A Quantitative Evaluation on the Software Use Experience with Electroencephalogram

Hitoshi Masaki[1], Masao Ohira[1], Hidetake Uwano[2], and Ken-ichi Matsumoto[1]

[1] Graduate School of Information Science, Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara, Japan
[2] Department of Information Engineering, Nara National College of Technology
22, Yata, Yamatokoriyama, Nara, Japan
{hitoshi-m,masao,matumoto}@is.naist.jp,
uwano@info.nara-k.ac.jp

Abstract. In usability testing, experimenters need to perform a pre-training, so as to control software-use experiences of subjects. The pre-training in usability testing is very important because subjects' software-use experiences have a large effect on a result of a subjective evaluation of software. This paper aims to evaluate the software-use experiences quantitatively using EEG. We have conducted experiments to observe the relationships between subjects' software-use experiences and EEG in using software. As a result, we found that there was a significant difference between them.

Keywords: EEG, Use Experience, Quantitative Evaluation, Usability Testing.

1 Introduction

In recent years, software systems have become much more functional but complicated. Users need usable interfaces to use software systems effectively. In order to develop such the interfaces, evaluating usability has been very important. Existing methods for usability evaluation include interviews, think-aloud protocols [1], questionnaires [2], and so on. These usability evaluation methods extract usability problems from subjective evaluations by users (subjects). They are widely used because they require no special measurement apparatuses and allow usability experts to know usability of software in a relatively easy way if software is completed.

However, we have to carefully consider that the differences of software-use experiences among users have a large effect on a result of subjective evaluation by users. In general, the mental workload of users in using a software system becomes large if the users do not sufficiently learn how to use the software. As a result, users sometimes have negative impressions on the software. Therefore, before usability evaluation, users are trained to learn usage of software in order to make the differences of software-use experiences among users constant. Even so, it is still difficult to adjust the mental workload and/or software-use experiences among users at a certain level since a capacity for learning software is large different among the users.

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 469–477, 2011.

Toward constructing a method to reduce the effect of software-user experiences on a result of usability evaluation, in this paper we would like to see whether we can measure software-use experiences quantitatively. As an approach to quantitative measurement of the software-use experiences, we use electroencephalogram (EEG) after users used software. In general, components of the alpha rhythm and the beta rhythm in EEG change by mental activities such as mental calculations, state of tension, condition of excitation, and so on. EEG measurement has some advantages as compared with other bioinstrumentations for the human central nervous system: EEG can be measured by using relatively cheap instruments [3], it imposes fewer limitations in measurement, it does not disturb subjects in using computers, and so forth.

2 Related Work

Measurable information to observe biological activities in the central nervous system includes EEG, magnetoencephalogram (MEG), functional magnetic resonance imaging (fMRI) and so on. In this paper, we use EEG as an indicator to measure the software-user experiences quantitatively. EEG is sometimes used to measure a psychological state of human. Xiaowei et al. [4] set up a learning system on a web site, and recorded EEG of learners visiting the web site. They found that the power of alpha rhythm of users who is concentrating in tasks is lower than that of user who is not distracted. Schier [5] recorded brain waves during a driving simulation task, using a driving simulator. The results showed that the power of alpha rhythm increased when the attention level of drivers decreased. Matsunaga et al. [6] developed a brain wave measurement system for evaluating one's satisfaction 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.

We also measure the power of alpha rhythm and beta rhythm for observing the physiological state of subjects. So our experimental results are easy to compare with the implications and insights from previous work.

3 Experiment

3.1 Overview

To investigate the relationship between software-use experiences and EEG, we have conducted two experiments where subjects used Microsoft Excel 2007. In one experiment, we recruited a group of ten subjects in 2007. Most of the subjects were inexperienced in using Excel 2007 because Excel 2007 had been a brand-new product. Most of them were still familiar with older version, Excel 2003. In another experiment in 2010, we also recruited a group of ten subjects who often used Excel 2007. All the subjects in our experiments were recruited from the graduate school of information science, Nara institute of science and technology. Table 1 shows subjects' usage frequency of Excel 2007. Each subject performed four kinds of tasks (eight tasks in total) described later. We measured EEG of each subject after s/he completed each task. We also asked the subjects to fill in a questionnaire on usability of Excel 2007.

