A First-personness Approach to Co-operative Multimodal Interaction

Pernilla Qvarfordt and Lena Santamarta

Natural Language Processing Laboratory, Department of Computer and Information Science,
SE-581 83 Linköpings universitet, Sweden
{perqv, lensa}@ida.liu.se

Abstract. Using natural language in addition to graphical user interfaces is often used as an argument for a better interaction. However, just adding spoken language might not lead to a better interaction. In this article we will look deeper into how the spoken language should be used in a co-operative multimodal interface. Based on empirical investigations, we have noticed that for multimodal information systems efficiency is especially important. Our results indicate that efficiency can be divided into functional and linguistic efficiency. Functional efficiency has a tight relation to solving the task fast. Linguistic efficiency concerns how to make the contributions meaningful and appropriate in the context. For linguistic efficiency user's perception of first-personness [1] is important, as well as giving users support for understanding the interface, and to adapt the responses to the user. In this article focus is on linguistic efficiency for a multimodal timetable information system.

1 Introduction

Brennan [2] argues that "human conversation is inherently cooperative" (p. 396). We agree that conversation is a co-operative activity, but using conversation in human-computer interaction will not inherently result in a good interaction. Maybury [3] argues that "fundamental to coopertative interaction are mechanisms that support media interpretation, generation, (language and representation) translation, and summarization" (p.15). In this paper we argue that having the required technology in order to offer users co-operative multimodal interaction is not enough. Instead focus on the users is equally essential when designing such systems. We agree with McGlashan that "from the user's point of view, whether the interaction is co-operative or not is judged solely on the basis of what system says" [4].

Closely related to co-operation is communication. Grice [5] talks about a co-operative effort in communication. A problem is how that effort is handled, one solution is to engage the user in a multimodal interaction, e.g. Fais, Loken-Kim and Park [6] stated that "one way to improve the performance of [natural language systems] ... is to supplement their processing capabilities with multimedia technologies designed to lessen the burden on the processing system" (p. 364). Their focus is, however, not to lessen the user's effort or engage in a dialogue on the user's term, but on the systems capabilities.

Another approach involves constraining the dialogue in telephone applications, in order to reduce the number of speech recognition errors.

In this paper we will discuss how to reduce the users' co-operative effort, by using Laurel's [1] concept of first-personness. Her article "Interface as Mimesis" is often used as an argument against natural language systems. Interestingly, Laurel does not argue against natural language systems. She even says that "good interfaces are ultimately multimodal" (p. 72). However, she do argue against poor design of interactive systems. She claims that the designer must let the user be in a first-person relation with the computer.

> "The first-personness is enhanced by an interface that enables inputs and outputs that are more nearly like in their real-world referents, in all relevant sensory modalities" ([1], p. 77)

If a system shall give real-world referents, in our view, the system must offer the users natural language interaction. Laurel also describes three aspects that contributes the first-personness; interactive frequency, interactive range, and interactive significance. Interactive frequency is a measure of how often the user can give input. Interactive range is a measure of the variety of choices available to the user at any moment, and finally, interactive significance is a measure of the impacts of user's choices and actions on the whole.

In what follows, we will discuss how these concepts can be used when designing co-operative multimodal interaction for a timetable information system. We will start by giving some background to co-operative interaction and communication, and then describe our investigations for gaining knowledge on how to design the interaction.

2 Background

In order to achieve a co-operative interaction the system needs to produce utterances which are perceived by the user as natural, coherent and helpful within the context of use [7]. The question is what makes utterances natural and helpful?
Clark and Shaeffer [8] argues that one function for a co-operative dialogue is to make sure that the participants have a common understanding of the dialogue, that they share a common ground. This theory has served as a basis for developing both spoken dialogue systems [9], and graphical user interfaces [10].

Brennan and Hulteen [9] believe that the response to the user should be adaptive to the system's processing. In this way the user knows where the interaction was "interrupted" and can easily correct the information given to the system. They also emphasise the importance to give both positive evidence and negative evidence of understanding in the response to the user.

