Toward Constructing an Electroencephalogram Measurement Method for Usability Evaluation

Masaki Kimura*, Hidetake Uwano, Masao Ohira, and Ken-ichi Matsumoto

Graduate School of Information Science, Nara Institute of Science and Technology
{masaki-k,hideta-u,masao,matumoto}@is.nasit.jp

Abstract. This paper describes our pilot study toward constructing an electro-encephalogram (EEG) measurement method for usability evaluation. The measurement method consists of two steps: (1) measuring EEGs of subjects for several tens of seconds after events or tasks that are targets to evaluate, and (2) analyzing how much components of the alpha and/or beta rhythm are contained in the measured EEGs. However, there only exists an empirical rule on meas-urement time length of EEGs for usability evaluation. In this paper, we conduct an experiment to reveal the optimal time length of EEGs for usability evalua-tion by analyzing changes of EEGs over time. From the results of the experi-ments, we have found that the time length suitable for usability evaluation was more than 0~56.32 seconds.

1 Introduction

Existing usability evaluation methods include interview, think-aloud protocols [1], and questionnaires [2, 3, 4]. These methods are widely used for usability evaluation because they require no measurement apparatuses and allow usability experts to measure usability of software systems in a relatively easy way. However, they also have some shortcomings for usability evaluation. For instance, usability evaluation using the above methods often requires a huge amount of time to analyze and evaluate collected data. Also, the results of the analysis are sometimes hard to replicate be-cause the collected data is based on qualitative and/or subjective evaluations from system users or subjects participating in usability testing. In order to complement the limitations of these methods, many studies try to quantitatively and objectively de-velop evaluation methods for measuring the mental or psychological state of users from biological information.

In this paper, we focus on the electroencephalogram (EEG) of users after using software as a quantitative measurement method of usability and aim to con-struct an EEG measurement method for usability evaluation. Alpha rhythm and beta rhythm, which are the frequency component of EEG, change according to mental condition [4], and EEG measurement methods have features that do not disturb sub-jects using a computer. In general, the EEG measurement method is used as follows:

---

* Graduate School of Information Science, Nara Institute of Science and Technology 8916-5,Takayama, Ikoma, Nara, Japan.

(1) experimenters (or usability experts) measure a subject's EEG for several tens of seconds after the subject finishes target tasks for usability evaluation, and (2) experimenters analyze how much the measured EEG contains the alpha and beta rhythm components, which respectively indicate a comfortable or uncomfortable state of the subject's mind. The results of the evaluation should change according to the time length to perform the EEG. After the tasks, subjects' EEGs return to the usual condition as time passes. Hence, the time length of the analysis is too long, and the ratio of the EEG changed by the tasks decreases. Conversely, if the time length is too short, evaluation accuracy decreases because the ratio of the noise to the entire the EEG increases. However, the measurement time at EEG analysis has been decided by experiential standards. So we need to analyze EEGs by time series for usability evaluation. In this paper, we try to obtain the proper time length of EEG data.

2   Related Research

In this paper, we quantitatively evaluate a psychological state of computer users during a system in-use, using alpha and beta rhythms composed of frequency components of brain waves. Power spectrums of alpha and beta rhythms, which are obtained by discrete Fourier transform, the ratio of alpha and beta rhythms in all brain waves, and beta/alpha, which is the ratio of alpha and beta rhythms, are often used as common indicators for observing the psychological state of human beings. Matsunaga et al. developed a brain wave measurement system for evaluating satisfaction of human beings and validated the hypothesis that people feel comfortable if the amount of information processing in the brain is small, while people feel uncomfortable if the amount of information processing in the brain is large [5]. In this paper, we also use the ratio of alpha and beta rhythms in brain waves and beta/alpha as indicators. These indicators are often used for studies using brain waves. So our experimental results are easy to compare with implications and insights from previous work.

3   Experiment

3.1   Overview

Using Microsoft Excel 2003 and Excel 2007, which are the most popular spreadsheet software, participants performed eight kinds of tasks operating spreadsheets, and we measured participants' EEGs after each task. Excel 2003 and Excel 2007 are different versions of software released in 2003 and 2007, respectively. Both versions have almost the same functions, but have a different *look and feel* in their graphical user interfaces (GUIs). Excel 2007 has a new GUI called "ribbon," which is designed to improve task performance and *user experience*. However, the newly designed ribbon interface introduced users to a lot of changes to menus, tool bars, and working windows. So, even if users would like to use a familiar function such as "Save As," they need to select a menu or button with different names and/or positions between the two versions. Furthermore, even if names and positions of menus are not changed, the design of working windows displayed after selecting a menu is often different from Excel 2003. In this way, not only new users but also existing users of Excel 2003 need to learn how to operate the new interface of Excel 2007.

