Making a Completely Icon-based Menu in Mobile Devices to become True: A User-centered Design Approach for its Development

Sabine Schröder
RWTH Aachen University
Jägerstr. 17-19
D 52056 Aachen, Germany
Tel. 0049-(0)241-8093535
Sabine.Schroeder@psych.rwth-aachen.de

Martina Ziefle
RWTH Aachen University
Jägerstr. 17-19
D 52056 Aachen, Germany
Tel. 0049-(0)241-8096130
Martina.Ziefle@psych.rwth-aachen.de

ABSTRACT
This paper investigates the feasibility and usefulness of completely icon-based menus in mobile devices. The underlying research was arranged in four steps: (1) Employing the sign production method, users' prototypic mental concepts for the pictorial representation of verbal menu terms were obtained. (2) Having selected and implemented the most prototypic ideas their potential was evaluated in the second step using the matching and the naming method. (3) In the third step an icon-based menu was constructed, programmed and implemented. (4) Finally, the utility of this prototype vis-a-vis a text-based menu was investigated on an experimental basis in a menu search task. The icon- based menu is shown to produce a significantly higher learnability thus outweighing initial differences. In general, results demonstrate that icon-based menus are basically viable and successfully applicable, *if* a users-centered procedure is pursued and *if* icons represent prototypical semantic knowledge.

Categories and Subject Descriptors
D.3.3 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces (D.2.2, H1.2, I.3.6).

General Terms
Performance, Design, Experimentation, Human Factors.

Keywords
Icon-based menu, text-based menu, population stereotypes, sign-production method, matching method, naming method, menu search task.

1. INTRODUCTION
Every day new products based on information and communication technology appear on the market: mobile phones, PDAs, pagers, communicators – all of them equipped with a text-based menu. This component is used in spite of its obvious disadvantages.

The linguistic terms used to indicate an item's functions, as well as their assignment to function classes (category headers) are oftentimes ambiguous and difficult to comprehend [16, 17]. Furthermore, the representation of linguistic terms is space consuming thus wasting a resource of increasing scarcity. Finally, text-based menu structures establish language barriers, a crucial disadvantage in times of increasing globalization.

By now popular devices (e.g. Ipod™) are distributed to more than 20 different language areas. Producers would be much better off, if they could confine themselves to the production of only one prototype instead of several models tailored to the respective languages of the target countries. Against this background it is tempting to investigate the feasibility of an icon based menu.

In several studies, icons have demonstrated their universal comprehensibility [6,8,14,20]. Furthermore, indications of their superiority with respect to usability have been found [5,22,23]. Thus, they improve the user's willingness for learning. A general disadvantage of icons, though, is their lack of grammar used in literary language to express relations between terms. This deficit is critical because menu items on different levels usually are linked to each other by hierarchical relations and located in a syntactical and semantical structure. Furthermore, it is doubtful whether adequate pictorial representations that differ both with respect to their content and their level of abstraction can be found for all menu items.

According to the present state of knowledge a strong interrelation exists between type of information and its optimal representation: Whereas the pictorial mode is superior in representing information about real objects and events, the text mode is superior when abstract content has to be represented and logical ways of thinking have to be supported [3,7]. Consequently, the results of studies comparing icons and text differed depending on the type and complexity of the tasks under observation. Whereas in non-hierarchical menu structures shorter search times resulted when information was represented by icons [1,18], in hierarchical menu structures either none [2] or the text format [21,23] was superior. The exclusive presentation of icons in non-hierarchical format has been employed oftentimes (e.g., in tool bars). In contrast, purely hierarchical icon structures have not been used so far. Therefore one could presume that the potential of pictorial menus is masked by lacking familiarity with the usage of icon-based menus. On the evidence of the studies presently available, this question – and

MobileHCI 2008, September 2–5, 2008, Amsterdam, the Netherlands.

hence the question of the superiority of one or the other format - cannot be answered definitely either way.

Therefore a series of studies was initiated with the intention to investigate the feasibility of an icon-based menu and to compare its effects to text-based menus. Thus, we installed an innovative research project with two goals: (1) to pursue a strict user-centered experimental and empirical design strategy and (2) to discover new ways of interface design for producers of small mobile technical devices.

2. PROCEDURE

As shown in table 1 the studies were conducted in four steps: (1) Icon representations were developed for each menu item. (2) The icons were evaluated and modified if necessary. (3) An icon-based menu was designed for a mobile phone. (4) Finally, the usefulness of the icon-based menu structure was experimentally compared to a text-based menu version. To integrate the users' cognitive models into the design process, we included participants (N=120 in total) in all stages of the design process. The single studies and the menu design will be discussed in more detail below.

Table 1. Research study process.

Step 1 – Icon Development	→	**Step 2** – Icon Evaluation	→	**Step 3** – Menu Development	→	**Step 4** – Menu Evaluation

3. STEP 1: ICON DEVELOPMENT

3.1 Objectives

Icon design was directed by the goal to find for each menu term its optimal pictorial counterpart, i.e. the representation which expresses the function involved as unequivocally as possible and which is comprehended by as many users as possible. This objective involved the question whether users agree about the mental model of a concept in terms of population stereotypes. Starting from the observation that mobile phone menus usually are composed of categories (e.g., messages), functions (e.g., text message), and actions (e.g., delete), it was another research question, whether such population stereotypes exist for each concept type.

