7468 4897

claudia.ehmke@lbi.com

Stephanie Wilson
Centre for HCI Design
City University
London EC1V 0HB
+44 20 7040 8152

steph@soi.city.ac.uk

ABSTRACT

Eye-tracking research is increasingly used to supplement usability tests in both commercial and academic practice. However, while there has been research into links between eye-tracking metrics and usability problems, this has so far fallen short of establishing a general correlation scheme between the two. Consequently, practitioners are left to make subjective judgements when interpreting eye-tracking data. We address the lack of general guidance by proposing an initial correlation scheme based on data from an exploratory study which aimed to find a wide range of possible correlations between usability problems and eye-tracking patterns. User testing of two websites was conducted and a set of diverse usability problems was extracted from the data; these were then analysed and some were correlated with users' eye-tracking patterns. In addition to this initial correlation scheme, a further finding from this study is that usability problems are connected to not just a single eye-tracking pattern, but to a specific sequence of patterns. This sequence of patterns seems to arise from different coping strategies that users develop when a problem is experienced.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces – *evaluation/methodology*.

General Terms

Human Factors.

Keywords

Eye-tracking, user testing, usability problems.

1. INTRODUCTION

The continuing penetration of the Internet into everyday life has led to the expectation that the user experience of any website should be a positive one. Poor usability is not tolerated by users who simply choose to go elsewhere. As a consequence, demand for usability analyses of websites is flourishing and it is becoming more common to include eye-tracking in the range of

techniques used for this purpose.

In brief, eye-tracking is a technique whereby eye movement is recorded whilst the user is looking at a stimulus. The eyes never rest on one position for long; they move several times per second, with micro-movements sometimes spanning only a few pixels. A fixation is a moment where the eye is relatively motionless, and a saccade is a quick movement between fixations to another element. Both fixations and saccades can be determined by eye-tracking software from data collected by the eye-tracker. A gaze plot may be used to show the succession of fixations and saccades on a screen or webpage for an individual user, while heat maps show how long each part of a screen has been looked at. These (and other) visualisations of eye-tracking data are interpreted by usability practitioners to identify confusion on the part of the user, reading or scanning behaviours, or simply, but interestingly, areas that users are not looking at.

The increasing interest in eye-tracking research has been stimulated in part by the eye-mind hypotheses, a principle formulated by Just and Carpenter [cited by 15] which assumes that what a person is looking at indicates what they are currently thinking about or attending to. Other studies have also shown a connection between eye-tracking patterns and users' decision making processes [8]. However, while eye-tracking is increasingly used during usability tests, as yet no correlation scheme has been established to link eye-tracking patterns to specific usability problems. Consequently, the analysis of patterns is mostly based on the opinion and interpretation of the individual evaluator. If eye-tracking is to become a serious tool in usability testing, we must move beyond the anecdotal and subjective to put the interpretation of eye-tracking data and its implications for usability on a more rigorous footing. The aim of research reported here is to establish an initial framework correlating eye-tracking patterns and usability problems. The constant movement of the eyes makes it difficult to analyse eye movement while a user is interacting with a system, hence we focus on post-test analyses.

The term eye-tracking pattern is used in this study as an umbrella term for visualisations of specific eye-tracking metrics. This means it includes single metrics such as a long fixation or a long saccade to another element, as well as combinations of metrics such as a specific scan path (a sequence of fixations and saccades). For further explanation of the single metrics, refer to Poole and Ball [15].

In this paper, we first summarise previous research in which eye-tracking metrics have been used to gain insight into the usability of an interface. The findings from these studies are organised as an initial framework correlating eye-tracking patterns and usability problems. Second, we describe an

exploratory empirical study undertaken to investigate further the relation between eye-tracking patterns and usability problems and use its results to propose a correlation scheme.

2. BACKGROUND

2.1 Eye-Tracking in Usability Studies

Usability companies are increasingly offering eye-tracking services. For example, Etre [6] enhance their website evaluation reports with 'session images and heat maps'. A usability practitioner experienced in eye-tracking stated that she looks at the following eye-tracking data [Armitage 2006, personal communication], and suggested some possible interpretations:

1. Long fixations. (Interest or confusion)
2. Back-track saccade. (Possibly confusion)
3. Not looking at elements of a page.
4. Scanning behaviour rather than reading behaviour, that is, fixations and saccades not in left to right order with sweeps. (What was the user looking for?)
5. Back and forth between two objects. (Trying to make a choice or comparison? Is it distracting?)
6. First place the user looks. (Why did this draw their attention?)
7. Last place the user looks. (Why did this lose their interest?)
8. When making a choice, fixations back to one item, then final scan before making choice.
9. Reading headings or subheadings, but no more. (Boring?)
10. Interaction, e.g., following an asterisk to a footnote, or a reference in the text to an image or other element.

Findings such as these are noted by the evaluator during a usability test and discussed with users afterwards. An extension of this approach is to ask users to provide retrospective protocols cued by a replay of their eye-tracking data to make it easier for them to explain their decisions and thoughts. This method, called PEEP (Post-Experience Eye-Tracked Protocol), is described further by Ball et al. [1]. They argue that concurrent think-aloud protocols can be incomplete and cause difficulties to the users as they have to verbalise ongoing cognitive processes that may be subconscious.

2.2 Related Research

While eye-tracking might be a relatively new technique for usability practitioners, made feasible by significant improvements in the technology, it has been used in other disciplines for some time. Cognitive psychology provides a rich background of research in this area because eye-tracking has the potential to offer insights into problem solving, reasoning, mental imagery and search strategies. Early studies using eye-tracking began before computers as we now know them were introduced. For example, Fitts, Jones, and Milton [cited in 9] tested the eye gaze of pilots back in 1950 and proposed fixation frequency and duration as important metrics.