Table 1. Subjects' usage frequency of Excel 2007

Usage Frequency	Experiment in 2007	Experiment in 2010
never	6	0
several times per year	1	2
several times per month	2	3
several times per week	1	5

3.2 Task

Table 2 shows a list of tasks used in the experiments. All the tasks can be performed on both Excel 2003 and Excel 2007, but we expected the group in 2007 confused with the tasks because the user interfaces were very different between Excel 2003 and Excel 2007. Under this experimental setting, the group of the subjects in 2010 should obviously have more positive impressions on Excel 2007 than that of the group in 2007 if software-use experiences have an effect on impressions of software usability.

We gave each subject a data file to make a grade report, which was used in all the tasks. Each subject could spend five minutes to complete a task. We counterbalanced the order of the tasks to minimize learning and remembering effects. The following are the details of each task.

Same Place Task. A subject selects a command with a same name and same position between the two versions of Excel. This task is completed when the subject selects the command.

Different Place Task. A subject selects a command with a same name and different position between the two versions of Excel. This task is completed when the subject selects the command.

Table 2. Task list used in experiment

Task Type	Task Name	Description
Same Place	Open Clip Art Pain	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 Macro	Run a recorder of macro.
Same Interface	Format Cells	Change date formats from Mar-01 to 03/01.
	Page Orientation	Change a page orientation to landscape and set margins.
Different Interface	Conditional Formatting	Indicate cells that have less than 60 or "Absence" as red font.
	Insert Bar Chart	Insert stacked bar chart of student's scores with chart/axis titles.

Same Interface Task. A subject uses a dialog box with a same composition between the two versions of Excel. In this task, menu name and position were given to the subject before the task. This task is completed when the subject achieved the given operations.

Different Interface Task. A subject uses a dialog box with a different composition between the two versions of Excel. As well As the Same Interface Task, menu name and position were given to the subject before the task. This task is completed when the subject achieved the given operations.

3.3 Environment

Emotional Spectrum Analysis System ESA-16 was employed to record EEG of subjects. After the task, we recorded subjects' EEGs for two minutes at 200Hz sampling frequency in eye-closing, 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 right earlobe (A2) as reference electrode. As ground electrode, center of the forehead (Fpz) was employed and center of the parietal (Pz) was used as exploring electrode to minimize electromyogram (EMG) artifact. We also recorded electrocardiogram (ECG) from both arms. To reduce artifacts in the measurement, we used headrest and elastic net bandage to fix electrode placed on the head. Before the first task, each subject adjusts height of chair and position of mouse/keyboard.

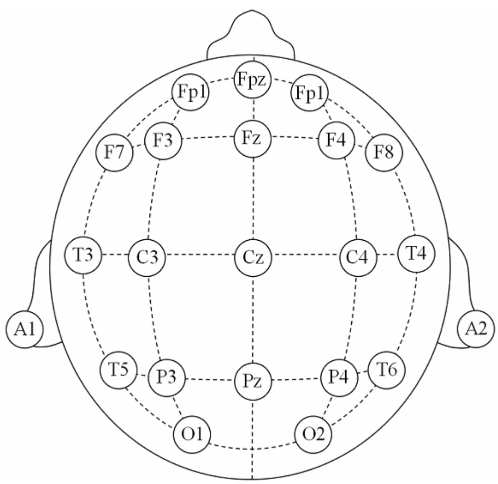

Fig. 1. Electrode locations in International 10-20 System

3.4 Questionnaire

After the all tasks, subjects were answered a questionnaire sheet for investigating subjective satisfactions of Excel 2007 and the 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 the usage frequency consists of four-point scale (from "Never" to "Few times per week") and seven-point scale (from "Strongly disagree" to "Strongly agree") items for evaluating subjective satisfactions.