3.2 Method

To collect ideas for the design of items the ***sign production method*** of Howell und Fuchs [10] was applied. According to this method respondents are asked to suggest pictorial representations for linguistic terms. The method relies on the assumption that icons which have been designed by users will have a higher chance of being correctly interpreted by them. This assumption is derived from the concept that every figure directly maps the user's mental model. A high degree of conformance in the icons suggested therefore is considered to be indicative for the existence of a populations stereotype. 56 persons – 32 females, 24 males – aged between 24-56 years participated. They were encouraged to submit as many suggestions as possible for various function and category terms of a mobile phone.

Due to the strong interrelations between the linguistic terms and their associations the study was based on a fictitious menu the terms of which had proven to be universally comprehensible in several previous studies [16, 17]. In addition, by using a fictious menu biases due to familiarity with specific brands were avoided. The menu comprised a total of 76 different items – among them 12 *category identifiers* (e.g. "message"), 30 *function names* (e.g. "text message"), 19 *actions* (e.g. "cancel") and 12 other items, which could not be assigned to one of these concept types (e.g., "level 1, level 2, etc." in the submenu "sound volume"). These menu items represented a total of eight different submenus consisting of three to five levels each and 103 mobile phone functions. Thus, the menu complexity was comparable to the complexity of actual mobile phones.

As we intended to explore the contextual dependence of the suggested icons and to realize a modular menu structure, in which icons can be combined in any order according to modular design principles, icons for functions and actions were not only inquired individually (e.g., "text message" or "delete"), but also in combination (e.g., "delete text message"). Therefore, in addition to the 76 individual concepts 116 combined ones were inquired.

In order to keep the burden on respondents as low as possible eight versions of the questionnaire were developed. Each version consisted of 24 concepts and was processed by seven participants, each. Terms of one submenu were queried jointly and not distributed across several questionnaires in order to enable participants to develop a logically consistent pictorial language. Figure 1 shows example items. Each questionnaire contained single as well as combined concepts; redundancies were avoided. The order of concepts within each questionnaire was varied to minimize sequence effects.

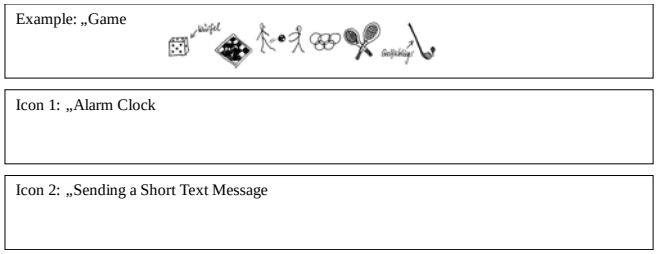


Figure 1. Exemplary items of the sign-production sheet.

Following Howell and Fuchs [10] data were collected on two ***dependent variables***: (1) the *number of suggestions per item* (V_i) in order to measure the ease of visualization of a concept, and (2) the *number of different ideas per item* (I_i). This measure required a classification of the ideas with regard to the similarity of their content.

A joint analysis of both dependent variables provides insight as to how close a given suggestion is to a population stereotype: the maximum is attained when all participants submit but one suggestion and the suggestions of all participants with respect to one concept are identical. At the other extreme, a very low degree of stereotypicality is indicated by a large number of suggestions per item which differ strongly in their content. Since a small number of ideas may be associated with a large, as well as with a small number of suggestions, and both situations involve completely different interpretations, an isolated analysis of only the number of ideas is insufficient.

Analysis focused upon as to how strongly the dependent variables reflected different concept types (e.g., category vs. function vs. action) and query formats (single vs. combined). In addition, distinctive qualitative features were looked for.

Four evaluators participated in the analyses who were not only responsible for counting and classifying the suggestions, but also had to select the best idea among the alternatives available. In order to perform these tasks properly they were provided with a list of criteria which required, amongst others, consideration of an idea's frequency of occurrence and of its feasibility.

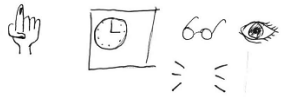
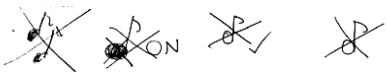
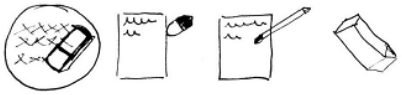
3.3 Results

ANOVA methods were used to analyze the data whenever appropriate. Otherwise the non-parametric Kruskal Wallis and Friedman tests were applied. The level of significance was set to 5 %; results reaching a level of 10% are referred to as marginally significant.

