NAIST-IS-DT0161040

Doctor's Thesis

Software Usability Evaluation Based on Quantitative Data of Brain Wave, Gaze Point, and First Impression

Jian Hu

February 1, 2004

Department of Information Systems
Graduate School of Information Science
Nara Institute of Science and Technology

Doctor's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Engineering

Jian Hu

Thesis committee: Ken-ichi Matsumoto, Professor
Masaki Koyama, Professor
Katsuro Inoue, Professor
Hajimu Iida, Associate Professor

Software Usability Evaluation Based on Quantitative Data of Brain Wave, Gaze Point, and First Impression*

Jian Hu

Abstract

This thesis proposes three methods for software usability evaluation respectively based on quantitative data of brain wave, gaze point, and first impression. The purpose is to improve evaluation efficiency and effectiveness of software usability by solving problems in existing three types of evaluation methods: operation evaluation, performance evaluation, and subjective operation. These three types of methods are complementary to each other and can be used separately or together according to evaluation purpose of software usability. The problems in existing methods include a great deal of time for data analysis, professional skill demand, difficulty to point out problem, and relying on user's memory.

The first method based on quantitative data of brain wave belongs to operation evaluation. This method hypothesizes causal relation between user's brain wave and emotion when using software. Through analyzing the specific scene pointed out by brain waves data, this method easily detects scene in which users feel difficulty in using software. The experiment confirmed that four out of five subjects statistically had a significant difference between the brain waves when the evaluated software was "easy to use" and the brain waves when the software was "difficult to use". The proposed method based on "Type II" improves usability evaluation efficiency in nearly three times than that of exsiting methods.

The second method based on quantitative data of gaze point belongs to performance

* Doctor's Thesis, Department of Information System, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DT0161040, February 1, 2004.

operation. It is hypothesized that a usability problem probably exist when a user's gaze point moves long distance, or when a users' gaze point moves slowly when using a web page. An experiment confirmed that usability problems could be found out by replaying the screen and gaze point motion when a user's operation efficiency is low. The result shows this method improve effectiveness by making it easier to find out usability problem in evaluation.

The third method based on quantitative data of first impression belongs to subjective evaluation. This method proposes a causal relationship between design factor and impression factor, which can indicate usability problem more clearly. Moreover, immediate comparison of first impression does not depend on user's memory in usability evaluation. The method decides design factors that elicit target impression based on statistical analysis of impression data of web pages. Three experiments were conducted in three countries considering internationalization of WWW. The experiment results clarified design factors that elicit good impression in audiences of three countries. The proposed method in this chapter can point out usability problem about first impression easily.

The thesis consists of five chapters. Chapter 1 introduces definition and importance of software usability. This chapter also clarifies existing software evaluation methods and their problems. This chapter also clarifies existing software evaluation methods and their problems. Chapter 2 proposes a method of usability evaluation by measuring brain waves. Chapter 3 proposes a method of web usability evaluation by gaze point information. Chapter 4 proposes a method of web usability evaluation based on audience first impressions. Chapter 5 concludes this thesis with a summary and future works.

Keywords:

Software usability, usability evaluation method, brain wave, gaze point, first impression

List of Major Publications

● Journal

1. **Jian Hu**, Kazuyuki Shima, Rudiger Oehlmann, Jiamin Zhao, Yasuhiro Takemura, and Ken-ichi Matsumoto, “An Empirical Study of Audience’s Impressions of B2C Web Pages in Japan, China and the UK”, Electronic Commerce Research and Applications, Elsevier, 2004 (in press).

● International Conference

2. Noboru Nakamichi, Makoto Sakai, **Jian Hu**, Kazuyuki Shima, Masahide Nakamura, “WebTracer: Evaluating Web Usability with Browsing History and Eye Movement,” Proceedings of the 10th International Conference on Human-Computer Interaction (HCI International 2003), Vol.1, pp.813-817, Crete, Greece, June 2003.
3. Makoto Sakai, Noboru Nakamichi, **Jian Hu**, Kazuyuki Shima, Masahide Nakamura, “WebTracer: A New Integrated Environment for Web Usability Testing,” Proceedings of the 10th International Conference on Human-Computer Interaction (HCI International 2003), Adjunct Proceeding, pp.289-290, Crete, Greece, June 2003.
4. Noboru Nakamichi, Makoto Sakai, **Jian Hu**, Kazuyuki Shima, Masahide Nakamura, “Development and Evaluation of A Usability Evaluating Tool: Webtracer,” Proceedings of the International Symposium on Empirical Software Engineering (ISESE 2002), Vol.2, pp.27-28, Nara, Japan, October 3-4, 2002.
5. **Jian Hu**, Kazuyuki Shima, Rudiger Oehlmann, Jiamin Zhao, Yasuhiro Takemura, and Ken-ichi Matsumoto, “An Investigation of Impressions of B2C web pages in China, Japan, and the UK,” Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2002), vol.I, pp.83-88, Florida, USA, July 14-18, 2002.
6. **Jian Hu**, Jiamin Zhao, Kazuyuki Shima, Yasuhiro Takemura, Ken-ichi Matsumoto, “Comparison of Chinese and Japanese in Designing B2C Web

Pages toward Impressional Usability," Proceedings of the 2nd Asia-Pacific Conference on Quality Software (APAQS 2001), pp.319-328, Hong Kong, December 10-11, 2001.

7. **Jian Hu**, Yasuhiro Takemura, Kazuyuki Shima, Ken-ichi Matsumoto, Katsuro Inoue and Koji Torii, "Analysis of relation between impressions and design of B2C web page," Proceedings of the 5th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2001), Vol.1, pp.286-293, Florida, USA, 22-25 July, 2001.
8. **Jian Hu**, Masahiro Nakanishi, Hirokazu Tagaito, Kazuyuki Shima, Ken-ichi Matsumoto, Katsuro Inoue and Koji Torii, "A method of usability testing by measuring brain waves," Proceedings of the International Symposium on Future Software Technology 2000 (ISFST 2000), pp.159-164, GuiYang, China, August 28-30, 2000.

Contents

List of Figures

List of Tables

1. Introduction

In recent years, computer users' strata are spreading with development of Information Technology (IT) and popularity of WWW. Novice users are dramatically increasing with the various IT products spreading in our daily life. An increase and complicatedness in software function is becoming a tense reality. Under these circumstances, we have to confront the questions that how easy the software is to learn and use, how productively users will be able to work, and how much support users will need. In other words, usability of software is therefore getting more and more important.

Usability is quality in use. However, many software developers would rather work with machine than work with people; they show little interest in issues in such as how much data should appear on the screen at one time. Additionally many designers do not realize that their perception of their creation does not provide much information about how others will react to it. That is why we get all those "perfectly obvious to the designer" creation. Usability should be regarded as one more quality attribute for consideration during software or website construction. Usability is a difficult attribute to embed in any system---not only software---and it requires specific knowledge and a lot of awareness about the user's likings, requirements, and limitations [17].

1.1 Definition of software usability

Definition of usability is "the extent to which a product can be used by specified goals with effectiveness, efficiency, and satisfaction in a specified context of use."(ISO 9241, part 11) Usability includes five basic attributes: learnability, efficiency, user retention over time, error rate, and satisfaction [5]. The five basic attributes play different role in various software. This thesis applies the definition to software usability when the product means software.

- **Learnability**: How easy it is to learn the main system functionality and gain proficiency to complete the job. People usually assess this aspect by measuring the time a user spends working with the system before that user can complete certain tasks in the time it would take an expert to complete the same tasks. This attributes is very important for a novice user.

- **Efficiency**: The number of tasks per unit of time that the user can perform by using the system. We look for the maximum speed of user task performance. The higher system usability is, the faster the user can perform the task can complete the job.
- **User retention over time**: It is critical for intermittent users to be able to use the system without having to climb the learning curve again. This attribute reflects how the system works after a period of no usage.
- **Error rate**: This attribute contribute negatively to usability. It does not refer to system errors. On the contrary, it addresses the number of errors the user makes while performing a task. Good usability implies a low error rate. Errors reduce efficiency and user satisfaction, and they can be considered as a failure to communicate to the user the right way of doing things.
- **Satisfaction**: This shows a user's subjective impression of the system [5].

1.2 Importance of software usability evaluation

Improving usability whether of IT systems, e-commerce Web sites, or shrink-wrapped software is not only highly cost-effective, but it can also reduce development, support, training, documentation, and maintenance costs. A system's usability does not only deal with the user interface; it also related closely to the software's overall structure and to the concept on which the system is based [17]. As IBM has stated, usability "makes" business effective. It makes business efficient. It makes business sense." [15] According to Jakob Nielsen, usability efforts can increase sales by 100 percent [30].

Good usability is gaining importance in a world in which users are less computer literate and cannot afford to spend a long time learning how a system works. Usability is critical for user system acceptance. A software product with better usability will result in reduced support costs in terms of customer support costs etc. For a software development organization operating in a competitive market, failure to address usability can lead to a loss of market share should a competitor release a product with higher usability. Poor usability can negatively influence efficiency, effectiveness, and satisfaction. Usability is a key aspect of a software product's success.

Web site and application shall be considered as special software in WWW. Web usability is taken as a branch of software usability in this thesis. Because of the Web's rapidly increasing significance in software development, the role of usability

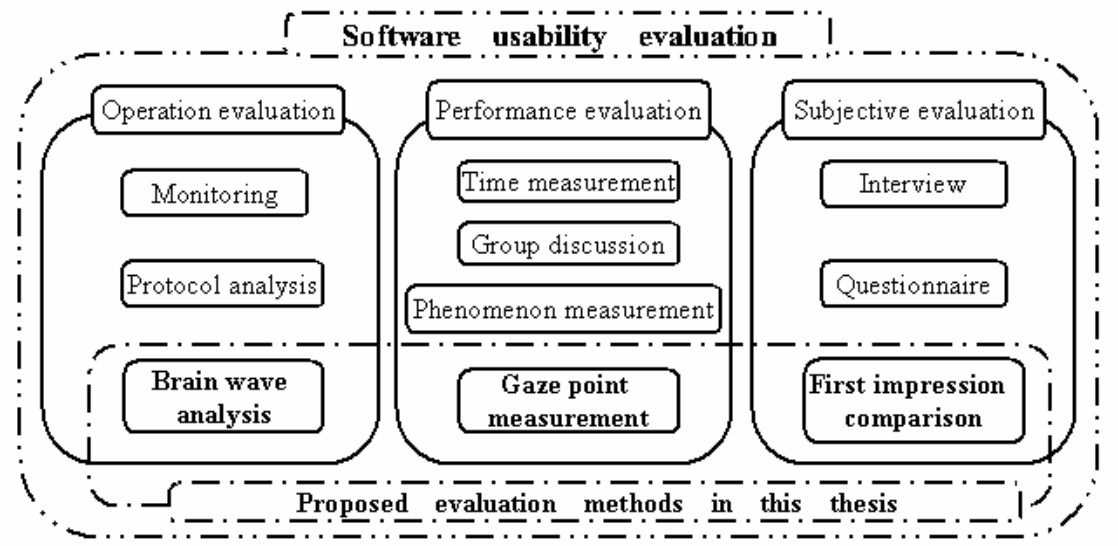


Figure 1. Placement of this research

engineering for Web application is becoming important. However, the usability of Web sites and application continues to be worse that that of more traditional software. However, to be competitive in e-business, usability is a must [17].

1.3 Existing usability evaluation methods and problems

In order to improve the software usability (quality in use), an iterative design of usability engineering is necessary. The usability process consists of three steps of trial, evaluation, and improvement. Usability evaluation is a central activity in the usability process. It can determine the current usability level and whether the design works [17].

1.3.1 Existing usability evaluation methods

In a large way, the conventional evaluation methods for software usability could be classified into three types [1]: operation evaluation, performance evaluation and subjective evaluation. Focusing on different attributes of software usability, these three types of methods are complementary to each other. They can be used separately or together according to evaluation purpose of software usability.

Operation evaluation is a type of method based on “observation” of dialogue

between system and users. This type of method mainly focuses on two attributes of software usability: learnability and user retention over time. Mainly, the evaluation object is “process of dialogue” between system and users in this method. Various factors consisting of the dialogue are examined in view of “ease to use”. Compared with the other two types of method, this type of method focuses on finding out the problem existed in a system. For example, in monitoring and protocol analysis method, the situation in which a user operates a system practically is recorded by digital video. After the task finished, the trouble of a user can be found out effectively by analyzing the recorded data.

Performance evaluation is a type of method to analyze parameters like execution time or error rate etc by which quantitative observation is possible after distributing some tasks to users. This type of method mainly focuses on two attributes of software usability: efficiency and error rate. Measured data are used to verity a hypothesis after statistical processing. The reliability and objectivity can be improved if measured population is taken largely. For an instance, key stroke model is well-known which estimates operation time by allocating prospected time taken in action factor of input (“push a key”, “see text” etc) [3].

Subjective evaluation is a type of method to analyze subjective impression of users. This type of method mainly focuses on the attribute of software usability: satisfaction. This method collects data by execute a questionnaire and interview after a user operates the evaluated system. Questionnaire procedure demands to set question items that shall fit the evaluation purpose. A representative example is interface satisfaction evaluation sheet-QUIS, which widely cover the items influencing satisfaction.

These three types of method focus on different attributes introduced in section 1.1. In addition, there is not one standard or well-established usability evaluation method. Various types of evaluation method have their merits and demerits. When we consider which types of method or which method to use in software usability evaluation, character of the software is important. Table 1 indicates the importance extent of three types of evaluation method for three kinds of software. The table also shows that different types of evaluation method are adapted to various software with different character.

1.3.2 Problems in existing evaluation methods

As shown in figure 1**,** operation evaluation contains two methods: monitoring and protocol

analysis. These two methods need a great deal of time for data analysis and demand much professional skill. Performance evaluation includes three methods: time measurement, group discussion, and phenomenon measurement. Problems of performance evaluation include it is difficult to point out usability problem, result is incapable to wide application, and comparison subject is necessary. Subjective evaluation mainly holds two methods: interview and questionnaire. Problems of subjective evaluation include that it is dependent on user's memory and difficulty to point out problem. Moreover, subjective evaluation has possibility of information distortion in statistic process. These problems in existing usability evaluation methods restrict efficiency and effectiveness of software usability evaluation.

