

Future Telecommunications: Exploring actual use

Lynne Baillie

FTW, Donau-City-Straße, Vienna, Austria

baillie@ftw.at

Abstract: The paper presents the initial findings of an exploratory study on the use of a multimodal mobile phone application in a real life setting and in a more traditional lab setting. Some established methods for evaluation were used, however new evaluation tools were also used in an aim to discover more about the users experience when using the application. This is a difficult area in which to undertake evaluations and one which offer opportunities and challenges to designers and evaluators, some of the challenges that we faced are discussed in this paper. The paper concludes with a short exploration of possible future avenues of research.

Keywords: interaction, affordance, multimodal, mobile, context, HCI,

1 Introduction

Does 'where we are' have an effect on the modality we choose? The initial findings from our study would indicate that it does. In this paper we briefly describe the prototype multimodal application that we built. However the focus of the paper is not on the application but on the ways in which we tried to evaluate it. The evaluation took place in Vienna, Austria and involved twelve users who were asked to undertake eight tasks. Four tasks were undertaken in a laboratory setting and four in a more realistic setting. The data was analyzed using grounded theory and ethnographic software. The results show that an investigation of this nature can show how use differs depending on 'where' the user is. This made us realize that new methods need to be developed to give directions to researchers who are hoping to design applications for multimodal mobile use.

The application

The application is a multimodal route finder. The user enters an address using either voice or text. The application then locates the address for them. A map is then delivered to their device showing their current location and how to get to the address from there.

2 Background

One of the main obstacles to undertaking an evaluation of a multimodal mobile application in a real life setting was the lack of appropriate methods and tools for undertaking an evaluation. There is a wealth of reported literature on how to undertake an evaluation of a mobile phone or a multimodal application on a PC, however, nearly all these evaluations took place in the laboratory, under standard user testing conditions. Some ethnographic studies have been undertaken, however, their focus was on the use of the whole phone and not just on one application available on it. It could be argued that none of these methods are ideal for undertaking an evaluation of an early prototype application. One exception to this lack of reported literature is a study by Pirhonen, et al (2002) in which they asked users to undertake tasks on a mobile device while using a stepper. However, one could argue that this is still not a very realistic environment. We decided therefore to use a two pronged approach for our evaluation of the prototype application. In the first part of the evaluation we asked the users to undertake four tasks in the laboratory and in the second part we asked them to undertake four tasks in a more realistic setting. We hoped that by doing this we would learn more about the affordances (Turner and Turner (2002) of the device. They proposed, in their paper, a three-layered model of affordance:

1. The innermost layer: basic usability or ergonomics
2. The middle layer: support for the user in undertaking task
3. The Outer layer: fitness for the purpose.

It was felt that this was a very appropriate model on which to base our evaluation as we wished to assess whether a user could complete a task, whether the application supported them in an efficient way while undertaking the task and whether the interactions were appropriate for the context in which the device would be used.

3 Method

3.1 Methods and Tools

The overall aim of the evaluation was to assess the usability of the prototype application. The device that will be used to access the application is a mobile phone. The evaluation aimed, in the first instance, to assess whether or not the user could obtain the correct routing information. The second part of the study investigated whether the user could undertake the task as easily in a natural setting as they could in a controlled environment. We did this by using and developing an evaluation method proposed by Petrie, et al. (1998) and extended by Brewster (2000, 2002). The aim of this method is to test a mobile device whilst the user uses it in a fairly realistic environment. The method enables the evaluators to keep enough control so that measures (e.g. time taken to complete individual tasks, errors, time taken up by various cognitive activities such as pausing within and between commands and reading or inspecting various parts of the interface display) can be taken to assess performance.

