Usability evaluation for mobile device: A comparison of laboratory and field tests

Conference Paper · January 2006

DOI: 10.1145/1152215.1152254 · Source: DBLP

3 authors, including:

Henry Been-Lirn Duh
La Trobe University
147 PUBLICATIONS **2,520** CITATIONS

SEE PROFILE

Vivian Hsueh Hua Chen
Nanyang Technological University
51 PUBLICATIONS **646** CITATIONS

SEE PROFILE

Usability Evaluation for Mobile Device: A Comparison of Laboratory and Field Tests

Henry Been-Lirn Duh
School of Mechanical and Aerospace Engineering
Nanyang Technological University
Singapore
65-67905151
mblduh@ntu.edu.sg

Gerald C. B. Tan
School of Mechanical and Aerospace Engineering
Nanyang Technological University
Singapore
65-97491110
S8039694E@ntu.edu.sg

Vivian Hsueh-hua Chen
School of Communication and Information
Nanyang Technological University
Singapore
65-67905833
chenhh@ntu.edu.sg

ABSTRACT
Usability testing of mobile devices is an emerging area of research in the field of Human-Computer Interaction. Guidelines had been established as to how usability tests should be conducted. However, there are limitations to the effectiveness of conventional usability tests, especially for mobile devices. Mobile devices typically are used in different situations, but current evaluation method cannot uncover problems in all type of situations. Hence, this study evaluates the effectiveness of conventional laboratory usability tests. It investigated the differences between usability tests on mobile phones conducted in laboratory and real life situation. Significant differences were found, including the frequency and severity of usability problems encountered, the users' behavior, and subjective responses to the device and the interaction.

Categories and Subject Descriptors
H.1.2 [**Information Systems**]: User/Machine Systems - *human factors*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - *Evaluation/methodology*

General Terms: Measurement, Documentation, Performance, Design, Verification, Human Factors

Keywords
Usability, mobile devices, dynamics environment.

1. INTRODUCTION
Mobile phones have become overwhelmingly popular in today's society. In Singapore, eighty-three percent of the total population, which is equivalent to 3.4 million people [1], subscribed to mobile phone service as of January 2004. Increasing sales of mobile phones therefore has become a major goal for many companies.

MobileHCI'06, September 12–15, 2006, Helsinki, Finland.

Usability and user friendliness are the two important determinants for the popularity of mobile phones. A good design is one of the critical factors that influence sales. A good design can also reduce mental and physical stress, reduce learning curve, improve user-device operability in using the device and thus improve overall product quality. In order to develop good designs, a series of rigorous design and development processes have to be undertaken, involving usability engineering as well as technological considerations. Usability Engineering is a systematic and practical approach that can be applied to mobile phone design. It consists of numerous methodologies, to be used in conjunction with one another, for designs of products. These methods include defining users, gathering design requirement specifications, prototype development and testing, usability evaluation and so on.

Conventional usability evaluation is mostly laboratory based. Most of them try to simulate scenarios of users' day-to-day usage of the mobile device for evaluation. This approach however cannot account for uncontrollable factors that influence mobile phone usage in real life situations. For mobile devices, due to the highly dynamic context of use, laboratory usability and field usability tests would involve different issues and might uncover different usability problems. In order to address and improve weakness of current usability evaluation, this study seeks to compare and contrast the results of laboratory and field usability tests on mobile phones. Specifically, it utilizes the same usability test in laboratory and field environment to understand the differences.

The current study looks at a scenario frequently encountered in Singapore, taking the MRT (Mass Rapid Transit train). The MRT is a form of public transport which serves up to more than 2 million commuters a day, approximately more than 50% of the total Singapore population according to the Land Transport Authority Annual Report [9]. The task scenario is set up based on the guidelines for setting the task scenario [3][17]. These guidelines include providing realistic scenarios and task analysis, arranging the tasks by the priority in which they are most likely to be performed, matching the tasks to the experience of the general users, incorporating sufficient mental and physical demands and amount of work to be done for each task. The scenario will be described in detail in the latter sections.

