

# EVALUATION OF USER INTERFACES FOR DISPLAYING TEXT MESSAGES ON MOBILE DEVICES

Olaf Thiele  
University of Mannheim, Germany  
Email: thiele@uni-mannheim.de

Pascal Sütterlin  
University of Mannheim, Germany  
Email: suetterlin@bwl.uni-mannheim.de

Lars Kaczmirek  
Centre for Survey Research and Methodology, Germany  
Email: kaczmirek@zuma-mannheim.de

## ABSTRACT

More and more information such as short text messages (SMS) or emails is stored on mobile phones. Retrieving the right information at the right time has become a difficult task when using traditional interfaces included in today's cell phones. We created two novel user interfaces (one textual and one graphical) and compared both to the classical interface for retrieving stored SMS messages in a 34-participant usability experiment. Both novel interfaces yielded significantly better results than the existing classical interface. Moreover, the results show that the novel textual interface, which groups messages by social contacts, outperforms the complex graphical interface, which displays messages in a chessboard-like grid. Both objective and subjective findings are in favor of the textual interface except for one result: The graphical interface is more enjoyable than the textual one.

## KEY WORDS

Mobile Computing Contexts, User Interfaces, Information Visualization, Evaluation

## 1 Introduction

The information stored on mobile phones has increased immensely in recent years. While in the 1990s mobile phones were mainly used for making phone calls, they have become the digital companion affordable to almost everyone in today's world. Most modern phones include a camera, can play charts as ring tones and have processor performance which is equal to basic game consoles. While new functionalities and memory capacity are constantly added, few considerations are made concerning the management of the large amount of data. For example, SMS message boxes with 300 stored messages have the same retrieval interface as ten-year old cell phones holding a maximum of 10 messages. While this was a technical necessity in the past, new programming packages introduce the possibility to create novel user interfaces for retrieving stored SMS messages.

Nevertheless, what should a new interface for mobile devices look like? Both technical possibilities and human preferences have to be taken into account. To find out more

about an optimal design, we created two new interfaces for accessing stored SMS messages and evaluated them. The first novel interface is the list interface, which is similar to the classical interface in use today, except for two differences. Incoming and outgoing messages are joined and subsequently ordered by communication partner. After clicking on a communication partner, the user sees a list of all inbound and outbound messages sorted by date. In other words, the list shows both sides of a "conversation" in the sequence in which it occurred as opposed to the classical interface, which shows individual transmissions one at a time. While the list interface looks similar to the classical one, we designed the other novel interface to be more complex. The graphical interface allows for sophisticated usage. It places all messages in a chessboard-like grid and thus offers a good overview of conversations with one or more partners over time. However, due to its high level of abstraction, the model behind it needs to be learned and internalized. Finally, we created a classical interface as a baseline for the experiment.

We evaluated the two novel interfaces and the classical one in a usability study. Test persons received mobile phones and could use them freely, but all worked on the same message basis for each of the six given scenarios. All events, such as message opening and closing times, or time spans between events, were logged for objective evaluation. Results show that the list interface concerned with objective and subjective usability as well as retrieval success outperforms the graphical interface. Our considerations and findings are laid out in detail in the sections below. The first section deals with background information on SMS communication, the second section describes the technical progress, and related work is discussed in the following section. Both novel user interfaces are presented thoroughly in section three, including design decisions and technical realization. Section four introduces the methodology of the experiment, followed by its results in section five. The implications are discussed in the subsequent section. The last section concludes with final thoughts based on our findings regarding real world adoption of mobile information retrieval interfaces.

## 2 Background

### 2.1 SMS retrieval problems

SMS messages are the predominant means of textual communication when using mobile phones. According to the Netsize Guide 2005, SMS volume increased from 16 billion messages in 2001 to 36 billion in 2003 in Germany alone [1, p.131]. The Guide states that today over 200 billion SMS messages are exchanged in Europe per month. These account for 70% of data communication over mobile networks [1, p.32]. However, the basic principle for accessing stored messages on mobile phones has not changed at all. First, incoming and outgoing messages are stored in two distinct lists. Second, the lists are sorted by date, i.e. when the message was sent or received, in descending chronological order. This retrieving principle has been a feasible solution for a long time. Explaining how messages are stored takes only seconds due to the simplicity of the concept. Furthermore, accessing messages on a different phone is in most cases very easy. We programmed our own classical baseline SMS interface for the study and were surprised how fast test persons learned to use it. This classical technique works well with small numbers of messages. If only 20 or 30 messages need to be screened, the interface works well. Problems arise, however, when larger numbers of messages need to be stored. SMS storage capacities of 200 or more are common today and are likely to increase in the future.