1.4 Research purpose and placement

Research purpose of this thesis is to establish effective and efficient evaluation methods of software usability by solving problems in existing three types of evaluation methods: operation evaluation, performance evaluation, and subjective operation. Focusing on different usability attributes of software, the proposed evaluation methods consist of a method of usability evaluation by measuring brain waves as operation evaluation, a method of web usability evaluation by gaze point information as performance operation, and a method of web usability evaluation based on audience's first impressions as subjective evaluation. Figure 1 shows the relation between existing evaluation methods and three proposed methods in this thesis.

1.5 Outline of this thesis

In order to improve usability relating to software quality, this thesis introduces the tries to make up for the problems in above three existing evaluation methods of software

Table 1. Relation between evaluated software and evaluation methods

Types of usability evaluation methods	**Evaluated Software**		
	Word processor	Mathmatica	Web site
Operation evaluation	Very important	Important	Very important
Performance evaluation	Important	Very important	Very important
Subjective evaluation	Important	Important	Very important

usability. In this thesis, three studies were conducted to improve existing evaluation method respectively based on brain wave analysis, proposal of two metrics about gaze point information, and bettered experimental procedures in user impression evaluation. The three studies respectively construct chapter 2, chapter 3, and chapter 4.

The thesis consists of five chapters. Chapter 1 introduces definition and importance of software usability. This chapter also clarifies existing software evaluation methods and their problems. This chapter also clarifies existing software evaluation methods and their problems.

In chapter 2, a method is proposed for evaluating software usability by measuring subjects' brain waves. This method hypothesizes causal relation between user's brain wave and emotion when using software. By analyzing the specific scene pointed out by brain waves data, this method easily detects scene in which users feel difficulty in using software. Based on experiments, it was verified that nearly 80% scenes in which users felt difficulty in using software could be detected with 73% analysis time cut down. It is concluded that by analyzing the pointed out part according to brain waves data, a person without expert skill can also detect usability problems efficiently.

Chapter 3 proposes a hypothesis that set up a causality relation between user's gazing point information and usability problems in software. The hypothesis is that in case that user's gaze point moves longer distance in a given web page, there is "difficult to use" problem in the page. Moreover, in case that users' gaze point moves slowly, there are usability problems in text of the web page. This hypothesis was suggested to detect quantitatively the characteristics of usability problems in given web pages. An experiment with five subjects was conducted. The experiment confirmed that usability problems could be found out by replaying the screen and gaze point motion when a user's operation efficiency is low. The result shows this method make it easier to find out usability problem in software.

In chapter 4, the proposed method decides design factors that elicit target impression based on statistical analysis of impression data of web pages. The method hypothesizes a causal relationship between design factor and impression factor, which can indicate usability problem more clearly. Immediate comparison of first impression in the method make the experiment result does not depend on user's memory in usability evaluation. Three experiments were conducted in three countries considering internationalization of WWW. As a conclusion, the design factors of web page design and the extent to which

those design factors can elicit good impressions in a web page become clear. There is a tendency in the subjects that cultural difference influence impressions more than sex difference does. In B2C page design and improvement considering impression, the common and different points of impressions of audiences in different culture could be considered important and put to practical use.

Chapter 5 concludes this thesis with a summary and future works.

2. A Method of Usability Testing by Measuring Brain Waves

In recent years, computer applications have become popular in many fields. A lay user can use a computer as well as use a common household electric appliance nowadays. This situation has made the demands for software usability stricter than in former times. Research on evaluating user interfaces relating to software usability evaluation is extensive [4, 16, 27, 40, 44, 46]. Usability testing means the activity of performing usability evaluation in a laboratory with a group of users and recording the results for further analysis. Nobody can predict a software system's usability without testing it with real users. Monitoring is a normal method of usability testing. As mentioned in chapter 1, monitoring method records situation when a user operates a software system practically by digital video. After the operation finished, the trouble of a user can be found out effectively by analyzing the recorded data. The monitoring method need a lot of time for data analysis and demand much expert skill. The two problems of monitoring limit evaluation efficiency of software usability.

This chapter proposes a method to evaluate usability quantitatively by measuring brain waves. An ESAM has been proposed as a method for measuring changes in emotions using brain waves [26]. Using this method, research on a plant operator's emotions was undertaken in an existing research [22]. The proposed method associates variations of brain waves with emotions or feelings when subjects use a software system. Based on existing monitoring method, this method uses brain wave data of subjects to identify "difficulty to use" scene in the mean time. The section 1.3 introduces that operation evaluation mainly focuses on two attributes of software usability: learnability and user retention over time. The two attributes are often mingled tightly in reality. This chapter use "difficulty to use" as substitute of the two attributes. In other words, "difficulty to use" means it is difficult to learn and remember usage of the object software. An experiment results that this method improves evaluation efficiency cutting down analysis time greatly. A person without expert skill can also find out the "difficulty to use" scene easily in a brain wave graph.

2.1 Measuring emotion by brain waves

Brain waves are generated by the activities of nerve cells called neurons. The neurons

Figure 2. Experimental view

normally maintain a state of electric potential from -60mv to -90mv. A neuron stimulated by other neurons will develop an action potential and brain waves occur as a result. Although the frequency of the brain waves ranges from 0Hz to hundreds of Hz, the frequency is usually measured in a setting such as a hospital setting, and ranges from 0.5Hz to 100Hz. We used an electroencephalograph (EEG) to measure brain waves. An EEG measures differences of electrical potential. Ten disc electrodes were affixed to a subject's head according to the International 10-20 System. The difference in electrical potential between each electrode was measured. We used the Emotion Spectrum Analysis equipment Ver1.0 manufactured by the Brain Functional Research Institute and an NF (neuron function) Circuit Design Block, Inc. The sample cycle was one percent per second. We measured the frequency of band waves, including theta waves (5Hz to 8Hz), alpha waves (8Hz to 13Hz), and beta waves (13Hz to 20Hz).

Musha et al. proposed the ESAM, in which a linear regression model is used to estimate emotional state [26]. Musha et al. chose four human emotional elements and defined them as anger/stress, sadness, joy, and relaxation. In our experiment, there are forty-five kinds of combinations for every two electrodes at random among a total of ten electrodes. Correlation coefficients of each combination are calculated three times because there are different frequencies of band waves consisting of theta, alpha, and beta waves in the experiment. As a result, 135 correlation coefficients (y_1, y_2, ⋯, y_{135}) are calculated. Assuming the four levels of emotional elements can be represented as (z_1, z_2, z_3, z_4), and then the following determinant is assumed in the ESAM.

$$z_1 = c_{1,1} y_1 + c_{1,2} y_2 + \cdots + c_{1,135} y_{135} \quad (1)$$
$$z_2 = c_{2,1} y_1 + c_{2,2} y_2 + \cdots + c_{2,135} y_{135} \quad (2)$$
$$z_3 = c_{3,1} y_1 + c_{3,2} y_2 + \cdots + c_{3,135} y_{135} \quad (3)$$
$$z_4 = c_{4,1} y_1 + c_{4,2} y_2 + \cdots + c_{4,135} y_{135} \quad (4)$$

y_1~y_{135} represents an “input vector”; z_1~z_4 represents an “emotion vector”, and $c_{1,1}$ $\cdots c_{4,135}$ represents an “emotion matrix”. The input vectors of subjects who can control their emotions are used to determine an emotion matrix. For example, input vectors are measured when the subject creates emotions of anger/stress, joy, sadness, and relaxation. Then the emotion matrix can be calculated using the determinant in the equations (1), (2), (3), and (4). The following emotion vectors are assumed for each input vector.

Anger /stress:	$z_1=1, z_2=0, z_3=0, z_4=0$	(5)
Joy:	$z_1=0, z_2=1, z_3=0, z_4=0$	(6)
Sadness:	$z_1=0, z_2=0, z_3=1, z_4=0$	(7)
Relaxation:	$z_1=0, z_2=0, z_3=0, z_4=1$	(8)

2.2 Usability testing method based on brain wave

To employ the emotion matrix, Musha et al. used the input vectors of actors who can control their emotions. Although this method may work in the case of actors, finding a subject who can control emotions related to software usability is difficult. Two phases were employed in our method. In the first phase, patterns of the subject’s brain waves are induced and measured when the subject uses reference software in which messages or functions of software menus were changed to evoke the user’s emotions related to software usability. In the second phase, the subject’ brain waves are measured when the subjects use the target software. Usability of software consists of various properties, such as ease of acquisition, efficiency, and appearance, to name a few. Our focus was on the intelligibleness of application software menus because menu is one important character of Windows 9x/NT /Me/2000/XP application software. We classified menu into two kinds: “easy to use” menu and “difficult to use” menu. The “difficult to use” menu was separated into two types according to two basic reasons when a menu is difficult to use: menu appearance (messages) and menu function. These three kinds of

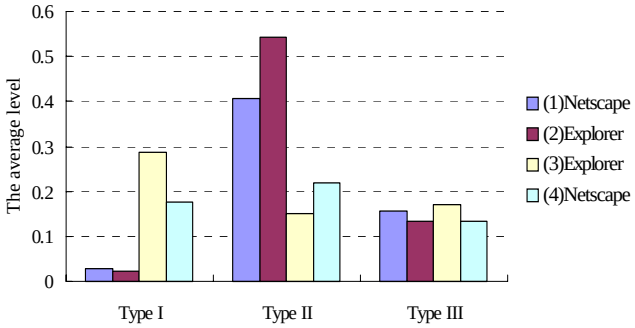


Figure 3. Average levels of brain waves

menus were used to bring out “easy to use” emotion and “difficult to use” emotions. Messages or functions of software menus were changed to evoke three emotional elements: 1) easy, intelligible, 2) deliberate, 3) frustrated and confused. These three emotional elements are related to usability. We proposed a method based on three types of brain waves and their corresponding relationship to software usability as follows:

Type I: The brain wave produced when the user employs familiar menus. The level of “Type I” was considered to be in direct proportion to the level of “easy to use” of software usability. These menus, which allowed users to easily employ the application software in a highly intelligible manner, were meant to evoke simple, intuitive emotions.

Type II: The brain wave produced when the user employed unreadable menus in which messages were written in mixed irregular Japanese Hiragana and Katakana letters. The Japanese messages in those menus appeared as “oPeN”, “cOpY”, or “foNtS” in English. The level of “Type II” was considered to be in direct proportion to the level of “difficult to use” of software usability. This kind of menu message was expected to evoke feelings of deliberation because the interface of the software was difficult for Japanese subjects to understand and use.

Type III: The brain wave produced when the user employs faulty menus in which the messages do not correspond with functions. For example, the function of “cut” was executed when subjects selected a message of “copy”. The level of “Type III” was considered to be in direct proportion to the level of “difficult to use” of software usability too. This kind of menu was expected to evoke in the subject feelings of

confusion because the subject could not understand the strange behavior of the faulty software.

Although we proposed our method based on three types of brain wave corresponding to the intelligibleness of application software menus in this research, both the number of brain wave types and the software feature like menu or button can be adjusted in the method.

2.3 Preliminary experiment

The purpose of the preliminary experiment was to confirm whether the phases proposed for evoking target brain wave were effective or not. We verified that an emotion matrix was practical for usability testing. We experimented on one subject B_0 (Figure 2). We gave the subject a task itemized in the 13 items listed in the instructions. The instruction paper was in the subject's left hand as shown in Figure 2.

The task was to make simple presentation slides with Microsoft PowerPoint 97. We allowed the subject to ask questions freely about the contents of the instructions. In the experiment, we collected brain waves' data, video images, audio, and the gaze point information of the subject. We collected the brain wave data using the Emotion Spectrum Analysis Equipment Ver.1.0 and recorded the video data and audio data on videotape with a video camera. We also collected the gaze data with an Un-contacting type of eye mark recorder. We did the experiment in a sequence of (1) training, (2) a phase for evoking feelings, and (3) a phase for evaluating usability. In order to avoid the influence of other unexpected factors, we fully trained the subject to make the subject accustomed to the experimental task before the experiment of evoking target feelings. We got emotional vectors of usability in the phase of evoking feelings. We measured four types of the subject's brain waves as follows: (1) we measured the subject's brain wave data as a value of "Type I" when the subject used the standard menu command smoothly. (2) We measured the subject's brain wave data as a value of "Type II" when the subject used the menu command with its allocations and names changed. (3) We measured the subject's brain wave data as a value of "Type III" when the subject used a menu command with its name and function mismatched. (4) Finally, we measured the brain wave data of the subject when rested quietly with eyes closed.

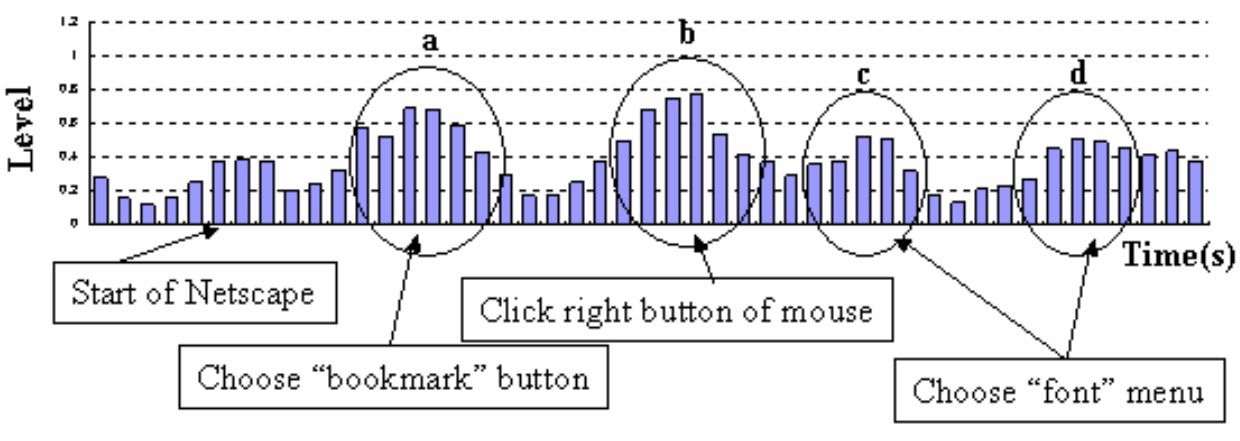


Figure 4. Levels of "Type II"

With these four values measured, we created an emotion matrix.

