Remote usability testing for information appliances through WWW - with the emphasis on the development of tools

K-P. Lee Korea Advanced Institute of Science and Technology, Korea

Abstract

As the importance of interface design in information appliances became recognized, usability testing has been widely introduced. However usability testing in the environment of a closed laboratory has been known to cause some significant difficulties: cost, time, unnatural environment, and lack of opportunity for idea generation. The goal of this study is to propose the new prototype of tool for 'remote usability testing for information appliances through World Wide Web'.

At first, existing usability testing methods for information appliances are reviewed to identify major problems. Based on the problems of existing usability testing, the concept of remote usability testing through WWW is established and the prototype called RUTIA is developed and introduced. In the new tool, the specially developed browser is distributed to users who are required to download. Once users download browser they are guided through stages of usability testing: introduction, identifying user himself, performing given tasks over computer-simulated information appliances, and generating user's ideas on interface design. After users finished the given tasks and generation of ideas, all the interaction data including time taken, operational path, think-aloud and generated ideas are saved on a server for further analyses and generating solutions. In the analysis module, a researcher can conduct diverse analyses with saved data. Analyses can be done with various forms: visualized user's operational paths and a table of statistics of time and pressed buttons. It was found that some further refinements of tool are required: product size limited less than screen size and limited user type.

Remote usability testing for information appliances through WWW - with the emphasis on the development of tools

Introduction

Since the introduction of computer technology, the fundamental nature of products has become changed: more interactive and less physical. This change, in turn, has made new type of product come into being: 'information appliance' which is defined 'a computer-enhanced consumer device dedicated to a restricted cluster of tasks' (Mohageg and Wagner 2000). Mohageg and Wagner (2000) argue that the design approach for information appliance should be differentiated from conventional products and computer for two main reasons: very wide base of non-expert-consumers and different characteristics of information appliance itself. That is, work of human being with information appliance has become less physical and more mental, and accordingly the key criteria of effective worker performance have shifted from the physical one like the speed or range of motion of their limbs to the quality and flexibility of their thinking (Adler et al. 1992). In addition, the substitute of microchips for mechanical parts, the product has become less tangible and 'black box', which makes the key success factor of product as 'the usability: the capability to be used by humans easily and effectively' (Shackel 1991). Particularly, the concept of usability has become highly valued in the area of information appliances and software where users' works are mainly mental. Other advantages of usability listed include 'reduced customer support, service and training costs', 'avoidance of costly delay in schedule', 'simpler-to-prepare product documentation' and 'accurate, ready-to-use marketing claims based on tests' (Wiklund 1994). This introduction of the concept of usability to information appliances has led to the wide application of 'usability testing' for ensuring the quality of usability before launching the product to the market. Usability testing employs techniques to collect empirical data while observing representative end users using the product to perform representative tasks (Rubin 1994). Typical methods for usability testing include interview, guidelines, heuristics, cognitive walkthrough, prototypes, protocol analysis, cognitive modeling, observations and so on (Mack and Nielsen 1994). Although most of usability testing methods are useful in their own context, the most valuable method of usability testing is to let users perform tasks and observe them for its rich contextual data and users' direct behavior. Usually this kind of testing is done in the closed laboratory equipped with one-way mirror for uninterrupted observation, video recording facilities, and data logging computers. However, despite advantages of usability testing, usability testing in the environment of closed laboratory has been known to cause some significant problems: high cost, unnatural environment and limited focus only on measurement. The goal of this study is to develop the prototype of a tool for remote usability testing for information appliances by using World Wide Web.'

Problems of usability testing in the closed laboratory

Among existing methods of usability testing, the most reliable and frequently used method is the empirical experiment done in the environment of closed laboratory with representative users of the target population. Normally the environment is set up and simulated so that the user feels as natural like real working environment as possible. Representative users are brought to the laboratory and are given tasks to perform for evaluating the degree to which a product meets specific usability criteria: efficiency, learnability, memorability, errors, and satisfaction (Nielsen 1993). User's interaction behavior with product while performing task is observed and recorded. The observation of tacit user behavior in usability testing is one of the strong advantages compared with conventional user-studies like questionnaire or focus group interview. The observation of behavior can reveal problems in performing tasks which even users cannot be aware of while opinion-oriented user studies can only show problems users can recognize in the conscious level. For securing the reliable quality of data, the usability testing should go through rigorous process of multiple stages: developing the test plan, selecting and acquiring participants, preparing test materials, conducting the test, debriefing the participant, transforming data into findings and recommendations (Rubin 1994).

