การทดสอบการใช้งานได้ของเว็บไซต์ด้วยระบบการติดตามดวงตาเฟซแล็บ

Web Site Usability Testing with faceLAB Eye Gaze Tracking System

*ปีลันธ ปิยศิริเวช
Pilun Piyasirivej

บทคัดย่อ

งานวิจัยนี้เป็นการศึกษาทดลองซึ่งใช้การติดตามดวงตาเป็นวิธีการประเมินความใช้งานได้ของเว็บไซต์ โดยมีผู้เข้าร่วมการทดลองเป็นจำนวนทั้งสิ้น 39 คน ผู้เข้าร่วมการทดลองแต่ละคนได้เข้าเยี่ยมเว็บไซต์สามเว็บจากสามประเภท (สารสนเทศ, การบันเทิง, การพาณิชย์อิเล็กทรอนิกส์) และสวมบทบาทของผู้ใช้ที่กำหนดให้ รูปแบบการเคลื่อนไหวดวงตาของผู้เข้าร่วมการทดลองทั้งหมดได้ถูกบันทึกไว้อย่างอัตโนมัติ ด้วยอุปกรณ์ติดตามดวงตาเฟซแล็บ ตัวชี้วัดที่ถูกพิจารณาและเปรียบเทียบระหว่างสามเว็บไซต์ได้แก่ ระยะเวลาฟิกเซชั่นเฉลี่ย ระยะเวลาฟิกเซชั่นรวม จำนวนฟิกเซชั่น และเพอร์คลอสเฉลี่ย จากการทดลองพบว่าตัวชี้วัดบางตัวให้ค่าที่แตกต่างกันระหว่างเว็บไซต์ โดยเฉพาะอย่างยิ่งระยะเวลาฟิกเซชั่นเฉลี่ย ซึ่งแตกต่างกันถึงสองครั้งจากการเปรียบเทียบสามครั้ง

---

*Assistant Dean and Director of Information Technology Program, Dhurakij Pundit University International College
E-mail : pilun@it.dpu.ac.th

Abstract

This paper describes an experimental study which utilised eye gaze tracking as a method for usability evaluation of web sites. There were 39 participants involved in this experiment; each participant visited the same three web sites from three categories (information, entertainment, e-commerce) and completed given use scenarios. All participants' eye gaze patterns were recorded automatically with the remote eye gaze tracking equipment called faceLAB. The following measures were investigated and compared across three web sites: average fixation duration, cumulative fixation time, number of fixations, and average PERCLOS. It was found that some of the measures were different across web sites. Average fixation duration, in particular, was found to be different in two out of the three comparisons.

คำสำคัญ : การติดตามดวงตา, การเฝ้าสังเกตดวงตา, เวิลด์ไวด์เว็บ, ความใช้งานได้, ปฏิสัมพันธ์มนุษย์กับคอมพิวเตอร์

Keywords : Eye Gaze Tracking, Eye Gaze Monitoring, World Wide Web, Usability, Human-Computer Interaction

1. Introduction

Eye gaze tracking is a relatively new technique used to observe users while browsing web sites. It is also non-intrusive (when remote eye gaze tracking equipment is employed) and provides statistical data promptly for usability analyses. For the past few years, it is considered by many theorists and practitioners to be one of the most promising techniques for usability evaluation. However, at present, the number of eye gaze tracking studies related to web site usability is fairly small; Stanford-Poynter (2000), Cowen, Ball, and Delin (2002), Goldberg, Stimson, Lewenstein, Scott, and Wichansky (2002), and Pan Hembrooke, Gay Granka, Feusner, and Newman (2004) are some examples of such studies. In order to explore this technique in further details, this study utilised the faceLAB eye gaze tracking system to compare the usability of three web sites.