In order to verify the effectiveness of the emotion matrix we created, we conducted an evaluating experiment. We selected Netscape Communicator 4.5 and Internet Explorer 4.0 as object software in the evaluating experiment. Our purpose was not to investigate which browser was better, but to confirm whether the emotion matrix was effective in usability testing. We made the subject carry out the task four times by using browser in the following order: (1) Netscape, (2) Explorer, (3) Explorer, and (4) Netscape. In case that a subject repeated to use same menu command, we used the brain wave data when the subject used a menu command for first time to calculate emotion matrix. We calculated "Type I", "Type II", and "Type III" using the emotion matrix. We analyzed the result by comparing three types of this brain wave data with the answers from an interview of the subject. In an interview after the preliminary experiment, the comments of the subject were as follows:

- It is easier to use Netscape because I usually use this browser.
- I am dissatisfied with the position of Netscape's bookmark because it differs from the one I usually use.

An average level of the subject's brain waves is shown in Figure 3. We define a horizontal axis to represent three types of brain waves. In each type of brain wave, the browser used to elicit each type of brain wave listed in the order of the conducted tasks. The levels of the subject's brain waves changed markedly between the task using (2) Explorer, and the task using (3) Explorer in both Type I and Type II. In Type I, the

level of brain waves signifies the degree of the subject's intuitive feelings in the same direction. Therefore, the surprising variance that occurred in Type I can be interpreted as: the subject became familiar with the task, and had a strong intuitive feeling in the task of (3) and the task of (4). In Type II, the level of brain waves represents the degree of the subject's deliberation. We may assume that the subject became skilled at Explorer in Type II. The brain wave data of Type I and Type II show the same result in different ways. On the other hand, the levels of "Type III" are mostly constant. The result of the subject's interview is in accordance with the result of Figure 3, except for (3) Explorer and (4), Netscape of "Type II".

In Figure 4, the peak area marked "a" with a relatively high level of brain waves is in accord with the interview of the subject. However, the subject did not express any information about areas of "c" and "d" in the interview, and these areas show that the subject's brain braves were high when he chose a font menu in the experiment. We asked the subject again about this, and he told that it was the first time for him to change a font size while browsing. He had forgotten to tell us this fact in the interview. This interesting fact shows that we can discover usability problems by brain waves even though the subject has not told us this information in the interview ("b" in Figure 4). Figure 4 based on the brain waves of the subject was generally in accord with the information obtained from an interview with the subject. As mentioned previously, brain waves can help us to find usability problems that cannot be found in an interview. Therefore, such a method is applicable to the usability testing of software, and high levels of brain waves may have important meaning for usability testing according to this research.

Table 2. The outline of task instructions used in the experiment

T_1. Go to the specified link.	T_2. Enlarge a font size.	T_3. Go back to the first page.
T_4. Go back to the link of T1.	T_5. Go to a favorite link.	T_6. Enlarge a font size.
T_7. Add the URL to a bookmark (or favorite).	T_8. Create a new window.	T_9. Go back to the first page.
T_{10}. Close the window.	T_{11}. Go back to the first page.	T_{12}. Make a font small.
T_{13}. End a browser.		

2.4 Main experiment

Considering the individual differences in brain waves, the purpose of the main experiment was to apply brain wave data to other subjects and validate the proposed method. We experimented on five subjects: (B_1, B_2, B_3, B_4, B_5). Other conditions were the same as in the preliminary experiment. Considering the effects of the different order of the experiment, we asked subjects (B_1, B_2, and B_4) to carry out the task four times in order of (1) Explorer, (2) Netscape, (3) Netscape, and (4) Explorer. For subjects B_3 and B_5, we asked them to conduct the experimental task in order of (1) Netscape, (2) Explorer, (3) Explorer, and (4) Netscape. However, we have not analyzed influence of the different order on our experimental results in this paper. To avoid the influence that a subject repeated to use same menu command, we also used the brain wave data when the subject used a menu command for the first time.

There were 13 task items from T_1 to T_{13} listed in the task instruction paper (Table 2). The maximum of brain wave data in each action during each task is taken as the analytic minimum element. Each action induced the subjects' three types of brain wave data according to "Type I", "Type II", and "Type III". We excluded the portion of time not related to the time necessary for usability testing, by analyzing image data and gaze data. We evaluated and analyzed video data and gaze data by the monitoring method without reference to brain waves, and made reasonable judgments on the subjects' face emotions, words, and behaviors during the experiments. First, we drew a detailed action table of subjects based on the video data and gaze data. Once the detailed action table was completed, we extracted scenes in which corresponding task actions seemed "difficult to use" or "easy to use" for the subjects. The extracted result is shown in Table 3. We investigated each action in order of the point of the high levels of the subjects' brain waves. If the result based on brain waves is in accord with the result of monitoring during the investigation, we will have succeeded in the detection of usability problem. This can verify the effectiveness of our method of usability testing by measuring brain waves.

In order to investigate the connection between the level of brain waves data and extracted scenes by monitoring, we conducted a t-test. The purpose for conducting the t-test was to confirm whether or not the average level of brain waves data significantly differ, for the extracted scenes based on monitoring and for the other scenes. We

measured the average difference of the three types of brain waves data, and assumed that Normal Distribution could be applied to the brain waves data. In addition, we assumed the average difference of three types of brain waves data is the same. We verified an opposition to the hypothesis.

For this research, we assumed that x in the formula of the t-test was the level of brain waves data of the extracted scenes based on monitoring, and y was the level of brain waves data of the scenes not extracted. n_1 was forty-five and n_2 was five. Therefore, the degree of freedom (n_1+n_2-2) was forty-eight. As results of the t-test, a significant level of alpha is shown in Table 4 and Table 5. The results are separated into two types of scenes; namely, “easy to use” or “difficult to use” and were extracted from monitoring. In Table 4 and Table 5, we defined a straight axis as subjects and types of brain waves as the horizontal axis. The constituents of the tables indicated the significant alpha level of the t-test. Moreover, the same t-test was conducted for scenes of “missed operations” where the subject made mistakes in operation during the experiments. The results of the t-test, therefore, are shown in Table 6.

Next, let us consider Table 4. If we assume “<10%” as a significant level alpha, we find a significant difference in the field of “Type I” of B_1 and B_5 in scenes which seemed “easy to use” based on monitoring. Nevertheless, only two subjects’ brain waves showed a significant difference. Therefore, we could not conclude there is a definitive connection between monitoring results and the level of brain waves data. In Table 5, if we took “< 10%” as a significant alpha level, we would find a significant difference for 4 out of 5 subjects in the field of “Type II”. “Type II” is equivalent to scenes which seemed “difficult to use” based on monitoring. Therefore, efficiency

Table 3. Result of our monitoring method on subjects

Subject	Scenes which seemed “Easy to use”	Scenes which seemed “Difficult to use”
B_1	(1) T_5, T_9, (2) T_2, T_5, (3) T_3, T_4	(1) T_2, T_6, T_7, T_{12}
B_2	(3) T_3, (4) T_4, T_5	(1) T_2, T_3, T_4, T_{12}, T_{13}, (4) T_{13}
B_3	(1) T_{12}, (2) T_{12}	(1) T_2, (2) T_2, T_7, (3) T_7,(4) T_{11}
B_4	(1) T_1, (2) T_2, (3) T_2	(1) T_2, T_3, (2) T_1, (4) T_1, T_7, T_9
B_5	(3) T_1, T_2	(1) T_{12}, (2) T_1, T_{12}, T_4

*We tested every subject four times with 13 task items. The number in ‘( )’ of this table represents one of the four tests.

Table 4. t-test result of scenes appearing "easy to use" based on monitoring

Subject	Type I	Type II	Type III
B_1	< 0.1%	< 20%	> 50%
B_2	> 50%	< 50%	< 20%
B_3	> 50%	> 50%	< 50%
B_4	> 50%	< 50%	> 50%
B_5	< 1%	< 40%	> 50%

Table 5. t-test result of scenes appearing "difficult to use" based on monitoring

Subject	Type I	Type II	Type III
B_1	> 50%	< 0.1%	> 50%
B_2	< 20%	> 50%	< 0.1%
B_3	> 50%	< 10%	< 40%
B_4	< 1%	< 1%	< 40%
B_5	< 30%	< 5%	< 5%

in finding a software usability problem should become easier by searching the high-level area of brain waves data of "Type II" than by searching video data for scenes that seemed "difficult to use".

Furthermore, we considered "misses in operations" based on Table 6. The influence of "Type III" became greater for subject B1. A reason for this result may be that the misses in operations occurred only one time in all samples, and the level value of emotional elements of the misses were not clearly reflected in the results of the t-test.

Normally to find out latent software usability problems from monitoring video, three to ten times length of monitoring time is necessary [1]. To find out how much the proposed method improve the efficiency in finding out scenes which may indicate usability problems in a software system, we analyzed monitoring video of five subjects in order of the high level of three types of brain wave. The result showed that nearly 80% scenes in which users felt difficulty in using software could be detected with 73% analysis time cut down.

Table 6. t-test result for scenes of "misses in operations" based on monitoring

"Difficult to use"	Type I	Type II	Type III
Subject B_1	> 50%	< 5%	< 1%
Subject B_3	> 50%	> 50%	> 50%

2. 5 Analysis of usability evaluation efficiency

The proposed evaluation method sets threshold value of emotional elements. The usability evaluation efficiency of this method was analyzed. The analysis method is to confirm whether usability evaluation efficiency improves by merely analyzing the video scene that corresponding emotion levels measured surpass the threshold value. Several analysis indexes and formulas were defined as below:

N: the sum of usability problems pointed out by monitoring.

T: experiment time

n: the sum of usability problems pointed out by monitoring scene when corresponding emotion levels measured surpass the threshold value.

t: experiment time that emotion levels measured surpass the threshold value.

$$\text{Coverage Rate} = n / N \quad (9)$$

$$\text{Usability Analysis Rate} = t / T \quad (10)$$

$$\text{Efficiency of Existing Methods} = N / T \quad (11)$$

$$\text{Efficiency of Proposed Method} = n / t \quad (12)$$

$$\text{Analysis Ratio} = (N / T) / (n / t) \quad (13)$$

Based on above formulas, three types of brain wave proposed in this chapter were analyzed by calculation results. The result of Coverage Rate and Analysis Ratio of “Type II” is shown in Table 7. The Analysis Ratio values in this table show that

Table 7. Analysis Ratio of “Type II”

	Coverage Rate		
	60%	*80%*	*100%*
B1	14%*	28%	46%
B2	24%	38%	46%
B3	19%	23%	25%
B4	23%	34%	43%
B5	7%	10%	14%
Usability Analysis Rate average	17%	27%	35%
Analysis Ratio	3.53	2.96	2.86

* Usability Analysis Rate

usability evaluation efficiency based on “Type II” is improved nearly three times.

2. 6 Conclusions

This chapter proposes a method for evaluating software usability with brain waves. This method hypothesizes causal relation between user’s brain wave and emotion when using software. Through analyzing the specific scene pointed out by brain waves data, this method easily detects scene in which users feel difficulty in using software. This method quantifies the emotions of subjects using the evaluated software. Experiments based on proposed method illustrated the following points:

1) The experiment confirmed that four out of five subjects statistically had a significant difference between the brain waves when the evaluated software was “easy to use” and the brain waves when the software was “difficult to use”.

2) The proposed method based on “Type II” improves usability evaluation efficiency in nearly three times than that of exsiting methods.It was verified that nearly 80% scenes in which users felt difficulty in using software could be detected with 73% analysis time cut down.

This research does not expect to find user interfaces that all users will feel easy to use. However, software may have some user interfaces that are difficult to use. This method is useful for finding problem when users feel difficult or easy to use evaluated software. Users may also feel difficulty in using software because of a lack of experience. In such a case, some users do not state outright that they felt difficulty because they cannot know whether the problem exists in the user interface or the problem is their experience lack. This method will give us some hints about which user interfaces should be improved. By analyzing the pointed out part according to brain waves data, a person without expert skill can also detect usability problems efficiently by this method. Practically, there may be not enough time to improve all problems of user interfaces found in usability testing. The quantified types of emotions are helpful for software developers in selecting the first priority solution for usability problem.

3. A Study of Web Usability Evaluating Method by Browsing History Including Eye Movement Tracking

Numerous companies have failed in developing online business application for a lack of corporate vision by not considering Web usability. Designing attractive Web sites is a crucial problem in business, since Web sites directly reflect the images and sales of companies [5]. Therefore, usability evaluation for web pages is now an important concern in finding flaws in the pages with respect to usability [32]. Web usability testing is becoming a popular way to conduct usability evaluation. Web usability testing requires subjects (users) to browse a target web site, and then evaluators get feedback from the users based on an interview.

This chapter studies an evaluation method of software usability by users' browsing history record, especially the gaze point tracking record. A web site was used as object software in experiment of this study. As mentioned in chapter 1, Performance evaluation includes three methods: time measurement, group discussion, and phenomenon measurement. Problems of performance evaluation include it is difficult to point out usability problem etc. This chapter proposes a method to solve the problem in existing method. Two metrics including movement speed and distance of gaze point are proposed to find out usability problem in web page. The movement distance of gaze point is the movement amount of a gaze point in a given web page by a unit of pixel. The movement speed of gaze point is value that the gaze point distance divides time consumed in a given web page. The unit of the movement speed of gaze point is pixel/second. This method assumes there is a relation between usability problem and the two metrics. According to experiment and reasonable conjecture, this study sets up a hypothesis that speedy gaze point motion and longer gaze movement distance suggest software usability problem (user trouble) in software such as a web sites.

This chapter firstly introduces a tool for web usability evaluation --WebTracer, which is use to record browsing history and operation. The tool can record user's gazing points, a user's operational data, and the screen image of browsed pages. In addition, the WebTracer can replay a user's browsing operations. In an evaluation experiment, WebTracer records five users' browsing operations without interruption.

Analysis is done by applying the usability testing support function of the WebTracer. The causality shown in analysis result verifies the hypothesis as mentioned above.

3.1 Outline of this thesis

Today, many individuals and enterprises apply web site as tool of sending information or business. The development of web sites cost a lot of such as time and labors. Especially, a web site shall be charming, intuitive to users. A Web site shall be easy to use for a number of users comparing with traditional software [18]. Web usability is the ease to use of a web site. Web usability is very important because of influencing sale of enterprises [6]. In addition, if the user has not understood the intention of the web developer, the information and function of the web site will not be applied very well. Usability evaluation is necessary in order to design charming web site. Web usability testing is widely used as evaluation method of web usability. Usability testing is a method that web developer/manager tests a site with help of normal users. Typical Users will access the object web site practically to give their comments about the usability. These comments will efficiently help to find out usability problem existing in the evaluated web site [31].