Goldberg and Kotval's [7] research offers one of the most influential frameworks when investigating correlations between eye-tracking metrics and usability problems. They proposed and evaluated a number of spatial eye-tracking metrics that are relevant to visual search (see Table 1 for a summary of these and other metrics reported in the literature). The metrics were tested in a study where a drawing tool selection program was configured with different interfaces. In one experimental

condition, the interface had randomly grouped tools, and in a second condition the tools were grouped by functionality. These set-ups were then evaluated by interface designers and typical users with the expectation that eye-tracking metrics would differ between the two interfaces. The poorly configured interface was expected to show 'more extensive search behaviour'. Some eye-tracking patterns showed significant differences between the interfaces. Amongst others, Goldberg and Kotval used the following metrics in making the comparison: 'Number of Fixations' (when searching for a single target, a large number of fixations indicates the user sampled many other objects prior to selecting the target); 'Fixation Duration' (longer fixations imply spending more time interpreting or relating the component representations to mental models and therefore can indicate less meaningful elements) and 'Fixation/Saccade Ratio' (higher ratios indicate that there was either more processing or less search activity). They also suggested some other metrics that might help when using eye-tracking patterns to investigate the usability of interfaces; these are listed in Table 1. Cowen [3] has criticised Goldberg and Kotval's study, first, because with a simple task, like choosing a tool from the selector, it is difficult to provide feedback on real tasks with more complex (usability) problems and, second, because in the comparison they used usability ratings obtained from the participants (interface designers and users) rather than actual performance measures.

Cowen et al. [4] analysed eye movements in the evaluation of website usability. They compared performance measures of four websites against different eye movement metrics to see whether they would indicate similar overall usability of the pages. To rate the overall usability, response scores (selection of correct link) and task completion times were used as performance measures. These were then compared against four eye-tracking metrics: Total Fixation Duration, Number of Fixations, Average Fixation Duration (all adapted from Goldberg and Kotval's study) and, additionally, Fixation Spatial Density. The latter was meant to provide a 'global measure of the total amount of processing performed on each page' [4]. Only the time-based eye-tracking metrics (Total Fixation Duration, Average Fixation Duration) showed the same significant difference as shown in the performance measures. The other eye-movement metrics, although not providing statistically significant results, showed patterns similar to the performance measures. Cowen et al. conclude that these results can only be taken as vague evidence that greater spatial densities are provoked by inefficient searching.

Goldberg et al. [8] studied different eye-tracking metrics during tests of a prototype for a web portal. As well as existing metrics such as Number of Fixations and Mean Fixation Duration, they also employed two more eye-tracking metrics: Saccade Amplitude and Scanpath Length. Larger saccades can indicate more meaningful cues, as attention is drawn from a distance and a longer scanpath can indicate less efficient searching (perhaps due to a sub-optimal layout). These metrics had been mentioned by Goldberg and Kotval [7], but with this study were shown to be clearly related to usability problems.

McCarthy et al. [11] explored another measure, 'Glance Frequency', defined as 'one or more successive fixations to the same screen object'. They tested specific design conventions such as the positioning of a menu by setting up usability tests to compare performance measures against eye-tracking results. However, Glance Frequency is used as frequency measure rather than as a link to a specific usability problem, and is therefore not included in Table 1. Renshaw et al. [17] explored the influence of visual design on eye movement. They

conducted a study to investigate an eye-tracking measure called 'Gaze Orientation'. This 'was developed to categorise gazes as being either vertical or horizontal' [17] and was derived from

the coordinates of final fixations of a gaze. They concluded that Gaze Orientation can help analyse the influence of alignment, proximity and other design features on eye movements.

Table 1. Summary of eye-movement metrics and related usability problems, reported in the literature.

Eye-movement metrics	Cognitive process or usability problem	Reference
Fixation-related		
Time to first fixation on target	Good (if short) or bad (if long) attention getting properties	Byrne et al., 1999 (cited by Poole and Ball, 2005 [15])
Fixation spatial density	Focussed efficient searching OR widespread inefficient search	Cowen, Ball, and Delin, 2002 [4]
Fixation duration, Fixation length	Difficulty in extracting information OR more engaging; voluntary (>320 ms) and involuntary (<240 ms) fixations; needs further investigation	Just and Carpenter, 1976 (cited by Poole and Ball, 2005 [15]); Graf and Kruger, 1989 (cited by Jacob and Karn 2003 [9])
Fixations on target divided by total number of fixations	Low search efficiency	Goldberg and Kotval, 1999 [7]
Number of fixations overall	Less efficient search due to sub optimal layout	Goldberg and Kotval, 1999 [7]
Repeat fixations (post-target fixation)	Lack of meaningfulness or visibility	Goldberg and Kotval, 1999 [7]
Fixations per area of interest	Element/area more noticeable OR element/area more important	Jacob and Karn, 2003 [9]; and Poole, Ball, and Phillips, 2004 [16]
Percentage of participants fixating on area of interest	Attention-getting properties of an interface element	Albert, 2002 (cited by Jacob and Karn, 2003 [9] and Poole and Ball, 2005 [15])
Fixations per area of interest adjusted for text length	Element harder to recognise	Poole, Ball, and Phillips, 2004 [16]
Saccade/fixation ratio	More processing or less searching	Goldberg and Kotval, 1999 [7]
Saccade-related		
Number of saccades	More searching if more saccades	Goldberg and Kotval, 1999 [7]
Saccades revealing marked directional shifts	User's goals changed OR interface layout does not match user's expectations	Cowen, 2005 [3]
Saccade amplitude	Meaningful visual clues if larger saccades	Goldberg, Stimson, Lewenstein, Scott, and Wichansky, 2002 [8]
Regressive saccades (backtracks/regressions)	No meaningful visual clues, changes in goals, mismatch between users' expectation and the observed interface layout	Sibert and Jacob, 2000 [18]; Poole and Ball, 2005 [15]; Goldberg and Kotval, 1999 [7]
Saccade duration	Low image quality such as blurred or low contrast	Vuori, Olkkonen, Pölönen, Siren, and Häkkinen, 2004 [20]
Scanpath-related		
Longer scanpath duration	Less efficient scanning	Goldberg and Kotval, 1999 [7]
Scanpath direction	Indication of search strategy	Altonen et al. (1998, cited by Poole and Ball, 2005 [15])
Longer scanpath length	Less efficient searching	Goldberg, Stimson, Lewenstein, Scott, and Wichansky, 2002 [8]
Small spatial density of scanpath	More direct search	Goldberg and Kotval, 1999 [7]
Scanpath regularity	Search problems due to lack of training or interface layout problems	Goldberg and Kotval, 1999 [7]
Transition matrix (back and forth between areas)	Uncertainty in search OR search order efficient and direct	Goldberg and Kotval, 1999 [7]
Transition probability between AOIs	Efficiency of arrangements of elements in user interface	Fitts, Jones, and Milton, 1950 and Hendrickson, 1989 (cited by Jacob and Karn, 2003 [9])
Gaze-related		
Gaze (dwell)	Measure of anticipation OR attention distribution between targets	Mello-Thomas et al., 2004; Hauland, 2003 (cited by Renshaw, Finlay, Ward, and Tyfa, 2003 [17])
Gaze orientation	Feedback about success of design features	Renshaw, Finlay, Ward, & Tyfa 2003 [17]
Gaze duration on AOI	Difficulty extracting or interpreting information from element	Several studies cited by Jacob and Karn 2003 [9]
Number of gaze per AOI	Possible importance of element	Several studies cited by Jacob and Karn 2003 [9]
Spatial coverage calculated with convex hull area	Scanning in a localised or larger area	Goldberg and Kotval, 1999 [7]