The strategies proposed by Brennan and Hulteen are developed from theories of human-human communication. Other results, however, show that users do not use the same language when talking to a computer as when talking to another human (cf. [11], [12], [13]). Cheepen and Monaghan [13] talks about the transactional and interactional goals in the discourse. Transactional goals are task-oriented and interactional goals are person-oriented, i.e. to promote a relationship between the speakers. They argue that the kind of goals pursued in dialogues with a computer are primarily transactional. The user

is interested in performing a particular task with the system, and not to small talk. For this reason Cheepen and Monaghan argues that what is natural in a human-human dialogue is not natural in a human-computer dialogue.

However, not only the type of spoken feedback given to the user is important, but also how the interaction between user and system is designed. Chu-Carroll and Nickerson [14] have noticed that in a telephone application mixed initiative and automatic adaptation of the dialogue strategies led to better performance and user satisfaction. Interesting to note is that the system's adaptations matched the expectations of the users.

3 Design Case

We have been working with the domain of timetable information of public transportation in the city of Linköping. The MALIN system (Multimodal Application of LINlin) is a prototype system under development, which consists of processing modules for interpretation, generation, dialogue management and knowledge manager. These in turn consult various knowledge sources such as the timetable database, a geographical information system, a domain model, dialogue models, lexicon, grammar etc. The system is describe in more detail in [15].

An evaluation of the first interface is presented in [16]. It allowed only multimodal input, and we are currently redesigning the interface to handle also multimodal output in addition to multimodal input. The system is designed as an information kiosk, and the spoken output from the system will be presented by an animated head.

The user shall be able to interact with the system using speech and either pointing or pen input. The system is based on the collaborative interface agent paradigm [17], [18]. Fill-in forms for questions, results and a map are all shown at the same time. Our system can be classified as a simple service system [19]. 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.

3.1 Expectation on the Interaction

We believe that how the system fulfils the user's expectations is an important factor when measuring efficiency and effectiveness. For this reason, one of the first steps in the design of our multimodal co-operative system was to investigate what expectations the users had. We therefore distributed a questionnaire with 41 different properties to students at Linköping University. A total of 114 students filled in the questionnaire, 74 students had experience from searching for bus timetable information on the Internet and 32 had not. The students had to rate how important they believed each property was for a timetable information system for public transportation (buses) on the Internet, on a 1-7 scale, where 1 was equivalent to "not important at all", and 7 "very important." We will not here presents the complete result, just the parts that are interesting in this context.

The five most important properties were (means in parenthesis); *trustworthy* (6.68), *efficient* (6.58), *relevant* (6.50), *stable* (6.41) *and fast* (6.39). Interesting to note, is that students, who had no experience in searching for bus timetable information on Internet, did rate the properties *simple, responsive, productive, spontaneous, willing to learn,*

tempting, and *supervising* as significantly (p<.05), see Table 1, more important than those students with experience. Non-experienced students also had a strong preference (p<.10) for the properties *comfortable*, *inviting*, *pleasant*, and *encouraging*. None of

Table 1. Compairison between students having experience from bus timetable on the Internet and those how have not for properties with a signigicant difference (p.<.05)

Property	Sig. (2-tailed)	Experienced		Unexperienced	
		Mean rating	Std. Deviation	Mean rating	Std. Deviation
Simple	.043	5.95	1.34	6.47	.80
Responsive	.021	4.58	1.63	5.34	1.23
Productive	.005	4.26	1.58	5.16	1.22
Spontaneous	.003	3.59	1.55	4.56	1.29
Willing to learn	.019	3.70	1.59	4.47	1.34
Tempting	.013	3.55	1.61	4.38	1.36
Supervising	.034	3.31	1.70	4.06	1.41

these properties are in the top ten in the average rating, simple is at place 11, however, most was rated above 4 amongst non experienced students. Worth noting is that many of these properties imply an agency from the computer, such as responsive, willing to learn.