2. Literature review

2.1 On Usability Evaluation Methods

There are many methods for usability evaluation. These methods can be categorized into four main areas, namely, Heuristic evaluations, Cognitive walkthroughs, Usability testing, and Comparison against guidelines [19]. To decide which evaluation method to use depends on the strengths and weaknesses of the method, as well as its applicability with regards to the researcher's objectives. There are increasing concerns and critiques of the validity and effectiveness of those methods from human factors and human-computer interaction aspects.

Usability tests, being part of the Usability Engineering process, are performance measurement to determine whether usability goals are achieved. It is usually conducted in laboratories with test participant(s) performing a pre-defined set of tasks while data on performance measures are documented [3][17] . Usability tests are conducted to collect quantitative data on usability problems (difficulties that users come across while using), performance of devices (in terms of accuracy, time to carry out functions, etc.) and mental/physical demands of using device, etc. Many types of usability problems will arise from the tests conducted and can be generalized according to their importance. Typical categorization is critical problems, then serious problems to the lesser cosmetic problems. Statistical analysis can then be done for comparison between tests. With such measures, the design of the evaluated device or system and its interface can be and improved based on the result of evaluation.

For laboratory based usability tests, the difficulties in conducting and collecting data are significantly reduced as compared to field based usability tests. However, those tests cannot address factors and issues that occur at the field. Research indicates some drawbacks in usability testing methods [9]. How much the simulated scenario of usability testing can represent real life situation is one critical factor in terms of the validity of the usability test. The issues involved are complications of simulating the scenario, the applicability of utilizing established evaluation methods in the field setting and data collected are subject to unknown variables which might not be uncovered and potentially affecting the results. To further understand those problem areas, this study seeks to use established evaluation method in both the laboratory and field situations. The usability tests are done on mobile phones.

2.2 Usability Studies on Mobile Phones

Usability research on mobile phones, especially WAP services, is an emerging area of interest. These usability studies utilized mainly conventional usability testing methods. Ramsay [16] conducted a usability study on WAP when it was first introduced to the market. He investigated the usability of WAP from a user's point-of-view and aimed at highlighting important factors to form the guidelines and reference for future WAP user interface designs. There were a lot of technical issues regarding WAP, such as connectivity rates, data access, etc and more information was needed from the usability point of view. Condos & et al. [2] study on WAP evaluated two main UK WAP portals based on a survey on usage and future trends as well as a WAP usability evaluation. They compared the methodologies and results of their study with past research and developed usability principles (a total of 10 guiding principles were written), thus contributing to further developments of WAP and M-commerce. Pascoe, Ryan and Morse [15] studied the effects of using mobile devices while "on the move", specifically HCI issues involved in using mobile devices in the dynamic/field environment. The mobile device used was a personal digital assistant (PDA) with a customized interface. The four main issues are dynamic user configuration, limited attention capacity, high-speed interaction and context dependency.

There were many issues regarding usability for users, mobile devices and their interactions in the dynamic/field environment. Typically, the features that influence the interaction between users and mobile device include contextual awareness, task hierarchy, visual attention, hand manipulation and mobility. These features may involve different elements in different scenarios but they are critical when considering the issue of usability and design. Key issues regarding mobility in the field context and user interaction are highlighted by Gorlenko and Merrick [4]. They investigated the issue of mobility in the dynamic/field environment. Mobility requires users to apply more cognitive abilities. Also, existing ergonomic usability design principles cannot be applied to mobile computing devices because mobility is the key to mobile devices. Mobility can be considered from the users' point-of-view, such as what constraints it would bring to the users, as well as from the technical perspective. Also, factors in dynamic/field environment have to be taken into consideration. They include fluctuation of temperature and lighting conditions, varying levels of noise and distractions, mobility of user, competition for resources in multitask mobile settings, and the need to manipulate other physical objects during interaction.

