System Usability Evaluation for Input Operation using Oculo-motors

Minoru Nakayama **Makoto Katsukura**

CRADLE(The Center for R & D of Educational Technology)
Tokyo Institute of Technology
Meguro-ku, Tokyo 152-8552
Japan
Email: nakayama@cradle.titech.ac.jp

Abstract

This paper investigates the relationship between oculo-motors, which consist of eye-movement and pupillary change, and the traditional subjective index for "usability", to determine the possibility of evaluating Human-Computer Interaction (HCI). An evaluation experiment was conducted by operating a target on a computer display using input devices: mouse, keyboard and key pad. The results show there is a significant correlated relationship between the pupil size and the SU-score, which is an established subjective evaluation index for system usability. These results provide evidence that pupil size can be used as an index of the system's usability, and also that the SU-score can be estimated from the pupil size. The indices of eye-movement, which consist of saccade frequency, saccade length and saccade time, indicate characteristics of the input operation behavior. These two results suggest that pupil size and index of eye-movement as oculo-motor indices, can provide information about a system's overall usability regarding the input operation task. Additionally, these indices show stability even during short observation periods. This suggests that it is possible to observe temporal changes of system usability.
The results provide evidence that oculo-motors can be an index of system usability.

Keywords: Human-Computer Interaction, Usability, Input device, Oculo-motors, pupil size, eye-movement, blink

1 Introduction

System usability is often discussed as a human-computer interaction (HCI) issue. Computers use various methods for input operations. To test the usability of a system, several evaluation methods have been proposed, such as performance evaluations that measure the time for a task, interviews of users that employ subjective evaluations and questionnaires, and so on (Kurosu 2003). Protocol-analysis, which is the analysis of user's continuous oral reports during designed problem solving, is based on the hypothesis of mental modeling known as systematic analysis. This is also used to discover difficulties of the operation process which concern the HCI issue (Kaiho & Harada 1993). Because they are based on the semantic analysis of each user's behavior, gathering data for these evaluation methods is costly. To simplify the processes, a normalized simple method based on the sum of 10 item scores, has been developed (Brooke 1996). According to previous research work, the major evaluation methods for HCI issues are based on subjective information.

These analyses are affected by the observer's criterion, however. Therefore, a more objective analysis is required. If temporal evaluation was possible in addition to these analyses, it could be applied to various methods of HCI evaluation. To develop objective and temporal evaluation for HCI, the user's physiological information has been used to measure the mental workload during situations where computers are being used (ISO 1991)(Osuga 1992). In particular, both pupil size and eye-movement have also been used to address HCI issues (Takahashi, Nakayama & Shimizu 2000*a*)(Goldberg, Stimson, Lewenstein, Scott & Wichansky 2002).

The eye pupil changes during problem solving (Hess & Polt 1964), so that pupil size can be an index of the task difficulty (Beatty 1982), and the mental workload (Kuhlmann & Boettcher 1999). Eye-movement can also be an index of the task difficulty (Takeda 1976) (Takahashi, Nakayama & Shimizu 2000*b*)(Nakayama, Takahashi & Shimizu 2002)(Nakayama & Shimizu 2004) and the vigilance level (Wright & McGown 2001). Pupil and eye-movement can be measured using equipment set at a distance from the subject, so that subject can perform the operations without distraction.

As a result, pupillary change and eye-movement, as oculo-motors, can be used as an index of the usability of HCI. To examine the possibility of whether the oculo-motors can reflect the usability, an evaluation experiment was conducted. Also, the results may help us understand the oculo-motor mechanism, and find out the possibility of applications with other issues.

In this paper the following topics are to be addressed:

1. To examine the relationship between oculo-motors and the subjective index of HCI.
2. To examine the possibility of using oculo-motors to evaluate the usability of a system.

2 Experimental method

2.1 Experimental task

To measure the system usability using ordinal devices for personal computers (PC), an experiment using an input operational task was designed, as shown in Figure 1. The task is simple: to move a block in a square matrix. Moving the block requires some input operation of the PC, however.

A system, whose purpose is to manipulate a block on a CRT display using input devices, was developed for this experiment. The system uses input operations which are typical for the devices, such as selecting menu options by making key depressions. The task illustrated in Figure 1 is performed on a two-dimensional 11×11 cell grid.

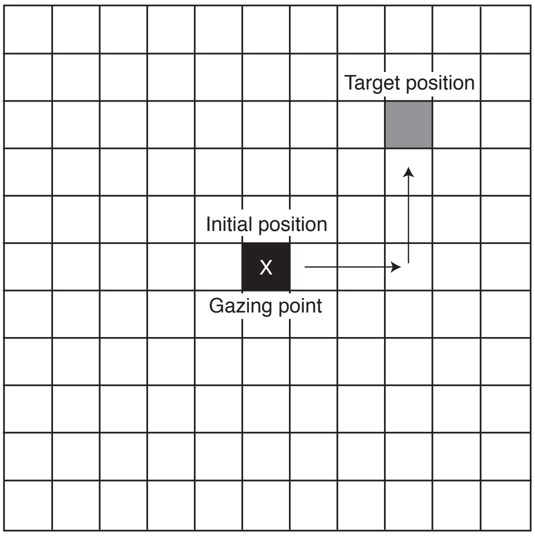

Figure 1: Input operation task.

