Identifying Website Usability Solely from Gaze Data of Visual Search

MARTIN SÖDERBERG
ABSTRACT

If researchers are able to derive usability simply by analysing gaze data it provides a quick, objective and potentially automatic way of measuring the usability of an interface. In order to do that it is essential to know which traits of the gaze data that have an impact on usability. This paper investigates these traits by analysing different eye tracking metrics in the data. The goal is to see which of these metrics have a general correlation with usability. Previous research provides a clue about which metrics are useful when analysing usability. However, much of the research is based on subjective analysis or lacks in general applicability. This research provides an objective analysis that is independent of characteristics of the interface.

A user study is done on 20 participants. They complete tasks on nine different ecommerce websites while their eye movements are recorded. Correlation is measured between usability and eye tracking metrics in order to investigate which metrics that are sensitive to changes in usability.

The results show that fixational backtracks and number of fixations has the strongest correlation with usability. Previous research did suggest that both of these eye tracking metrics have an impact on usability.

SAMMANFATTNING


En användarstudie utförs på 20 deltagare. De utför uppgifter på nio olika websidor för e-handel medan deras ögonrörelser spelas in. Korrelation mäts mellan användbarhet och mätt i ögondata för att undersöka vilka mätt som är känsliga för förändringar i användbarhet.

Resultatet visar att tillbakabildande sackader och antal fixeringar har starkast korrelation med användbarhet. Tidigare forskning visade att båda dessa mätt påverkades av användbarheten.
1. INTRODUCTION
In today’s society people use computers and handheld devices for an ever increasing number of tasks such as banking, accessing news, shopping and communicating. Thus, the supply of websites and applications that provides assistance with these tasks has never been bigger. The tough competition between providers of these services leaves no room for usability issues as users will simply choose to go elsewhere. To guarantee highly usable products there is a great need for fast, reliable and cost effective ways of identifying usability problems.

The usual way of identifying usability problems is to conduct a user study where the participants use the interface of the product under supervision of a usability expert who gathers information about user behaviour in order to make an usability analysis (Leroy). This can be costly and time consuming as it requires usability experts who recruit users, conduct the study and analyse the findings, often by making a subjective analysis of recordings and participant responses.

In these user studies, tracking eye movements of users is commonly used to gather objective data about cognitive processes while using the interface. The eye tracking is done by an eye tracking device which records sample points many times each second of where the user is looking on the screen. The data is then processed into gaze data; data about when, where and for how long the user fixated their gaze. The data is usually presented as a heatmap image or a scanpath recording which lets the usability expert know which parts of the screen the user was looking at and in which order. The heatmaps and scanpaths can then be analysed subjectively by experts based on the visual stimuli fixated upon. (Poole, Ball).

In this paper, a method is tried to measure the usability of several e-commerce sites by only analysing certain eye tracking metrics in the gaze data. Usability is defined as the subjective assessment of ease of use and satisfaction when using the websites. Eye tracking metrics are quantifiable traits extracted from the gaze data such as fixation durations and distance between fixations. As the analysis is based solely on eye tracking metrics no subjective analysis of the visual stimuli are taken into consideration. So far, research has fallen short in its attempt to make a general correlation scheme between usability and traits of the gaze data. This has forced researchers to use subjective judgement when analysing the gaze data (Ehmke, Wilson).

The goal of this paper is to see which eye tracking metrics correlates with perceived usability. Research has been done to identify how gaze behaviour and traits in gaze data is affected by usability. However, this research has been based on subjective measures and analysis of the interface, or has been to restricted in its problems in order to apply in a more general context. The hope is that this paper can help usability practitioners to derive the level of usability solely from gaze data. The paper can also serve as a foundation for future automatisation of usability analysis by companies that seek to reduce costs of usability studies. If it were possible to build software that can reveal usability problems and user experience by simply analysing the gaze data developers can use this software to gain fast and objective insights about future improvements of their systems.

2. THEORY AND RELATED RESEARCH

2.1 Classification of fixations and saccades
Most eye tracking metrics are based upon the properties of recorded fixations and saccades. A saccade is the rapid eye movements between two fixations. When analysing saccades the length and directional shifts are essential properties. A regressive saccade, also called a backtrack, is a saccade with a larger than 90° directional shift from the preceding saccade (Sibert et al). A fixation is generally defined as a period of relative stability between eye movements. It is during the fixation that most of the visual information is gathered from the visual stimuli fixated upon. The duration of this period can vary between 50ms and longer depending on the stimuli (Salthouse, Ellis). 100ms has been set as a standard threshold for a short fixation in attentional studies because the eye needs this time in order to interpret the visual stimuli. Discarding fixations below 100ms is therefore recommended (Komogortsev et al.).

