Evaluation of a Multimodal Dialogue System for Small-screen Devices

Holmer Hemsen

Natural Interactive Systems Laboratory (NISLab)
University of Southern Denmark
Campusvej 55, Odense, Denmark
hemsen@nis.sdu.dk

Abstract

The small screen size of mobile phone devices introduces usability problems for the development of graphical user interfaces and applications beyond pure telephony applications. The question that arises is, in which way a multimodal application including speech interaction can circumvent these restrictions and usability problems. In this article problems during the development of the user interface for a prototype application in the real estate domain are described and design decisions taken are discussed. A first user test is described, which shows the acceptance of the system, but also usability problems that still needs to be solved.

1. Introduction

The mobile phone coverage in most European countries has reached 70 percentage (REGTP, 2002). Because of their discrete size and little weight mobile phones are taken along everywhere and have thereby become ubiquitous. Together with the increasing computational power of the devices, the mobile phone is as a consequence a promising device for developing other applications (e.g. games, video viewer, calendar) than those directly related to telephone communication.

Graphical only applications based on the Wireless Application Protocol (WAP) have shown usability problems (Ramsay and Nielsen, 2000; Buchanan et al., 2000). Also according to efficient interaction graphical only applications for small screen devices like mobile phones are problematic, e.g. according to higher interaction times (Jones et al., 2002). Even though recent advances in development of mobile phone technology makes it possible nowadays to implement speech recognition facilities for these devices. Based on the fact that speech is a volatile medium and is not well suited for some tasks (e.g. spatial descriptions), using only speech as interaction medium introduces usability problems as well.

A promising solution for developing usable applications for small screen devices is by enabling multimodal input, cf. (Almeida et al., 2002; Pieraccini et al., 2002)

The integration of speech and graphical interface under the constraint of an extremely small screen space poses the following usability questions:

- Can usability and effectiveness of a spoken dialogue system be enhanced by adding a graphical user interface, even though the GUI screen is small?
- In which way can speech input/output circumvent the restrictions and usability problems of graphical user interfaces for small screen devices?

In order to answer these questions an evaluation environment has been developed as a demonstrator (Hemsen, 2002). The test application that has been chosen is an information system for the real estate domain.

2. The System

The demonstrator uses a commercial telephone-based speech recogniser, enhanced by components for multimodal interaction handling and a graphical user interface that simulates the mobile phone device (see fig. 1).

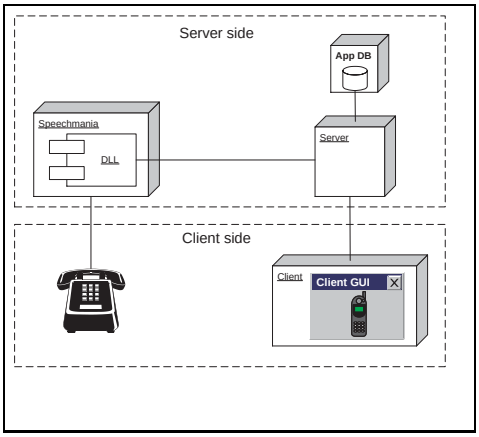

Figure 1: The general architecture of the demonstrator.

Based on the implementation of the demonstrator, an application for information retrieval for the real estate domain has been developed (Hemsen, 2003). The system is used for evaluating different visual modalities well suited for small screen devices, as well as, for evaluating combinations of speech and visual modalities that have shown to be effective under these constraints.

2.1. The application

Starting from analysing how people look for a house in newspaper advertisements, the application design should enable the user to perform similar tasks.

Even though the system is only a demonstrator it is necessary to mention that the test application – an information system for realties using a mobile phone – is not meant as a substitution of existing other information sources as Internet, newspaper, contacting a real estate agent etc. The system is rather a supplement to these ways of getting information on houses for sale in situations where these are not available.

However, similar to searching for a house in a newspaper the system is designed to enable the following tasks:

1. The user specifies a basic set of criteria for the house of choice, e.g. area, price level.
2. The system retrieves from the database information on realties that match these criteria and presents it in form of a list consisting of basic data for each house.
3. A list showing on the screen is the starting point for the user to inspect more detail about the house.
4. Another possibility of inspecting the items is by browsing photos of the realties.
5. Additionally, the user has the possibility to mark certain items by double clicking on them in the list and ask the system to show only the highlighted items.
6. Similar to switch board applications the system includes the functionality that the user can ask the system to call the realtor for the chosen item. As part of the demonstrator this function is only simulated.