### 2.2 Technological foundation

Early models of mobile phones could only be used to make phone calls and send SMS messages. Today's phones have enough processing power to encrypt phone calls in real time without using special hardware. This development offers the possibility to program complex applications like the graphical interface used in this study.

The Java 2 Micro Edition (J2ME) from Sun is the predominant programming model for mobile phones at the moment [2]. It is used to build small programs called MIDlets. A MIDlet can be seen as a classic Java applet that is loaded onto a mobile phone. MIDlets use a subset of the Java language and offer a wide range of functionality. We used the MIDlets to create all three SMS interfaces employed in our study.

### 2.3 Related Work

We found the following three fields of research to be inter-related with our study: information retrieval interfaces for mobile devices, general visual information retrieval and visualizations for the semantic web. Only few evaluations of graphical information retrieval interfaces for mobile devices have been published. Marcus complains about the lack of research in this area: *“One aspect that has received*

*relatively little attention is information visualization”* [3, sec.5]. One of the few studies was carried out by Masoodian and Lane who evaluated travel itinerary visualizations [4]. They compared a textual travel itinerary application with a visual one based on the Wireless Application Protocol (WAP). Masoodian and Lane conclude that the textual representation was more effective than the visual one due to the laboratory setting [4, sec. 6].

The field of general visual information retrieval has received more attention from researchers. Hearst gives an extensive overview of available techniques in 1999 [5, chap. 10]. In recent years, Donath developed the “Loom Project” which shows graphical illustration of Usenet conversations [6, p. 48]. Conversations are drawn as circles, with e-mails building the arch. The more e-mails a conversation comprises, the bigger the circle. Furthermore, colors are used to differentiate between different speakers in a conversation.

The field of semantic web visualization gives even more suggestions on how semantic information might be appropriately visualized. Fluit, Sabou and van Harmelen use given ontologies to draw interrelations between resources [7]. Again, dimensions like size, shape, position or color are employed to represent abstract relationships between entities. A good overview of other visual interface approaches for semantic information retrieval is given by Börner [8].

## 3 SMS retrieval interfaces

Three different interfaces were applied in the study: the classic interface representing the predominant user interface found in most modern mobile phones today, the list interface with a moderate level of abstraction showing messages ordered by communication partner and finally a complex visual interface with a chessboard-like structure. The two novel interfaces were derived from research on existing interfaces and SMS messaging usage. While related work has been presented in Subsection 2.3, the main functions and contexts of SMS messaging are presented in the following subsection. Subsequently, the two novel interfaces and the classical interface are presented together with design descriptions and alternatives.

### 3.1 Functions and contexts of SMS usage

In order to find a new user interface for SMS messages, one needs to take the main functions of SMS usage into account. Döring's reports five functions of SMS messages: contact, information, appeal, obligation and declaration [9, p. 27], with contact and information being the most frequent functions (combined usage 88%).

In a study of mobile phone usage among teenagers, Pedersen found that most users currently send ordinary SMS messages instead of multimedia messages or other SMS-based commercial services [10, p. 66]. He also

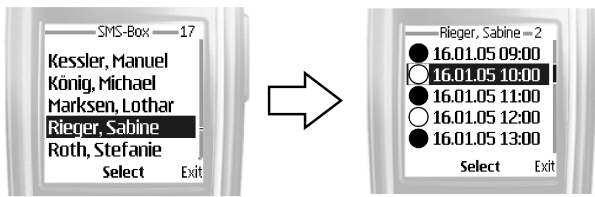


Figure 1. Screenshots list user interface

states that applications or services for sharing messages with friends or organizing important messages is still missing [10, p. 67].