The number of suggestions per item and respondent (V_{ij}) ranged between 0.4 and 2.5 (M_{ij}=1.2, SD=.36). The lowest number of suggestions was obtained for the menu items "equipment" (M_{ij}=0.4) and "menu view" (M_{ij}=0.5), for which only half of the participants suggested a pictorial representation. On the other side, the largest number of suggestions resulted for the menu items "display brightness" (M_{ij}=2.5) and "sound volume" (M_{ij}=2.33). The total number of different ideas per item varied between 1 and 13. No effects on the number of suggestions per item were observed for concept type and query format.

A high degree of stereotypicality was observed particularly for concepts associated with objects and events, which are part of users' experience world, such as "meeting", "camera", "calendar", "alarm clock", "time of day", "portrait photo" (Table 2, line 1). At least 75% of participants used the objects' visual appearance for their pictorial representation. Thus, the icon was equated with the term's content.

Table 2. Examples of produced icon types.

	Example 1	**Example 2**
(1) **Icons with high stereotypicality**	"Meeting" 	"Calendar" 
(2) **Icons with low stereotypicality**	"Delete an email" 	"Show" 
(3) **Combined icons of concrete and abstract elements**	"Set Silent" 	"Add" 
(4) **Metaphorical icons**	"Read" 	"Delete" 
(5) **Icons from other systems**	"Save" 	"Contrast" 
(6) **Context dependent icons**	"Edit" 	"Alarm clock" 
(7) **Outdated icons**	"Alarm clock" 	"Telephone calls" 

A low degree of stereotypicality in contrast was found for concepts, which tended to be abstract such as “show”, “communication”, and “settings”, for which respondents came forward with several different proposals in each case (Table 2, line 2). Howell and Fuchs [10] ascribe this phenomenon to the fact that abstract concepts defy immediate physical perception and therefore have to be represented metaphorically. By this, the Whereas for the first item, however, all proposals were based on the same idea (perfect agreement) there was no agreement at all concerning the ideas for the latter item (perfect disagreement).

A significant effect was observed for the type of query on the number of ideas, which was higher for queries of combined concepts than for those with single concepts ($\chi^2(1)=9.14$, $p<0.05$). Obviously the combined presentation causes a broader activation of the semantic memory, which supports the generation of more ideas. In contrast, no significant effects of the type of concept were found. degrees of freedom with respect to their representation are increased. Therefore as a rule a direct link between icon and underlying concept does not exist. The following examples may illustrate these situations: For both menu items “calendar” and “delete an email”, seven proposals were collected in total.

Several additional insights were generated by a qualitative analysis of the responses:

- Participants preferred figures of concrete items derived from their visual-spatial experience world (material objects, people, animals, etc.). Irrespective of the level of abstraction, abstract elements (e.g., arrows) merely served to clarify certain features of the items (Table 2, line 3). This confirms findings of Rogers [15], according to which abstract concepts are best represented by combinations of concrete and abstract elements.
- In many cases, verbal concepts were represented metaphorically by selecting associated objects involved in the actions (e.g., “read” by “eyeglasses”) because they lack physical perception (Table 2, line 4).
- Again and again icon representations were resorted which were familiar to participants from other systems. The disk, for example, used by WORD™, was frequently proposed for the concept “store” (Table 2, line 5). This pattern of behaviour demonstrates the strong influence of learning patterns on mental models and the importance of consistent applications of icons in different systems.
- With strikingly frequency icon representations were found to be context dependent (Table 2, line 6). Obviously typical utilization settings are automatically associated with concept queries.
- Interestingly, icons suggested oftentimes drag behind the current state of technology. Good examples for this are the proposals for “alarm clock” and “telephone”, which are outdated from today’s perspective (Table 2, line 7). Obviously, time and the associated modernization are not indicative for the user’s mental model.

In spite of these peculiarities, the icons with the highest frequency of occurrence had to be used for the design of the mobile phone menu. Four evaluators chose icon representations for each verbal concept independent from each other. Interrater reliability reached an impressively high value of r=.86 - which shows the prototypicality of the proposed icons from another side. Deviant choices were reconciled by open discussions and evaluators made proposals to improve representations with low degrees of prototypicality.

Overall, 13 items were revised by either changing their verbal description or by redesigning their pictorial representation. The final selection of items was realized using Adobe Photoshop 7.0™. A pictorial representation in black and white-format was generated for each of the 76 menu terms and 116 combinations (each consisting of functions and actions). The combinations were modularly assembled. Examples of the final icon representations are shown in Table 3 for each concept type and each query format. The results show that population stereotypes do exist even in evolutionary young areas such as micro technology.

Table 3. Examples of icon prototypes

		Single Icons	Combined Icons	
Category	Camera		-	-
Function	Picture		-	-
Action	Send		Send a Picture	
	Save		Save a Picture	
	Zoom		Zoom a Picture	
	Delete		Delete a Picture	

In addition, these stereotypes do not depend neither on concept type nor on query format. This finding is supported by the – sometimes remarkably high – consistency of the proposals, but also by the high agreement among the evaluators in icon selection. Thus, the sign production method was found to be a valid method for uncovering population stereotypes and their underlying mental models.

