Correlation between the gazing point movement and usability for the purpose of Web page selection containing usability problems

Shota Fukushima[1], Noboru Nakamichi[1], Makoto Sakai[2], Kazuyuki Shima[3],
Ken'ichi Matsumoto[1]

[1]Graduate School of Information Science, Nara Institute of Science and Technology 8916-5, Takayama, Ikoma, Nara, 630-0101, JAPAN
  {shota-f, noboru-n, shima, matumoto} @is.aist-nara.ac.jp
[2]SRA Key Technology Laboratoly, Inc. Marusho Bldg. 5F, 3-12, Yotsuya, Shinjuku-ku, Tokyo, 160-0004, Japan
  sakai@sra.co.jp
[3]Department of Computer and Media Technologie Faculty of Information Science, Hiroshima City University 3-4-1, Ozuka-Higashi, Asaminami-ku, Hiroshima, 731-3194, JAPAN
  shima@computer.org

Abstract

We examine whether the usability of Web pages is correlated with various quantities such as the amount of operations and the movement of the gazing point. We empirically show that the Web pages for which the user's gazing points move quickly tend to have a high degree of usability. A gazing point is the intersection of the user's sight line and the monitor. An expansion of this work will enable even non-experts to use quantitative data such as the moving speed of the gazing point to find the Web pages with low usability and to efficiently evaluate the usability of each Web page.

1   Introduction

The usability of a Web site is so important that it can influence the amount of sales (Goto & Cotler, 2002), because users are unwilling to read Web pages with low-usability, such as the pages hard to operate or understand, or the pages reacting differently from expected (Jakob, 1993). In order to create easy-to-use Web pages, an evaluation of the usability is required. Among various methods for usability evaluation, usability testing (J. S. Dumas & J. C. Redish, 1993) is widely used. One implementation of usability testing is utterance analysis, in which subjects are supposed to tell their impression on the target Web page, as they are reading the page. In another implementation of usability testing, the eye-movement of subjects is used, which has been empirically proved to express the subjects' mental states.

Although usability testing using information on the movement of the gazing point has the advantage that the test procedure does not interrupt the operation of subjects, only a qualitative evaluation is mainly conducted in the testing. Mori et al. (Mori et al., 1995) evaluated usability without interrupting the operation of subjects, by recording the location of the gazing point, and by using the location information for analysis. Sakai et al. (Nakamichi et al., 2003) proposed to support usability testing, by recording the users' operation log including the location of the gazing point and by replaying and visualizing the operation log. Although various types of knowledge are acquired in their work, only qualitative evaluation is conducted in the existing usability testing using information on the movement of the gazing point. Therefore, experts' knowledge and experience is required for conducting the evaluation of usability. However, as the number of Web pages grows, manual evaluation of usability by experts is becoming difficult in terms of time and cost (Okada, 2001).

One way to alleviate this problem is to evaluate only suspicious Web pages; in other words; those which seem to contain a problem, but not all the pages the subject visited. In this method, after hard-to-use Web pages are efficiently collected using the quantitative operation data obtained while subjects are reading Web pages, evaluation is conducted for those collected Web pages. However, no qualitative investigation has been done on the relation between usability and data of the users' gazing point or operation.

In this paper, we examine whether there is correlation between the usability of Web pages and the quantitative data including the users' gazing point or operation. As a result, we show that there is correlation only between the speed of the gazing point and the usability. In addition, we examine which problem usability of the speed of the gazing point is related to. The result shows that the average speed of the gazing point changes when the Web page has one of four usability problems: "The layout of the current page is different from that of the previous page", "the layout is bad", "the anchor text does not describe the content of the linked page", and "the linked text is hard to read".

2   Related Work

2.1   Effectiveness of screen design using a trajectory of the gazing point

Focusing on human interface, Mori et al. (Mori et al., 1995) analyzed the movement of eyeballs and revised the prototype screen in order to improve the effectiveness of prototyping with screen design in the development of an information system. They showed that revising redundant movements of the gazing point improves the operation speed and the users' satisfaction of usability.

Although the quantitative trajectory of eyeball movement is used in the analysis of eyeball movement, only the qualitative analysis by experts of usability is conducted. The trajectory of eyeball movement must be analyzed for the entire screen. Therefore, the effectiveness of the sight line information in the usability evaluation has been proved, but the quantitative relation of the gazing point and the usability has not.

2.2   WebTracer: A tool for an recording operation log including sight line information

WebTracer (Nakamichi et al., 2003) can collect the operation log of users on the Web pages. Collectable data include the information on users' sight line (the coordinates of the gazing point on the computer screen), operation log of a mouse, and the displayed screen images, together with their time information.