However, usability testing traditionally consists of “Think aloud”, “group interview”, and “heuristic” evaluation techniques. These qualitative evaluation methods demand professional knowledge and skills of usability. Those qualified expert is very limited compared with the explosive number of web sites. Moreover, the increased cost and time necessary for web usability evaluation is one problem. As solution to these problems mentioned above, several usability testing approaches using quantitative data are proposed. Most of these approaches employed data of server side based on access log, such as evaluation of web page shift approach. Evaluation based on browsing action of user has not been applied yet.

This study applied quantitative information of gaze point of eye to show the close relation between usability problems and user’s behavior in web browsing. Concretely, a device of gaze point tracking records coordinates of the point that a user’s eyeball is looking at. Then the movement speed and movement distance shall be calculated. The WebTracer can supply the necessary data to complete the calculation. It is supposed that the proposed usability evaluation method do not demand much professional knowledge

and experience. The two metrics supply qualitative measurement possibility of web usability problem. This method may decrease evaluation cost of software usability in practice.

3.2 Related research

3.2.1 Usability testing

There are many methods of usability evaluation developed until now, one typical method category is usability testing. Usability testing is a general term of the usability evaluation method in which users practically operate various machines and systems or its prototype. Users are the most important center of usability testing. Recently, the traditional usability testing is applied in web usability evaluation. The applied methods include performance measurement based on quantitative data such as operation time or operation times of users, “think aloud” in which what users say is analyzed to identify the usability problem.

However, the expenses of preparing test user and analyzing data, the cost of time, available device make it difficult to apply usability testing. In addition, usability testing cannot be executed before system/prototype is finished because of process limit of the development engineering of product. In order to improve efficiency and decrease cost in usability testing, automation evaluation method, evaluation support tool, and computer software tool are studied and developed. The below is an introduction of several methods to support usability evaluation of GUI application.

Guzdial’s method applies Markov chain analysis to find out continuous two operations used very frequently by calculating probability that an operation continuously conducts after another operation. If finding out two continuous operations, the position of corresponding GUI parts should be closer to make the mouse movement distance less between the two operations.

According to the method of Kishi [21], operation history of user and standard operation history are compared in several separated phases. In operation of GUI application software, it is possible that mouse or keyboard can operate same operation part. It is also possible several parts have same function. By different comparison criterion like function or part, there are various levels to compare standard operation history and user operation history, such as comparison level between different executed

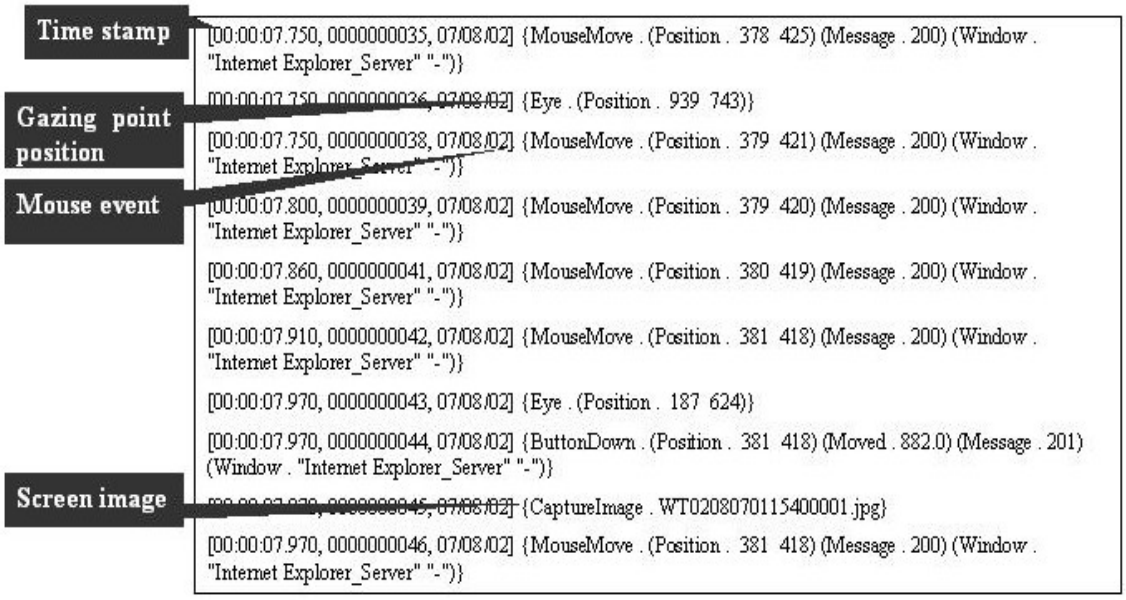
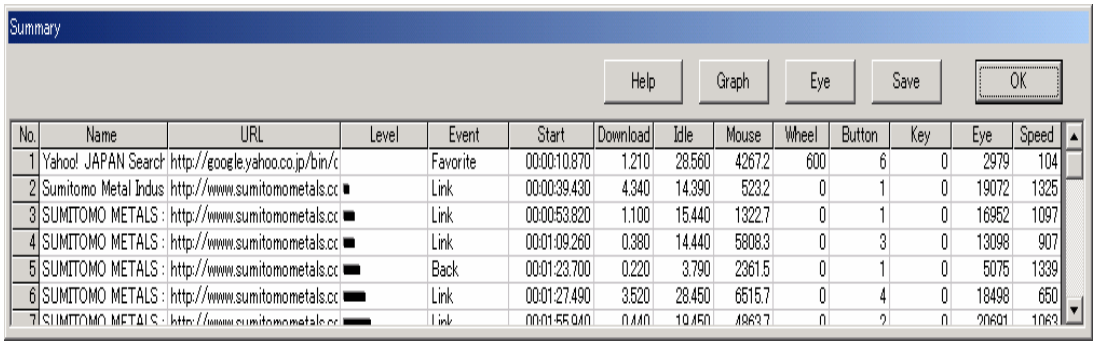


Figure 5. Example of data collected by the WebTracer

functions (commands) disregarding difference of operating parts, comparison level between differences of operated parts disregarding difference of input devices, comparison level between differences of input device. Comparing standard and user history after separating them in multi phases shows the difference level between the two histories. These comparisons make it easy to identify whether those differences suggest usability problems.

Ikemoto's method [14] can detect the operation that takes longer time than predicted time by compare predicted time with time interval of operations that select menu or button by mouse. In case that time difference is big, it is possible that the operation is difficult fro users to understand or it takes users much time to find out next operation part because of the complicated screen layout of system.

"UI Tester" and "GUI Tester" developed by Okada el. are tools to evaluate software of FAX device and GUI application [35, 37]. The common characteristic of the two tools is to find out the common mistaken operation by extract common operation pattern from operation histories of more than one user. To minimize effluence to evaluation result from individual difference, the analysis of common operation pattern is effective with more users' operation histories collected. It is possible to apply above methods to evaluate single web page because these methods evaluate object limited a few screens.



No.	Name	URL	Level	Event	Start	Download	Idle	Mouse	Wheel	Button	Key	Eye	Speed
1	Yahoo! JAPAN Search	http://google.yahoo.co.jp/bin/c		Favorite	00:00:10.870	1.210	28.560	4267.2	600	6	0	2979	104
2	Sumitomo Metal Indus	http://www.sumitomometals.cc		Link	00:00:39.430	4.340	14.390	523.2	0	1	0	19072	1325
3	SUMITOMO METALS :	http://www.sumitomometals.cc		Link	00:00:53.820	1.100	15.440	1322.7	0	1	0	16952	1097
4	SUMITOMO METALS :	http://www.sumitomometals.cc		Link	00:01:09.260	0.380	14.440	5808.3	0	3	0	13098	907
5	SUMITOMO METALS :	http://www.sumitomometals.cc		Back	00:01:23.700	0.220	3.790	2361.5	0	1	0	5075	1339
6	SUMITOMO METALS :	http://www.sumitomometals.cc		Link	00:01:27.490	3.520	28.450	6515.7	0	4	0	18498	650
7	SUMITOMO METALS :	http://www.sumitomometals.cc		Link	00:01:55.940	0.440	19.450	4963.7	0	2	0	20691	1063

Figure 6. Example of a summary (summarized browsing history)

However, it is difficult to detect usability problem from web site consisting of linked web pages by existing methods.

3.2.2 Application example of gaze point information

Mori el. [28] proposed a method to improve usability prototype based on screen design in information system development. They focused on human interface and analysis of eyeball movement, and tried a study repairing prototype screen. The experiment result showed that both operation speed of screen processing and satisfactoriness of user were obviously improved under the method.

In research of Mori el., movement of subject's gaze point firstly was recorded, and then the track of subject's gaze point was drawn in prototype screen. They set up a hypothesis that smooth motion of gaze point shall be movement from upper side to lower side, or movement from left side to right side. They checked out opposite motion of user's gaze point and modified the position of items in the screen. They compared the operation speed and user satisfaction between original design and modified design. The effectiveness of screen design using track of gaze point was verified. However, in case of usability evaluation using track of gaze point, the knowledge and experience is necessary to find out problem from the track of gaze point. Therefore, this method cannot realize the target to improve evaluation efficiency and decrease evaluation cost.

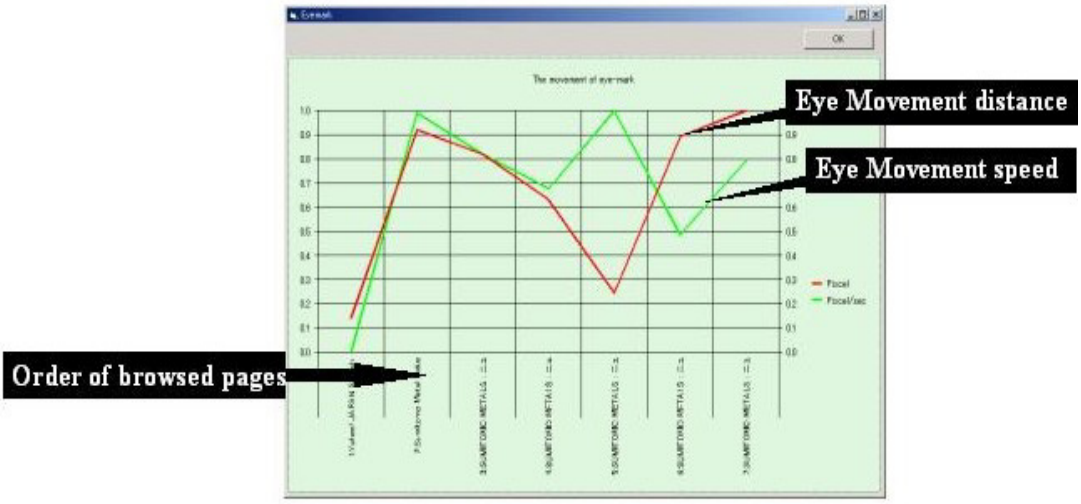


Figure 7. Example of statistics graph of eye movement

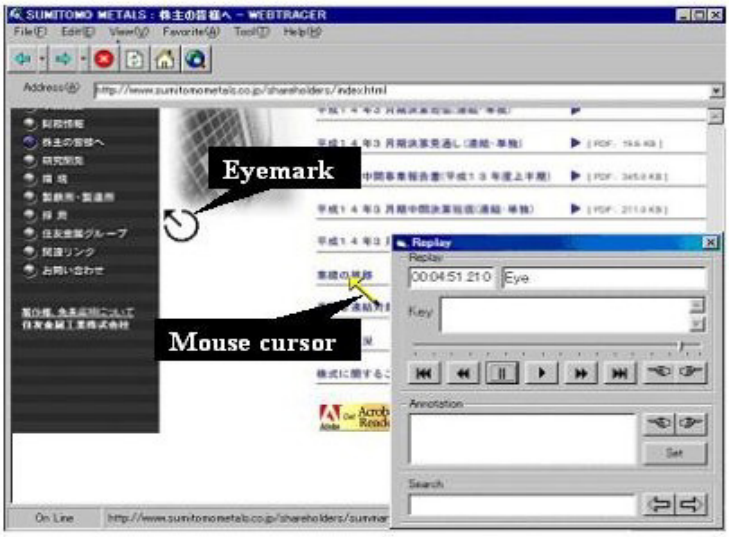


Figure 8. Example of replay screen with eye mark

3.3 WebTracer

WebTracer is an integrated environment for web usability evaluation. It can record a user's browsing operations, replay recorded browsing history, and provide analysis

tools that can depict graphs and calculate statistical equations. WebTracer gets information of gaze point position by sampling way in an interval (every tenth of a second). The gaze point movement is complicate that there are many patterns. Simply, the movement pattern of gaze point can be divided into delicate movement and leap movement (e.g. saccade movement). WebTracer cannot record delicate movement in tenth of a second. Thus, the distance of this kind delicate movement could be considered as acceptable errors. Moreover, spatial errors of gaze point measurement are about a character because of environment factor etc. WebTracer is optimized especially in the following two features.

3.3.1 Recording web operation

WebTracer records the various user operational data needed for replay and analysis. Specifically, WebTracer records user's gazing points via the camera eye, mouse movements and clicks, keyboard inputs, and the screen image of the browsed pages. An example of data collected by WebTracer is shown in Figure 5. Unless the appearance of the browsed page changes, WebTracer does not record browsed screen image. The image is captured only when a transition of the browsed page is triggered by a user's events (e.g., mouse click to follow the next links). Thus, the size of the recorded image can be significantly reduced to 1/10 to 1/20 of the size of recorded data when compared with data recorded in an Mpeg-2/4 format.

3.3.2 Replay and summary functions

WebTracer can support usability testing by using a replay of the user's operations, summarized data, and graphs derived from the recorded data. By using the summarized data, we can capture the characteristics and statistics of each page, which helps with the analysis of a web site. Recorded data are summarized in the form of a table for every page, as is shown in Figure 6. The data can also be shown in graph form. An example of an eye movement statistics is shown in Figure 7. In addition, an example of the replay screen with the eye mark of the user (the user's gazing point) is shown in Figure 8. The replay feature reproduces operations, such as the eye mark and mouse cursors, operations performed when the page is being browsed. In another window, WebTracer can display other events, such as a keystroke. Moreover, at any time during the recording, we can insert annotations and replay these annotations later. An experiment has been conducted to evaluate the effectiveness of

WebTracer in a Web usability evaluation.

3.4 Usability testing using gaze point information

Web usability is the ease of use of the Web. Problematic page design, such as “inconsistency between link titles and target pages” and “ugly menu layout” etc, decreases the usability of the web page, and thus should be detected and revised. However, managing such usability problems often requires qualitative evaluation by experts with enough knowledge and experience.