Figure 2: A numeric key set.

The problem solving task is as follows. First, a subject is asked to gaze at the center position of a screen, marked with an 'x', and then asked to move the block the 'x' is in from the initial position to a target position as quickly as possible, once the targeted position is displayed. Because the distance to move the 'x' is fixed at a 6 city-block distance, the task used requires some operation of each input device. There are 20 possible positions which are 6 city-blocks distant from the initial position. For each trial, a target was assigned randomly from the set of 20 positions.

Each of the 3 input devices required a different operating procedure to move the block.

1. Mouse: A mouse is given to the subject. To move the block to an assigned position, the subject is asked to position the mouse above the target location and to leftclick the mouse button.
2. Keyboard (KeyBD): A numeric key set (10 key pad) is used to move the block. The keys for moving the block are placed on a three by two square, with the "4" and "6" controlling horizontal movement and "5" and "8" controlling vertical movement, instead of using the arrow keys. During the task, depression of an assigned key moves the block in the corresponding direction. Many systems use this method, or one similar to it.
3. Key pad (KeyPAD): The remote controller of a DVD player was modified as a pointing device, with a wired connection to the PC, and was used to move the block in the same way as the numeric keys did. Most remote controllers are built with many buttons, and four of these were used to control block movement. The button assignment was as follows: Play = upwards, Stop = downwards, Rewind = to the left, and Fast Forward = to the right. These button locations are the same as for numeric keys on a keyboard.

Figure 3: Key pad as a remote controller.

To complete the task as soon as possible, the task execution time limit was set at three lengths: 1.0, 1.5, and 2.0 seconds. If the time limit was too short, all trials would fail. Also, the time limit depended on the input device. As a result, three time limit lengths were set in accordance with the results of a preparation experiment.

2.2 Experimental procedure

The experimental conditions, which are based on the input devices, are Mouse, Keyboard (KeyBD), Key-pad (KeyPAD), and Control (Ctrl), which does not require any operation by the subjects, except their gazing at the target position on a CRT display.

In each condition, 20 successive trials were conducted, and the duration for each trial was 5 seconds which consisted of the task execution time and interval, though it was not dependent on the task execution time. Audible feedback information, which notifies a subject as to whether task completion was a success or failure, is given immediately after the end of the task execution time. The sequence of experimental tasks was randomized to avoid the effects of sequences. A short time interval was given to subjects to refresh themselves between experiments.

The subjects were 6 university students from the Faculty of Engineering. They have normal visual acuity and some computer operation skills, and they took part in all of the experiments.

2.3 "Usability" evaluation

The usability of the three input devices, Mouse, Keyboard (KeyBD) and Key pad (KeyPAD), was evaluated using a subjective method. All three are ordinal input devices. Most people use a mouse and keyboard for daily PC operation. Also, key pads are always used as remote controllers for TVs or with cell phones.

(1)	I think that I would like to use this system frequently	Strongly disagree	1 2 3 4 5	Strongly agree
(2)	I found the system unnecessarily complex	Strongly disagree	1 2 3 4 5	Strongly agree
(3)	I thought the system was easy to use	Strongly disagree	1 2 3 4 5	Strongly agree
(4)	I think that I would need the support of a technical person to be able to use this system	Strongly disagree	1 2 3 4 5	Strongly agree
(5)	I found the various functions in this system were well integrated	Strongly disagree	1 2 3 4 5	Strongly agree
(6)	I thought there was too much inconsistency in this system	Strongly disagree	1 2 3 4 5	Strongly agree
(7)	I would imagine that most people would learn to use this system very quickly	Strongly disagree	1 2 3 4 5	Strongly agree
(8)	I found the system very cumbersome to use	Strongly disagree	1 2 3 4 5	Strongly agree
(9)	I felt very confident using the system	Strongly disagree	1 2 3 4 5	Strongly agree
(10)	I needed to learn a lot of things before I could get going with this system	Strongly disagree	1 2 3 4 5	Strongly agree

Figure 4: The system usability scale questionnaire items [Brooke 1996].

John Brook has established a measuring procedure using a simple questionnaire, which is often selected for evaluation of human computer interactions (Brooke 1996). This is defined as the system usability scale (SUS). The SUS is derived from responses to the questions which are listed in Figure 4.

The SUS is designed as an attitude scale consisting of users degree of preference in response to 10 questionnaire items, similar to the Likart scale with 5 levels. A "Strongly disagree" response gives 1 point, and a "Strongly agree" response gives 5 points. For Japanese subjects, a Japanese version, which was translated by the authors from the original SUS questionnaire, was used. All subjects could respond easily the translated items.