With regards to mobile phone studies associated with the dynamic/field environment, Nunes and Recarte [14] studied mobile phone use and road safety. The purpose is to understand how cognitive demands of conversation, using a mobile phone and in person, affects driving. Their laboratory simulation tests emulated the physical and mental demands of the real scenario. A visual distraction task was set up with targets and response button locations for the test subjects to react to the stimuli. A total of four experiments were done to investigate how each of the factors correlates to each other. They find that there is an inverse relationship between cognitively demanding phone conversation tasks and visual processing capacities. As cognitive demands get higher, the visual processing capacity will be reduced. Another research by Monk, Fellas and Ley [11] investigated how the interaction of two parties involved in a phone conversation affects the notice-ability and intrusiveness of mobile phone conversations. Questions were raised regarding the testing to be done in the "natural" environment of the train. The real environment meant that test subjects would behave as they would normally, to the extent that they need not be told to do so. Many factors come into play as it was not done in a controlled environment. For instance, restrictions were being placed by the rail company. The social context also influence the results.

2.3 New usability evaluation techniques for Mobile Devices

A recent survey by Kjeldskov and Graham [5] showed that 71% of all mobile HCI evaluations are laboratory evaluations and very few conventional usability tests are modified. These generally

included studies on different usability evaluation methods and comparisons between methods, usability studies on mobile phone services such as WAP and phone conversation, and studies on issues associated with the dynamic/field environment. There is no agreed consensus on the better evaluation techniques for mobile phones and little research has been done on evaluating the evaluation methods specific to devices and their context of use.

There were few papers that had looked into this area of developing new techniques for evaluating mobile devices. Kjeldskov and Stage [7] searched literature from Proceedings of the International Conference Mobile HCI: 1998, 1999 and 2001, Proceedings of the ACM Conference of CHI: 1996-2002 and ACM Transactions on CHI: 1996-2002. Only 114 out of 636 papers deal with HCI for mobile systems and only 6 out of the 114 papers try new techniques for usability evaluations. Among the 6 papers, the new usability evaluation methods could be categorized into two types. One type uses a treadmill, pre-defined track or real world route to simulate walking conditions and the other uses a car simulator to simulate driving scenarios. They conclude that traditional techniques were insufficient for evaluating usability for mobile devices at this point of time.

New techniques have been introduced but they are still in the stage of been validated. More research need to be done to prove that those new techniques are superior in evaluating usability for mobile devices due to the complexity of the dynamic/field environment. Kjeldskov and Stage [6] proposed a series of new usability techniques to evaluate mobile devices and conduct tests both in the laboratory and dynamic/field environment. Their discussion only focus on usability problems, mobility and data collection in the dynamic/field environment. Factors such as the user's subjective opinions, and behaviors were not addressed.

Kjeldskov et al. [7] also did a similar study on mobile device, investigating the effects of users' immediate surroundings. They also compared laboratory and field usability tests, heuristic walkthrough and rapid reflection. The analysis compared different usability problems found. There are some innate differences, such as the social comfort of use, and "system in the world" issues. Generally, most basic usability issues are similar. They did not address how such differences would affect the test results. The result shown allows one to select the more appropriate method of evaluation for different purposes, rather than improve the evaluation method itself. In order to fill in the gaps in current literature, this study attempts to address how the difference between laboratory and field setting affect usability test results as well as users' subjective opinions and behaviors in using mobile phone.

3. USABILITY TESTS AND PROCEDURE

The objective of this study is to identify the differences between usability tests conducted in laboratory and in dynamic/field environment. A series of usability test were done in Center for Human Factors and Ergonomics (CHFE) laboratory at Nanyang Technological University's (NTU) and on MRT in Singapore. The tests were done on the mobile phone use only. Specific objectives include:

- To find out the quantity and quality (in terms of severity) of the usability problems and for each task in both settings.
- To find out the usability performance measures of each task done in both settings.
- To find out users' satisfaction and subjective feelings in using the mobile phone for each task in both settings and the reasons.
- To find out users' behavioral patterns while using the mobile phone for each task.

