Usability in Location-Based Services: Context and Mobile Map Navigation

Kristiina Jokinen

University of Tampere and University of Helsinki, Finland
kristiina.jokinen@helsinki.fi

Abstract. The paper discusses usability and communicative capability of mobile multimodal systems. It reports on the evaluation of one particular interactive multimodal route navigation system and discusses the challenges encountered in this task. The main questions concerned the user's preference of one input mode over the other (speech vs. tactile/graphics input), usefulness of the system in completing the task (route navigation), and user satisfaction (willingness to use the system in the future). The user's expectations and real experience of the system were analysed by comparing the users' assessments before and after the system use. Conclusions concerning system design are drawn and discussed from the perspective of the system's communicative capability, based on the view of the computer as an interactive agent.

Keywords: usability, human-computer interaction, mobile context.

1 Introduction

Evaluation of speech-based interactive systems is needed for forming and refining system design as well as for assessing the impact, usability and effectiveness of the overall system performance. It is important for potential purchasers of commercially available systems, funding agencies who want to develop core technology, and researchers to guide and focus their research. However, the more complex systems we need to evaluate, the more complicated it is to determine appropriate features that determine the user's assessments. Especially for systems that exploit such highly human-related skills like talking, providing information, and presenting visual pictures and maps, evaluation is not a straightforward matter of getting a task done but also involves the user's experience of the system and interaction itself.

The usual way of evaluating interactive systems is to interview users after they have used the system for particular tasks, collect the evaluation forms, and use a weighted function of task-based success and dialogue-based cost measures to assess the system's functioning and suitability. The standard criteria are usefulness (what the system is for), efficiency (how well the system performs), and usability (user satisfaction). The system's performance and the user's subjective view of the usage of the system are taken into account, with the goal to maximize the objective function of usability. For instance, following the Paradise framework [1], the overall objective is to maximize user satisfaction by maximizing task success on the one hand and

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2007, LNCS 4555, pp. 401–410, 2007.

minimizing the costs on the other hand (efficiency and qualitative analysis). On the other hand, [2] points out that interaction design is moving from standard usability concerns towards aspects of interaction that deal with fun, enjoyment, aesthetics and user experiences: ”attractive things work better”. Also various evaluation studies (see below) show that the users seem to tolerate difficulties such as long waiting times and even mere errors, if the system is interesting and the users motivated to use it

Considering practical systems, [3] points out the need to measure the quality of the service. The users also evaluate what is the value of the system for the users in helping them to complete the task, and to the evaluations need to find ways to quantify the value. One way to do this is to find out what is expected of the system by the users, i.e. evaluation does not only concern what is the quality of the system’s performance perceived by the users, but what quality features are desired from the system. The main question focuses on selecting appropriate parameters: what are the quality features that provide reliable information of the system design and the user perception of the system. It is not unproblematic to determine the features, and their direct impact on the system may be difficult to operationalise. Once the features are determined, however, their impact can be quantified by calculating the difference between the user’s expectations and experience before and after the use of the system, and checking how the different features have contributed to the overall change.

In multimodal and ubiquitous computing scene, location-based services are a growing service area. The ubiquitous computing paradigm [4] envisages that pervasive and mobile communication technologies will be deployed through a digital environment which is aware of the presence of mobile and wireless appliances, is sensitive, adaptive and responsive to the users’ needs, habits and emotions, and is also ubiquitously accessible via natural interaction. In this context, multimodality provides realistic interaction means for practical applications: the use of text, speech, pictures, touch, and gaze allow natural input/output modalities for the user to interact with the back-end application. Multimodal systems are considered more natural than unimodal systems for several reasons. Besides bringing flexibility to interaction as the users can choose the modality that best suits to their particular situation, the interaction is also easier and enjoyable, since the users can exploit the interaction strategies they have learnt in human-human communication. Different modalities also bring in different benefits: it is easier to point to an object than refer to it by speaking, and in noisy environments it is useful to combine speech and graphics (see an overview in [5]).