This research conducted a usability testing of a web site applying WebTracer (developed by group) and gaze point tracking device. The gaze point information of five subjects was collected. These subjects undertook diagnosis of the web site and presented comments of usability in an interview. The purpose of the experiment is to verify the hypotheses set up in above mentioned part. The gaze point information, users’ comments, and the checklist result of subjects would be used to contrast whether the gaze point speed and distance implies the usability problem in a web page.

3.4.1 Outline of the experiment

Firstly, the object web site and task was assigned to five subjects. WebTracer recorded their browsing operation. After task finished, an interview was conducted. The subjects were asked about points that they felt difficult to use the software in experiment. Finally, compared with the interview results, data analysis were conducted based on recorded gaze point information. The consistency between recorded data and subject comment was checked.

3.4.1.1 Operation record including subjects’ gaze point information

WebTracer can collect user’s operation history (event) in a web page. These events could be eye gaze point information (coordinates of gaze point in screen measured by gaze point tracking device), key stroke, mouse operation, state of web application, image of browsed web page, shifting time among various web pages etc. Time information has been added in all events record.

Figure 9. A page with long movement distance of gaze point

WebTracer can show the outline of user' browsing information based on collected history data. Figure 7 shows an example of the movement speed and distance of gaze point when browsing web page. In addition, the motion history of the subjects' gaze point in computer screen in operation can be replayed again (Figure 8). The same as replaying digitized video, operation history also can be replayed by various operations like "fast-forwarding", "rewind", "stop", and specifying replayed position by slide bar etc.

3.4.1.2 Subjects and tasks

Five subjects conducted the task in the experiment. These subjects apply Internet in daily life. Four subjects often use the object web site in the experiment. One subject uses the object web site for the first time. The task of five subjects is to find out information from web site of our university (www.aist-nara.ac.jp) as below.

Task 1: investigating premise knowledge of a class.

Task 2: finding out telephone and fax number of office of Graduate School of

information Science.

Before the task started, content of task was explained. The state when the subjects browsed web pages was observed during tasks. During subjects performing assigned tasks, WebTracer recorded the browsing operation (including gaze point information) in the background. It is confirmed there was no break in the process of task in order to record a user's browsing operation as usual.

3.4.2 Analysis of browsing history

According to the recorded browsing data of five subjects, the movement distance and speed was calculated out automatically in WebTracer. Based on the calculated value, record of the gaze point tracking was replayed repeatedly to find out coincidence in accordance with the hypothecation in this chapter. The record of gaze point data in accordance with the hypothesis was found in the experiment (Figure 9, Figure 10). Figure 9 shows the scene in which a subject's gaze point moved at the longest distance record: 16, 929 pixels in a browsed web page. The confusion of recorded tracks of gaze point suggests a probable problem in page layout. Tracks of gaze point in Figure 10 seem congested around several links. This may indicate usability problems in color or texts of page links design. Gaze point in Figure 10 moved in slow speed. These records help to point out usability problem in a given page easily. This shows that hypothesis gets support from the experiment.

3.5 Conclusion

In order to achieve more effectiveness of software usability evaluation, an empirical study was conducted and the result shows that user's eyes (i.e., gazing points) could supply useful information that is quite relevant to usability problem. In the experiment, five subjects were assigned a task to browsing a web site, gaze point tracking device and Webtracer recorded the tracks of subjects' gazing points. Based on the recorded data of gaze point motion, gaze point speed and gaze point distance were calculated. Based on the calculation and replay of the browsing history, an analysis was conducted to find out web pages that have usability problems. Finally, we categorized the usability problems according to speed and distance of the gazing point movement.

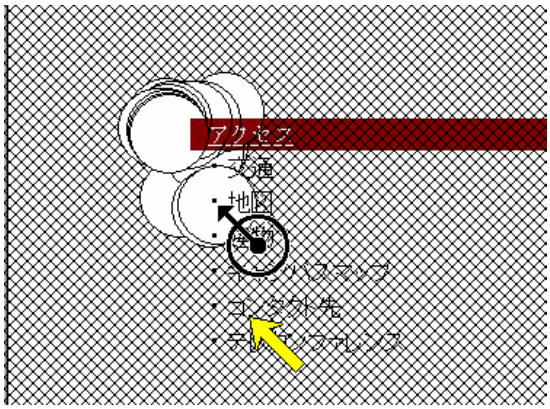


Figure 10. A page with slow movement speed of gaze point

The result verifies when a user browses a web page, the longer the gaze point distance is, the more probable usability problems exist in the browsed web page. The problem may exist in the web page layout. In addition, usability problems such as color or text in links tend to exist in the given web page where a user's gaze point moves slowly. Although a web site is taken as evaluated object in this study, the gaze point information is also applicable to software usability evaluation of other software.

4. An Empirical Study of Audience Impressions of B2C Web Pages in Japan, China and the UK

Subjective evaluation mainly holds two methods: interview and questionnaire. The two methods are dependent on user's memory and difficulty to point out problems in software usability. This chapter proposes a method to solve the problems. The proposed method was applied in a study of audience impressions. Negative impressions that arise during a first interaction with a Business-to-Customer web page often have the unpleasant side effect to destroy a firm's efforts in achieving B2C electronic commerce on the WWW.

This part of thesis verifies the relation between audience impressions and the visual style of a B2C web page. In comparison to previous work, the experimental procedure was greatly improved. It was therefore expected that this change leads to improved results with higher reliability. Moreover, this study considered the impressions of Japanese, Chinese, and English subjects to investigate differences and consistencies in impressions, which are based on the underlying culture. Three empirical studies based on self-report questionnaires were conducted in Japan, China, and the U.K. The studies measured the subjects' impressions of various B2C web pages that showed eight design factors. The evaluation values for seventeen impression factors and their antonym terms were collected in the questionnaires. The studies in China and the U.K were conducted using the same procedure as in Japan. Sign tests of the results show a significant difference in subjects' impressions corresponding to changes in design factors. Moreover, the results show cross-cultural consistencies in various impressions but also several differences between the subject groups. This study concludes by discussing the implications of the empirical results for the visual design of international B2C web pages in terms of target impressions.

4.1 Web usability and audience impressions

4.1.1 The importance of the impression of a business-to-customer web page

The Internet profoundly changes the way in which commerce is conducted. In some ways, Internet commerce seems deceptively simple [31, 45]. However, there are many factors affecting the success of electronic commerce. One of these factors is web

usability, which becomes increasingly important for the Internet society. One day it may become a reality that usability drives the Internet economy [31]. Many Web sites confront usability problems that shall be solved by usability evaluation, web improvement, and redesign.

Usability of IT applications should display five major attributes: learnability, efficiency, memorability, errors and satisfaction [32, 41]. “Satisfaction” shows a user’s subjective impression of a system. “Satisfaction” requests the system should be pleasant to use so that users like it [1, 41]. For business to customer electronic commerce (B2C EC) on the WWW, satisfaction of customers appears more important. However, satisfaction is not intuitive and has often been ignored by designers.

First impression is a subattribute of satisfaction [5], which decides a consumer's image of a product or a company. Impression has the same role in B2C EC on WWW. The positive/good impression of a B2C web page is an important component of audience satisfaction. With a bad impression of a B2C web page, audiences will stop browsing or will not return any more. There are so many similar B2C web pages on WWW.

The importance of impressions has been emphasized in the design of physical products [9, 24, 25]. In real world commerce, the impressions elicited by a sales agent or a commercial organization as a whole influence the overall satisfaction of the customer [20]. The feelings that are aroused in interacting with a system are especially important for systems that are used on a discretionary basis such as EC [47]. In the case of EC, impressions of the web pages will influence the audience desire to purchase. Impressions can be expected to play a similarly important role in the design of B2C Web pages just as they do for physical products. The impressions created in interacting with a B2C web page are especially important for EC systems, which are used on a voluntary basis. People do not have to use such a system if they dislike it [31]. The same conditions apply to a B2C web page.

In B2C EC on the WWW, services or products are supplied to customers through web pages, which are the interface between the seller and the buyer. A positive impression can play an important role in attracting audiences to a web page and turning them into customers. Therefore, the research is needed to analyze the relation between audience impressions and the visual style of a B2C web page. However, little research has been conducted regarding the impressions of B2C web pages. Nielsen's

research on web usability merely mentions that the first impressions an audience obtains from a given web page are important [31, 33, 34]. A systematic methodology that takes in consideration impressions will be helpful in the design of B2C web pages.

Kim and Moon have conducted experimental research on the feeling of trustworthiness [20], which especially focused on the feeling of trustworthiness that the interface of a cyber-banking system should elicit in customers who carry out financial transactions. Forty terms for emotions were identified to indicate emotions elicited by the user interfaces of cyber-banking systems. Fourteen design factors were concluded to describe the studied user interfaces. The results of their research indicate that it is possible to design customer interfaces of cyber-banking systems, which will elicit target emotions, such as trustworthiness.

4.1.2 Cross-cultural impressions

Culture is always viewed as a collective phenomenon. It represents “mental programming”, which is partially predetermined by the collective values of their local community [38]. Nielsen advocates that web usability shall consider international use that serves a global audience [31]. Barber and Barde [2] argue the success of a global interface may only be achievable when the interface design reflects the cultural nuances of the target audience. Negative and positive consumer reactions become more understandable and predicable when a person’s cultural context is taken into account [38]. It is expected that people with different cultural backgrounds would respond differently to a globally generic Web site. Different cultural responses would have important implications for the corresponding Web interface design. This is important for building electronic commerce systems that offer global usability [7]. We therefore hypothesize that culture differences may be reflected in the relationship between the design of a B2C web page and audience impressions of that web page. In this research, we study the differences and consistencies of impressions resulting from three diverse cultures.

Newsbytes Asia reports that the number of online users in Asia is expected to reach 228 million by 2005. Most of Asia's users are in Japan. We conducted first a controlled experiment in Japan. In addition to Japan, we selected China and the UK for the following reasons: Newsbytes Asia reports that China is expected to surpass all other countries in Asia by 2005. 37.6% of Asia's online users will be Chinese in 2005; this

signifies 85 million users. The huge population and the remarkable growth of the Chinese economy make it undoubted that China will also take an important position in the world's B2C EC in the near future. As for the UK, of the approximately 215 million current global Internet users, 57.4% use English as their primary language [42].

The main objective of this research is to study the relation between impressions and visual design of a B2C web page base in an improved experimental environment. This research also compares subjects from Japan, China, and the UK to identify differences and consistencies in impressions across different cultures [10]. Based on this research result, it is expected to find an implication for the optimal design of B2C web pages that are intended to elicit certain target impressions of audiences while they interact with B2C web pages. This paper especially focuses on the positive impressions that a B2C web page should elicit in audiences for the first time. This will help a designer to improve the usability of B2C web pages in terms of audience impression.

4.2. Studies based on three controlled experiments

4.2.1 Definition of terms

Electronic commerce (EC): has been defined as the delivery of information, products and services, or payments via telephone lines, computer networks or any other electronic means [20]. However, this paper restricts the meaning of this term to business that is processed by the World Wide Web. Here EC includes business transactions like online shopping, online securities, online banking [29].

Culture: Culture is always viewed as a collective phenomenon. People learn patterns of thinking, feeling, and potential acting from living within a defined social environment, normally typified by country [38].

B2C web page: A web page used for B2C EC in the WWW. In this research, we just consider static web page as objective web page constituted by eight design factors (Table 9). The eight types of design factor are still the basic components in a B2C web page despite the growth and popularity of dynamic web pages.

Web page design: In this research, we refer to the visual style of a web page based on available design factors such as title, background color etc. The B2C web pages were designed based on various layouts of eight design factors in this research.

Impression: An impression describes an emotion state or feeling of an audience, which is elicited by a B2C web page when the audience visits the web page for the first time. In this research, impression is expressed by affective terms like “charming”, “boring”, “likable” etc, which are referred to as impression factors (see below). For the purpose of this paper, emotion and feeling is viewed as synonym for the term “impression”.

Impression factor: In this research, impression factors point to seventeen impression terms (Table 8) which are considered the most important ones while audiences browse a B2C web page. These seventeen factors, together with their antonyms, are used as impression dimensions to evaluate the audience impressions of a B2C web page. It is assumed the seventeen impression factors construct the impression space of audiences of a B2C web page.

Usability: The usability of a computer product is the extent to which the product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use (ISO 9241; 11; 1994). Usability of IT applications should display the following five major attributes:

1. Learnability - easy to learn;
2. Efficiency - efficient to use, making it highly productive;

Table 8. Seventeen impression factors

No.	**Impression factors**
1	Awkward
2	Brief
3	Boring
4	Charming
5	Cluttered
6	Soulful
7	Unpleasant
8	Consistent
9	Epochal
10	Exciting
11	Likable
12	Opulent
13	Progressive
14	Reliable
15	Simple
16	Vibrant
17	Witty

3. Memorability - easy to remember, so that a casual user is able to use the system easily after a period of non-use;
4. Errors - relatively error-free, so that users make few errors and recover easily from those they made;
5. Satisfaction - pleasant to use, so that users like it [32, 41].

Design factor: visual style elements/components, which a B2C web page consists of. We apply eight design factors in this research. Each design factor has several choices to be selected in web page visual design (Table 9).

Choice: In this paper, choice means the available selections or options included in a design factor. This term is also used to represent a version of a web page, which features a particular option of a given design factor. The objective B2C web pages used in this research were designed based on choices of eight design factors (Table 9).

4.2.2 Design of the improved experiment

The results of Kim and Moon's research [20] are not immediately applicable to the actual design of customer interfaces due to some limitations. For example, the emotions indicated were the result of passive exposure to those visual interfaces, not

Table 9. Twenty-nine versions of web page

Design factors	**Choices of design factors**			
	A	*B*	*C*	*D*
Title format	Bar	Clipart	Text	No Format
Title position	Top	Middle	Bottom	No Format
Menu size*	>1/16	>1/32	>1/64	No Format
Clipart size*	>1/2	>1/4	>1/16	No Format
Main color	Primary	Pastel	No Format	-
Background color*	White>1/2	White<1/2	Color I	Color II
Color brightness	High	Medium	Low	-
Color harmonization	Harmonized	Multiple	Single	-

- "No Format" means no special format was used to indicate this design factor.
- 29 versions of the original web page were designed to represent 29 choices of the eight design factors.
- Underlined choices are used in the original B2C web page.

* The value such as "1/16", "1/2" means percentage of a screen size a design factor takes in a B2C web page.

