Digital Museum Usability:
a cross cultural study with elderly users

Mitwa Kaemba
167 864

6. 2. 2008
Master's thesis
University of Joensuu
Department of Computer Science and Statistics
111 Joensuu, 80101 Finland

Abstract

The digitization of museum wares has seen to the expansion of museum viewership. Taking the form of digital museums, exhibitions are cutting across social, cultural and geographical borders. Digital museum visitors are increasingly differing in background, age and experience all of which need to be considered in the design of such applications. This would ensure ease-of-use by all.

However, members of staff of the National Museum of Japanese History (Rekihaku) identified that elderly users were having problems in the use of digital museum applications running on touchscreen systems. This research therefore took an investigative approach to formally reveal the nature and source of these problems and evaluate the suitability of the current interaction technique for the elderly. The inclusion into the study of the North Karelian Museum with which Rekihaku and the Infotonics Center cooperates allowed for the investigation of the role culture plays in application use and evaluation.

An application with common digital museum features was evaluated in usability tests involving Finnish and Japanese participants. It was found that the application was not usable by the elderly. The study revealed the nature and source of the typical problems that elderly users experience and found the application in violation of established design guidelines. Findings suggest that culture has an impact on the usability evaluation process and raises the question of result comparability.

Things to be considered in designing for a diverse population taking the elderly into account are highlighted. Recommendations are made concerning cross cultural usability evaluations with emphasis on the need for an evaluation framework that would allow for the accurate comparison of results among different cultural groups. The research provided insufficient evidence to conclude and thus recommend the most suitable interaction technique for the elderly as most problems observed were as a result of application design flaws.

ACM Computing Classification System, 1998 version: H.5.2. [User Interfaces]

Keywords: digital museum, usability, culture, elderly, user interaction technique

Contents

1 Introduction

1.1 Overview

Technology and the use of computers today is rapidly increasing, crossing over borders, cultures, age groups and types of industry. It could be assumed that technology is being used in each and every industry ranging from education, medicine, governance and agriculture to mention a few. The study and conservation of art is one field that is also benefiting from the ground being gained by technology. In order to allow for remote access, preserve the appearance of exhibits and expand viewership, some museums have computerized their wares in form of digital museums. A digital museum can be looked at as a computer based reproduction of the physical one. The digital museum stores and displays information pertaining to exhibits and artifacts possibly including their 2- or 3-dimensional representations as well as multimedia. Digital museums may be web based or ran as local applications.

Digital museums have a diverse viewership with visitors being of different backgrounds, ages and cultures. On a seemingly unrelated note, it is estimated in [11] that by 2050 a significant proportion of the world's population will be 65 years and older. This implies that by then most computer and internet users and thus digital museum visitors will belong to this age group. As such, this group of users which was formally ignored is being involved more in the design and development of applications. This is important as it is proposed that elderly users have special needs and cannot interact with systems designed for younger users as easily as designers would like them to.

Digital museums may be of cross cultural content. It is thus possible that its visitors may not only be of different cultures from the application content but also from each other. As cultures hold different values, interpretation of information is different for each. For instance, the cow which is valued for its dairy products and meat in Christian cultures is seen sacred by others such as the Hindu. Careful consideration must be made when developing applications that span different cultures to avoid offending or insulting anyone.

This careful consideration of aspects that are intrinsic parts of the user are tied together in a concept known as User Centered Design. This concept has also be coined the term Usability Engineering, usability referring to the effectiveness, efficiency and user satisfaction of a system [22]. It is on this concept that this study is built.

Through collaborative efforts between the North Karelian Museum (Joensuu, Finland), Infotonics Center (University of Joensuu, Finland) and the National Museum of Japanese History (Sakura, Japan) a usability study was proposed to investigate the usability of digital museum applications deployed in the two museums. The National Museum of Japanese History which exhibits materials, replicas and restorations of artifacts from prehistoric and ancient periods to modern times is made up of five physical galleries. It has also digitized some information, an online version being available at [37]. The North Karelian Museum holds two types of exhibitions; a permanent one displaying local history and folk-tradition and a periodically changing one which is usually owned by other museums. A digital form was created for a particular exhibit in this museum known as the 'Shades of Color' which can be found at [52], the museum as a whole has not launched a digital museum yet.

1.2 Problem Statement

The most common interaction technique in the National Museum of Japanese History is the touchscreen system. It was identified by Rekihaku staff that elderly visitors were having problems in the use of digital museum applications particularly with the touchscreen, usually showing frustration and eventually giving up on tasks they were carrying out on the computers. The purpose of this study was therefore to evaluate the usability of a digital museum application on a touchscreen system with elderly users, to formally reveal the problems being experienced and the reasons behind them. As an added dimension it was decided that the same study be conducted in the North Karelian Museum. This was seen necessary because through their collaboration, digital exhibitions from Rekihaku have and will continue to be displayed in the North Karelian Museum and vice versa. Also, this would allow for the identification of the effect of culture on the use of technology, seeing if the problems being experienced in Rekihaku are culture dependent. It was hoped that the results of the usability study would also guide the selection and implementation of future applications and interaction techniques in both museums that are suitable for all.

1.3 Study Objective

Having identified the scope of the problem it was possible to define the main objective of the study to be:

'To investigate the usability issues associated with a digital museum application running on a touchscreen system that affects its use by the elderly.'

Specific research questions that were to aid the fulfilling of this objective were as follows:

- What problems do elderly users face during their use of software applications?
- What problems do elderly users face in the use of the touchscreen as an interaction technique?
- Does culture have an effect on the use of technology and its usability evaluation?
- What interaction technique would be most suitable for elderly users?

It was hoped that these research questions be answered through the usability study described in later sections so as to achieve the overall objective.

1.4 Thesis Organization

The rest of this thesis is organized into six chapters. A review of literature for the understanding of key concepts is conducted in Chapters 2 to 5. Chapter 2 defines Usability highlighting its benefits and beneficiaries. Also, this section compares and contrasts the five most popular usability evaluation methods. Chapter 3 identifies who the elderly are, explaining the age related impairments they experience which affect their use of technology and outlines design guidelines recommended in designing applications for this age group. In Chapter 4 I look at the impact culture has on the use of applications and their usability evaluation. The details of the usability tests conducted in Finland and Japan can be found in Chapter 5, which also contains an analysis of results. Chapter 6 shares important things learned and thus to note when conducting usability tests with the elderly of different cultures. Conclusions are drawn and recommendations made in Chapter 7.

2 Usability

This section aims at defining key concepts in the area of usability. It begins by outlining the different definitions that have been pegged to this term, describing the various schools of thought. This section also explores avenues such as the benefits and beneficiaries of a 'usable' system and how usability can be ensured and measured through different evaluation methods.

2.1 Usability defined

Prior to the birth of Usability as a field some authors such as [8] argue that focus was mainly placed on Human Computer Interaction (HCI) when it came to user-focus in system design. Usability which was mainly heard of in terms of testing was seen to be encompassed in software tests of the user interface which employed the use of the software community's guidelines of good practice.

It is claimed in [62] that the pioneers of usability were John Whiteside and John Bennett of Digital Equipment Corporation and IBM respectively. In 1988 [62] they published a paper proposing the concept together with Holtzblatt. In this paper they described new ways of approaching system design which place more focus on the context in which the user works. They argued that the yardstick of measuring usability is ultimately the users' experience in the use of a product. In a definition by Dumas et al. [9] usability is also looked at in terms of the users' experience, where a usable system is one which allows users to use it quickly and easily to accomplish their own tasks. Nielsen [38] in his definition states that usability comprises learnability, efficiency, memorability, errors and satisfaction. Similar to this definition is that given by Rubin [49] who sees usability to encompass learnability, satisfaction, effectiveness and usefulness. Most of the definitions in literature basically point to the same concept though the definition has evolved over time.

True to its purpose of standardizing terms, the International Standards Organisation (ISO) cemented and reiterated this notion in the establishment of the International Standard ISO 9241-11 [22] which gives guidance on usability. The standard defines usability as the

'Extent to which a product can be used by specified users to achieve

specified goals with effectiveness, efficiency and satisfaction in a specified context of use'

The terms effectiveness,efficiency and satisfaction are further defined as

- Effectiveness: Accuracy and completeness with which users achieve specified goals
- Efficiency: Resources expended in relation to the accuracy and completeness with which users achieve goals
- Satisfaction: Freedom from discomfort, and positive attitudes towards the use of the product

This is the definition of usability that is adopted in this study. It is assumed that having addressed these three factors, what is developed is a usable product. However, in practice it is usually the case that evaluators of usability test only one or two of these factors and make an umbrella claim on the usability of a product [17]. This means that they assume that these three factors are somewhat related where, if one is addressed so is another. Experiments conducted to investigate these claims such as [17] and [39] prove to the contrary that each of these factors are independent and should each be included in usability tests. In an experiment involving information retrieval it is statically proven that efficiency (which is measured by task completion time) and effectiveness (indicated by quality of solution) are not correlated [17]. In the event that they are, this correlation is negligible. In a study of the relation between efficiency and user preference (satisfaction) it was found that in 25% of the cases users did not prefer a system that they were more efficient in using [39]. The percentage though less than half, still shows that the dependancy between efficiency and satisfaction is questionable.

It is for this reason that for the purpose of this study usability will be looked at through all three dimensions of efficiency, effectiveness and user satisfaction.

2.2 The benefits and beneficiaries of usability

Having a clear definition of usability, one may still ask what the point of all of it is. After all usable or not, a determined user can use a product to

complete a task, producing a low quality solution but a solution all the same. What then is all the hype around usability?

We could consider a printer produced by a hypothetical company X. This printer is bought with company funds by user, Mary, who has a report to print out and give to her superior in 2 days. The printer arrives without an instruction manual, button labels or even an interactive display to show printing status. Three days later by virtue of her experience with printers Mary finally discovers how to operate the printer. She prints out the report only to find that the font is too faint. As no buttons are labeled and the report is already late she submits it as it is. It is rejected by her superior and she takes another 3 days to figure out how to change the printer settings. What Mary does not know however is that this report was to serve as the basis for an application for funding to keep the company for which she works operating. They fail to meet the application deadline, the company loses funding, Mary is out of a job and her superior demands the return of the printer to its manufacturers who in turn are losing business as they have a complete production line of such printers which will potentially be returned as well.

This hypothetical situation though portrayed in an extreme manner may well be the situation with an information technology community that ignores usability. Had the manufacturers of the printer paid close attention to usability factors the scene may have played out differently. Had they ensured effectiveness, efficiency and user satisfaction by labeling the printer, including a user manual, having a toll free support phone number pasted on the box or even set the default printer settings to allow for printing of standard font appearance alot of unnecessary time and resources which were wasted could have been saved.

The moral of the story being that the beneficiaries of usability are both the users and manufacturers of the products. The overall benefits of usability include;

1. Increased Productivity: Landauer [29] claims that on average products have 40 flaws resulting from poor design (that ignores usability) that hinder users' progress in their use. He further claims that the resulting cost of lost productivity can be up to 720%. Reiterating the benefits of usability he states that productivity would increase 4 to 9% per annum if the usability factors are carefully taken into consideration during design.

2. Decreased costs of development and support: If a product is designed with usability in mind it potentially eliminates the need for its redesign. Also with a usable product, users can carry out their intended tasks with little or no support. These savings in cost can allow manufacturers to channel funds into alternative ventures.

3. Increased Return on Investment: With a usable product in the market a manufacturer may stand head and shoulders above its competitors. Karat [27] emphasizes how usable products are most desirable and differentiate themselves in a highly competitive market. He further states that ' ...those (companies) committed to ease of use do more than meet customer expectations, they can actually exceed anticipated earnings.'

4. Increased customer satisfaction: It is no doubt that a user having completed a task using a reasonable amount of time and other resources to produce quality results will be highly satisfied with the product in question. Research such as [4] has shown that with emphasis on usability, user satisfaction ratings can increase by as much as 40%. It follows also that a satisfied customer is a returning customer. Satisfied with a product from a particular company, a customer will not hesitate to do business with that company again should the need arise.

The benefits of usability can be summarized in a statement taken from [21] ; 'developing easy-to-use products makes business effective. It makes business efficient. It makes business sense.'

2.3 Ensuring usability through Usability Design and Evaluation methods

Professionals in the field emphasise the need to focus on the user during design in order to ensure usability. Gould and Lewis [18] claim that 1. Focusing early and continuously on users, 2. Integrating consideration of all aspects of usability, 3. Testing versions with users early and continuously and 4. Iterating the design, leads to the development of usable products. Dumas [9] makes statements (amongst others) that to ensure usability users have to be involved throughout the design process, usability and users' needs have to be the driving force behind design decisions and commitment has to be made to making technology work for people. Others such as Jokela *et al.* [25] go

further to state that the paradigm of user centered design was established for the sole purpose of improving usability of software and information systems.

In reviewing literature it was noticed that no clear distinction could be made between design and evaluation methods. That is, the methods considered to be design methods by some were addressed as usability evaluation methods by others. This is understandable as emphasis is placed on the use of these usability methods from early design phases through to late testing phases of development. What remains clear however among them all is the need to set quantitative usability goals which will direct the (design/evaluation) process.

Therefore, for the purpose of this study an umbrella term, Usability Design and Evaluation methods, will be used to refer to the methods that can be used to ensure usability. Methods falling under this term have by others been divided into broad categories. One such categorization was made by Bednarik [2], who places them into the categories of usability testing, usability inspection and usability inquiry.

For uniqueness and argument's sake the approach taken here is to compare and contrast the most popular usability design and evaluation methods (from hereon referred to simply as methods). In the past, research has been conducted in order to determine the most popular methods. In a survey initiated by Gunther et al. [19] it was discovered that the most successful methods are Usability testing, Paper or other prototyping and Heuristics evaluation. These results were obtained in a web survey on the opinions of 100 usability practitioners. According to Rohn et al. [48] the most popular methods are Heuristics Evaluation, Field studies, Surveys, Usability testing, Focus groups and Usage scenarios.

Furthering research in this area were Mao et al. [33] who conducted a survey in order to determine which methods are known to work best in practice over those that do not. Obtaining 103 responses to an email questionnaire, they (in [33]) site iterative design, usability evaluation, task analysis, informal expert review and field studies as the most commonly used by the respondents. The respondents also considered these methods, apart from informal expert review, as having the most important impact in practice. Mao et al. [33] identify low cost as a reason for the common use of methods such as informal expert review, which they claim do not necessarily work well in practice.

More informally, usability specialists (e.g. [57]) have described the five most popular methods to be Focus groups, Usability testing, Card sorting, Participatory design, Questionnaires and Interviews.

There are five methods that are common to these and other pieces of literature even though they may be named differently in each case. These methods are Usability testing, Heuristic evaluation, Customer interviews, Focus groups and Questionnaires & Surveys. These five methods are the ones considered as the most popular in this review.

Literature reveals that partial comparisons of different methods have been made. Jääsko and Mattelmäki [23] compare Contextual design (observation) and Self Documentation & Interviews. They base their comparison on a case study of two companies; Thermo Clinical Labsystems and Instrumentarium Corp. Datex-Ohmeda division. The two companies made use of the methods for the first time and were successful in their quest. The authors look into how each method is carried out and the kind of data that they produce. It was established that Contextual design was useful in developing work flows and understanding interaction between people and their environment. However Contextual design was not able to gather more personal information on users such as personality, which Self-documentation did. It was concluded that the two methods were complimentary.

Straub [54] compares Expert reviews to Usability testing. She argues that while Expert reviews identify fundamental challenges within the user experience Usability Testing highlights mismatches between the user and product model. Other authors such as Ryu et al. [50] describe subsets of methods, and give their comparative advantages. In their article they compare the use of laddered grid, focus groups, usability testing, and expert reviews in the development of a menu system for mobile phones. They do not however detail each method nor give advice on which is most appropriate.

In as much as this topic appears to have received a lot of attention over the past couple of years, comparison among the five most popular methods to the best of my knowledge has not yet been conducted. This section of the review may therefore also serve to extend the various comparisons that have been made, by providing an integrated and comprehensive version. Aspects being compared and contrasted are the basic principles, type of data collected, skills required, circumstances under which the method could be adopted and the weight of associated costs.