Despite of advantages of usability testing in the laboratory, there are some critical known problems. At first, setting up usability testing laboratory and running the usability testing cost lots of money and takes quite long time and significant efforts. Setting up usability testing laboratory requires spaces for testing room and observation and control room, full set of video recording equipment, computer and video monitors, video editors, time generators, intercoms, data logging software and other state of art electronic devices. This kind of setup requires a large capital outlay and commitment to testing by management. Without appropriate management of the laboratory, highly costing and sophisticated usability laboratory can be easily operated as *the world's most elaborate storage rooms* (Rubin 1994). Conducting usability testing also requires time, cost, and effort. Although 4 to 5 participants are known to be enough for a less formal usability test covering 80 percent of the usability deficiencies of a product (Virzi 1990), for a true formal experimental design, a minimum of 10 to 12 participants per condition must be utilized (Spyridakis 1992). However if a researcher wants to find some differences between groups like novice vs. expert, then the number of participants should be increased. All the subjects should be physically brought in the laboratory one by one and each of them should spend at least one to two hours for answering pre-test questionnaire, performing the number of given tasks and joining debriefing sessions. In addition, while subjects are performing tasks, there should be (usually in the other side of one way mirror) other people such as test monitor, data logger, timers, video recording operator, product/technical expert, and test observers. After recording of all user performance is done, the data should get through exhaustive analysis process: measuring time, picking up errors, logging the data, transforming data into findings and recommendations and so on.

Secondly, problem lies in the unnatural atmosphere of laboratory where users participate the testing. The closed environment of usability testing room equipped with one-way mirror and video cameras are very impersonal. Except for special cases where moderator joins the testing with subject, in usual cases, a user is left alone to perform the tasks according to the instructions given through intercom. Although there are number of techniques to soothe subject's uncomfortable emotional state, this kind of 'prison-like environment' can intimidate inexperienced users and they can easily get nervous. This so called 'guinea pig' syndrome makes the subject feel overly self-conscious during the test, which prevents them from showing natural responses and performance.

Finally, a session of usability testing focuses mainly on 'measuring' aspects: i.e. a researcher focuses on evaluating the degree to which a product meets specific criteria. As a result, types of data from a session of usability testing include data on time duration for performing tasks, number of errors, percentage of tasks completed successfully, ratings or rankings of the product, and number of negative references to the product. However, users' suggestions for new idea are also as much important as measuring the usability. For user's participation to idea generation, some methods have been developed such as 'exploratory test (Rubin 1994)', 'card sorting', 'scenario-based design (Carroll 1995)', 'collaging', 'velcro modeling', and 'cognitive mapping' (Sanders and Williams 2001). These kinds of user-participatory design methods should be more systematically incorporated in the process of usability testing. These three main problems should be considered in the development of new tool for usability testing.

Development of tool for remote usability testing through WWW

There have been developed various ways to solve above-mentioned problems in conventional laboratory based usability testing. Those include 'third-party laboratory evaluation', 'third-party usability inspection', 'remote questionnaire or survey', collaborative remote evaluation', 'video-conferencing-supported evaluation', 'instrumented or automated data collection for remote evaluation' and 'user-reported critical incident method' (Castillo 1997). The common point those methods share lies in the 'remoteness' of location where evaluators are separated in space and/or time from users. The first problem of expensive cost for setting up usability laboratory and running usability can be solved by 'automating' implementing 'on-line remote usability testing'. Castillo further classified those remote usability testing methods by number of dimensions: types of users involved, time of evaluation, user location during evaluation, person or role who identifies critical problems, types of tasks, level of interaction between user and evaluator, types of data gathered, type of equipment used for collecting data, cost to collect data, cost to analyze data, and quality or usefulness of collected data. Particularly, by the major concern of cost-effectiveness and quality of data, methods are mapped like Figure 1. According to Figure 1, the most reliable quality of collected data, the method that costs less, is identified 'instrumented or automated data collection'.