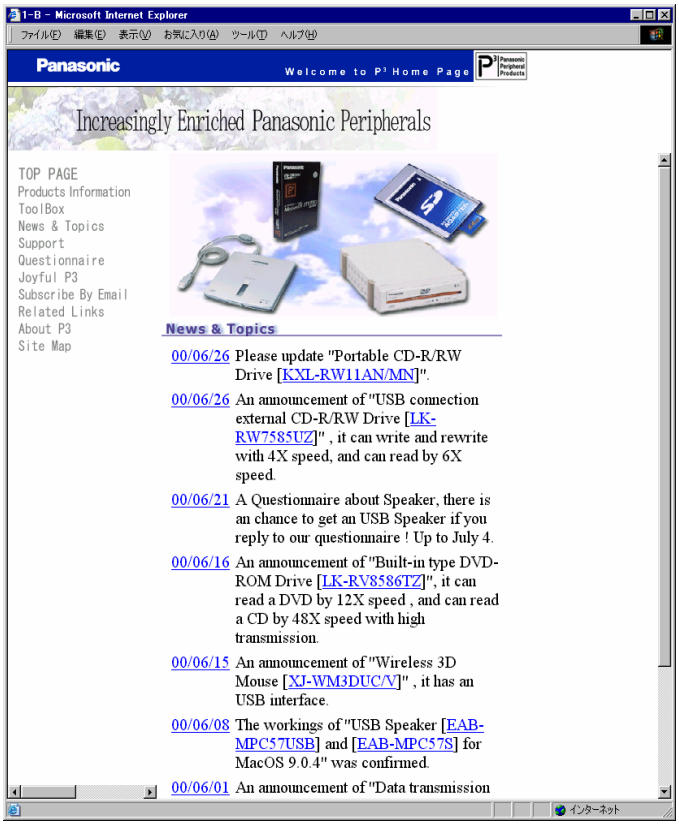


Figure 11. The original B2C web page

the result of actual usage of the cyber-banking system. Therefore, future studies should investigate the emotional usability of the customer interface while subjects are actually using the system.

Improvements of the experimental procedure in the current study in comparison with previous research [20] include:

- The method of eliciting impressions was changed from merely presenting slides to actual browsing of a B2C web page. It is expected that this change will improve the results related to impressions and therefore the reliability of the conclusion.
- Based on Kim's research and preliminary studies in the current research, the redundancy of the impression scales (terms) was reduced. Several new impression terms were added to the impression terms.
- Experimental conditions were designed that are near the real environment in which normal audiences browse B2C web pages. Every subject in the experiments

was able to browse independently web pages using separate computers.

- The evaluation of the impressions was based on comparisons between different web pages. Subjects evaluated the intensity of their impressions of various versions of web pages when these different web pages are juxtaposed on the same computer screen.
- In addition, the statistical analysis method was changed to Sign test. The sign test only utilizes the numerical relation of larger or smaller for matched pairs of variables rather than the actual values of the variables. It is expected that the sign test method will clarify what factor or choice can elicit a target impression efficiently.
- The subjects in this research come from three different cultures. This will help to identify differences or consistencies in impressions of B2C web pages in terms of international use.

4.2.3 Preliminary study

In our empirical investigation, we used a list of forty impression factors and fourteen design factors based on Kim's research [20]. As using all these factors in our investigation may have imposed too large a cognitive load on the subjects, we decided to identify the most important impression factors and design factors. In cooperation with a web design company in Japan, we first conducted a preliminary investigation to limit the number of impression factors and design factors. In this preliminary study, Web designers were asked to vote, which factors are important to evaluate B2C web pages from the audience perspective. These important factors were then used in the main studies [10, 11, 12, 20]. Furthermore, the designers could add new design factor and impression factor, which they considered as important for the design of a B2C web page. We asked seven experienced B2C web designers to score the list of impression factors and design factors according to importance from integer value 1 to 5. The higher value indicates an increased importance. Based on the scores and subsequent ample discussions with the seven designers, seventeen representative impression factors (Table 8) and eight design factors (Table 9) were selected for use in the main investigation. The seventeen representative impression factors and their antonyms were used as bipolar dimensions to evaluate various B2C web pages in the main studies.

Figure 12. Choices of title format

4.2.4 Main studies across three countries

This research targets the first impression of B2C web pages. Therefore, an original homepage of a Japanese B2C website, which is based on the eight design factors, was selected. This original web page contains information about computer products in Japanese (Figure 11). The web pages and the questionnaires in the study were generated in Japanese, Chinese, and English versions according to the native languages of subjects in the studies. This enabled the accurate measurement of subtle differences in the impression terms.

The purpose of the studies was to verify the relations between impressions and the design of B2C web pages by measuring subjects' impressions of various B2C web pages in an improved experimental environment. The same material was used in all

the studies in Japan, China, and the UK. However, it was translated into the appropriate language. The questionnaire consisted of the seventeen impression factors and their antonyms, altogether thirty-four affective terms. A two part self-report questionnaire was used in this phase. Six questions in the first part focused on the subjects' knowledge of and experience with B2C web pages. In addition, subject age and gender information was also collected. This part was also intended to prepare subjects for the second part of the main study. The second part included eight pages, one for each of the eight design factors. On each page, the seventeen impression factors and their antonyms were arranged in a form. Self-reports are most commonly used as the measures of emotion under laboratory conditions [8, 23, 43]. A questionnaire with a bipolar seven-point Likert scale with integer values from -3 to 3 was set to evaluate different intensities of the audience impressions elicited by the various B2C web page designs.

4.2.4.1 The various B2C Web pages

The eight design factors embodied twenty-nine possible combinations of choices. Twenty-nine web pages were designed by applying different combinations of the design factor options. The possible choices of design factors were juxtaposed as shown in Figure 12. The subjects compared different web pages on a single screen and noted their impression values for each web page in the questionnaire.

Table 9 describes the twenty-nine web page designs consisting of the eight design factors, which were used in the main studies. In the heading of Table 9, the letters of "A", "B", "C" and "D" each represent one given version of a possible web page that corresponds to a given design factor. In the second row in Table 9, four choices of "Bar", "Clipart", "Text" and "No Format" represent four versions of web pages corresponding to the design factor: Title format. In addition, there are eight design factors in the design factor column. In this paper, the twenty-nine choices such as "Bar", "Top", ">1/16" etc are related to the twenty-nine different versions of web pages. For example, corresponding to the design factor of "Title format" in Table 9, this paper uses the notion "Bar" to represent the web page design that uses a title in bar format.

A web page to illustrate each design factor was made to prevent confusion in some subjects. The difference between options embodied in a design factor was emphasized graphically. When the subjects did not understand the difference between the displayed web pages, they could refer to the illustrative web page just by clicking on a hyperlink. These eight illustration web pages were linked together with the twenty-nine web pages in a platform web page, which was used to start the task. All web pages needed in the experiment were linked to the platform web page. The experiment in Japan was carried out in a college computer center. The computers were connected in a

Table 10. Evoking probability of three target impressions
regarding choices of the design factor: *Title format*

Awkward	A (Bar)	B (Clipart)	C (Text)	D (No Format)
A (Bar)	0	1	9.98E-1	1.11E-2
B (Clipart)	1.36E-8	0	4.51E-5	4.86E-11
C (Text)	8.21E-4	1	0	2.26E-7
D (No Format)	9.78E-1	1	1	0

Brief	A (Bar)	B (Clipart)	C (Text)	D (No Format)
A (Bar)	0	8.04E-2	1.53E-4	1.75E-1
B (Clipart)	8.69E-1	0	3.95E-2	4.47E-1
C (Text)	1	9.28E-1	0	8.04E-1
D (No Format)	7.48E-1	4.47E-1	1.26E-1	0

Boring	A (Bar)	B (Clipart)	C (Text)	D (No Format)
A (Bar)	0	1	7.34E-1	4.01E-4
B (Clipart)	1.02E-11	0	6.92E-8	4.54E-13
C (Text)	1.74E-1	1	0	2E-8
D (No Format)	9.99E-1	1	1	0

- "No Format" means no special format was used to indicate this design factor.
- As result of the sign test, values in this table represent the probability that a choice in a column elicits a target impression (such as "awkward", "brief" or "boring" underlined in the table) more strongly than a choice elsewhere in a row. (This table shows "Title Format" as an example of the design factors).

server-based network. All the web pages for the experiment were stored in the server and given a URL. Every subject could access the server to browse the B2C web pages using a separate computer. In order to keep consistency between the experiments in Japan, China, and U.K, we selected a similar experimental environment in Japan, China, and U.K. We did the other two experiments in computer centers in Jinan University (South China) and Kingston University (U.K).

4.2.4.2 Procedure

The main study was conducted in several group sessions for a total of sixty-nine subjects in Japan. All subjects were students of two junior colleges. Most subjects had used Web browsers, but only few subjects had experience with browsing B2C web pages. The subjects were aged from nineteen to twenty-two. They showed great interest in Web pages. The eighty-nine Chinese subjects were third-grade students, aging from nineteen to twenty-three. They had greater Web experience than the Japanese subjects. The sixty-eight U.K subjects were first-grade students of the School of Computing and Information Systems, with an age range from eighteen to forty. However, the age of most subjects ranged from eighteen to twenty-four years. Only five subjects were over twenty-six year old. Chinese and UK subjects had richer experience in browsing the Word Wide Web than Japanese subjects. Most subjects belong to the same generation and were assumed to become potential customers of B2C electronic commerce in the near future.

In all the CNNIC surveys beginning from 1997, young users aged 18-24 always account for the highest proportion, which is much higher than the other age groups. The results of the CNNIC 2002 survey show that student users account for the highest proportion of 26.2% among Internet users. Users with university education or junior college education account for the proportion of 55.5%. These features are in accordance with that of subjects in China.

The procedure of the experiment was the same in all three countries. The subjects were given the nine questionnaire-pages and were specifically requested to mark their first impressions about the B2C web pages. The tasks were illustrated carefully to all subjects, and any questions from the subjects were welcomed throughout the whole experiment. After all the subjects understood their tasks, they were instructed to finish the first part of the questionnaire, which collects subject profile information. At the

end of this part, the subjects prepared for the next part. Every subject was asked to start Internet Explorer (IE) in his/her computer. In the next step, the subjects entered the URL to browse the web page that was designed as the work platform.

The opening web page of the platform contained hyperlink buttons to the twenty-nine web pages with the eight design factors. On each page of the questionnaire, the subjects filled out their evaluation values for the seventeen impressions evoked by each web page that featured a specific design for that design factor. The subjects followed the instructions to open the web pages for the eight design factors. They compared all versions of B2C web pages that embodied one design factor, and filled out their impression values in the forms on the task sheets. This study provided three or four choices of web pages for each design factor.

4.3. Analysis and results

Sixty-nine Japanese questionnaires, eighty-nine Chinese questionnaires, and sixty-eight UK questionnaires were collected in the main studies. Although seven

Table 11. Elicitation probability of impressions by the design factor: *Title format*

Impression Items	**Choices of Title format**			
	A (Bar)	**B (Clipart)**	**C (Text)**	**D (No Format*)**
Awkward	1.11E-2	2.98E-23	1.85E-10	*9.78E-1*
Not awkward	1.09E-11	*1*	4.50E-5	1.22E-19
Brief	2.15E-6	1.53E-2	7.46E-1	4.23E-2
Not brief	6.49E-1	3.33E-2	7.63E-7	6.28E-2
Boring	2.94E-4	3.22E-31	3.49E-9	9.99E-1
Not boring	1.78E-12	1	5.08E-8	3.65E-24
Charming	2.32E-18	1	8.39E-13	2.05E-30
Not charming	2.82E-5	2.95E-43	4.81E-13	1

- "No Format" means no special format was used to indicate this design factor.
- Values in this table show the probability that a choice can elicit an impression more intensely than other choices of a design factor. For a given impression item, the most effective choice in a design factor has the biggest elicitation probability value.

values were used to evaluate the intensity of subjects' impressions, the difference among individuals in the intensity of their impressions both was important and unavoidable. Different subjects would have their own quantitative standard to express the intensity of the impression they felt in the studies. Since no definition was given for the numerical values of impression intensity, the mean value of impressions does not reflect the important individual difference. Sign Test was used to analyze the evaluation values of the impressions of the subjects from three cultures based on Level of Confidence.

4.3.1 Sign test

The sign test is a test that shows whether a tendency exists in matched pairs of data, such that one of the variables tends to have larger values than the other [13]. In matched pairs of variables, A represents one of the variables, and B represents the other variable. Then the values of A and B are compared for every matched pair of data. The number of pairs (L) for which A is smaller than B is taken as statistical sum of the test. The number of pairs (W) for which A is bigger than B is taken as statistical sum of the test. This test assumes that the probability that A is smaller than B is the same as the probability that A is greater than B. Here, W+L is the number of all matched pairs other than those where A is equal to B. The test is carried out with the above conditions. The probability P that the differences between data are significant is calculated by the following equation.

$$P = \sum_{x=W}^{W+L} \binom{W+L}{x} \frac{1}{2^{W+L}} \tag{14}$$

A procedure corresponding to above sign test definition and procedure was developed to fit the needs of the analysis of the impression values. The purpose of applying the sign test in this research is to identify the correlation between designs based on design factors and user's impressions. In other words, we wanted to know which choice of design factors is most effective for eliciting various impressions when an audience interacts with a B2C web page. We also expected to discover the intensity of the impression effects caused by the various page designs. A model was constructed to apply the sign test to the results of the main study. Any two choices of given design factors were taken as matched pairs. In a selected matched pair, one choice of web

page, such as version “Bar” (Table 9), could be optionally taken as one of the variables, the other choice of web page, such as version “Text”, could be assumed as the other variable. The observation was labeled as strong/weak impressions elicited by the matched pair of two web pages. A hypothesis was suggested as follow: in terms of given impression factors, the version “Bar” of a web page could bring out a stronger impression than the version “Text”. The level of confidence R represented the probability that the hypothesis was correct. This equation (15) reflected the relation between level of confidence R and Significant Probability P. Equation (16) was used to calculate R.

$$R = 1 - P \tag{15}$$

$$R = \sum_{x=0}^{W-1} \binom{W+L}{x} \frac{1}{2^{W+L}} \tag{16}$$

Here, W means the number of subjects who thought that the version A (web page) elicited a given impression intensely than version B did. L means the numbers of subjects, who thought that the version B elicited a given impression intensely than the version A did, in terms of that impression factor. The values of W and L in terms of a given impression could be gained from our experiment results. The equation would then be used to calculate the evoking probability values of 17 impression factors with respect to choices of the eight design factors (shown in Table 10). The evoking probability indicates for two given choices A_1 and A_2, the probability that A_1 elicits a given impression more intensely than A_2.