3.2 Expert Evaluation

We have also developed a lofi-prototype of our multimodal timetable information system, MALIN. The prototype was evaluated in an expert evaluation. Three experts participated in the study, all had experience from designing interactive systems and conducting usability studies, but none had participated in the design of Malin or had any experience in designing multimodal interfaces. The experts tried the prototype using a scenario. They were encouraged to give comments about the prototype during usage, and to think-aloud. Afterwards they were given some questions. The question of interest here is "What would you like system to say to you?"

The experts came with several suggestions, both on the type of information the system should give and how it should behave. The comments can be divided into five groups:

- *Initiate a dialogue:* "It could say 'Where do you want to go from?' and things like that."
- *Give further/extended information:* "It could say: 'If you are in a hurry, you can walk somewhat longer' and things you do not think of when using the system." "Give a short introduction, 'With this system you can see where thingis are', both in text and in speech."
- *Clarify concepts used in the interface:* "I believed that 'nearest' meant 'fastest', it could clarify things like that." "It could clarify what is shown in the interface, e.g. 'The best time is shown in the grey field.'"
- *Clarification:* "Asking for clarification, when I have underspecified a question."

- *Behaviour:* "It should be supportive"

These suggestions have in common that the spoken comments from the system give some added value to the graphical interface. Interesting for this conclusion is that the subjects did not like the system to say something after every action, instead they said: "You can see what it has filled in."

4 Suggestion for Co-operative Dialogue

As shown from the results of the student questionnaires, users expect a timetable information system to be efficient. We can also see that speed is very important, but that efficiency is something more than just speed. In analogue with Cheepen and Monaghan's [13] discussion about transactional and interactional goals in conversation, and the fact that our system is a simple service system, the transactional goals should be of more importance for the users.

We would, however, like to focus on the part of efficiency that is not explained by speed. We believe that efficiency in co-operative multimodal dialogue systems can be divided into two, functional efficiency and linguistic efficiency. Functional efficiency means that the system is offering appropriate functions for solving a task efficiently. Linguistic efficiency means that responses are meaningful and are adapted to the context. This means that e.g. information given visually does not need to be repeated verbally. As shown from the expert evaluation, this is also something proposed by the usability experts. We believe that linguistic efficiency is especially important to reduce the co-operative effort of the user. Since linguistic efficiency also strive for adapting to the context, we suggest that, this also can contribute to the property *trustworthy.*

4.1 Input

As mentioned above, Laurel [1] discerns three interactional aspects that affect first-personness: Interactive frequency, Interactive range, and Interactive significance.

Interactive range is, as stated above, a measure of what choices are available for the user at any moment. In our case, all choices are available at any. The system is also designed to enable user input whenever, even while the system is speaking, which means that the user may interrupt the system or change her mind in the middle of a task. As in human communication barge-ins and changes of subject do always result in a momentary poorer interaction from which the dialogue should recover as soon as possible. However, this could require some additional turns. We believe that this can give users some of the wanted added value. Barge-ins also preserves the user's first-personness, since she is always in control of the dialogue.

One problematic feature of natural language systems and the aspect of interactive range is language understanding capabilities the systems have. We have for that reason given visual cues about what kind of questions the systems can answer. The interactive range of the system is also delimited by the "world" described in the application, i.e. geographical information and public transportation in Linköping. Thus the system cannot answer question about tourist information in the same city, or public transportation in another city.

The users' choices and actions have maximal significance in our design, as it is the user who decides (within the range of the context) what the system will do. Adaptability, both of interaction and language gives a maximal significance not only to what to user do but also to how it is done.

4.2 Output

Laurel [1] focuses her discussion about first-personness on the input to the system and on the constraints of the system. However, we believe that preserving the user's perception of first-personness is also a feature of the output of the system. In this section, we put forward a proposal on what is required of the output from a co-operative multimodal system in order to maintain first-personness and how spoken output supports it.

