Evaluating the Siemens C10 Mobile Phone:  
- Going beyond „quick and dirty“ Usability Testing

Martin Böcker and Anne Suwita  
Siemens ICP CD TI 3, Hofmannstr. 51, 81359 Munich, F.R. Germany.  
martin.boecker@mch.siemens.de, anne.suwita@mch.siemens.de

Abstract
In spite of the general appreciation of the need for usability evaluation, systematic evaluation studies of new products are the exception. This paper describes the evaluation conducted at Siemens of the prototype for a novel interface of a cellular telephone. The interface design itself is presented followed by a description of the test method, the 80 subjects sample, and the results. The conclusions focus on improvements in the design based on the study and on the need of integrating usability evaluation in the product cycle.

1 Introduction
It is a commonplace demand that products and in particular those with a complex user interface should be empirically evaluated at an early stage (many text books on user interface design dedicate a considerable percentage of the content to the issue of usability testing, e.g. [1]). Still, in spite of the plethora of things that can be learnt from evaluating products, systematic empirical evaluation studies are conducted only rarely. The current practice of product evaluation is one of not evaluating at all or „quick and dirty“ tests („three of my colleagues and the cleaning woman had a look at it“).

Product development processes rarely accommodate (a) the time for the evaluation study itself and, more importantly, (b) the time for making changes based on the experience from the study (cost does not seem to be the main constraint for most companies). Furthermore, user interface designers may dislike learning that some of their design decisions may be wrong.

The trade-off is one between investing resources early, and consequently delaying the completion of the technical specification, and completing technical specifications faster and learning only after market introduction of the product what the (user interface) bugs of the product are.

While decisions on technical details can be made such that the impact on performance can be tested in the lab with standardised equipment and often minimum resources, decisions on the design of a user interface for a product have to be based, at least to a limited degree, on good practice and experiences gathered from earlier products. Human Factors is no exact science leading to reliable predictions on the usability of a given user interface solution. From this follows, that user interfaces have to be tested, not applying criteria of technical performance but of usability for the target user population. User interface designers evaluate their solutions not because „they don’t know what they are doing“ but because they have to make sure that the solutions, and in particular innovative ones, meet the usability goals defined for the product.

At an early stage, a quick, low-cost approach („discount usability methods“) to evaluating particular aspects of an interface may often be enough and, of course, is better than not testing at all. Usually, many of the more fundamental bugs become
obvious after watching five or six subjects having trouble. However, a more substantial evaluation study does not only indicate whether a product hits a pre-defined usability criterion and where the most critical interface bugs are. It also allows the evaluator to measure additional usability-related dimensions ranging from the ease of use to the device’s appeal and the joy of using it. Aspects influencing usability such as previous experience, attitude towards technology („techno-freaks“ and „techno-phobes“) can also be captured. Even though all this is well known, evaluation studies are rarely performed in industry. This paper tries to illustrate the importance and benefits of in-depth product evaluation by describing a study conducted at Siemens of a novel user interface concept of a cellular telephone, the Siemens C10. In total, Siemens invested six weeks within the design process. The results of the study were taken as the basis for the decision to go ahead with the further development and production of the C10 mobile phone.

2 The design rationale of the C10 mobile phone

It is no longer technically difficult to pack a telephone with every feature a user could possibly imagine or want. Nowadays, usability is the main constraint on which features to add to a telephone. The small screen and the limited room for input devices, normally only buttons, present unique challenges to the designer. The design of the C10 for Siemens focused on one design aspect of the interaction with mobile telephones, namely the menu structure.

Arguably the central design element of every mobile telephone is the menu structure. All mobile telephones that we are aware of are organized around a hierarchical menu structure as opposed to other interactive techniques such as a command line interface (e.g. the user could type in „set incognito = off.“). This is probably due to the limited input and output ability of the mobile telephone and to the enormous success of the hierarchical menu structure as used in WIMP interfaces such as Windows in recent years.

The user navigates this menu structure in order to access the settings and features of the telephone. A typical mobile telephone has on the order of one hundred features and settings plus values for the settings. Given the C10’s functionality, its complete menu structure is rather daunting to navigate with the tiny screen provided by a mobile telephone. The values for the settings can number over one hundred. An example setting is the tone to use when a call arrives and the values for this setting are the different electronic tones that have been loaded into the telephone.