The system allows keyboard input, input from a pointing device, combined speech and pointing input, as well as speech only input. Both visual output and combined visual and spoken output is used.

2.2. Challenges for design and usability

Not only the user test itself, but also implementation and technical testing have revealed challenges for the design of the system, as well as problems regarding usability.

2.2.1. Specifying criteria in an effective way

In an ideal system using speech only the user could specify the criteria the house should have, e.g. by saying: ”I would like to have a house in Bolbro, with minimum 100 square meters living area and it should cost no more than 200000DKK. . . ”. However, speech recognisers are not perfect and the difficulty of filtering the extreme variant background noise in mobile situations (Dobler, 2000) will with a high probability lead to speech recognition errors, mostly without knowing where the error is. Recovering from these errors is difficult and inefficient, since each criteria has to be verified. Therefore the system is designed in a way that the user specifies the criteria of choice sequentially. Yet, the system tries to offer the user with the most effective user interface for the given task, which either is pure graphical, pure speech or combined speech and graphics. For example, for specifying the area(s) the house should be situated in, the user can either use speech or choose areas from a list. Specifying a single area is faster using speech, taking into account that the screen of the mobile phone only can present four items out of around twenty and scrolling is needed.

However, for specifying more than one area, the advantage of producing more reliable input than via the speech recogniser overrides the scrolling disadvantage of the list, particularly if error recovery is taken into account.

For specifying the price level a graphical user interface with two text fields (min/max value) is used, instead of using speech only interface which would require a specify value - confirm value - specify value - confirm value sequence. After typing the values and pressing a send button the values are presented on the screen and confirmed by the user using speech.

2.2.2. Navigation and grouping

One of the challenges for the design of interfaces for small screen devices consists of finding effective ways for navigation between the information provided, as well as for grouping related information (cf. e.g. the focus+context approach by (Holmquist, 2000)).

With respect to the real estate information system the user interface design challenge has been to enable the user to quickly choose items of interest and inspecting details of the real estate without loosing to which item the details relate to. After retrieving the data for the houses that satisfy the user specified criteria, a list is presented (fig. 2).

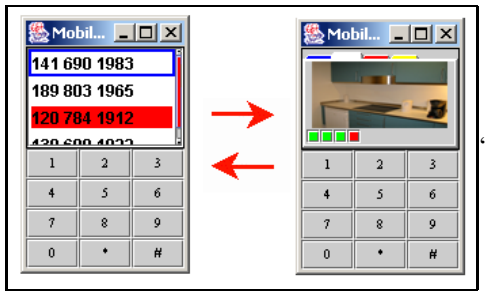

Figure 2: The list view (selected item: framed; marked item: with background color) and related tabbed view.

Each item represents a house that satisfies the user criteria and shows in form of numbers the size of the living area, size of the realty and age of the house. This basic data enables the user to already exclude some of the items. By pointing on the item and asking for e.g. a photo of the house, details of the realty can be inspected. The data for each house are grouped by using a tabbed view (see fig. 3). Browsing photos of the selected houses can be done by directly selecting the next or previous item using speech or in a less efficient way: asking for the list view, selecting the item and asking for a photo of this item.

2.3. Results from the implementation process

The concrete implementation of an arbitrary chosen application revealed several difficulties in the design of the user interface as discussed in section 2.2. The proposed design decision for these challenges by using interfaces using alternative input modalities or by combining speech and visual interface to circumvent the small screen space showed to be adequate. In the following sections, results from the first user test are described, which in addition present difficulties and advantages of the proposed application implementation and user interface design.

3. The Experiments

On the basis of demonstrator application a first user test has been made.