The basic principles will give an overview of what the method involves. Having carried out the various steps outlined in the basic principles, it would benefit the reader to be aware of the type of data these steps collect. Some special skills may be required to carry out certain methods therefore it is important for an individual to be aware of these to avoid rude awakenings during its use. In the event that a single method cannot be proved to be the overall best it may be important to determine in which kind of settings each method works at its optimum. Costs also play a major role in the selection of an appropriate method to adopt for design. It is common knowledge that limited resources lead organizations to select methods whose costs match their profile. As such, comparison of costs is important as well.

2.3.1 Usability Testing

Usability testing (UT) involves the selection of a representative group of users, whose use of the product is observed. These users are presented with tasks that are typical to the use of the product and are observed by the design team. Nielsen [41] recommends the use of only 5 users as he discovered that after observing 3 users you begin to observe the same mistakes thus wasting resources. During observation, the development team collects quantitative data such as time taken to complete tasks, usability data such as number of mistakes made and data on the users' attitude during use. Data on the users' experience can be collected from the user by asking them to think aloud, or allowing them to discuss in pairs as they complete tasks [53].

During Usability testing a great number of severe problems that are directly experienced by the user are identified. As such, there is no need to sort or filter the problems according to their predicted impact on users. It is indeed necessary for evaluators to be sensitive to usability issues that may arise. It is thus recommended that Usability specialists be recruited for the purpose of usability testing. The specialists should be given time to familiarize themselves with the product.

This method however can only be used when there is some form of the product already in place. This form can range from a simple prototype to a fully functional product; as such can be deployed from early through late stages of development. Costs incurred using this method are tied to costs of recruiting/training a Usability specialist, rewarding participants, time spent identifying and rallying participants and time spent on familiarizing the specialist with the product.

2.3.2 Heuristic Evaluation

In **Heuristic Evaluation (HE)** a group of evaluators are presented with the product to inspect. They compare various aspects of it to a list of heuristics. This list of heuristics is a set of rules that govern what the product (usually an interface) should possess. The person monitoring the evaluation need not make inferences about the actions made by the evaluator but merely collect the comments made.

The data collected reflects mainly the evaluators' opinions of what is wrong with the product and may not be exhaustive of the typical use of the product. Nielsen and Molich [44] recommend the use of about 5 evaluators with varying levels of expertise. Not all the evaluators need to be experts in the product's domain and may not even be representative of the target users. In an experiment they conducted they discovered that even 'poor' evaluators could uncover some of the most difficult problems.

The costs of heuristic evaluations are relatively low as they do not strictly require the participation of usability specialists. However, research (e.g. [24] and [44]) has shown that the involvement of usability experts increases the number of problems identified but also significantly raises costs. Therefore, heuristic evaluation is suitable for projects that have limited resources. This method can be employed in various stages of development but is mostly used in early stages so as to get feedback on preliminary design.

2.3.3 Interviews

An **interview** is a method by which one person tries to extract knowledge or information from another by asking questions [31] and may be used across various disciplines for various purposes. In interviews for usability evaluation, stakeholders are asked questions in order to understand their current work practices. The interviewer is free to ask anything that is felt would be relevant to design or the investigation of usability. Emphasis is placed on an interview being a discussion rather than of question-answer format. In conducting an interview the interviewees are first selected possibly by conducting an initial round of interviews. The interviewer is to make an outline of the questions that he intends to ask beforehand so as to direct the conversation from a clear understanding of the project focus. It is recommended that two people from the design team are present during the interview [5], this way they can

support each other in asking questions and interpreting the results.

Data collected includes details about the tasks that the client carries out, the purpose of these tasks, problems they face and how those problems are currently tackled. From this data it is possible for the design team to identify areas of weakness and thus opportunity for improvement. Interviewers should be able to decipher meaning from customer responses and be vigilant in their note taking as such interviewers should be trained because if not then the interview may not produce any valuable information.

Costs related to interviews are relatively low since only two people require training and it is usually not necessary to pay participants, as they are the actual stakeholders. This method is thus suitable in situations where resources are limited and can be used in initial stages of development for the elicitation of requirements.

2.3.4 Focus Groups

Employing the use of **Focus groups** involves inviting a group of about 6 to 8 people that are representative of the user population and discussing various issues with them. The moderator of the focus group sits down with the participants and guides the conversation to cover all topics that were initially planned. During Focus group sessions a general overview is obtained about the users' feelings and attitudes toward the envisioned product, on current work practices and/or on an initial version of the product. With focus groups it is said that rather than a one on one interview, a group of users can give a wide variety of views, each person filling in where someone else falls short.

The moderator of the focus group needs to be someone who has detailed knowledge of the product as well as experience as a moderator. This experience comes in handy for instance when the moderator needs to determine if an argument is disrupting the session or leading to a new and innovative idea. This person need not necessarily be a usability specialist. For this reason as well as the fact that they are quick and fairly easy to assemble [61], Focus groups run on low costs. On a limited budget this method could work well. Focus groups are mainly used as an initial step in the design process in order to uncover issues that could be explored in more detail through other methods.

2.3.5 Questionnaires & Surveys

Questionnaires administered in **surveys** present a sample of potential users with a predefined set of structured questions regarding the product. These can be used prior to design in order to elicit requirements or after the release of an initial version in order to obtain feedback. Questionnaires can be administered via the Internet, traditional post or even in person.

With unbiased questions, questionnaires have been known to provide good statistical data. It may be necessary though that the person analyzing the data have sufficient background knowledge in Statistics to ensure that the results are reliable.

When it is necessary to get information from a large number of users this may be the best method to use. It also eases the obtaining of information from users in remote areas. Apart from that related to the printing/ deploying of the questionnaire this method does not incur much cost.

The information presented above is summarized in Table 1. The Rank* is assigned in such a way that 1 is the most expensive and 5 the least expensive method.

Table 1: Comparison of popular methods

Method	**Type of data**	**Skills required**	**Costs(Rank*)**	**Suitability**
Usability test	Usability problems	Usability specialist	High(1)	-Flexible budget. -All dev stages.
Heuristic evaluation	Usability problems	None(Though Usability specialists improve results.)	Considerably low (2)	-Fairly flexible budget. -Early dev. stages
Interviews	Details on current work practice	Trained interviewer	Considerably low (mostly related to time invested) (3)	Restricted budget but ample time available
Focus groups	Users' attitude towards work practice and envisioned product	Experienced moderator	Low(4)	-Restricted budget -Early dev stages
Question -naires & surveys	Statistical data; type of which depends on reqs. of team	Statistical analysis of results	Low(5)	-Restricted budget -Inaccessible users. -When large sample is required.

2.3.6 Summary

Questionnaires and surveys being the only method that provides direct statistical data at extremely low costs could be assumed to be the best available option. However, with the use of questionnaires it is impossible to get information related to the actual use of the product without presenting the users with some form of the product to try out. Presenting the users with some form of a product as earlier explained could be seen to constitute either Heuristic evaluation or Usability testing. These two methods differ on account of the people that are involved in the evaluation. In HE, evaluators need not necessarily be representative of the actual users where as in UT they are.

It appears that UT and HE emerge as the 'best' methods to adopt among those reviewed and it is interesting to note that this somewhat theoretical way of arriving at this conclusion is consistent with empirical findings of research into what works best in practice (e.g. [19] and [33]). Even though UT is of high cost, its benefits seem to compensate for this. With UT, the team is assured of accurate and more reliable results as it directly involves the user. In comparison to HE, UT uncovers fewer problems (only by a small margin however [30]) but identifies the most severe. On the other hand HE is known to identify more problems, including some of the severe ones. However since the evaluators are not necessarily the actual users HE has been known to uncover problems up to 50% of which are irrelevant to the user [1].

Decisions on which method to adopt between these two could be based on a project's budget and/or the phase of development that the application has reached. In early stages of development, HE could be used to identify and eliminate the greater part of the usability problems, but this would have to be accompanied with some form of 'low-cost' user testing prior to product release to identify other problems that may have been missed. If on the other hand, one has a flexible budget and seeks consistency in the method to be used across all phases of development, UT would be a better approach. A recommendation would therefore be to look into ways in which the costs related to UT can be reduced. One such way could be for instance to officially publish guidelines on conducting a UT, so that even people who are not fully fledged usability specialists can be able to successfully conduct a UT session.

3 Elderly Users and Technology

From the previous section on usability it seems possible to determine the most appropriate method to adopt in the evaluation of usability. Irrespective of the method selected, one of the first steps in usability evaluations is to identify the participants of the study. As this study focuses on elderly users, it was seen relevant to review literature written on their use of technology. This section investigates who elderly users are exactly, age related impairments that hinder their use of technology and design guidelines established for applications targeted at this group.

3.1 Who are the elderly?

According to the American Heritage Dictionary [56] the elderly are those that are past middle age approaching old age or those far advanced in the years of their lives. As dictionary definitions of the word still do not give a distinct age group it was necessary to review various pieces of literature to determine which age group constitutes the elderly. According to the European Commission [11] three persons out of ten will be aged 65 years or over by 2050. It would as such seem reasonable to use this age group in studies relating to the use of technology as they will make a siginificant fraction of technology users in the future. Further, studies undertaken in the past that involve elderly users such as [46] and [51] on average use the age group of 65 and above as well. It is assumed that by this age individuals begin to display significantly different behavior in their interaction with technology than younger users.

3.2 Age related impairments; hindrances to the use of technology

It is common belief that elderly users have a dismissing attitude towards the use of technology and are incapable to keep up with the dynamic waves of change. In literature reviewed however it was noticed that this incapability has not been empirically proven. Infact, the elderly have been identified as potential enthusiastic computer users. In a study conducted by Ogozalek and Praag [46] it was discovered that in as much as these social conclusions have been drawn, there were in actual fact no differences between the performance

and preference of elderly and younger users in carrying out a computer based composition task. The task involved the use of a keyboard editor and a simulated listening typewriter for the composition of letters. Supporting these findings was another experiment [32] which compared performance between younger and elderly users in a computer based virtual driving environment. It is even concluded in this study that virtual reality is an appropriate technology for use by elderly users.

This in itself discounts any myth of elderly users' incapability to use technology. It should be noted, however, that the elderly do face barriers to computer use that are unique to age. Particular to the use of the internet, these barriers have been categorized (as in [10]) into 3; normal age-related changes, cohort differences, and stage of life. Even though these barrier categories are not investigated in this study it can be noticed that all these barriers are those that cannot be avoided and occur naturally. Observation of the elderly shows that they mainly experience age related impairments in the areas of vision, movement (as a result of motor function) and memory, abilities that are the most used in user interaction with technology.

3.2.1 Vision

It is known that visual capabilities decline with age. Elderly people are not able to perceive objects as accurately as younger ones. This makes it difficult for the elderly to read text off a computer monitor especially if it is small and against the standard white web page background [28]. It may be assumed that this decline in vision may also hinder their use of certain computer input devices. One such device is the eye tracker which uses gaze as input. The eye tracker may require fixation on a particular point on the screen to invoke certain actions but Fukuda and Bubb [16] claim that fixation for elderly users takes longer than it does for younger users. This may prove to be a problem in gaze driven input devices as the user may become frustrated in its use if their gaze takes a long time to invoke an action. On the other hand, Murata [36] states that with the age-related decline in motor function, the eye tracker leads to faster pointing times in comparison with mouse input. This decline in motor function is described next.

3.2.2 Motor function

The coordination between the mind and body governing movement and posture is what is referred to as human motor function (or control) by both [16] and [6]. It has been established that elderly users tend to complete movements much slower than younger users which could affect their use of scroll bars or image maps [28]. Apart from being slow in movement it is pointed out in [16] and [6] that elderly users experience difficulty in making fine motor movements such as cursor positioning, which would create difficulties in the use of pointing input devices. It is for this reason that in [6] given a particular point-and-click and click-and-drag task the elderly had to be slower in movement in order to achieve the same (or more) accurate results as younger users in the use of a mouse and trackball. It is was also found that the mouse requires more effort in terms of force application and as such the authors suggest that a trackball may be a better option for elderly users performing frequent, repetitive actions for sessions of long duration. With a decline in motor function, it may also be thought that the elderly experience a significant amount of difficulty in the use of direct manipulation devices such as the touchscreen. However, research has shown that the touchscreen works particularly well with the elderly [58].

3.2.3 Memory and Intellectual skills

Apart from vision and motor function impairments, the elderly also suffer from a decline in the operation of their memory and ability to absorb new information. Zajicek [64] explains that this loss is mainly related to short term memory, where knowledge and memories accumulated over the years remain relatively unaffected. In the experiment she carried out, she found that elderly users had difficulty in remembering sequences of actions and the results that they produced in tasks. Meyer *et al.* [35] discovered in an experiment that elderly users had a difficult time in remembering which web pages they had seen or how they arrived at a particular page. Some authors (e.g. [64]) for this reason recommend the use of voice help though others argue that the benefits of audio output are lost to older people who are also hard of hearing or operating in a noisy environment [28].

3.3 Usability guidelines in designing for the senior citizen

It is apparent that the impairments that come with ageing affect the use of computer applications by the elderly. However this is rarely taken into account as research has shown that present day applications are twice as usable by younger than older users [43]. This is attributed to the fact that designers of these applications are usually young and unaware of the impairments that characterise old age. For even though most seniors are retired without having used computers during the course of their careers, poor design makes it even harder for them to use applications.

Design guidelines that would increase the usability of applications by the elderly have been proposed. Authors such as Nielsen [43] and Zhao [65] recommend the use of at least 12-point font for text, with a provision allowing users to increase text size whether or not the application targets senior citizens. For increased readability and clickability they recommend the use of even larger text for hypertext links which should be well spaced to avoid erroneous clicks. For web applications different colors should be used to distinguish visited and unvisited links to prevent users from losing track of where they have been.

A more comprehensive list of guidelines is provided in [28]. The guidelines are categorized into; Target design (e.g. older adults should not be expected to double click), Use of graphics (e.g. graphics should be relevant and not for decoration), Navigation (e.g. provide location of current page), Browser window features (e.g. avoid scroll bars), Content layout design (e.g. language should be simple and clear), Links (e.g. links should not be tightly clustered), User cognitive design (e.g. provide ample time to read information), Use of color and background (e.g. blue and green tones should be avoided), Text design (e.g. avoid moving text), Search engine (e.g. search engines should cater for spelling errors) and User feedback and support (e.g. provide online help tutorial).

It is believed that adherence to these guidelines would improve the usability of applications for the elderly user.

3.4 Summary

From this section it has been established that the elderly are generally individuals in the age group of 65 years and above. The impairments that they experience as a result of their age have also been described. These impairments have been discovered to hinder their interaction with technology. However it has been duly noted that elderly users would also like to reap the benefits that technology has brought with it. For instance in [16] it is said that elderly users (would like to) use technology for the same things that younger users do.

The elderly belong to a special group of users with special needs just as those with natural disabilities e.g. the blind or the deaf. In a society that is committed to 'digital accessibility' for all, it would be best to find ways that could improve the elderly users' experience in the use of technology. One way is to adopt the design guidelines that have been proposed by authors as highlighted above. Another is also to take into consideration these age related impairments in the selection of an interaction technique to be used.

4 Culture and Technology

Hofstede [20], a renowned cultural anthropologist defines culture to be 'the collective programming of the mind which distinguishes the members of one group of people from another'. This is a definition that has been adopted in most of the pieces of literature on culture and technology. Since literature reviewed in this section references these dimensions, their definitions are included here.

Cultural dimensions [20]:

- Power distance (PA)-the extent to which less powerful members expect and accept unequal power distribution within a culture.
- Individualism vs. Collectivism- individualism in cultures implies loose ties; everyone is expected to look after one's self or immediate family but no one else. Collectivism implies that people are integrated from birth into strong, cohesive groups that protect them in exchange for unquestioning loyalty.
- Masculinity vs. Femininity (MAS)- Masculinity and femininity refer to gender roles, not physical characteristics. There is a traditional assignment to masculine roles of assertiveness, competition, and toughness, and to feminine roles of orientation to home and children, people, and tenderness.
- Uncertainty Avoidance (UA)- the extent to which people feel anxiety about uncertain or unknown matters, as opposed to the more universal feeling of fear caused by known or understood threats.
- Long- vs. Short-term Orientation - orientation to practice and the search for virtuous behavior vs. orientation to belief and the search for truth.