The faceLAB system, marketed by an Australian company named Seeing Machines, combines head, face, and eye tracking into one. The faceLAB system is non-intrusive to the user because it does not require any device to be worn by the user. Instead of the usual single camera, the faceLAB system uses dual cameras. The following quote from Seeing Machines explains how faceLAB works: *"faceLAB is a monochrome video-based tracking system, that is able to find a face in an image sequence and then home in on facial landmarks, such as the lips, nose and eyes, to obtain accurate position estimates of the features in each image. Data is then combined from two video sources (stereo), to generate 3D measurement information. Once the head-pose is located, the eye images are examined and the gaze direction is inferred from pupil orientation."* (Seeing Machines, 2003)

2. Details of the faceLAB system

The version of faceLAB used in this experiment is faceLAB 3.2.1, which contains dual cameras, PC, and accessories. The dual cameras are mounted on a stand (called Stereo-Head). The distance between the two cameras and their tilt angle can be adjusted easily according to the position of the user whose eye gaze is to be tracked. It is recommended that the distance between the cameras and the user be between 0.5 to 2.0 metres, while the tilt angle of the cameras must not exceed 15 degrees. It is best to select a user distance and a tilt angle that enable the cameras to capture the whole face of the user. During the tracking process, faceLAB can measure head-pose, gaze direction, saccades, attention, what

objects people are looking at, eye closure, PERCLOS fatigue measure, and eye blinks.

To track the user's face and head position, faceLAB first builds a model of the user's face. This process relies on unique features of the face, such as eye and lip corners, or artificial markers that can be applied to increase tracking accuracy. Therefore, it is necessary to build a unique model for each user under specific lighting conditions and camera settings. If any of these factors change (e.g. marker position, lighting, camera tilt angle, etc.), the model previously created cannot be used any longer and a new model has to be built. It is simple to build a model as faceLAB provides a built-in semi-automated application for this purpose. The application takes snapshots of the user and allows the experimenter to select unique features of the user's face. Two types of models can be created within faceLAB: full head and front-only. While a full head model is more suitable in the case where the user's head rotates more than 30 degrees to the left or right during the tracking process, it requires more unique features to be tracked, usually around the sides of the head, e.g. ears, side-burns, earrings, etc. In addition to face/head model building, faceLAB also requires the calibration of eye gaze in order to track the user's eye gaze accurately. This can be done by having the user look into each camera for a few seconds. The following parameters can also be adjusted to increase the accuracy: pupil radius, iris radius, and eye opening. Further details of the procedures for using the faceLAB gaze tracker are provided in the next section.

3. Experimental procedures

In this study, an eye gaze tracking experiment was conducted with 39 participants, all of whom were undergraduate students. Three Web sites were selected for the experiment to keep the duration within reasonable limits, and the selection was made so that one Web site was chosen from each of the three categories (information, entertainment, e-commerce). The three web sites selected were:

A) Full Sail (information site)

B) Madonna (entertainment site)

C) Chronicle Books (e-commerce site)

The eye gaze tracking experiment was conducted in a specialised laboratory equipped with Seeing Machines' faceLAB 3.2.1 head/eye tracking system (released in 2004), which included faceLAB PC (for running the tracking software), Stereo-Head (dual cameras for captur-

ing face/eye images), and the Overlay Unit. The installation of the faceLAB system was completed one week prior to the experiment. In addition to the faceLAB PC, another PC with a 17-inch monitor (User PC) was used to display web sites to the user during the experiment. While eye gaze tracking was in process, faceLAB PC also generated screen Intersections that displayed the approximate area on the screen where the user was looking at, represented as a red dot on the screen. screen Intersections were also shown to the user during the screen calibration process, just before the start of the experiment. During screen calibration, a white dot appeared on the screen and moved to different points around the screen. The user was required to stare directly at the white dot, so that the cameras could capture images of the user's eyes when he/she looked at different parts of the screen. When the screen calibration finished, the user could no longer see screen intersections. Instead, he/she was presented with stimulators (Web sites). However, both screen intersections and web sites were visible to the experimenter at all times. They were also combined (using the Overlay Unit), and then recorded for future analyses. Figure 1 depicts the schematic diagram of the connection of all equipment within the laboratory.