The hierarchical menu structure used in mobile telephones is not too different from the menu structure used by software in Windows pull-down menus. Many papers have been written on how to organize the categories of a menu structure and how to order and name the entries in a menu structure [2]. There has also been recent research on how users search a menu structure [3]. But, our focus here is not on the organization of the menu hierarchy nor on exactly how users search a menu structure, but rather on the limitations that the small screen of a mobile telephone impose on the presentation of the menu hierarchy.

By cascading a number of menus and submenus, a Windows menu indicates (a) the path

---

1 The study was a joint effort by G. Grace who designed the C10 user interface, M. Böcker who co-ordinated the study for Siemens and A. Suwita who conducted the study at the Human Factors Dept. of the Heinrich-Hertz Institut in Berlin (Ms. Suwita has joined and Mr. Grace has left Siemens since).
taken so far through the menu hierarchy and (b) the options (menu items of further submenus) available (Fig. 1a). The menus typically employed in mobile phones (Fig. 2a) usually only provide the latter – i.e. in the menu tree of Fig 1a, only the currently available items 2.2.1, 2.2.2, etc. are being displayed and no information is given as to which branch of the menu the user is currently in.

The design of the C10 menu remedies this deficit by two means. The top level menu is represented by an icon line. Upon accessing the menu (by means of a ‘Menu’-Softkey in the phone’s idle state), the user can navigate with the two left softkeys through the seven symbols (i.e. top level menu entries). A text label in the second and third display line indicates the symbol’s meaning. When one of the symbols is selected, all other symbols disappear, leaving an indication about the current top level branch.

At the next level down, one item is being displayed at a time together with an indication on the position of the list and the length of the list (e.g. in Fig. 2c, Item 2.2.2 is the second of four menu items in the current menu branch, see also Fig. 1b). The total menu was designed such that the user would rarely be in the menu hierarchy more than two levels with the C10. The C10, while not providing as much information about the menu path as a cascading Windows menu, still makes most of the precious screen real estate offering more orientation than a conventional menu of a mobile phone. The question to be addressed empirically was whether this new approach increases usability and the users’ acceptance of the phone.

3 Procedure
In order to examine the C10's usability, its user interface was compared with an identical interface without icons. The usability test addressed the following questions: Does the icon line interface reach an a-priori usability criterion of objective performance? Does the icon line interface compare favourably with its text-only sibling? What is the appeal of the new interface? In addition, the study was supposed to help identifying possible usability bugs.

Moreover, it was interesting to what extent the user manual is helpful or even necessary. Therefore half of the subjects were asked to read the user manual before starting the experimental tasks whereas the other half had the possibility to look up the manual during the tasks. The experimental design thus comprised two between-subjects factors (Icon vs. Text and Manual mandatory vs. Manual optional) with 20 subjects in each of the four cells.

As the mobile phone did not exist yet at the time of testing, the study was carried out by means of a computer simulation, i.e. computer prototypes of the two navigation methods were built. The only difference between them was that one used icons on the icon line at the top level for the menu structure and for status information (battery charge and signal strength) whereas the other used icons only for status information. The test subjects made button selections by using the mouse to click on an image of a button. Testing the designs at an early stage with computer prototypes seemed acceptable since several recent studies show that there is little or no advantage to building physical models for the evaluation of telephones [4].

Altogether 80 subjects took part in the experiment with 40 subjects testing each prototype. The large number of subjects was necessary to allow for inferential statistical analyses testing for differences between the two designs and to match certain requirements to the sample: As it could not be excluded that previous experience with mobile phones as well as the nationality of the test subjects may influence their performance or their rating of the mobile phones, German and foreign test subjects as well as subjects with and without mobile phone experience were assigned to the two prototypes as far as possible according to their distribution in the entire sample.

As experimental tasks the subjects had to test particular functions of the mobile phone within the framework of pre-defined scenarios. In total, the test subjects had to solve the following seven experimental tasks:

1. To make a simple phone call.
2. To select a network.
3. To call a person out of the „phone book“ of the mobile phone.
4. To look up a phone number in the „phone book“ of the mobile phone during a phone call.
5. To call a number from the call list.
6. To input a quick dial number.
7. To dial using a quick dial number.

Objective observation data were collected by the experimenters (by means of observation sheets), as well as subjective ratings by the test subjects themselves (questionnaire, interview).

Types of observation data were whether the tasks were solved, how many trials and how much time test subjects spent, errors (e.g. how often a wrong menu or key was chosen) and how often the test subjects looked up an information in the user manual.