With culture being an intrinsic part of one's being it may be assumed that it is one of the major considerations in the design of computer applications through the user centered design process. However, authors such as Sun [55] claim that culture has been ignored in the design and usability evaluation process. In this article it is claimed that the engineering approach to

usability evaluation is adopted in most situations which treats users as test subjects from a mechanical view. The author also reviews that this approach decontextualises the use of products. Sun therefore states that to explore cultural factors new frameworks are needed and goes on to propose a model for Cultural Usability. Despite Sun's claims of culture being ignored or wrongly addressed the fact still remains that culture is an area in human computer interaction that is steadily gaining popularity [26]. As this study relates to the use of the digital museum in both Finland and Japan, countries of two different cultures, it was seen necessary to be aware of the potential influence that culture has on technology use and its usability evaluation.

4.1 Influence of culture on technology

In literature reviewed, I noticed that culture and technology are usually talked about in light of user interfaces and their design. This is understandable as irrespective of the input device used, interaction between the user and an application is mainly via the interface. Further, it is assumed that focus has taken this direction because of the cross-cultural use of global interfaces such as the World Wide Web. Authors such as Ford and Gelderblom [15] have even tried to make deductions about human performance by studying cultural differences in the use of web sites. In the tests conducted insufficient evidence was found to conclude that culture had any impact on the users' accuracy, speed and satisfaction with an interface.

More conclusive results were reported by Jantawan and Norcio [45] who studied the effects animated graphic colors has on attention and perceived usability across cultures. They found that differences existed between users of the two cultures of America and Thailand. They suggest that culture influences overall performance (i.e. Thai participants took longer to complete tasks), overall retention (i.e. American users were observed to retain the color of website banners more than the Thai users) and overall self-reports on usability (i.e. Thai users were less inclined to give negative comments about the interface).

Similar research was conducted by Fang et al. [13] and Rau et al. [47]. In [13] it was established that differences existed in user satisfaction and the steps performed to complete tasks between Chinese and American participants. In [47] it was found that 1) lack of localized content about services and products affect users' browsing experiences, 2) Asian users need more

time browsing English content and 3) the information on the website used in the experiment was too rich for non-native English speakers. Their findings emphasise the need to take culture into consideration during design.

Taking a more comprehensive approach Marcus and Gould [34] study the effects of all the cultural dimensions defined by Hofstede [20] on web user interface design. They find significant differences in the design of interfaces across different cultures. For instance, in studying masculinity vs. femininity it is stated that masculine cultures focus on assertiveness, competition and toughness while the feminine focus of orientation to home and children, people and tenderness. It was found that Japan (which has the highest masculinity index [20]) has websites that orient search portals toward a specific gender. They compare Japan to other countries with a low masculinity index such as Sweden whose websites make no distinction in gender or age. Other dimensions studied in this research are power distance, individualism vs. collectivism, uncertainty avoidance and long vs. short-term orientation, all in which differences across cultures were found.

It is seen from here that culture has a great impact on user interfaces and their design. One would infer a similar impact of culture on the usability evaluation of these applications. How culture affects the usability evaluation process is studied next.

4.2 Effects of culture on usability evaluation

Apart from the considerations highlighted in Section 2.3, culture also needs to be considered in the selection of a usability evaluation method. It would be incorrect to assume that an evaluation method proven to work in one society will automatically work in another. For instance, the co-discovery approach to usability testing has been proved not to work well in Japan [14]. It was discovered that when people of different status were put in the same room, those of a lower status (particularly women) talked less. Still in Japan, it was also found that women spoke softly and as such it was recommended that great consideration be placed on the quality of the microphone. Also, as the Japanese are a society that are not accustomed to expressing feeling and emotion, questions regarding how they 'felt' about a product did not work well among them.

Research conducted in Malaysia [63], a country also in the Asia Pacific

region showed that a cultural factor affecting usability evaluation was power distance (i.e. the effect to which less powerful members ... expect and accept that power is distributed unequally [20]). It was because of this that in the think-aloud and questionnaire approach that was used in the research, individuals of lower rank than the experimenter were found to give neutral comments compared to the harsher ones made by those of higher rank. The author also suggests from experience in Malaysia that it is ideal to conduct tests in a native language in order to reduce cognitive load (for example in the use of a language other than one's own in the think-aloud process).

In an investigation of the cross-cultural applicability of user evaluation methods by Evers [12] she found that different usability evaluation methods pose different problems within different cultures. Her study involved participants from America, United Kingdom, the Netherlands and Japan evaluating a website. A summary of her findings is shown in Table 2.

Table 2: Methods that posed problems for different cultures (indicated by X).*Adapted from [12]*

Method	**US**	**UK**	**Netherlands**	**Japan**
Questionnaire	X			X
Task observation (think aloud)		X	X	X
Interview	X			

It was found that Americans and Japanese had difficulty in answering the questionnaire. For the Americans this was attributed to the fact that it is a culture that does not weave into its people a common understanding of their cultural background. For instance, when asked which language they spoke some ticked both English(American) and English(British) or referred to themselves as white/christian as opposed to the commonly used American. For the Japanese it was mainly attributed to the ambiguity of the questions that was introduced by translating the questions from English to Japanese. Participants from UK, Netherlands and Japan experienced problems with Task observation. The British and Japanese needed alot of reassurance for them to finally feel comfortable to think aloud. The author recommends the use of well-matched pairs or focus groups in these two cultures. The Dutch tended to use alot of humor and sarcasm making the extraction of relevant information difficult. Interviewing Americans caused problems in the sense that they gave responses that were inconsistent with their observed behavior. They felt more inclined to give the 'right' answer than that which

was true. This which was attributed to their competitive nature could dent the reliability of a study making use of interviews.

Vatrapu and Quinones [60] explore the concept of international usability. In their research they evaluated a website with a group of Indian students using structured interviews. The participants were divided into two groups, one interviewed by an Indian and another by an Anglo-American. It was found that more usability problems were uncovered and more suggestions made to the Indian interviewer (i.e. an individual of the same culture) than to the Anglo-American. Vatrapu and Quinones attribute their findings to India being a high power-distance country. It can also be explained by the fact that Indians (coming from a country with a low individualism index [20]) tend to stick together rather than taking an individualistic viewpoint and as a result, a more objective stance. In all, the article concludes that 'Culture might influence the efficacy of a usability method that involves a high degree of social interaction' such as structured interviews.

4.3 Summary

From this section the influence of culture on technology is apparent. Different things relay different meanings across cultures especially those of the Western vs. Eastern parts of the world. It is necessary therefore in trying not to offend people from different groups, to take their culture into consideration in the design and evaluation of applications. Also, as seen from above it would be important to be aware of the cultural values of a group one intends on including in a usability study. Once that is done it would be possible to select or even tailor a usability evaluation method that is most appropriate for a given culture.

5 Usability Testing of a digital museum application with elderly users

A usability evaluation was planned and executed to reveal the nature and source of the problems reportedly experienced by elderly users of digital museum applications. The application selected for this purpose was one which was in use at the North Karelian Museum as part of the Shades of Colors exhibition held between 25.01.2007 and 02.09.2007. The application was web based running locally on a touchscreen system and included some multimedia developed at Rekihaku. The touchscreen was 14" in size and had a screen resolution of 1024 x 768 pixels. This screen was the interface to a computer running on a Windows XP operating system. In Finland the touchscreen was enclosed within a casing about 1.2 metres high, and faced upwards allowing users to stand during its use. The Japanese participants on the other hand had a seated 'desktop' arrangement. The two setups are illustrated in Fig. 1.

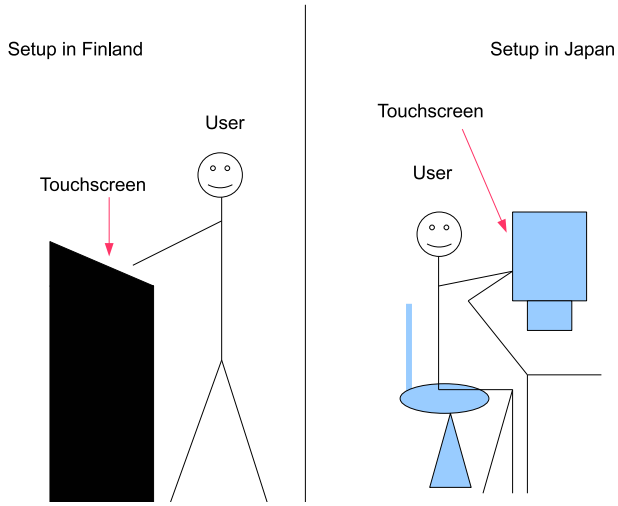

Figure 1: Setup of touchscreen system.

After careful consideration of the five most popular evaluation methods described in the Usability section of this document, it was decided to make use of Usability Testing. Taking into account the cultural dimensions of the two countries, discussions were held with native Finnish and Japanese individuals to make sure the questionnaires and tasks were appropriate. Additionally, a representative from each culture was present during each test.

As the implementation of innovative interaction techniques was foreseen in both museums it was hoped that the findings of this usability test would also serve as a starting point in the selection and implementation of an innovative interaction technique suitable for all.

5.1 The test plan

Prior to the test a thorough plan was put into place (a detailed description of which is attached as Appendix A). As earlier reviewed, when the elderly are spoken of it usually refers to people who are 65 years and above. Since there are no hard and fast rules to this definition, and for the sake of availability of participants the elderly in this study include those who are 60 years and over. It is assumed that this did not have a significant impact on findings as physical and psychological changes in the elderly are evident from the age of 55 [7].

It was planned that the tests be conducted in actual galleries of each museum. The test consisted of predefined user tasks and pre- and post-test questionnaires. The pre-test questionnaires were to collect demographical information as well as information pertaining to the participants' computer and museum experience. From an initial assessment of the application it was felt that the features that elderly users were most likely to face problems with were navigation, manipulation of 3D graphics and zooming. In the navigation tasks the users were supposed to find their way to a page containing flash animations of how Japanese kimono materials are weaved, dyed and embroidered. From this page they were supposed to select and watch the animation of how weaving is done and then return to the home page. For the manipulation of 3D objects, they were supposed to use a particular control to rotate a 3D representation of a head gear left/right and up/down to view it from different angles. The zoom task involved selecting a small image of a kimono, enlarging it and using the custom made zoom application to view it in detail. Screen shots from the 3 tasks are shown in Figures 2, 3 and 4

below.

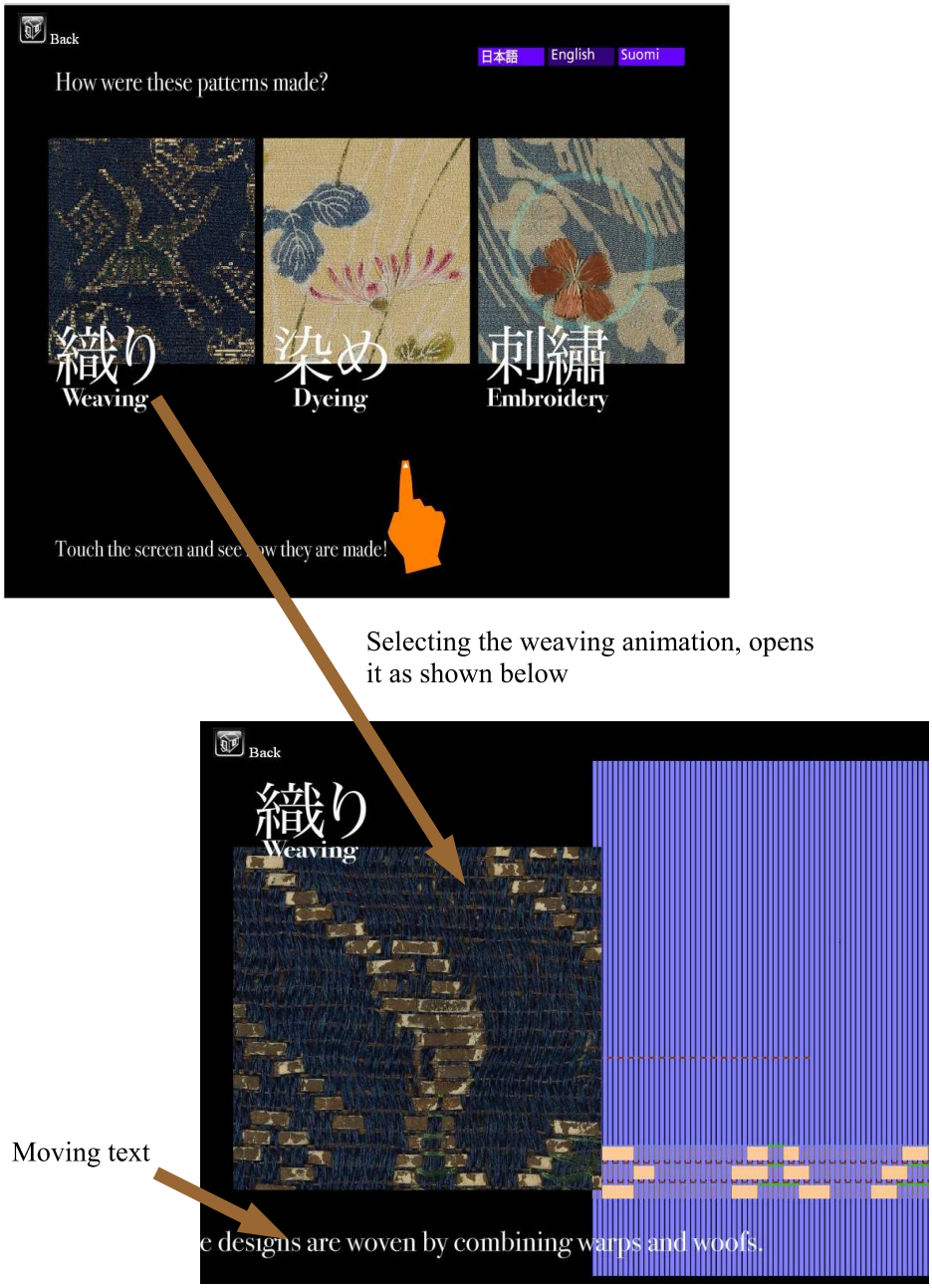

Figure 2: Screen shots from navigation task.

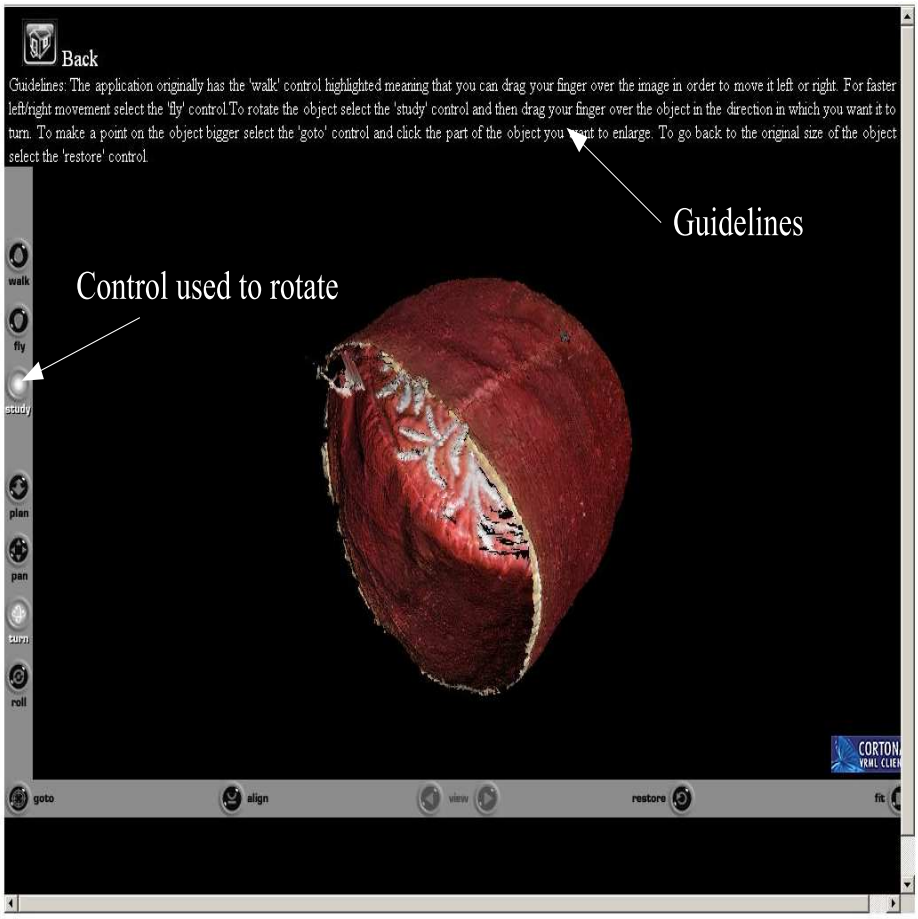

Figure 3: Screenshot of 3D manipulation task.

