EFFECTS OF USING SPEECH IN TIMETABLE INFORMATION SYSTEMS FOR WWW

Pernilla Qvarfordt and Arne Jönsson

Department of Computer and Information Science
Linköping University, S-581 83, LINKÖPING, SWEDEN
perqv@ida.liu.se arnjo@ida.liu.se

ABSTRACT

Design of information systems where spatial and temporal information is merged and can be accessed using various modalities requires careful examination on how to combine the communication modalities to achieve efficient interaction. In this paper we present ongoing work on designing a multimodal interface with timetable information for local buses where the same database information can be accessed by different user categories with various information needs. The prototype interface was evaluated to investigate how speech contributes to the interaction. The results showed that the subjects used a more optimal sequence of actions when using speech, and did fewer errors. We also present suggestions for future design of multimodal interfaces.

1. INTRODUCTION

A recent development in interactive systems is to combine WWW-interaction with speech (e.g. WebGalaxy [8]). Our aim is to develop such a multimodal dialogue system. The application domain for our study is local bus timetable information. This is a suitable domain for research on human computer interaction as it combines a variety of problem areas such as temporal and spatial reasoning [2]. Furthermore, as it is a public information system, it involves various user categories such as experts and novices, with different familiarity of the domain. One way to meet this challenge is to allow different combinations of modalities for different users.

However, it is still an open question how communicative modalities are to be selected and combined to address various information needs and support efficient interaction. Especially interesting is an investigation on the properties of allowing spoken interaction as one modality. Oviatt [12] has shown that when using a multimodal interface with maps, the task completion time was reduced and that users preferred to use a multimodal interface, instead of only pen based input or speech input. This paper presents a study on if and how the use of speech can improve the usability of an Internet application.

To guide the design of the information system, knowledge on the users and what they want from the system is important. It is not sufficient, nor advisable, to design dialogue systems that resemble human interaction behaviour [1]. Such attempts will provide models that are inaccurate and computationally ineffective and based on the erroneous assumption that humans would like to communicate with computers in the same way as they communicate with people. On the contrary, the language that humans use when they are interacting with a computer differs significantly from the language used between humans [1, 11, 4, 3].

Furthermore, the complexity of the application also affects the requirements on the system. The knowledge the interface must have in order to be an effective collaborative partner is determined by the application and the role of the agents and distinguish three different types[10]: Task dialogue, where the system guides the user's actions, Planning dialogue, where the system assists in planning the user's actions, and Parameter dialogue, where the user's task is not known to the system, an example of this is database access. Another classification is provided by Hayes & Reddy [5]. They define the class Simple Service Systems which can be said to incorporate both Planning and Parameter dialogues of Van Loo & Bego [10]. Such systems require in essence only that the user identifies certain entities, parameters of the service, to the system providing the service. Once they are identified the service can be provided. A local bus timetable information system belongs to this latter class.

In our application the users are all travellers, the main differences are traveling frequency and knowledge of the domain. Each user category also has its own requirements on the interaction and different combinations of interaction modalities address different information needs. If the user, for instance, does not know the name of the actual bus stop but only knows that it is in a certain area or near some other place, filling in a form is not of much help. In such cases a map might be more useful. A map on the other hand requires that the user knows the geographic location of a bus stop. This is not always the case, especially if the user is not familiar with the town. In such cases it might be better to enter the name using for example speech input.

2. EXPERIMENT

Based on two pre-experimental investigations, one conducted in a telephone setting between travellers and a timetable informant and one of travellers using regular timetables, a prototype was developed. The prototype was evaluated in an experiment with the aim to investigate if the use of speech makes it more useful. Thus, in the experimental setting the subjects could either interact with the interface using mouse and keyboard to enter data, or speak to the interface. The prototype did not have a speech recognizer, instead a Wizard listened to the subject's and performed a key-based command to enter data based on the subject's spoken interaction.

The prototype interface had four different parts, a fill-in form for asking questions to the database, a map that could be used for pointing out points of arrival/departure, timetable questions, and finally an area for messages from the system. The map consisted of a an overview map and a map showing magnified parts of the overview map. The magnified map had two fixed magnification factors, that also showed different amounts of details. For example, in the map with the largest magnification all the street names were visible.