4. STEP 2: ICON EVALUATION

4.1 Objectives

Cognitive task requirements differ fundamentally for the development and the use of icons: Whereas the designer has to translate a concept into an icon the user has to translate an icon into a concept. These processes are symmetrical to each other if and only if the mental models of the designer and the users are in conformance; otherwise asymmetries will occur. Therefore, the quality of the icon selection process was evaluated in the second step of the study.

Since icon development was guided by the mental model of potential users a high transparency was expected for the majority of icons. The validity of this assumption was checked by an additional analysis. Furthermore, the comparability of the visual and textual items had to be ensured in order to eliminate qualitative differences between the modes of representation as potential interfering variables. Finally, the question was addressed, whether the icons presented simultaneously on the same menu level could be distinguished from each other.

4.2 Method

To explore the questions stated above the ***matching method*** and the ***naming method*** were employed. Both methods reflect different aspects of the evaluation process: Using the matching method, the suitability of an icon has to be evaluated in comparison to other icon variants. Within the naming procedure, only the icon at issue is presented and evaluators have to indicate what the icon means in the respective context. Overall, the naming method is more critical, since the evaluation is not supported by comparison processes as it is the case in the matching method, where evaluators may be biased by the context information presented by distractor items. Within the naming procedure, evaluators must rely on a full understanding of the icon.

Concerning the ***matching method*** participants were given a limited selection of icons and menu items and asked to link pictorial and verbal representations to each other. Each list of menu items included two dummies that were very similar to target concepts, in order to impede guessing. Figure 2 shows an excerpt from a questionnaire. Within the ***naming method*** solely the icons were presented to the subjects and their meaning had to be specified verbally.

			3	Edit
1			*2*	Write
2				Format
3			*1*	Delete
				Record

Figure 2. Examplary items of the matching sheet

For each of the two methods two questionnaires were developed in order to query single menu items (N=76) separately from combined menu items (N=116). To avoid sequence effects an additional version was developed for each of these questionnaires presenting the icons in reversed order. Adobe Designer 7.0™ was used to generate the electronic format of the questionnaires which were sent to participants by e-mail.

48 subjects – 22 male and 26 female, aged between 18-62 years – participated, each 24 of them processed the matching and the naming task.

The number of correct links and the number of correct replies respectively were selected as *dependent variables* in the matching and the naming task. The use of the latter required a distinction between "correct" and "wrong" replies which was supplied by two judges each acting independent from each other.

The data were analyzed with respect to effects of concept type (category vs. function vs. action) and query format (single vs. combined).

4.3 Results

Even though the results of both methods basically showed a high transparency of icons, the number of correct replies differed significantly between them ($\chi^2(1)=10.57$, $p<0.05$): Whereas 89% (SD=.18) of icons were correctly identified in the matching task , only 73% (SD=.29) were correctly classified in the naming task. Furthermore, 99 icons in the first case compared to only 32 icons in the latter case scored a perfect recognition rate, i.e. were correctly interpreted in all cases. In both cases such icons mostly represented categories and functions, e.g. the icons for the menu items "ringing tone", "birthday", "alarm clock", and "keypad tones". Icons with low recognition rates (below 50%), in contrast primarily addressed functions or actions, as "voice recorder", "loudness", "enlarge", and "show". No effects of concept type and query format on the number of correct assignments were revealed for either of the two methods.

The results suggest that the matching method tends to overestimate the comprehensibility of icons. This bias may be due to the fact that the method does not capture the understandability of the icon within the concept space but rather the recognition of its possible meaning from a small set of choices. The naming method in contrast does not confine the number of possible replies but calls for spontaneous associations. Since it therefore better fits the task of assessing the comprehensibility of icons its results are used as a benchmark for the evaluation.

In accordance with recommendations by ISO 9186 [11] icons with recognition rates below 66% were regarded as not sufficiently well-defined and revised. This applied to 14 icons. With these revisions all the raw materials for the construction of an item based menu were available.

5. STEP 3: MENU DEVELOPMENT

For the simulation of a virtual mobile phone menu a specific software system was developed using Flash und Java. By virtue of this software icons could be arranged in a hierarchical tree structure which corresponded to the linguistic menu with respect to both appearance and content. Whereas on the highest level solely categories were presented, lower level icons represented functions and actions. Participants could freely navigate within this tree structure and search for information. For data presentation and collection a touchscreen (Iyama TX 3841™) was used, which enabled direct menu item selection with a pencil. Each selection was followed by a new sub-menu appearing on the screen. "Return" and "Start Menu"-soft buttons allowed direct access to the next higher level and the start menu, respectively.

A mobile phone corpus was used as screen background. On its display the items of each level were presented as lists. Menu items were displayed in 14-sized Arial font. Icon size was 69x34 pixels. Figure 3 shows a participant using the simulated mobile phone within the experimental set-up.

Figure 3. Simulated mobile phone menu

Furthermore, the software included a logfile tool for parallel data recording and navigation performance logging.