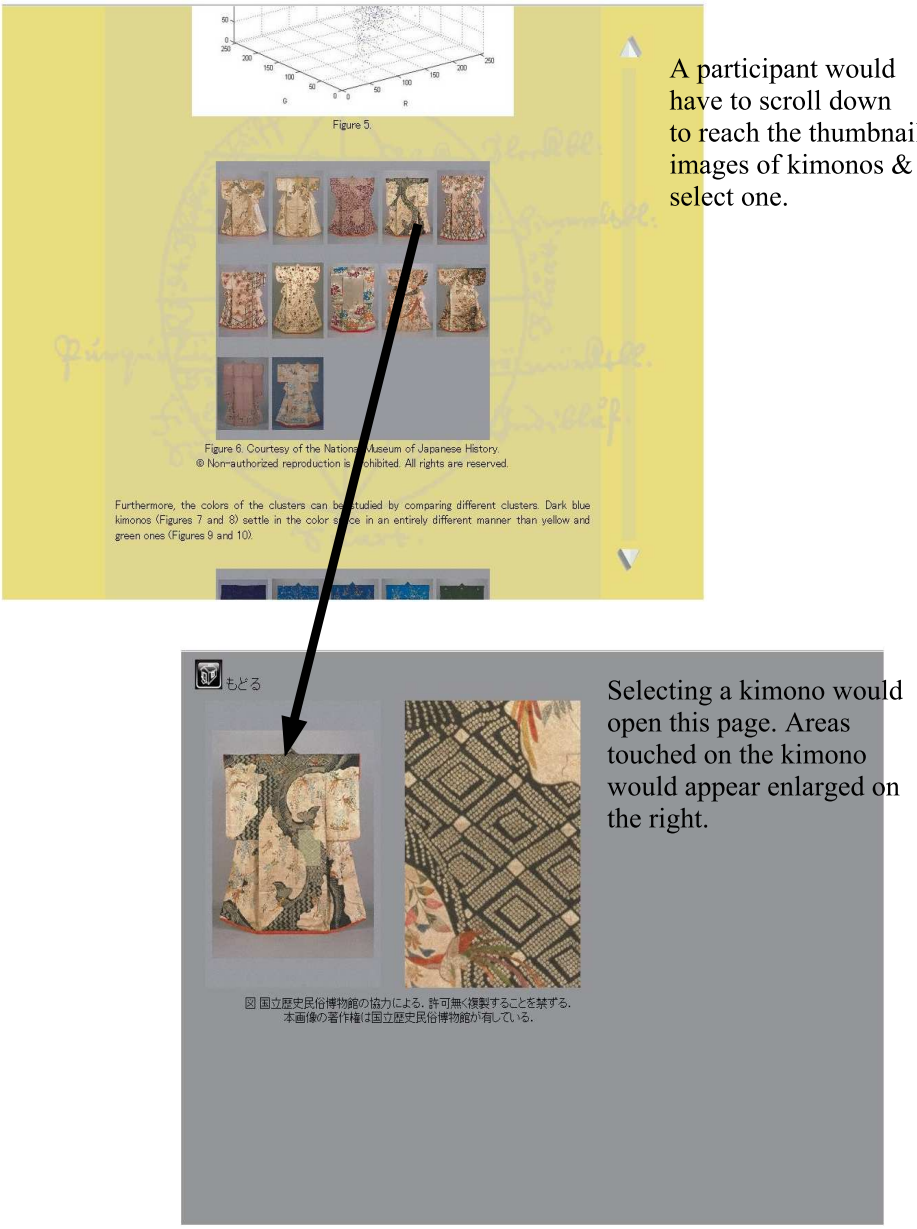

Figure 4: Screen shots from Zooming task.

These tasks were to allow for the collection of data pertaining to task completion times, errors and number of times assistance was sought. Usability problems were to be identified mostly by observation. Errors encountered during tasks were categorized into two (as suggested in [49]); **Recoverable errors**- where the user performs an action not expected to assist in the com-

pletion of the original task (e.g. opens wrong page) but manages to correct their mistake and **Unrecoverable errors**- where the user does not realise their mistake or cannot continue the task and has to be assisted. User satisfaction and general feeling toward the application and test were collected by the post test questionnaire.

Note: In sections to follow, reported are the average number of errors and assists. The averages were calculated **only** over the number of participants who actually encountered errors/sought assistance. This was done so as to avoid over-looking a large number of errors/assists which could potentially be minimised by averaging over all participants (including those who did not encounter any). Completion times were on the other hand averaged over all participants.

5.2 Results from the North Karelian Museum; Joensuu, Finland.

A pilot test was conducted prior to the main test to ensure that the tasks and questionnaires were easily understood and that the test as a whole was collecting the desired information. The pilot test revealed that there were some questions in the questionnaires that were ambiguous and in some cases required more options as potential answers. The pilot test also provided some logistical information such as an estimate into how long the tests would take, equipment that would be needed, environment specific issues such as lighting and also gave direction as to the sequence of events during the tests. A suitable area in a gallery of the museum was also reserved for the tests. With these results at hand the test plan, questionnaires and tasks were edited where seen necessary to optimise the test result, a summary of which is presented next.

5.2.1 Participants' profile

A total of seven participants were tested, all were 60 years of age or over. The tests ran over five days with a maximum of two participants tested per day. Of the seven participants, six were female and one male. It was seen important to collect information that would potentially reveal the different cultures that participants had been exposed to. When asked, none of the participants had lived out of the country for an extended period of time

whereas all of them had been on holiday out of the country at least once. A summary of their language skills averaged (and rounded up) over the 3 aspects of speaking, reading and writing is shown in Table 3.

Table 3: Number of Finnish participants with various levels of language skills:1=very poor & 5=very good.

Language	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**
English	1	2	0	1	0
Finnish	0	0	0	1	6
Swedish	0	1	3	0	0
German	1	2	2	0	0
Portuguese	1	0	0	0	0

Five of the participants had been visiting museums for over three years and two for less than one year. In terms of frequency of museum visits two visited at least once a month, two at least once every six months and three at least once a year. Majority of the participants said during their visits they would view physical exhibits and follow guided tours with only two of them saying they made use of the computer applications provided as supplement to exhibitions.

When asked how they felt about the use of computers generally, none of them said they were not interested in their use, but instead responses were evenly distributed among 'Highly interested and eager to learn more', 'Don't mind using computers but could do without them' and 'Have studied the basics of computers but do not use them in everyday life'. Following those responses it was no surprise that four out of the five participants who said they use computers had been using them for over three years. All computer users said they used computers for typing and email and with only two using them for internet surfing.

5.2.2 Summary of Task Measurements

For the **navigation task** completion times ranged between 3mins 38secs and 12mins 12secs with the average being 6mins 48secs. The average number of assists sought was 2.14, with the majority of participants asking how to begin, that is - how to open a page, start an animation and also how to return to the home page after viewing the animation. The average number of recoverable

errors was 1.75, where participants without realising mostly opened a wrong page or viewed pages in the wrong language. Not many unrecoverable errors were noted but where they were, were usually related to the opening of wrong animations deviating from the original task. These findings are summarized in Table 4 below (Note: results for each participant are given in a fixed order i.e. Participant 1 in Table 4 corresponds to Participant 1 in Table 5 and so on.)

Table 4: Navigation task results summary (Finland).

Participant	**Completion time (min:sec)**	**Num of assists**	**Recoverable errors**	**Unrecoverable errors**
1	12:12	2	1	1
2	3:38	1	0	0
3	8:02	4	3	0
4	7:33	2	2	1
5	5:38	2	1	0
6	5:20	2	0	0
7	5:18	2	0	1
Avg	**6:48**	**2.14**	**1.75**	**1**

The 3D-manipulation task had completion times ranging between 5mins 31secs and 8mins 56secs with an average of 6mins 33secs. The average number of assists sought was 3.25 mostly related to which control to invoke to rotate the headgear, how to do the actual rotation even after entering the correct mode and how to restore the object to its original form. The average number of recoverable errors was 2, and were mainly observed where a participant zoomed into the object by accident and managed to restore it without any help. There were on average 2.50 unrecoverable errors recorded for this task where a participant zoomed into the object or moved it out of visible range and failed to restore it. Details of this second task are summarized in Table 5.

In the zooming task, completion times ranged from as low as 3mins 49secs to as high as 17mins 26secs, averaging at 6mins 56secs. The average number of times that assistance was sought was 2.50, mostly related to how to use the animated scroll bar, how to open a bigger image of a kimono and how to actually use the zoom facility. Recoverable errors were not common but where observed were related to 'clicking' on text rather than an image to enlarge it, selecting wrong images and dragging along the scroll bar as opposed

Table 5: 3D Manipulation task results summary.(Finland).

Participant	**Completion time (min:sec)**	**Num of assists**	**Recoverable errors**	**Unrecoverable errors**
1	6:27	5	0	3
2	5:43	2	2	0
3	7:12	3	2	0
4	8:56	3	2	2
5	6:28	0	2	0
6	5:31	0	3	0
7	5:39	0	1	0
Avg	**6:33**	**3.25**	**2**	**2.50**

to clicking on arrows to move the page up/down. Also uncommon, were unrecoverable errors which mainly occurred when a user tried and failed to stop the moving of the page once the animated scrolling was begun. Results of the zooming task are given in Table 6.

Table 6: Zooming task results summary(Finland).

Participant	**Completion time (min:sec)**	**Num of assists**	**Recoverable errors**	**Unrecoverable errors**
1	17:26	4	0	1
2	4:44	0	1	0
3	5:09	2	1	1
4	3:45	0	1	0
5	5:49	2	0	0
6	4:04	0	0	1
7	7:41	2	1	2
Avg	**6:56**	**2.50**	**1**	**1.25**

5.3 Results from the Rekihaku Museum; Sakura City, Japan.

In Japan, it was requested that besides the pilot test, a 'pre-pilot' test be conducted as well in order to familiarize the translator with the entire study.

From this test and discussions held after it, it was decided that some parts of the questionnaires were too personal for the Japanese, such as those pertaining to educational level and whether a participant had a known or diagnosed memory problem. For that reason these questions were excluded from the questionnaire. It was not possible to conduct the tests in an actual gallery as such they were held in an office space taking the user out of context. It must be noted also that the computer used in this set of tests was characterized by slow system response.

When it came to the actual pilot test, as most of the issues had been ironed out in the pre-pilot test the only concern that was raised was related to the think-aloud process. It was noticed that even though the pilot test participant had been encouraged to verbalize his thoughts he was very quiet during the test. In discussions held with the translator and another local contributing source, they were of the opinion that he was actually quite 'chatty' for a Japanese person. This was an indication that thinking aloud would not work as well as anticipated in this particular context. A note was thus made to be extremely observant of the actual test participants' behavior during the tests. The translator was also asked to probe the users' actions by asking questions at times when they seemed to be lost in thought.

5.3.1 Participants' profile

Similar to the previous test sessions I had a total of seven participants in Japan. Among these were five women and two men. All of whom were again above the age of 60. In investigating the type of cultures that they had been exposed to it was found that all but one of the participants had been on holiday outside Japan, with none of them staying out of the country for long periods of time. The various languages that they spoke were self assessed as shown in Table 7.

Table 7: Number of Japanese participants with various levels of language skills:1=very poor & 5=very good.

Language	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**
English	5	1	0	1	0
Japanese	0	0	3	2	2

It turned out that six of the participants had been visiting museums for over three years with only one having a history of museum visits for less than

one year. In terms of frequency of museum visits, responses were distributed among at least once a month (three participants), at least once every six months (two participants) and at least once a year (two participants). Most of the participants said they visited the museum to view physical exhibits and take meals in the museum restaurant, with a couple saying they used the computer applications provided by the museum as exhibit supplements, visited the library, attended lectures and visited the museum shop.

Their attitudes to computer use were split between 'Don't mind using computers but could live without them' and 'I am interested in computers and would like to learn more about them'. The majority however leaned more towards enthusiasm towards computer use. When asked about their experience with computers 5 out of 7 reported using computers for over 3 years with most using them for between 4 and 6 hours in a week. The whole group used computers for word processing tasks, five participants used computers for email and general internet surfing and one even watched online television.

5.3.2 Summary of Task Measurements

In the navigation task, completion times ranged between 2mins 2secs and 4mins 37secs, with an average of 2mins 53secs. The average number of assists sought was 2 and these were noted when a participant asked whether to touch the screen to use it and how to touch it i.e. 'Should I double click or press on a link for some time?'. The average number of recoverable errors was also 2 involving a participant opening a wrong page or changing the application's language but successfully correcting it on their own. There was only one unrecoverable error noted and this was where a participant had to be told that she was on the wrong page and had to be shown to the correct one. These results are presented in Table 8.

In the second task which involved the manipulation of the 3D head gear, completion times which were centered around an average of 5mins 41secs ranged between 2mins 16secs and 8mins. The average number of times assistance was sought was 1 pertaining to how to restore the object to its original form when it was zoomed into incorrectly. In some cases the participants asked how to do the actual rotation even after they had managed to invoke the correct control. In such cases they were told to try and read the instructions again or if they still did not understand were told to try and drag their finger over it. On average, 1.25 recoverable errors were noted the most com-

Table 8: Navigation task results summary(Japan).

Participant	**Completion time (min:sec)**	**Num of assists**	**Recoverable errors**	**Unrecoverable errors**
1	3:18	0	0	0
2	3:09	1	2	1
3	4:37	2	2	0
4	2:13	0	0	0
5	2:02	3	0	0
6	3:41	2	0	0
7	1:12	0	0	0
Avg	**2:53**	**2**	**2**	**1**

mon being that of selecting the wrong control and changing to the correct one after seeing that rotation was not happening. Only two participants encountered unrecoverable errors bringing the average to 1.50. Unrecoverable errors were mainly related to their failure to restore the object to its original form on their own or trying to rotate it in the wrong mode for an extended period of time queuing the test monitor to intervene. Measurements for each participant are shown below in Table 9.

Table 9: 3D manipulation task results summary(Japan).

Participant	Completion time (min:sec)	Num of assists	Recoverable errors	Unrecoverable errors
1	3:25	1	1	0
2	7:26	1	0	2
3	7:28	0	0	0
4	6:57	1	1	0
5	4:21	1	1	0
6	8:00	1	0	1
7	2:16	0	2	0
Avg	**5:41**	**1**	**1.25**	**1.50**

For the final task, completion times lay between 4mins 27secs and 10mins 26secs. The mean task completion time was 6mins 40secs. The average number of assists was 2.20, sought by those participants who asked for help and was mostly related to the meaning of 'scroll' in the task instructions,

how to scroll the page, and how to select areas on a kimono image to view it in more detail. The number of recoverable and unrecoverable errors averaged at 2.17 and 1 respectively. Recoverable errors were mainly tied to dragging along the scroll bar as opposed to touching the arrows to move the page; touching wrong images in trying to open an enlarged kimono image and touching the zoomed image trying to make it bigger. Unrecoverable errors were noticed where one participant kept scrolling up and down bypassing the target kimono images over and over again until the test monitor had to step in. An unrecoverable error was also noted where one participant kept touching an image which would not open an enlarged image of a kimono and had to be told this by the evaluator when she began to appear frustrated. These findings are summarized below.