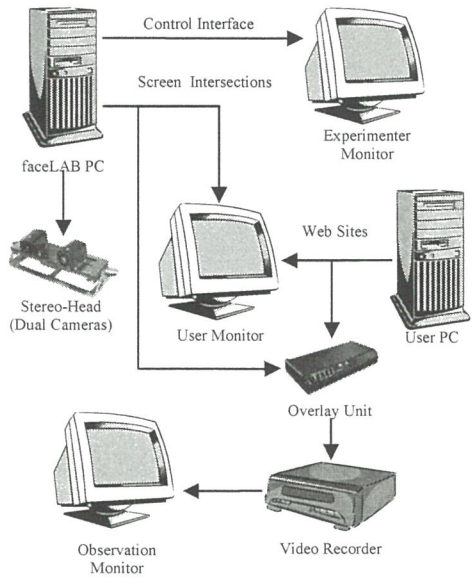


Figure 1 Schematic diagram showing eye gaze tracking equipment and its connection

During the two-week experimental period, each of the participants visited the eye tracking laboratory one by one. Each participant was given a one-hour time slot for the experiment, and he/she was monitored by the researcher who was also present in the laboratory while the experiment was taking place. When entering the room, the participant was asked to put four markers on their face (as shown in Figure 2) and sit on an adjustable chair in front of the cameras and PC monitor.

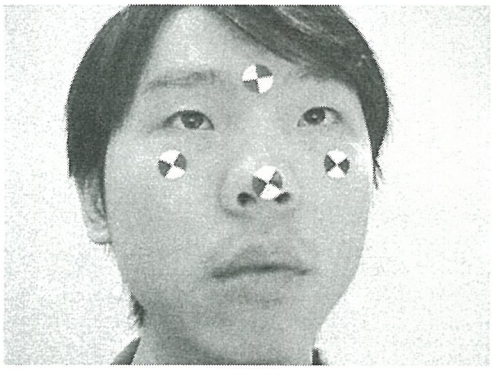


Figure 2 Positions of four markers on the user's face

Markers were used on all participants (except one participant who was allergic to Latex material used in the markers) because they were recommended by Seeing Machines for increasing tracking accuracy. After the user applied the markers and adjusted the height of the chair, their face model was created as per faceLAB guidelines, and then eye gaze calibration and screen calibration took place.

After the screen calibration, the participant was asked to view a brief PowerPoint presentation on the screen to measure the accuracy of the tracking process. The presentation displayed multiple five-digit numbers on the screen, one by one, on different parts of the screen, and the user was asked to read the numbers out loud in order to aid the experimenter in verifying the accuracy of eye gaze tracking. When the presentation finished, the user was directed (via an onscreen message) to a web page where they were presented with links to the three web sites mentioned earlier. The user then followed use scenarios for each web site. While the participant was browsing the web sites, their eye movement data was collected automatically by the faceLAB system.

4. Data collection and analysis

Raw eye gaze data collected from the faceLAB system is in a binary form and contains various eye/head movement measurements. Therefore, it is necessary to extract only some data fields that are relevant to this study as the first step towards quantitative data analysis. The

faceLAB eye tracking system records the user's eye gaze position at 60 Hz frequency, i.e. it captures the user's eye gaze every 1/60 of a second. Each time faceLAB captures eye gaze data, it writes that data to disk, together with a unique numeric identifier called frame. Frame number increases as more data is captured over time, and it is reset once the system is restarted (in this experiment, this was done every time a new participant came in, to prevent overflow in the frame number). During the experiment, a binary file with FLL extension (short for faceLAB Log) is created for each participant. It contains all data captured in a proprietary format (already compressed). The size of the data file depends on the duration of the experiment. The longer the experiment, the larger the file. On average, the size of a faceLAB log file is approximately one megabyte per minute recorded.