In a recent study Tamminen et al. state that the context of mobile phone usage should be considered when designing a user interface [11]. They point out that mobile applications should be especially aware of the social contacts of the user [11, p. 141]. Rodden et al. support this view in their work on design for mobile contexts [12]. They argue that a successful novel software design needs to combine the contexts of both the virtual and real world.

### 3.2 The classical interface

In order to have the same look and feel for all three interfaces, we rebuilt a standard SMS interface as the baseline for the novel ones. Other advantages of programming our own classical interface are the possibility to record user behavior through the application and using the same messages for all three interfaces.

### 3.3 The list interface

The list interface is the first novel interface and was designed to give a comprehensive overview of the underlying message base. In contrast to the classical interface, both incoming and outgoing messages are combined and then sorted by communication partner for the list interface. The starting screen of the application shows an alphabetical list of communication partners. If one of them is selected by clicking on the name, a list of messages sorted chronologically is displayed. Each line represents one message and is comprised of an icon showing the direction of communication, together with the date and time the message was sent. Figure 1 shows two screenshots: the list of all communication partners and list of messages from and to a single communication partner.

The list interface was designed to ease navigation. As described in Section 3.1, Tamminen et al. ask for applications to take the social contacts of the user into account [11, p. 141]. By organizing the list interface around social contacts, queries where the communication partner is obvious but the exact time frame unknown are solved easily (e.g. "He sent me his address, but I can't remember when."). Furthermore, conversations are much easier to comprehend because no interfering messages are displayed.

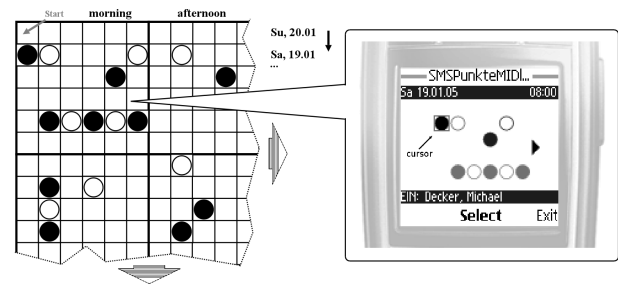


Figure 2. Basic principle and screenshot of the visual interface

### 3.4 The visual interface

The design rationale for the second novel interface - the visual interface - was to build a graphical information retrieval application that facilitates the comprehension of conversations as well as showing interrelations of messages over time.

The visual interface is composed of a chessboard-like grid where messages are represented by a circle. For this study, the grid was comprised of (18x18) 324 fields, each field containing only one or no message. Each row of the grid represents one day, the top row is the present day and the bottom row is 18 days before that. Columns were grouped into three categories: the leftmost columns represent the morning of a day, the middle columns show afternoon messages and columns on the right display evening messages. The ten most frequent communication partners have a distinct color, all others are black. Incoming messages are depicted by a shaded circle, while outgoing messages are indicated by an empty circle (analogous to the list interface). Figure 2 depicts the basic principle and a screenshot of the graphical interface.

The grid is split into 9 screens because not all 324 fields of the model can be drawn together on the display of a mobile device. Each screen contains (6x6) 36 fields. Users navigate through the screens by using a small joystick found on most modern mobile phones, moving one field at a time and allowing movements into the four main directions: up, down, left and right. Clicking on one of the circles opens the corresponding message.

To facilitate navigation between grids, small arrows at the four sides indicate whether there is another screen in that direction or not. The cursor is depicted by a small square. If the cursor is placed on a position holding an SMS, the date and time, communication partner and direction of communication (incoming or outgoing) is shown in the bars at the top and bottom. If the cursor is placed on an empty position, date and timeframe (morning, afternoon or evening) are displayed. The application always starts in the upper left screen (morning of present day).

## 4 Experimental evaluation

### 4.1 Hypotheses

We conducted an experiment to explore the usability differences between the classical, list and graphical interface. Following our lines of thought in designing the novel interfaces we formulated two hypotheses:

- H1: Both novel interfaces are equal or better than the existing classical interface because of enhanced functionality and better data representation.
- H2: The list interface is superior to the graphical interface due to its reduced complexity and its focus on social relations.