Accurate gaze drift correction is highly important in the classification process (Vadillo et al.). One algorithm that has drift correction is an algorithm proposed by Olsson (Olsson, 2007). The algorithm divides gaze data into subsets no shorter than 100ms. 100ms will therefore be the lowest possible gaze event that is classified as a fixation. Subsequent data points that are below a certain velocity threshold, measured in pixels/ms, are bundled up and an average of their positions is calculated. This position is the location of the fixation.

During the classification it is also important to discard gaze data when the eye tracker fails to determine the location. Olssons algorithm uses the previous and following fixation to interpolate a value to the discarded fixation between these fixations. This interpolation is common during fixation classifications. The motivation is to prevent one fixation as being classified as two separate fixations (Komogortsev et al.).

2.2 Usability in user studies
Usability can be measured by both objective and subjective methods. The objective methods look at
factors such as success rate and effectiveness of human performance e.g. completion time, number of errors and if the intended sequence of activities is followed. The subjective methods evaluate user impression on effectiveness, efficiency and satisfaction of the system. It is commonly measured as the subjective assessment of ease of use and satisfaction when using an interface (Shackel, Richardson).

In user studies, a common way to evaluate the usability of a system by subjective methods is to have the users answer a usability scale questionnaire. A classic and reliable tool for such measurements is the SUS usability questionnaire which measures usability in any system (Jordan et al.). The SUPR-Q questionnaire is another tool that measures the user experience when using a website in the factors usability, trust, appearance, and loyalty. SURP-Q has a high internal consistency and strongly correlates with the SUS-usability scale. The mean usability score of 2513 SUPR-Q questionnaire responses across 70 websites is 4.06 out of 5 (J. Sauro).

2.3 Usability and eye tracking metrics

Some eye tracking metrics are known to affect usability. Goldberg and Kotval investigated two very similar interfaces with different grouping and visibility of the tool icons. According to their definition of usability, a high usability interface had the buttons grouped reasonably by components with highlighted target buttons. Low usability interfaces had random groupings, incorrect highlighting and target buttons were greyed out. They found that an increased amount of overall fixations when completing a task indicated poor usability. They stated that longer saccades were an indication of meaningful visual queues in the interface. The fixation/saccade ratio they found to be a content independent ratio where a high ratio indicates that there was either less search activity or more processing. In the low usability interfaces users exhaustively searched the interface with small local saccades between fixations. They also found that repeat fixations on a target would indicate lack of meaningfulness or visibility. All of the target independent search measures that shown to affect usability were linearly dependent on usability to various degrees (Goldberg, Kotval).

Poole and Ball stated some additional measures that can generally be used to identify usability problems. They stated that longer fixations on a target indicate difficulty in extracting information or that the target is engaging in some way. If the user has difficulty in interpreting the stimuli of a page it could affect its usability. They also stated that regressive saccades, or backtracks, could mean that the user have trouble encoding the interface, that the interface layout does not match the expectations of the user or that the user’s goals have changed. (Poole, Ball).

When analysing four different web pages with very similar content and functionality Cowen et al. concluded that the measures total fixation duration and average fixation duration were significantly affected by usability. However, the average fixation duration measure was only sensitive to usability in a general sense but was highly dependent of which task was performed on a page. The usability level was measured as a series of performance measures including task completion time and whether or not the correct link was clicked (Cowen et al.).

2.4 Online shopping behaviour

There are two types of visual behaviour customers show when entering a web page, visual search and browsing. Visual search means that the user has a specific goal in mind that the user plans to accomplish. Browsing refers to when the user visually inspects a web page without a specific target goal (Lu et al.).

Three steps are included in completing a purchase on an e-commerce site. These are selecting an order, putting in personal information data and confirming the order with the provision of credit card data. Each of these steps are important as they can correspond to significant visitor loss (Sismeiro, Bucklin).

3. Research question

If it is possible to look solely eye tracking metrics when evaluating the usability of a website the usability evaluation could become much faster and more cost effective. In order to look solely at gaze data, a model that predicts usability from gaze data must work independently of interface design and must be able to identify and cope with different user behaviours. For development of such a model it is highly important to identify which properties in the gaze data that are sensitive to usability level and how sensitive they are. For the purpose of this paper a limitation is set as defined by this research question.