The major role of spoken language in our system, besides dialogue supporting, is to help the user to achieve a fast and easy understanding of the graphical information. This means that the system directs the user's attention to where important information is presented and complements the graphical information, e.g. by informing how graphical symbols are to be interpreted. To do this in an effective way, the spoken output has to be tailored towards the user's information needs and linguistic preferences. This implies that the same information can be presented in different ways, but a certain form will be more suitable as an answer to a question than to another.

Dialogue openings and closings, as well as giving feedback, is done via speech and gestures (compare with the expert evaluation above). The system may also ask for a confirmation or a repetition to reduce uncertainty. It is important that multimodal feedback is not perceived as intruding or irritating by the user. Instead the feedback should be perceived as co-operative. Therefore, the spoken feedback should be short and use the everyday words of the user.

The system has a sophisticated domain knowledge management capable to handle complicated spatial and temporal references [20]. Most of the database information is presented in the graphical interface, therefore it is very important that the spoken output supports its interpretation. There are two important points: first to direct the user's attention to critical data and to complement the graphical representation with explanations when needed. When the information given by the user have been relaxed in order to find a solution to the task; this is conveyed to the user as the relation between what the user said and the solution. For instance, if the user said she wanted to go to "Skäggetorpskyrkan" (The Church of Skäggetorp) and there is neither a suburb nor a bus stop with that name, the system realises that it is a church in the suburb "Skäggetorp" that is close to a bus stop. The system finds the nearest bus stop and indicates both on the map highlighting them. Then using speech and gestures, the system shows the geographical relation between the two.

Another important factor is for the system to use the referent expressions chosen by the user to avoid misunderstanding, i.e. to tailor its language to the use's linguistic preferences. For example, if the user refers to the railway station as "the railway station" the system should do so too, although the official name is "the central station". When the system refers to items on the screen, pronouns (and/or other appropriate referring expressions) are used. If needed they are accompanied by diectic actions.

To generate co-operative utterances in context, it is also very important to ensure coherence between the user's and the system's turns; i.e. to ensure that the same concepts are in focus. This is conveyed by means of prosody as well as lexical choice and syntactic structure, and results in the fact that the same data is presented in different ways depending on which information request they have to fulfil.

The students questionnaires showed that depending on user's experience, their expectations differ, and that non experienced users expects more agency from the interface. Results from the first interface [16] show that users' performance and preference also differ depending on their domain knowledge. This suggest that the spoken responses should be flexible and adapt to the user. For example, a user that barges in frequently can be given less spoken feedback, while long silences from the user can evoke more spoken initiatives from the system.

5 Conclusion

In this article we have argued that the user's co-operative effort needs to be reduced in co-operative multimodal interfaces. In order to do so, the efficiency of the interface plays an important role, especially in dialogues with transactional goals in simple service systems. The efficiency can be divided into functional and linguistic efficiency. Functional efficiency has a tight relation to solving the task fast. Linguistic efficiency concerns how to make the contributions meaningful and appropriate in the context.

We believe that a co-operative multimodal dialogue system working as we propose supports first-personness, as it always adheres to the user and it does not impose a way of interacting or of expressing. However, this is still up to the user do decide. The next step is therefore to conduct an experiment, in order to see if the user perceive the system as efficient, and co-operative, and if the user perceives first-personness in the interaction.

We have focused on discussing linguistic efficiency, which consist of preserving users' experience of first-personness, giving users support for interpretation the interface, and to adapt the responses to the user. In this article we have giving examples of linguistic efficiency of a multimodal timetable information system.