### 4.2 Design and procedure

The experiment was comprised of six different scenarios that were grouped into two task types: group contact and single contact. For the first task group contact, scenarios were constructed as a conversation happening within a short period of time (e.g. one afternoon) with two communication partners. Scenarios for the second task (single contact) consisted of comprehending a conversation over a long period of time with only a single communication partner. Messages were scattered over the whole 18 day time period (e.g. finding a birthday present for a friend). To be able to show that differences in the measures are due to the experimental design it is important to control other sources who could have an impact on the dependant variables. Therefore, all participants worked on the same text message base.

For training purposes, participants were asked to solve the first scenario with the classical interface. Scenarios 2-5 were solved using the list interface and the graphical one respectively. Scenario 6 served as the baseline with participants using the classical interface for solving a more simple problem. Each scenario was built in the following way: participants received contextual and chronological information. Afterwards, they were asked to solve two problems related to the scenario. Finally, test persons answered a questionnaire after completing each scenario. As for materials, two Nokia 6230 mobile phones were used for the study. As for time, most experimental sessions lasted about 40 minutes.

### 4.3 Measures

Time to complete a given task and retrieval success were used as objective measures, while a questionnaire served as a subjective measure in the experiment. The data collected from the phones following the experiments was the main data base for the objective assessment. We stored the following four user events in the log file of the application: beginning and end of the scenario, opening and clos-

ing times of a message. The retrieval success data was derived from automatically counting how many messages were opened in total and how many of those were related to the given task. After they had performed the respective scenario, participants were asked to complete a questionnaire containing five items typically used in usability testing: perceived ease of use, perceived usefulness, enjoyment, attitude toward behavior and intention to use. The items draw on the work of Venkatesh, Morris and Davis and incorporate the Technology Acceptance Model which is supplemented by intrinsic motivation [13, p. 428]. The answer categories consisted of a 7-point-scale with labels at both endings. Participants had to rate in how far they felt given statements in the questionnaire are applicable to the interface at hand.

### 4.4 Sample

The sample is comprised of 34 first year university students who participated in early 2005 in the experiment. The majority of test persons report sending SMS messages on a daily basis or at least once a week (82%). They receive a median of 66 text messages per month, and they report sending the same number of messages to other contacts. A quarter of the participants reopen messages stored on their phones on a weekly basis (30% on a monthly basis). The median number of presently stored incoming messages is 40. Asked for experience in playing mobile phone games (e.g. Snake or Tetris), the majority of participants stated to play never (n=22) or to play less than once a month (n=8). However, nearly every second participant reported to be a skilled desktop computer game player (n=16).

## 5 Results

### 5.1 Methodology

The design comprises two factors (task and design) with two levels each. The task factor differentiates between the single contact and the group contact scenarios. The factor design consists of the graphical user interface and the list interface. This sums up to four independent variables. To test for main and interaction effects we calculated one model for retrieval success, one for the time spent to complete a task and one for the usability questionnaire. The items of the usability questionnaire were included as different measures in one model.

### 5.2 Objective results: Time and retrieval success

The mean completion times for all five scenarios of the experiment are displayed in the column "time" of table 1. Results show clearly that the fastest way to solve a scenario was the list interface for both tasks.

Table 1. Objective results (mean values, smaller time spans and larger percentages represent better results)

Task type	Scenario & interface	Time sec.	Rec. in %	Prec. in %	F1 in %
Group contact	2 graphical	115	66	90	75
	3 list	98	64	96	76
Single contact	4 graphical	101	72	90	80
	5 list	53	99	98	98
Baseline	6 classical	80	67	91	58

