

Usability Benchmark Study of Commercially Available Smart Phones: Cell Phone Type Platform, PDA Type Platform and PC Type Platform

Jeroen Keijzers¹, Elke den Ouden^{1,2} and Yuan Lu¹

¹Eindhoven University of Technology
Department of Industrial Design,
Sub department of Business Process Design
P.O. Box 513, 5600 MB Eindhoven
The Netherlands

²Philips Applied Technologies
Industry Consulting
High Tech Campus 5
5656 AE Eindhoven
The Netherlands

J.Keijzers@tue.nl; E.d.Ouden@tue.nl; Y.Lu@tue.nl

ABSTRACT

The upcoming of smart phones are the result of consumers' preference for high-feature products: manufacturers are lured into integration of a growing number of technologies and features to provide attractive and competitive models. At the same time usability of such complex products becomes an increasing problem. This study aims to provide designers more insight into the consequences of emerging technologies on the usability of smart phones with different platforms. We conducted a usability benchmark study and tested 4 different features on 3 smart phones with 3 different platforms (Nokia E6li (Symbian S60, cell phone type platform), HTC S710 (Windows Mobile 6, PC type platform), Palm Treo680 (Palm OS, PDA type platform)) with in total 43 subjects in a between-subjects design. The results show significant differences in usability of the smart phones for the selected functions. For all platforms, the common design guidelines for usability still apply, but are restricted by the desire to integrate more functionality to create competitive products.

Categories and Subject Descriptors

H5. [Information Interfaces and Presentation (e.g., HCI)].

H5.2 [User Interfaces]

General Terms

Design, Experimentation, Human Factors.

Keywords

Smart phone, Usability Evaluation, Design, User Interface.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MobileHCI 2008, September 2–5, 2008, Amsterdam, the Netherlands.
Copyright © 2008 ACM 978-1-59593-952-4/08/09...\$5.00.

1. INTRODUCTION

The market for mobile communication products (e.g. cell phones, Personal Digital Assistants (PDAs) and smart phones) is one of the most dynamic and competitive markets in the consumer electronics industry. The latest trend is the upcoming of smart phones which, in addition to traditional voice communication and messaging functionality, provide web access, personal information management, multimedia and business applications [34]. The smart phone is the future of today's cell phone. The market is booming: up 75% in 2006 compared to 2005. Forecasts show that in 2009 more than 200 million smart phones are expected to be sold worldwide [23]. While the smart phone used to appeal mainly to businesses, nowadays the smart phone costs have dropped significantly which make it easily affordable, even for the consumer market.

Technology advances at an exponential rate, making possible solutions and products not even dreamed of a decade ago [8]. Research has shown that consumers buy products based on the number of features provided [3, 31]. Fierce competition in the smart phone market has accelerated speed-to-market of new features and applications, because companies assume people will adopt new products that deliver more value or utility than the existing ones [7, 14]. As technology advances, it becomes more feasible to load products with a large number of features, each of which individually might be seen as useful [25, 31]. It seems very likely that the high rate of introduction of new features and applications has increased the speed of adoption of smart phones.

Research into "feature fatigue" (increasing the number of useful functions at the expense of usability) shows that even though consumers know that products with more features are harder to use, they initially choose high-feature models [31]. They even pile on more features when given the chance to customize a product for their needs. Once consumers have actually worked with a product, however, usability starts to matter more to them than capability [7, 14, 31]. The increasing number of consumer complaints that are not related to physical product failures shows that consumers do experience problems with usability of products with many features [10, 27]. A study of mobile device returns in the United Kingdom shows that 1 in 7 cell phones is returned

within the first year of purchase as faulty [27]. Of these returns, about 63% has no hardware or software fault but the reported problems relate to usability problems, mismatch with user's expectations or issues relating to the configuration of the handset.

Although usability is an increasing problem in smart phones, usability in itself is surely not a new topic. User-centered design and goal-directed interaction design have increased the usability of products by restricting the number of features to the goal of the product, and improving the User Interface (UI) by reducing the number of keystrokes to perform a task and better design of menus, keys and icons [7, 24, 32]. These design guidelines drive towards products with fewer features, and target at a smaller user group or market segment. However, this presents a dilemma to manufacturers of smart phones [28]: Should they maximize initial sales by designing high-feature models, which consumers consistently choose, or should they limit the number of features to enhance the lifetime value of their customers?