The SUS is derived using the following procedure, according to the original method of calculation (Brooke 1996). Each item's score is changed into a numeral ranging from 0 to 4.

- items 1,3,5,7 and 9: the converted value equals the response value minus 1.
- items 2,4,6,8 and 10: the converted value equals 5 minus the response value.
- multiply the sum of the converted values by 2.5, so that the SUS has a range of 0 to 100.

It is suggested that the easier a user feels it is to operate a system, the higher the SU-score becomes. All subjects are required to answer a questionnaire after their session with each input device.

2.4 Error rate

Failure of a task trial, wherein a subject can not reach the target or selects a position which is not the target, is defined as an error. The causes may be that a subject has made an operational mistake, or there was not enough time for the subject to operate the input device. According to usability testing, operation times and error rates are often used as indices for performance evaluations (Kurosu 2003). Because the times for the trials are regularized in this experiment, the error rate is a unique behavioral measure for each task execution condition.

Here, the error rate is defined as the percentage of errors per 20 trials for each subject in each experimental condition. The error rates were summarized by each input device used in the experiment. In general, the rate is higher when the operation of the input device is more difficult.

2.5 Oculo-motor measure

The experimental environment is shown in Figure 5. The stimuli is displayed on a 17 *inch* CRT monitor positioned 50 *cm* from the subject. The oculo-motors, which consist of pupil size, blink and eye-movement, are measured by a video-based eye-tracker (nac:EMR-NL8). The subject is seated and does not wear or put any measuring equipments on, and rests his or her head on a chin rest during the experiment. The hands are always free, so that the subject is not restricted during the task (Nakayama et al. 2002).

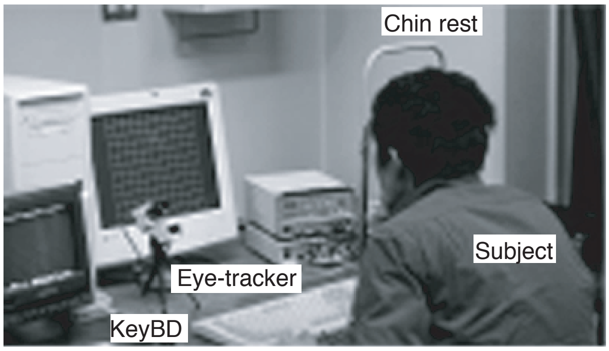

Figure 5: Experimental setup.

2.5.1 Measuring pupil size

To capture left eye images, an small infra-red camera is positioned between the subject and the monitors, 40 *cm* from the subject. An eye-tracker detects pupil shape as a ellipse, then measures the long and short diameters at 60 *Hz*. Eye-movements are also measured using the pupil shape. The eye-tracker also produces a status code, which is based on the ratio of the short diameter to the longer one. When the ratio is lower than 0.7–0.8, the status code becomes an error code. In most cases, the error means an eye blink, suggesting that the system can detect blinks.

The blinks are defined as a series of error codes. The pupil sizes decrease during a blink drop off, because the pupillary images are obscured by the eyelid. To resolve this issue, an estimation method has been developed (Nakayama & Shimizu 2004). We use a simple interpolation method which replaces the dropped pupil size with regular data from before the status code indicated an error. Although it is suggested that blink drops affect the results of frequency analysis of pupillography, the average pupil sizes for the condition do not depend on the estimation of pupil size during blink (Nakayama & Shimizu 2004). The blink rate can be extracted from the experimental measurements of the pupil size.

The pupil sizes are calculated from the diameter measurements, and are standardized for each subject, since pupil sizes of all individuals are different.

2.5.2 Measuring eye-movement

Eye-movements are measured as the coordinates of a CRT screen size (640×480), and then converted into visual angle degree values for the horizontal axis and the vertical axis, according to each experimental condition. The sampling rate is 60 *Hz*.