6. STEP 4: MENU EVALUATION AND PERFORMANCE

6.1 Objectives

Finally, the usefulness of the icon-based menu was experimentally evaluated against a text-based menu. In addition, two variants of orientation guides to support the logical linkage of items located on different levels were tested: Selected functions were either transferred to the head line of the subjacent action menu or presented by combining them directly with their respective actions. Thus, the question of single or combined information presentation was reformulated and studied from the viewpoint of the better orientation guide.

6.2 Method

Two independent variables were defined: (1) the mode of presentation of a menu structure (pictorial vs. textual presentation) and (2) the type of orientation guide (headline vs. item combination). Complete permutation of variables yielded a 2x2 design consisting of four combinations of conditions, resulting in four different menu structures (Figure 4):

- a pictorial menu presenting functions as headlines
- a pictorial menu presenting functions in combination with actions
- a textual menu presenting functions as headlines
- a textual menu presenting functions in combination with actions.

Each menu contained 257 entries resulting from multiple use of icons on different levels. The usefulness of the menu structures was tested with an information search task administered to a sample of 40 subjects aged between 21-32 years (17 males and 23 females). 10 subjects were randomly assigned to each of the four menu conditions. Participants had to find eight different mobile phone applications in the menu hierarchy: (1) Setting the alarm clock, (2) Zooming a portrait photo, (3) Listening to a voice message, (4) Switching off the keypad tones, (5) Using the voice recorder (6) Selecting a new ringing tone, (7) Storing a private number and (8) Adding a meeting.

Pictorial menu: Functions as headlines	**Pictorial menu: Functions combined with actions**
	
Textual menu: Functions as headlines	**Textual menu: Functions combined with actions**
	

Figure 4. The four different menu structures

Since icons are pictorial representations of concepts, task actions were phrased such that the mobile phone functions sought were termed in accordance with the terms used in the production sign method whenever icons were used for concept representation. When textual representations of concepts were used, however, instructions were phrased differently because the terms used in the sign production method were integrated into the textual menu and their prior mentioning presumably would have biased search outcomes. Thus, differently phrased task instructions were used for the pictorial and the textual menu. This procedure may be exemplified by the function “Switching off the keypad tones”. The instructions for both menu types were:

> **Icon menu:** “You want to switch off the tones which occur when letters or numbers are entered”.
>
> **Text menu:** “You want to enter letters and numbers without a sound”

Subsequently to a training task participants processed all eight tasks sequentially each starting from the start menu. The order of tasks was varied across participants. The minimal number of steps required for the solution of a task varied between three and five. If a participant believed that (s)he had made one or more incorrect choices while moving in the menu (s)he could move backwards again via the soft buttons. Maximal processing time per task was limited to three minutes.

Measures of effectiveness, efficiency and satisfaction were used as ***dependent variables***. In detail they were

1. effectiveness (per task and in total)
2. processing time (per task and in total)
3. item selection time, i.e., the average time required for the selection of an item which is computed by dividing total processing time by the number of steps executed; Empirical evidence suggests that item selection is longer for pictorial representations [5, 23] as the "visual appeal" provokes to abide for a longer time;
4. the number of detour steps (per task and in total), defined as the difference between the actual and the minimal number of steps required for task solution (without consideration of returns in order to avoid double counting of errors)
5. user acceptance of the respective menu type. After task completion, participants processed an acceptance questionnaire with eight statements addressing the unambiguousness of icons or terms and menu structures (e.g., "I had a clear idea of the meaning of the icon / menu term" or "In general menu navigation was self explanatory"). To evaluate the statements respondents supplied ratings ranging each from 1 (="completely false") to 4 (="completely true").

6.3 Results

In spite of the newness of pictorial menu structures and in spite of the high familiarity with text menus in mobile phones effectiveness revealed to be high for both modes of presentation. 39 out of 40 participants managed to solve all eight tasks within the time limit. Therefore no further results about processing outcome will be presented in the following. Efficiency and acceptance measures will be focused instead.

Data were analyzed including efficiency measures by a MANOVA. With respect to presentation mode a marginally significant omnibus effect ($F(2,20)=2.9$, $p<0.1$) was observed which resulted from differences in total processing time ($F(1,32)=3.8$, $p<0.1$) and the number of detour steps ($F(1,32)=5.7$, $p<0.05$).

The values of both variables were smaller for the text mode, for which the eight tasks on the average were completed after 207 seconds (compared to 231 seconds for the pictorial mode) with an average of 1.9 detour steps per task (as compared to 2.2 detour steps for the pictorial mode). Item selection time did not differ significantly between the two modes.

Type of orientation guide did not produce any significant effects neither for overall nor for individual efficiency measures. Factor interactions were not statistically significant either. Figure 5 shows the differences in total processing time and detour steps for the presentation forms.

Although the outcomes indicate a superiority of text-based menus, differences have to be rated as small in view of the fact that participants did not have any experience with completely icon-based menus. Therefore the question arises whether and to what extent use of icon-based menus affects efficiency. In order to answer this question the efficiency measures collected for the individual tasks were compared to each other for the total processing period.