Table 1. Subjects' usage frequency of Excel 2003 and Excel 2007

	Excel 2003	Excel 2007
never	0	6
several times per year	2	1
several times per month	3	2
several times per week	5	1

Using Excel 2003 and Excel 2007, we investigate the relationship between the experiences of software and the attributes of brain waves without the effect of functional differences. In our experiment, we also analyzed the relationship between the results of subjective evaluation by questionnaire and the attributes of brain waves.

3.2 Participants

Ten master's students from the graduate school of information science participated in the experiment. Table 1 shows participants' usage frequency of Excel 2003 and Excel 2007. All participants had experience in using Excel 2003 and understood basic operations and functions, but half of the participants had never used Excel 2007.

3.3 Task

Participants performed eight tasks (four types of tasks for each version of Excel) operating the spreadsheets given in advance. Table 2 shows a list of tasks used in the experiment. All tasks can perform both Excel 2003 and Excel 2007. The content (a grade report) of the data file that is used in this experiment is the same in all tasks. Participants can continue the tasks until the task time exceeds five minutes. We counterbalanced the order of the tasks to minimize learning and fatigue effects. The following are the details of each task.

Same Place Task
In this task, the participant selects a particular menu that has the same name in the same position in different versions of Excel. The task is completed when the participant selects the objective menu.

Different Place Task
The participant selects a menu that has a different name in a different position in the two versions of Excel. This task is completed when the participant selects the objective menu.

Same Interface Task
The Same Interface Task consists of functions that have the same interfaces of modal dialog in different Excel versions. In the task, menu name and position were given to the participant before the task.

Different Interface Task
In this task, functions that have different dialog interfaces in Excel 2003 and Excel 2007 were used. As with the Same Interface Task, menu name and position were given to the participant before the task.

Table 2. A task list used in the experiment

Task type	Task name	Description
Same place	Open Clip Art Pane	Open clip art pane to select clip art from a list.
	Filter Setting	Set options for data filtering.
Different place	Display of version information	Display the version information of Excel.
	Record of macros	Change date formats from Mar-01 to 03/01.
Same interface	Format Cells	Change a page orientation to landscape and set margins.
	Page Orientation	Change a page orientation to landscape and set margins.
Different interface	Conditional Formatting	Indicate cells that have less than "C" or "Absence" as red font.
	Insert Bar Chart	Insert stacked bar chart of student's scores with chart/axis titles.

3.4 Environment

The Emotional Spectrum Analysis System ESA-16 was employed to record EEGs of participants. After the task, we recorded participants' EEGs for two minutes at 200Hz sampling frequency in the eyes-closed, resting condition. Electrode locations are based on the International 10-20 System, shown in Figure 1. We adapted referential derivation to observe the EEG, and used the right earlobe (A2) as a reference electrode. As a ground electrode, the center of the forehead (Fpz) was employed and the center of the parietal (Pz) was used as an exploring electrode to minimize the electromyogram (EMG) artifact. We also recorded electrocardiogram (ECG) from both arms. In addition to this, we used a headrest and elastic net bandage to secure electrodes placed on the head. Before the first task, each participant adjusted the height of chair and position of the mouse/keyboard.

3.5 Procedure

The procedure of the experiment was as follows.

1. **Preparation:** Authors informed participant about experiment and EEG measurement.
2. **Environment setting:** Put the electrodes on the participant at the points described in Section 3.4, then set up the EEG analyzer.
3. **Practice tasks:** The participant performed two practice tasks to understand the procedure of EEG measurement. These tasks were excluded from analysis.
4. **Perform a task:** The participant performed a main task described in Table 2.
5. **EEG measurement:** After each task, the EEG of the participant was measured.
6. **Perform eight tasks:** The participant performed tasks repeatedly until finishing eight tasks and EEG measurements.
7. **Questionnaire:** After the tasks, the participant filled out the questionnaire described in Section 3.6.