Subjective ratings were collected by questionnaires and an interview. After each task the
subjects had to rate on 5-point agreement scales three items concerning the usability of the particular function they just had tested:

„This task could be performed well using this mobile phone“ (from „agree“ to „disagree“).
„How often was the next step not clear in this task?” (from „very often” to „very seldom”).
„For solving this task, too many steps were necessary.” (from „agree” to „disagree”).

After the block of tasks, the subjects filled in a questionnaire concerning general aspects of the mobile phone such as display and keys as well as the general usability, acceptance, satisfaction, appeal, purchasing intention and purchasing criteria. Moreover, the questionnaire included control variables such as experience with Windows programs respectively with telephone comfort functions as well as personal data.

In order to test whether the test subjects associate the right functions to the icons, they filled in an icon evaluation questionnaire, which was based on the MIA (Multiple Index Approach) developed by ETSI [5, 6]. In the MIA questionnaire, the task consists of assigning the correct icons to seven functions (quick dial list, phone book, call list etc.). After their choice test subjects indicate how sure they are about their choice (5-point scale from „very sure” to „very unsure”) and how they rate usefulness of the particular icon in terms of representing the function (5-point scale from „very good“ to „very poor”).

Finally, subjects were interviewed about particular problems they may have encountered, and about any suggestions they may have had for improvement with regard to the usability or design of the mobile phone. The subjective impressions of the experimenters were summarised in a group discussion with all experimenters.

In both versions of the user interface design, each half of the test subjects were recorded by video with split screen presentation: both the face of the test subject with all mimic reactions as well as the mobile phone and the movement of the cursor were seen.

In each experimental session one test subject and two experimenters took part. The first experimenter led the test subject through the experiment collecting the observation data mentioned above. The second experimenter recorded the test subject’s questions and comments and clocked the time. An experimental session lasted about 1½ hours.

4 Results

As stated in Section 3, the questions addressed by the study were (a) whether the C10 icon interface reaches an acceptable performance level (tasks performed and time taken) both in absolute terms and in relative terms compared to the text-based interface, (b) how the icon interface is rated in terms of "joy-of-use" parameters, and (c) which possible usability bugs are hidden in the design? The following presentation of the results focuses on statistically significant results.

The effects of the subjects’ nationality and time of icon test (before or after the main test) are not reported here because there were no systematic significant effects of these variables.

Objective performance-related data: Table 1 shows the percentages of subjects who successfully solved the tasks. The data for the seven tasks are broken down by the version of the phone used (icon versus text) and by handy-ownership. On average, an a-priori criterion of usability (50% of the non-owner subjects) was reached for all seven tasks. For owners of mobile phones, the pre-set criterion of 80% was not reached in
tasks 2, 4, 5, and 6, though only in Task 2 the mean failed to reach 70%. Most subjects needed between one and two trials to solve the task, the average time needed to solve the tasks was between one minute (Task 1 and Task 7) and three and a half minutes (Tasks 2 and 6).

<table>
<thead>
<tr>
<th>Task</th>
<th>Owners of mobile phones</th>
<th>Non-owners of mobile phones</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) to make a simple phone call</td>
<td>96%</td>
<td>88%</td>
</tr>
<tr>
<td>2) to select a network</td>
<td>67%</td>
<td>67%</td>
</tr>
<tr>
<td>3) to call a person from the &quot;phone-book&quot;</td>
<td>92%</td>
<td>96%</td>
</tr>
<tr>
<td>4) to consult the &quot;phone-book&quot;</td>
<td>82%</td>
<td>64%</td>
</tr>
<tr>
<td>5) to call a phone number out of the call list</td>
<td>67%</td>
<td>80%</td>
</tr>
<tr>
<td>6) to input a quick dial number</td>
<td>82%</td>
<td>75%</td>
</tr>
<tr>
<td>7) to dial a quick dial number</td>
<td>89%</td>
<td>84%</td>
</tr>
</tbody>
</table>

Table 1 Percentage of subjects who successfully solved the tasks

In terms of differences between the performance of the subjects in the two experimental groups, no clear trend emerged. Icon interface users performed better in Tasks 4 and 6, in Task 1 only for mobile-owners. In all other cases, the text interface led to better results. An overall analysis of variance with User interface, Handy-ownership, and User instructions as main predictor variables and Task achieved as criterion showed no significant main effects of User interface, nor of User instructions (for the effects of mobile-ownership, see below).