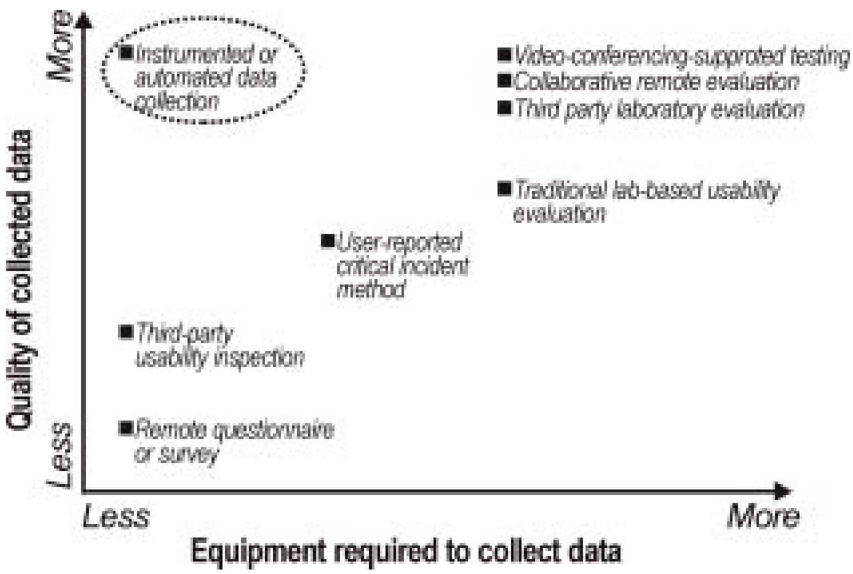

Figure 1: Remote usability testing methods in terms of quality of data and equipment required (Castillo 1997)

Instrumented or automated data collection method refers to instrument some application program to automate the collection of a log of data occurring as a natural usage in users' normal working environment. Once user downloads and installs an application program in his or her computer, all user should do is just to work normally as usual in his own environment. Then the application program takes care of collecting and reporting data such as program usage, project time, internet usage, comments to usage, keystrokes and mouse movements, and any other activity. Nielsen (1996) also mentioned the advantage of collecting usability data through Internet. These advantages of 'remoteness', 'asynchrony', 'natural environment', and 'simple management' cover almost all the problems identified above in the traditional usability testing. However there still remain some problems uncovered. At first, it still does not allow users to participate in idea generation. User is only 'using' the testing product without having 'testing and generating ideas' in his mind. All the data evaluator can get is users' usage pattern rather than their conscious effort to reflect their ideas on interface design. Secondly, major area of application of instrumented or automated data collection method is limited in software or web. Information appliance requires different attributes to be added to the instrumented or automated data collection method. The method should allow a researcher to have more control and collaborative attribute. Keeping these in mind, the new tool of remote usability testing for information appliance was developed with the following objectives:

- Use Internet for automating usability session.
- Make users' testing environment very comfortable and natural.
- Let users participate to generate ideas for interface design.
- Make the way of collecting user's performance data as simple as possible.
- Make the tool work with the collaboration of other related tools like ordinary statistical programs or word processor.

Structure of Tool-RUTIA

The tool called RUTIA (Remote Usability Testing for Information Appliance) was developed based on objectives mentioned above. RUTIA has the structure comprising of three main modules: testing module, idea module and analysis module. These three modules go through the process shown in Figure 2.