The features, performance, security and add-on applications in smart phones are determined by the Operating System (OS) [22]. Currently, the smart phone market is dominated by 6 OS's: Symbian, Blackberry OS, Windows Mobile, Palm OS, Linux, and since the launch of the iPhone: Mac OS X. Symbian is the market leader, installed on 100 million smart phones worldwide. In the United States, Blackberry and Windows Mobile are market leaders [22].

Combined with a set of related software development libraries, application programming interfaces and programming tools, an OS is also referred to as a mobile software platform [34]. These platforms differ from each other not only regarding their technical specifications [34]. They are either proprietary for special devices (e.g. Blackberry and Mac OS X) or open to all independent software providers (e.g. Symbian, Windows Mobile, Palm OS and Linux). Moreover, these platforms differ on historical basis. Palm OS and Blackberry OS originate from PDA products, Symbian from cell phone products and Windows Mobile and Linux originate from general purpose OS's.

The starting point of our study is that we assume that the complexity in smart phones is unavoidable since buyers of these products strongly prefer high-feature models. At the same time manufacturers aim to achieve higher levels of user satisfaction by improving usability. Because smart phones integrate PDA and cell phone features, conflicting design goals arise [26, 34]. PDA features require, for example, a different display size and data input method to achieve easy to use products, than the traditional cell phone functionality. Users might have a different mental representation of a smart phone when they think it to be either an enhanced cell phone or an enhanced PDA. The mental representation might hamper them in using the product. The goal of this study is to investigate how different mobile software platforms influence the usability of smart phones when taking into account different features. We set up an experiment to compare 3 commercially available smart phones with different mobile software platforms: Symbian S60 (cell phone type platform), Palm OS (PDA type platform) and Windows Mobile 6 (PC type platform). By analyzing the data both quantitatively and qualitatively, we aim to provide designers insight into the consequences of emerging technologies on the usability of smart phones with different mobile software platforms.

2. RELATED WORK

In this section we discuss literature dealing with the usability of mobile devices and summarize a selection of the most relevant findings. Usability in this paper is defined 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 [18]. Zhang and Adipat [32] show that there is a vast amount of literature available on UI design, usability and related topics for mobile devices, in particular for cell phones and, to a smaller extent, for PDAs. They discuss that mobile devices have unique features which pose a number of significant challenges for examining usability, such as mobile context, connectivity, small screen size and restrictive data entry methods.

In literature, both empirical testing and model-based evaluations are used for the usability testing of mobile devices [30]. Model-based evaluations such as GOMS (goals, operators, methods and selection rules), keystroke-level models or derivatives of these models are used to predict menu performance and support UI design (e.g. [1], [6] and [16]). Research shows that design choices in breadth versus depth [13], as well as sequential versus simultaneous menu presentation influence usability regarding menu navigation [15].

Empirical research by Ziefle and Bay [38] shows that the awareness of the hierarchical structure of the menu is of central importance to use a cell phone properly: the better the mental map of the menu, the better the performance when using the device. However, difficulties users experience with technical devices cannot be explained by defining cognitive complexity only by the number of production rules [37]. Ambiguous naming and allocations of functions as well as functioning and design of keys negatively influence usability [2, 32, 36, 37, 39]. Even though expertise helps in using complex cell phones, these aspects of the UI design have proven to weaken or even cancel the advantage of expertise. Interfaces that rely on contra-intuitive learning rules that defy learnability and "fool" the user each time s/he tries again, as well as interfaces that interfere with basic constraints of the information processing system perform less on usability [39]. Interface design could be improved by consistency with the analogues product's interface design concept that is familiar to the users [19, 32], and navigation aids can help prevent disorientation in cell phone menus [32, 35].

One specific feature of mobile applications which especially challenges UI design is mobile Internet. Mobile Internet browsers often break the underlying interaction style consistency [20]. Furthermore, because of the limited screen size and the increasing complexity of mobile tasks, significant usability problems arise [4, 5]. We expect that these design issues might be more severe for smart phones because features from different types of mobile devices are integrated into these products.

3. RESEARCH QUESTION

On the one hand, many different smart phone platforms are introduced in the market and consumers desire to know their usability performance before making purchasing decision. On the other hand, it is important for designers to review the usability performance of different platforms in order to develop better smart phones in the future. In this study we approach the problem by investigating the following research question:

- How do different mobile software platforms differ in their usability performance in the first consumer use phase, in terms of effectiveness, efficiency and satisfaction?

The research methods, research results and discussions are presented in the following sections.

4. METHODS

4.1 Overview

In order to answer the research questions, we designed a between-subjects experiment in which participants were asked to perform 4 tasks with a different level of complexity on 1 of the 3 smart phones selected for this study.