3.6 Questionnaire

After the eight tasks, participants answered the questionnaire sheet to investigate subjective satisfactions for each version of Excel and usage frequency of each function that was used in the tasks. The questionnaire was created by the authors based on the Questionnaire for User Interaction Satisfaction (QUIS). Each question about usage frequency consists of a four-point scale (from "Never" to "Few times per week") and seven-point scale (from "Strongly disagree" to "Strongly agree") for subjective satisfaction. Figure 2 shows a part of questionnaire sheet that used in the experiment.
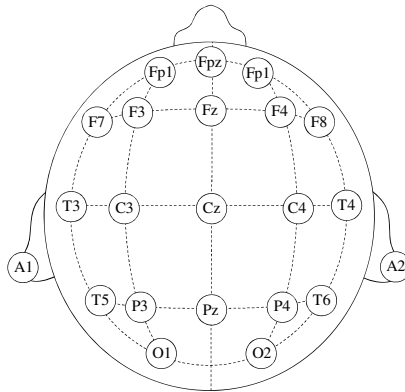
Fig. 1. Electrode Locations in the International 10-20 System
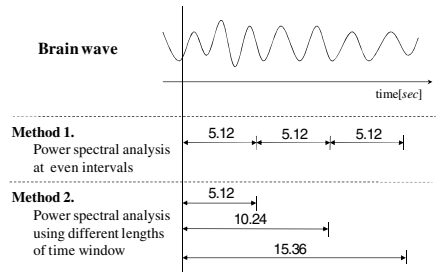

Fig. 2. Questionnaire Sheet

Fig. 3. Two Kinds of Analysis Methods for Electroencephalogram

4 Analysis for EEG

We applied power spectral analysis to EEG data we collected at a sampling frequency of 200Hz. To have a clear understanding of how frequency components of brain waves changed over time in the setting of our experiment and how analysis results varied according to lengths of the analysis window, we used two analysis methods for the EEG data as follows. Figure 3 illustrates the difference between the analysis methods.

Method 1. Power spectral analysis at even intervals

This analysis aims to observe how brain waves change over time. We analyzed EEG data with a time interval of 5.12 seconds by cutting out the entire EEG data every 5.12 seconds so as not to overlap analysis data. The EEG data with 19 intervals (from 0 to 92.16 second) was analyzed.

Method 2. Power spectral analysis using different lengths of time window

This analysis aims to observe how analysis results differ according to lengths of analysis window. We analyzed EEG data by increasing time length of the analysis window by 5.12 seconds, without changing the start position of our analysis. The time length was increased by 5.12 seconds (min.) to 97.28 seconds (max.) (i.e., 0~5.12 sec., 0~10.24 sec , …, and 0~97.28 sec.).

Next, the target data was filtered to reduce the artifacts from eye blinking, myoelectric activity and so on. We used a high-pass filter (HPF, 3Hz cutoff frequency, +6*dB/oct* attenuation factor), a low-pass filter (LPF, 60Hz cutoff frequency, -6*dB/oct* attenuation factor), and a band-elimination filter (BEF, 60Hz central frequency, 47.5Hz~72.5Hz stopband, second order). The band-elimination filter was used to remove the influence of an alternating-current power supply. After the EEG data was multiplied by the Hamming window and processed with the fast Fourier transform (FFT), we obtained the power spectrum from the EEG data. From the obtained power spectrum, we calculated the respective proportions of alpha rhythm and beta rhythm to all brain waves, and also calculated beta/alpha, which divided the proportion of alpha rhythm into the proportion of beta rhythm. In accordance with the classification of the international 10-20 system, we set the frequency components of alpha rhythm and beta rhythm to 8~13Hz and 13~30Hz respectively. We also set the range of all brain waves to 3~30Hz. Since the proportions of alpha and beta rhythms to all brain waves are widely used for observing various activities in the brain, we also decided to use them as indexes for measuring the physiological state of subjects after the tasks.

However, because the proportions and intensity of alpha and beta rhythms vary from individual to individual, comparisons of brain waves with the absolute value would be inappropriate. In this paper, we normalized EEG data by an average value of each subject's power spectrum and compared it with each data.

5   Results

5.1   Results of the Power Spectral Analysis at Even Intervals

Figures 4, 5, and 6 respectively show the mean and the standard deviation of the alpha rhythm, beta rhythm and beta/alpha in the power spectral analysis at even intervals. In each graph, the left x-axis, the right x-axis, and the y-axis are the mean, standard deviation, and time respectively.

Figure 4 indicates that the mean of alpha rhythms for Excel 2003 were larger than that for Excel 2007 after 56.32 seconds and that the difference of the alpha rhythms between Excel 2003 and 2007 was largest at 81.92 seconds. The standard deviation was always comparatively small and lowest at 56.32 seconds. Figure 5 shows the mean of the beta rhythms for Excel 2007 was larger than that for Excel 2003 after 46.08 seconds, and the difference of the beta rhythms between Excel 2003 and 2007

was greatest at 87.04 seconds. The standard deviation was higher on the whole than the standard deviation of alpha rhythms. The lowest standard deviation was observed at 40.96 seconds. Figure 6 presents that the mean of beta/alpha in Excel 2007 was larger than that in Excel 2003 after 46.08 seconds and that the difference of beta/alpha between Excel 2003 and 2007 was at 81.92 seconds. The standard deviation is larger than the results of alpha rhythms and beta rhythms and smallest at 40.96 seconds.