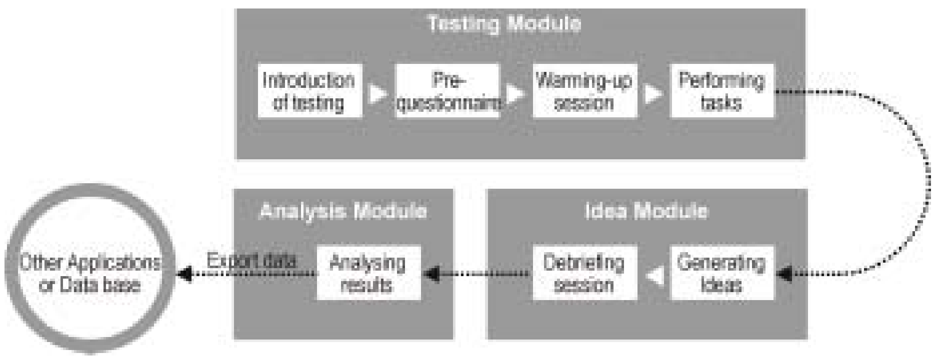

Figure 2: The structure and process of the tool, RUTIA

Testing module

In the testing module, at first, the overall purpose and process of the usability testing is introduced and then user is guided to input their demographic data like gender, age and so on as pre-questionnaire before actual usability testing. In the pre-questionnaire other items than demographic data can be included for basic data for later analysis: e.g. user's experience with testing product or general use-behavior. After understanding user's basic profile, a warming-up session is given to user for familiarizing herself or himself with on-line usability testing. The warming-up task is usually very simple like setting up alarm of digital clock. If user feels confident enough to her or his capability of interacting with the warming-up task, then user can go ahead to start main usability testing. Various tasks are provided one by one and user performs the tasks by operating the computer-simulated information appliances. User uses mouse to press control buttons, for which the product responses exactly same as real product: display, sound, or other various states. While performing tasks, user can refer user's manual for help or skip the task if she or he cannot continue the task for its difficulties at anytime. In addition, if user is equipped with microphone she or he can perform 'think aloud'. The sample screen of testing module is composed as shown in Figure 3. As shown in Figure 3, the computer-simulated product for testing is shown on the screen with the task bar, other control buttons for skipping task and opening user's manual.

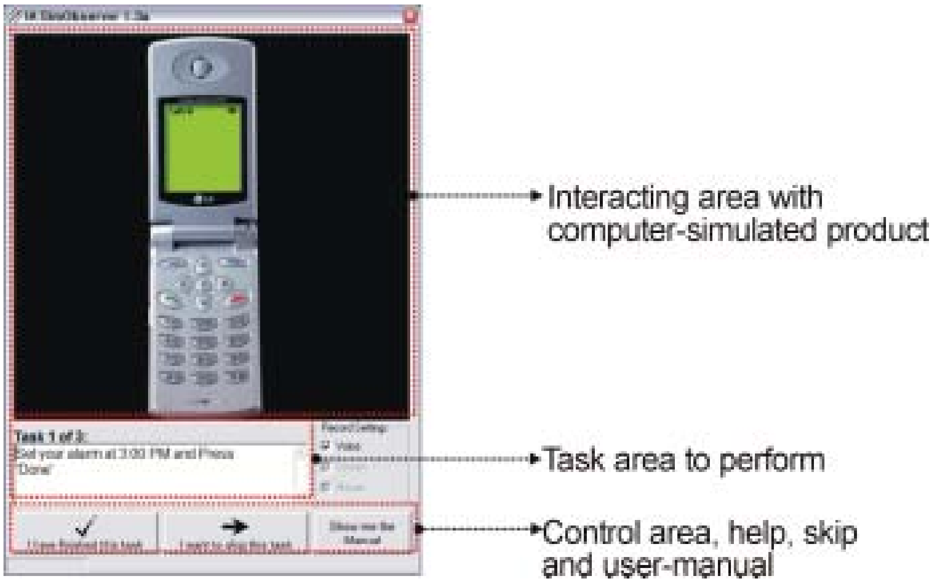

Figure 3: Sample screen of testing module

Idea module

After finishing all the given tasks, user is guided to participate in the idea-generation session. In the session of idea-generation, user can actively generate her or his own ideas regarding layout of control buttons, grouping menus, arranging interacting process and organizing interface structure. Figure 4 shows sample screen of user's idea generation of layout of control button. User can drag control buttons and configure his preferred way of layout. Or he can comment new ideas for improving the usability in 'idea box'. These kinds of various ideas generated by users themselves can be important means to understand their conceptual model on user-interface of testing product. After ending up with the session of idea-generation the debriefing session starts to ask few more questions regarding test itself or to get other feedbacks from users. Or some other additional questions can be given to users: asking some reasons for particular behaviors.

Figure 4: Sample screen of idea module.

Analysis module

As soon as the user finishes the usability-testing session and idea-generation session, all the usability data including time taken, operational path, voices recorded from ‘think aloud’ and user’s idea generated in the idea-generation session are transmitted to and saved in the evaluator’s server. These data are analyzed in various ways for finding problems and insights for generating solutions.

At first, all the interacting processes by users while they were engaging the usability testing are replayed with exactly the same operational paths, sequences and time. The operational traces are visualized in line over the product so that analyzer can easily see how user interacted, moved around, made errors and so forth in sequence. This replay is done with the interface like a VCR: a researcher can stop, pause, play, fast forward or rewind by clicking control buttons. In addition, user’s protocol data from think aloud is also replayed. If analyzer needs to analyze user’s particular interaction in detail, he can always stop and resume the play.