A fairly similar picture emerged from the analysis of the time required for performing the task. Again, the data seem to favour the text version with the exception of Tasks 4 and 5. To summarise the results presented so far: as far as objective data of task achievement and time required are concerned, no advantage of using the icon-line interface could be identified. If anything, it tended to receive less favourable results.

A post-hoc analysis looked at the effects of the subjects' familiarity with Windows-type computer interfaces. Windows-experienced subjects (i.e. those who rated their experience with Windows as "a lot", "some" or "middle") needed less time and fewer attempts to solve the tasks, they consulted the manual less frequently, and they indicated less frequently that the next step was unclear than for Windows-inexperienced subjects ("low" or "very low" experience). Generally, knowing Windows gave a head-start to the subjects under the icon condition. E.g. icon interface users tended more to agree with the statement that the task could be "handled easily with this mobile phone". In turn, windows-inexperienced subjects under the text interface condition agreed with the statement that "too many steps are necessary for solving this task".

A second post-hoc analysis tested the influence of the experience with using mobiles. Icons proved to be more advantageous for non-owners than for owners of mobile phones. Subjects who were experienced with mobile phones seemed to be rather confused by the icons when referring to particular functions (select network, quick dial list), possibly because their own mobile phone works differently.

In Task 2 (to select a network), non-owners using the text version agreed more frequently with the statement that "for solving this task too many steps are necessary"
than the users of the icon version. Such differences in ratings between the two versions of the mobile phone did not occur in the subgroup of mobile phone owners.

During Task 6, owners of mobile phones consulted the user manual more often than non-owners. One possible explanation of this result may be that owners of mobile phones draw on the experience with their own mobile phones which may have a different procedure for inputting quick dial numbers (i.e. a case of negative transfer). In addition, cases of owners of mobile phones choosing the wrong menu in Task 3 occurred more frequently within the icon version sub-sample than within the text version sub-sample. In the case of the non-owners, it was exactly reversed: these subjects chose less often the wrong menu within the icon version than within the text version. In Task 7, within the icon version, owners of mobile phones chose more often wrong keys than non-owners, however, within the text version experience with mobile phones did not have this effect. These results indicate that while suiting novice users, the icon interface may initially produce instances of negative transfer for experienced mobile phone users.

**Analysis of subjective ratings:** A second batch of analyses looked for main effects of the key predictor variables on the subjective ratings. A number of significant differences between the icon line version and the text version were identified (see Figure 3): subjects rated the icon version significantly more frequently as "user-friendly" and agreed more often, that "using the mobile phone was fun". With the text version subjects tended to think that "user steps for the comfort functions are confusing". This also had an effect on the acceptance values: subjects would rather use the icon version than the text version "for private use".

![Subjective ratings by interface type (icon and text)](image)

Concerning the effects of experience with Windows, a number of questionnaire items led to significant differences between those who were experienced Windows and those who do not. Windows users agreed that the mobiles were easy to use, easy to understand, familiar, using them hardly requires much thinking, and they would like to use such a phone. They disagreed more frequently than the Windows-novices that the mobiles were too awkward to use, difficult to explain to others and that a detailed user manual is necessary in order to be able to use them. On a seven-point Kunin scale ("smileys") they were significantly more pleased with the phones (5.3) than the Windows-novices (4.5). Generally, for many variables the analysis of Windows experience and phone type indicated that the icons compensate to a certain degree for
not knowing Windows.
The effects of owning or not owning a mobile on the subjects' subjective ratings indicate that the two designs appeal to the mobile-experienced subject. Mobile owners rated the prototypes more frequently as user-friendly and fun to use, and they indicated more frequently that they would like them for private use than non-owners. However, there were no significant differences between the two interface types.
Subjects for whom the reading of the manual was mandatory rated the mobile phone as being more “easy to understand”, as being less ”awkward” and they wished less often "more feedback” as compared to those subjects for whom the reading of the user manual was optional.