5.2 Results of the Power Spectral Analysis Using Different Length of Time Window

Figures 7, 8, and 9 respectively show the mean and the standard deviation of the alpha rhythm, beta rhythm and beta/alpha in the power spectral analysis using different lengths of time window. In each graph, the left x-axis, the right x-axis, and the y-axis are the mean, standard deviation, and time respectively.
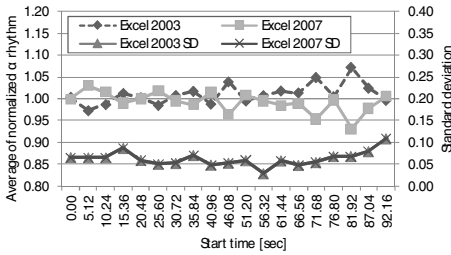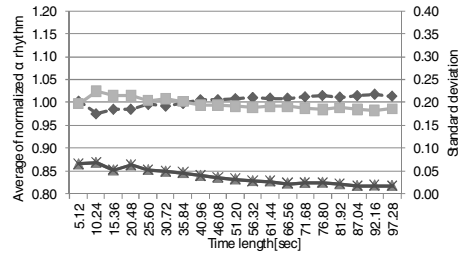

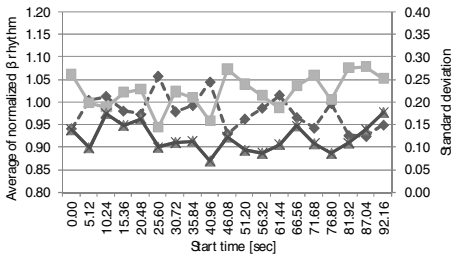**Fig. 4.** Alpha Rhythm in Method 1

Fig. 7. Alpha Rhythm in Method 2

Fig. 5. Beta Rhythm in Method 1
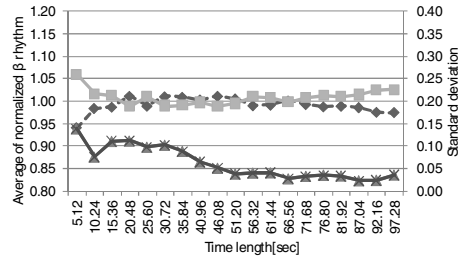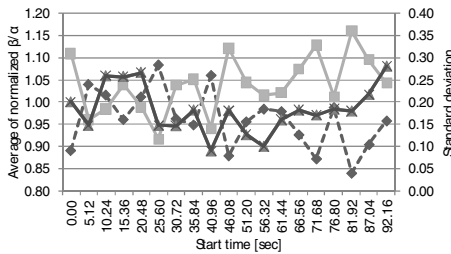

Fig. 8. Beta Rhythm in Method 2

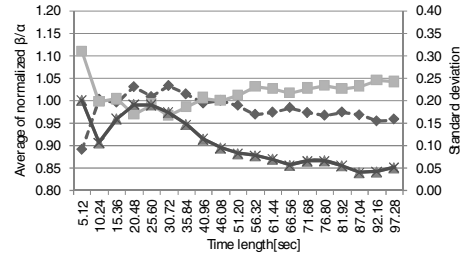**Fig. 6.** Beta/Alpha in Method 1

Fig. 9. Beta/Alpha in Method 2

Figure 7 shows that the mean of alpha rhythms in Excel 2003 was larger than that in Excel 2007 in the case of using time windows over 40.96 seconds. As lengths of time window became longer, the alpha rhythms in Excel 2003 tended to increase and the standard deviation decreased. Figure 8 presents that the mean of the beta rhythms in Excel 2007 was larger than that in Excel 2003 in the case of using time windows over 56.32 seconds. As lengths of time window became longer, the standard deviation tended to decrease as well as the results of the alpha rhythms. Figure 9 indicates that the mean of beta/alpha in Excel 2007 was larger than that in Excel 2003 in the case of using time windows over 40.96 seconds. As lengths of time window became longer, the standard deviation tended to decrease as well as the results of the alpha and beta rhythms.