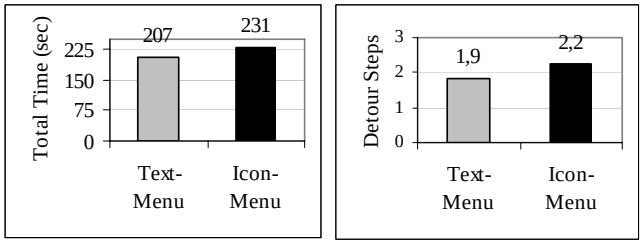


Figure 5. Differences in total time and detour steps between the two modes of representation

Analyses showed that the minor superiority of text-based menus vanished with increasing experience in menu handling. While significant learning effects occurred for both modes of representation between the first and the last task, they were much larger in the pictorial than in the textual mode. Processing time, e.g., was reduced in the text mode by 50% from 48.7 to 24.2 sec, whereas it was reduced in the pictorial mode by almost 75% from 48.7 to 16.2 sec ($\chi^2(7)=33.1$, $p<.001$) (Figure 6). In both cases reductions occurred continuously. The data make clear that respondents provided with pictorial menu structures performed worse only in the beginning whereas they outperformed respondents provided with textual menu structures in the end. These observations suggest that the initial disadvantages in performance result from lack of familiarity with pictorial menus. Additional studies are needed to clarify whether the ease and the success of learning of pictorial menus results in better performance in the long run, too.

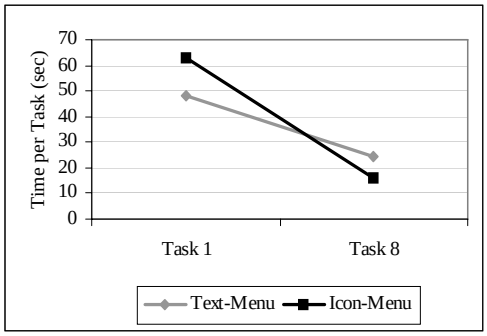


Figure 6. Learning effects during the processing period

Only weak support for the alleged superiority of the textual modus was also found when performance on the individual task level was analyzed. Significant differences in the number of detour steps between the two modes of representation occurred only for the tasks "set the alarm clock" ($\chi^2(1)=9.93$, $p<.01$), "zooming a potrait photo" ($\chi^2(1)=11.10$, $p<.01$), and "select a new ringing tone" ($\chi^2(1)=7.47$, $p<.05$), and it was only in the last task that the difference was in favour of the textual mode.

By way of qualitative analyses of detour steps menu items were identified for both modes of representation that tended to pose problems in task processing. These problematic items, however, did not have any regularities that might be interpreted as weaknesses of one or the other mode or would even challenge its usefulness for menu representation. It was striking, on the contrary, that the problematic menu items were totally different from each other.

Thus, not only cases could be observed where textual representations performed better but also ones with superior pictorial representations. In the textual representation, e.g., the search for the alarm clock-function proved to be problematic, because the superordinate category “date/time” was not always recognized. Instead participants selected various other categories such as “phone settings” or “communication”. On average they made two false selections before they identified the correct category “date/time”. The associated icon, in contrast, was recognized immediately. The superior performance of the icon-based menu is particularly amazing in view of the fact that icon development was based on verbal descriptions of the function. A potential explanation for this phenomenon might be the fact that by using concrete icons for all menu items a higher degree of unambiguousness was achieved than available for their verbal counterparts which in part represented abstract concepts.

Finally, in the last step, acceptance ratings were considered. As figure 7 shows the - basically positive - assessments of items and menus differed only marginally between the two modes of representation.

A significant effect on user acceptance, however, resulted from the type of orientation guide employed in the textual representation mode ($F(3,36)=3.6$; $p<.05$): The combined presentation of functions and actions ($M=3.49$, $SD=.13$) showed a significant lower acceptance rate than the presentation of functions in the headline of the pertaining actions ($M=3.76$; $SD=.17$).

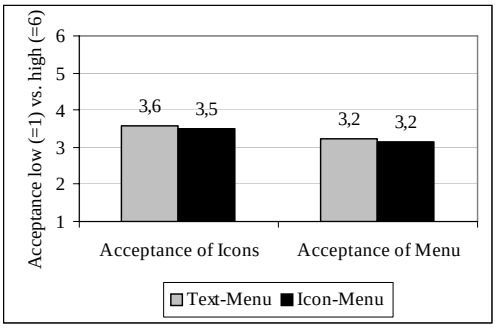


Figure 7. Acceptance ratings

7. DISCUSSION

The research reported in this paper represents a first attempt to investigate the feasibility and usefulness of completely icon-based menus for mobile phones. The latter stand for a whole class of devices which are characterized by the necessity to present a multitude of information on a limited, very small space. A positive evaluation might yield two advantages: First, from a methodological point of view, icons may serve as an interculturally understandable language meeting demands of user diversity. Second, from an economic perspective, the usage of icons in menus could effect considerable cost reductions for small scale devices on global markets.

Since no completely icon-based menus are available at present, this research intended to present a methodology for its development. The methodology is based on several consecutive steps, from icon design and evaluation to icon-based menu design and evaluation. Following strictly demands of user-centred design, potential users participated in all steps of this process, except for the process of software development and construction of the icon-based menu.