To reveal problems in the dialogue structure and missing output sentence recordings, a preliminary user test with

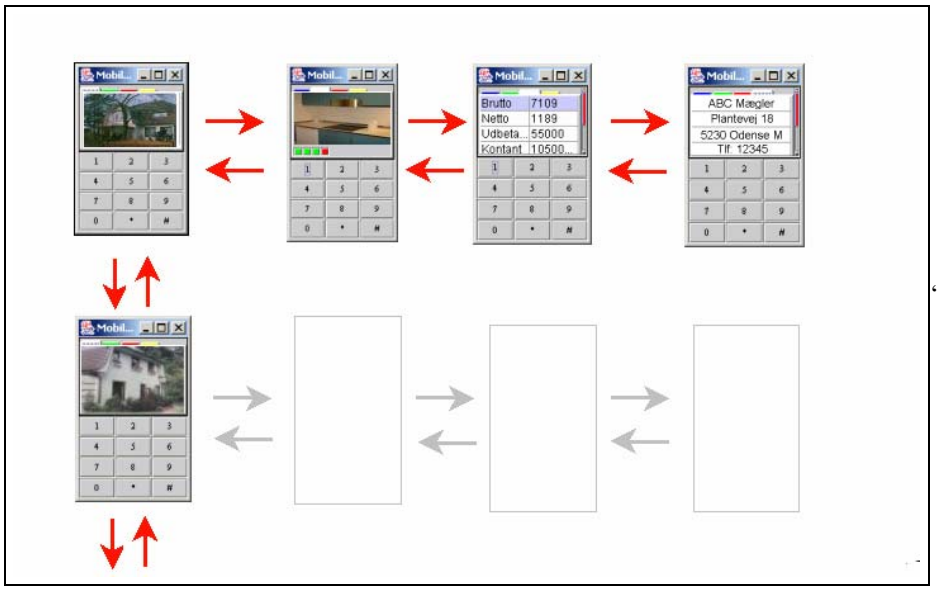

Figure 3: Tabbed view and possible interaction structure.

three colleagues has been made. It also showed that the system is usable by persons unknown to the system, with only basic instructions given and mainly guided by the system itself.

The user test has been made with four persons. The users were recruited at the university and were completely new to the system. Figure 4 shows the setup for the user test.

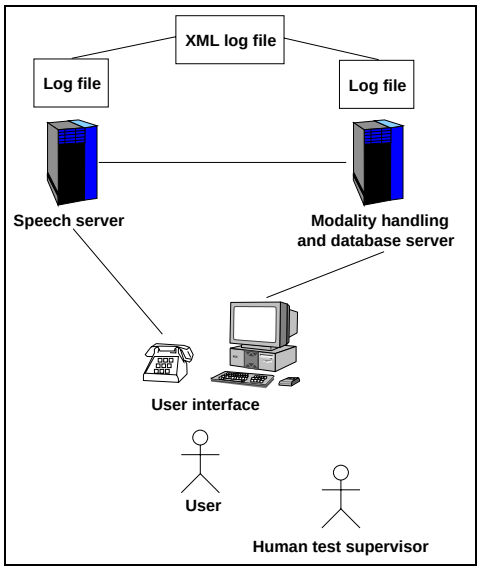

Figure 4: The test setup.

In the test setup a log file output is produced by SpeechMania as well as a log file generated by the modality handling and database server. The logs are transfered by a script to a single XML log file. Additionally SpeechMania records the speech input and produces a file containing a word hypothesis graph for each input.

The user test consisted of three scenarios that the user were asked to execute.

- The first scenario consisted of finding realties in a specific area and within a defined price limit. The user were further asked to inspect pictures of the houses and the rooms and call the real estate agent for a particular item.
- In the second scenario the user has to specify two areas, and the house of choice should have been built before specific year.
- The third scenario specified an upper price limit and asked the user to find the fastest way to inspect fotos of all the houses as well as finding the gross price of a specific house. Additionally the user should explain which data is related to which house.

Before the test the users were told that the speech input is command based and no barge-in is allowed. Further instructions were not given, apart from the scenarios that were handed out.

A questionnaire in which the test users could express their opinion about the system supplemented the test. The questionnaire contained 23 questions distributed according to information about the user, multimodal interaction, usability evaluation and suggestions for further improvements of the system. Nine of the questiones used a 5-point Likert scale, the others were free text or yes/no questions.

Observations made by a human, supervising the test, lead to additionally evaluation results.

4. Results

Even though the ground for analysing the system is too small and presenting the final results is too early at this stage of the evaluation process, the 12 scenarios collected so far have been useful for making some observations.

According to a preliminary analysis of the log files the transaction success of the scenarios given, was 50%.