Table 10 shows a part of the evoking probability values based on the experiment results in Japan. The complete evoking probability values of an experiment include seventeen sub-tables corresponding to the seventeen impression factors for each design factor. Only a part of the original table is shown here due to a lack of space. Table 10 consists of three independent sub-tables for three target impression factors “Awkward”, “Brief”, and “Boring”. In Table 10, the choices of design factor of “Title format” include A (Bar), B (Clipart), C (Text), and D (No Format). Among the four choices, the evoking probability was calculated with respect to each pair of two

choices. In each sub-table, the left top column of each sub-table shows the target impression factor in bold font. The values in the sub-table indicate the probability that one optional choice in the left column such as “No Format” evoke a given impression with higher intensity than another optional choice in the top row such as “Text”. Note that the value of one in Table 10 is just an approximate value influenced by the calculation precision of the used spreadsheet system. The values in Table 10 merely reflect the more/less relation for each pair of choices in terms of a given design factor. One choice that evokes a given impression more/less intensely is one more or less effective choice. The values in Table 10 could help to find out the most/least effective choice in a given design factor.

Based on the values of Table 10, we calculated the elicitation probability values (shown in Table 11) that a given choice elicit a target impression with the highest intensity than all other choices of a given design factor. The elicitation probability values were calculated by using the multiplication principle of independent probabilities shown in Table 10. Table 11 shows an example of the elicitation probability values. The complete elicitation probability values of an experiment include seventeen impression items for each design factor. Only a part of the original

Table 12. Proximate value of elicitation probability of choices in design factor: *Background color*

Impression Items	**Choices of “Background color”**			
	A (White>1/2)	**B (White<1/2)**	**C (Color I)**	**D (Color II)**
Brief	***3.77E-1***	***5E-1***	5.84E-8	1.26E-10
Not boring	1E-20	2.79E-4	***4.32E-1***	***3.77E-1***
Epochal	1.47E-12	1.64E-3	***4.82E-1***	***3.13E-1***
Exciting	7.98E-24	1.02E-6	***6.81E-1***	***2.13E-1***
Likable	2.30E-11	***5.26E-1***	***2.88E-1***	2.07E-3
Opulent	4.56E-14	3.51E-5	***6.77E-1***	***2.10E-1***
Progressive	2.30E-13	9.29E-3	***5.08E-1***	***2.35E-1***
Simple	***3.23E-1***	***5.55E-1***	2.59E-6	2.17E-8
Witty	2.64E-10	***2.21E-1***	***2.20E-1***	***1.34E-1***

- The underlined ***bold italics*** indicate corresponding choices in this design factor have proximate probability to elicit a given impression shown in the left column.

table is shown here due to a lack of space. In Table 11, the four numerical values in the row of “Awkward” indicate the probability that a given choice of "Title format" elicits the "awkward" impression with the highest intensity in the subjects. For example, the italicized numerical value of 9.78E-1 indicates the probability that the "D (No Format)" choice could elicit the “Awkward” impression with the highest intensity among the four choices of the design factor "Title Format".

In the row for “Not awkward” in Table 11, the italicized numerical value of one indicates the probability that the "B (Clipart)" choice could elicit the "not awkward" impression in the subjects most intensely among the four choices of "Title format”. The value of one in Table 11 was also an approximate value influenced by calculation precision of the used spreadsheet. Therefore, the conclusion of the above analysis for the design factor of “Title Format” is that a designer of B2C web pages should apply the “Clipart” choice in a B2C web page to elicit the impression of “not awkward” in audiences. Moreover, the "No Format" version is the worst option for the design factor "Title Format" because the “No Format” version will probably elicit an "Awkward" feeling among the audience.

4.3.2 The most effective choice of eight design factors

Table 11 shows an example of the analysis result that explains the mutual relation between the thirty-four impression items (the seventeen factors and their antonyms) and the eight design factors. This example includes the most effective choices of the eight design factors. The probability values of one reveal the correlation between impressions items and the corresponding web page versions representing the choices of eight design factors. If this relation can be regarded as causal relation, the findings of this research can be used to construct an interacting model for B2C web page design in terms of target impressions of audiences. When B2C web page designers are aiming at a given audience impression, they can refer to the research results that show the relation between the visual design of a B2C web page and the audience impressions.

The most effective choices of the given eight design factors were discussed above. The ranking of choices for a given impression item was determined by the elicitation probability values of each choice (Table 11). In Table 11, the differences among the elicitation probability values of various choices were significant. Therefore, it is easy

to rank the various choices in the order of the elicitation probability with respect to the various impression items. The ranking also indicates an order of choices for a given design factor when a B2C web page designer wishes to achieve a target impression in visual page design.

In fact, the probability values in a given row in Table 11 happened to be significantly different in terms of the design factor “Title format”. For other design factors, the analysis result will be slightly different. As an example, Table 12 shows a part of the elicitation probability values for the choices in the design factor “Background color”. Unlike the values in Table 11, the differences of the probability

Table 13. The best choices of eight design factors for each country

Design factors	**China**	**Japan**	**UK**
Title format	B	B	B
Title position	A	A	A
Menu size	AB	A	A
Clipart size	AB	A	B
Main color	A	C	C
Background color	D	*	*
Color brightness	B	B	B
Color harmonization	A	A	*

* This mark means that statistically there is not best choice considering significant difference.

Table 14. The worst choices of eight design factors for each country

Design factors	**China**	**Japan**	**UK**
Title format	A	D	D
Title position	CD	D	CD
Menu size	D	D	D
Clipart size	D	D	D
Main color	B	B	B
Background color	A	A	C
Color brightness	C	C	A
Color harmonization	B	B	B

values for these choices are not significant. Therefore, it appears to be more difficult to rank the various choices according to the probability values. From the view of the design of B2C web pages, the most effective choice or the ranking of choices derived from Table 12 are not obvious because of the proximate value of the elicitation probability of choices in the design factor “Background color”.

Doubtless, it will be difficult for a designer to choose a choice for the visual design of a B2C web page when two or three choices of a design factor have no significant differences, such as the two values of 2.21E-1 and 2.20E-1 in the bottom row of "witty" in Table 12. In other words, the choice of “B (white<1/2)” or “C (Color I)” will have almost the same probability of eliciting the impression of "witty". On the other hand, the differences among the other choices embodied in a design factor should be considered too. For example, suppose that A and B are the first and the second choice of a design factor respectively, and the difference between A and B is not significant. Web designers usually select A. However if the choice A causes a practical problem, e.g. an increase of cost or loading time, then the choice B can be a substitute. Therefore, it is useful for B2C web designers to consider not only the first choice but also the second and any other choices if the differences among them are not significant.

In particular, two choices with proximate probability values can be used as substitute for each other in order to elicit a target impression. In Table 12, the

Table 15. The best and worst choices of eight design factors
for China, Japan, and UK

Design factors	**Choices of design factors**			
	A	B	C	D
Title format	~~Bar~~	Clipart	Text	~~No Format~~
Title position	Top	Middle	~~Bottom~~	~~No Format~~
Menu size	>1/16	>1/32	>1/64	~~No Format~~
Clipart size	>1/2	>1/4	>1/16	~~No Format~~
Main color	Primary	~~Pastel~~	No Format	-
Background color	~~White>1/2~~	White<1/2	~~Color I~~	Color II
Color brightness	~~High~~	Medium	~~Low~~	-
Color harmonization	harmonized	~~Multiple~~	Single	-

- Choices meshed are the best for at least one of three countries.
- Choices crossed out are the worst for at least one of three countries.

proximate relation of choices was indicated by emphasizing the values of proximate choices in bold font. Consider the impression item “Brief” in Table 12 as an example: the values of 3.77E-1 and 5E-1 indicate that the choice “B (white<1/2)” can elicit the “brief” impression more easily than the choice “A (white>1/2)”, however the difference between the eliciting probabilities is small.

4.3.3 Comparison of the results from three countries

Based on prior statistic analysis, Table 13 shows the best choices of design factors for each country, whereas Table 9 shows the worst choices of design factors for each country. In Table 13 and Table 14, A, B, C and D represent the choices of eight design factors in Table 14. The results in Table 13 and Table 14 are gained based on the results of the sign test with 95% confidence level. The “*” mark in Table 13 indicates that statistically there is not a best choice considering the significant differences. The best or worst choices theoretically depend on the given impression factor. For example, the choices “AB” were shown in the cell in the “China” column across the “Menu size” line of Table 13, the choice “A” of the design factor “Menu size” is the best choice when the target impressions are “Opulent” and “Reliable”. However, the choice “B” becomes the best choice when the target impressions are “Charming” and “Not boring”. In such a case, we denoted two best choices (like “AB”) in Table 13 and Table 14. According to the results in Table 13, an “ideal” B2C web page for Chinese subjects is shown as Figure 13 in terms of the target impressions such as exciting, soulful, and witty.

Furthermore, Table 15 shows the best and worst choices of design factors for China, Japan, and the UK based on the results of Table 13 and Table 14. Choices in meshed fields are the best for at least one of three countries. Choices crossed out are the worst for at least one of three countries. With the exception of the design factor “Background color”, the original choices (underlined) are the best for three countries. For “Background color”, choice “Color II” is the best for Chinese and not the worst for other two countries. Although we cannot conclude that “Color II” is the best “Background color” for all subjects in three countries, however the result reveals it is possible that the original page can be improved by changing the background color.

Figure 13. The “ideal” B2C web page for Chinese subjects

Barber et al. contend that culture and usability are intertwined into a single entity: culturability where cultural preferences and biases affect the degree of the friendliness of an interface such as background color, graphics, and spatial orientation. For instance, the Japanese associate white color with death. In Chinese culture, the red color represents happiness [2]. This research result also suggests different color and spatial effect in impression across culture.

4.3.4 Comparison based on genders and countries

In order to consider gender difference in subjects from three countries, choices in experiment results were analyzed as figure 14 shows. In this thesis, only best choices are analyzed because of page space. There are seventeen impression factors and eight design factors in this study. There are one hundred and thirty-six best choices for each group of subjects. The percentage of same best choice between different groups is used to describe consistency among groups.

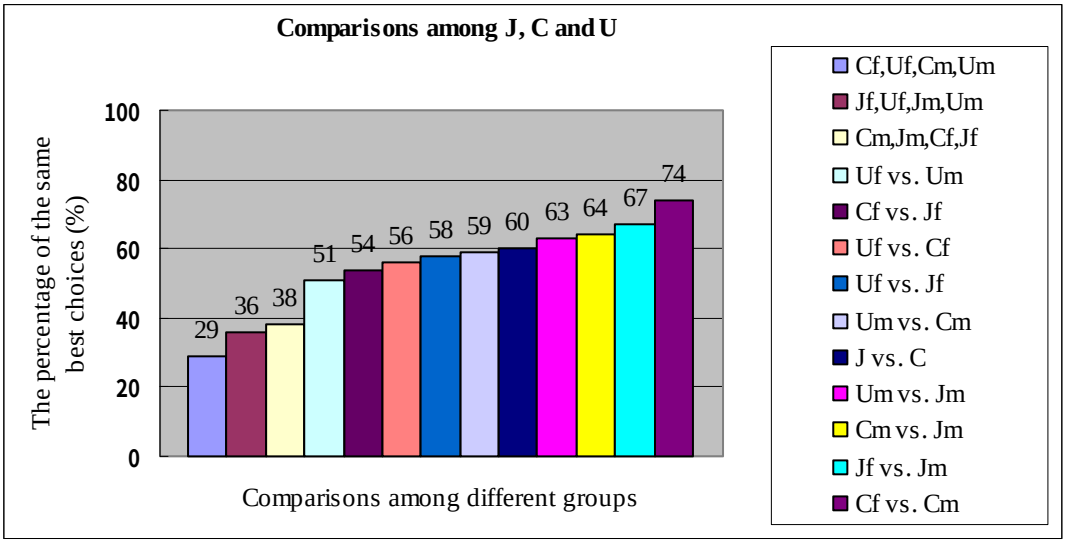


Figure 14. Comparison based on genders and countries

The subjects from different country are classified in male (m) and female (f) group. J, C, and U represent Japan, China, and United Kingdom in figure 14. Thus, there are six groups respectively represented as Jf, Jm, Cf, Cm, Um, and Uf. The value in figure 14 shows the percentage of same best choices among compared groups. Among the six groups based on three countries, the percentage of the same best choices is 24%. The other percentage values (from small value to big value) are shown in Figure 14 from left to right (29% among Cf, Cm, Um, and Uf, 74% between Cf and Cm.). Figure 14 shows a tendency in all subjects that cultural difference influence impressions more than gender difference does.

4.4 Conclusion

Based on the improved experimental design in comparison to previous research [20], this research confirms the causal relation between impression and visual design of a B2C web page. The analysis results uncover the probability ranking of choices in eight design factors in terms of target impression. An interesting finding of proximate choices suggests a possible trade off in the visual design of a B2C Web page. Generally, subjects of three countries have common first impression on design factor of “Title format”, “Title position”, and “Color brightness”. The subjects have different first impression on other five design factors.

We have conducted studies in three countries to evaluate the relation between audience impressions and the visual design of B2C web pages. The results indicate that different B2C web page designs will elicit related impressions in Japanese, Chinese and UK subjects. Moreover, the elicitation probability differs between different visual designs related to different target impressions. This can help the designer to understand interactions between audiences and the B2C web pages of an EC web site. The research demonstrates that a specific choice of design factors applied to B2C web pages elicits positive or negative audience impressions. The research also provides the more important conclusion that we can trade off different design options to create an optimal visual design that can realize various target impressions.

The result shows the proposed method in this chapter can point out usability problem about first impression easily and supply definite solution. Table 9 in section 4.2.1 indicates the eight choice of the original B2C web page by underlined way. According to the study result, the best choices of Japanese subjects are mostly in common with the original design except “background color”. This shows the proposed method can detect usability problem easily. Moreover, the solution is clear: to change the choice from original “White>1/2” to “Color I” based on the study result.

Generally, the comparison of impressions of subjects from Japan, China, and the U.K shows most design factors can elicit the same impressions in three groups of subjects from different culture. This result suggests that many problems in the visual design of B2C web pages have general solutions even in terms of international use. The comparison of the evaluation of impressions based on three groups of subjects shows that some design factors have special culture-dependent characteristics. For this kind of design factor, the optimal design or improvement of B2C web pages (e.g. in multilingual website of global company) must consider the localization in visual design. Japan’s culture had been affected by China’s culture since ancient times; Japan’s culture was also affected by European culture since the Meiji era. It is assumed that this impact on Japan’s culture may be reflected in this research, e.g. the best or worst choices in the Japanese results are expected to be more similar to the Chinese results than to the results obtained in the UK. The assumed tendency was not confirmed by the results of the current study (Table 13, Table 14).