4.2 Experimental Variables

The independent variable under study is the mobile software platform of the smart phones selected for this study. For reasons of simplicity, we will use the term “OS” to refer to the smart phone mobile software platforms.

As dependent variables the standard usability measures effectiveness, efficiency and satisfaction were used [11]. Effectiveness was measured as the percentage of tasks solved. Efficiency was measured by 3 parameters: task completion time (in seconds), number of hierarchical levels in the menu used to complete the task and number of detour steps (number of returns

to a higher level in the menu). As qualitative measures the keys used, the steps taken in the menu and the user complaints were recorded. Satisfaction was measured per task by using the After-Scenario Questionnaire (ASQ) which addresses 3 components of user satisfaction with system usability: ease of task completion, time to complete a task and adequacy of support information [21].

4.3 Apparatus and Materials

Smart phone manufacturers offer their smart phones on the market with either their own OS or a commercially available OS. Consequently, it is not possible to compare different OS's without using different brands of smart phones. In this study, 3 commercially available smart phones were selected. They were Nokia E61i (Symbian S60, cell phone type platform), HTC S710 (Windows Mobile 6, PC type platform) and Palm Treo680 (Palm OS, PDA type platform). All 3 smart phones offered similar functionality and were configured for use by the same mobile network operator. An overview of the selected smart phones is shown in Table 1. This table includes a comparison of the most important aspects of the design of the UI, input method and key mapping, which in addition to the type of platform, might influence usability.

Table 1. Comparison of smart phones

Smart phone	Nokia E61i	HTC S710	Palm Treo 680
Operating system	Symbian S60	Windows Mobile 6	Palm OS
Physical appearance			
Main menu presentation	Desktop menu with shortcuts + application grid	Desktop menu with shortcuts + application grid	Desktop menu + application grid
Input method	Qwerty keyboard	Qwerty keyboard / keypad	Qwerty keyboard / touch screen
Softkeys	yes	yes	no
'Navigation' key mapping	4-way rocker key	4-way rocker key / arrows on qwerty keyboard	4-way rocker key / touch screen
'Select' key mapping	Center of rocker key and / or left soft key 'ok'	Center of rocker key and/or left soft key 'done' and/or 'enter' key	Center of rocker key or touch screen
'Cancel' key mapping	right soft key	right soft key or with 'del' key with left pointing arrow	'cancel' key on touch screen
'Global exit (to idle)' key mapping	red key or hard menu key (this key alternates between desktop and main menu)	'home' key or red key	'home' key (red key switches off device)

For readability reasons, we will refer to the brand names of the smart phones only, where in fact we refer to the combination of the OS and the selected branded product.

The tests were performed in a usability lab. Manuals were provided and the basic first use installation settings such as location and time were already set. Both the participant interacting with the smart phone and the UI of the smart phone were recorded on video cameras.

4.4 Experimental Tasks

Smart phones have numerous features ranging from the standard cell phone features to advanced business applications. Because the research objective here is to compare different smart phone platforms that originate from cell phones, PCs and PDAs, we set several criteria for the selection of features for testing: the features (1) should encompass typical cell phone as well as typical PDA features; (2) should have a different level of task complexity to cover multiple design aspects of the UI and; (3) should be practically suitable for testing in a laboratory environment.

Based upon these criteria, we selected the following tasks for the participants to solve during the tests: (1) Calling a number which is stored in the contacts list; (2) Retrieving information from a website; (3) Entering an appointment in the calendar and (4) Setting up a push mail account and using this account to send an email. We expected that the origin of the smart phones (cell phone, PDA or PC) would influence the related usability. The smart phone evolving from a cell phone type platform was expected to perform well on the Phone task, whereas the smart phone evolving from a PDA type platform was expected to perform well on the Calendar task. For the newer functions Internet and push mail, we are interested in discovering the differences in usability performance among these smart phones.

The complexity of each task per smart phone, in terms of the minimum number of menus and keystrokes needed to complete the task, is shown in Table 2.

To account for differences in speed of typing the participants were instructed to enter the appointment and the email with a minimum amount of text. Although the smart phones differ regarding the level of detail of information required for setting up the push mail account, the same information regarding the account settings was provided to all participants independent of the smart phone tested. The task order was varied to account for possible learning effects. Due to the restricted availability of the usability lab and to keep the total test duration per participant within 1 hour, a time limit was set for each task. This limit was based on durations measured during the pilot tests. The time limits for the tasks in consecutive order were 4, 7, 10 and 14 minutes.