However, their evaluation of such systems is often difficult due to many system components, and also due to problems in assessing system quality, performance and usability in objective terms, and providing clear metrics for system comparison. The system architecture can be distributed in the network, and the users need not sit in front of one computer terminal but operate with the system in a mobile context. Furthermore, the usage situation is challenging because of the novelty of the devices and because of the systems can be used both as traditional information searching tools and as applications enabling communicating with other, human and computer agents. The system should be useful and work well, but also questions concerning the user's expectations of the application as an enjoyable and understanding partner are brought in, together with the user’s experience of the quality of service. In all these cases, an important requirement for usable and enjoyable interaction seems to be the system’s communicative competence [6]: success of interaction measured in terms of dialogue and cooperation principles. As a promising approach to designing interactive systems, it is also critical when assessing trust and reliability of the system in general. It measures the system's capability to provide services and appropriate information that the user finds useful and reliable, in a manner that that takes into account the user's focus of attention: it does not interrupt the user's task goals nor intrude their privacy.

The paper will be structured as follows. Section 2 provides arguments for the "system as a communicating agent" view, and also gives a short overview of the Constructive Dialogue Model as a design framework. Section 3 discusses special aspects of location-based services in relation to the system's understanding of the context in order to provide contextually approriate information for supporting users in their tasks. In Section 4, the route navigation system MUMS and its evaluation results are briefly reported as an example case of a system usability. Finally, Section 5 draws conclusions concerning system design and the system's communicative capability.

2 Complex Systems as Communicating Agents

Due to technological and social development, our interaction with the environment has become more complex, and automatic systems require more sophisticated knowledge management and adaptation to various user needs. The challenge that speech and language technology faces in this context is not so much in producing tools and systems that enable interaction in the first place, but to design and build systems that allow interaction in a natural way: to provide models and concepts that enable experimentation with complex natural language interactive systems, and to test hypothesis for flexible human-computer interaction. Since interaction can take place via speech, text, and signing, special attention is also to be paid to multimodality.

Context-aware communication promises to build environments that understand social context and learn the preferences of individual users, adapting its decisions accordingly. Research projects are concerned e.g. about how to make communication with the house reality, while chatbots and conversational avatars are introduced as the future web surfing method. There are also many initiatives which focus especially on mobile, ubiquitous, and location-aware services, enabling "invisible intelligence" in systems and applications of our everyday life (such as cars, home appliances, factory automation, mobile phones), as well as reinforcing research and development in these areas, and trying to establish successful global standards.

Ubiquitous computing paradigm changes communication, too. Novel aspects for human-human as well as human-computer interaction are brought forward: systems exploit such highly human-related skills like talking, providing information, and presenting visual pictures and maps, and the computer is not only a tool but an agent which the users interact with. For instance, [6] argues in favour of the change of design paradigm for interactive systems, while the studies by [7] show how the users anthropomorphisize, i.e. assign human characteristics to the computer.

Also social interaction changes. The users may want to share their own (digital) data among friends and colleagues, and learn from the other members of the community by navigation, intelligent browsing, and direct interaction. Virtual communities are created where interaction is rapid but not necessarily face to face, and one's identity may be hidden behind different roles. This kind of on-line communication presupposes off-line processing of vast data which consists of various types of digital data such as texts, music, photos, videos, and emails. The organisation of data should be automatic and fast, and allow human intervention in directing and guiding the process according to individual preferences and needs. Interaction management with the application should thus support off-line organisation of the data and its retrieval according to some topical principles which relate to the conversational topic that the speakers are talking about. Recognition of the dynamically changing topic is thus of vital importance, in order for the system to provide rapid information retrieval and appropriate monitoring of and assistance in the conversation.

3 Context Understanding

Rational cooperation can be seen as emerging from the partners' communicative capabilities that maintain interaction on the basis of relevant and truthful information. Communicative behaviour is based on the speakers' observations about the world and on their reasoning, within the dialogue context, about the new information being exchanged in the dialogue contributions. In human-human interaction, this is related to the partners' shared goal, consideration on each other, and mutual trust that they both follow the same communicative principles. In human-computer interaction, cooperation manifests itself in the system properties that enable the user to interact with the system: robustness of data processing and appropriate presentation of the information user [8].