4 Analysis for EEG Data

We applied power spectral analysis to EEG data we collected at a sampling frequency of 200Hz. Firstly we cut out each subject's EEG data for 81.96 seconds in eye-closing and resting condition after each task. 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 classification of the international 10-20 system, we set 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.

The proportions of alpha rhythm and beta rhythm to the entire brain waves are often used for observing various activities in the brain [7, 8]. We also use the beta/alpha as indicators for measuring the physiological state of subjects after the tasks. Since the proportions and intensity of alpha rhythm and beta rhythm vary from individual to individual, comparisons of brain waves with the absolute value would be inappropriate. In this paper, we normalized the EEG data from each subject by an average value of each subject's power spectrum and compared it with the each EEG data.

5 Results

5.1 Correlation between EEG and Questionnaire Result

Table 3 shows correlations between each indicator of EEG and each questionnaire item. This table shows that there were significant correlations between each indicator of EEG and each questionnaire item.

5.2 Relationship between EEG and Usage Frequency

Figure 2 shows the median of beta/alpha by each usage frequency. The vertical axis is the normalized power spectrum of the EEG data. If the value is higher than 1.0, it means the mental workload of the subjects is large. In contrast, if the value is lower than 1.0, it means that the mental workload is small. As shown in Figure 2, the median of beta/alpha of the subjects who often use Excel 2007 (several times per year/month/week) was lower than 1.0. The median of beta/alpha of the subjects who never used Excel 2007 was higher than 1.0. The correlation coefficient between the values of beta/alpha and the usage frequency was -0.599 and there was a significant difference between them (pearson's r, p<0.01).

Table 3. Correlation between EEG and questionnaire result

		Simple to Use	Productivity	Clarity of Information	Pleasant Interface	Easy to Use	Degree of Satisfaction
beta/alpha	pearson's r	-0.616	-0.654	-0.502	-0.672	-0.613	-0.645
	p-value	0.004	0.002	0.024	0.001	0.004	0.002

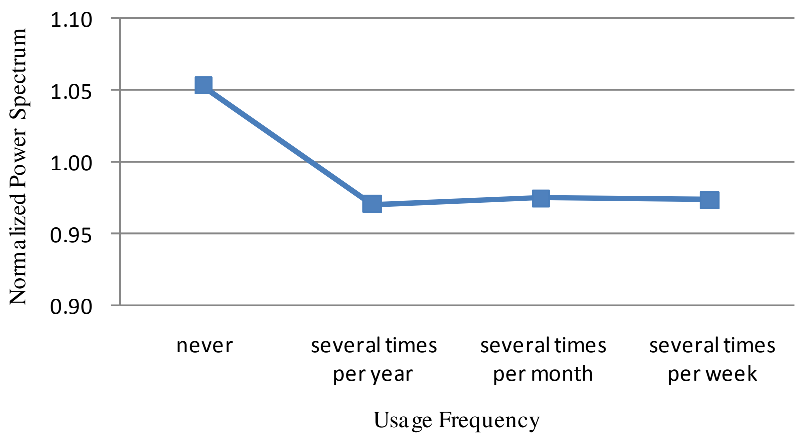

Fig. 2. Normalized value of beta/alpha and usage frequency

5.3 Differences of EEG by Tasks

Same Place/Interface Task. In the experiment in 2007, we hypothesized that the median of beta/alpha would be lower than 1.0 because knowledge and experiences of using Excel 2003 could be applicable to perform Same Place Task and Same Interface Task of Excel 2007 which used menus and dialog boxes as same as Excel 2003. As a result, beta/alpha in Same Place Task was lowers than 1.0. However, beta/alpha in Same Interface Task was over 1.0, contrary to our expectation.

In the experiment in 2010, we hypothesized again that the median of beta/alpha in both Same Interface Task and Same Place Task would be lower than 1.0 because the subjects in the experiment in 2010 were well accustomed to using Excel 2007. Comparing the results in 2007 and the results in 2010, we would like to show the reason why our hypothesis was not supported in 2007.