Jacob and Karn [9] give a useful overview of eye-tracking and eye-tracking metrics in HCI research. They list previous studies with details such as number of participants and tasks, and also which eye-tracking metrics were used. They analysed 24 studies that were conducted between 1950 and 2002 to show the six most commonly used metrics: overall number of fixations (used in 11 studies), gaze on each area of interest (7 studies), overall fixation duration mean (6 studies), number of fixations on each area of interest (6 studies), gaze duration mean on each area of interest (5 studies) and the overall fixation rate (5 studies). They also suggest that these metrics are sometimes not the most suitable and recommend others, including Scan Path and derived metrics such as the Transition Probability between AOIs (Area of Interest), to indicate the efficiency of the arrangement of elements in the user interface.

Finally, Poole and Ball [15] provide an extensive overview of eye-tracking metrics that have been used in either HCI research or cognitive psychology. They summarise metrics from previous studies related to fixations, saccades and scanpaths, but do not provide a critical analysis of the different metrics.

2.3 Summary of Relations between Eye-Tracking Patterns and Usability Problems

Several studies have compared the results of traditional usability evaluation against different eye-tracking metrics. Amongst these, Goldberg and Kotval's study [7], which compared the greatest number of metrics against standard usability evaluation outcomes, was highly influential. More recent studies have added further eye-tracking metrics. Table 1 provides an overview of these metrics reported in the literature in relation to the usability problem or cognitive process that they might indicate (either validated or hypothesized).

However, there remains a need for studies that relate eye-tracking patterns to specific usability problems (by indicating cognitive processes). Previous research has compared eye-tracking metrics to general interface usability, either as assessed by usability experts or as calculated from performance measures such as task completion times. However, expert reviews and performance measures can only give an overview; they fail to establish direct links between metrics and usability problems. In particular, expert reviews have no possible direct relation to users' eye-tracking patterns and consequently have not been used for the identification of usability problems.

In the research reported here, we explore correlations between the appearance of a specific eye-tracking pattern and a usability problem. This means that specific usability problems from different contexts (where a context is defined by the user, task and system) are investigated for their related eye-tracking patterns.

3. METHOD

We conducted an empirical study to collect data from which to analyse possible correlations between eye-tracking patterns and usability problems. The exploratory nature of the study meant that it was important to identify a diverse set of usability issues (rather than to collect a lot of data about just a few problems) from a variety of data sources.

In summary, participants undertook tasks with either one or two websites. We collected verbal protocols (either concurrent think-aloud and/or retrospective protocols) and observational data to identify usability problems, as well as eye-tracking data from which to derive eye movement patterns; correlations between the two were then analysed.

3.1 Tasks

As stated above, it was important to obtain a rich and diverse set of usability problems, not to undertake a thorough evaluation of any one system or of a specific usability problem. Hence, two different websites were selected for use in the study: bbc.co.uk and thetrainline.com, with one specific task for each of them. The task on the BBC website was to find the surfing conditions in Wales (therefore focussing on using navigational elements and the scanning of pictures and text elements) and the task on thetrainline.com was to find a specific train journey (therefore focussing on text input and scanning of information). Consequently, these website-task combinations provided coverage of a range of different kinds of interactions, including scanning text, reading, text input, usage of navigation elements, scanning of pictures, searching etc., that were expected to yield a broad range of usability problems.

3.2 Participants

19 participants (9 female and 10 male) were recruited to participate in the study. No specific recruitment criteria were applied. Several participants had glasses or contact lenses but this was not an impediment to tracking their eye movements. Nine had background knowledge of human computer interaction; however none had previously participated in an eye-tracking study. Their ages ranged from 22 to 55, although only two participants were above 36. English was the native language for less than 50%, however all were working and living in the UK and able to communicate effectively in English. All participants reported that they worked with the Internet on a daily basis. They had all used the BBC website before. Only two had used thetrainline.com previously, though all but one had used similar ticket-booking websites before.

3.3 Apparatus

The study was performed with a Tobii x50 [19] eye-tracker, a free standing, non-invasive device which can be set up in front of any interface. The x50 tracks both eyes with a data rate of 50Hz, returning a tracking status every 20ms. The tracker works with two infrared light sources, the reflection of which from the retina is recorded by a camera. Consequently, the participants could move freely in the limited area that the tracking system can record accurately. Two 21 inch flat screen monitors were attached to the computer: one showing the websites to the participant, the other one showing the eye movements simultaneously to the facilitator. A webcam, two keyboards and two mice were attached to give control to both participant and facilitator, without needing to move devices.