Human-human communication involves smooth coordination of a number of knowledge sources: characteristics of the speakers, topic and focus of the conversation, meaning and frequency of lexical items, facial expressions, prosody, communicative context, physical environment, world knowledge, etc. The following human-human dialogue between a service agent and a customer (Interact corpus, [9]) exemplifies these aspects: the overall dialogue structure is non-deterministic, and the agent's guidance shows flexible and considerate interaction strategy.

A: I'd like to ask about bus connection to Malmi hospital from Herttoniemi metro station – so is there any possibility there to get a bus?
L: Well, there's no direct connection – there's the number 79 that goes to Malmi but it doesn't go to the hospital, it goes to Malmi station
A: Malmi station? oh yes – we've tried that last time and it was awfully difficult
L: well, how about taking the metro and changing at Sörnäinen, or Hakaniemi if that's a more familiar place
A: Well Hakaniemi is more familiar yes
L: Ok, from there you can take the bus 73
A: 73?
L: Yes it leaves Hakaniemi just there where you exit from the metro to the bus stops, next to the market place
A: So it's by the market place that 73 leaves from?
L: Yes
A: And it's not there where the other buses leave from in front of Metallitalo?
L: No, it's there right when you come out from the metro

A: And it goes to the hospital?
L: yes, it has a stop just by the hospital
A: Ok, it must be a better alternative than the bus we took to the station, we didn't know which way to continue and nobody knew anything and we had to take the taxi…
L: what a pity – there would have been the number 69 though. It leaves close to the terminal stop of number 79 and goes to the Malmi hospital.
A: I see, so 79 to the station and then 69?
L: yes
A: Are they on the same stop?
L: well not on the same stop but very close to each other anyway
A: close to each other? Ok, well thank you for your help.
L: thank you, goodbye
A: goodbye

Since the customer had found the route via Malmi station "awful", the agent senses her frustration and introduces a simpler route via Hakaniemi using metro and bus. When the customer returns to the earlier frustrating experience, the agent provides information of this option, too, following the customer's requests. It is this kind of sensitiveness to the partner's needs and attention to their emotional state that characterizes cooperation in human communication.

The dialogue also shows another type of cooperation, related to the context in which the dialogue takes place. The speakers make frequent references to the physical environment (*change in X*, *close to each other*, *Hakaniemi*, *Malmi station*), and the spatial and visual environment directs their interpretation and generation of the language. In other words, language is *grounded* in the communicative context. Grounding is a part of natural communication, exemplified by frequent situational and contextual references, and the whole range of different modalities used in processing and manipulating information (gestures, pointing, nodding, gazing, etc.). Grounding is regarded as one of the most important problems in AI, see [10] for a discussion.

In ubiquitous, location-based services, such as map navigation, way-finding, information search, spatial information is frequently referred to. From the technical point of view, location-aware systems can be seen as responding to the grounding problem in that they enable the environment and the user's location be taken into account in the interaction design. However, the necessary context is usually included by the designers in the general usage scenario, or the system explicitly asks the user for relevant contextual information. These techniques can lead to user frustration, since the assumed context may not be the one the user is in, and the system's repeated requests for the same information sound unnatural and irritating. The systems thus need to interact by making dynamic references to the physical environment and by presenting information in accordance with the information flow of the dialogue and the knowledge of the speakers. Moreover, multimodal pointing, gestures and gazing are used in efficient and accurate reference expressions since purely verbal expressions may become rather clumsy.

Moreover, in mobile situations, the user's full attention is not necessarily directed towards the device and the service that the device provides; rather, it is often divided between the service and some primary task such as meeting people. The applications thus need to be tailored so that they are easily available when the users want them and when they can use them. This requires awareness of the context and the user's emotional state, so as to support their needs but intrude them as little as possible.