- How does eye tracking metrics correlate with website usability when a user visually searches the interface with set goals and further traits about web page design or user behaviour is unknown?

The goal of the study is to identify which, if any, eye tracking metrics that can be used generally as an indication of the usability level.

4. METHOD

A user study was conducted with 20 participants, 14 males and 6 females, between the ages of 19 and 29. The participants were instructed to take a seat on the chair in front of a screen with a Tobii X2 60Hz eye tracker connected. They were instructed to adjust their seating...
into a comfortable position within the angular and distance range of the eye-tracker. A calibration of the eye-tracker was done and accepted if the gaze accuracy was sufficient at all calibration points according to the calibration software and the researcher.

Once the setup was completed the participants were supposed to complete simple tasks on three, out of nine, different e-commerce websites. The order of these websites was evenly distributed among the participants and the websites were rotated so that the same number of participants solved tasks on each site. Task instructions were shown before each website was displayed. The tasks consisted of different types of order selections and information gathering. The aim was to let the user perform order selection and then stop just before the personal information input and confirmation stage of the purchasing process. A task was set to be done when the proper order had been put inside the shopping cart and when the information had been found. The participants were instructed to memorize the tasks and then proceed to execute them in silence while their gaze data was collected. In order to allow the participants to solve the tasks as they would have done naturally the participants were completely free in their strategy as long as they stayed within the website. This meant that participants could utilize the search engine of the website or navigate on the menu as they pleased.

Raw gaze data was collected once the page was fully loaded and until the participant indicated that all of the tasks on a site were completed by pressing a button. The gaze data was processed and filtered by an implementation of Olsson’s filter called Tobii fixation filter to be able to categorize the data into individual fixations.

To measure the usability of the site participants answered a SUPR-Q usability questionnaire after each website session before proceeding to the next website. The aim was to measure usability with a subjective method to be able to compare gaze data with how the user perceived the site in terms of usability.

After every task on each website was done the participants were interviewed about their e-commerce usage and impressions of the study. The video recordings and data of each participant was inspected to make sure the participant had made an effort to complete the tasks and that no segments of data was lost due to loss of signal.

When data collection was complete, eye tracking metrics known to have a relationship with usability were investigated for correlation. As a first determination of correlation between variables their scatter plot was examined for any apparent linear or nonlinear relationships. If no apparent nonlinear relationship was found the data was analysed by linear regression. The Pearson Product Moment Coefficient was used to determine the correlation between variables. If a sufficiently high linear relationship was found its confidence was determined with a requirement of $p<0.05$ to be valid relationship. A Pearson Product Moment Coefficient of at least ±0.1, ±0.3 and ±0.5 was used as thresholds for a weak, moderate and strong relationships as generally proposed in behavioural sciences (Cohen).

Below is a list of the eye tracking metrics used in the examination.

- **Fixation duration**, the duration of a single fixation.
- **Fixation count**, the number of fixations for one participant on a single site.
- **Saccadic amplitude**, the length in distance of one saccade.
- **Fixational backtracks**, also known as regressive saccades, is a change in gaze direction of 90° or more between fixations. A hypothesis was also tested changing the definition to a change in gaze direction of 180±5° to identify back and forth scanning along objects or navigation bar elements to find repeat fixations on a target.
- **Fixation/saccade ratio**, the ratio between the time the eye spent fixating and the time spent in saccadic motion.

5. RESULTS

Moderate correlation was found between usability and total fixation count and the total amount of backtracks (90° and 180°). There was also a weak-moderate correlation between the likelihood of a fixation being a backtrack (90°) and higher usability.

The measures fixation count, average fixation duration, average saccadic amplitude and fixation/saccade ratio showed no correlation with usability. A fixation was no more likely to be a backtrack (180°) regardless of the usability of the website.

5.1 Usability of the websites

The mean usability score of all of the websites was 3.53, the median was 3.6 and the internal variance was 0.313. It is slightly lower than the SUPR-Q mean of 4.06. The highest average usability score between sites was 4.25 and the lowest was 2.5.

5.2 Fixation duration and fixation count

Moderate correlation was found between fixation count and usability. No correlation was found between the average fixation duration and usability.

The Pearson product-moment correlation coefficient ($r$) between the average fixation duration and the usability score is $r = 0.024$. This is interpreted as no linear
correlation. As Cowen et al. concluded that average fixation duration is generally affected by usability this finding is in conflict with their conclusion.