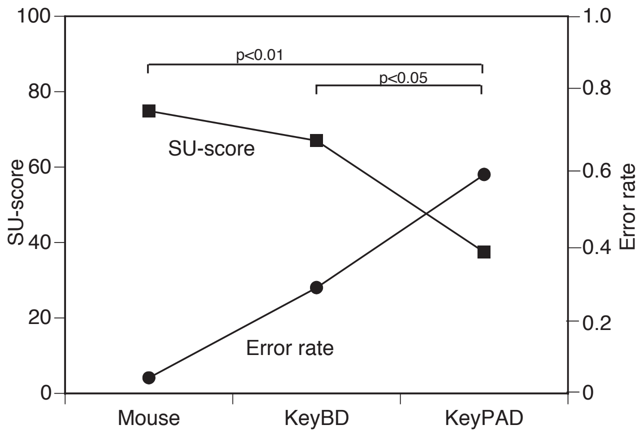

Figure 6: SU-scores and error rates for the input devices

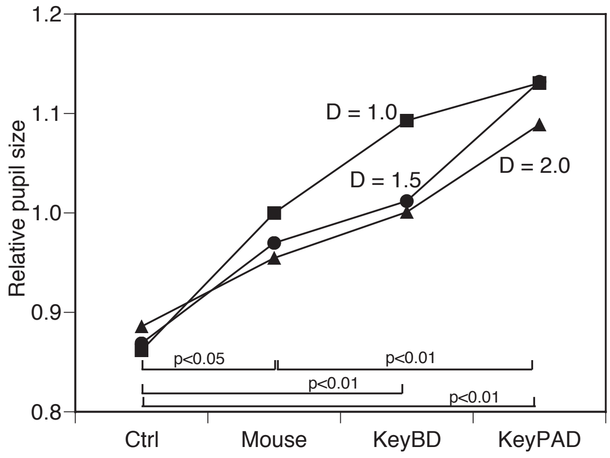

Figure 7: Relative pupil sizes for the input devices

Eye-movements were divided into gazes and saccades, using the following procedure (Takahashi et al. 2000*b*). Saccade is the ballistic movement of the eye, typically taking only about 30 *msec*. (Palmer 1999). Eye-movement is detected by temporal changes in the tracking position, and the speed is evaluated as a differential angle for each 60 *Hz* frame. To extract saccades from eye-movements, saccades are defined as the eye-movement velocity above the threshold. It is supposed that a saccade is eye-movement velocity over 40 deg/sec. (Ebisawa & Sugiura 1998), and the rest are gazes. Saccades are detected as components which have a velocity over the threshold between the frames. Here, the series of saccades are defined as a single saccade. The saccade frequency is evaluated as an occurrence rate of saccades. Each saccade length is computed by differential summation of the x–y coordinates of the saccade. Also, the incidences of saccades and the duration of saccades are summed up as well. As blinking prevents the detection of eye-movement, the data for eye-movement during blink is replaced with data from before the status code of the measurement becomes an error.

As a result, saccade frequency per second, saccade length in degree, and the saccade duration rate are extracted. Here, the saccade time rate is defined as the rate of duration of saccades. The phenomenon that visual blurring is not perceived during saccades is well known as "saccade suppression" (Palmer 1999). Therefore, the saccade time rate suggests there is a degree of unviewable time.

3 Results

3.1 SU-score and error rate

The SU-scores for each input device were measured. Means of the SU-scores are summarized in Figure 6. The horizontal axis shows the input devices, and the vertical axis at the left side shows the SU-score. According to the figure, Mouse scored the highest and KeyPAD the lowest.

The error rate, which is defined in section 2.4, is also illustrated in the same format in Figure 6. The vertical axis on the right side shows the error rate. The rates depend on the task execution time limit, but it is not easy to remove influence of the task execution time limit factor. Therefore, a rate is calculated for each input device. When combined, the error rate for KeyPAD is the highest, and the rate for Mouse is the lowest. This suggests that there is a significant relationship between the two indices. The negative correlation coefficient was calculated ($r = -0.86$, $p < 0.01$).

To examine the differences of the indices amongst input devices, a within-participant multiple comparison test was conducted. From the results, there are significant differences for both SU-scores and error rates between Mouse and KeyPAD ($p < 0.01$), and between KeyBD and KeyPAD ($p < 0.05$).

These results seem quite reasonable. The Mouse, which can be pointed directly at the target, is easier to operate, however the KeyPAD operation is not as easy because the KeyPAD keys must be depressed with a finger, such as in the operation of a cell phone. The results of both indices reasonably reflect their performance features.

3.2 Pupil size

3.2.1 Mean pupil size

To examine the possibility of evaluating system usability by pupil size, mean pupil sizes were compared amongst the four experimental conditions. The means are calculated from all trials, including the error responses. Mean pupil sizes are illustrated in Figure 7. The horizontal axis shows the four experimental conditions, the vertical axis shows the relative pupil size. The means are summarized by the task execution time limit (D).

According to the figure, all pupil sizes for the three operational conditions are larger than the ones for the Control condition. The order of the pupil sizes corresponds to the results of the SU-scores and error rates. Here, the sources of variance are the input devices and the task execution time limit. To determine the significance of the factors (the input devices and the time limit), a subject repeated two-way ANOVA was conducted. In the results, the factor of the input devices is significant ($F(3,15) = 10.9, p < 0.01$), but the factor of the time limit is not significant ($F(2,40) = 2.4, p = 0.11$). According to these results, the factor of the task execution time limit was pooled with the error term, and it was neglected in the following analysis. In a multiple comparison between the input devices, there was a significant difference between Control and KeyBD or KeyPAD ($p < 0.01$), and also between Mouse and Control ($p < 0.05$).