4 Evaluation of an Interactive Navigation System

Let us briefly study one location-based service, the multimodal map navigation system MUMS and its evaluation as an example of the usability issues dealing with cooperation in context. The MUMS system [11] is a multimodal route navigation system which aims at providing the user with real-time travel information and mobile navigation in Helsinki. It has been developed in a technological cooperation project among Finnish universities, supported by the Finnish Technology Agency TEKES and several IT-companies. The main goal was to build a robust practical application that would allow the users to use both spoken language commands and pen-pointing gestures as input modalities, and also output information in both speech and graphical modes. The system is based on the Interact-system [9] which aimed at studying methods and techniques for modeling rich natural language based interaction in situations where the interaction had not been functional or robust enough.

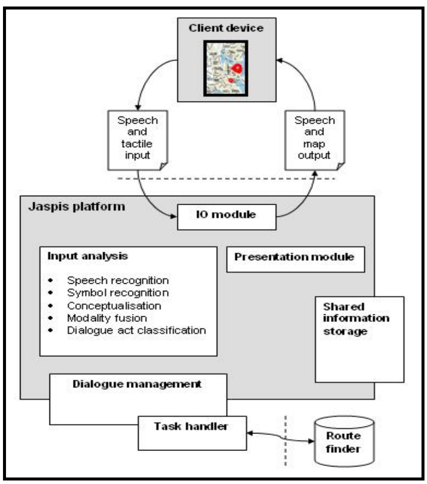

Fig. 1. General architecture of MUMS [11]

Figure 1 presents the general architecture of the system. The PDA-client only has a light-weight speech synthesizer, while the server handles all processing of the information. The touch-screen map can be scrolled and zoomed, and the inputs are recorded simultaneously and time stamped for later processing. The map interprets all pen input as locations: a tap denotes a coordinate location and a circled area a number of possible locations. The architecture is described in more detail in [11].

A screenshot of the system output is given in Figure 2. It corresponds to a situation after the user asked for information (*I want from the Opera to …<here>*), and the system provides spoken and graphical answer "*The tram 7 leaves from the Opera stop at 13:46. There are no changes. The arrival time at the Railway Station is at 13:56*".

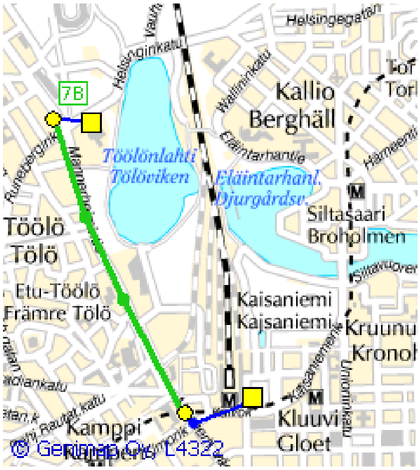

Fig. 2. MUMS example screen shot

The evaluation consisted of 17 users who had varying levels of experience with speech and tactile interfaces but none had used the MUMS or a similar system before. Before the actual evaluation sessions, they all were given a 15 minute crash-course in using the system. The users were given various scenario-based tasks ([12]) that dealt with using the system to find means of public transportation to get to certain places. They were also divided into two groups depending on the background information given to the users about the system they were to use: one group was instructed to interact with a speech interface that also had a tactile input option, while the other group was instructed to interact with a tactile system with spoken dialogue capabilities. The tasks given to both groups were exactly the same, but a priming effect on the users' behavior and expectations was expected: advanced knowledge would have impact on the users' evaluation. The expected and observed performance of the system was measured by asking the users to fill in the evaluation form twice. First they were asked to fill in their expectations of the system performance after the crash course before the actual tasks, and then again after the finishing the tasks. The same evaluation form was used to evaluate the system performance.

The evaluation questions were organized into six groups: the user's perception of the system's speech and graphical interface, the user's perception of the system's functionality and consistence, the user's perception of the system's responses and their appropriateness, the user's perception of system taking the user into account and the easiness of completing the task, the user's eagerness in future use of the system, and general questions of the overall assessment of the system. The results concern the differences in the perceived performance measures and the overall change between expectations and actual use. The latter corresponds to the actual differences in the measurement of how big the users' surprise or disappointment was as compared to what they had expected. The evaluation is reported elsewhere in detail (see e.g. [13]).