The data collected by WebTracer characterize Web pages and have the possibility of being used for supporting the usability evaluation. However, the relation between such data and the problems in the Web usability is merely an example of the characteristics of the Web pages. No quantitative evaluation on the relation to the usability of Web pages has not been done.

3   Experiments

3.1   Outline of experiments

While subjects are conducting previously set tasks, several types of data are recorded including the gazing points of subjects during browsing. WebTracer (Nakamichi et al., 2003) is used for recording the operation log on Web pages. After doing the assigned tasks, the subjects are asked to answer questionnaire on the usability of each visited Web page.

Subjects
15 frequent users of the Internet. They have never visited the sites used in the experiment.

Experimental setting
Display: 21 inches
(viewable screen size: H30 x W40cm)
Resolution: $1,024 \times 768$ pixels
Distance from subject's face to display: approx. 50cm
Device for measurement of sight line: NAC, EMR-NC
(view angle: 0.28, resolution on the screen: approx. 2.4mm)
Recording and playing of sight-line data: WebTracer
(sampling rate: 10 times per second)

Tasks
Subjects are asked to find the following two pieces of information from the site of a school:
Task 1: the prerequisites of a class
Task 2: the phone number and the fax number of the secretary

3.2   Experiments

While subjects are doing the previously set tasks, several types of data are recorded, including the gazing points of subjects during browsing. WebTracer (Nakamichi et al., 2003) is used for recording the operation log on Web pages. After doing the assigned tasks, the subjects are asked to answer questionnaire on the usability of each visited Web page.

Procedure 1: While subjects are doing the tasks, several types of data regarding browsing are recorded.
While subjects are doing the tasks, several types of data regarding browsing operations such as the gazing points of subjects are recorded with WebTracer. We do not take any interruptive measure, such as asking a question to the subjects.
Procedure 2: Subjects evaluate the Web pages.
The Web pages that subjects visited are displayed. Those pages are listed on the operation log recorded in Procedure 1. The subjects are asked to write on the questionnaire (Figure 1) answers to questions regarding the usability of each visited Web page.

4   Analysis on the result

The data pieces, for which the gazing point was not successfully measured, are removed from the entire sample consisting of 195 data pieces. For the remaining data, the relation between the recorded quantitative data and the usability is analyzed. In addition, we investigate what types of usability problems are related to the speed of the gazing point movement, which has been correlated with the usability.

4.1   Removal of deficient data

There are some cases where the gazing point cannot be measured due to unexpected behaviors of subjects such as the eye blinking of subjects, movement of head position outside of the measurable region, and gazing at a point outside of the display. The sampling rate of WebTracer is 10 times per second. If the measurement failed more than 5 out of 10 times, the case was removed from the analysis target. There are 39 such cases. Table 1 shows the number of cases recorded in the experiment, the number of cases of deficient data, and the number of remaining cases, for each degree of usability.

---

1. Impression of usability
Do you think this web page is easy to use? Please choose best one which expresses your impression.
  1.   hard to use
  2.   relatively hard to use
  3.   relatively easy to use
  4.   easy to use
  5.   don't know

2. Please check following boxes which explain the reason why this page is hard to use.
Page Title
( ) Naming of the page title is bad.

Layout
( ) The layout of the current page is different from that of the previous page.
( ) I don't understand the difference between the previous page and the current page.
( ) The layout is bad.
      Please specify the part you feel bad.
      ( )impossible to specify   ( )menu   ( )contents   ( )images

---

Figure 1: Questionnaire (excerpt)

Table 1: The number of cases of each usability

Usability	record cases	deficient cases	Analysis cases
"Hard to use"	24	5	19
"Relatively hard to use"	83	20	63
"Relatively easy to use"	38	7	31
"Easy to use"	29	4	25
"Don't know"	21	3	18
Summary	195	39	156

4.2 Analysis on the relation between quantitative operation data and usability

We investigate the relation between quantitative operation data and usability. Specifically, we split the data cases into two types: the cases with high usability and the cases with low usability. Next, we examine whether, for each type of operation data, there is a difference between the two types of cases. We also investigate the correlation between the values of operation data and usability.

Following the answers in the questionnaire, we classify the cases into two types: the cases with high usability and the cases with low usability. If the subject in a case answers in the questionnaire for a Web page, "easy to use" or "relatively easy to use", then we regard this case as a case with high usability. If the subject answers "hard to use" or "relatively hard to use", then we regard this case as a case with low usability.