Seeing Machines, the provider of faceLAB, offers a software application named FaT (faceLAB Toolbox) which allows selected data (user-defined frame numbers and data fields) to be exported into an ASCII (text) file, which can be further analysed with other software applications, e.g. Microsoft Excel, SPSS. In this study, each log file (for each participant) was exported into three ASCII files, one for each Web site visited. The following data fields were selected for exporting because they were crucial to the analysis of web site usability: frame number (for reference only), saccade, PERCLOS, and gaze-screen intersection. The saccade field contains a Boolean value (0 or 1), where 0 means there was no saccade detected during that frame, and 1 means there was a saccade. PERCLOS is a fatigue measure endorsed by the National Highway Traffic Safety Administration (U.S. Department of Transportation), which is the percentage of the intervals with closed eyes in a fixed time window (disregarding regular blinks). PERCLOS values range between 0 and 1. For example, if a user has 5 eye closures of 3 seconds each (for a total of 15 seconds) within a 60-second period, his/her PERCLOS value is 0.25 (=15/60). The gaze-screen intersection field contains a Boolean value, where 0 means that the user's eye gaze did not intersect with the computer screen, while 1 means the opposite. Therefore, it is particularly useful for checking whether the user is looking at the screen or away from it.

Once the required data was extracted, it was imported into Microsoft Excel for further analysis. First, the gaze-screen intersection field was used to filter out any frame that the user's

eye gaze fell out of the screen area (gaze-screen intersection = 0). After the unwanted frames were removed, fixations were extracted by using the information from the saccade field. This method was chosen because fixation information was not available directly from the faceLAB system. However, Seeing Machines, the provider of faceLAB, suggested that at any time when there was no saccade (saccade = 0), there had to be fixation. Therefore, the number of consecutive frames with no saccade was counted as one fixation and the duration of each fixation (in seconds) was calculated by dividing the number of frames in the fixation by 60 (as the gaze data was recorded at the rate of 60 frames per second). Fixation before the first saccade or after the last saccade was omitted because its length could not be determined (incomplete fixation). This makes the first real fixation happen after the first saccade and the last real fixation happen before the last saccade. For the same reason, any incomplete fixation occurred immediately before the user looking out of the screen or immediately after the user looking back into the screen was also removed. A Visual Basic macro was written to help automate the extraction process within Excel.

For the usability evaluation of a computer user interface, including web sites, several eye gaze measures can be used. However, only the following measures were selected for this study because they were effective in measuring web site usability and supported by the faceLAB system:

Average fixation duration. This is the average duration of all fixations over the period of viewing the web site. Average fixation duration is the most frequently used measure for usability evaluation, as it is accepted that more complex information/representation usually requires a higher level of processing and therefore leads to longer fixation duration time. This is emphasised by Goldberg and Kotval (1999) that representations which require long fixations are not as meaningful to the user as those with shorter fixation durations.

Cumulative fixation time per second. Cumulative fixation time is the sum of duration of all fixations over the period of viewing the web site. Cowen et al. (2002) has suggested that cumulative fixation time (they call it total fixation duration) is a global measure of the total amount of processing performed on each page (or in this study, each web site), rather than just

the mean amount of processing on each part of a page (or a web site). However, since different users spend different amounts of time viewing the same web site, this cumulative fixation time is divided by the total amount of time viewing the web site, so that the cumulative fixation time of each person is comparable to others. Cumulative fixation time per second can also be calculated by summing the total number of fixation frames and dividing it with the total number of all frames; this method was implemented in the analysis of this experiment.

Number of fixations per second. The number of fixations is the total number of individual fixations over the period of viewing the Web site. Tzanidou (2003) suggests that an increase in the number of fixations indicates the difficulty in extracting information or the frustration of the interaction with the user interface. However, the number of fixations could also indicate the amount of interest in a specific area of the screen. For the same reason as the previous measure, the total number of fixations is divided by the time used for viewing the web site, to provide the number of fixations per second.

