

# Challenges in Early-Phase Usability Evaluation – A Case-Study of an Adaptive Telecommunicative Agent in Nokia Phones

*Juha Syri, Heikki Hyötyniemi and Antti Salovaara*

*Nokia Research Center, P.O. Box 407, 00045 Nokia Group, Finland  
juha.syri@nokia.com, heikki.hyotyniemi@hut.fi, antti.salovaara@hut.fi*

## Abstract

Even though the field of usability has seen a notable growth during the past decade and usability-related issues are generally regarded as a main factor in product acceptability and marketing, usability-evaluations in the earliest phases of product-development aren't very common. In this study, accomplished in Autumn 2000, we assessed usability factors regarding an adaptive message agent for mobile phones, focusing on features that had not yet been realised in the agent. The study questions were the following. What would the end-user value of adaptation be in the message agent and what expectations have the end-users got regarding their interaction with the agent?

The study was organised as three group interviews for proof-users of the then version of the message agent. This approach was selected, because the data was relatively rapidly analysable. All of the interviewees (N=16) worked for the company Nokia, Finland. Three out of four of them had a university-level education and nearly all of them were technologically advanced and active and had a technologically favourable attitude measured with a pre-questionnaire. The groups didn't differ from each other in these respects. The semistructured thematic interviews covered the study questions and were presented to the subjects in the form of scenarios that described the presumed product-development. The interview protocols were videotaped and the discussions analysed using the method of qualitative content analysis.

The main results were that the users were quite unified in their skepticism against adaptivity of the message agent. The reason for this, stated by some respondents, was most likely the personal and delicate nature of mobile message usage. Any personal adaptivity would rapidly lose its value if non-optimal function would occur. The users would only want to use the adapted version of the message agent if the interaction were in their control and non-automated. Still, even this would be a good solution only if the interaction didn't demand much effort from the users.

Even though the most likely motivation for the respondents' skepticism was the delicacy and the private nature of message use, the study setting was a compromise between various goals and thus some other contributors can be speculated about. Firstly, there is some pressure toward social desirability in a group interview situation. More diverse opinions could have emerged in a single-interview setting. Secondly, because of information-security reasons, all the interviewees had to be selected within Nokia. There wasn't much divergence in their attitudes even though this was deliberately sought with the pre-questionnaire. The respondents were highly familiar with technology, which may have rendered them more realistic about

what can be accomplished with technology. Their skepticism might finally result from the fact that they had not tested the service type they were questioned about. This study suggests that research methodology should be further developed for early-phase usability evaluations. Still, it is concluded that even approximate results in the first phases of product-development are valuable.

**Key words:** usability, mobile phone, message, agent, study methodology

## 1. Introduction

During the past few decades, there have been certain transformations in the field of product development. Especially in companies with huge volumes, product development has become a more open process. Potential customers of the product, often called end users, are more and more taken into account before the actual product is launched on the market. Considerations of who the end users will be, what needs they have and what their use context is have entered the process of product development during the last few decades more and more (Preece (ed.), 1993). This has certain justifications such as cost efficiency. By taking the users' preferences into consideration in the product development, producing an unsuitable product becomes less likely and so product development costs are diminished (see Bias et Mayhew (ed.), 1994).

The development has taken place under many, somewhat different frameworks, which all have their advocates. One framework is for example the so-called contextual design (Bayer et Holtzblatt, 1998) where the means with which the product development is started are analyzing market trends and other relevant products, their respective market niches, the users of the products and what the users would hope for the new, assumedly better product, that is, the product in its overall context. The aspired result would be a more usable product than otherwise would be the case: a product the users of which perform faster and with fewer errors and are more satisfied with the product (Committee Draft International Standard ISO/DIS 9241-9, 1998).

Still, there are certain problems when including the end users in the product developmental cycle. Especially in the very earliest phases of product development, the idea of the aspired product isn't often very precise. No working prototype might be available, sometimes only a vague demo of it. Usability evaluations of www pages get performed with paper models of the pages, models that are far from the actual product (Grady, 2000).

In the present study, the focus was a usability assessment of a mobile telecommunication agent to be implemented in Nokia mobile phones. A usability evaluation was performed in the early product development phase. In order to find guidelines for the most reasonable development paths, different functionalities of the assumed service were presented for the likely target group of the service and opinions about its usage and usability were enquired with the aid of use scenarios. More specifically, we wanted to know whether the end users of the mobile telecommunication agent (for agents, see <http://www.agentlink.org/>) perceived the value of the new feature and whether the planned technical improvement of the service would result in greater usability. This presentation describes the study, its findings and implications as well as discusses challenges in this type of usability evaluation.