In general, the results with respect to icon generation strongly support the feasibility of a completely icon-based menu. Support in particular comes from the astoundingly high agreement in users’ associations when translating verbal concepts into pictorial representations thus suggesting the existence of population stereotypes. Utilization of such population stereotypes for the design of icon representations has the potential for high and universal comprehensibility as confirmed by the evaluation. In addition, feasibility of icon-based menus is supported by the finding that population stereotypes were found for all menu terms, independently whether used for category headers, functions names or actions. Yet there were differences between concrete and abstract concepts: Whereas stereotypicity of the suggestions was high for the former it tended to be rather low for the latter. Abstract concepts do not have a visual counterpart that might be used for icon design. They have to be represented metaphorically instead; the associated wide range of possibilities comes along with a low degree of stereotypicity of the proposals. This problem may be overcome by the development of universal standards for the pictorial representation of menu items. Traffic signs are a good example for the possibility to develop a universally comprehensible design pattern for abstract concepts, provided that they are consistently applied. Moreover, it should be kept in mind that abstract concepts effect ambiguity whether represented pictorially or textually [16]. Therefore, problems associated with abstract menu items are general, not icon specific.

In spite of these problems an attempt was made to develop as unambiguous icon representations as possible for all menu items, irrespective of their being categories, functions or action terms. Based on an iterative process first of all the best from a set of alternatives generated by respondents was selected by four evaluators and drafted in electronic form. Subsequently the quality of each icon was evaluated and icons were revised if necessary. Not until then was the icon based menu constructed and its usefulness evaluated vis-à-vis a standard text based menu.

Within the experimental framework of a simulated phone menu, text-based menus tended to perform better with respect to both processing time and detour steps. This finding, however, proved to be short-lived. In the course of the experiment users of icon based menus were able to increase their navigation efficiency much faster than users of textual menus with the effect that the former performed even better than the latter in the end.

According to the authors’ judgment the superiority of pictorial representations in the end outweighs their inferiority in the beginning of the experiment. This evaluation is primarily based on the long experience in using text-based menus which is associated with a knowledge advantage in the beginning. Users of icon based menus, on the other hand, could not resort to any experience and therefore at first had to become familiar with the task requirements. Delays resulting from this lack of experience decay with increasing practice. Therefore a better comparability of the two menu structures may be assumed to exist at the end of the task processing. Because of these considerations our findings not only demonstrate the competitiveness of the icon based menu designed but also suggest the possibility that icon based menus may perform better than text based menus in the long run. However,

this suggestion rests on the critical condition that potential users' are incorporated into the design of the icon-based menu.

Yet, the time and effort required for the development of the icon based menu poses the questions of the method's cost-effectiveness and of the possibilities of shortcuts. From our point of view the use of the sign production method for icon development and the execution of an information search task for icon evaluation are necessary and sufficient for the construction of an icon based menu. Icon evaluation, however, may be waived if comprehensibility is ensured by having respondents participate in the development work.

All in all the findings are favorable for both modes of representation as demonstrated by the overall high performance (97.5% of participants solved the tasks and required less than 1/8 of the time available). However, it should be noted that some kind of ideal users participated in the evaluation study: young, bright and technology-prone students, which are not representative for the whole range of users. The good performance outcomes therefore have to be taken with caution considering the practical case of less gifted users of mobile devices [25, 26].

8. LIMITATIONS AND IMPACT FOR FUTURE RESEARCH

In principle, our research argues in favor of an icon-based menu for mobile devices and suggests a methodology for its development. In order to define the potential as well as the restrictions of completely icon-based menus more clearly, however, some limitations should be considered as well:

(1) On the basis of our findings we cannot definitely say that icon-based menus are truly internationally and -culturally comprehensible. Though the existence of cognitive universals across cultures had been confirmed in several studies [6,13,14], indications for culture-specific stereotypes are also available [13]. Hence, before the icon-based menu can be finally recommended for international adoption its usefulness for different language regions and cultures should be clarified.

(2) For legibility reasons, comparatively large-sized icons were used in the menu search task. Reality in contrast is characterized by an ever increasing miniaturization of display symbols causing a reduced visibility [21]. Thus, the outcomes presented here should be confirmed with smaller icon sizes.

(3) We used black-and-white icons throughout, maximizing luminance contrast and minimizing confounding effects stemming from the usage of different colors. Empirical studies, however, show that colors are important and distinctive icon attributes for information selection and recognition [9,12]. Inclusion of colors might therefore further improve the efficiency of icon based menus.

(4) Finally, the learning effects reported here strictly speaking reflect more likely retention effects than stable learning effects. Thus we will have to show in future research that performance advantages do persist after an extensive familiarization to icon-based menus and that this holds for a different user group.

9. ACKNOWLEDGMENTS

Many thanks to Mine Kögel for her research assistance and to Thomas Brambring, Kresimir Spisic and Tobias Vaegs, who developed the simulated mobile device.