Table 10: Zoom task results summary(Japan).

Participant	**Completion time (min:sec)**	**Num of assists**	**Recoverable errors**	**Unrecoverable errors**
1	7:23	3	4	1
2	5:06	1	3	1
3	10:26	3	1	0
4	7:28	0	2	0
5	4:27	2	0	0
6	5:44	0	2	0
7	6:11	2	1	0
Avg	**6:40**	**2.20**	**2.17**	**1**

5.4 Analysis of results

With task summaries at hand authors such as Rubin [49], outline the following steps to take in analyzing data obtained from Usability Tests; identification of and focus on non-criterion tasks, identification of usability problems and their sources, prioritization of problems by criticality and the analysis of differences between groups (or product versions). This is the format that is adopted in this section. We will also take a look at the users' stated preferences and satisfaction with the application.

5.4.1 Non-criterion tasks

Even though it is not a user's capability under scrutiny in usability tests, Nielsen [42] and Rubin [49] recommend the calculation and use of user success rates in test result analysis. They argue that these numbers help to clearly point out the tasks that were particularly 'difficult' for users, allowing a design team to focus on them during improvement. These success rates may be calculated as the percentage of participants who completed a task correctly. It is also noted that the success of participants may be weighted; for example with two successful participants, one who completes a task within a time limit is ranked higher than one who exceeded a maximum time.

For the purpose of this study time was not a factor but rather, a task was considered successfully completed if the participant finished a task without any assistance (100%). Partial success referred to those tasks that users completed correctly but required assistance to do so (50%). A task was considered unsuccessful if it was incorrectly completed (i.e. result of task was not one asked for in the instructions) or if the participant gave up (0%). The time taken to complete a task was not taken into consideration mainly for the following reasons:

- The system speed may not have been constant across all participants
- The amount of time taken by the monitor to respond to a participant who sought assistance was not kept constant
- The amount of time taken by participants to ask questions during the tests was unique to each individual

- The reason for taking an amount of time on a task varied for each individual i.e. out of interest or experiencing difficulty
- It was felt that for museum applications the ability of a visitor to use an application on their own was more important than the speed with which he/she uses it as museums rarely have people stationed at computers to assist them

The success rates for each participant in each task were calculated and are presented in Tables 11 and 12 for the Finnish and Japanese participants respectively. The average success rates for each group per task and per user are also shown.

Table 11: User success rates (%) for Finnish participants

Finns	**Navigation**	**3D Manipulation**	**Zooming**	**Avg per user**
1	50	0	50	33.3
2	50	50	100	66.7
3	50	50	50	50
4	50	50	100	66.7
5	50	100	50	66.7
6	50	100	0	50
7	50	100	50	66.7
Avg. per task	50	64.3	57.1	**57.1**

Table 12: User success rates (%) for Japanese participants

Japanese	**Navigation**	**3D Manipulation**	**Zooming**	**Avg per user**
1	100	50	50	66.7
2	50	50	50	50
3	50	100	50	66.7
4	100	50	100	83.3
5	50	50	50	50
6	50	50	100	66.7
7	100	100	50	83.3
Avg. per task	71.4	64.3	64.3	**66.7**

In [49] Rubin goes on to define non-criterion tasks as those that do not meet a predetermined success rate, which is recommended to be 70%. Tasks

with success rates lower than 70% are considered to be problematic. For the Finnish participants all tasks received success rates lower than 70%. This was the same for the Japanese participants with an exception of the first task which slightly passed the benchmark (receiving a success rate of 71.4%). These low success rates show that all tasks were generally difficult and that all three areas of the application involved need to be focused on during improvement. It appears however that most difficulty was experienced in the first and third tasks reasons of which are explored next.

5.4.2 Nature and source of usability problems.

During the test sessions both the test monitor and translators were observant of participant behavior and the problems that they experienced. Problems observed could be divided into two; those that would affect the general user and those that are possibly age related. Also highlighted here are the possible reasons behind (sources) the problems experienced.

General problems. When shown to the touch screen most participants simply did not know how to begin using it. The problem of not knowing how to start was magnified by the fact that on the home page there was no distinguishing feature between plain text and text which when touched would open another page. It is assumed (in the case of the Finnish participants that is) that because of this some partipants after standing at the screen for a minute or two would touch the thing that appeared most familiar to them i.e. the Finnish flag icon, which among other flag icons serve to translate the application into different languages. Even by so doing it must be noted that the participants did not know the purpose of these flag icons as they kept touching the Finnish flag even though it was the Finnish version of the application running. Some participants recovered from this minor setback on their own by realising that touching different pieces of text would open their respective pages, while others had to be assisted.

While participants were watching animations of how kimono textiles are made (as part of the navigation task), it was noticed that some did not wait for them to complete and exited the page while others repeated them two or three times. After taking a closer look at the profile of one Finnish participant who particularly watched each animation twice it was noticed that she had a genuine interest in Japanese history, and once interviewed even said that she had on occasion visited Japan and hosted Japanese students in her home. Of the 7 Japanese participants, all but one watched the animations more

than once confirming a link between the number of times an animation was watched and genuine interest on the subject. This inspired the thought that the animations were potentially too long for those that did not have a great interest in the subject matter but also too short for those who were.

For some participants who watched the animation more than once it was noticed that in the first round they were focusing on text that was moving at the bottom of the page and by the time they finished reading this distracting text, the animation was complete and having not watched it they would play it again. This would be the case for a user of any age group as it is simply impossible to read text and watch an animation at the same time.

A navigation loop hole was noted when those not interested in watching the whole animation pressed 'Back' to exit. Doing so took them to the main page of the application and not the main page of the Rekihaku Museum. This caused participants to go back and forth between pages at times even losing them in the application.

As the third task involved scrolling down over a text filled page to get to small kimono images, some participants dragged along the scrollbar or pressed on the text rather than on the arrows of the scroll bar. Pressing on text was more common among the Finns who it is assumed had gotten accustomed to clicking on text to open a page from previous tasks. The Japanese were more inclined to drag along the scrollbar instead of pressing on the arrows, which would have worked was it a normal scrollbar and not an animated one. In either case, when a participant successfully pressed on the downward facing arrow and the page began to move most were fascinated, but panicked when they saw that the scrolling would not stop on its own. Several made exclamations such as 'Stop! Stop!' or 'Oh no!' before they were assisted or discovered that clicking anywhere on the page would stop the scrolling. They were a bit overwhelmed by such overly complicated features. Most participants for this reason selected the first kimono they came across for the zooming task.

Having enlarged a kimono image, it was necessary for the participant to drag their finger over or press on different areas of the image in order to study it in detail. When on this page most participants did not know what to do, some just 'trying' to press on the image and exclaiming in surprise when they saw that doing that zoomed into the area that they pressed. Something unique to the Japanese group however was that most of them pressed on the enlarged section of the kimono trying to make it even bigger. When asked

they said that the 'zoomed' image was still too small and that they would have preferred greater detail in a larger image. The majority of this group also unveiled another navigation loophole when trying to view one kimono image after another. In order to reach the thumbnail view of kimonos so as to select one, a participant who clicks back would be taken to the top of the previous page and would have to go through the process of scrolling again. In most cases the participants would have by this time forgotten how they had successfully scrolled the page and would once again go through the grueling task of trying to discover how to.

During the zooming task, a particular problem was observed among the Japanese participants. This was the misunderstanding of the task instructions. In the English version of the task the phrase '... scroll down the page to view small kimono images...' had been used. During task translation the word scroll had been 'borrowed' from English as there was no single Japanese word that could describe the action. For this reason most participants did not understand this somewhat technical term and were not quite sure what was being asked of them.

Another problem was observed where participants at times were not sure of what was happening with the application. When the system did not respond immediately to their action they were not sure why. This on occasion resulted in them touching the screen more times than was necessary to invoke a particular control.

Also, some participants commented that they did not understand the overall purpose of the application. They said they were carrying out the tasks quite alright, but did not understand the relationship among them. In other words, they did not see the overall theme of the content.

Age related problems. It was noticed that most of the participants, were leaning in towards the screen when using it. This was particularly so during the 3D manipulation task which required them to read through a short paragraph instructing them on how to use the 3D viewer controls. When questioned, most complained that the text was too small and too long causing them to squint and strain their eyes. The texts had a standard 14 point font size appearance and according to the design guidelines listed in Section 3.3 of this thesis this text size should have worked fine but no option was given for the enlarging of text which is also part of the guideline. This inability to read the text clearly could be associated with the decline in vision with age described in the literature review of this thesis. It is possible that for

this reason, some participants did not read through the instructions resulting in their selection of wrong controls and failure to successfully rotate the 3D object. This was one the major reasons cited for the failure to complete this task successfully among the Japanese participants. Most of them admitted to skipping the instructions and just trying to use the controls on their own and read the instructions only after they failed to do so. This was a clear indication that the text was truly difficult to read as the Japanese (being of a culture of high Uncertainty Avoidance [34]) would have been expected to read quite thoroughly information presented to them.

Most participants were also observed to trace a finger over the text line by line whilst reading it out loud. Having read the instructions (or not) it was still quite difficult for the participants to find and invoke the correct control for the rotation of the object. Most participants complained that the controls were too small making it hard to select them. The elderly participants' failure to zone into these small controls by touching them sheds light on the decline in fine motor movements in the elderly described in [6]. It did not help either that the controls were many in number and were labeled in English (for both groups).

Once in the correct (rotation) mode, some participants managed to successfully rotate the object while others had trouble dragging their finger left/right which was required for the object to move. For some this was because they were not able to exert enough pressure on the screen for their finger to be sensed. One participant in particular explained that she had a medical problem with the joints in her fingers and found dragging difficult but had no problem with 'touching' the screen. It was also noticed that in this task some participants made a rotation/circular movement with their finger rather than the left/right movement that was explained in the instructions. When probed some explained that they felt it common sense that if something is to be rotated then the rotation movement should come from the finger, reinforcing the idea that the elderly associate tasks they are confronted with with something familiar to them.

By the third task, it was noticed that some participants began shifting from leg to leg giving the impression that they were tired of standing (this was only true for the Finnish group as the Japanese group had a seated arrangement). However during the debriefing sessions it was a surprise that only one participant, complained about being tired of standing and would have preferred sitting down. All others said they had no problem with standing and did not even notice how time flew as they were captivated by the

tasks and application.

5.4.3 Problem prioritization

Several authors such as Rubin [49] and Dumas *et al.* [9] in outlining the analysis of usability test results recommend the prioritization of problems identified. They say this presents the problems in an organized manner. Also apart from the use of user success rates this makes it is easy for the design team to work on the most critical problems first. A general problem severity ranking is presented by both authors and the one to be used here (presented in Table 13) is tailored from the one given in [9].

Table 13: Problem severity ranking

Severity Ranking	**Description**	**Definition**
4	Unusable	Problem prevents task completion
3	Severe	Problem creates significant delay and frustration
2	Moderate	Problem has minor effect on usability
1	Irritant	Problem is more subtle pointing
		to future enhancement

The problems were ranked taking into consideration how severe they were and how often they occurred. This was made possible by the use of the number and nature of the recoverable and unrecoverable errors recorded in the task summaries. Observation also made the ranking easy as the tasks were quite short and the number of participants few. The problems described in the previous section and their sources are presented in order of severity in Table 14.

Table 14: Problems identified with their corresponding severity rankings

Rank	Problem	Source
4	Not knowing what to do	-No visible instructions. -'Active' links/icons did not have button-like appearance. -Non-uniform use of touchscreen.
4	Difficult to understand use of controls	-Labeled in foreign language. -Were too many in number. -Were too small in size. -Were different from 'normal' ones e.g. animated scroll bar.
4	Difficult to read text	-Font too small. -Texts too long.
4	Difficult to understand text	-Use of technical terms.
3	Unsure of what is happening	-No feedback from system to confirm action. -No change in cursor shape or sound. -Slow system response.
3	Going back & forth between pages	-Navigation loopholes. -No way of keeping tab on where one is.
2	Did not understand general application purpose	-Lack of general theme tying sections together. -Unexplained Color Research information.
1	Early exit from & repetition of animations	-No playing options. -Too many things happening on page.

The problem of the user not knowing what to do is ranked as one of the highest as it would be pointless to have an application that will not be used as people do not know how to. Ranked with equal severity are difficulties in using given controls, reading from the screen and understanding text. In improving the application the sources of these four problems should be tackled first. While some of these problems such as font size would be very easy to fix, others such as the non-uniform use of the touch screen may require more effort.

5.4.4 User preference and satisfaction

With the problems observed it could be said that the elderly generally did not find the application easy to use. This was reflected in the post test questionnaire where none of the participants in either of the groups strongly agreed with the statement about the application being easy to use. Only 29% of the participants agreed to some extent, the rest remaining neutral or disagreeing.

The severity ranking in the previous section is mirrored by the comments that the participants made about what they liked least about the application. Figure 5 shows the distribution of participants across the various comments made. The most common being the understanding of instructions. This however is a combination of the task instructions as well as guidelines available in the application. Either way this shows that technical terms may not be easily understood by users. Having observed that participants were having difficulty reading from the screen, about 15% of them actually complained about the pieces of text being too long and fonts too small.

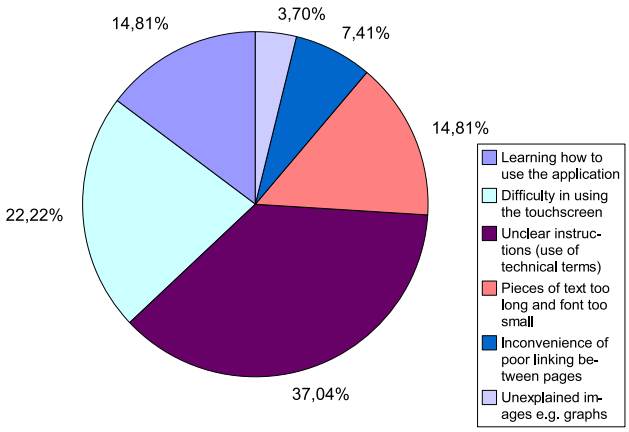

Figure 5: Distribution of what was liked least about the application.

Regarding the interaction technique; even though there was the occasional participant who '... liked using the touchscreen because I could touch things directly...', 22% of the participants did not enjoy using it. They complained that they just did not know how to use it; whether a single-touch, double-touch, light/strong tap or holding onto a link was the correct way of using it. Others complained that dragging with the finger was laborious.

It was noticed that some participants were continually glaring at or wiping their fingers. Comments were made such as 'The screen just won't obey my finger', 'My nails are too short that's why the screen is not responding' and 'Is my touch too weak? The movement of the object on the screen is not proportional to the effort I am putting into dragging'. This attitude towards touchscreen use must however be shared with the system's slow and lack of response. These comments were usually made where the user was observed to have touched the screen correctly but as the system did not give an immediate response began to doubt their actions. Some authors such as [40] say that the maximum time a user should be made to wait for a response is 10 seconds. With this application however, response times were in most cases more than that.

The participants' views of the ease of use of the application however should be distinguished from their attitudes towards it, as 76% of them actually enjoyed the experience and 93% felt that the use of computer applications was a good way for museums to provide extra information about exhibits. They were also allowed to freely comment on things they liked most about the application. Their comments are summarized in Figure 6.

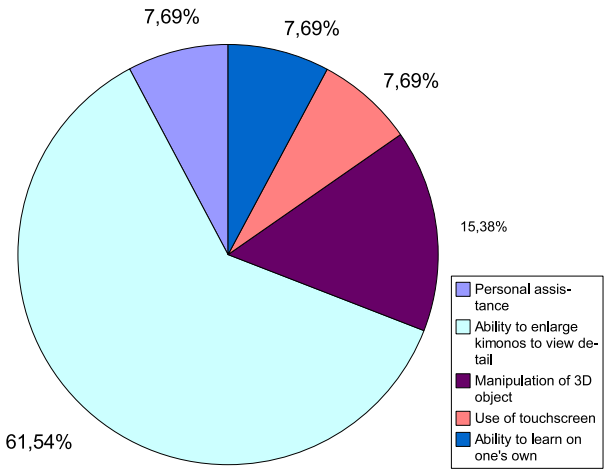

Figure 6: Distribution of what was liked most about the application.