Previous research and usability guides conclude that four or five participants will reveal approximately 80% of the usability problems in any product. Similarly, using ten participants will usually result in 90% of all usability problems detected. [3][12] [17]. Therefore, ten participants were recruited from NTU. Similarly, the usability test in the MRT environment itself also involved 10 other participants for consistency in comparison. Two testers were involved for the laboratory test, a test monitor and a correspondent who communicate with the participant for the phone conversation task. An additional test monitor assistant was involved in the tests done on MRT.

The mobile phone used is Nokia 6220, equipped with all the functions required for finishing the tasks. For both the laboratory and dynamic/field usability tests, other equipments included:

- 2 Digital Video Cameras (Panasonic Model NV-GS70)
- CMOS Video Camera (Model 502C)
- Correspondent Mobile Phone, subscribed to Mobile One (service provider) phone Network

Additional equipments were used for the laboratory usability tests:

- Color Quad Processor (Model AVC 703)
- Digital Video Camera Tripod Stand

The pre-test questionnaire include demographics, participants' experience of using mobile phones, level of familiarity of using mobile phones and the four types of tasks, the usage frequency of phone and SMS functions. This information is useful for comparing "familiar" and "not familiar" (regarding the phone service such as SMS, MMS etc.) participants, which could influence the result of usability testing. An additional question regarding the study's main objective was included to make sure participants know the research objective and gather feedback from them.

The post-test questionnaire was mainly used to gather information on the participant's subjective response to the mobile phone functions tested, and information that could not be obtained from the audio/video data. It includes questions on the ease of performing the tasks, the user-friendliness of the functions, the satisfaction of performing the tasks, opinions on the functions tested, key usability problems encountered and lastly on the differences between two environments. A five point Likert scale and open-ended questions were both used.

Each participant was given a scenario for performing the tasks. The Task Scenario is "You are on the MRT, taking the NS line from Raffles Place to Toa Payoh (or opposite direction) as you would everyday taking the MRT from school back home. You always carry your mobile phone around everyday. You are standing and leaning against the side wall in the area between 2 MRT cabins".

The tasks were designed based on the "prioritize by frequency" principle. The "prioritize by frequency" was developed from NTU's Center for IT Services (CITS) in 2003 [13], as well as a pre-test survey conducted on 100 student samples from NTU by researchers in this study. There are six tasks (see Table 1).

Table 1. The Test Scenario and Tasks involved

Task	Scenario of action	Task Description
1	You need to inform your friend about your personal particulars as he needs to fill up a form for you. You decide to call out.	1. Dial out to contact Gerald from mobile phone contacts list. 2. Start a conversation upon pick up as you normally would. 3. Verbally inform the contact your full name, NRIC, address and date of birth
2	You receive a call from a friend on your mobile phone. You answer the phone call.	1. Answer phone call as you normally would. 2. Start a conversation with the friend.
3	You need to inform your friend about your personal particulars information as he needs to fill up a form for you. You decided to SMS out.	1. Compose a SMS including the following information: your full name, NRIC, address and date of birth. 2. Send SMS to Gerald from mobile phone contact list. 3. Reply again to Gerald if necessary, i.e. if Gerald replied your message.
4	You received a SMS earlier and stored in the mobile phone's Inbox. You decide to reply now.	1. Go to mobile phone SMS Inbox. 2. Reply to the SMS from Gerald. 3. Reply again to Gerald if necessary, i.e. if Gerald replied your message again.
5	It is your friend's birthday and you decide to create a Multi-media Message (MMS) and save it, to be sent out later in the day.	1.Create a MMS with the following parameters: 10s Slide timing, Text on top, insert Image001.jpg from Images folder, and text content "Happy Birthday" 2. Preview the MMS and save it. 3. Access the saved items folder and check the content of the MMS.
6	You decided to surf M1 WAP Home website.	1. Go to Services and Home 2. Go to M1 Services 3. Go to M Coupons 4. Read the M Coupons available and take note of what it is. Disconnect GPRS after.