Of course, there are still some limitations in this research that will be addressed in the future. We applied eight design factors and seventeen impression factors to

construct an evaluating system of impression usability based on the related research and preliminary studies. However, with the further development of web page design techniques, it is necessary to integrate new design factors into further research. Moreover, B2C web pages have to address different target customers and cultures, which require the web designer/developer to adjust the impression factors to achieve their usability objective. Although the seventeen impression factors and the eight design factors may not fit the needs of some visual designs of B2C Web pages, given the novel developments of the World Wide Web technology, such as Flash animation or other visual effects based on DHTML technology. Whereas the eight design factors used in this study are essential components even in today's B2C page design, those new design factors may be considered in future research.

To improve good impression usability for B2C EC web pages, a web designer should have a concrete objective of impression usability and clear choices of design factors in mind. Designers can use the approach described in this paper to identify causal relations between their design factors and target impressions. This can be achieved in three stages. Designers should

- Decide about the target impressions and available design factors.
- Select appropriate subjects from target customers based on design usability testing.
- Conduct the experiment and clarify the causal relations between given design factors and impression factors.

They then can realize the actual B2C web page design based on good practice of usability.

5. Summary and Future Works

5.1 Summary

This thesis consists of three studies of evaluation methods for software usability. In the first study, a method is proposed for evaluating software usability by measuring subjects' brain waves. This method contains inducing phase and evaluating phase. Preliminary experiment and main experiment were conducted to certify effectiveness of the proposed method. In the preliminary experiment, patterns of the subject's brain waves are induced and measured when the subject uses reference software. The messages or functions of the software menus were changed to evoke the user's emotions related to software usability. In the main experiment, the subjects' brain waves were measured when the subjects use the target software of evaluation. The result confirms that the change in emotion is reflected in the subjects' brain waves. Consequently, the experiment confirmed that four out of five subjects statistically had a significant difference between the brain waves when the evaluated software was "easy to use" and the brain waves when the software was "difficult to use". The proposed method based on "Type II" improves usability evaluation efficiency in nearly three times than that of exsiting methods.

The second study proposed a hypothesis that set up a causality relation between metrics applying information of user's gazing points and usability problems. The hypothesis is that in case that user's gaze point moves longer distance in a given web page, there is "difficult to use" problem in the page. Moreover, in case that users' gaze point moves slowly, there are usability problems in text of the web page. This hypothesis was suggested to detect quantitatively the characteristics of usability problems in given web pages. An experiment with five subjects was conducted. After analyzing the browsing history of the subjects, result was found to support the hypothesis. The result of this study is helpful to set up a quantitative model to evaluate web usability in future work.

In the third study, in comparison to previous work, the experimental procedure was greatly improved to lead to results with higher reliability. In addition, this study considered the impressions of Japanese, Chinese, and English subjects to investigate differences and consistencies in impressions, which are based on the underlying

culture. Three experiments based on self-report questionnaires were conducted in Japan, China, and the U.K. The experiments measured the subjects' impressions of various B2C web pages that showed eight design factors. The evaluation values for seventeen impression factors and their antonym terms were collected in the questionnaires. The experiment results show a significant difference in subjects' impressions corresponding to changes in design factors. Moreover, the result shows cross-cultural consistencies in various impressions but also several differences between the subject groups. The proposed method can point out usability problem about first impression easily and supply definite solution.

In this thesis, application of quantitative data makes the proposed methods show efficient and effective good point proved by experiments. These new evaluation methods of software usability can improve evaluation efficiency by decreasing evaluation time and professional skill demands. The proposed evaluation methods also supply new viewpoint to solve the problems in existing evaluation methods.

5.2 Future works

Usability evaluation will take more and more important role in information technology progress. Present evaluation method shall be improved to catch up with the progress of technologies. Each evaluation method of software usability has both good and bad points. In this thesis, three studies proposed improvement to the usability evaluation methods. Brain wave, gaze point, and first impression were firstly proposed to use in usability evaluation. Experiments results show the efficiency of these new proposals. However, there are many works to do in the future to bring the proposed methods to completion. Such as more proposals of quantitative metrics and experiments with more subjects is necessary. In addition, the future work includes refining the evaluation procedure, as well as comparing the proposed method with other available methods.

Acknowledgements

During the course of this work, it is very fortunate to receive lots of assistance from many individuals. I would deeply like to thank Professor Ken-ichi Matsumoto of Software Engineering Laboratory for his precious supports and many valuable suggestions.

I am also very grateful to Professor Masaki Koyama of Information Technology Center for his invaluable comments and helpful suggestions concerning this thesis.

I also wish to thank Professor Katsuro Inoue for the valuable suggestions and warm discussion.

I would like to give thanks to Associate Professor Hajimu Iida of Information Technology Center for the useful leading and heated discussion.

I would especially like to thank President Koji Torii for his continuous support, encouragement, and guidance for this work.

I would like to express my thanks to Research Associate Kazuyuki Shima of Software Engineering Laboratory for his continuous supports, stimulating discussions and very helpful criticism.

I would like to acknowledge also the valuable suggestion of Research Associate Akito Monden of Software Engineering Laboratory for his help and instructive suggestions.

I would like to give thanks to Research Associate Masahide Nakamura of Software Engineering Laboratory for his kindness and helpful advices.

My grateful thanks are due to Rotary Yoneyama Memorial Foundation and International Communication Foundation for financially supporting my doctor course.

I would like to give my special thanks to other members of Software Engineering Laboratory for their kindful help. Specially, I would like to thank my senior of Software Engineering Laboratory: Dr. Makoto Sakai, Dr. Shuuji Morisaki, Dr. Yasuhiro Takemura, Dr. Masatake Yamato, and Mr. Tomonori Kumashiro who gave me kindful help. I also like to thank my junior of Software Engineering Laboratory: Mr. Noboru Nakamichi for his help in this work.

Finally, I would like to thank my family for their endless love and great support.

References

1) T. Asahi, “Research on technology of usability testing of software,” Ph.D. Thesis, Graduate School of Information Sciences, Nara Institute of Science and Technology, Japan, 1998 (In Japanese).
2) W. Barber, A. Badre, “Culturability: the merging of culture and usability,” The 4th Conference on Human Factors and the Web, Basking Ridge, New Jersey, USA, http://www.research.att.com/conf/hfweb/proceedings/barder/index.htm, 1998.
3) S. Card, T. Moran, A. Newell, “The Psychology of Human-Computer Interaction,” Hillsdale, NJ, Lawrence Erlbaum Associates, 1983.
4) A. Dix, J. Finlay, G. Abord and R. Beale, “Human Computer Interaction,” Prentice Hall International, New Jersey, 1993.
5) X. Ferré, N. Juristo, H. Windl, L. Constantine, “Usability Basics for Software Developers,” IEEE software, pp. 22-30, January/February 2001.
6) K. Goto, E. Cotler, “Web ReDesign,” Pearson Education, 2002.
7) E. Galdo, J. Nielson, “International User Interfaces,” Wiley, New York, 1996.
8) J.J. Gross and R.W. Levenson, “Emotion elicitation using films, Cognition and Emotion 9 (1) pp. 87-108, 1995.
9) G. Hofmeester, J. Kemp, A. Blankendaal, “Sensuality in product design: a structured approach,” Electronic Proceedings of the Computer Human Interaction ’96, 1996.
10) J. Hu, K. Shima, R. Oehlmann, J. M. Zhao, Y. Takemura, and K. Matsumoto, “An Investigation of Impressions of B2C web pages in China, Japan, and the UK,” Proceedings of 6th World Multiconference on Systemics, Cybernetics and Informatics, vol. I, pp.83-88, Florida, USA, 2002.
11) J. Hu, Y. Takemura, K. Shima, K. Matsumoto, K. Inoue, and K. Torii, “Analysis of relation between impressions and design of B2C web page,” Proceedings of the 5th World Multi-Conference on Systemics, Cybernetics and Informatics, Vol. I, pp.286-293, Florida, USA, 2001.
12) J. Hu, J. M. Zhao, K. Shima, Y. Takemura, K. Matsumoto, “Comparison of Chinese and Japanese in Designing B2C Web Pages toward Impressional Usability,” Proceedings of the 2nd Asia-Pacific Conference on Quality Software, pp. 319-328, Hong Kong, China, 2001.
13) H. Ikeda, “Guidebook of statistics,” Shin You Sya Inc, Tokyo, 1989.
14) H. Ikemoto, “Operation Evaluation of GUI Using Operation History,” The 10th Human Interface Symposium, pp.447-454, 1994.
15) IBM, “Cost Justifying Ease of Use,” www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/23 (current 2 Jan. 2001).
16) B.E. John, D.E. Kieras, “The GOMS Family of User Interface Analysis Techniques: Comparison and Contrast,” ACM Trans, Computer Human

Interaction 3, 4, pp.320-251, 1996.

17) N. Juristo, H. Windl, L. Constantine, “Introducing Usability,” IEEE software, pp. 22-30, January/February 2001.

18) N. Kawasaki, “Cope with web usability,” Journal of Information Processing Society of Japan, Vol.44 No.2, pp.163-168, 2003 (in Japanese).

19) M. kurosu, M. Itou, T. Tokitsu, “A guide to user engineering,” Kyoritsu shuppan co., ltd, 1999 (in Japanese).

20) J. Kim and J. Y. Moon, “Designing towards emotional usability in customer interfaces trustworthiness of cyber-banking system interface,” Interacting with Computers, vol.10, pp.1-29, 1998.

21) N. Kishi, “Development of Tool for Comparing User Operation History in X Window,” Information Processing Society of Japan, Human Interface Report, No.53, pp.41-46, 1994.

22) T. Kurooka, M. Kisa, Y. Yamashita, and N. Nishitani, “Application of mind state estimation to plant operators,” Seventh IFAC/ IFIP/ IFORS/ IEA Symposium on Analysis, Design and Evaluation of Man -Machine System, pp.16-18, Kyoto, Japan, 1998.

23) R. W. Levenson, “Human emotion: a functional view,” P. Ekmanand, R.J. Davidson, (Eds.) The Nature of Emotion, Oxford University Press, pp.123-126, Oxford, UK, 1994.

24) R. Logan, S.Augaitis, T. Renk, “Design of simplified television remote controls: a case for behavioral and emotional usability,” Proc. CHI’91, Human Factors in Computing Systems, ACM Press, pp.365-369, Boston MA, 1994.

25) S. Y. Lee and M. Nagamachi, “Kansei Human Engineering,” Yangyonggak, Seoul, 1996.

26) T. Musha, H. Terasaki, H.A. Haque and G.A. Ivanitsk, “Feature extraction from EEGs associated with emotions,” Artificial Life and Robotics 1, pp.15-19, 1997.

27) B. Myers, M. Rosson, “Using GOMS of User Interface Design and Evaluation: Which Technique?,” ACM Trans, Computer Human Interaction 3, 4, pp.387-319, 1996.

28) M. Mori, T. Ui, “Research on effectiveness of gaze point movement analysis in screen design,” Office Automation, Vol.16 No.3, pp.49-56, 1995.(in Japanese)

29) D. Minoli and E. Minoli, “Web Commerce Technology Handbook,” McGraw-Hill Companies Inc, London, 1998.

30) J. Nielsen, “Web Research: Believe the Data.” Alertbox, www.useit.com/alertbox/990711.html, 1 July 1999. (current 3 Jan, 2001)

31) J. Nielsen, “Designing Web Usability,” MdN Corporation, Tokyo, 2000. (In Japanese)

32) J. Nielsen, “Usability Engineering,” Academic Press, New York, 1993.

33) J. Nielsen, “Report from a 1994 web usability study,” http://www.useit.com/paper/1994_web_usability_report.html, 1994.

34) J. Nielsen, “User interface directions for the web,” Communications of the ACM, vol.42, No.1, 65-72, 1999.

35) H. Okada, T. Asahi, “GUITESTER: A log-based usability testing tool for graphical user interfaces,” IEICE Trans. on Information and Systems, Vol.E82-D, No.6, pp.1030-1041, 1999.

36) H. Okada, “Usability and evaluation method, system/control/information," Transactions of the Institute of Systems, Control and Information Engineers, the Institute of Systems, Control and Information Engineers, Vol.45 No.5, pp.269-276, 2001 (in Japanese).

37) H. Okada, R. Matsuda, T. Asahi, O. Izeki, “Usability evaluation by simulator correspondence UI test,” Human Interface Society Report, Information Processing Society of Japan, No.54, pp.25-32, 1994 (in Japanese).

38) Y. K. Patrick, “Cultural Differences in the Online Behavior of Consumers,” Communications of the ACM, Vol.45, No.10, pp.138-143, 2002.

39) T. Shinohara, “Web usability rule book,” Impress, 2001. (in Japanese)

40) C. Sugahara, “Making and Evaluation of a Guideline of Experimental Usability Evaluation,” Master Thesis, Nara Institute of Science and Technology, NAIST-IS-MT 9651059, 1998.

41) B. Shneiderman, “Design the user Interface,” Addison Wesley Longman Inc., California, USA, 1998.

42) S. J. Simon, “The Impact of Culture and Gender on Web Sites: An Empirical Study,” The Data Base for Advances in Information Systems, Vol. 32, No.1, pp.18-37, 2001.

43) N. Schwarz, G. L. Clore, “Mood, misattribution, and judgments of well-being: informative and directive functions of affective states,” Journal of Personality and Social Psychology 45, pp.513-521, 1981.

44) K. Torii, K. Matsumoto, K. Nakakoji, Y. Takada, S. Takada, K. Shima, “Ginger2: An Environment for CAESE,” IEEE Trans, Software Eng, 25, 4, pp.474-491, 1999.

45) G. W. Tresse, L. C. Stewart, “Design Systems for Internet Commerce,” Addison Wesley Longman Inc, New York, 1998.

46) H. Tagaito, “The Evaluation of the Usability with the User’s Electroencephalogram (EEG),” Master Thesis, Nara Institute of Science and Technology, NAIST-IS-MT 9751061 1999.(in Japanese)

47) R. Virzi, “A preference evaluation of three dialing plans for a residential, phone-based information service,” Proceedings of CHI’91, Human Factors in Computing Systems, ACM Press, 1992.