While the user’s interaction is replayed, at the same time, the interaction process is visualized in another way. This time, the user’s interaction is visualized in terms of interface structure. In parallel with a small window of product where user’s interaction is replayed over actual physical product, there is another small window where interface structure of product is shown. A researcher can see user’s interacting behavior in terms of the structure of interface: how deeply user went down, how frequently user changed the level of interface depth and so on. The play in the window of interface structure is synchronized with the play in the product window so that a researcher can get the view of a user’s interacting behavior on product and interface structure simultaneously.

Finally, all the data are summarized in the table: pressed buttons, time taken, user’s action, sequence, user’s protocol data, and researcher’s comment. In this summarizing table, a researcher can sort out the time and easily search for specific interaction by simple click of relevant cell of table. For further analyses, researcher can cross-tabulate between different elements. For example, researcher can find out what type of users made specific types of errors by cross-tabulating element of 'user' and 'error'. All the data can be exported to other conventional software like statistical program or word processing software for further analyses. These data can be accumulated for certain period of time to make database for usability. The sample screen of analysis module is shown in Figure 5.

Figure 5: Sample screen of analysis module

Conclusion and further study

The current tool is still in the stage of development of prototyping. The tool shows potential advantages in several respects: low cost of management, easy collection of testing data, short time to conduct usability testing, provision of natural atmosphere to user to test the usability, user's active participation in idea generation and availability of diverse ways of insightful analyses.

However, even with those various advantages of on-line remote usability testing, it needs further refinements in several respects. At first, a product bigger than the size of computer monitor-screen can cause a problem because the product shown on the screen is shown smaller than actual size. This problem can seriously reduce 'the reality' of product. Secondly, since the tool is working on World Wide Web, types of users participating in usability testing can be limited only in those who can access and use Internet without any serious difficulties. For effective implementation of the tool these problems should be further improved.

References

Alder, Paul S. & Winograd, Terry A. 1992. “The Usability Challenge” In *Usability: Turning Technologies Into Tools* edited by P. S. Alder & T. A. Winograd. New York: Oxford University Press.

Carroll, John M. eds. 1995. *Scenario-Based Design: Envisioning Work and Technology in System Development*. John Wiley & Sons, Inc. : New York.

Castillo, J. C. 1997. *The User-Reported Critical Incident Method for Remote Usability Evaluation. Unpublished Thesis.* Virginia Polytechnic Institute and State University. Blacksburg, VA.

Mack, Robert L. and Nielsen, Jakob. 1994. “Executive Summary” in *Usability Inspection Methods* edited by Jakob Nielsen and Robert L. Mack. New York: John Wiley & Sons.

Mohageg, Michael F. and Wagner, Annette. 2000. “Design Considerations for Information Appliances.” In *Information Appliance and Beyond*, edited by B. Eric. San Francisco: Morgan Kaufmann Publishers.

Nielsen, Jakob. 1993. *Usability Engineering*. Boston: Academic Press, Inc.

Nielsen, Jakob. 1996. “International Usability Engineering”, In *International User Interfaces* edited by del Galdo, E. M. & Nielsen, J. New York: John Wiley & Sons.

Rubin, Jeffrey. 1994. *Handbook of Usability Testing*. New York: Wiley & Sons, Inc.

Sanders, Elizabeth and Williams, Colin. 2001. “Harnessing People's Creativity: Ideation and Expression through Visual Communication”, In *Focus Groups: Supporting Effective Product Development* edited by Langford J and McDonagh-Philp D. Taylor and Francis

Spyridakis, J. H. 1992. “Conducting Research in Technical Communications: The Application of True Experimental Designs.” *Technical Communications*, Fourth Quarter: 607-624.

Shackel, B. 1991. “Usability-Context, Framework, Design, and Evaluation”, In *Human Factors for Informatics Usability* edited by Shackel, B. & Richardson, S. Cambridge: Cambridge University Press.

Virzi, R. A. 1990. “Streamlining in the Design Process: Running Fewer Subjects.” *Proceedings of Human Factors Society*: 291-294.

Wiklund, Michael E. 1994. “Introduction”, In *Usability in Practice*, edited by Michael E. Wiklund. Boston: AP Professional.