The general tendency in the evaluation is clear: user expectations were fulfilled and the system seems to have offered a genuinely positive experience. Speech and tactile input were considered natural, and especially the combination of synthetic speech and graphical map and route representations as a means of system output was praised. The biggest disappointments were experienced with the system's speed and accurate indication of how quickly it presents responses. Although it is possible to speed up the system somewhat, it is difficult to shorten the time needed for input analysis and data transfer over the mobile GPRS network: more powerful mobile devices are needed. However, the system's novelty value often seemed to overcome technical problems.

As expected, the users' prior knowledge of the system influenced their evaluations. The speech group gave very positive reviews for the system's multimodal aspects: they enjoyed using the tactile interface, and they felt that the system's speech and graphical representation contribute to the intelligibility of the system's output. They also seemed to be more willing to use a tactile interface unimodally than the tactile group, i.e. they believed that a tactile interface was more usable even if not combined with speech. The tactile group also felt that several modalities make interaction flexible and the system easy to use, but they were happier with the system's overall performance, and in particular with the rate of how often the system interprets user input correctly. The tactile group seem to believe that both tactile and speech are important, and a unimodal tactile interface would not be as usable as the multimodal one. There was also some evidence that the tactile group was more willing to use a unimodal speech system than the speech group, but these results were not significant.

It is interesting how the priming effect comes to play a role in the evaluation. For instance, the speech group seemed to feel, more than the tactile group, that the system was slow, even though the response time of the system is not affected by the form of user input. This points to the fact that speech presupposes rapid and understandable communication, and especially, if the users' prior expectations lead them to believe that the system is meant for spoken interaction, the speed of the system is important. The speech group also perceived the use of the map more positively than the tactile group (it was pleasant to look at, intuitive to use). The tactile group was more critical towards the map qualities, and in fact, the difference between the user's expectations and perceived system qualities was in absolute terms negative. Analogously, the tactile group was slightly more positive at the use of speech input and output than what they had expected, whereas the speech group was disappointed with the use of speech only system. Both groups, however, expected the system to take the user into consideration in a more individual way: this shows that the users expected a more personalized and "intelligent" system. As a whole, however, users felt the system was very easy to learn to use, and were very enthusiastic about using it again in the future.

5 Conclusions

The paper has discussed natural language communication as an important aspect of an interactive system's functionality and usability, and has also presented some evaluation results of a multimodal route navigation system to support these claims.

Evaluation focused especially on the user experience, and compared the users' expectations with real experience concerning the system properties and functionality. It is shown that the users' predisposition towards the system and prior information about the system capabilities affect the evaluation. The study also found interesting group differences that are useful to keep in mind when designing applications. For instance, younger users (under 35 years of age) were more disappointed at the interest value of the system, whereas the older ones did not consider this so important but assessed the system more on the basis of its technical features. Prior experience also seemed to make the users more critical: less experienced users regarded the novelty value of the system high and did not pay so much attention to the practical usability issues, as did the more experienced users. Although the evaluation differences may not always be pinpointed down to prior knowledge, age, or gender differences, it is important to notice that the goal of building one single practical system that would suit to most users is not reasonable. Rather, there is a need for adaptive systems that allow the users to use different modalities when interacting with the system, and can also tailor their responses according to user preferences.

User's perception of the system also depends on the system's communicative capabilities related to the task: natural intuitive interaction vs. quick and simple prompts. In general, system properties like understandability and pleasantness of the speech output are especially highly evaluated after the task. The users were unanimous that the system with both speech and tactile information is preferable whereas unimodal input and output would not be as good. However, compared with the speech group, the tactile group was surprised at the system capabilities and also had a positive experience concerning the system's functionality: the system is seen as helpful and considerate, it is cooperative and functions in a consistent way, it understands what is spoken and interaction succeeds with the first try. These aspects are commonly related to speech interfaces where inadequate performance of the speech recognizer causes problems. In our evaluations, however, speech recognition worked in a similar fashion for both groups, so the differences cannot be associated solely to speech recognition problems. Rather, the answer seems to lie in the predisposition of the users and their expectations concerning whether they interact with a speech interface that uses tactile input, or with a tactile interface that can speak as well. As mentioned above, the users automatically adjust their assessment of the system communicative capabilities with respect to what can be expected from a speech system as opposed to what can be expected from a tactile system that just happens to have speech capability. The use of (spoken) language seems to bring in tacit assumptions of fluent human communication and expectations of the system possessing similar communication capabilities. The results thus corroborate with the findings of [7] that the users project human-like properties to the computers, and more so if the system speaks and can be spoken to, supporting the observations in [8] that spoken interactive systems are regarded as agents.