Even though the zooming task proved to be one of the most difficult to complete, from Figure 6 we see that about 62% of the participants enjoyed viewing the details of the kimono over other parts of the application. This could **possibly** be because elderly people are more comfortable looking at magnified rather than small images. Also the response of this part of the

application was immediate i.e. touching an area on a kimono would immediately magnify it without the user having to wait not knowing what the system is doing.

About 15% of the participants enjoyed the ability to view the 3D object from different angles. Participants also valued the availability of personal assistance (i.e. the test monitor and translator). During the debriefing sessions this saw to most of the participants passing comments such as 'I do not think I would have managed if I was on my own' or 'This may be easy to use for young people but for myself and other elderly people I think someone would have to be there with me'. This shows that the elderly may be a group of users that require more assistance than others. In the case where it is not possible to have somebody physically standing by to give assistance one participant suggested 'Maybe it would have been easier if there were audio instructions'.

It is worth noting here that the elderly who are assumed to be negative towards the use of technology, actually enjoyed using parts of it that they did not necessarily find easy to use emphasizing their enthusiasm.

5.4.5 Differences between groups

In this study we had groups distinguishable by their culture. In the pre-test questionnaire participants were asked whether they had lived in countries other than their countries of origin. This was seen important as it was felt that living in another country for extended periods of time would split ones cultural background. As none of the participants had lived outside their own countries, we see only two distinct cultures; Finnish and Japanese.

The content of a museum application and its user may be of different cultural backgrounds. It would be ideal that a user of a different culture from the application content does not necessarily experience more difficulty in the application's use than someone who shares its cultural background. It was noticed during the test sessions that some differences existed between the Finnish and Japanese participants' use of the application. This section however does not serve to compare performance but rather just highlight the differences between the two groups.

For instance, in the first task there is a significant difference between the average time taken to complete a task by the Finns (6mins 48secs) and the

Japanese (2mins 53secs). While one may argue that these differences can be attributed to differences in testing conditions, system response times and time taken up to give assistance; we cannot ignore the fact that one timing is less than half of the other. Also, it must be noted that the Japanese were at more of a disadvantage when it came to system response times as the computer they used was several times slower than that used by the Finns. The Japanese however took up less time to complete the task. Why was this so? Firstly, it was noted that the Finns had a harder time actually starting the task. Though the Japanese also had trouble starting out, their questions took the form 'Is this a touchscreen?' as opposed to the Finns who tended to ask 'What am I supposed to do to open a page'. It can be seen from here that the Japanese showed a higher level of familiarity with this type of interaction technique. It could be assumed that Japan is a more technologically advanced country and as such sees the use of touchscreen applications not only in museums but also public areas such as train stations. This could possibly explain why they were more at ease with the interaction technique used in the test.

Another reason for the better performance of the Japanese could be attributed to what the first task entailed for both groups. For the Finns they first had to change the default language of the animation page which was in Japanese. Some Finnish participants for this reason incorrectly watched animations of how kimono textiles are weaved, dyed and embroidered in the wrong language. In this case the Japanese were at some sort of advantage as they did not have the shock of opening a page that was in a language other than their own. The inability of some Finnish participants to realize that they could actually change the animation language could be attributed to the fact that the text which was to be touched to change the language did not have a 'button like' appearance. It could also be assumed that the Finnish participants were not greatly interested in the text but rather the animations, thus not noticing that the page was in Japanese. On the other hand some participants (4 out of 7 to be precise) realised that the language needed to be changed and either managed to do so on their own or sought assistance.

While the average completion times are not significantly different between the two groups for Tasks 2 and 3 some differences were still noted. Firstly, the Finnish participants spent the amount of time that they did on tasks because of learning how to use the application and touchscreen, while the Japanese spent the same amount of time on tasks merely by choice. After completing what the task asked of them (particularly in the manipulation

of the 3D object), they went on to explore the different controls that were available. This could be because the Japanese participants were not 'burnt-out' by difficulties experienced in the use of the interaction technique and application. Apart from familiarity with the interaction technique, a closer look at their background information revealed that the Japanese were about 1.75 times more experienced in computer use than the Finnish participants. The Japanese participants were also more conversant with web applications where 5 of them when asked said they used computers for general internet surfing (this was true for only 2 of the 7 Finnish participants). This potentially suggests two things; firstly that the application was easier to use for those with more computer experience and secondly that intrinsic to the Japanese culture is a higher computer literacy level among the elderly in comparison to the Finns.

It was also noticed that the Japanese participants were quite inquisitive, not wanting to just look at an image but also wanting to understand what it was. For instance the 3D object which was a spectral image of an ancient headgear had a hole in it due to the limitations of the imaging technique used. Most of the Japanese participants stopped mid way through the task (while the clock was still ticking) to ask whether the hole in the hat was because it was old. An almost insignificant but important observation. This attitude of 'wanting to know exactly what I am doing' closely resembles the cultural dimension of Uncertainty Avoidance defined in [20]. Japan ranks 7th (quite high) in this dimension and it is explained in [34] that people from countries with a high Uncertainty Avoidance index would much rather fully understand the consequences of what they are doing not leaving anything to chance. This somehow explains the participants' attitudes during the tests. Their focus on the meaning of content contrasted the Finns focus on appearance of content earlier highlighted (where I have proposed that one of the reasons the Finnish participants watched an animation in a wrong language was that they were more focused on watching it rather than understanding the accompanying text).

It was also noticed that the Japanese participants spent more time on the tasks than they had to because most of them did the tasks more than once before saying that they were finished. Having met difficulty in a first attempt and seeking assistance they would go through it once or twice more trying to get it done on their own. On the other hand the Finnish participants were quite content on completing a task with assistance and as such only went through the tasks once. One reason for this amongst others could be that the Japanese were more interested in the content of the application as it

was mostly Japanese oriented i.e. kimono images, animations of how kimono textiles are made and so on.

Irrespective of the reason for one cultural group interacting with the application more easily than the other, the fact remains that the application is biased along this dimension. Cross cultural applications such as digital museum ones should be designed in such a way that people of different cultures do not necessarily experience more difficulty than others in their use.

6 Lessons learned from Usability Testing with the elderly of different cultures

With the uniqueness of each individual, no two usability tests can ever be the same. The type and number of usability problems revealed are diverse and vary considerably. The type and number of problems usability tests reveal in my opinion depend greatly on the suitability of the type of test selected for the target group and in the case of cross cultural studies how well different cultures converge in the testing environment. Lessons learned from dealing with a sensitive group such as the elderly and two different cultures are described next.

6.1 Testing with elderly users

Irrespective of culture it was observed during this study that the elderly are indeed a very sensitive group of individuals. They tended to blame their own capabilities for the inadequacies of the application. As such, care had to be taken to encourage and reassure them constantly that it was not them but the application that was being evaluated.

The main part of the usability evaluation method selected that worked well for the elderly was having someone there to assist them when they got completely stuck. This came in handy as it was noticed that they tended to seek assistance quite often. Even when participants were performing a task correctly they would still seek reassurance of this fact. This also had a down side in that providing the participants with so much help potentially masked usability problems that they would have experienced had they been on their own.

One part of the test plan which did not seem to work well for this group was the use of questionnaires. This was pronounced in open ended questions to which the elderly mostly gave short answers. However, during debriefing sessions when asked the same questions they said much more than they had written down. Some also spent quite alot of time trying to remember things asked about in the pre-test questionnaire such as how long they had been using computers. They appeared to recall things quite easier when they spoke and not wrote about them. Another problem that was tied to the use of the questionnaire was rooted in an individual's culture as described in the

next section.

6.2 The merging of different cultures in usability testing; the test, the translator, the evaluator and the participant

The idea of this study was simply to 'translate' a usability test conducted in one country for use in another in hope of collecting the same kind of data. Crossing borders with this usability study showed the naivety of this idea.

Firstly, there were some practical issues of ***the test*** that had to be considered. In Finland, a society that is not as rooted in strict rules as Japan, it was easy at a short moment's notice to set up a testing environment in the actual museum. In Japan on the other hand, setting up the test in actual context was not possible as all the touchscreen systems in the galleries were in use and 'borrowing' one required clearance which would have taken a long time. It was then decided to have the tests in an office space. This unforeseen circumstance was the first hurdle.

In Japan, having settled on where the tests would be conducted it was then necessary to translate the questionnaires and other test related material. The question of whether to have word-for-word or literal meaning translations loomed. Having a word-for-word translation would mean that the material would not be easily understood by the Japanese participants, but on the other hand a literal meaning translation would potentially provide this group with different information than that provided to the Finns. In as much as it was desired that the tests be as close a match as possible it would not have made sense to have a translation that the participants would not understand. As a result, for the most part, the translations gave the literal meaning of text. However, in the task instructions where literal meaning would have given away how to actually **do** what was being instructed some word-for-word translation was used. The problem with doing this revealed itself during test sessions when some participants did not understand some directly translated words like 'zoom' in the instructions. This potentially split the identified problem of 'use of technical words' between the application and the test itself.

During translations, it was brought to light that the Japanese do not take well to personal questions such as those regarding education level or memory

problems. These questions which were successfully answered in Finland were as such removed from the questionnaire in Japan. This made it impossible to explore and/or compare the relationship if any between education level and how well the participant interacted with the application.

The Japanese have been known to experience problems in the use of questionnaires in the form of ambiguity of questions introduced by translations from English to Japanese [12]. In this study another problem was identified. It was noticed that while the Finnish participants had no problems evaluating their skills be it language or other, the Japanese seemed hesitant to do so. We see them being modest giving themselves low rankings. As an illustration it can be pointed out that 6 out of the 7 Finnish participants gave their Finnish language skills the highest possible ranking where as only 2 out of the 7 Japanese participants did the same for Japanese. We must remember that all participants were native to their group. We see that careful consideration also has to be placed in the type of questions asked across cultures. Whether this would have been the same result had interviews rather than questionnaires been used is subject to further research.

The presence of ***the translator*** introduced its own problems. In trying not to disturb the participant during the test things that they may have said were only translated after they had left, relying alot on the memory of the translator. Even though audio recordings were taken which could be played back later on, I feel that some information may have been 'lost in translation' where certain things could not be accurately expressed in English. One Finnish/Japanese statement could be translated in different ways by different people. For instance a Japanese word that could be translated to interesting or fun (which are not the same in English) could have two very different implications in terms of usability. The fact that the translators were not professional translators nor usability specialists seemingly added to this problem.

Some authors such as Vatrapu et al. [60] describe the effect of ***the evaluator*** in cross cultural usability testing where more problems are uncovered when the evaluator is of the same culture as the participant. How does it work though, when the evaluator is of a different culture but has a native as a translator? Being a Zambian national, not able to speak Finnish or Japanese, I had present with me a Finnish female translator (in Finland) and a male Japanese translator (in Japan). The presence of a person of the same culture seemed to put the participants at ease in both countries. However, I felt that my relationship with the participant was different in the two

cultures which I try to describe in Figure 7.

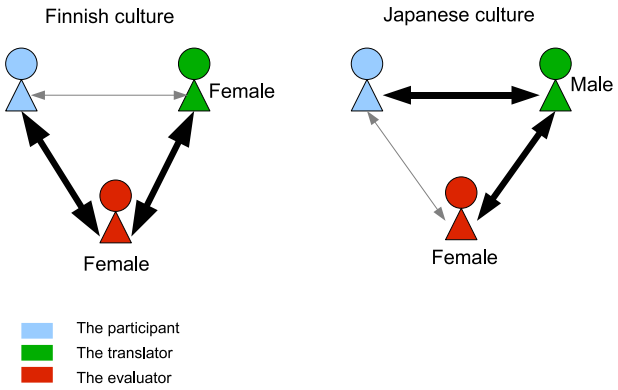

Figure 7: The relationship among the test participant, translator and evaluator.

In Finland I felt more control over the test, with ***the participant*** recognizing that I was the evaluator and feeling comfortable with it. It was possible to ask the participants questions (which were translated) in an informal manner to which they responded enthusiastically. On the other hand, in Japan I felt that the participants recognized the translator more as the evaluator. They addressed questions and comments directly to him as opposed towards me which was the case in Finland. As such in Japan I had less control over the information that was relayed to participants especially in terms of assistance. I am sure that this wall that built between myself and the participant was not deliberate but rather cultural. Since the Japanese value respect for the elderly, the translator felt the need to shield the participant from anything I would possibly say which in my culture would be fine but would translate to being disrespectful amongst the Japanese. With Japan being a highly masculine society, the fact that the translator was male could have also led to the natural assumption that he had a greater role to play in the test than I did. This could also have been because most participants were aware that the translator was a professor in academic standing and was older than I was.

It would seem that with this barrier the test in Japan would reveal very few usability problems but this was not the case. Even though the participants formed a bond with the translator this worked out fine for this particular case as the translator having a direct stake in the project was also very observant of the usability problems experienced. In the event that the

translator did not have any ties with the research and saw his duty just for what it was- translation, I feel that the tests would not have been as revealing as they were.

In summary, for the sake of data comparability it can be seen that it is not enough to design a test for one culture, translate it and deploy it it another. We also see that in the case where the evaluator is not only from a different culture but speaks a different language the translator plays quite a vital role in the tests. For some cultures things that would seem unimportant in others such as the translator's gender or age may greatly influence the outcome of the test.

7 Conclusions and Recommendations

7.1 Conclusions

This study aimed at investigating the usability issues related to a digital museum application running on a touchscreen system as experienced by elderly users. Having conducted usability tests with groups of elderly users and observing their use of the application it can be said that the application in question was not usable for this particular group.

The application was flawed mainly on counts of efficiency and satisfaction. It may be effective in the sense that users at the end of the day can accomplish a task even though they may require a great amount of assistance and time. The use of unreasonable amounts of resources such as time and effort to complete simple tasks makes the application inefficient. As users were observed to be frustrated more times than comfortable, this questions the application's ability to satisfy its users. According to ISO 9241-11 [22], to be considered usable a product would have to be effective, efficient and satisfying. As two of these conditions were not met it can be concluded that the application tested is not usable by elderly users.

The first thing that this research was to reveal were the problems that elderly users experience in the use of digital museum applications running on a touchscreen system. The list of problems and sources given in Appendix D are summarized here.

The first problem observed was that users simply did not know how to start using the application. The source of this problem lies in the lack of instructions on how to use the interaction technique and application in general. Contributing to this was also the difficulty in distinguishing certain pieces of text as buttons which when pressed perform some action.

Another problem was that it was difficult for users to understand how to use features made available to them. For the 3D manipulation controls this was mainly because they were labeled in a foreign language, too many and too small. Difficulty in the use of the animated scroll bar was mainly because of its difference from the conventional scroll bar that participants from both groups said they were used to using.

Also, pieces of text were hard to read and understand mainly because of

font size, text length and the use of technical terms. Even though the font was of a reasonable size according to guidelines, there was no way for users to enlarge it which they felt necessary.

Users were also noticed to get lost in the application usually going back and forth between sections, mainly because of the poor linkage between pages. Most of the time the participants were unsure of what the application/system was doing as it gave no visible feedback such as a change in shape of the hourglass.

The general purpose of the application was not clear as such users did not understand the relation among the tasks. It was also noticed that participants either exited or repeated certain tasks a number of times before saying that they had completed them (such as the playing of animations). This could be as a result of not being given more than one option for carrying a task out.

For the problems they encountered, all participants were glad they had assistance in form of the test observers and commented that they would not have managed on their own.

In answering the question about whether culture has an impact on computer use, this research revealed that the Japanese participants found it easier to use the application than the Finnish participants. Irrespective of the reason, this shows that people of different cultures may have different skills and respond to the use of technology differently.

This study does not compare results between the two groups but only highlights the differences. This is mainly because the testing conditions were different, system speed and amount of assistance not constant and culture seen to have an impact on the usability evaluation process.

The original plan was to simply translate the usability test used in Finland for use on Japan, but this turned out to be problematic and the amount of information obtained from the two groups differed.

Also, we see the translator playing an unforeseen role in the tests. As the dynamics of the relationship among the participant, translator and evaluator was different in both cultures, we see the two groups being exposed to different amounts of information and assistance which specifically in Japan, the evaluator had little control over. It is for this reason that one of the the first recommendations I would like to make concerns the usability evaluation

of applications across cultures.