2.1. Subjects

A total of 12 subjects, 6 male and 6 female, participated voluntarily in the study. The subjects were divided into three groups, corresponding to their domain knowledge, i.e. knowledge on local buses in Linköping. The different groups were called Good, Moderate and Weak, where Good stands for good knowledge of the domain.

2.2. The Wizard and his environment

The Wizard was a native citizen of Linköping and also helped to develop the prototype, so he had good knowledge of both the city of Linköping and of the limitations of the prototype.

The Wizard was in one room and listened to what the subjects said over a telephone via a loudspeaker. To be able to interact with the prototype, a special program was running on the subjects' computer. The program created a virtual desktop that both the subject and the Wizard could see and control. When the user said something the Wizard typed it in and it became visible after some delay on the subjects' screen.

2.3. Material and Procedure

The study was divided into three parts. First, the subjects were asked to give some statements about their background, such as age and knowledge of Linköping.

In the main study, each subject used two different interfaces; one multimodal and one unimodal. Using the multimodal interface the users could do all the things they could using the unimodal interface plus speak to the system. The subjects were first given a short introduction to the prototype and then had to solve different scenarios. When they were finished using one interface, the procedure was repeated for the other. Each subject was given three different scenarios and solved them using both systems.

Table 1: Average amount of errors done by the subjects

	Multimodal	*Unimodal*
Total	2.50	7.92
Good	2.50	8.00
Moderate	3.75	6.25
Weak	1.25	9.5

The scenarios were simple or complex. Simple scenarios include just one task whereas complex scenarios include three tasks. Each of the tasks had different characteristics, such as time constraints, vague descriptions of the geographic location and the need to find best route. One task may include several searches for buses. Scenario 1 contained a task with time constraints, Scenario 2 vague description of a geographic location, and Scenario 3 all three characteristics. The reason for using different situations was to investigate how to modify the prototype to best fulfill different user needs.

Finally, the subjects were given a questionnaire about their attitudes toward the prototype, with emphasis on efficiency.

3. RESULTS

In order to investigate the efficiency of the prototype, three different aspects of the usage were measured. The first was the amount of errors. The second, was if the users sequence of actions was optimal when solving the scenarios. Finally, the number of times the subjects zoomed out in the map, as a measure of how lost they were. We also studied the subjects' attitudes toward the prototype using the questionnaire.

3.1. Errors

The number of errors done by the subjects are presented in Table 1. The subjects did significantly more errors (t=-3.285, p<.01, one-tailed) using the unimodal interface than when using the multimodal. The most common type of error using the multimodal interface was that the subjects forgot to change some of the input when posing a new question. The most common error using the unimodal interface was that subjects by mistake clicked in the map. A click in the map means that the point of arrival/destination is changed to that point.

3.2. Sequence of actions

The deviation from the optimal sequence of actions to perform for a scenario was calculated, see Table 2. In Scenario 1, with time constraints, subjects kept to an almost significantly better sequence of actions (t=-1.693, p<.056, one-tailed) using the multimodal interface compared to using the unimodal. In Scenario 2, with vague geographic location descriptions, subjects kept to a significantly better sequence of actions (t=-2.549, p<.05, one-tailed) using the multimodal interface. In Scenario 3, with time constraints, vague geographic locations and instructions to find

Table 2: Mean number of deviations from optimal sequence of actions

		Scenario1	*Scenario2*	*Scenario3*
Multi- modal	Mean	1.25	3.17	6.75
	Good	0.50	3.75	3.00
	Moderate	2.25	3.25	7.50
	Weak	1.00	2.50	9.50
Uni- modal	Mean	3.67	5.67	7.83
	Good	3.00	7.50	6.00
	Moderate	2.25	4.75	8.75
	Weak	5.75	4.75	8.75

Table 3: Mean number of times subjects zoomed in the map

	Same area		*Other area*	
	Multimod.	*Unimod.*	*Multimod.*	*Unimod.*
Mean	0.25	0.75	0.33	1.58
Good	0.00	0.50	0.25	1.50
Moderate	0.50	0.75	0.25	1.50
Weak	0.25	1.00	0.50	2.00

the best route, there was no significant difference between the modalities. However, taking into account the subject's knowledge of the domain we found, for Scenario 3 when using the multimodal interface, that the subjects with good knowledge of Linköping followed, a significantly better sequence of actions (t=-1.856, p<.05, one-tailed) than the subjects with weak knowledge. Using the unimodal interface the same difference was noticeable, but far from significant.