![Figure 1. Scatter plot with linear trendline of the relationship between usability score and average fixation duration.](image1)

The correlation coefficient between fixation count and usability score is $(r = -0.378, p<0.01)$. This is interpreted as a moderate negative correlation between fixation count and usability score.

**5.3 Saccadic amplitude**

The correlation coefficient between average saccadic amplitude and usability score is $r=0.057$. This is interpreted as no linear correlation.

![Figure 2. Scatter plot with linear trendline of the relationship between usability score and the total fixation count.](image2)

**5.4 Fixation/saccade ratio**

No correlation was found between the fixation/saccade ratio and usability $r = 0.01$. More processing always requires more searching regardless of the usability of the site.

![Figure 3. Scatter plot with linear trendline of the relationship between usability score and average saccadic amplitude.](image3)

**5.5 Fixational backtracks**

This section is divided into two parts. In the first part, a fixational backtrack is defined as a change in gaze direction of 90° or more, hereafter referred to as backtrack (90°). In the second part, the change in gaze is restricted to 180±5°, referred to as backtrack (180°).

The results suggest that there are more backtracks (90°) and backtracks (180°) when usability is low. This is likely due to there being less total fixations for the higher usability cases. The results also shows that a fixation is more likely to be a backtrack (90°) if usability is high.
5.5.1 Backtracks (90°)
The correlation coefficient between total amount of backtracks (90°) and usability score is ($r = -0.342$, $p<0.01$). A greater usability score leads to fewer total backtracks (90°).

![Figure 5. Scatter plot with linear trendline of the relationship between usability score and amount of backtrack (90°) fixations.]

The correlation coefficient between the percentage of fixations being a backtrack (90°) and usability score is ($r = 0.293$, $p<0.05$). The percentage of fixations being backtracking fixations increases as usability increases.

![Figure 6. Scatter plot with linear trendline of the relationship between usability score and backtrack (90°) fixations as a percentage of fixation count.]

The total amount of backtracks (90°) is likely to be strongly affected by the total amount of fixations. Therefore, high usability interfaces will produce less backtracks. However, the percentage of fixations being backtracks (90°) increases with usability, as illustrated by Figure 6. This is remarkable since Poole and Ball stated that a backtrack (90°) could be an indication of user’s expectation not being met.

5.5.2 Backtracks (180°)
The correlation coefficient between total amount of backtracks (180°) and usability score is ($r = -0.360$, $p<0.01$). A greater usability score leads to fewer total backtracks (180°).

![Figure 7. Scatter plot with linear trendline of the relationship between usability score and number of backtrack (180°) fixations.]

The correlation coefficient between the percentage of fixations being a backtrack (180°) and usability score is $r = -0.07$. This is interpreted as no linear correlation.

![Figure 8. Scatter plot with linear trendline of the relationship between usability score and backtrack (180°) fixations as a percentage of fixation count.]

Since the likelihood for a fixation to be a backtrack (180°) is not affected by usability the conclusion can be made that the negative correlation between backtracks (180°) and usability is, like proposed in the case of backtrack (90°), simply due to there being more total fixations on the less usable sites.
5.6 Participant interviews
This section presents the results from the concluding interview where the participants were asked about their e-commerce usage behavior and where they had the opportunity to raise concerns about the study.

All of the participants reported having used similar e-commerce websites before. However, the participants were not familiar with the specific websites they used in the study except for in a few cases. 17 out of 20 participants reported visiting or purchasing from an e-commerce website at least one time every other month. Only one participant reported having trouble at memorizing the given tasks. Most of the participants agreed to the notion that the study setup let them use the website as they would have done during normal circumstances. However, many participants felt that they would have taken more time to browse the site under normal conditions. Some participants also reported using price comparison sites and search engines like Google to navigate directly to an item in the webshop.

“I felt a little rushed to complete the given tasks. If I was at home I probably would have taken more time to browse the site.”

A majority of the participants used the sites search functions as a first tool to navigate in the site. Only when the search function failed the participants started to navigate through the navigation bars.

5.7 Individual site data
Here, data is divided into groupings representing the individual sites. An observation suggests that even when comparing sites with similar usability level, the deviation between eye tracking metrics is large. This makes it infeasible to compare usability between websites solely by analysing these eye tracking metrics.