Figure 3 and Figure 4 show comparisons of beta/alpha in the two experiments. In Same Place Task, it is not difference between the median of beta/alpha in the two experiments. In Same Interface Task, beta/alpha in the second experiment was lower than the beta/alpha of the first experiment, and there was significant difference in beta/alpha between the two experiments (t-test, p=0.012). This indicates software-use experiences have an effect on beta/alpha.

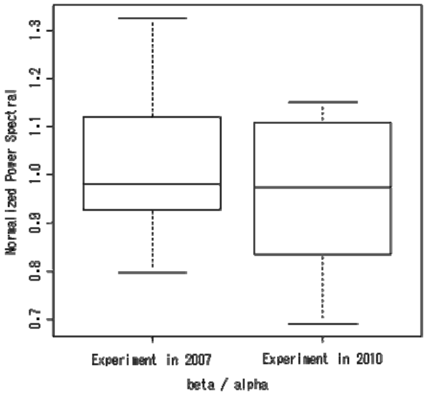

Fig. 3. Difference of beta/alpha at Same Place Tasks

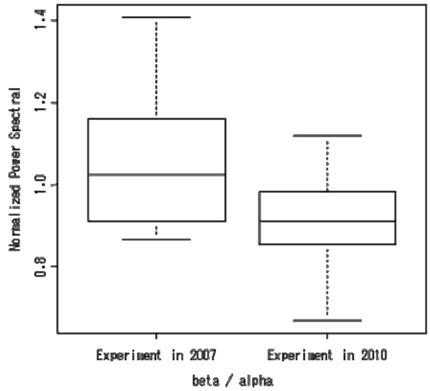

Fig. 4. Difference of beta/alpha at Same Interface Tasks

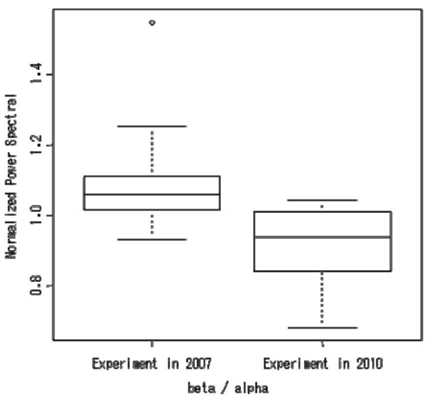

Fig. 5. Difference of beta/alpha at Different Place Tasks

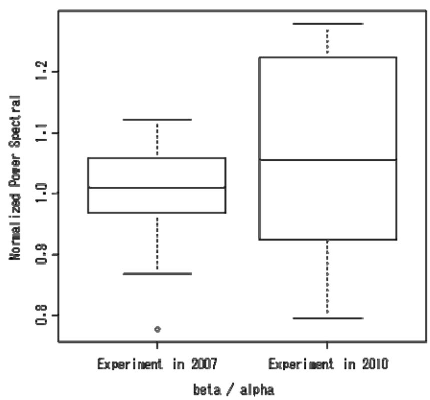

Fig. 6. Difference of beta/alpha at Different Interface Tasks

Different Place/Interface Task. In the experiment in 2007, we hypothesized that the median of beta/alpha would be higher than 1.0 because knowledge and experiences of using Excel 2003 could not be applicable to perform Different Place Task and Different Interface Task of Excel 2007 which used different menus and dialog boxes from Excel 2003. As a result, beta/alpha in both Different Place Task and Different Interface Task was higher than 1.0 as we expected.

In the experiment in 2010, we hypothesized that the median of beta/alpha would be lower than 1.0 because of the same reason in Same Place/Interface Task.

Figure 5 and Figure 6 show comparisons of beta/alpha in the two experiments. In Different Place Task, beta/alpha of the experiment in 2010 was lower than that of the experiment in 2007, and there was a significant difference in beta/alpha between the two experiments (t-test, p=0.041). In Different Interface Task, the beta/alpha of the two experiment were higher than 1.0.