For the laboratory usability test, the scenario was made known to the participant via the briefing script and the task scenario description sheet. The participants would have to think that he or she was in that situation. For the field usability test, the participants were in the scenario setting itself but had to act according to the instruction for each task.

The correspondent received a set of instructions for the correspondent to react or initiate an interaction with the participants. The test monitor prepared a set of scripts used to brief the test subjects for consistency. The participants received the Consent Form, Task Scenario, Pre-test and Post-test Questionnaire. The various documents were designed based on the various guides and references from research.

The usability testing and data-logging method used was thinking aloud. This method was introduced to each participant and they get a chance to practice think aloud prior to the actual test. The entire test session was audio and video recorded.

In the laboratory setting, two digital video cameras were linked to a color quad processor. The CMOS Camera was linked to one of the digital video camera and attached via the flexible support stand and positioned to capture the mobile phone screen interface. The other digital video camera would be used to capture the facial expression of the participant. Using the quad processor, the two captured video images would be linked side by side for ease of post-test analysis (See figure 1).

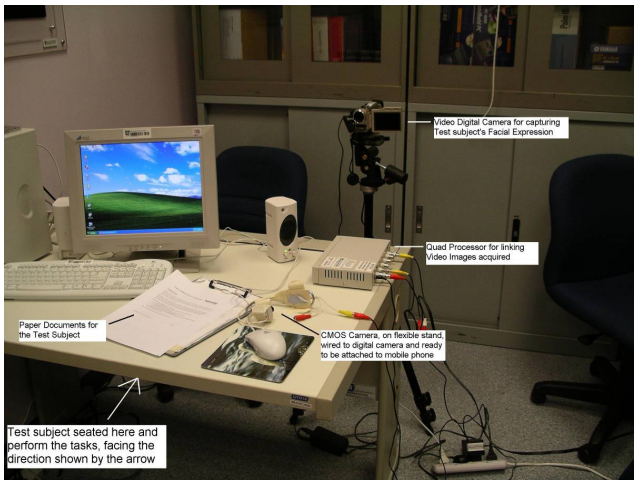

Figure 1. Usability Test Setup

For the dynamic/field environment setting, an additional role of test monitor assistant was added. The test monitor, test monitor assistant and participant would all board the MRT train at Marina Bay Station, and the test commenced upon reaching Raffles Place Station. The CMOS Camera would still be linked to one of the digital video cameras and attached to the mobile phone via the flexible support stand to capture the mobile phone interface. The test monitor assistant started recording the test session once the train reaches Raffles Place Station. The test monitor assistant was also in charge of holding the task list board at a comfortable viewing position for the participant to read. The correspondent was not present at the actual testing place. The test monitor would capture the participant's facial expression using one of the video digital cameras and audio-recorded his observation.

4. DATA ANALYSIS & RESULTS

Based on the data from the pre-test questionnaire, there were no significant statistical differences for demographic information between the two groups of participants. Participants are more

familiar with Nokia 6220 phone and SMS function than MMS and WAP for both sets of participants.

Usability problems found between the Laboratory and the MRT tests were analyzed using t-test. Usability problems were categorized according to the standard provided by Molich [10]. The criteria are illustrated below.

Critical problems

- Prevented test subject from completing tasks and/or
- Recurred across all test subjects

Serious problems

- Increased test subjects' time to complete task of test subjects severely and/or
- Recurred frequently across test subjects and
- Test subjects still managed to complete task eventually

Cosmetic problems

- Increased test subjects' time to complete task slightly and/or
- Recurred infrequently across test subjects and
- Test subjects could complete task easily

There is significant difference in the total number of usability problems revealed between the two usability tests ($t[17] = 9.19$, $p < 0.05$). The total number of usability problems for lab test is 92 and 171 for MRT test. The mean for all usability problems is 9.2 for lab test and 17.1 for MRT test. The standard deviation of all usability problems is 1.87 for lab test and 1.97 for MRT test. Figure 2 demonstrated the quantity of usability problems. There is an increase in the number of usability problems across all three types of usability problems on MRT. There were a lot more critical problems associated with phone conversation in the MRT environment, comparing to serious and cosmetic problems. Some problems were specific to the environment itself, meaning that there were no corresponding occurrences in the laboratory tests.