From both SU-score and Error-rate, we can find out that there is a difference in system usability between KeyBD and KeyPAD, but there is no significant difference in pupil size between the two devices during a test. This lack of a significant difference comes from the statistical error term issue. Therefore, if the same analysis were conducted on the data when the Ctrl condition was omitted, there would be a significant difference between KeyBD and KeyPAD ($p < 0.05$).

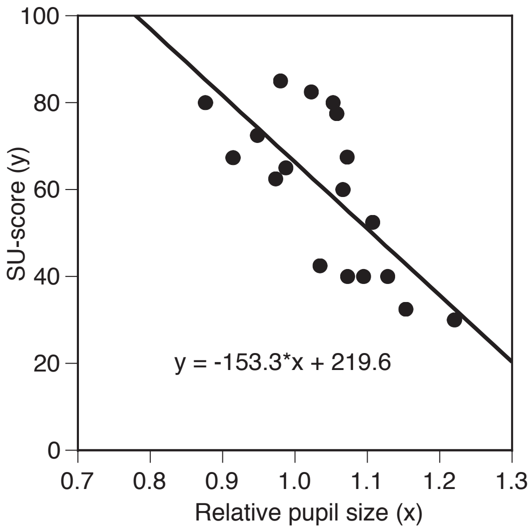

Figure 8: The relationship between pupil size and SU-score

The mean pupil sizes, which were calculated from correct responses, show the same results. It has been suggested that the pupil size reflects the task difficulty (Beatty 1982), and mental workload (Kuhlmann & Boettcher 1999). Therefore, the mean pupil size shows that the operational task difficulty is affected by the input devices.

3.2.2 Relationship between SU-score and pupil size

The SU-score is defined as the degree of usability, and pupil size reflects the task difficulty. These two measurements are definitely based on different concepts, because of different definitions. If these two indices measure a similar aspect of the operation, then they may have a correlation to each other. On the other hand, it is interesting to examine the relationship between the SU-score as a subjective evaluation and the pupil size as a physiological objective measure. The relationship of the SU-score and the average pupil size among the input devices is summarized in Figure 8. Figure 8 shows a scatter gram between the pupil size and the SU-score. The horizontal axis shows the pupil size, and the vertical axis shows the SU-score.

To examine the relationship between them, the correlation coefficient was extracted. There is a significant negative correlation ($r = -0.72$, $p < 0.001$). The relationship can be presented mathematically, because the correlation coefficient is so high. The regression equation is extracted and is shown as follows:

$$[the\ \ SU - score] = -153.3 * ([pupil\ \ size]) + 219.6$$

Here, y is a value of $[the\ \ SU - score]$, x is a value of $[pupil\ \ size]$. Using regression analysis, both parameters are significant ($p < 0.01$). This equation is illustrated as a bold line in the figure. The line is shown as summarizing both distributions.

This result suggests that the SU-score can be estimated from the pupil size, as a regression.

3.3 Saccade frequency as eye-movement

To examine the change in eye-movement while the tasks are undertaken, saccade frequency is measured. According to the definition of a saccade, the number of saccades per second is summarized as the saccade frequency. The frequency according to input device is indicated as bars

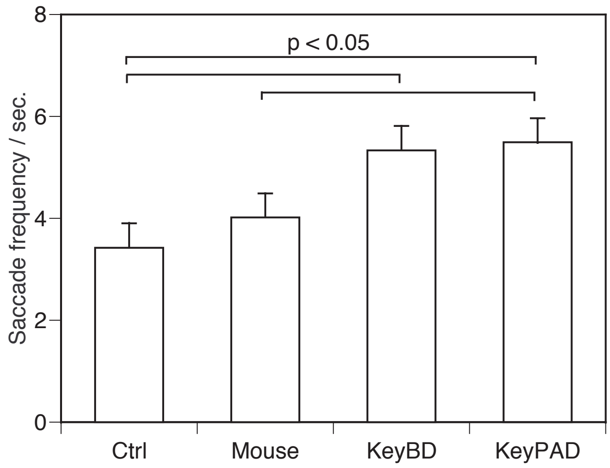

Figure 9: Saccade frequency according to input device

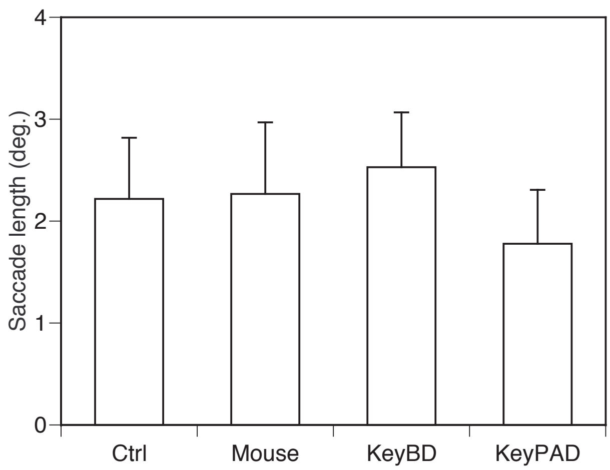

Figure 10: Saccade length according to input device

with standard errors, in Figure 9. The horizontal axis shows the input device, and the vertical axis shows the saccade frequency in average number of saccades per second. Because a single saccade takes only about 150 – 200 *msec.* to plan and execute (Palmer 1999), the maximum number of saccades is 5 – 6.7 per second.