When dealing with the present-day interactive services, the users are often forced to adapt their human communication methods to the needs of the technology. Simultaneously, with the development of ubiquitous environment, a growing number of interactive applications will appear, and expectations for fluent and intuitive communication get higher. For intelligent interaction, it is thus necessary to support intelligent information processing and the system's communicative capability: to understand and provide natural language expressions, recognise what is new and old information, reason about the topic of interaction, and adapt to the user's different dialogue strategies.

Acknowledgments

The author would like to thank the PUMS project partners for their collaboration in the system development and Topi Hurtig for the implementation of the system. Thanks go also to the test users, especially the group from the Helsinki City Transportation Authority for their helpful comments and participation in the evaluation.

References

1. Walker, M., Litman, D., Kamm, C., Abella, A.: PARADISE: A framework for evaluating spoken dialogue agents. In: Procs of the 35th Annual Meeting of the Association of Computational Linguistics, pp. 271–280. Madrid, Spain (1997)
2. Norman, D.A: Emotional Design: Why We Love (Or Hate) Everyday Things. In: Basic Books, Cambridge, Mass (2004)
3. Möller, S.: A New Taxonomy for the Quality of Telephone Services Based on Spoken Dialogue Systems. In: Jokinen, K., McRoy, S. (eds.) Procs of the 3rd SIGDial Workshop on Discourse and Dialogue, pp. 142–153. Philadelphia, U.S (2002)
4. Weiser, M.: The Computer for the Twenty-First Century. Sci Amweican, pp. 94–10 (1991)
5. Jokinen, K.: Interaction and Mobile Route Navigation Application. In: Meng, L.A., Zipf, S. (eds.) Map-based mobile services - usage context, interaction and application, Springer series on Geoinformatics, Springer, Heidelberg (2006)
6. Jokinen, K.: Communicative Competence and Adaptation in a Spoken Dialogue System. In: Procs of the 8th Int Conf on Spoken Language Processing (ICSLP), Jeju, Korea (2004)
7. Reeves, B., Nass, C.: The Media Equation. Cambridge University Press, Cambridge (1996)
8. Jokinen, K.: Constructive Dialogue Management - Speech Interaction and Rational Agents. John Wiley & Sons (forthcoming)
9. Jokinen, K., Kerminen, A., Kaipainen, M., Jauhiainen, T., Wilcock, G., Turunen, M., Hakulinen, J., Kuusisto, J., Lagus, K.: Adaptive Dialogue Systems: Interaction with Interact. In: Jokinen, K., McRoy, S. (eds.) Procs of the 3rd SIGDial Workshop on Discourse and Dialogue, pp. 64–73. Philadelphia, U.S (2002)
10. Harnard, S.: The Symbol Grounding Problem. Physical D42, 335–346 (1990)
11. Hurtig, T., Jokinen, K.: On Multimodal Route Navigation in PDAs. In: Procs. of the 2nd Baltic Conference on Human Language Technologies Tallinn, pp. 261--266 (2005)
12. Kanto, K., Cheadle, M., Gambäck, B., Hansen, P., Jokinen, K., Keränen, H., Rissanen, J.: Multi-Session Group Scenarios for Speech Interface Design. In: Stephanidis, C., Jacko, J. (eds.) Human-Computer Interaction: Theory and Practice (Part II), Mahwah, New Jersey, June. Lawrence Erlbaum Associates, vol. 2, pp. 676–680 (2003)
13. Jokinen, K., Hurtig, T.: User Expectations and Real Experience on a Multimodal Interactive System. In: Procs of the Interspeech Conference, Pittsburgh, U.S (2006)