Table 2. Minimum number of menus (keystrokes) needed to execute the task

Task	Nokia	HTC	Palm
Phone call	1(2)	1 (2)	1 (3)
Internet	2(6)	1 (7)	1 (5)
Calendar	3 (19)	4 (18)	3 (8)
Pushmail	7 (27)	7 (15)	5(16)

4.5 Participants

We decided to execute the experiment with students of a technical university, as they are generally familiar and confident with using related technologies and therefore resemble the profile of potential professional smart phone users. To ensure that differences in usability would not be attributable to differences of the personal characteristics among the participants, we used a survey to select participants with regard to familiarity with cell phones and computers, subjective expertise of cell phones and computers; and intention to use smart phones. Only students who did not own a smart phone or a PDA were selected for the experiment. Measurements of familiarity [9] and subjective expertise [12] with both cell phones and computers were used to ensure that there were no significant differences between the groups in terms of level of contact with related technology and confidence in using this technology.

In total 43 participants were randomly assigned to either one of the smart phone groups. In the Nokia group 10 males and 4 females with a mean age of 21.1 years participated. In the HTC group 11 males and 4 females with a mean age of 19.8 years participated and in the Palm group, 12 males and 2 females with a mean age of 21.0 years participated. Non-parametric Kruskal-Wallis tests showed no significant differences between the personal characteristics of the participants in the 3 smart phone groups.

4.6 Procedure

At the beginning of the experiment the participants were instructed that the experiment was set up to examine the ease of use of the smart phone and not to test the participants. The participants were provided with information on the capabilities of smart phones in general and were asked to read through a list of instructions. They were asked to perform the tasks in the predefined order and to fill-in the ASQ questionnaire after completing each task. Furthermore, they were asked to go back to the main menu of the smart phone before starting a new task to ensure that every task in every test was started from the same position in the menu. The participants were asked to think-out-loud while performing the tasks as well as during completion of the questionnaires. We did not notice any significant influence of this protocol on the time measures of the tasks (participants were instructed not to interact with the observer).

The participants were not instructed to solve a task within a certain time limit. However, after spending more time than the predefined time limits discussed above or when the participant gave up because the task was too difficult, the participant was asked to proceed with the next task. Except from the Calendar task, the tasks were set up as such that participants could verify whether they successfully completed the task. The whole experiment lasted 30 to 50 minutes depending on the working speed of the participant. After the experiment the participants received a token of appreciation for their participation.

5. RESULTS

In this section, we will analyze effectiveness, efficiency and satisfaction measures for each task separately. In addition, the users' actions and problems will be discussed in more detail for each task to give insight into the influence of smart phone platforms on usability.

The measures of efficiency and satisfaction did not meet the assumption of normality and equal variances between the different groups which are a prerequisite for being able to use parametric tests. In order to compare differences in efficiency and satisfaction measures per task across the different smart phones; one-way Kruskal-Wallis analyses were run. In order to compare differences between groups, separate pair-wise Mann-Whitney U tests were used. Holm's sequential Bonferroni correction [17] was used to determine the corrected significance level per comparison. The level of significance was set at $p = 5\%$. Results within the less restrictive 10% level are indicated as marginally significant.

5.1 Phone Task

A comparison of the performance of the smart phones for the Phone task in terms of effectiveness, efficiency and satisfaction is shown in Table 3. Since not all users were able to complete the task, the efficiency measures shown in this table are only reported for users who were able to complete the task. Both from a user and design perspective we expected that making a phone call would generate few problems in usability as this is a basic functionality of any particular phone. However, from the results shown in Table 3 it can be seen that even for this task differences occur in time and number of menus used. The Kruskal-Wallis tests show significant differences for time ($\chi^2(2) = 9.11, p < 0.05$), for number of menus used ($\chi^2(2) = 14.99, p < 0.01$) and for satisfaction ($\chi^2(2) = 6.16, p < 0.05$). As shown in Table 2, the complexity of the Phone task on the Palm evolving from PDA systems is slightly higher than that on the phones evolving from a cell phones or a PC (requiring 3 keystrokes against 2 for the other phones). The test results show that the effectiveness is lower and time required to complete the tasks is higher for the other two phones. Separate pair-wise Mann-Whitney U tests show that using the Palm to make a phone call takes approximately 75% more time than using the HTC (significant with $p=0.004$). Furthermore, users of the Palm used significantly more menus compared to users of the Nokia ($p=0.000$) and the HTC ($p=0.001$). It was also found that users were more satisfied when performing this task with the HTC than with the Palm ($p=0.016$). No significant paired comparisons for satisfaction were found between the use of Nokia and the other two smart phones.