7.2 Recommendations

It would be ideal to have a local usability specialist to evaluate an application as he/she would be sensitive to a group's culture and being able to speak the same language would eliminate the use of a translator. In the event that an application is designed by one country for deployment in another I understand that it would probably be cheaper for a company to use their own specialist rather than a local one. In this case I would recommend that a new approach to usability testing be sought which evens out the differences in culture. For instance shortly training the translators involved would not only make them fully understand their role in the test and the type of information they pass on to the participants but would also most likely increase the number and type of problems observed as there would be two and not one person with an observant eye.

For the Japanese group, another approach would be to change the usability evaluation method used. As the stakeholders of the museum application correctly identified most of the problems that they felt elderly users would experience through the 'pre-pilot' test a better method for this group could possibly be Heuristic Evaluations. These evaluations could involve museum staff members and could also be coupled with observations of actual usage which could potentially reveal problems missed earlier (as suggested also in Section 2.3.6). In the event that two different methods are used in two different groups that need to be compared, a framework would then have to be sought for the standardization and comparison of results.

Also, in future collaborative efforts among the North Karelian Museum, Rekihaku and Infotonics Center it should be noted that it is not enough to translate the language of an application developed in one country for deployment in another. In addition, assumptions made by the designer on the type of interaction technique that is familiar or easy to use for a different culture should be avoided. People of different cultures have different skills and attitudes towards computer use as has been suggested in this research. A better approach would be to present the content of the application to two designers from the two cultures and have them develop culture specific applications. As this may be resource consuming another way would be to have a designer study deeply the cultures of the countries for which he/she is

designing an application. This way, even if an application's content is specific to a particular culture (which is to be expected with museum applications) it is at least easy enough to use by people of another culture. This studying of cultures could also work well for the usability specialist involved in the evaluation of the designed application.

In designing for elderly users I would recommend careful consideration of their age related impairments and full adherence to the reviewed design guidelines. More specifically, it would be advisable not only to use fonts and images that are large enough but also give the option of increasing font/image size. As most of the elderly users preferred not to read long pieces of text seeking other ways of giving adequate help and instruction such as voice or movie clips would also be advisable. In designing for a diverse population such as that of museum visitors the use of complicated graphical features and technical jargon should be avoided or where used should be accompanied by adequate explanation.

Apart from the use of technical terms, pieces of text should also be kept to a minimum and alternative ways sought to put a message across to users, such as pictures. However, care must be taken not to oversimplify such an application for fear of boring other user groups that would have interacted easily with the application. One way to ensure this is to present the user with more than one option for carrying out a task; where some options would be easier than others.

A developer of a digital museum application must be careful when translating content from one language to another and ensure that all sections are in a language understandable by the target group of users. Where more than one language is used, options to change the language must be given.

To avoid confusing the user, constant navigation terminology should be used. Applications should be designed in such a way that the use of the interaction technique is constant for invoking the same type of action e.g. scrolling should be done by dragging or tapping but not by both in different situations.

Fast system response is also vital, therefore rather than having large images that take a while to load, the use of smaller ones which allow fast system response would be better. However, we cannot ignore the number of participants who enjoyed the 3D manipulation, so to do away with 'heavy' images may not entirely be a good solution. A better way would be to use

as many small images as possible and where it cannot be avoided use others such as 3D objects but have the system give the user constant feedback by the use of message alerts and/or status bars in the event that system response is not immediate.

The results obtained were insufficient to conclude on whether the touchscreen as an interaction technique is suitable for the elderly or not. This is simply because most of the problems observed were associated with application design flaws and not the input device. Also, the participants were genuinely enthusiastic about computer use and it would be wrong to jump to the conclusion that this group is incapable of using this interaction technique or others for that matter. The first step would therefore be to improve the application and carry out another set of tests with this group of users.

In the event that the same type of problems are observed with an improved version of the application we could then think about an interaction technique that would be more suitable. For instance if it is noticed again that the elderly are more comfortable with things that they are familiar with, an innovative method of interaction would be to provide them with input devices which mimic things in their everyday life. These may range in simplicity from a simple television remote control to the steering wheel of a car. This modelling of the user's natural environment closely resembles the concept of the Tangible User Interface (TUI) described in [59], through which physical objects act as an interface between the user and an application. Another idea for the TUI would be to have miniature physical forms of exhibits loaded with sensors which when held and rotated would rotate its 3D representation on a large screen. This approach would however be hardware intensive and thus potentially costly.

Interaction through presence or body gestures would also be worth looking into with this group. These types of interaction techniques may prove beneficial as they involve natural body movement while remaining non-laborious (also depending on how digital museum applications are designed).

Also, if after testing an improved version of the application (and maintaining the same interaction technique) strong comments are made about dislike for the touchscreen then a thought would be to eliminate the use of a direct manipulation device replacing it with an indirect one such as the mouseball. This could possibly lessen the blame that the elderly tend to put on themselves when the application does not do as expected; attention is shifted from 'What is wrong with my finger/foot/head etc?' to 'What is

wrong with this input device?'.

This research has raised quite a number of questions. These questions include; whether or not it is feasible to compare usability test results across different participants and cultures, how (if possible) these comparisons can be done, whether the presence of a local mediator as is recommended by most HCI specialists is actually beneficial and whether it is actually possible to clearly establish which interaction technique is best. It is my hope that these questions trigger further research and are answered in the near future.

References

[1] Bailey, R. W. 2001. Heuristic evaluations vs usability testing. Sited on http://www.humanfactors.com/downloads/jan01.asp on 6th April, 2007

[2] Bednarik, R. 2002. Master Thesis: Evaluation of Educational Environments: The TUP model. Sited on ftp://ftp.cs.joensuu.fi/pub/Theses/2002_MSc_Bednarik_Roman.pdf on 7th October 2007.

[3] Bevan N. 2005. Cost benefits evidence and case studies. Serco Usability Services. Sited on http://www.usabilitynet.org/papers/Cost_benefits_evidence.pdf on 6th October, 2007.

[4] Bias, G. and Mayhew, D.1994. Cost-Justifying Usability. Academic Press, NewYork.

[5] Bodker, K., Kensing, F. and Simonsen, J. 2004. Participatory IT design: designing for business and workplace realities. Pages 228-233. MIT Press.

[6] Chaparro, A., Bohan, M., Fernandez,J., Kattel, B. and Choi, S.,D. Is the Trackball a Better Input Device for the Older Computer User? Journal of Occupational Rehabilitation, Volume 9, Number 1.Pages 33-43.

[7] Czaja, S.J. and Sharit, J. (1998) Age differences in attitudes toward computers (1998) Journal of Gerontology Psychological Sciences, 53B(5). Pages 329 340.

[8] Dumas J. 2007. The Great Leap Forward: the birth of the usability profession (1988-1993). In the Journal of Usability Studies, Vol. 2, Issue 2. Pages 54-60. USA.

[9] Dumas, J. and Redish, J. C. 1999. A practical guide to Usability Testing, revised edition. Intellect Books. USA.

[10] Dutton, W.H. 2006. The internet generation: Digital divides and choices shaping diffusion and use. Talk presented at the Rob Kling Center for Social Informatics, Indiana University, Bloomington, IN.

[11] Eurostat. 2006. EU25 population aged 65 and over expected to double between 1995 and 2050. Sited on http://epp.eurostat.ec.europa.eu on 5th February, 2008.

[12] Evers, V. 2002. Cross-cultural applicability of user evaluation methods: a case study amongst Japanese, North-American, English and Dutch users. In CHI '02 Extended Abstracts on Human Factors in Computing Systems (Minneapolis, Minnesota, USA, April 20 - 25, 2002). CHI '02. ACM Press, New York, NY.Pages 740-741.

[13] Fang, X. and Rau, P., P. 2003. Culture Differences in Design of Portal Sites. Ergonomics, 46 (1-3). Pages 252-254.

[14] Fernandes, T. Global Interface Design. Academic Press, Chestnut Hill, MA, 1995.

[15] Ford, G. and Gelderblom, H. 2003. The effects of culture on performance achieved through the use of human computer interaction. In Proceedings of the 2003 Annual Research Conference of the South African institute of Computer Scientists and information Technologists on Enablement Through Technology (September 17 - 19, 2003). ACM International Conference Proceeding Series, vol. 47. South African Institute for Computer Scientists and Information Technologists, 218-230.

[16] Fukuda, R. and Bubb, H. 2003.Eye tracking study on Web-use:Comparison between younger and elderly users in case of search task with electronic timetable service. PsychNology Journal, Volume 1, Number 3, .Pages 202- 228.

[17] Frokjaer, E., Hertzum, M. and Hornbaek, K. 2000. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In the Proceedings of the SIGCHI conference on Human factors in computing systems. Volume 2, Issue 1. Pages 345-352. The Hague, The Netherlands. ACM Press.

[18] Gould, J. D. and Lewis, C. 1985. Designing for usability: key principles and what designers think. Commun. ACM 28, 3. Pages 300-311.

[19] Gunther, R., Janis, J., and Butler, S. 2001. The UCD Decision Matrix: How, When, and Where to Sell User-Centered Design into the Development Cycle, Sited on http://www.ovostudios.com/upa2001/ on 12th March 2007.

[20] Hofstede, G. 1991. Cultures and Organizations: Software of the Mind. McGraw-Hill, London, UK.

[21] IBM. 1999. Cost Justifying Ease of Use. Sited on http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/23 on 1st September 2007.

[22] ISO/IEC. 1998. 9241-11 Ergonomic requirements for office work with visual display terminals(VDTs)-Part 11 Guidance on Usability, ISO/IEC 9241-11:1998(E).

[23] Jääsko V. and Mattelmäki T. 2003. Observing and probing. In the Proceedings of the 2003 international conference on Designing pleasurable products and interfaces. Pages 126-131. ACM Press. New York, NY, USA.

[24] Jefferies, R. and Desurvire, H. 1991. User interface evaluation in the real world: a comparison of four techniques. In the Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology. Pages 119-124. ACM Press.

[25] Jokela, T., Iivari, N., Matero, J., and Karukka, M. 2003. The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In Proceedings of the Latin American Conference on Human-Computer interaction (Rio de Janeiro, Brazil, August 17 - 20, 2003). CLIHC '03, vol. 46. Pages 53-60. ACM Press, New York, NY.

[26] Kamppuri, M., Bednarik, R., and Tukiainen, M. 2006. The expanding focus of HCI: case culture. In Proceedings of the 4th Nordic Conference on Human-Computer interaction: Changing Roles (Oslo, Norway, October 14 - 18, 2006). A. Mørch, K. Morgan, T. Bratteteig, G. Ghosh, and D. Svanaes, Eds. NordiCHI '06, vol. 189. Pages 405-408. ACM Press, New York, NY.

[27] Karat, C. 1997. Cost-justifying usability engineering in the software life cycle. In Helander, M., Landauer, T., and Prabhu, P. (Eds), Handbook of Human-Computer Interaction. Elsevier Science, Amsterdam.

[28] Kurniawan, S. and Zaphiris, P. 2005. Research-derived web design guidelines for older people. In Proceedings of the 7th international ACM SIGACCESS Conference on Computers and Accessibility (Baltimore, MD, USA, October 09 - 12, 2005). Assets '05. ACM, New York, NY. Pages 129-135.

[29] Landauer, T. 1995. The Trouble With Computers. MIT Press.

[30] Law, L. and Hvannberg, E. 2002. Complementarity and convergence of heuristic evaluation and usability test: a case study of universal brokerage platform. In the proceedings of the second Nordic conference on Human-computer interaction. Pages 71-80. ACM Press.

[31] Lindgaard, G. 1994. Usability Testing and System Evaluation. Chapman & Hall, UK.

[32] Liu, L., Watson, B. and Miyazaki, M. VR for the elderly: Quantitative and Qualitative differences in performance with a driving simulator. Cyberpsychology & Behavior, Vol 2(5). Pages 53-67.

[33] Mao, J., Vredenburg, K., Smith, P. and Carey, T. 2001. User-centered design methods in practice: a survey of the state of the art. In the proceedings of the 2001 conference of the Centre for Advanced Studies on Collaborative research. Page 12. IBM Press.

[34] Marcus, A. and Gould, E. W. 2000. Crosscurrents: cultural dimensions and global Web user-interface design. interactions 7, 4 (Jul. 2000). Pages 32-46.

[35] Meyer, B., Sit, R. A., Spaulding, V. A., Mead, S. E., and Walker, N. 1997. Age group differences in world wide web navigation. In CHI '97 Extended Abstracts on Human Factors in Computing Systems: Looking To the Future (Atlanta, Georgia, March 22 - 27, 1997). CHI '97. ACM, New York, NY. Pages 295-296.

[36] Murata, A. 2006. Eye-Gaze Input Versus Mouse: Cursor Control as a function of age. In the International Journal of Human-Computer Interaction, 21(1). Pages 1-14.

[37] National Museum of Japanese Histort, available at http://www.Rekihaku.ac.jp

[38] Nielson, J. 1993. Usability Engineering. Academic Press, Inc., San Diego.

[39] Nielsen, J. and Levy, J. 1994. Measuring usability: Preference vs. performance, Communications of the ACM 37, 4. Pages 66-75.

[40] Nielsen, J. 1994. Response Times: The Three Important Limits. Sited on http://www.useit.com/alertbox/ on 5th February 2008.

[41] Nielsen, J. 2000. Usability Testing with 5 users is enough. Sited on http://www.useit.com/alertbox/ on 27th March, 2007.

[42] Nielsen, J. 2001. Success Rate: The Simplest Usability Metric. Sited on http://www.useit.com/alertbox/ on 8th January, 2008.

[43] Nielsen, J. 2002. Usability for Senior Citizens. Sited on http://www.useit.com/alertbox/ on 10th January, 2008.

[44]] Nielsen, J. and Molich, R. 1990. Heuristic evaluation of user interfaces. In the proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people. Pages 249-256. ACM Press.

[45] Noiwana, A., J. and Norciob, F. A. 2006. Cultural differences on attention and perceived usability: Investigating color combinations of animated graphics. In the International Journal of Human Computer Studies, 64. Pages 103-122.

[46] Ogozalek, V. and Van Praag, J. 1986. Comparison of elderly and younger users on keyboard and voice input computer-based composition tasks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Boston, Massachusetts, United States, April 13 - 17, 1986). M. Mantei and P. Orbeton, Eds. CHI '86. Pages 205-211. ACM Press, New York, NY.

[47] Rau, P., P. and Liang, S., M. International and Localization: Evaluating and Testing a Website for Asian users. Ergonomics, 46 (1-3). Pages 255-270.

[48] Rohn, J. A., Rosenbaum, S., Humburg, J. 2000. A toolkit for strategic usability: Results from workshops, panels, and surveys. In the proceedings of CHI'200), Amsterdam. Pages 337-344.

[49] Rubin, J. 1994. HANDBOOK OF USABILITY TESTING: How to Plan, Design, and Conduct Effective Tests. John Wiley & Sons, Inc. USA.

[50] Ryu, Y., S., Lee, Y., Smith-Jackson, T., Nussbaum, M., and Tomioka, K. 2006. User Centered Design and Testing Methods for the development of the Menu System for Mobile phones. Sited on http://uweb.txstate.edu/ỹr12/Papers/HFES2006_Ryu.pdf on 16th March 2007.

[51] Sayago, S. and Blat, J. 2007. A preliminary usability evaluation of strategies for seeking online information with elderly people. In Proceedings of the 2007 international Cross-Disciplinary Conference on Web Accessibility (W4a) (Banff, Canada, May 07 - 08, 2007). W4A '07, vol. 225. ACM Press, New York, NY, Pages 54-57.

[52] Shades of Color Exhibition available at http://www.digimuseo.joensuu.fi

[53] Sinkkonen, I. Kuoppala, H. Parkkinen, J. and Vastamaki, R. 2006. Psychology of Usability. Edita Publishing Oy, Finland.

[54] Straub, K. 2004. Cleaning up for the housekeeper or why it makes sense to do both Expert Review and Usability Testing. In the UPA Voice sited on http://www.upassoc.org/upa_publications/upa_voice/volumes/2004 /nov/cleaning_up.html on 16th March, 2007.