The system was running the eye-tracking software ClearView 2.5.0b on a PC desktop with Windows XP operating system. To ensure correct recording of eye-tracking patterns of different tasks (text reading, scanning of pages etc.), the minimum fixation duration was set to 30ms and the gaze point to deviate not more than 30 pixels during a fixation (as fixations would be shorter for reading than for scanning tasks).

3.4 Testing Procedure

Each session began with the participant giving informed consent. A short questionnaire was then administered to collect demographic data and the session was introduced. The eye-tracking equipment was calibrated for the participant, the tasks were explained and the participant was asked to undertake the tasks.

Depending on time constraints with different participants, some undertook both tasks, some only one task. Nine participants did

both tasks, 7 participants undertook the task with only the BBC website and 3 participants with only thetrainline.com. Eye-tracking and observational data of their interactions were recorded for all participants using Tobii ClearView. Of the 16 participants who used the BBC website, 4 gave a concurrent protocol, 1 gave a retrospective protocol and 11 were just observed and eye-tracked. Of the 13 participants who used thetrainline.com, 1 gave a concurrent protocol, 7 gave a retrospective protocol and 5 were just observed and eye-tracked.

This mix of protocols was used to obviate the effect that single protocols would have on the study outcomes. Concurrent protocols ('think-aloud' reports) are commonly used in user testing, however as they were likely to affect the eye-tracking patterns by resulting in more fixations per element and page and longer fixations, only a limited number of participants were asked to do this. Participants giving think-aloud protocols during the sessions were asked to provide details about what they were doing and why, what they thought of the design, if they were looking for something, if they were unsure about actions or if they found elements unclear.

Retrospective protocols followed the approach described by Ball et al. [1]: participants were shown their eye-gaze recording and were asked to describe why they were doing things or what had attracted their attention. The following details were specifically queried: long fixations or big saccades; not looking at specific elements that were considered important for the task; text scanning rather than reading, and backtracking saccades. The retrospective protocols were recorded using Camtasia, which records a video of the screen, the mouse position, keyboard input, the webcam and the audio channel.

Some participants did not give any protocol to increase the number of eye-tracking patterns that could be matched to provide further insight.

The participants spent on average 20 minutes performing the tasks. For all participants, ClearView recorded the screen, including the participant's eye movements, as well as the webcam, audio from the microphone, mouse and keyboard input. Recording the eye movement included the following every 20ms [19]: the positions of both eyes' gaze on screen in x and y coordinates; the validity of the gaze data (error, i.e., a reflection, a closed eye etc. causing tracking problems); events such as a URL becoming or ceasing to be visible.

3.5 Extraction of Usability Problems

The first stage in the data analysis was to identify the usability problems experienced by participants as follows:

1. Usability problems were extracted from the raw data (the observations, concurrent protocols and retrospective protocols)
2. The problems were matched across participants to determine where more than one participant had experienced the problem
3. The problems were matched across protocols to determine the unique problems for each task/site.

Usability problems were identified from the raw data using a set of 'problem extraction criteria'. These are statements that, if found true during the analysis, are considered to be evidence of a usability problem. This helped to ensure a standard way of identifying usability problems across the individual participants. The criteria included: participant shows surprise or frustration; participant fails at the task or leaves the site;

participant uses the browser's back button, etc. The outcome was a set of tables of usability problems for each task/site that contained separate data for each participant and each source (concurrent think-aloud protocol, retrospective protocol, observation). Each usability problem was described in terms of:

- ◇ The number of the specific participant (e.g., A2)
- ◇ A unique number for this usability problem (e.g., 1)
- ◇ A description of the usability problem (e.g., 'Participant mentions that the holiday page does not contain any useful data for her task "finding information about surfing in South Wales"')
- ◇ Location (URL) where the problem was discovered (e.g., www.bbc.co.uk/holiday/)

There were 74 instances of usability problems on the bbc.co.uk and 75 instances of usability problems on the trainline.com. To determine where there were multiple instances of the same usability problem, matching criteria were established:

- ◇ **Page problem:** The problem explained or observed is experienced with a specific element on the same page and results in the same specific behaviour of the participant.
- ◇ **Site problem:** The problem explained or observed is experienced with a specific element on different pages of the site but results in the same pattern of behaviour.

This led to the identification of 48 unique problems on the BBC site and 44 on thetrainline.com. The breakdown of these is given in Table 2, where 'initially extracted usability problems' refers to the problem sets where duplicates are still included.

Table 2. Numbers of usability problems per task/site.

Unique problems per site, aggregated across participants and protocols		
Total	BBC: 48	Trainline.com: 41
Initially extracted usability problems, including duplicates		
Think-aloud	BBC: 31	Trainline.com: 11
Observation	BBC: 38	Trainline.com: 20
Retrospective	BBC: 5	Trainline.com: 44

3.6 Correlation with Eye-Tracking Patterns

Having obtained a set of usability problems, two distinct approaches were adopted for correlating these with eye-tracking patterns. In both cases, the correlation was controlled by either of the two correlation criteria:

- ◇ **Page problem:** The usability problem and the eye-tracking pattern are connected to the same element on a page.
- ◇ **Site problem:** The usability problem and the eye-tracking pattern are connected to the same element that is present on different pages of the site.

An element in this context is referred to as an item on the page that users interact with using an interaction device or their eyes, such as a text block, a picture, or an interaction or navigation element.

In the first approach to establishing correlations, eye-tracking patterns were identified, aggregated across participants, and then linked to the set of usability problems. Two identification criteria were set up to control the extraction of eye-tracking patterns: specific behaviour for an element on a page and specific behaviour on a page. The eye-tracking patterns were

entered with textual descriptions into a spreadsheet [5]. After aggregating them across participants, 23 eye-tracking patterns were found for both tasks. The patterns were examined in correlation with the usability problems (using the correlation criteria mentioned above) whereby 12 of them correlated directly to a usability problem and 4 correlated with a few usability problems, caused by broad pattern descriptions that were not sufficiently specific for one usability problem. The other 11 eye-tracking patterns had no correlating usability problems. Given the limited success of this method, a second correlation strategy was used.