For each type of data, we perform a statistical test with the hypothesis that there is a difference between the two classes of cases: the cases with high usability and the cases with low usability. Specifically, we perform an F-test (a test assessing whether the variances of two groups are statistically different from each other) for each type of operation data. If the result of F-test indicates that these variances are different, then we perform a Welch's t-test (a test assessing whether the means of two groups with different variances are statistically different from each other). If the result of the F-test indicates that these variances are not different, then we perform the usual t-test (a test assessing whether the means of the two groups with the same variances are statistically different from each other). Table 2 shows the average values for each type of operation data for the cases with high usability and the cases with low usability, together with the results of the above statistical tests.

The results of the statistical tests in Table 2 show that the mean of each type of operation data for the cases with high usability is statistically different from that for the cases with low usability. However, only the speed of the gazing point has both significantly different variances and means. The distance of gazing point movement has significantly different variances, but not means. The staying duration of the gazing point, the distance of the mouse movement, and the speed of the mouse movement have neither significantly different variance nor means.

We also analyzed the relation between the values of each type of operation data and the Web page usability data collected in the questionnaire to subjects. We calculated a Kendall's rank correlation coefficient for all cases except for the 18 cases where subjects answered, "I don't know" in the questionnaire.

Kendall's rank correlation coefficients in Table 3 show that only the speed of the gazing point is correlated with usability. Since the value of Kendall's rank correlation coefficient is positive, it is shown that faster movement of the gazing point means higher usability and slower movement of the gazing point means lower usability. Thus, we conclude that the usability becomes apparent in the speed of the gazing point among the several types of operation data we analyzed.

Table 2: t-test of user's operational data

Operation data	Average values of High usability	Average values of low usability	t-test
Distance of gaze-point movement (Pixels)	5243	5337	0.4573
Speed of gaze-point movement (Pixels / sec)	421	359	0.0004
Browsing time (sec)	13.1	15.6	0.1491
Distance of mouse movement (Pixels)	1946	2107	0.2972
Speed of mouse movement (Pixels / sec)	186	166	0.1704

Table 3: Correlation between user's operational data and usability

Operation data	Kendall's rank correlation coefficients
Distance of gaze-point movement (Pixels)	-0.025 (negative)
Speed of gaze-point movement (Pixels / sec)	0.205 (positive)
Browsing time (sec)	-0.120 (negative)
Distance of mouse movement (Pixels)	-0.015 (negative)
Speed of mouse movement (Pixels / sec)	0.098 (positive)

4.3 Analysis on the relation between the speed of the gazing point and the problems in Web usability

We analyze the relation between the speed of the gazing point, in which the usability is apparent, and the problems in the Web usability. Specifically, we split the cases into two groups: the cases having problems in Web usability and the cases having no such problems, and then examine whether there is a difference in the speeds of gazing point movement of the two groups.

Following the check items regarding the problems in Web usability in the questionnaire, we split the cases into two groups. Subjects are supposed to check items for the problems in Web usability, if they find some problems while browsing. If an item is checked, then the case is classified as a case having problems in Web usability.

For each problem in Web usability, we perform a statistical test with the hypothesis that there are differences in variances and means of the speeds of the gazing points for two classes of cases: the cases with a problem in Web usability and the cases without a problem in Web usability. Specifically, we perform an F-test (a test assessing whether the variances of two groups are statistically different from each other) with 5% significance level, for each problem in Web usability. If the result of the F-test indicates that these variances are different, then we perform the Welch's t-test (a test assessing whether the means of two groups with different variances are statistically different from each other). If the result of the F-test indicates that these variances are not different, then we perform the usual t-test (a test assessing whether the means of the two groups with the same variances are statistically different from each other). For each problem in Web usability, Table 4 shows average values of the speeds of gazing point movements for the cases with a problem and the cases without a problem, together with the results of the above statistical tests.

The results of the statistical tests in Table 4 show that, for some problems in Web usability, both the mean and the variance of the speed of the gazing point movement for the cases with a problem are statistically different from those without a problem. There are three Web usability problems that produce a significant difference in the statistical test for both variances and means; "The layout of the current page is different from that of the previous page", "the layout is bad", and "the anchor text does not describe the content of the linked page".

Table 4: F-test and t-test of the move speed of the gazing point of each Web usability problem.

The problems in the Web usability (the number of pages pointed out)	Speed of gaze-point movement (Pixels / sec)		the means of two groups with variances (Significant probability P)
	the cases with a problem in Web usability	the cases without a problem in Web usability	
The layout of the current page is different from that of the previous page. (20)	509	370	**0.0003**
The layout is bad. (40)	354	402	**0.0136**
The width of web page height is large. (17)	420	386	0.1584
The anchor text does not describe the content of the linked page(40)	347	403	**0.0033**
The linked text is hard to read. (29)	352	398	**0.0450**
There are a lot of characters. (21)	415	385	0.1129
The character is small. (10)	419	387	0.2333
The character is hard to see. (39)	366	398	0.0978

The Web usability problem, "the linked text is hard to read" has no significant difference in variance, but has a significant difference in means. The other problems "the page is too long in the vertical direction", "there are too many characters", "the characters are too small", and "the characters are hard to read" have no significant difference for variances and means.