The user was given a questionnaire after each speech task and asked to fill it in. The questionnaire followed the same structure as the ones used in the Paradise method (Walker, et al 1997). This method assists in the evaluation of a systems performance. This method had been developed to aid performance evaluation of spoken dialogue applications on PCs. The method has not been used to assess the use of a mobile device. We hoped that the method would provide us with information about the performance of the application and the performance of the user. The users were also encouraged to talk to the evaluator and discuss freely the problems they were having with the application (Monk et al, 1993) as they used it. This technique has the following benefits according to Dix (1993): the process is less

constrained, the user is encouraged to criticize the system and the evaluator can clarify points of confusion at the time they occur. We could only use this method when the user was using text input, therefore another way of capturing this data was required. It was felt that action scenarios/descriptions may be an efficient way of eliciting the information required.

3.2.1 Action scenario/descriptions

In an effort to elicit more information from the users we asked them to show and describe to us the different menus, icons and so on that had annoyed them during the tasks. We further asked them to tell us how they would structure the task, menus and icons differently. We told them that they could move freely while they did this to aid the description. We hoped that by doing this we would have a richer description of the user's feelings about their interaction with the application.

The user was videoed and the evaluator took additional notes with paper and pencil. The video was analysed using conversational analysis techniques and by the noting of incidents by the evaluators. The users were video-recorded as they worked through the eight tasks given. Video-recording the user has the advantage of capturing very fine-grain data. The evaluator also kept a paper note of special events and circumstances.

The methods used in this evaluation, have been chosen because they reflect the current research knowledge in this area. That is not to say that these methods are the only applicable methods for evaluating a prototype application of this nature only that they were thought to be the most appropriate. However, the Paradise method has not been used to assess the use of a spoken dialogue system in a mobile environment. The results could highlight; that the method is not appropriate for this type of application or that it requires some adjustment. Pirhonen et al (2000) method may also need some adjustment as it was not used to assess a multimodal device. Also our idea of using: Action Scenarios may not be embraced by the users.

4 Evaluation

The number of users used in the study was twelve. Many usability experts believe that twelve is an adequate sample size for a controlled experiment (Dix, 1993). The users were split evenly along

gender lines and reflected the general user population.

The user's percentage preferred walking speed (PPWS) was measured as it has been found in previous studies (Pirhonen et al, 2002; Petrie et al, 1998) that this measurement is very important to the evaluation of a mobile device in context.

The users were given four tasks to complete while static and mobile.

Task 1: The user was asked to choose the modality they would like to find an address.

Task 2: The user was asked to find an address using voice interaction.

Task 3: The user was asked to find an address using text input.

Task 4: The user was asked to choose the modality they would like to find an address.

The evaluation took place at the research centre there were two settings in which the user undertook the tasks:

- 1) In a room free from interruptions and noise (the user was sitting while undertaking the tasks).
- 2) In the research centre building and adjacent to it.

It was hoped that we would discover additional information about the user's interaction with the application by undertaking part of the study in a natural setting. It was also anticipated that we would discover differences between the choices of interaction with the device when moving, as apposed to when the user was static. This environment, being a natural setting, was not completely free from noise or interruptions.

5 Discussion

The users, in the main, completed all the tasks given to them. There was some problem with the voice recognizer taking slightly longer to recognize the commands from a non-native speaker. There was no significant time difference between the tasks which had been undertaken in the laboratory and the ones undertaken in the natural setting. Pirhonen (2002) Petrie et al, (1998) method helped us to understand more about the problems a user had when they undertook a task. If a user had taken longer than average to complete a task we could go back to the video or notes taken during the evaluation to try and understand what was happening at that time. The number of errors made by a user also helped us to understand more clearly the problems they were having with the application.

The analyses of the Paradise questionnaire is still ongoing and as yet has not fed into our redesign. This may be a problem with the method itself as it takes a considerable amount of time to undertake the analyses.

Undertaking part of the evaluation in a natural setting turned out to be more of challenge to the evaluators than to the users. The users felt comfortable wandering through the research center buildings and wandering outside. The evaluators had more problems as they had to carry cameras, notes and so on. We found that the users did use different interaction modes when mobile than when static and could articulate the reason for their choice:

Researcher: Can you say why you chose that mode of interaction?