As a second possible way forward, the 48 plus 41 unique usability problems per task were examined and the eye-tracking patterns of participants that experienced each problem were recorded in a spreadsheet (for details see [5]). As an example, the usability problem ‘expected information missing (on page, area or subsection)’ was linked with two pattern descriptions from different users: ‘The participant scans across the whole page without finding details he is looking for’ and ‘There are a lot of fixations on various items on the page without any particularly long fixations’ (see Figure 1).

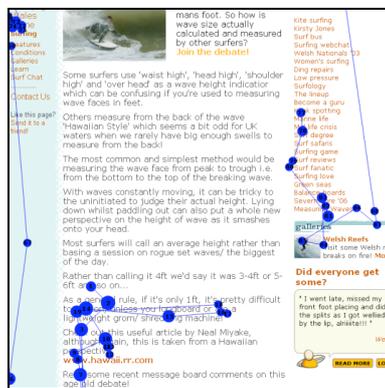


Figure 1. Typical eye-tracking pattern for usability problem - ‘Missing information on page’.

A table of correlations (between patterns and problems) was produced (see Table 3) using the correlation criteria outlined above. Each of the correlations was described by a generalisation of the usability problem, an example usability problem, an eye-tracking pattern generalisation, and the number of participants experiencing the usability problem.

A few steps were undertaken in the generation of the table:

- ◇ Creating a generalisation of the usability problem.
- ◇ Creating a short eye-tracking pattern generalisation: This was important to provide a quick overview of the pattern.
- ◇ Aggregating eye-tracking patterns across participants: After recording textual descriptions of eye-tracking pattern sequences for all participants, the patterns were aggregated in a similar manner to the usability problems beforehand - using aggregation criteria to reduce their numbers. This meant that sometimes several different descriptions had to be combined. For example, the pattern mentioned above was combined with: ‘The participant scans around the different options in the middle of the page, followed by the right hand link list, back to the middle where she finds “Surf reports” which she selects’, forming the general description: ‘Participants are going through different options in the middle and right hand side of the page,

before selecting a link’. In cases where the eye-tracking pattern descriptions varied considerably from each other, both descriptions were maintained next to each other in the table.

The correlations between patterns and problems were reduced to 53 from a total of 89 for both tasks, by using the aforementioned aggregation criteria. Of these, 17 are included in Table 3. The remainder were excluded for the following reasons:

- ◇ Insufficient data: 25 usability problems were experienced only once in the tests and are not included for this reason.
- ◇ Usability problems that were very specific to the task, such as ‘lack of trust in functionality of the search box on BBC’, or problems that were related to the participant’s interaction method, e.g., the user looking at the keys when keyboard input was required (no eye-tracking pattern was recorded).
- ◇ Eye-tracking patterns that were clearly related to the protocol the user provided, e.g., a lot more short fixations for participants who were giving a concurrent think-aloud protocol.
- ◇ No eye-tracking data: some usability problems such as ‘Default font size too small’, ‘expected option missing’ (that were verbalised by the participant) or “scripting error” cannot be recorded by eye-tracking and are not included.
- ◇ Some usability problems that only became apparent for the user on the next page, such as a misinterpretation of a link name – where the user goes directly to a link, but only realises once on the next page that the link name was misleading.

The eye-tracking patterns that were identified and recorded in Table 3 turned out to be combinations of the following metrics:

- ◇ Fixation related: Short fixation (shorter than 100ms); long fixation (longer than 800ms); high number of fixations (more than 5 fixations on an element); missing fixation (no fixation, usually in combination with ‘on a specific element or area’); fixation on a specific element or area; specific order of fixations; reading pattern (short fixations following the text flow); scanning pattern (fixations across the page, order is not following the text flow).
- ◇ Saccade related: Layout or content-related patterns (specifically wide saccades indicate jumps to specific elements that attract attention); regressions or repeated fixations (or backtracking saccades to one element); high number of saccades (more than 5 with only short fixations in between); horizontal or vertical saccades (usually following the layout of specific elements); long saccades (across a whole page vertically or horizontally, these are usually layout oriented or can indicate jumps to specific browser elements such as the status bar);
- ◇ Change over time: Where the pattern consists of one type of metric and than changes to another type.

4. DISCUSSION

This study should be seen as exploratory research rather than definitive prescription as it looked to provide an initial top-level view across a lot of usability problems and their relations to eye-tracking patterns. This work also takes a new perspective on research into eye-tracking patterns by relating sequences of patterns to usability problems.

Table 3. Correlation of usability problems to eye-tracking patterns (Freq = number of occurrences of usability problems).