## 2. Study

The background of the study was the fact that presently, there are a lot of applications that deliver messages to a mobile device such as a cellular phone (<http://www.emsd.org/>, <http://www.iobox.fi/>, [http://www.skytel.com/products/services\\_mobilemail.htm](http://www.skytel.com/products/services_mobilemail.htm)). Such applications offer certain benefits: mobile devices are more common among the population

than computers in many countries and thus are more often available (for the situation in Finland, see Kopomaa, 2000; Nurmela et al., 2000). They can also be used in a wider range of situations – for example in a bus it is presently easier to get connected in the web on a phone than on a computer.

Currently, the applications for gaining www contents such as messages on a phone take care of the delivery of contents into the mobile device, but the information delivered isn't processed contentually. Still, the content that has to be delivered from the computer application to the mobile device doesn't often fit into the memory of the mobile device. For the purpose of message delivery on the phone, the amount of contents that can be presented via the present channels, especially the SMS (Short Message Service) channel, isn't very notable: maximum 3200 (20x160) characters in most present Nokia phones, which typically leads part of the contents being mechanically left out (see for example <http://www.iobox.fi/>). For a comparison, this study report has 7940 characters until now.

For this reason, an improvement over the current means to deliver messages into a mobile device was planned. The idea of the service was that the contents to be delivered are analysed and the most essential information of it only given to the end user, if the space limitations don't allow everything to be shown. An example of such a situation would be a user having 30 new mails in the mailbox. When receiving the mail on the mobile device, only the most important ones would be delivered, based on the contentual analysis. Already in the planning stage of the service, these improvement ideas were usability evaluated during the Autumn 2001. The study was conducted as follows.

## **2.1 Subjects**

The subjects of the study belonged all to the target end-user group of the service where the improvement was aimed, that is, they already had in use the then version of the service. The user pool of approximately 200 users was contacted per email where their voluntariness of participating in the study was enquired. Of the potential 200 users a sample of 19 users was formed. Finally, due to schedule reasons, 3 respondents couldn't participate. In the study participated sixteen respondents, who all worked within Nokia Research Center, Helsinki, Finland.

The subjects were all Finnish except for one South African. 5 of the subjects were female, 11 male. Most of the subjects had a university-level degree, the second most common level of education were university students. The subjects mainly worked within research and development, one of the subjects was a secretary.

Because the study was an interview study, in which not the objective performance but opinions about future performance were enquired, the participants' motivations and opinions about communication technology were also measured. These more common level opinions about communication technology as well as the respondents' command of it were seen as a potential factor contributing to the results (Bereiter et Scardamalia, 1993, Hakkarainen et al., 2000).

The subjects were on the whole very technology oriented: the subjects' self-estimation on a ten-point interval scale of mastery of communication technology was 7.6, one being the lowest possible value, ten the highest. The subjects were so divided into three interview groups that no significant differences between the three groups existed. The grouped averages of the three groups were 7.8, 7.0 and 8.4. So on the average, the subjects viewed themselves as computer-savvy.

As for their attitudes toward communication technology, the situation wasn't quite as favourable. The average of the respondents' attitudes toward technology and its development was 5.4 on a 10-point scale, 1 meaning a non-favourable and 10 a favourable attitude. In the interview groups, the averages were 6.0, 5.6 and 4.5, so all the three groups had respondents that reported notable mastery of communicative technology but didn't feel very enthusiastic about it and its development. This fact will be discussed in the concluding section of this presentation.

## **2.2 Study procedure**

The first stage of the study was recruiting the voluntary subjects. They were promised in the first contact letter that all data gathered in the study would be analysed with respect to respondents' privacy and that the data was only gathered for research and development purposes of the company Nokia. The subjects were also told that they could interrupt the study whenever they wanted.

The subjects first filled in a preliminary questionnaire where information about them such as the age was collected. Based on this, the subjects were divided into three groups of 5, 6 and 5 participants. The aim was to have the groups as similar one with another as possible and on the other hand, have so small groups that everyone could have the chance to participate in the discussions.

In the next stage, the study was conducted as three semistructured thematic group interviews. The interviews lasted about 1,5 hours in each group. In the interviews, there were two researchers present in the same room as the subjects. One of the researchers lead the discussion and the other one followed it and took notes. During one group's interview, there were additional two product development experts following the discussion from another room behind a video camera – television monitor connection.

During the discussions, the principal interviewer presented two use scenarios for the respondents. In the scenarios, some benefits of the improved version were described in a real-life situation. After the scenarios, some issues related to the improvements were discussed. These were among others security and privacy of interaction and user control in the functioning of the application. Also, the respondents' present usage of the service was discussed. The researchers didn't deliberately try to lead the conversation on any particular detailed level issue. Rather, the principal researcher presented each main issue and the interviewees could present their opinions in the order and breadth they themselves wanted.