Table 1. Individual site averages of eye tracking metrics.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Usability score</th>
<th>Fixation duration</th>
<th>Fixation count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adlibris.se</td>
<td>3.82</td>
<td>335</td>
<td>3</td>
</tr>
<tr>
<td>Babyländ.se</td>
<td>3.6</td>
<td>368</td>
<td>2</td>
</tr>
<tr>
<td>CDON.se</td>
<td>4.04</td>
<td>327</td>
<td>2</td>
</tr>
<tr>
<td>Clas Ohlsson.se</td>
<td>3.82</td>
<td>370</td>
<td>3</td>
</tr>
<tr>
<td>Elgiganten.se</td>
<td>3.85</td>
<td>313</td>
<td>2</td>
</tr>
<tr>
<td>Scrapweb.se</td>
<td>2.5</td>
<td>344</td>
<td>3</td>
</tr>
<tr>
<td>Sportgym.se</td>
<td>4.25</td>
<td>311</td>
<td>2</td>
</tr>
<tr>
<td>Tekunjen.se</td>
<td>3.07</td>
<td>320</td>
<td>3</td>
</tr>
<tr>
<td>Zara.se</td>
<td>3.66</td>
<td>314</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2. More site averages of eye tracking metrics.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Saccadic amplitude</th>
<th>Backtracks 90°</th>
<th>Backtracks 180°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adlibris.se</td>
<td>5.00</td>
<td>160</td>
<td>32</td>
</tr>
<tr>
<td>Babyländ.se</td>
<td>5.00</td>
<td>124</td>
<td>24</td>
</tr>
<tr>
<td>CDON.se</td>
<td>5.55</td>
<td>120</td>
<td>24</td>
</tr>
<tr>
<td>Clas Ohlsson.se</td>
<td>4.99</td>
<td>131</td>
<td>30</td>
</tr>
<tr>
<td>Elgiganten.se</td>
<td>6.02</td>
<td>115</td>
<td>18</td>
</tr>
<tr>
<td>Scrapweb.se</td>
<td>4.97</td>
<td>148</td>
<td>30</td>
</tr>
<tr>
<td>Sportgym.se</td>
<td>5.57</td>
<td>108</td>
<td>22</td>
</tr>
<tr>
<td>Tekunjen.se</td>
<td>5.49</td>
<td>167</td>
<td>32</td>
</tr>
<tr>
<td>Zara.se</td>
<td>8.28</td>
<td>125</td>
<td>38</td>
</tr>
</tbody>
</table>

6. DISCUSSION

6.1 Findings and prior research
The study managed to find linear correlation between some eye tracking metrics and usability. Total fixation count does have moderate correlation with usability. This finding is in line with Goldberg and Kotval’s conclusion. The reason for this correlation is likely to be that total fixation duration is strongly related to time spent on site. It was also shown that total amount of backtracks (90° and 180°) are correlated with usability in similar fashion. If the user is spending more time looking at a site it is very likely that the data will contain more fixations, and thus also more backtracking fixations. The fact that users spend more time on sites with low usability can seem strange since low usability are known to repel customers. However, in this case, the users were visually searching the interface with set goals without any real option to close the site and walk away. The longer the user spent navigating the site the longer it took to complete the tasks.

When only the metrics related to time spent on site affect usability one might simply use task completion time as a measurement of usability instead. However, the study found that there are more backtrack (90°) fixations as a percentage of all fixations if usability is high. This metric is not dependent on the time spent on site. These results were surprising as Poole and Ball stated that a backtrack (90°) could happen due to the interface not meeting user expectations. There could be many reasons for this correlation. Perhaps a backtrack (90°) can be used to describe other traits of user cognition than what Poole and Ball suggested. One hypothesis is that a high backtrack (90°) percentage indicates that users effectively identifies that an area is not of interest and quickly adjust their gaze to another location instead of exhaustively searching that area.

No correlation was found between average fixation duration and usability. This is in conflict with the conclusions found by Cowen et al. The reason for this could be due to the additional complexity of the stimuli or due to the different ways of measuring usability used in the studies. It could also be due to the low amount of
different webpages used by Cowen et al. As displayed by Table 1 and Table 2 there are variations in the average fixation duration, even among pages with similar usability. Therefore, when comparing only four different pages it is likely to come up with similar conclusions.

There was also no correlation to be found between backtrack (180°) and usability. There are no consensus that backtracks (180°) can be used as an effective way of measuring back and forth fixations between objects or navigation bar elements like proposed in this paper.

Further investigation would be required in order to find if any target independent eye tracking metric could be used to describe this behaviour.

The findings suggest that it is hard to confidently predict what a single user thinks about a system’s usability by solely examining these eye tracking metrics. Prediction is hard due to the large deviation in eye tracking metrics between users as well as large variations between the websites, even when corresponding usability scores were similar. However, the findings can help usability practitioners identify which traits in gaze data that are sensitive to usability.