Freq.	Usability problem generalisation	Usability problem	Eye-tracking pattern generalisation
12	Expected information missing (on page, area or subsection)	A user goes to a page on the site, expecting to find specific details which are not provided.	Many short fixations across page where information is expected.
6	No error prevention through incorrect pre-filling (unclear functionality)	Drop downs are provided and are required to be used without being clearly marked as required so that users receive error messages.	Short fixation on specific element followed by some regressions without long fixations on element.
5	Ineffective presentation	The search box draws the attention of visitors more than the actual link to the item that was looked for.	Very few fixations in general. No fixations on quickest link element. Some interface objects draw more attention.
5	Misleading element, unclear target of link	Users have problems distinguishing between interaction elements with slightly different behaviours (internal vs. external links).	High number of short fixations across page (scanning and reading), no long fixations.
5	Overloaded, ineffective presentation	Most participants find the page too crowded and not easy to read. They cannot easily grasp the concepts or how things are interlinked on the page or how they have to interact with it.	A lot of short fixations on single areas followed by longer saccades and regressions to elements.
5	Hard to see error message	An error message that comes up on the screen after an incorrect input is not reacted to.	There are no fixations on element but long fixations on other elements.
4	Unclear input format	Input fields are not clearly labelled to support the user by advising which format data should be entered in. Users react with confusion.	A lot of short fixations on page but not specifically on this element. Once coming back to the page after having seen the error message, there are less fixations across the page and more on the problematic field.
4	Specific information or links not provided	A user is coming to a page that does not contain options relevant for his task, makes an unhappy comment and decides to go to the Top Search Box to go further.	There are many fixations across the page without very long fixations and no fixations on some important links.
3	Missing functionality	A functionality is expected (either because it would be useful or because it is provided on similar pages) but is not provided on the page. Consequently, participants reported they were looking for it on the page.	High number of fixations in task specific area, followed by less spatially dense fixations.
3	Unclear interaction mode	Text elements appear as clickable but are not interaction elements.	Long fixations on misleading element.
3	Too much information	A content page contains too much information in a big block of text and readers drop out after the first line.	Depending on page content: reading and then scanning pattern or short fixations on limited parts of the page.
3	Functionality unclear and missing output or error message	Specific functionality is expected as buttons are also provided for it. However, the buttons work in a limited range of conditions. Participants got confused as the conditions for the buttons to work are not displayed.	The fixations are focussed on two decision points, combined with short fixations across site.
2	Unclear grouping - not matching mental model of user	Geographical grouping on page was not expected by users who therefore tried finding information elsewhere.	Many fixations overall (different areas on the page), no fixations on the menu area where unclear grouping is presented.
2	Design of element unclear (layout and colour coding)	The design of a specific element and layout of the page are confusing the user in finding the required information.	No fixations on problematic element in the beginning. Re-visits on page shows longer fixations on it.
2	Non-obvious interaction because of design issues	Interaction elements are displayed as buttons and text links, mixed on one page and cause problems for the users who don't seem to be interacting with the text links or are delayed in their interaction.	No fixations on element, a lot of fixations on area and page.
2	Terminology unclear	The wording or layout seems to be unclear as the requested information is displayed but the participant still goes further to another page.	Fixation on target but further fixations across other elements on page and further scanning with regressions back to unclear element.
2	Unclear item grouping (ineffective presentation)	The presentation form and grouping of options does not match the mental model of the user so they react with confusion or cannot find the information although it is provided.	Distinct fixation order on elements (with specific saccades following the layout down to the bottom of the page). Followed by fixations on specific element.

Earlier studies have suggested that one eye-tracking pattern is connected to one cognitive process, for example, a long fixation on an element shows that the user has difficulty in extracting information from it or that the element is more engaging (Just and Carpenter, cited by [15]). However, looking at the complexity of human behaviour, and especially the coping strategies that users adopt once a problem is encountered, would suggest that users' behaviour, and with that their eye-tracking patterns, can be linked back to usability problems. Consequently, our focus is on the analysis of eye-tracking pattern sequences rather than just single patterns.

4.1 Comparison with Other Studies

No previous study has attempted to create a general correlation scheme that links usability problems to eye-tracking patterns. We suggest that there are two main types of study in this area at present:

◇ **Design specific studies**, such as 'Eyetrack III' by the Poynter Institute [14], which address issues such as 'the typical scanning pattern on a homepage' and 'how are headlines or text blocks read by users'. With the help of eye-tracking, these studies investigate where people are looking and which design changes might affect their behaviour. These studies provide insight into specific design elements that might cause usability problems, however they do not address the relation between specific usability problems and the eye-tracking patterns.

◇ **Academic studies** which investigate specific eye-tracking metrics and their relation to usability problems [4, 17, 20]. In general, these use performance measures to determine the usability of the stimuli (website, graph or illustration) and then compare these against eye-tracking data. The data are usually statistically evaluated to see if any significant correlations can be found.

Although both types of study analyse the relation of usability problems to eye-tracking patterns, neither of them address the challenge of creating a more comprehensive correlation scheme. Furthermore, from a practical point of view, the academic studies may not be relevant to the specific needs of usability practitioners who are running standard usability evaluations of websites with the enhancement of eye-tracking technology in order to gain insight into usability shortcomings. Only a few of the academic studies have used a similar setup. One example is the study by Bojko [2], which compared a website to a redesign proposal using eye-tracking with real users and authentic tasks. Other studies, such as that by Goldberg and Kotval [7], were run with an adjustable testing interface where a tool had to be discovered in differently configured interfaces. On the one hand, this approach means that not only can other factors be controlled, but also that the results can be seen more clearly as the difference between poor and good layout is more distinct. However, real tasks or websites would not have such simplistic usability problems.

Additionally, there are some studies that bridge between these two categories, such as those reported by McCarthy et al. [11, 12]. The first of these is a study to verify existing design conventions for menu placement and how it can affect the usability of a site using performance measures versus eye-tracking metrics; the second investigates the connection between expectations of users regarding the positioning of specific elements on a site. These studies provide a different perspective as they help practitioners by evaluating existing design norms with real websites, and consequently give specific feedback that can be incorporated into practice.

The intended beneficiaries of the work reported in this paper are usability practitioners, and therefore our study focussed on real tasks (and websites) so that more complex usability problems could be evaluated. The setup made it possible to see combinations of patterns, for example, if information was expected by the participant but not provided, the eyes would first fixate on the most likely places (jumping with bigger saccades), followed by short fixations all over the page, in order to look for the item. Analysing a sequence of patterns therefore relates more directly to the coping strategy of the user when a problem is experienced, although this could also make it more difficult as coping behaviours could differ from user to user depending on their experience and background. The relation of pattern combination to usability problems has not been analysed by previous studies. This pattern of coping strategies seems to be related to the 'Geometry of web search' [12] and the order of typical user fixations on a page.

Furthermore, from a very practical point of view, although previous research has yielded a long list of possible eye-tracking metrics (Table 1), it became clear during this study that not all of these can easily be used by any practitioner with 'standard' eye-tracking software. For example, while the ClearView application offers a mechanism to export data to Excel in order to analyse 'Areas of Interest' (AOI) and transitions between them, the functionality in the version we used was not sufficiently reliable (see [5] for more details). Consequently, the list of metrics that make up the eye-tracking patterns identified here focuses on combinations of fixations and saccades.