Inspection of the usage patterns and user complaints revealed that the longer time and higher menu count for the Palm was mainly caused by the fact that users thought that pressing the call-button after selecting a contact person would result in the call. However, not the contact person but the telephone number had to be selected.

Table 3. Performance when making a Phone call

	Nokia	HTC	Palm
Effectiveness [%]	100	100	92.9
Time [s]	65.6	61.4	105.9
Number of menus used	2.3	3.2	10.8
Number of detour steps	0.3	0.7	1.1
Satisfaction	6.0	6.3	4.9

Another problem they faced was that they expected to return to a previous menu when selecting the red key, but instead the phone

was switched off. Consequently, the users spend more time and detour steps when browsing through the menus to find a solution.

5.2 Internet Task

The performance of the smart phones when users were asked to retrieve information from a website through the mobile Internet browser is shown in Table 4. Kruskal-Wallis tests show significant differences for all the measures: time ($\chi^2(2) = 13.81, p < 0.001$), number of menus used ($\chi^2(2) = 32.67, p < 0.001$), number of detour steps ($\chi^2(2) = 5.89, p = 0.05$) and satisfaction ($\chi^2(2) = 8.54, p < 0.05$). From the results shown in Table 4 it can clearly be seen that for this task both the Palm and the Nokia outperform the HTC on almost all the measures. For both smart phones almost all users were able to complete the task while for the HTC surprisingly low effectiveness percentages were observed.

Separate pair-wise Mann-Whitney U tests show that a high number of paired comparisons are significant. For the time measure, users of the Nokia took significantly less time to complete the task than users of the HTC ($p=0.001$). For the Palm, differences in task completion time are significantly different with the HTC ($p=0.008$). Roughly speaking, users of the HTC took twice as much time to complete the task than users of the Nokia and Palm. For the measures of the number of menus used, all the differences are significant except from the difference between the Nokia and the HTC. Finally, users take significantly less detour steps for the Palm compared to the Nokia and the HTC.

As shown in Table 2, the smart phones do not differ much in terms of task complexity for this specific task. Consequently, it can be concluded that the differences in usability are caused by other factors than task complexity. The Nokia caused problems for the users by an unclear status or lack of feedback (users got unclear error messages that did not give them sufficient information to solve the problems). Nevertheless this phone scored second in effectiveness and even first in time needed. We discovered that the guidance in the Nokia phones helps users to complete tasks successfully. The analysis of the usability problems with the HTC showed some remarkable results. As can be seen in Table 4, the number of menus used and the number of detour steps do not differ between the Nokia and the HTC while the users of the HTC experienced far more difficulties when trying to complete this task. We expected that the HTC has an advantage because the Internet function is called "Internet Explorer" and this is familiar to most users.

Table 4. Performance when using the Internet browser

	Nokia	HTC	Palm
Effectiveness [%]	92.9	33.3	100
Time [s]	195.9	400.9	239.9
Number of menus used	7.2	7.0	1.5
Number of detour steps	1.7	1.9	0.2
Satisfaction	5.3	4.3	5.2

The usage patterns and user complaints reveal that the HTC users were confused, because after entering the web address the phone displayed a search screen, while they expected to immediately get the requested website.

Finally, pair-wise comparisons for satisfaction show that compared to the HTC, the Nokia (significant with $p=0.003$) and the Palm (significant with $p=0.018$) score higher on satisfaction.

5.3 Calendar Task

A comparison of the performance of the smart phones for the Calendar task is shown in Table 5. Kruskal-Wallis tests show there are only significant differences for the number of menus used ($\chi^2(2) = 7.12$, $p < 0.05$). Separate pair-wise Mann-Whitney U tests show only one significant differences between the HTC and the Palm ($p=0.009$). For this task we expected the Palm to perform better because the calendar functionality originates from PDAs. Furthermore, the Palm is exceptional on task complexity as it only needs 8 keystrokes, where the other phones need 18 and 19. As Table 5 shows, the results of the experiment showed otherwise. Despite the advantage in complexity, the Palm scores low on effectiveness and high on time needed to complete the task. Most of the problems were related to setting the reminder: users had difficulty to allocate this function, and they did not understand that an “alarm” was to be set, they were looking for a function named “reminder”. The Nokia scores the best on effectiveness and it scores even better on time. Furthermore, the number of problems with the Nokia was equally low. This is remarkable as the Nokia is the phone that requires the highest number of keystrokes to complete the task. When analyzing this into more detail, we discovered that, although more keystrokes are required, they are quite logical in sequence as the user is guided through the fields to be filled in.