The retrieval success is presented by stating the recall, precision and F1 values using means. Precision is defined the ratio of scenario-related messages read compared to all messages read in the scenario. Recall represents the relation of messages belonging to the scenario compared to the total amount of relevant messages found by the participants (eight messages for scenarios 2-5, four messages for the last scenario). Thus, precision represents a good measure how many false messages were read. Typically, recall and precision are combined in the F1 measure, describing the overall performance ( $F1 = 2rp/(r+p)$ ,  $r = recall$ ,  $p = precision$ ). Concerning retrieval success, the main effects for task and design and the interaction effect were significant ( $F_{task;1;33} = 58.8$ ;  $F_{design;1;33} = 28.6$ ;  $F_{task*design;1;33} = 16.4$ ;  $p_{all} = .000$ ). In addition, all effects concerning time spent to complete a task were significant as well ( $F_{task;1;33} = 20.0$ ;  $F_{design;1;33} = 34.5$ ;  $F_{task*design;1;33} = 6.5$ ;  $p_{task} = .000$ ;  $p_{design} = .000$ ;  $p_{task*design} = .016$ ).

Overall, the list interface performed better than the graphical interface and the single contact scenarios performed better than the group contact scenarios. If we look at the interaction effects, we see a highly better performance with the list interface than with the graphical interface for the single contact scenarios, but only a minor difference within the group contact.

### 5.3 Subjective results: Usability questionnaire

Concerning the usability, the interface designs had a significant effect ( $F_{design;5;29} = 4.3$ ;  $p = .005$ ) but the tasks had not ( $F_{task;5;29} = 2.4$ ;  $p = .062$ ). The interaction effect was also highly significant ( $F_{task*design;5;29} = 6.2$ ;  $p = .001$ ). Figure 3 shows the results and demonstrates the interaction effect. The graphical design produced significantly better results with the group contact tasks than with the single contact tasks. This effect was the other way round for the list design: single contact tasks were significantly better evaluated than group contact tasks. Similarly to before, the categorized list design got better scores than the graphical design and the single contact scenarios were better than the group contact ones.

Interestingly, an interaction effect exists for the item

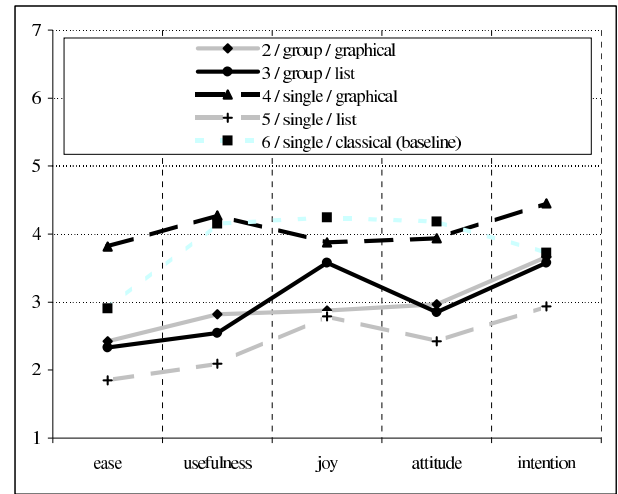


Figure 3. Subjective results (mean values, 1 = best, 7 = worst, captions represent the number of the scenario, the task type and the interface design)

enjoyment. A clear deviation of means for the list interface is visible in figure 3. Participants enjoyed the graphical interface much more than the list interface to solve the group contact scenarios, but took more pleasure in using the list interface over the graphical one while working on single contact scenarios.

## 6 Discussion

As for hypothesis H1, both novel interfaces received better objective and subjective evaluation than the existing classical interface. Even though the overall time for solving scenario six with the classical interface was shorter than the graphical one, it should be mentioned that only four instead of eight messages had to be found. The same argumentation applies to the subjective results: The classical baseline interface might have received even worse results if all eight messages had to be read.

Hypothesis H2 can be accepted as well. The list interface is superior to the graphical interface, except for the variable intrinsic motivation. The graphical interface is more enjoyable. This interaction effect might also be interpreted through the high rate of skilled computer game players in our sample.

Interestingly, it can be clearly seen from our findings presented in table 1 and figure 3 that the novel interfaces differ significantly more for the single contact scenarios than for the group contact scenarios. As it is reasonable to state that SMS conversations generally take place in short time frames, the broader differences for the single contact scenarios might not be that relevant.

Furthermore, the twofold experimental setup might have had an influence on how much better a new interface is in comparison to the classical one. During the planning phase of the interface design we discussed a combination

of both interfaces where users of the graphical interface can browse to the preceding or following message from the same communication partner (similar to the list interface). We finally decided to use a two-interface study in order to learn more about the differences between the tasks. Combining the features of both interfaces might still yield an even better evaluation of the novel interfaces.