The major difference is that there was a generic increase in the occurrence usability problems for the MRT usability tests and associated issues in the severity of the usability problems found. Usability problems also existed that were unique to the usability testing environment.

The participants' behaviors are classified into 3 categories, namely positive, negative and neutral behavior. Behaviors could be related to usability problems, the context of testing, or the task itself. Positive behaviors include friendly facial gestures such as smiles and relaxed expressions. Negative behaviors refer to the opposite of positive comments and behavior. Neutral behavior tended to be more environmental and task related. For example, how a participant reacts to reading of the stored Inbox SMS in Task 4 (Reply SMS). Researchers went through the behavior logged data for analysis.

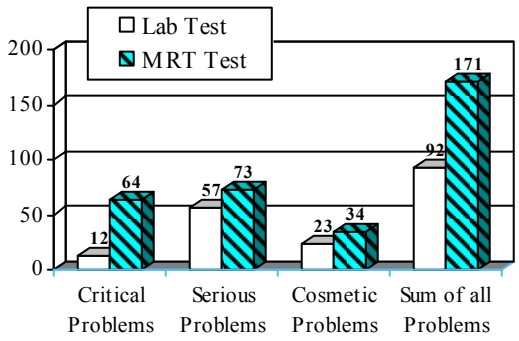

Figure 2: Comparison of Usability Problems between the Laboratory and MRT Usability Test

There is significant difference in the pooled sum of the behaviors notions between usability tests done in two different settings. ($t[17] = 9.30$, $p < 0.05$). Figure 3 compares differences in participants' behaviors while performing tasks. The participants in the MRT behave less positively and more negatively. There were key differences in behaviors exhibited which are unique to the usability testing environment. Participants also generally took longer time to perform the tasks, especially for MMS and WAP tasks.

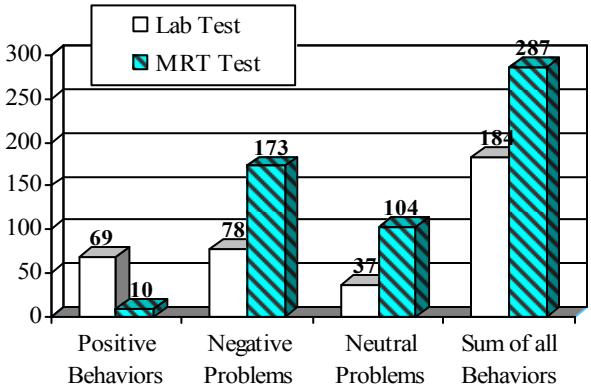

Figure 3 Comparisons of User Behaviors between the Laboratory and MRT tests

For the participant's time to perform tasks, the total time to perform all tasks were significantly longer for MRT Tests and only phone calls & MMS tasks exhibit significantly longer for MRT Tests based on t-tests. For the post-test questionnaire responses, there were generally significantly poorer responses for MMS and WAP Tasks.

Above discussion indicates that the laboratory usability test was insufficient in finding usability problems and evaluating the mobile device in its actual context of use. Since the participants' behaviors, while performing the usability tests, in the laboratory were different from their behaviors in the actual situation of mobile phone use. This would in turn affect the results of the usability tests, usability problems encountered and performance.

Based on analysis of pre and post-test questionnaires, participants identified reasons for differences listed above. They are listed in the order of importance below.

1. Noise level of the MRT environment
2. Moving MRT Setting
3. Lack of Privacy in the MRT
4. More effort needed to perform tasks
5. Additional stress/nervousness

5. CONCLUSION

The analyses of the comparison between usability testing done in two different settings revealed that there were many more types and occurrences of usability problems found in the field than in the laboratory. Those problems discovered tend to be critical issues regarding usability. Some of these usability problems are only related to the device being used in the field, which could not be found using conventional laboratory usability tests. With regards to the users' behaviors, users behave less positively and more negatively in the field than in the laboratory. Some behaviors can only be observed in the field. Users also take longer time to perform certain tasks and also reflected more negative feelings, such as dissatisfaction and difficult of use, to the use of the device in the field.