### 6.2 The subjective nature of usability

Although there are many ways of measuring usability objectively, usability is in its essence about people’s experience when using a system. Any objective method that claims to measure usability must agree with people’s opinion about the usability of the system. Otherwise, it is not meaningful as it does not reflect the perception of those who are going to use the system.

If usability is subjective in its essence, how is it possible to develop an objective measurement that reflects people’s opinion? Perhaps it is unthinkable that an objective method that predicts usability can correspond flawlessly with the experience of the user for each case. Perhaps the aim should be to find the methods that objectively can measure usability and have the highest correspondence with people’s perception about usability.

### 6.3 Method criticism

An objective method similar to how Goldberg and Kotval and Cowen et al. could have been used when evaluating the usability of the websites. Although, there are many objective methods to measure the effectiveness of human performance there are still limits to how well we can measure the user’s own experiences objectively. In development of such methods subjective measurements are needed in order to develop this method. However, it would be interesting to compare the results using different usability evaluation methods and to see how it affects the conclusions.

The method of use was a compromise between controlled conditions and general applicability where applicability was of greater importance. It was important that no restrictions were set on page design or on participants in terms of behaviour to be able to answer the research question. However, as not many studies had been done looking solely at gaze data to determine usability one might have been better of starting with more controlled conditions. For example, the use of on-site search functionality could have been restricted to force users to use the menus, tasks and stimuli could have been selected so that there are not multiple ways of solving a task. Also, simpler and more controlled visual stimuli could have been selected.

The dynamic nature of the websites were found to be troublesome when looking solely at the gaze data. The visual stimuli provided by the page was constantly changing. Dynamic drop down menus, scrolling abilities and clickable elements made it unrealistic to use regular area of interest metrics such as repeat fixations on an area of interest or time spent fixating on a certain area. In order to investigate area of interest metrics without further analysis of the interface it must be guaranteed that the interface is not changing, or that some identification process is used to identify when the interface changes between two fixed states. The eye tracking metrics could then be split up between these states for separate analysis. When manually analysing gaze recordings dynamic elements are less of a problem as the researcher can identify changes in the interface and interpret the visual stimuli.

### 6.4 Future research

There is a need to develop a functioning methodology for predicting usability solely from gaze data that can be applied in practice. There is also possible, perhaps very likely, that gaze data and usability are related in a more complex way than this paper attempts to look at. In order to identify these complex and possibly unexpected relationships techniques such as machine learning could be applied. However, in order to train an algorithm to predict usability based on many variables a lot of data is required for confident predictions. This paper can help researchers with feature selections for these models by selecting data that affects usability.

### 6.5 Ethics and Sustainability aspects

Usability studies is something that can be overlooked at the development stage as many see it as too time consuming, costly and ineffective at giving insight. A more efficient and effective process of identifying usability issues makes it economically sustainable to perform usability studies regularly in the development process. Effectivizing usability identification will lead to systems that are more effective at their purpose and easier to use. This would limit the time workers and
people in general unnecessarily spend navigating sub-optimal systems. An overall improved usability could also provide greater opportunity of involvement in the systems by people who are excluded because of their inability to handle the systems.

At some point computers might automate the process of pointing out design flaws and making improvements. At this point human influence in the systems supposedly used by humans is limited. A computer might not have the same considerations of ethics and moral aspects that are essential for the design. The future role of usability practitioners might therefore be to control this process to make sure that the system is nondiscriminatory and that everyone has an equal opportunity of using the systems.

7. CONCLUSION
Correlation was found between some eye tracking metrics and usability. Fixation count and the number of backtrack (90° and 180°) fixations were shown to have moderate negative correlation with higher usability. These metrics are all strongly related to time spent on site. The conclusion made here is if task based goals are set and the user spends a long time completing these goals it is an indication that the usability of the site is low.

In contrast to previous research, a weak-moderate positive correlation was found between percentage of fixations being backtrack (90°) fixations and usability. The reason might be that a backtrack (90°) is describing other traits than expectations not being met.

While all of the metrics can be applied to get information about user cognition some metrics did not correlate with usability. The metrics that showed no significant correlation with usability were average fixation duration, average saccadic amplitude, fixation/saccade ratio and percentage of fixations being a backtrack (180°). Some of these conclusions stand in contrast to prior research.

8. REFERENCES
4. Ehmk Co., Wilson S., Identifying web usability problems from eye-tracking data. Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it, September 03-07, 2007, University of Lancaster, United Kingdom