[55] Sun, H. 2002. Exploring cultural usability. In the Proceedings of the Professional Communication Conference, 2002. IPCC 2002. IEEE International. Pages 319 - 330.

[56] The American Heritage Dictionary of the English Language, Fourth Edition. Houghton Mifflin Company.

[57] The usability & accessibility specialists. 2005. User-centered design-6 methods. Sited on http://www.webcredible.co.uk/user-friendly-resources/web-usability/user-centered-design.shtml on 12th March 2007.

[58] Tobias, C. L. 1987. Computers and the elderly; A review of the literature and directions for future research. In Proceedings of the Human Factors Society 31st Annual Meeting (pp. 866-870). Santa Monica, CA. Human Factors Society.

[59] Ullmer, B. and Ishii, H. 1997. The metaDESK: models and prototypes for tangible user interfaces. In Proceedings of the 10th Annual ACM Symposium on User interface Software and Technology (Banff, Alberta, Canada, October 14 - 17, 1997). UIST '97. ACM, New York, NY. Pages 223-232.

[60] Vatrapu, R. and Perez-Quinones, A., M. 2006. Culture and Usability Evaluation: The Effects of Culture in Structured Interviews. In the Journal of Usability Studies, Issue 4, Volume 1. Pages 156-170.

[61] Webcredible: the usability and accessibility specialists. 2006. Focus Groups; how to run them. Sited on http://www.webcredible.co.uk/user-friendly-resources/web-usability/focus-groups.shtml on 29th March 2007.

[62] Whiteside, J., Bennett, J. and Holzblatt, K. 1988. Usability engineering: Our experience and evolution. In M. Helander (Ed.) Handbook of Human-Computer Interaction.Pages 791-817. Amsterdam: North Holland.

[63] Yeo, A. 1998. Cultural effects in usability assessment. In CHI 98 Conference Summary on Human Factors in Computing Systems (Los Angeles, California, United States, April 18 - 23, 1998). CHI '98. Pages 74-75. ACM Press, New York, NY.

[64] Zajicek, M. 2001. Interface design for older adults. In Proceedings of the 2001 EC/NSF Workshop on Universal Accessibility of Ubiquitous Computing: Providing For the Elderly (Alcácer do Sal, Portugal, May 22 - 25, 2001). WUAUC'01. ACM Press, New York, NY.Pages 60-65.

[65] Zhao, H. 2001. Universal Usability Web Design Guidelines for the Elderly (Age 65 and Older). Sited on http://www.otal.umd.edu/uupractice/elderly/ on 5th February 2008.

APPENDIX A: USABILITY TEST PLAN

USABILITY TEST PLAN FOR THE "SHADES OF COLOR" DIGITAL MUSEUM APPLICATION

Prepared By: Mitwa Kaemba

This plan to evaluate the usability of a digital museum application in use at the North Karelian museum includes the following sections:

- Purpose
- Problem Statements
- User Profile
- Methodology
- Task List
- Test Environment and Equipment requirements
- Test Monitor Role
- Evaluation Measures
- Test Report contents and presentation

Purpose

Having identified a bottleneck in the use of digital museum applications by elderly users in RekiHaku, the first step in resolving it would be to assess the usability of a current application. The application selected for this purpose is one which was in use at the North Karelian Museum as part of the Shades of Colors exhibition (25.01.2007 – 02.09.2007).

The goal of the study is to reveal the nature and source of the problems being experienced by elderly users. The findings of this usability test will also serve as a starting point in the selection and implementation of improved applications making use of innovative interaction techniques.

The usability test will consist of predefined user tasks allowing the measurement of task completion times, error rates, number of assists sought and user satisfaction.

Problem Statements

The main concerns that are to be addressed in the test include:

1. Are elderly users able to navigate through the application quickly and easily?
2. Are elderly users able to understand and use the zoom controls easily?
3. Are elderly users able to use the touch panel system to make selections on the screen?
4. Is it possible for elderly users to use the application without online help?
5. What obstacles prevent the successful completion of tasks?

User Profile

A total of 8 participants are required for the test, with two being tested per day. For Joensuu, recruitment of participants is scheduled for Tuesday 9th November, at the Open University. For the study in Joensuu, participants will be tested beginning Tuesday 13th November 2007 and at RekiHaku beginning Tuesday 27th November 2007. Two test sessions will be held in a day one at 10am and the other at 2pm.

Characteristics that are relevant to the test are age, educational level, computer and museum experience. Of these characteristics the users will have age in common, as they are all to be 60 years of age or older. These participants may have varying levels of education, museum and computer experience. Highest level of education will be an option among the following: vocational training, high school, college/polytechnic, university and other (specified by the participant). Computer experience will be defined by the amount of time that the participant has been using computers for, that is, first time users, less than 1 year, 1 to 3 years and more than three years. The types of applications that non-first-time users make use of may also vary from one participant to another.

Methodology

1. Introduction

   The session will start by each participant being greeted and assured that the test is not a complex one in order for them to relax. After seating the participant down at a desk, a short verbal introduction explaining the purpose of the test and the role that they will play will be read out from a script. The script is also to include the reassurance that though the participants are being observed it is the application and not them who are under scrutiny.

2. Background Questionnaire
   The participants will then be assured of their anonymity in the test and at this point given the pre-test questionnaire to fill in. This questionnaire which is to collect basic background information will be filled in on the desk at which the participant is seated.
3. Test Scenario
   Having filled in the pre-test questionnaire the participant will be led to the touch panel system on which the tasks will be performed. The user will be handed the tasks one at a time and asked to read through them before beginning. Having read through it the test will begin and timers set. During the test, task completion time, number of repeat trials to perform a task, number of times verbal help is sought, general user behavior and any unusual circumstances will be noted by the test monitor. When one task is complete the test monitor will ensure that the second task begins from the home page of the application.
4. Debriefing
   When the tasks have been completed the participant will be led back to the chair and desk where they will be seated for a short debriefing session. The participant will first be asked to fill in a post-test questionnaire at which point the monitor will leave the desk to give the participant some privacy. When the questionnaire is filled the participant will be asked if there is something in particular they would like to say about the application and test. Also, the test monitor can raise questions about unusual behavior during the test. The participant will than be thanked, offered coffee and a free museum ticket as gratitude and released from the study.

Task List

It was noticed that the features of the application that users would mainly have a problem with are navigation and zooming. As such two tasks have been defined to explore these as described below:

Task 1: Navigation

User Instructions:

1. Go to the **RekiHaku Museum** page. 2. See how the weaving of kimonos is done. 3. Return to the main page.

Requirements:

- Running application displaying the home page
- Touchscreen

Successful completion criteria

Participant navigates to correct page, waits till weaving demonstration completes and successfully returns to the main page

Task 2: 3D manipulation

User Instructions:

1. Go to the **Chiba University** page. 2. Read the instructions. 3. Rotate the head gear up/down and left/write to view it from different angles 4. Return to the main page

Requirements:

- Running application displaying the home page
- Touchscreen

Successful completion criteria

Participant selects the correct rotation control, drags their finger over the object in order to rotate it and returns to the main page.

Task 3: Zooming

User Instructions:

1. Go to the **Colors of Kimonos** page. 2. Scroll down and select one small kimono image to view a bigger picture of it. 3. Select areas on the dress to view it in more detail 4. Return to the main page

Requirements:

- Running application displaying the home page
- Touchscreen

Successful completion criteria

Participant enlarges a small kimono image and drags their finger over it to view detail in zoom window .

Test environment and equipment requirements

The test sessions are to be held in actual context, that is in the North Karelian and RekiHaku Museums, as such no artificial conditions need to be created. Due to the language barrier there needs to be a native language speaker present throughout the test to assist in translation during introduction and when questions are raised.

Equipment:

- Touch Panel System running Windows XP
- Two chairs and a desk in close proximity to the touch panel system
- Pens, pencils and erasers

Test Monitor Role

The test monitor will be responsible for administering the questionnaires and tasks, and observing participant behavior throughout the duration of the test. As the test progresses the test monitor will time tasks, record number of errors and assists, and respond to questions asked. The translator present will be responsible for reading out the introductory and debriefing scripts as well as translating any questions that may arise.

Evaluation Measures

The measures to be collected and calculated in line with effectiveness, efficiency and user satisfaction include:

- Completion rate: the percentage of participants that successfully and correctly fulfill each task goal.
- The number of errors: which are being classified into two categories
  - Recoverable errors: where participant makes a mistake but manages to correct it on their own
  - Unrecoverable errors: where the participant cannot recover on their own and fails to complete a task
- Average time taken to complete each task.
- User satisfaction to be collected by general observation and the post-test questionnaire.

Test Report Contents and Presentation

Upon completion of all test sessions the findings will be reported in a document which based on the International Standard ISO/IEC 25062, will consist of the following sections:

- Executive Summary
- Introduction
  - Product description
  - Test objectives
- Method
  - Participants
  - Context of product use in the test: tasks, test facility, participant's computing environment, test administrator tools.
- Experimental design
  - Procedure
  - Participant instructions
  - Usability Metrics
- Results
  - Data analysis
  - Presentation of the results

- Appendices
  - Questionnaires
  - Participant general instructions
  - Participant task instructions
  - Release notes

The findings will be preliminarily discussed with supervisors in meetings to be held after each study (that is, in Joensuu and at RekiHaku). The formal report and recommendations will then be submitted to all concerned approximately two weeks after my return from Japan.

APPENDIX B: PRE TEST QUESTIONNAIRE

Usability Study: Background Questionnaire

Please help us understand your background and experience by answering the following questions. Mark the appropriate answer(s) with an X.

1. Gender:

Male		Female	

2. Age:

Younger than 55 years	
55- 64 years	
65- 74 years	
75 years or older	

3. Highest educational level attained:

Vocational Training	
High School	
College/ Polytechnic	
University	
Other (please specify) _____.	

4. Which best describes you (read through all options before making your selection)?

I am Finnish & have never been out of the country	
I am Finnish & have been on holiday outside the country	
I am Finnish but have lived (for studies/work/country of original citizenship/other purpose) outside the country	
If you have lived outside the country, please state where _____.	

5. On a scale of 1 to 5 how would you rate your skills in the following languages (1=poor and 5=excellent)

Language	**Speaking**	**Reading**	**Writing**
English			
Finnish			
Japanese			
Specify other: _____.			

6. Do you have any known or diagnosed memory problem?

Yes		No	

If your answer is Yes, please briefly describe how this problem affects your memory

_____

_____

MUSEUM EXPERIENCE

7. How long have you been visiting museums for?

This is my first time	
Less than 1 year	
1-3 years	
More than 3 years	

8. If you are not a first time visitor,
   a. How often do you visit museums?

Every week	
At least once a month	
At least once in 6months	
At least once in a year	
More seldom than once a year	

   b. What do you usually do when you visit a museum (mark all that apply)?

View physical exhibits	
Use computer applications provided by museum as part of exhibitions	
Follow guided tours	
Attend lectures	
Visit the museum cafeteria	
Other (please specify) _____ _____ _____	

COMPUTER EXPERIENCE

9. Which statement best describes how you feel about using computers (mark only one)?

I am not interested in using computers	
I do not mind using computers but I could do without them	
I am interested in using computers and would like to learn more about how to use them	
I cannot live without using a computer	
I am not sure how I feel about using computers	

10. How long have you been using computers?

This is going to be my first time	
Less than 1 year	
1-3 years	
More than 3 years	

11. If you are not a first time user,
    a. How many hours do you use computers for in a typical week?

Less than 1 hour	
1-3 hours	
4-6 hours	
6-10 hours	
More than 10 hours	

    b. What type of tasks do you generally use a computer for (mark all that apply)?

Typing documents	
Sending and receiving email	
Internet surfing	
Other (please specify)	

-Thank You-

APPENDIX C: POST TEST QUESTIONNAIRE

Usability Study: Post-test Questionnaire

Please help us understand your experience with the application by answering the following questions. Mark the appropriate answer(s) with an X.

1. Was this the first time that you have used this application?

Yes	
No	

2. Using the application during the test was ...

Very difficult	
Difficult	
Neither easy nor difficult	
Easy	
Very easy	

How much do you agree/disagree with the following statements:

3. I enjoyed using the application.

Strongly disagree	
Disagree	
Neither agree nor disagree	
Agree	
Strongly agree	

4. While using the application I always knew in which part of the application I was and where I wanted to go.

Strongly disagree	
Disagree	
Neither agree nor disagree	
Agree	
Strongly agree	

5. It was easy for me to use the zoom tool to get a detailed view of the kimono images.

Strongly disagree	
Disagree	
Neither agree nor disagree	
Agree	
Strongly agree	

6. It was easy to use the touch screen to select items on the screen.

Strongly disagree	
Disagree	
Neither agree nor disagree	
Agree	
Strongly agree	

7. There were adequate instructions and help provided on the screens of the application.

Strongly disagree	
Disagree	
Neither agree nor disagree	
Agree	
Strongly agree	

8. What did you like best about the application?

_____

9. What did you like least about the application?

_____

10. Do you feel that the use of computer applications is a good way of displaying extra information about exhibits in the museum?

Yes	
No	

If your answer to question 10 is No then, in what ways would you rather have extra information about exhibits displayed in the museum (Mark all that apply)?

Posters	
Pamphlets	
Through guided tours	
Lectures	
Please specify other:_____ _____	

11. Please add any comments and/or recommendations for the improvement of the application.

_____

_____

_____

-Thank You-

APPENDIX D: LIST OF USABILITY PROBLEMS, SOURCES AND RECOMMENDATIONS FOR IMPROVEMENT.

Given here is the list of usability problems identified during the tests with their corresponding severity rankings (4= unusable, 3= severe, 2= moderate and 1= irritant) . The sources of these problems and recommendations for improvement are also shown.

Severity ranking	Problem	Source	Recommendation
4	Not knowing what to do	- No visible instructions. - Active links/icons did not have button-like appearance - Non-uniform use of touchscreen	- Provide adequate instructions for application and input device use - Seek other ways of providing instructions other than text e.g. sound or movie clips - Give all icons standard button-like appearance - Ensure methods of using the input device is the same for invoking similar kind of actions
4	Difficult to understand use of controls	- Labeled in foreign language - Were too many in number - Were too small in size - Were different from 'normal' ones e.g. animated scroll bar	- During translation of application ensure all its parts are in a language understood by target users, or give option to change language - Develop custom made controls rather than making use of browser plug-ins that may not come in all languages - Custom made controls should however resemble as closely as possible to the ones that people are used to - Controls and icons alike should not be small in size - Avoid clustering controls and links leaving enough space between them
4	Difficult to read text	- Font too small - Texts too long	- Even though font may be adequately large always provide option to increase size further - Keep pieces of text to a minimum
4	Difficult to understand text	- Use of technical terms	- Avoid use of technical terms - In the event of translations, the literal meaning should be translated rather than having a word-for-word translation - Presentation of research results should be explained in a way that people without any background in Color research can understand
3	Unsure of what is happening with system	- No feedback from system to confirm action - No change in cursor shape or sound - Slow system response	- Give user constant feedback in form of sound/messages/a status bar/ change in shape of hour glass - Keep "heavy" images to a minimum to allow fast page loads - Provide online help for the user outlining steps to take in the event that something is really wrong with the application

3	Going back & forth between application sections	- Navigation loopholes - No way of keeping tab on where one is	- Ensure links take user where they are supposed to by conducting series of user tests prior to application release - Avoid the use of long pages that a user will have to scroll through - Keep to a minimum the “distance” that a user can travel through pages away from the home page - Provide a way for users to keep track of where they are, possibly by a visual hierarchy showing where they came from
2	Did not understand general application purpose	- Lack of general theme tying sections together - Unexplained Color Research information	-Avoid heaping unrelated information into a single application out of convenience - Group similar sections together under simple themes - Provide extra information about application content possibly in the form of printed brochures than can be placed near the computers
1	Early exit from & repetition of animation	- No playing options - Too many things happening on page	- Provide different versions of animation and/or video clips, some possible longer than others - Provide conventional start/stop/rewind/forward options - Avoid too many things happening on pages such as moving text