Table 5. Performance when using the Calendar

	Nokia	HTC	Palm
Effectiveness [%]	92.9	93.3	64.3
Time [s]	190.9	241.6	254.5
Number of menus used	9.1	13.8	6.8
Number of detour steps	1.6	3.2	1.0
Satisfaction	5.8	5.1	4.9

5.4 Push mail Task

The performance of the Nokia, the HTC and the Palm smart phones for a push mail task are shown in Table 6. When performing this task, users were asked to set up and use a push mail account. The Push mail task was the most complex of all 4 tasks. Even the simplest set-up, the Palm, requires 5 menus and no less than 16 keystrokes. Kruskal-Wallis tests show there are significant differences for number of menus used ($\chi^2(2) = 18.01$, $p < 0.001$) and number of detour steps ($\chi^2(2) = 9.73$, $p < 0.01$). The results show that for none of the phones a high effectiveness percentage was achieved. Furthermore, the Palm outperforms the other two phones on all the measures.

Inspection of the pair-wise Mann-Whitney U tests show that with the Palm a significant lower number of menus is used ($p=0.000$) and the difference regarding the number of detour steps with the Nokia is significant ($p=0.001$) and with the HTC marginally significant ($p=0.026$).

As predicted by the minimum number of menus and keys needed to complete the task, users of the Palm needed much less menus and steps to complete the task than the other phones. Although the Nokia provides an installation wizard for the set-up of the account, nevertheless it is asking detailed input information that is not clear to the users (e.g. asking to select POP3 or IMAP4). The users of the other phones were able to complete the set up without this specific information. Additionally, users of the HTC were struggling with the automatic dictionary which was set up for English, while they were typing their email in another language. They could not find a way to switch off the dictionary function. Overall, it seems that for all the phones users needed more confirmation on whether they were doing the right thing when executing this complex task. The feedback the phones provide is not clear enough (e.g. users did not know if the email was sent).

Table 6. Performance when setting up a Push mail account

	Nokia	HTC	Palm
Effectiveness [%]	50	60	71.4
Time [s]	689.6	705.6	490.2
Number of menus used	26.1	18.3	6.9
Number of detour steps	6.0	4.3	0.9
Satisfaction	3.9	4.0	4.3

6. CONCLUSION AND DISCUSSION

This study examined the usability of 3 different smart phones with different OS's. In order to find out if users experience differences in ease of use of important functions, an experiment was set up to test the usability of 4 different functions with a homogeneous group of users. The results showed significant differences in usability of the smart phones for the selected functions. Especially the Internet and Push mail task showed significant differences for most of the measurements and for both tasks significant usability problems were identified. As is shown in Table 2, the selected functions have an overall increasing level of complexity based on the minimum number of menus and key strokes required to complete the tasks. Except from the Internet task, the results indicated that a higher level of complexity results, on average, in lower effectiveness and efficiency scores. These overall results confirm earlier research on the influence of task complexity on performance (e.g. [5] and [39]) and earlier research on usability of mobile Internet [4, 5, 20], discussed in section 2.

However, the differences in performance are not consistent across the selected OS's for the selected tasks. Furthermore, except for the Push mail task, task complexity differences between the phones when compared for each specific task, do not explain the performance differences. We expected that the origin of the smart phones (cell phone, PDA or PC) would influence the design decisions and related usability. The Nokia OS evolving from a cell phone was expected to perform well on the Phone task, whereas the Palm OS evolving from a PDA was expected to perform well on the Calendar task. For the Phone task the results are consistent with our expectations, i.e. the problems with the Palm are caused by both the platform (the phone call functionality is not designed as a primary function) and ambiguity of key

design. For the other tasks, the results are not consistent with our expectations. Although for the Calendar task the Palm is less complex, it is outperformed by the Nokia and the HTC, mainly due to the task guidance provided.

When taking a bird's eye view on the usability problems users encountered during the experiment it becomes clear that several of those problems could have been avoided. For example, the functionality of the red key on the Palm. Users expected to return to a previous menu when pressing this key, but instead the phone was switched off. In other words, this design is not consistent with the commonly used interface in cell phones.