As Figure 9 shows, the frequency increases according to the input device, from Ctrl to Mouse to KeyBD to KeyPAD ($F(3,15) = 4.5, p < 0.05$). This suggests that many saccades happened during the operation task. There are significant differences between Ctrl and both KeyBD and KeyPAD ($p < 0.05$). Also, there is a significant difference between Mouse and KeyPAD.

According to the data for SU-scores and pupil sizes, the usability of KeyPAD is not higher. The data for the saccade frequency shows a similar tendency. However, there is no difference between KeyBD and KeyPAD.

This may reflect the characteristics of eye-movement. Because subjects have to look at their hands to operate these devices, eye-movement between the target on the screen and their hands may appear.

3.4 Saccade length

To examine the characteristics of the saccade, the average saccade lengths are summarized according to the input device. Here, the saccade length is represented as a visual angle in degrees. It is not a saccade speed. The results are summarized as bars with standard errors, in Figure 10.

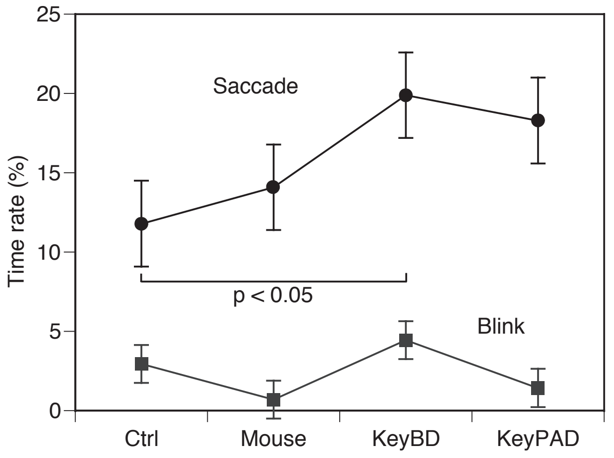

Figure 11: Time ratio of saccade and blink

The horizontal axis is the input device, the vertical axis is the saccade length in degrees.

According to the results, the saccade length for KeyBD is longer than others and the saccade length for KeyPAD is shorter than others. There is no significant difference amongst the four conditions, however ($F(3,15) = 0.3, p = 0.81$).

In the KeyBD condition, subjects might have looked at the target and the keyboard buttons, therefore the saccade length was longer. On the other hand, if subjects might have gazed downwards to operate the key pad buttons, then the saccade was restricted to a narrow area, in KeyPAD condition. Because subjects can hold the KeyPAD in their hand, the average saccade length shows a different tendency than with a KeyBD.

According to the above analysis, both saccade frequency and saccade length show features of the input operation. Therefore, a difference of saccade behavior may provide information about the characteristics of input operation.

3.5 Saccade and blink time ratio

Eye-movements are divided into saccade and gaze, and observation during the saccade is suppressed, as "saccade suppression". We can get visual information from the point that is being viewed during the gaze, however.

The eye-blink also prevents observation, because the eyelids cover the eyes. As a result, viewing time is the remainder of the saccade and blink. Here, the blink also shows the vigilance level, or the task difficulty (Tada, Yamada & Fukuda 1991). To examine the viewing time, the saccade and blink time rates are summarized for each of the conditions, in Figure 11. The horizontal axis shows the input devices, the vertical axis shows the time rate.

The solid square symbols and standard error bars indicate the time rate of blink. All rates of blink are less than 5%. This suggests that the blinks are suppressed because of the short task time. In particular, the blink is suppressed for the Mouse and KeyPAD input devices. The blink rate for KeyBD is relatively higher than for the others. There is no significant difference amongst the conditions, however.

The solid circle symbols and standard error bars indicate the time rate of saccades. The rates of saccade for KeyBD and KeyPAD are higher than for the others, as shown in Figure 9. There is a significant difference between KeyBD and Ctrl ($p < 0.05$).