6 Discussion

From the results of the two experiments, we found there is a significant correlation between EEG and subjective evaluation of Excel 2007. The experienced subjects seemed to complete the tasks easily and rated Excel 2007 higher than the inexperienced subjects. Thus, the software-use experiences have a large effect on the subjective evaluation. We also found that there was a significant correlation between beta/alpha and use experience of Excel 2007. The median of beta/alpha of the experienced subjects was under 1.0. In other hand, that of the inexperienced subject was over 1.0. From these results, we conclude that the software-use experiences can be measured by analyzing beta/alpha of subjects.

We would like to discuss the reasons that beta/alpha of Same Interface Task and the Different Interface Task was different from our hypothesis. In Same Interface Task, there was a significant difference in beta/alpha between the two experiments. However, there were not significant differences in the each questionnaire item between the two experiments. Although the usage frequency of Excel 2007 was high, the subjective evaluation of Excel 2007 did not change. We thought that the time to open a dialog box has an effect on EEG. In Different Interface Task, beta/alpha of the experiment in 2010 was higher than 1.0 as well as the experiment in 2007, and there was no significant difference in beta/alpha between the two experiments. However, there were significant differences in the questionnaire items of “Productivity” and “Pleasant Interface” between the two experiments. Furthermore, the all questionnaire results of the second experiments were higher than the all questionnaire results of the first experiments, but there were not significant differences in the questionnaire results of “Simple to Use”, “Clarity of Information”, “Easy to Use” and “Degree of Satisfaction” between the two experiments. From these questionnaire results, we consider that the subjects in 2010 have been accustomed to using the interface in Different Interface Task. We also consider that the interface in Different Dialog Task had a problem on interactivity and it resulted in over 1.0 beta/alpha value.

7 Conclusion and Future Work

In this paper, we had conducted the two experiments to investigate whether we can measure software-use experiences quantitatively or not. As an approach to quantitative measurement of the software-use experiences, we used electroencephalogram (EEG) and analyzed the relationships between EEG and software-use experiences. As a result of the analysis, we found that beta/alpha of experienced subjects become lower than 1.0 while that of inexperienced subjects exceed 1.0. From the findings, we conclude that beta/alpha can be an indicator for measuring the software-use experiences quantitatively.

In our future work, we would like to clarify how long do users spend to accostom use of software, comparing short time use at regular intervals with continuous, long time use. We also would like to show EEG measurement is effective to evaluate *learnability* of software.

References

1. Ericsson, K.A., Simon, H.A.: Protocol analysis: Verbal reports as data. MIT Press, Cambridge (1993)
2. Chin, J.P., Norman, K.L., Shneiderman, B.: Subjective user evaluation of CF PASCAL programming tools. Technical Report (1987)
3. Lee, J.C., Tan, D.S.: Using a Low-Cost Electroencephalograph for Task Classification in HCI Research. In: Proceedings of the 19th ACM Symposium on User Interface Software and Technology, pp. 81–90 (2006)
4. Xiaowei, L., Bin, H., Tingshao, Z., Jingzhi, Y., Fang, Z.: Towards Affective Learning with an EEG Feedback Approach. In: Proceedings of the First ACM International Workshop on Multimedia Technologies for Distance Learning, pp. 33–38 (2008)
5. Schier, M.A.: Changes in EEG alpha power during simulated driving: a demonstration. International Journal of Psychophysiology 37, 155–162 (2000)
6. 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. Conference on Manufacturing Milestones Toward the 21st Century, pp. 217–222 (1997)
7. Hjelm, S.I., Browall, C.: Brainball - using brain activity for cool competition. In: Proceedings of NordiCHI 2000 (2000)
8. Goo, J.J., Park, K.S., Lee, M., Park, J., Hahn, M., Ahn, H., Picard, R.W.: Effects of Guided and Unguided Style Learning on User Attention in a Virtual Environment. In: Pan, Z., Aylett, R.S., Diener, H., Jin, X., Göbel, S., Li, L. (eds.) Edutainment 2006. LNCS, vol. 3942, pp. 1208–1222. Springer, Heidelberg (2006)