## **2.3 Measurement and analysis**

The study used mostly qualitative analysis methods. The prequestionnaire had open-ended, qualitative items such as respondents' education and occupation and two quantitative scales for users' command of communication technology and their opinions about it and its development. The scales were specifically constructed for the purpose of this study and they were interval scales between 1 and 1. The first scale had four items and the second had seven. An excerpt of the questionnaire items is given below in the Table 1.

Please indicate below the alternative that closest resembles You.	
	I highly disagree      I highly agree
I use communication technology very widely	1 2 3 4 5 6 7 8 9 10
A great part of my communication with other people takes place via email or mobile phone	1 2 3 4 5 6 7 8 9 10
The constant transformation of technology causes me extra labour	1 2 3 4 5 6 7 8 9 10

*Table 1. An excerpt of the prequestionnaire items.*

In the analysis stage, the values of each scale were calculated as an average of the scale items. The negative statements such as the last one in the Table 1 were converted in this stage in order that a greater sum value in the scales would uniformly mean a more skilled and a more computer-favourable respondent.

The interview protocols were recorded as the assistant researcher's notes and on video. After that the interview data was segmented (divided in brief fragments) and categorized according to the principles of qualitative content analysis (Chi, 1997) and the frequencies of each response category calculated.

### **3. Study results and their implications**

The main results of the study were that the users didn't very much like the development ideas that were thought to improve the application. The users would rather do without any improvements over the existing form of the application. Even though the space limitations in mobile delivery of messages were notable in the users' opinion, they would still want to receive all the mail contents as such and sooner delete the less important contents manually. They preferred a very high degree of user control to any system control. A stated reason for this was possible cases of non-optimal function in the application. The labour of deleting unnecessary message items was regarded as smaller than the possible harm of the application analyzing the message contents and wrongly estimating some of it as less important and not showing it to the end-user. The users were also quite contented with the then version of the application and hadn't got any urge to have it improved. In the next section, the results are presented in more detail.

#### **3.1.1 Then version of application, users' satisfaction with it**

The users were on the whole quite contented with the service. About three out of four respondents said the service was useful and worked well for their purposes. The two commonest opinions about good in the service were its availability: the mobile message delivery service can be used in a great variety of situations. Of all comments on the good sides, the availability was mentioned in 75%. The second commonest opinion about the good sides in the service was its use discreteness: the use is disguised so that no other people than the end-user can see the messages. The use discreteness was mentioned in 25% of all positive comments regarding the good sides of the service.

Even if on average the users were satisfied with the service, some negative statements about it were also mentioned. The commonest bad side of the service was the difficulties of interaction with the service. The commands for using the service were difficult to remember. 33% of all cases where users said something negative about the service were such statements. The next two commonest categories were unoptimal security solutions in the service and the fact that with long messages, the service didn't function optimally. 24% of all negative statements were in these categories. The fourth negative answer category was slowness of interaction, represented by 19% of all negative user statements.

### **3.1.2 Developmental ideas of showing only important mail items, users' satisfaction with them**

It was planned that in the application, a user's longer-term usage would be monitored and the mail contents analysed across many use instances and a longer-term use profile created on this basis. The arriving new messages would have been compared with the importance criteria thus formed and the most likely important items preferred in mobile delivery. Still, the users' opinions about important messages and forming criteria for them didn't quite match this. 75% of all statements about importance criteria represented situation-boundedness of the criteria: it cannot be said across different message reading instances what the criteria of important messages are. Only 25% of the statements concerning importance criteria indicated that some stable criteria could be given. Typically mails arrived from one's spouse or children were regarded as important.

It was also planned that within any single mail, the most important information could be detected. The users were quite worried about getting a non-representative summary. The users would sooner read through the original mail all the time than have it summarized and sometimes get an unoptimal summary. The subjects didn't want any important mails to be summarized. They would have let a less important mail to be summarized more easily. Some respondents mentioned a few criteria of important information in any given mail. They were work-related specific vocabulary such as 'a meeting'. Also in work-related mails, important items within one mail could be recognized from a language change: in an otherwise Finnish mail, any English words would more likely be important, signifying project names and such (the users typically used English for project names and such).

## **3.2 Result implications**

In general it can be said that the users weren't very favourable about the developmental ideas of only showing the more important contents of mail in mobile situations. They would sooner go through the manual labour of deleting a less important mail than let a device do such filtering automatically.

It has to be remembered that the service in question was a very personal and intimate service. The application delivers the user's own mail on a mobile device. It isn't very surprising that in such a situation, the users wouldn't like the idea of having a separate analysator in between deciding for them, what information could be important and what not. Other researchers have discussed the same issue, which has to do about user comprehending and trusting how a given device works (see for example Höök, 2000).