User: Hm..m it just seemed to me(.2) that when I was walking down the stairs, that I wanted to be able to see fully my surroundings and not be trying to text at the same time as walk.(.2) so I chose to use the voice mode instead. (Excerpt from Evaluations EDSCH.1.4.03)

This was a typical comment made by the users. This was interesting as we had not been sure how the more noisy environment i.e. lifts going up and down, people walking in and out of offices, people going past on bicycles, would affect the users choice of modality. The users were free to wander where they pleased and some wandered around quite a lot and some only a little. Sometimes the users leaned on railings, others decided to walk to the nearest underground station! We found that the mode of interaction was quite dependent on the surrounding terrain.

Our second finding re: natural setting was that in this setting user's became more quickly frustrated with the application, if it did not work (compared to when they were in the quiet laboratory) as quickly as they wanted it to.

Action Scenarios/Descriptions

We found that this was the most useful part of the evaluation in terms of aiding us in our redesign of the application. The users could clearly point out to us the problems with the menu structure. For example, the application structured the input form in the following way:

District, street and house number

Whereas the users wanted the order to be:

Street, house number and district.

The users also pointed out that some landmarks that they were used to seeing in real life would help to orientate themselves when using the maps, such as underground station stops or chemist shops. In the voice recognition part of the application the users wanted to be able to say help after each part of the task rather than wait to the end to say that part of the address was wrong. These and other problems aided our redesign.

6 Conclusion

It can be concluded from the findings that users can take an active part in an evaluation of a multimodal mobile application and can provide valuable insights. That it is problematic to undertake studies in a truly mobile environment as it is difficult for the evaluator to see what the user is doing due to the small screen size. However, it is still worthwhile carrying out studies in the field as it gives valuable insights into possible uses and interactions with the application. The action scenario descriptions aided us in our redesign and proved to be a valuable tool. We hope by describing our evaluation and its early results we stimulate and add to the discussions already taking place about, which methods and tools are appropriate for the evaluation of multimodal mobile applications.

References

- Brewster, S.A., Crossan, A. and Crease, M. (2000). *Automatic volume control for auditory interfaces*. In Volume II Proceedings of BCS HCI 2000 (Sunderland, UK), pp 17-18.
- Brewster, S.A. Overcoming the Lack of Screen Space on Mobile Computers. (2002) *Personal and Ubiquitous Computing*, 6(3), pp 188-205.
- Carroll, J.M. 1996. Becoming Social: Expanding Scenario-based Approaches in HCI. *Behaviour and Information Technology*, 15(4), 266-275.
- Dix, A., Finlay, J., Abowd, G., & Beale, R., (1993). *Human-Computer Interaction*. Wiltshire: Prentice Hall. Downs, R. and Stea, D (1973). Cognitive Representations, pp.79-86, in R.Downs and D.Stea (eds). *Image and Environment*, Chicago: Aldine.
- Monk, A., Wright, P., Haber, J., and Davenport, L (1993). *Improving your Human Computer Interface: A Practical Approach*. Hemel Hempstead: Prentice Hall International UK.
- Petrie, H., Furner, S., & Strothotte, T. (1998) Design Lifecycles and Wearable Computers for Users with Disabilities. Paper presented at HCI Mobile Devices Workshop, Glasgow, 2002.
- Pirhonen, A., Brewster, S., & Holguin, C. (2002) Gestural and Audio Metaphors as a Means of Control for Mobile Devices. Paper presented at CHI letters 2002, pp 291- 298.
- Turner, P & Turner, S (2002) An Affordance-based framework for CVE evaluation. Paper presented at HCI 2002, London, UK, pp.230-251.
- Walker, M.A., Litman, D.J. Kamm, C.A. & Abella, A. (1997) Paradise: A Framework for Evaluating Spoken Dialogue Agents. In *Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics*, ACL 97.