The total of unviewable time is the highest for the KeyBD condition. Also, there is a significant difference between KeyBD and Ctrl ($p < 0.05$). The total of the unviewable

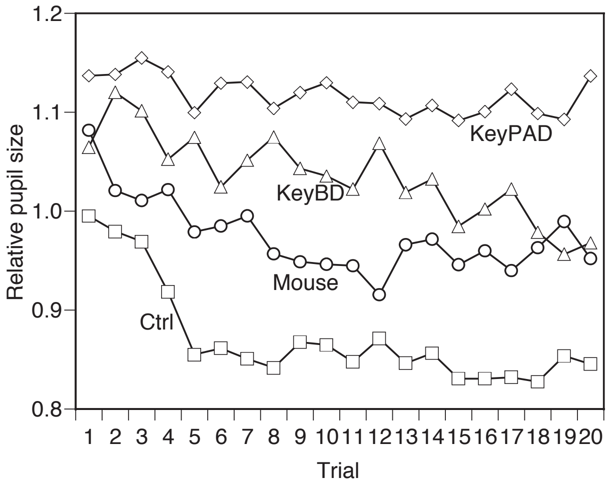

Figure 12: Temporal pupillary changes according to input device

time for the KeyPAD is shorter than that for the KeyBD, although the SU-score for the KeyPAD is the lowest.

For operations using the KeyBD, users had to look at two disparate areas, the target on the display and the location of their hands. The saccade time increased, and the blink time increased somewhat, also. As a result, the total unviewable time for the KeyBD was the highest.

It seems that this result reflects the operational characteristics of the input devices.

3.6 Relationship between eye-movement and usability

According to the above results, some indices reflect the characteristics of the input operation task. Therefore there may be a relationship between indices and SU-scores. To examine the relationship between them, the correlation coefficients were extracted. The coefficients are relatively small and vary between -0.04 and -0.24 ($p > 0.10$). There is no significant relationship in the results.

The results are different in the case of pupil size. The pupil size increases with the task difficulty of the input devices, however the indices of eye-movement depend on the characteristics of the input operation, rather than the task difficulty. This suggests that indices of both pupil size and eye-movement help to understand the total usability of the system.

4 Discussion

4.1 The relationship between trials and pupil size

The above results suggest that it is possible to measure usability and to observe operational characteristics, using oculo-motors. However, these results were based on the total average across the trials. If the pupil size order amongst the experimental conditions does not depend on the trials, those measures can be used for observations in the trial sequence, such as time series analysis.

To examine the changes in the trial sequence, pupil sizes for each trial in the four experimental conditions are shown in Figure 12. The horizontal axis shows the trial sequence, the vertical axis shows the relative pupil size. The order of pupil size is maintained in each trial, and pupil size increases in the order: Ctrl, Mouse, KeyBD and KeyPAD. There are small differences in pupil size amongst

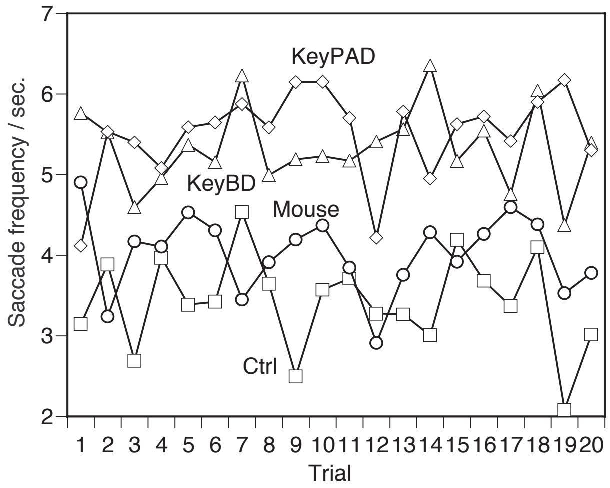

Figure 13: Temporal saccade frequency according to input device

trials for Ctrl and Mouse. The series of the change of KeyBD shows the subject's adaptation. This indicates that the usability of KeyBD is increasing in during the trials, but pupil sizes for other devices do not change during the trials.

These results provide evidence that pupil size reflects the task difficulty of the input operation for each trial. According to the relationship between SU-scores and pupil sizes, the SU-scores for each trial can be estimated. This suggests the possibility of obtaining the SU-score temporally.

4.2 The relationship between trials and saccade frequency

During the trial, the series of pupil sizes, as well as the average, change as the task difficulty changes. An index of eye-movements may offer similar changes during the trials. These suggest that some indices can be extracted from eye-movement. Here, the saccade frequency is examined as an index of eye-movement to observe the operational characteristics. The saccade frequencies for each trial in the four experimental conditions are shown in Figure 13. The horizontal axis shows the number of trials, the vertical axis shows the saccade frequency in degrees.

The saccade frequency changes in accordance with the trials. The pattern in Figure 13 tends to follow the pattern in Figure 9. The saccade frequencies for KeyBD and KeyPAD are higher than the ones for Ctrl. However, it seems that there are no significant differences between KeyBD and KeyPAD, or between Mouse and Ctrl. The variances of saccade frequency and other eye-movement indices are not small, therefore the results depend on the trials, to a small extent.