Finally, the design of both novel interfaces had to be slightly simplified. The list interface in its current form, for example, would lead to problems with unread text messages. If two or more unread SMS messages exist in the message store, users can hardly access them. This could be solved by allowing users to browse such messages by pressing the buttons left or right. The graphical interface, on the other hand, allows for displaying only one message per hour in its current form. This problem could be solved by reducing the size of the messages displayed or by applying zooming techniques.

## 7 Conclusion

Both novel interfaces used in our experiment yielded better results than the classical one. Some details of our findings might have implications for future mobile user interface development.

Today's mobile applications seem to be built around technologies (e.g. incoming and outgoing text messages are separated). Novel interfaces could be built around social contacts. Many participants stated that they especially liked sorting messages by social contacts. This supports Tamminen et al., who formulated that applications should adapt to social contexts of users [11, p. 138]. The graphical interface might benefit from social awareness as well. The interface could be enhanced by assigning unique icons to well-known contacts. Depending on display resolutions this could be everything ranging from small figures to large photographs.

Commercially, ring tones, sound files and games are a profitable business. The graphical interface might serve as a kind of mobile game. It could be played while traveling or waiting. Moreover, new technical developments might allow for incorporating SMS messages, e-mails and other associated documents. Novel interfaces (both textual and graphical) could facilitate to find the right information at the right time.

## References

- [1] Netsize Group. *The netsize Guide 2005*. available online: <http://www.netsize.com>, 2005.
- [2] Sun Microsystems. *Java 2 Micro Edition Specifications*. available online: <http://www.sun.com/software/communitysource/j2me/>.
- [3] A. Marcus. Babyface design for mobile devices and the web. In *Proceedings, Vol. 2, Human Computer Interface International (HCII)*, pages 514–518. New Orleans, August 2001.
- [4] M. Masoodian and N. Lane. An empirical study of textual and graphical travel itinerary visualization using mobile phones. In *Proceedings of the Fourth Australian user interface conference on User interfaces(AUIC)*, pages 11–18. Australian Computer Society, Inc., 2003.
- [5] M. Hearst. *User Interfaces and Visualization*. In: *Modern Information Retrieval*, Eds: Ricardo Baeza-Yates, Berthier Ribeiro-Neto, Chapter 10, Addison Wesley, 1999.
- [6] J. Donath. A semantic approach to visualizing on-line conversations. *Communications of the ACM*, 45(4):45–49, April 2002.
- [7] C. Fluit, M. Sabou, and F. van Harmelen. *Ontology-based Information Visualization*. In: *Visualizing the Semantic Web*, Eds: Vladimir Geroimenko and Chaomei Chen, 36-48, Springer, 2002.
- [8] K. Börner. *Visual Interfaces for Semantic Information Retrieval and Browsing*. In: *Visualizing the Semantic Web*, Eds: Vladimir Geroimenko and Chaomei Chen, 99-115, Springer, 2002.
- [9] N. Döring. Kommunikative Funktionen von Kurzmitteilungen (SMS). *Zeitschrift für Medienpsychologie*, 14(3):118–128, English summary available online: <http://www.receiver.vodafone.com/06/articles/index05.html>, 2002.
- [10] P. E. Pedersen. *The adoption of text messaging services among Norwegian teens: development and test of an extended adoption model*. SNF-Report No. 23/02, Samfunns- Og Naeringslivsforskning As, 2002.
- [11] S. Tamminen, A. Oulasvirta, K. Toiskallio, and A. Kankainen. Understanding mobile contexts. *Personal Ubiquitous Computing*, 8(2):135–143, 2004.
- [12] T. Rodden, K. Cheverst, N. Davies, and A. Dix. Exploiting context in hci design for mobile systems. In *Proceedings of the First Workshop on Human Computer Interaction for Mobile Devices*, pages 12–17. University of Glasgow, May 1998.
- [13] V. Venkatesh, M. Morris, G. Davis, and F. Davis. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478, 2003.