The second scenario was the less successful (completed only by one user). The explanation of why most of the users

Question	Value
How good do you know the real estate area?	good–neutral
How easy was it to use the system?	easy–neutral
What do you think about the system in its current state?	useful
How easy was it to correct errors?	easy
How easy was it to understand what the system said?	easy
Did the system provided sufficient information about which modality to use?	sufficient
How did it feel to talk to the system?	neutral
What do you think about speech based systems in general	very useful
How was it to interact with the system?	neutral

Table 1: Questionnaire results according to Likert scale questions

failed can be that, the users were asked to look for houses built before a specific year. Since the construction year of the house was not part of the criteria that users could specify, this property only could be observed by looking at the list item. This task lead to confusion, which resulted in two of the users entering the subdialogue for specifying 'properties' which only contains check boxes for properties like garage, central heating, etc. Further system instructions are therefore needed to clearly state which criteria can be specified and which data can afterwards be inspected. Additionally, with respect to the second scenario the following observation could be made. While all of the users in the first scenario used speech for specifying the area, in the second scenario a majority of users chose the list view to select the two areas, using the mouse. All of the test persons had certain difficulties to carry out the first test scenario, but showed better performance in the following scenarios. In fact the third scenario was completed by three users.

With respect to the usability evaluation based on the questionnaires, the following conclusions can be made. Two of the test persons evaluated the system as easy to use, one as neither difficult nor easy and one as difficult. With respect to the user which rated the system negatively, the spoken input was not flexible enough, which can be explained by only few command words currently accepted by the system. Average answers to some of the questions are presented in table 1.

5. Conclusion

In this paper a simulation environment for a multimodal speech centric dialogue system for mobile phones has been presented. The test application for this demonstrator, an information system for the real estate domain, has been described, and design decisions and the choice of modalities used for the different tasks of the application have been explained. Furthermore, the results of a first user test with the system have been discussed. The user test discussed in this article showed the feasibility of the approach of building a multimodal speech centric application and the ability of the users to interact with the system. In general, the positive impression of the system by the users uttered in the questionnaire supports this statement. Missing command word alternatives implemented in the system still makes the system difficult to use and further work is needed to improve the system. So far, however, the current results are encouraging.

6. References

Almeida, L., I. Amdal, N. Beires, M. Boualem, L. Boves, E. den Os, P. Filoche, R. Gomes, J. E. Knudsen, K. Kvale, J. Rugelbak, C. Tallec, and N. Warakagoda, 2002. Implementing and evaluating a multimodal and multilingual touris guide. In J. van Kuppevelt, L. Dybkjær, and N. O. Bernsen (eds.), *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*. Copenhagen, Denmark.

Buchanan, G., S. Farrant, M. Jones, H. Thimbleby, G. Marsden, and M. Pazzani, 2000. Improving mobile internet usability. In *Proceedings of the Tenth International World Wide Web Conference*. Hong Kong.

Dobler, St., 2000. Speech recognition technology for mobile phones. *Ericsson Review*, 3.

Hemsen, H., 2002. A testbed for evaluating multimodal dialogue systems for small screen devices. In *Proceedings of he ISCA Tutorial and Research Workshop Multi-Modal Dialogue in Mobile Environments*. Kloster Irsee, Germany.

Hemsen, H., 2003. Designing a multimodal dialogue system for mobile phones. In P. Paggio, K. Jokinen, and A. Jönsson (eds.), *Proceedings of the 1st Nordic Symposium on Multimodal Communication*, number Report No. 6. September.

Holmquist, L.E., 2000. *Breaking the Screen Barrier*. Ph.D. thesis, Göteborg University, Departement of Informatics, Sweden.

Jones, M., G. Buchanan, and H. Thimbleby, 2002. Sorting out searching on small screen devices. In F. Patern (ed.), *Proceedings of the 4th International Symposium on Mobile HCI*. Pisa, Italy: Springer.

Pieraccini, R., B. Carpenter, E. Woudenberg, S. Caskey, St. Springer, J. Bloom, and M. Phillips, 2002. Multimodal spoken dialog with wireless devices. In L. Dybkjær, E. André, W. Minker, and P. Heisterkamp (eds.), *Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments*. Kloster Irsee, Germany.

Ramsay, M. and J. Nielsen, 2000. WAP Usability Déjà Vu: 1994 All Over Again. Technical report, Nielsen Norman Group.

REGTP, 2002. Marktbeobachtungsdaten der Regulierungsbehörde für Telekommunikation und Post. Technical report, Regulierungsbehörde für Telekommunikation und Post.