Some problems, however, are inherent to the required number of features that the smart phone needs to include, such as having to navigate to allocate a specific function. Given the complexity, we observed that certain design guidelines are more important to improve usability. For example, a high number of menus and keystrokes required to complete a task does not automatically correlate with efficiency. The results suggest that clear task guidance and navigation aids are a more important determinant for ease of use. Consequently, the commonly advised design guidelines for mobile applications (such as the guidelines discussed in [32]) still apply but can only be implemented to a certain extent. Further research is needed to investigate how the design guidelines should be adapted to apply to high-feature mobile applications.

7. Limitations and further research

In this study we used carefully selected students of a technical university as test participants. It is expected that a more diverse group of participants in terms of knowledge, age and education will generate even more diverse usability problems. Furthermore, we have tested only the first use phase; it is expected that some learning will take place during extended use [39]. During extended use, users might want to adapt the menus to their personal needs, e.g. by creating a personalized "start menu" which show the frequently used applications and functions. We have not tested the ease-of-use in the extended use phase, so we cannot comment on the usability performance of these smart phone platforms in the longer term.

We have only tested products that are currently available on the market. It is expected that soon new technologies will become available for integration in next generation smart phones. An example of a function that becomes possible with broadband wireless communication is the use of off-board applications that push information upon request to the user's phone. Off-board GPS-navigation will be one of the interesting applications as it can automatically include up-to-date information on road works, traffic information, etc. Another example is ambient intelligent technologies providing applications and information dependent on the location of the user [29]. These technologies will challenge designers of smart phone OS's and applications to run on those systems much more than before, as the amount of features and options for users will increase even further, and at the same time consistency is at stake as many applications will be designed by different parties. Further research into usability of complex products such as smart phones is needed to ensure good usability and satisfied users.

8. ACKNOWLEDGMENTS

This work has been carried out as part of the TRADER project under the responsibility of the Embedded Systems Institute. This project is partially supported by the Netherlands Ministry of Economic Affairs under the BSIK program. We would like to thank Jakajima who provided the smart phones, SIM cards and Bluetooth headsets. Furthermore, we would like to thank the students Jeroen Cox, Yvonne Kleuskens, Lonke Bakx, Jeroen Bovend'eerdt, Geoffrey van IJzendoorn and Dianne Soons for their help in setting up and executing the test and analyzing the data.

9. REFERENCES

- [1] Amant, R. St., Horton, T.E. and Ritter, F.E. Model-based evaluation of cell phone menu interaction. In Proc. CHI 2004, ACM Press (2004), 343-350.
- [2] Bay, S. and Ziefle, M. Children using cellular phones: The effects of shortcomings in user interface design. *Human Factors* 47, 1 (2005), 158-168.
- [3] Brown, C.L. and Carpenter, G.S. Why is trivial important? A reasons-based account for the effects of trivial attributes on choice. *Journal of Consumer Research* 26, (2000), 372-385.
- [4] Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G. and Pazzani, M. Improving mobile internet usability. In Proc. 10th Int. Conf. on World Wide Web, ACM Press (2001), 673-680.
- [5] Chae, M. and Kim, J. Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour & Information Technology* 23, 3 (2004), pp 165-181.
- [6] Cockburn, A., Gutwin, C. and Greenberg, S. A predictive model of menu performance. In Proc. CHI 2007, ACM Press (2007), 627-636.
- [7] Cooper, A. *The Inmates are Running the Asylum*. Macmillan Computer Publishing, Indianapolis, USA, 1999.
- [8] Cooper, Robert G. *Winning at new products: accelerating the process from idea to launch*. 3rd edition. Perseus Publishing, New York, 2001.
- [9] Cordell, V.V. Consumer knowledge measures as predictors in product evaluation. *Psychology & Marketing* 14, 3 (1997), 241-260.
- [10] Den Ouden, E., Lu, Y., Sonnemans, P.J.M. and Brombacher, A.C. Quality and reliability problems from a consumer's perspective: an increasing problem overlooked by businesses? *Quality and Reliability Engineering International* 22, 7 (2006), 821-838.
- [11] Dumas, J.S. and Redish, J.C. *A Practical Guide to Usability Testing*, Intellect Books, Portland, USA, 1999.
- [12] Flynn, L.R. and Goldsmith, R.E. A short, reliable measure of subjective knowledge. *Journal of Business Research* 46, (1999) 57-66.
- [13] Goldstein, M., Anneroth, M. and Book, R. Usability evaluation of a high-fidelity smart phone prototype: Task