Average PERCLOS. This is the average PERCLOS value over the period of viewing the web site. PERCLOS is an effective indicator of fatigue (or alertness), as it is the percentage of eyelid closure over the pupil over time and reflects slow eyelid closures (droops) rather than blinks (Federal Highway Administration, 1998). The higher the value, the greater tiredness indicated. PERCLOS measurement is often used in detecting a driver's alertness while driving a motor vehicle. However, it is hypothesised that PERCLOS could also be used in user interface evaluation as well, because the effectiveness of a user interface should have a negative relationship with the user's boredom (lack of engagement), whose effect is very similar to that of the user's fatigue.

5. Results of the study

After the measurement data was calculated for each participant, quantitative analysis was performed by considering Web site category (information, entertainment, e-commerce) as independent variables, while eye gaze tracking measures, as discussed in the previous section, were regarded as dependent variables. The following research question and hypothesis were formulated and tested according to the quantitative eye gaze data collected from the experiment:

Q1: Are users' eye gaze measures affected by category of web site version viewed (information, entertainment, or e-commerce)?

H1: Users' eye gaze measures can be affected by category of web site version viewed (information, entertainment, or e-commerce).

Table 1 depicts descriptive statistics of eye gaze measurements considered in this study.

Table 1 Descriptive statistics of eye gaze variables

Variable	N	Min	Max	Mean	SD
Average Fixation Duration	117	.629	4.787	1.834	.771
Cumulative Fixation Time per Second	117	.790	.985	.941	.035
Number of Fixations per Second	117	.204	1.336	.601	.238
Average PERCLOS	117	.000	.140	.022	.032

In order to test hypothesis H1 , cases were grouped according to the Web site visited, as there were three Web sites, each of which falls in a different category. In this case, paired-sample t-tests were performed because each participant had visited all three Web sites. Comparisons between the information site and the entertainment site (A-B), between the entertainment site and the e-commerce site (B-C), and between the information site and the e-commerce site (A-C) were made, as shown in Tables 2, 3, and 4, respectively.

Table 2 Results from paired t-tests comparing information and entertainment web sites (A-B)

Variable	Mean Diff.	SD	SEM	t	Sig. (2-tailed)
Average Fixation Duration	.019	.486	.078	.244	.808
Cumulative Fixation Time per Second	-.007	.028	.004	-1.661	.105
Number of Fixations per Second	.010	.176	.028	.367	.716
Average PERCLOS	-.004	.021	.003	-1.234	.225

Table 3 Results from paired t-tests comparing entertainment and e-commerce web sites (B-C)

Variable	Mean Diff.	SD	SEM	t	Sig. (2-tailed)
Average Fixation Duration	.240	.505	.081	2.967	.005*
Cumulative Fixation Time per Second	.013	.032	.005	2.500	.017*
Number of Fixations per Second	-.061	.142	.023	-2.661	.011*
Average PERCLOS	-.005	.020	.003	-1.495	.143

* = significant at .05 level

Table 4 Results from paired t-tests comparing information and e-commerce web sites (A-C)

Variable	Mean Diff.	SD	SEM	t	Sig. (2-tailed)
Average Fixation Duration	.259	.682	.109	2.368	.023*
Cumulative Fixation Time per Second	.005	.037	.006	.912	.367
Number of Fixations per Second	-.050	.213	.034	-1.471	.150
Average PERCLOS	-.009	.024	.004	-2.303	.027*

* = significant at .05 level

As shown in Table 2, no significant difference was found in the eye gaze variables between information and entertainment web sites. However, there are significant differences in average fixation duration between entertainment and e-commerce web sites, and also between information and e-commerce web sites, i.e. participant's fixation duration was shorter in the e-commerce Web site than the other two Web sites. This could be because the e-commerce Web site is more organised, so that the user does not need to fixate on specific areas of the site for so long to understand what they represent. The significant difference in cumulative fixation time between entertainment and e-commerce Web sites can also be interpreted the same way, as cumulative fixation time also indicates the amount of processing required to understand the representations. The lower number of fixations in the entertainment Web site (compared to the e-commerce Web site) could be resulting from longer fixation durations, because when fixations are counted during a specific period of time, longer fixations usually

generate a smaller counted number. In addition, the e-commerce web site incurred a significantly higher average PERCLOS value than the information web site. This could indicate that the e-commerce web site is less interesting to these users than the information web site, and therefore engages the user less strongly.