Table 3. A Result of Questionnaire

		Usage frequency	Understand	Productivity	Simple to use	Interface	Easy to use	Satisfaction
Excel 2003	Average	3.3	5.0	5.4	4.8	4.9	5.3	5.0
	SD	0.82	1.33	1.26	1.32	1.45	1.34	1.15
Excel 2007	Average	1.8	3.5	3.5	4.0	3.1	3.4	3.3
	SD	1.14	2.17	1.72	1.63	2.13	1.90	1.95
	$p < 0.05$	**yes**	no	**yes**	no	**yes**	**yes**	**yes**

5.3   Results of Questionnaire

Table 3 shows the mean, standard deviation and the result of the two-sample t-test of each questionnaire item. In the table, there were significant differences in "Productivity," "Interface," "Easy to use" and "Satisfaction" between Excel 2003 and Excel 2007. Our subjects gave Excel 2007 lower scores than Excel 2003.

6   Discussions

From the results of the power spectral analysis at even intervals, we could confirm that all three indexes tended to be relatively stable after 56.32 seconds. In fact, the alpha rhythms in Excel 2003 were larger than those in Excel 2007. The beta rhythms and beta/alpha in Excel 2007 were larger those in Excel 2003. The standard deviations of all three indexes also indicated lowest or relatively lower from 40.96 seconds to 56.32 seconds. These results would imply that EEG data from 56.32 seconds to 61.44 seconds is stable among all subjects, is less influenced by artifacts, and appropriately reflects the influence of the difference between Excel 2003 and 2007.

However, for all indexes, the difference of the influence between two versions of Excel and the standard deviations constantly fluctuated. We consider that this might occur due to individual differences of subjects' brain waves, fatigue from the long experiment duration, myogenic potential by attitude variation, and so forth. Therefore, it is necessary not to use a time window with short length, but to use one with long length, in order to conduct accurate usability evaluation.

From the results of the power spectral analysis using different lengths of time window, we could observe that the alpha rhythms in Excel 2003 were stably higher in the case of using time windows over 40.96 seconds, the beta rhythms in Excel 2007 were stably higher in the case of using time windows over 56.32 seconds, and beta/alpha in Excel 2007 was stably higher in the case of using time windows over 40.96 seconds. For all indexes, the standard deviations tended to become small as the time window became longer. This might be because the rate of EEG influenced by artifacts became small as the time window became longer. These results suggest us that we could analyze EEG data with no influence of artifacts by using a time window over 56.32 seconds.

The results of our analysis showed that the alpha rhythms in Excel 2003 were larger than those in Excel 2007, and the beta rhythms and beta/alpha in Excel 2007 were larger than those in Excel 2003. It is clarified by previous studies on EEG measurement methods that the amount of alpha rhythms is decreased and the amount of beta rhythms and the value of beta/alpha are increased when a subject's mental work load is high. The results of the questionnaire showed our subjects preferred Excel 2007 to Excel 2003. These results of questionnaire and our analysis agree with previous work. Since the results of our analysis and questionnaire also supported the results of the past studies on EEG, we can conclude EEG measurement would be useful for evaluating software usability.

7 Conclusion and Future Work

In this paper, we have conducted an experiment to have a clear understanding of the appropriate timing and lengths of time window in order to analyze EEG data for accurate usability evaluation. From the results of the experiments, we could obtain the following insights.

- A short length of time window (e.g. 5.12 seconds) is not suitable for usability evaluation because the frequency components of brain waves constantly fluctuate.
- The accuracy of usability evaluation can be improved by using length of time windows of 56.32 seconds.

In this experiment, we could not observe normal state of EEG data but EEG influenced by the tasks. In the near future, we need to conduct another experiment to observe it.

References

1. Ericsson, K.A., Simon, H.A.: Protocol analysis: Verbal reports as data. MIT Press, Cambridge (1993)
2. Osgood, C.E., Suci, G.J., Tannenbaum, P.H.: The measurement of meaning. University of Illinois Press, Urbana (1957)
3. Chin, J.P., Norman, K.L., Shneiderman, B.: Subjective user evaluation of CF PASCAL programming tools. Technical Report (CAR-TR-304) (1987)

4. Hart, S.G., StaveLand, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Handcock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. Elsevier, Amsterdam (1988)
5. Matsunaga, H., Nakazawa, H.: A Study on Human-Oriented Manufacturing System (HOMS) – Development of Satisfaction Measurement System (SMS) and Evaluation of Element Technologies of HOMS using SMS. In: Int. Conf. Manufacturing Milestones toward 21st Century, pp. 217–222 (1997)