Still, the methodology used in the study partly contributed to the results found. The study used a group interview setting. In the video protocol analysis it was seen that the quieter respondents quite often rephrased the same answers that somebody else had already said. This had the effect of diminishing variance in the responses and lead to the uniformity in users' skepticism toward the improvement ideas. Such social desirability, according to which the

respondents adjust their answers toward a supposed norm is a well-known phenomenon and exists in interview and questionnaire formed studies (Nancarrow et Brace, 2000). In a single user interview setting such social desirability between the participants would not have existed, but the effort needed to conduct the study would have been greater. It was also probable that the users didn't hope for the improved version because they were quite happy with the existing one.

The subjects were selected from the employees of Nokia and were premeasured on their technical command. All the subjects were very familiar with technology and this can have contributed to the results. They could easily see the potential flaws of the improved application. Most respondents had used and developed many information technology applications in their work and had had experience about the problems involved in this.

This calls for certain improvements in the study methodology. First, the likely target group of the users should have been better represented in the respondent sample. Even if in the then stage of the product development, the service was in use only within the company Nokia, the final target group was larger and also people outside the company should have been enquired. Also, the problem of the users not testing the actual improved version was present. A richer picture of users' opinions would have been possible to get if we had had a working prototype for the users to test. The minor reluctance of the users toward the improvements can be reflected in the following light: very often consumers' needs aren't present to start with but a new technology adapts its environment and creates novel needs in users. This idea of the pervasiveness of technology and its effect on the whole system where it exists has also been discussed in the literature (Nardi et O'Day, 1999).

The focus of this presentation was in demonstrating as a case study some challenges that this kind of research methodology contains. There are a lot of pay-offs between for example the cost of the study and the time that can be allocated in it. In our opinion, every study method has some advantages and some drawbacks, and researchers have to be aware of these. As a conclusion it can be said that the results obtained in this study can be used as a rough guideline for eliminating the less successful possibilities of development, as is often the case in user evaluations within product development (see Grady, 2000).

## References

Bayer, H. et Karen Holtzblatt. (1998). *Contextual design. Defining customer centered systems*. Morgan Kaufmann Publishers, Inc.: San Fransisco, CA.

Bereiter, Carl et Marlene Scardamalia. (1993) *Surpassing ourselves: an inquiry into the nature and implications of expertise*. Chicago, IL: Open Court.

Bias, Randolph G. et Deborah J. Mayhew (ed.). (1994). *Cost-Justifying Usability*. Academic Press Limited: London.

Chi, Michelene T. H. *Quantifying Qualitative Analysis of Verbal Data: A Practical Guide*. (1997). The Journal of the Learning Sciences. Vol. 6, No. 3, pp. 271-315. Lawrence Erlbaum Associates, Inc.

Grady, Helen M. (2000). *Web Site Design: A Case Study in Usability Testing Using Paper Prototypes*. In IPCC/SIGDOC 2000. Proceedings Technology & Teamwork. Pp. 39-45. Department of Tech. Commun., Mercer University: Macon, GA, USA.

Hakkarainen, Kai, Liisa Ilomäki, Lasse Lipponen, Hanni Muukkonen, Marjaana Rahikainen, Taneli Tuominen, Minna Lakkala et Erno Lehtinen. *Students' skills and practices of using*

*ICT: results of a national assessment in Finland.* (2000). Computers & Education No 34, pp. 103-117. Elsevier Science Limited.

<http://www.agentlink.org/>. A collection of web pages on intelligent computer agents.

<http://www.emsd.org/> A service where mail contents are delivered on the mobile phone via the SMS channel.

<http://www.iobox.fi/>. A service where mail contents are delivered on the mobile phone via the SMS channel.

[http://www.skytel.com/products/services\\_mobilemail.htm](http://www.skytel.com/products/services_mobilemail.htm)). A service where mail contents are delivered on the mobile phone via the SMS channel.

Höök, Kristina. *Steps to take before IUIs become real.* Interacting with Computers. (2000). Vol. 12, No. 4, pp. 409-426.

Kopomaa, Timo. (2000). *The City in Your Pocket. Birth of the Mobile Information Society.* Gaudeamus: Helsinki.

Nancarrow, Clive et Ian Brace. *Saying "the right thing": Coping with Social Desirability Bias in Marketing Research.* (2000). Bristol Business School Teaching and Research Review. No. 3. MCB University Press: England.

Nardi, Bonnie et Vicki L. O'Day. (1999) *Information ecologies: Using technology with heart.* MIT Press: USA.

Nurmela, Juha, Risto Heinonen, Pauli Ollila et Vesa Virtanen. (2000). *Mobile Phones and Computer as Parts of Everyday Life in Finland.* Reviews 2000:5 of Statistics Finland: Helsinki.

Preece, Jenny (ed.). (1994). *Human-Computer Interaction.* Addison-Wesley Publishing Company: Wokingham, England.