3.3. Zooming out in the map

The number of times a subject zoomed in the map is shown in Table 3. We distinguish the case where subjects zoom out from the same area that they hade previousely zoomed in from, from cases where they zoom out from another area than they zoomed in from. An example of the latter is that the subjects could zoom in one area and then move the visible part of the detailed map to another area.

The subjects zoomed out significantly more in the map (t=-0.172, p<.01, one-tailed) using the unimodal interface than when using the multimodal. When discriminating between cases of zooming out was same tendency visible. Zooming out when having zoomed in from the same area was significantly more frequent (t=-2.171, p<.05, one-tailed) using the unimodal interface than when using the multimodal. Zooming out without previously having zoomed in in that area was significantly more frequent (t=-4.103, p<.01, one-tailed) using the unimodal interface than when using the multimodal.

3.4. Ratings

The subjects rated the multimodal interface to be faster compared to a paper-based timetable. In comparison between the multimodal and the unimodal interface, subject's thought that the multimodal interface was slightly faster than the unimodal. However, when asking for efficiency the subjects thought that the unimodal interface was slightly more efficient than multimodal.

4. DISCUSSION

Despite the fact that the subjects made fewer errors, used a more optimal sequence of actions and needed to zoom out less when using the multimodal interface, they still thought that it was less efficient to use the multimodal interface than the unimodal. This might be explained by the fact that much of the work done when interacting with the computer using multimodal interfaces was carried out by the "computer" and not by the subjects. If this is a problem, one way to solve this might be to indicate that the system is working, for example moving around the map and when the system has finished parsing a subject's utterance showing the arrival/departure point in the map. This might be irritating in the long run, but the system is intended to be a walk-up-and-use system so long run usage is not anticipated.

The effect mentioned above has been described by Laurel[9]. She argues that the drawback with natural language interfaces is that the users experience that they give commands to a hidden intermediary which then performs the actions for the user. In the multimodal interface the subjects might have experienced themselves giving commands, instead of having a dialogue. Our suggestion is to develop the copperation between the user and the system, i.e. to use a dialogue metaphor, as also suggested by Hugunin and Zue[6].

The subjects made different kinds of errors using the two systems. Using the multimodal interface the most common error was to forget to change all the parameters. In order to support the user in what to say next, the system can somehow make the user notice what have been changed and what have not been changed since the last search, for example through highlighting.

The difference between the two interfaces in keeping to an optimal sequence of actions increased with geographic complexity, that is, when the task requires searching in the map. Thus, multimodal interfaces are suitable for systems including usage of maps (c.f. [12]). One reason for this in our prototype, may be that the subjects thought that the multimodal interface had additional functionality. It was possible to say to the prototype: "Show me Arrendegatan (a street name)". The same functionality existed when using the unimodal interface, but none of the subjects used it. Clearly this functionality was more obvious when using the multimodal interface than when using the unimodal.

It seems to be harder for subjects with weak knowledge of the domain to keep to the optimal sequence of actions than for subjects with good knowledge, especially using the multimodal interface. A reason for this can be that it is harder for subjects with weak domain knowledge to express and estimate spatial relationships. The question is if there are other differences in the knowledge condition, and if a certain modality is more suitable for users with a limited or wide knowledge of the domain.

The subjects seems to loose their way easier in the map using the unimodal interface compared to using the multimodal, when they for example are zooming in the map. This might be explained by the fact stated above, that users experience the multimodal interface to have extra functionality. But it may also be that navigating in the map using the unimodal interface places higher demands on the cognitive abilities. This effect should be especially prominent in the case when users zoom out in the same area as they have zoomed in, because then the user hopefully had some idea on how the area looked liked before they zoomed in and then they could use that memory when navigating in the zoomed map.