10. REFERENCES

[1] Arend, U., Muthig, K.P. & Wandmacher, J. Evidence for global superiority in menu selection by icons. *Behaviour and Information Technology*, 6 (1987), 411-426.

[2] Benbasat, I. & Todd, P. An experimental investigation of interface design alternatives: icon vs. text and direct manipulation vs. Menus. *International Journal of Man-Machine Studies, 38* (1993), 369-402

[3] Booher, H.R. Relative Comprehensibility of Pictorial Information and Printed Words in Proceduralized Instructions. *Human Factors, 17* (3) (1975), 266-277.

[4] Davis, S. A. & Bostrom, R. P. Training End Users: An Experimental Investigation of the Roles of the Computer Interface and Training Methods. *MIS Quarterly*, 17 (1) (1993), 61-85.

[5] Egido, C. & Patterson, J. Pictures and category labels as navigational aids for catalog browsing. *Proc. CHI 88*, ACM Press (1988), 127-132.

[6] Ferguson, A. & Brown, W. (2003). Standard symbols - using the iso test method. In: P. T. McCabe (ed.), *Contemporary Ergonomics*, Taylor & Francis (2003), 531-536.

[7] Fisk, A.D. & Kobylak, R.F. Relative value of pictures and text in conveying information: perfomance and memory evaluation. *Proc. 30th Annual Meeting Human Factors Society 1986*, 1269-1272.

[8] Garcia, M., Badre, A.N., & Stasko, J.T. Development and validation of icons varying in their abstractness. *Interacting with Computers, 6* (1994), 191-211.

[9] Helbing, K. G., Jenkins, J., Kim, Y.S., and Miller, M.E. (1993). Influence of Icon Detail, Color, and Perspective on Preference, Recognition Time, and Search Time. *Proc. Interface '93*

[10] http://www.frontiernet.net/~mkmiller/Prof/Icons/icon.ht

[11] Howell, W.C. & Fuchs, A.H. Population stereotypy in code design. *Organisational Behavior in Human Performance*, 3 (1968), 310-339.

[12] ISO 9186. International Organization for Standardization. Graphic Symbols. Test methods for judged comprehensibility and for comprehension (2001).

[13] Lansdale, M. On the memorability of icons in an information retrieval task. *Behaviour and Information Technology*, 7, 2 (1988), 131-151

[14] Pappachan, P. & Ziefle, M. Cross-Cultural Comprehensibility of Icons in Mobile-Computer-Interaction. *Behaviour and Information Technology* (in press).

[15] Piamonte, D. P. T., Ohlssen, K., Abeysekera, J. D. A. Across the Seas: A User-Based Evaluation of Candidate Telecommunication Icons. *Research report of the department of Human Work Sciences of Luleá Tekniska Universitet*, (1999), 4-135.

[16] Rogers, Y. Pictorial Communication of abstract verbs in relation to human-computer interaction. *British Journal of Psychology*, 78 (1987), 99-112.

[17] Schröder, S. and Ziefle, M. Semantic transparency of cellular phone menus. Comparing users from different age groups. In B. Fisseni, H.-C. Schmitz, B. Schröder & P. Wagner (Eds.)

Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen, Peter Lang (2005), 302-315.

[18] Schröder, S. & Ziefle, M. Evaluating the usability of cellular phones' menu structure in a cross-cultural study. In R.N. Pikaar; E.A. Konigsveld & P.J. Settels (Eds.) *Proc. IEA* 2006, Elsevier (2006).

[19] Schröder, S. & Ziefle, M. Icon design on small screens: Effects of miniaturization on speed and accuracy in visual search. *Proc. of the 50th Conference on Human Factors and Ergonomics Society* (pp. 544-549). Santa Monica: Human Factors and Ergonomic Society (2006).

[20] Schröder. S. & Ziefle, M. Semantic transparency of icons in mobile devices- a comparison of younger and older users. *Human Computer Interaction for Aging*. LNCS. Heidelberg, Springer (in press).

[21] Wandmacher, J. & Müller, U. On the usability of verbal and iconic command representations. *Zeitschrift für Psychologie*, Suppl. 9 (1987), 35-45.

[22] Whiteside, J., Jones, S., Levy, P.S. & Wixon, D. (1985). User performance with command, menu, and iconic interfaces. In L. Borman und B. Curtis (Eds.), *Human Factors in Computing Systems*, ACM Press (1985), 185-191.

[23] Wiedenbeck, S. & Davis, S.: The Influence of Interaction Style and Experience on User Perceptions of Software Packages. *International Journal of Human-Computer Studies*, *46*, 5 (1997), 563-588.

[24] Wiedenbeck, S. The use of icons and labels in an end user application program: an empirical study of learning and retention. *Behaviour & Information Technology*, *18*, 2 (1999), 68-82.

[25] Ziefle, Schroeder, Strenk & Michel. How young and older users master the use of hyperlinks in small screen devices. *Proc. SIGCHI 2007*, ACM Press (2007), 307-316.

[26] Ziefle, M. & Bay, S.. Transgenerational Designs in Mobile Technology: In J. Lumsden (Ed.). *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*. Hershey, PA: Information Science Reference (in press).