Comparing the changes in pupil size and saccade frequency for each trial, the results of pupil size measurements can provide more stable information as to system usability. However, eye-movement information may show operational characteristics, therefore the saccade frequency during the trials may show the change of characteristics for each trial, despite the large variance. Again, it is suggested that eye-movement observations can obtain the temporal changes of the characteristics of operation for each task.

5 Conclusion

This paper investigates the relationship between oculo-motors and the traditional subjective index for "usability", to determine the possibility of evaluating Human-Computer Interaction (HCI). An evaluation experiment was conducted by operating a target on a computer display using input devices: Mouse, KeyBD and KeyPAD. The results indicate the following:

Firstly, there is a significant correlated relationship between pupil size and the SU-score. SU-scores are an established subjective evaluation procedure for system usability. These results provide evidence that pupil size can be used as an index of system usability, and also that SU-scores can be estimated from pupil size.

Secondly, the indices of eye-movement, which consist of saccade frequency, saccade length and saccade time, show the characteristics of the input operation behavior.

Thirdly, both oculo-motors of pupil size and indices of eye movement can provide the information about the system usability of an input operation task. Additionally, these indices can be applied to time course analysis.

The results provide evidence that oculo-motors can be an index of system usability.

To examine the possibility of various evaluations for Human Computer Interaction issues, such as the usability of web page design work, will be a subject of further study.

References

Beatty, J. (1982), 'Task-evoked pupillary resposes, processing load, and the structure of processing resources', *Phychological Bulletin* **91**(2), 276–292.

Brooke, J. (1996), Sus: A quick and dirty usability scale, *in* P. Jordan, B. Thomas, B. Weerdmeester & I. McClelland, eds, 'Usability Evaluation in Industry', Taylor & Francis, London, UK.

Ebisawa, Y. & Sugiura, M. (1998), 'Influences of target and fixation point conditions on characteristics of visually guided voluntary saccade', *The Journal of the Institute of Image Information and Television Engineers* **52**(11), 1730–1737.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N. & Wichansky, A. M. (2002), Eye tracking in web search tasks: Design implications, *in* R. Vertegaal & J. W. Senders, eds, 'Eye Tracking Research and Applications Symposium 2002', ACM, ACM Press, New York, USA, pp. 51–58.

Hess, E. H. & Polt, J. M. (1964), 'Pupil size in relation to mental activity during simple problem solving', *Science* (132), 1190–1192.

ISO (1991), *Ergonomic principles related to mental work–load —General terms and definitions*, International Organization for Standardization.

Kaiho, H. & Harada, E. (1993), *Introduction to protocol analysis (in Japanese)*, Shinyosya, Tokyo, Japan.

Kuhlmann, J. & Boettcher, M., eds (1999), *Pupillography: Principles, Methods and Applications*, W. Zuckschwerdt Verlag, Muenchen, Germany.

Kurosu, M. (2003), *Usability testing (in Japanese)*, Kyoritsu-shuppan, Tokyo, Japan.

Nakayama, M. & Shimizu, Y. (2004), Frequency analysis of task evoked pupillary response and eye-movement, *in* A. T. Duchowski & R. Vertegaal, eds, ‘Eye Tracking Research and Applications Symposium 2002’, ACM, ACM Press, New York, USA, pp. 71–76.

Nakayama, M., Takahashi, K. & Shimizu, Y. (2002), The act of task difficulty and eye-movement frequency for the ’oculo-motor indices’, *in* R. Vertegaal & J. W. Senders, eds, ‘Eye Tracking Research and Applications Symposium 2002’, ACM, ACM Press, New York, USA, pp. 37–42.

Osuga, M. (1992), ‘Evaluation of mental workload using physiological responses’, *Journal of the Society of Instrument and Control Engineers* **31**(2), 324–325.

Palmer, S. E. (1999), *Vision Science –Photons to Phenomenology*, The MIT Press, Cambridge, Massachusetts, USA.

Tada, H., Yamada, F. & Fukuda, K. (1991), *Psychological blink (in Japanese)*, Kita-Ouji Shobo, Kyoto, Japan.

Takahashi, K., Nakayama, M. & Shimizu, Y. (2000*a*), ‘Eye-movements and pupillary changes during simulated driving’, *The Journal of the Institute of Image Information and Television Engineers* **54**(9), 1323–1329.

Takahashi, K., Nakayama, M. & Shimizu, Y. (2000*b*), The response of eye-movement and pupil size to audio instruction while viewing a moving target, *in* A. T. Duchowski, ed., ‘Eye Tracking Research and Applications Symposium 2000’, ACM, ACM Press, New York, USA, pp. 131–138.

Takeda, M. (1976), ‘Effect of mental activity in problem solving on eye movements’, *The Japanese Journal of Ergonomics* **12**(5), 175–181.

Wright, N. & McGown, A. (2001), ‘Vigilance on the civil flight deck: incidece of sleepiness and sleep during long-haul flights and associated changes in physiological parameters’, *Ergonomics* **44**(1), 82–106.