Analysis Purchasing intention: For the question "Would you buy this particular mobile phone?", the icon interface mobile received more positive answers (23% of the sample answered with "yes" whereas only 8% of the text interface users gave that answer).
In order to find out what the subjects' relevant purchasing criteria for a mobile phone are, 12 product characteristics were rated in terms to their importance for purchasing a mobile phone (quality, ease of use, stand-by time, talking time, robustness of case, size, weight, price, sound quality, design, size of display, brand). Quality was rated the most important criterion followed by ease of use and the telephone's stand-by time. At the other end of the scale, the producer's brand was considered of little relevance for choosing a product.
To analyse to what extent the purchasing intentions of the icon-line interface users can be predicted on the basis of the purchasing criteria, a multiple regression analysis was carried out for the group of subjects who tested the icon version, using the ratings of the purchasing criteria as a predictor for the dependent variable "purchasing intention”. For this analysis, the answer alternatives of the dependent variable "purchasing intention” were dichotomised since the original answer alternatives cannot be considered equidistant: The two categories "purchasing intention in principle” (answer "yes” or "possibly") and "no purchasing intention” (answer "no") were formed. The results of the regression analysis yielded two statistically significant purchasing criteria, namely "ease of use” as a positively related predictor and the brand as a negatively related one. Those who would buy a mobile phone with icon-line interface think that ease of use is important whereas the brand is not.

Analysis Icon Test: MIA The icons used in the icon-line interface were tested with all subjects using the ETSI's symbol testing method, the MIA [5, 6]. Generally, the results were fairly positive with 80% of the subjects and more making the right associations. Two symbols, however, those for Last Call List and Quick Dial received low results (34% and 37%). Handy-owners tended to be more successful in associating the symbols to the corresponding function, this was presumably the case because mobile owners have a clearer understanding of those functions. In all cases, the subjective ratings (the subjective certainty of the symbol association and the rating of how well the symbol represents the function) were very positive.

Usability bugs: Not surprisingly, more than twenty usability problems could be found and there were also a couple of new design elements that were tested that proved not to cause any problems. Below is a list of some of the finding is no particular order.
The on-hook (“end”) key was also assigned the function to return from a menu-level. Despite instructions about how the key operated and despite using the key to return from
several menus, some people could just not learn to use the on-hook key to return from a menu level. It may be that the on-hook key has such a serious meaning of ending a call that the subjects had a very difficult time associating another less critical function to this key.

A design change that proved to cause no usability problems was that the telephone book was turned into a flat list with the names and the commands together in one list. The commands became the first items in the list and were surrounded by „<“ and „>“. We feared that the users would not understand the meaning of the brackets and might not scroll passed the commands to find the names in the telephone book, but all of these fears proved to be unfounded.

Test subjects complained that after changing the network provider they should have been brought back directly to idle and should not have to press the return key several times.

Some subjects could not recognize the idle display after changing the network. The network name is prominent in the idle display and once changed the idle display was not recognized.

The icon version of the mobile telephone scrolled the menu lists horizontally and the non-icon version of the mobile telephone scrolled the lists vertically. The only difference is in the direction of the arrows that are presented in the soft-key labels. The subjects were not confused by either version and it appeared that the only important aspect to the user was pressing a button moved them to the next item.

Conclusions

The above results can be summed up by saying that in terms of objective performance data (tasks solved, time needed), the icon-line interface did not turn out to be more successful – for some tasks, the results were actually unfavourable. As far as subjective data of "joy of use" and the general appeal is concerned, a quite different picture emerged: the icon interface was seen as user friendly and more fun to use than its more traditional text-only sibling, and consequently received better ratings in terms of purchasing intentions: not only did significantly more subjects indicate that they would buy the icon phone, but also those who indicated that they would like to buy it did so because to them, the ease of use of a phone matters whereas the brand does not.

The overall line of these results, i.e. no improvement in performance but strong marks for appeal, remind of the introduction of the first graphical user interfaces: the first computer user interfaces using the desktop metaphor led to a speed of learning, a speed of task completion, a level of comprehension, and error rates that were indistinguishable from more traditional approaches; they were, however, preferred by the users because they were more interesting and fun to use [7].

Siemens invested a considerable amount of manpower, time, and money into the usability test of the C10. Given the results we obtained, this effort was well justified: some, maybe even the most crucial usability bugs may also have surfaced using a "discount usability method" [1] with a handful of subjects and a less systematic approach. It is unlikely that the whole range of problems would have been identified. In addition, the larger number of subjects allowed those statistical analyses that showed what the strengths of the new design are, namely not a phone that is quicker to use but one with the appeal of being easy to use and fun to use. The full-scale evaluation study provided a sound basis for the decision to go ahead with the development and introduction of the C10 as a product that appeals to the first time buyer who looks for an
inexpensive mobile phone which does not frighten him off. Furthermore, the C10 study provides a precedent or model for future product evaluation studies in Siemens.

**Literature**