This study also investigated reasons to account for the differences summarized above. These included the external factors associated with the environment of use, such as noise, moving train environment, the privacy of using the device in a crowded place, the mental and physical resources and extent to which the user were affected by these factors. Although there is no unified ways to explain how these factors would affect everyone, they cannot be ignored, especially for the purposes of conducting accurate usability evaluation. This study helps to understand limitations of current evaluation method and contributes to the improvement of designing mobile devices suitable for the social context of use, considering the vast range of lifestyles of users. Future research will address how to develop a valid or modified usability evaluation for mobile devices.

6. ACKNOWLEDGMENTS

Our thanks to anonymous reviewers' comments and Jazmin Chong, for all her assistance throughout the design of the usability tests and pilot testing

7. REFERENCES

[1] Channelnewsasia report on Mobile Phone Usage in Singapore, 2004. Retrieved from: http://www.channelnewsasia.com/stories/singaporelocalnews/print/65398/1/.html

[2] Condos C., James A., Every P., and Simpson T. Ten Usability Principles for the Development of effective WAP and M-commerce services. *Aslib Proceedings: New Information Perspectives*, 54, 6, 2002, 345-355.

[3] Dumas J.S., and Redish J.C. *A Practical Guide to Usability Testing. Intellect*, Exeter. 1999.

[4] Gorlenko L., and Merrick R. No Wires Attached: Usability Challenges in the Connected Mobile World. *IBM Systems Journal*, 42, 4, 2003, 639-651.

[5] Kjeldskov J., Graham C. A Review of MobileHCI Research Methods. *Proceedings of the 5th International Mobile HCI 2003 conference*, 2003, 317-335.

[6] Kjeldskov J., Stage J. New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies*, 60, 2004, 599-620.

[7] Kjeldskov J., Graham C., Pedell S., Vetere F., Howard S., Balbo S. and Davies J. Evaluating the Usability of a Mobile Guide: the Influence of Location, Participants and Resources. *Behavior & information Technology*, 24, 2005, 51-65.

[8] Krippendorff K. *Content Analysis: An Introduction to its Methodology*. Thousand Oaks, CA. 2004.

[9] LTA (Land Transport Authority) Annual Report. 2004. Retrieved from: http://www.smrt.com.sg/annualreport2004/pdf/connectivity/rail-operation.pdf

[10] Molich R. (2000). *Usable Web Design*. Ingeni ren|b ger. (In Danish).

[11] Monk A., Fellas E. and Ley E. Hearing only one side of Normal and Mobile Phone Conversation. *Behavior & information Technology*, 23, 2004, 301-305.

[12] Nielsen J. *Usability Engineering*. Academic Press, San Diego. CA.1993.

[13] NTU's Student IT Background Survey. 2003. Retrieved from: http://www.ntu.edu.sg/CITS/Home/What+is+New/student+it+background+survey+2003.htm

[14] Nunes L. and Recarte M. A. Cognitive Demands of Hands-free Phone Conversation while Driving. From the 3rd International Conference on Traffic & Transport Psychology, *Transportation Research Part F: Traffic Psychology and Behavior*, 5, 2002, 133-144.

[15] Pascoe J., Ryan N., Morse D. Using While Moving: Human-Computer Interaction Issues in Fieldwork Environments. *Annual Conference Meeting Transactions on Computer-Human Interaction*, 7, 3, 2000, 417-437.

[16] Ramsay M. Mildly Irritating: a WAP usability study. *Aslib Proceedings: New Information Perspectives*, 53, 2001, 141-158.

[17] Rubin J. *Handbook of Usability Testing*. Wiley, New York. 1994.