According to these significant differences found in some eye gaze measures across the Web sites, it could be concluded that hypothesis H1 is partially supported. Although the order of viewing the web sites was the same for all participants (web site A, then B, then C), the possible effects of web site order on eye gaze behaviours were found to be minimal, because there were only three web sites used in the experiment and no significant difference in eye gaze variables was found between the first two web sites.

6. Conclusion

Results from the eye gaze tracking experiment show some differences in eye gaze patterns across web sites. The difference in average fixation duration was found in two out of three comparisons, while the difference in other measures was found in one comparison only. This could be due to the nature of eye tracking technologies which are not yet in their prime and still evolving, or it could be because the difference is too subtle for the eye tracking method to perceive. This could be confirmed by future research. In addition, one web site selected for each category may seem to be a small representative of web sites in that category; therefore, further investigation with a larger number of web sites could be undertaken, so that the chosen web sites could better represent their category.

It is also worth noting that the participants in this study were all undergraduate students. Therefore, the findings in this study are only able to be generalised to parts of the user population with the same characteristics. However, if the results need to be applied to different population groups, such as children or working professionals, it is recommended that further research be undertaken with samples from those population groups, in order to gain results which are more relevant to the groups in consideration.

Regarding the eye gaze tracking technique in general, more research studies have to be undertaken, possibly with different tools and/or methodologies. This is because there are many eye tracking systems available, which can support different kinds of measurements. For

example, gaze trails and look zones, which may be beneficial for the usability evaluation of web sites, are not supported by the faceLAB system utilised in this study. It is also inevitable that further research needs to be conducted with various eye tracking systems, so that the eye gaze tracking technique can be evaluated more effectively, as results from several studies can be compared and comparative performance can be measured, which in turn will benefit practitioners who are seeking suitable eye tracking systems for usability applications.

References

L. Cowen, L.J. Ball, and J. Delin, "An eye movement analysis of webpage usability," **Proceedings of HCI 2002,** Springer-Verlag, London, 2002, pp. 317-335.

Federal Highway Administration, "PERCLOS : A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," 1998; **http://www.fmcsa.dot.gov/ pdfs/tb98-006.pdf.**

J.H. Goldberg and X.P. Kotval, "Computer interface evaluation using eye movements: Methods and constructs," **International Journal of Industrial Ergonomics, vol. 24**, Elsevier Science, New York, 1999, pp. 631-645.

J.H. Goldberg, M.J. Stimson, M. Lewenstein, N. Scott, and A.M. Wichansky, "Eye tracking in web search tasks: Design implications," **Proceedings of ETRA 2002,** ACM Press, New York, 2002, pp. 51-58.

B. Pan, H.A. Hembrooke, G.K. Gay, L.A. Granka, M.K. Feusner, and J.K. Newman, "The determinants of web page viewing behavior: An eye-tracking study," **Proceedings of ETRA 2004**, ACM Press, New York, 2004, pp. 147-154.

Seeing Machines, "faceLAB frequently asked questions," 2003; **http://www.seeingmachines.com.**

Stanford-Poynter Project, "Eyetracking online news," 2000; **http://www.poynterextra.org/et/i.htm.**

E. Tzanidou, "Eye tracking as a complementary usability evaluation technique for e-commerce sites," 2003; **http://computing.open.ac.uk/interact2003/Papers/ ETzanidouPP.pdf.**