4.2 Differences Between Participants

During the analysis of the eye-tracking data, it became clear that certain differences between the participants were relevant, notably:

◇ **Language**: Participants varied in their fluency in English. Some of those who were less fluent showed many more fixations on the page, which could have been related to them looking around more to find specific terms. Although questions tried to be addressed retrospectively, the eye-tracking patterns were already different. Consequently, when looking back at the table of correlations, it is important to recognise that usability problems which were extracted from only the patterns of these participants might have been affected by other factors.

◇ **Browsing behaviour**: The results for specific eye-tracking patterns clearly showed that different users follow different browsing schemes. Some find the first link, select it and see whether it brings them the information they were looking for, whereas others scan every page in more detail to find a link that best suits their needs. Similar behaviour has been reported by Nielsen [13] who separates users into 'search-dominant' users and 'link-dominant' users. Krug [10] analyses this as some people will 'look for a search box as soon as they enter a site, whilst others 'will almost always browse first'. Krug also identifies a third category of users where it 'depends on their current frame of mind' whether they start browsing or searching first. Such specific behaviour will again influence the findings and should be noted clearly.

◇ **Internet experience**: Although all participants stated in the questionnaire that they worked daily with the Internet, there was a clear gap in how websites (even known ones) were used by different participants. It might have been better to ask participants to rate their Internet experience on a scale from 1 to 10. This would have asked them to express their knowledge

rather than behaviour and might have given a better understanding of their background.

4.3 Limitations of the Method

This was an exploratory study and the method we adopted was to some extent also exploratory. It is useful to recognise some limitations of the method and the impact they may have had on the results:

◇ The analysis was conducted by one evaluator. Although using only one evaluator might have made results more consistent, it also means that some usability problems may have been overlooked or their interpretation may have been one-sided. Furthermore, the decision as to when an eye-tracking pattern sequence for one element started or ended was made by one evaluator, making it less generic and standardised than would be the case with more evaluators.

◇ Eye-tracking patterns make thought processes clearer (Just and Carpenter, cited by Poole and Ball [15]) but not necessarily the specific usability problem. Accordingly, relying on observation alone might have provided additional eye-tracking patterns, however, without making the specific usability problem explicit. This is especially important when comparing sequences of eye-tracking pattern across participants from different backgrounds: a problem encountered by one user does not necessarily affect another user or even create the same behaviour.

◇ Patterns were identified on an element or page basis and described with a textual description. Therefore, an analysis of the metrics identified by previous research was not possible. This identification method was mostly limited by the software and its possibilities and by time. Given the limitations, it seemed to be an efficient way of comparing patterns across different participants. However, a question for the future is whether there may be more effective approaches to analysing or reproducing correlations.

4.4 Problems with the Correlation Scheme

Apart from differences between the participants highlighted above, other issues with the correlation scheme became apparent:

◇ Handling of different patterns for different participants: how good were the matching criteria? For example, one problem ('Overloaded ineffective presentation') has different patterns across different participants. The pattern used in the correlation table ('A lot of short fixations on single areas followed by longer saccades and regressions to elements') was chosen as it was visible for all but one of the participants. However, this means that sometimes a pattern could have been typical of that one user but not incorporated into the table. It might be better to use several patterns for one usability problem, rather than one standard pattern.

◇ Limitations of listed metrics for combination: not sufficiently clear. Several patterns describe a change over time in the pattern. Although change over time was one of the factors to describe a pattern, it is limited in that a very specific order of patterns cannot be described.

◇ Strength of indication for usability problem: the broad identification of eye-tracking patterns affected the actual analysis of the patterns and time constraints did not allow checking the reliability of a pattern in relation to the existence of a usability problem.

◇ Some patterns relied on the participant voicing their thoughts during the study, for example, "[...] fixations across page where information is expected", or on establishing an understanding of important interface elements upfront as the correlations require specific attention to certain elements.

5. CONCLUSION

Previous research has established metrics that relate a single eye-tracking measure, such as 'fixations per area of interest' to different possible usability problems, in this case 'element is more noticeable' or 'element is more important'. However, with the complexity of human behaviour and the coping strategies that users exhibit when they encounter a problem, combinations of patterns which reflect the structure of user behaviour need to be studied in more depth.

This research has explored how user behaviours and their related eye-tracking patterns are linked to specific usability problems. The sequences of eye-tracking patterns consist of a series of different metrics, such as 'a high number of fixations across the page and navigation, followed by fixations on one element only' correlating with the usability problem 'ineffective presentation through unclear item grouping'.

Some serious technical problems were experienced with the setup of the system and the eye-tracking software. Although this had an effect on some aspects of the study, a comprehensive analysis of the data was conducted and a new approach for investigating the relations between eye-tracking patterns and usability problems has been suggested [5]. When analysing the study results, the complexity of the research required became clear. Several of these complexities have been discussed here and we hope this will benefit others who undertake similar research in the future.

5.1 Correlation Schemes Created by Previous Research

A review of previous research revealed six studies that were most influential in their investigations of eye-tracking metrics. First of all, Goldberg and Kotval [7] provided an important insight into 15 eye-tracking metrics. Other work, by Cowen et al. [4], Goldberg et al. [8], Renshaw et al. [17], Vuori et al. [20] and Poole et al. [16] gave further insight into different metrics. Furthermore, two papers, by Poole and Ball [15] and Jacob and Karn [9], offered a useful overview of previous studies and current eye-tracking metrics, providing a basis for this study.

A table of 28 eye-tracking metrics was compiled from this earlier work, linking the metrics to the cognitive processes or usability problems to which they are related. With its coverage of recent studies, in combination with the previous overview studies, this provides an updated list of eye-tracking metrics used in research at the moment and consequently follows the direction of the other overview studies in pointing out possibilities for eye-tracking metrics to be used in future research.