5   Discussion

In this section, we discuss the results of the above experiments such that only the speed of the gazing point is correlated with usability, and such that the speed of the gazing point changes if there is a problem in Web usability. Although the distance of the gazing point movement, the duration of stay of the gazing point, and the distance of the mouse movement are larger on average for the cases with low usability, the differences are not statistically significant. The reason why those values are larger for the cases with low usability is presumably because the subjects cannot make a quick decision what to do on those pages and those pages contain too much content. However, when a page contains too much content, the distance of the gazing point movement, the duration of the stay of the gazing point, and the distance of the mouse movement will be large without yielding an impression of low usability. This would be the reason why there was no significant difference for those values.

The speed of the gazing point for each Web page is correlated with the usability, but the speed of the mouse movement is not. When searching Web pages for some information, each subject keeps the gazing point moving. Some subjects put the mouse pointer to their gazing point, while other subjects put the mouse pointer outside of the Web page window. For this reason, no correlation was observed between the speed of mouse movement and usability.

We empirically showed that the average speed of the gazing point significantly changes if the Web page has one of the four following problems: "The layout of the current page is different from that of the previous page", "the layout is bad", "the anchor text does not describe the content of the linked page", and "the linked text is hard to read". When the layout of the current page is different from that of the previous page, the speed of the gazing point becomes high. The reason for this is that the subjects first read through the Web page as if they were visiting the Page for the first time. This problem can be found with heuristic evaluation by experts of usability. This problem can also be easily found with a "GUI tester II", which is a tool for assessing the consistency of a GUI design proposed by Okada et al.

The problems "The layout is bad", "the anchor text does not describe the content of the linked page", and "the linked text is hard to read", often occur when the speed of the gazing point is low. The reason for this could be that subjects have difficulties in finding an appropriate link leading to the target page and keep gazing at anchor texts, since the subjects are confused by a bad layout such as unordered menu, an anchor text that do not look like an anchor text, or an anchor text that does not have an appropriate information on the linked text. The method for finding these problems in Web pages has been to interview subjects. However, the result of this paper shows that we can reduce the cost for assessment by selectively evaluating the Web pages with a low speed of the gazing point We can also predict what types of problems the current Web page has.

We observed that a lower speed of the gazing point indicates low usability. However, this result may come from problems in the anchor texts in the Web pages that we used in the experiments.

The result of analysis shows that we can efficiently perform usability evaluation, by using the speed of the gazing point to find Web pages with low usability, or by using another quantity to find Web pages with a certain type of problem in Web usability.

6   Conclusion and Future Work

In this paper, we statistically showed that the Web pages with a high speed of the gazing point tend to have high usability. We also showed that the average speed of the gazing point significantly changes if the Web page has one of the four problems: "The layout of the current page is different from that of the previous page", "the layout is bad", "the anchor text does not describe the content of the linked page", and "the linked text is hard to read".

We have so far had to assess all the Web pages that subjects visited in the usability testing for Web pages. However, the cost required for assessment is getting larger, and the assessment method must become efficient. If we can efficiently find Web pages with low usability, an efficient assessment will be realized. This paper shows that some quantitative information such as the speed of the gazing point can be used for efficient detection and assessment of Web pages.

More detailed experiments in future research will clarify the relation between the various types of quantitative data and the various issues regarding problems in Web usability. Such future research should enable the quantitative detection of problems.

Acknowledgment

References

Goto, K. & Cotler, E. (2002). Web ReDesign. Peason Education.

Nielsen, J. (1993). Usability Engineering. Academic Press.

Masatoshi Moi, Tetsuo Ui (1995). Research on effectiveness of gazing point movement analysis in screen design, Japan Society for the Study of Office Automation (JSSOA), Vol.16, No.3, 49-56 (In Japanese)

Hidehiko Okada, 2001. Usability and its evaluation technique, The Institute of Systems, Control and Information Engineers, Vol.45 No.5, pp.269-276. (In Japanese)

Noboru Nakamichi, Makoto Sakai, Jian Hu, Kazuyuki Shima, Masahide Nakamura. (2003). WebTracer: Evaluating web usability with browsing history and eye movement, Proc. 10th Int'l Conference on Human - Computer Interaction (HCI International 2003), Vol.1, pp.813-817, Crete, Greece.

Hidehiko Okada, (2001). Usability and its evaluation technique, The Institute of Systems, Control and Information Engineers, Vol.45 No.5, pp.269-276. (In Japanese)