Our design suggestions for multimodal interfaces are of the conceptual kind, so that they can be applied on multimodal interfaces in other domains. As stated above we argue for the use of a dialogue metaphor. The use of such a metaphor can also reduce the most common type of error found using the multimodal interface; forgetting to change some input parameters. If using a dialogue metaphor the system can alert the user on what has been change since the last search. We also suggest that the interface should draw the users attention to other things than just waiting, e.g. showing that it is processing a name of a bus stop by moving around the map.

5. FUTURE WORK

Future work in this field could throw some light on if users with different domain knowledge use certain modalities more efficiently than others. As part of that a future direction could be to investigate to what extent the use of different modalities influence the cognitive load.

Another important issue is the complexity of the computational mechanisms needed to allow for multimodal dialogue. Further work requires investigating to what extent simple dialogue models, which has proven sufficient for spoken or written natural language interfaces [7], are sufficient also for multimodal interfaces.

From the users point of view, some other properties, not purely computational, are important. In a system for public use the usability of the system is important. For example, it must be efficient to use the system, otherwise the users will not use the system a second time. A public timetable information system has more in common with a walk-up-and-use systems, than a database interface in an office environment.

6. ACKNOWLEDGMENTS

This work is supported by The Swedish Transport & Communications Research Board (KFB). Thanks to the Wizard, Björn Johansson.

7. REFERENCES

1. Nils Dahlbäck and Arne Jönsson. Empirical studies of discourse representations for natural language interfaces. In *Proceedings from the Fourth Conference of the European Chapter of the association for Computational Linguistics, Manchester*, 1989.
2. Annika Flycht-Eriksson and Arne Jönsson. A spoken dialogue system utilizing spatial information. In *Proceedings of ICSLP'98, Sydney, Australia*, 1998.
3. Raymonde Guindon. Users request help from advisory systems with simple and restricted language: Effects of real-time constraints and limited shared context. *Human-Computer Interaction*, 6:47–75, 1991.
4. Raymonde Guindon, K. Shuldberg, and J. Conner. Grammatical and ungrammatical structures in user-adviser dialogues: Evidence for suffciency of restricted languages in natural language interfaces to advisory systems. In *Proceedings of the 25th Annual Meeting of the ACL, Stanford*, pages 41–44, 1987.
5. Philip J. Hayes and D. Raj Reddy. Steps toward graceful interaction in spoken and written man-machine communication. *International Journal of Man-Machine Studies*, 19:231–284, 1983.
6. Jim Hugunin and Victor Zue. On the design of effective speech-based interfaces for desktop applications. In *Proceedings of Eurospeech97, Rhodes, Greece*, pages 1335–1338, 1997.
7. Arne Jönsson. A model for habitable and efficient dialogue management for natural language interaction. *Natural Language Engineering*, 3(2/3):103–122, 1997.
8. Raymond Lau, Giovanni Flammia, Christine Pao, and Victor Zue. Webgalaxy - integrating spoken language and hypertext navigation. In *Proceedings of Eurospeech'97, Rhodes, Greece*, pages 883–886, 1997.
9. Brenda Laurel. *User Centered Systems Design: New Prespectives on Human Computer Interface Design*, chapter Interface as Memesis, pages 87–124. Lawrence Erlbaum Associates, 1986.
10. W. Van Loo and H. Bego. Agent tasks and dialogue management. In *Workshop on Pragmatics in Dialogue, The XIV:th Scandinavian Conference of Linguistics and the VIII:th Conference of Nordic and General Linguistics, Göteborg, Sweden*, 1993.
11. William C. Ogden. Using natural language interfaces. In M. Helander, editor, *Handbook of Human-Computer Interaction.* Elsevier Science Publishers B. V. (North Holland), 1988.
12. Sharon Oviatt. Multimodal interfaces for dynamic interactive maps. In *Proceedings of Conference on Human-Factors in Computing Systems: CHI'96*, pages 95–102. New York, ACM Press, 1996.