5.2 Future Research

This work represents an initial step towards developing a general scheme for correlating eye-tracking data with usability problems. An ultimate goal could be to use such a scheme as the basis for automated analysis of eye-tracking data in the context of usability evaluation. However, further research is required before this can be realised.

Firstly, 'pattern-problem' correlations proposed here need to be validated using focussed experimental studies. This means that

a limited number of usability problems should be analysed in greater depth to find all their related eye-tracking patterns. If the usability problems are to be identified in the same study as the eye-tracking data is collected, we would recommend using the PEEP method by Ball et al. [1]; this will ensure that the analysis focuses on what each participant actually perceives as a problem as opposed to stand-alone eye-tracking data. Secondly, further research should also look for additional usability problems that have to be correlated to their specific eye-tracking patterns. This implies undertaking studies of participants doing different tasks on a range of systems. Related to this, there should be a more detailed investigation of whether specific patterns can also be found if the usability problem does not exist, for example, if pattern x (described by 'y number of fixations longer than 600ms on a specific element') usually indicates usability problem z – are there any occurrences of pattern x without it being the usability problem z? Furthermore, it was helpful for the analysis reported here to have highly defined criteria for the extraction of usability problems. However, there was less clarity in the extraction of eye-tracking pattern sequences; investigating this would be helpful for further research. Finally, once correlations and an indication of their strength are validated, automated analysis should be tested. However, this needs to take into account differences between users as outlined above, as their cultural background and experience will influence their eye-tracking patterns.

6. ACKNOWLEDGMENTS

Many thanks to Ursula Armitage from Bunnyfoot for the suggestion of the research project (there was no formal agreement over the study) and to my friends and family who supported me during the process. Thanks also to the anonymous reviewers and Linden Ball, our meta-reviewer, for their detailed comments and encouragement.

7. REFERENCES

- [1] Ball, L. J., Eger, N., Stevens, R., and Dodd, J. Applying the PEEP method in usability testing. *Interfaces*, 67 (Summer 2006), 15-19.
- [2] Bojko, A. Using eyetracking to compare web page designs: A case study. *Journal of Usability Studies*, 3, 1 (2006), 112-120.
- [3] Cowen, L. *An Eye Movement Analysis of Web Page Usability*. Masters Thesis, Lancaster University, UK, Retrieved 03/07/06 from http://lauracowen.co.uk/download/mres_dissertation.pdf, 2005.
- [4] Cowen, L., Ball, L. J., and Delin, J. An eye-movement analysis of web-page usability. In *People and Computers XVI (Proceedings of HCI 2002)*. Springer, London, 2002, 317-335.
- [5] Ehmke, C. *Identifying Web Usability Problems from Post-Test Eye-Tracking Analysis*. MSc Thesis, City University London, UK, 2006.
- [6] Etre. *See Your Site through your Users' Eye: Eye Tracking*. 2006. Retrieved 27/09/06 from <http://www.etre.com/usability/eyetracking/>.
- [7] Goldberg, J. H., and Kotval, X. P. Computer interface evaluation using eye movements: Methods and constructs, *International Journal of Industrial Ergonomics*, 24, 6 (1999), 631-645.
- [8] Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., and Wichansky, A. M. Eye tracking in web search tasks: Design implications. In *Proceedings of the Eye tracking research and applications symposium (ETRA 2002)*. ACM Press, New York, NY, 2002, 51-58.
- [9] Jacob, R. J. K., and Karn, K. S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyona, R. Radach, and H. Deubel (eds.), *The mind's eye: Cognitive and applied aspects of eye movements*. Elsevier, London, UK, 2003, 573-605.
- [10] Krug, S. *Don't Make Me Think! A Common Sense Approach to Web Usability (Second Edition)*. New Riders, Berkeley, 2006.
- [11] McCarthy, J. D., Sasse, M. A., and Riegelsberger, J. "Could I have the menu please?": An eyetracking study of design conventions. In *People and Computers XVII (Proceedings of HCI 2003)*. Springer, London, 2003, 401-414.
- [12] McCarthy, J. D., Sasse, M. A., and Riegelsberger, J. The geometry of web search. In *People and Computers XVIII (Proceedings of HCI 2004)*. Springer, London, 2004, 249-262.
- [13] Nielsen, J. *Alertbox: Search and You May Find*. 1997. Retrieved 22/09/06 from <http://www.useit.com/alertbox/9707b.html>.
- [14] Outing, S., and Rual, L. *The Best of Eyetrack III: What We Saw When We Looked Through Their Eyes*. Published on Poynter Institute (not dated). Retrieved 20/06/06 from <http://poynterextra.org/eyetrack2004/main.htm>.
- [15] Poole, A., and Ball, L. J. Eye tracking in human-computer interaction and usability research. In C. Ghaoui (ed.), *Encyclopedia of human computer interaction*. Idea Group, Pennsylvania, 2005, 211-219.
- [16] Poole, A., Ball, L. J., and Phillips, P. In search of salience: A response time and eye movement analysis of bookmark recognition. In *People and Computers XVIII (Proceedings of HCI 2004)*. Springer, London, 2004, 363-378.
- [17] Renshaw, J. A., Finlay, J. E., Ward, R. D., and Tyfa, D. Designing for visual influence: An eye tracking study of the usability of graphical management information. In *Proceedings of the IFIP conference on Human-computer interaction (INTERACT 2003)*. IOS Press, London, 2003, 144-151.
- [18] Sibert, L. E., and Jacob R. J. K. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human factors in computing systems (CHI 2000)*. ACM Press, New York, NY, 2000, 282-288.
- [19] Tobii 2006. *Downloads: User Manual Tobii and ClearView 2.6*, Retrieved 18/07/06 from <http://www.tobii.com/support/downloads/>.
- [20] Vuori, T., Olkkonen, M., Pölonen, M., Siren, A., and Häkkinen, J. Can eye movements be quantitatively applied to image quality studies? In *Proceedings NordiCHI 2004*. ACM Press, New York, NY, 2004, 335-338.