- navigation depth affects effectiveness. In Proc. of the HCI Int. 1999, ACM Press (1999), 38-42.
- [14] Gourville, J.T. Eager sellers and stony buyers – Understanding the psychology of new-product adoption, *Harvard Business Review* 84, (2006), 98-106.
- [15] Hochheiser, H. and Shneiderman, B. Performance benefits of simultaneous over sequential menus as task complexity increases. *International Journal of Human-Computer Interaction* 12, 2 (2000), 173-192.
- [16] Holleis, P., Friederike, O., Hußmann, H. and Smidt, A. Keystroke-level model for advanced mobile phone interaction. In Proc. CHI 2007, ACM Press (2007), 1505-1514.
- [17] Howell, D.C. *Statistical Methods for Psychology*, Fifth Edition, Thompson Learning, Pacific Grove, 2002.
- [18] ISO 9241-11:1998 Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability, 1998.
- [19] Jeong, S.-H. and Lee, K.-P. The effects of experience with a PC on the usability of a mobile product, In Proc. 6th Asian Design International Conference, 2003.
- [20] Kiljander, H. Evolution and Usability of Mobile Phone Interaction Styles. Thesis for Doctor of Science in Technology, Helsinki University of Technology, December 2004.
- [21] Lewis, J.R. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. *SIGCHI Bulletin* 23, 1 (1991), 78-81.
- [22] Malykhina, E.. Analysis: How smartphone platforms compare. *InformationWeek*, 20 January 2007. <http://www.informationweek.com/story/showArticle.jhtml?articleID=196902226> (accessed August 2007).
- [23] Malykhina, E.. Smartphone buyer's guide. *InformationWeek*, 30 October 2006. <http://www.informationweek.com/story/showArticle.jhtml?articleID=193402812> (accessed August 2007).
- [24] Mohageg, M.F. and Wagner, A. Design Considerations for Information Appliances. In *Information Appliances and Beyond*. Interaction Design for Consumer Products, ed. Bergman, E. Academic Press, San Diego, 2000, 27-51.
- [25] Norman, D.A. *The Invisible Computer*. The MIT Press, Massachusetts, 1998.
- [26] Nyberg, M., Björk, S., Goldstein, M. and Redström, J. Handheld applications design: Merging information appliances without affecting usability. In Proc. IFIP TC.13 Conference on Human Computer Interaction. IOP Press (2001), 391-398.
- [27] Overton, D. 'No Fault Found' returns cost the mobile industry \$4.5 billion per year. July 2006. <http://www.wdsglobal.com/news/whitepapers/20060717/20060717.asp> (accessed August 2007).
- [28] Rust, R.T., Thompson, D.V., Hamilton, R.W. Defeating Feature Fatigue, *Harvard Business Review* 84, 2 (2006), 98-107.
- [29] Svendsen, B. Mobile computing. In *True Visions, The emergence of ambient intelligence*, Aarts, E. and Encarnação, J.L. (Eds.), Berlin, Springer-Verlag, 2006.
- [30] Tarasewich, P., Wireless devices for mobile commerce: User interface design and usability. In *Mobile Commerce: Technology, Theory and Applications*, ed. Mennecke, B.E and Strader, T.J. Idea Group Publishing, Hershey, 2002, 26-50.
- [31] Thompson, D.V., Hamilton, R.W. and Rust, R.T. Feature fatigue: When product capabilities become too much of a good thing. *Journal of Marketing Research* 42, 4 (2005), 431-442.
- [32] Weiss, S., *Handheld Usability*, John Willey & Sons, LTD, Chichester, 2002.
- [33] Zhang, D. and Adipat, B. Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction* 18, 3 (2005), 293-308.
- [34] Zheng, P. and Ni, L.M. *Smart Phone & Next Generation Mobile Computing*. Morgan Kaufmann Publishers, San Francisco, 2006.
- [35] Ziefle, M. and Bay, S. How to overcome disorientation in mobile phone menus: A comparison of two different types of navigation aids. *Human-Computer Interaction* 21, (2006), 393-433.
- [36] Ziefle, M., Bay, S. and Schwade, A. On keys' meanings and modes: The impact of different key solutions on children's efficiency using a mobile phone. *Behaviour & Information Technology* 25, 5 (2006), 413-431.
- [37] Ziefle, M. and Bay, S. How older adults meet complexity: aging effects on the usability of different mobile phones. *Behaviour & Information Technology* 24, 5 (2005), 375-389.
- [38] Ziefle, M., and Bay, S. Mental models of a cellular phone menu. Comparing older and younger novice users. In Proc. MobileCHI 2004, Springer-Verlag (2004), 25-37.
- [39] Ziefle, M. The influence of user expertise and phone complexity on performance, ease of use and learnability of different mobile phones. *Behaviour & Information Technology* 21, 5 (2002), 303-311