References

1. Laurel, B.: Interface as Mimesis. In D. A. Norman and S. W. Draper (Eds.) *User Centered Systems Design, New Perspectives on Human-Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum (1986) 67-85
2. Brennan, S.: Conversation as Direct Manipulation. In B. Laurel (Ed.) *The Art of Human-Computer Interface Design. Reading,* MA: Addison-Wesley Publishing Company (1990) 393-416
3. Maybury, M.: Toward Cooperative Multimedia Interaction. In H. Bunt, R.-J. Buen, and T. Borghuis (eds.) *Multimodal Human-Computer Communication, Systems, Techniques, and Evaluation. Lecture Notes in Artificial Intelligence 1374.* Springer Verlag (1998) 13-38
4. McGlashan, S.; Towards Multimodal Dialogue Management. In *Proceedings of Twente Workshop on Language Technology 11,* Enschede, The Netherlands (1996)

5. Grice, H. P.: Logic and Conversation (From William James Lectures, Harvard University, 1967). In P. Cole, and J. Morgan (Eds.) *Syntax and Sematics 3: Speech Acts.* New York: Academic Press (1975)
6. Fais, L., Loken-Kim, K.-H., and Park, Y.-D.,:. Speaker's Responses to Requests for Repetition in a Multimodal Language Processing Environment. In H. Bunt, R.-J. Buen, and T. Borghuis (eds.) *Multimodal Human-Computer Communication, Systems, Techniques, and Evaluation. Lecture Notes in Artificial Intelligence 1374.* Springer Verlag (1998) 264-278
7. McGlashan, S., Fraser, N. M., Gilbert, G. N., Bilange, E., Heisterkamp, P., and Youd, N. J.: Dialogue Management for Telephone Information Systems. In *Proceedings of the International Conference on Applied Language Processing*, Trento, Italy (1992)
8. Clark, H. H., and Schaefer, E. F.: Contribution to Discourse. *Cognitive Science.* Vol. 13, (1989) 259-294
9. Brennan, S. E. and Hulteen, E. A.: Interaction and feedback in a spoken language system. In *AAAI-93 Fall Symposium on Human-Computer Collaboration: Reconciling Theory, Syntesizing Practice. AAAI Technical Report FS-93-05* (1993) 1–5
10. Pérez-Quiñones, M. A., and Sibert, J. L.: A Collaborative Model of Feedback in Human-Computer Interaction. In *Proceeding of Conference on Human Factors in Computing, (CHI 96),* Vancover, Canada (1996) 316-323
11. Jönsson, A. and Dahlbäck, N.: Talking to a Computer is not Like Talking to Your Best Friend. In *Proceedings of The first Scandinivian Conference on Artificial Intelligence,* Tromsø, Norway (1988)
12. Cheepen, C., and Monaghan, J.: Designing for Naturalness in Automated Dialogues. In Y. Wilks (Ed.) *Machine Conversation.* Boston: Kluwer Academic Publishers (1999) 127-142
13. William, D. and Cheepen, C.: "Just speak naturally": Designing for naturalness in automated spoken dialogues. In *Proceedings of Conference on Human Factors in Computing Systems (CHI'98),* Los Angeles (1998) 243–244
14. Chu-Carroll, J., and Nickerson, J. S.: Evaluating Automatic Dialogue Strategy Adaptation for a Spoken Dialogue System. In *Proceedings of 1st Meeting of the North American Chapter of the Association of Computational Linguistics,* Seattle, WA (2000) 202-209
15. Dahlbäck, N., Flycht-Eriksson, A., Jönsson, A, and Qvarfordt, P.: An Architecture for Multi-Modal Natural Dialogue Systems. In *Proceedings of ESCA Tutorial and Research Workshop (ETRW) on Interactive Dialogue in Multi-Modal Systems,* Germany (1999) 53-56
16. Qvarfordt, P., and Jönsson, A.: Effects on Using Speech in Timetable Information System for the WWW. In *Proceedings of the ICSLP'98,* Sydney, Australia (1998) 1635-1638
17. Bunt, H., Ahn, R., Beun, R.-J., Borghuis, T., and van Overveld, K.: Multimodal Cooperation with the DENK system. In H. Bunt, R.-J. Buen, and T. Borghuis (Eds.) *Multimodal Human-Computer Communication, Systems, Techniques, and Evaluation. Lecture Notes in Artificial Intelligence 1374.* Springer Verlag (1998) 39-67
18. Sidner, C. L., Boettner, C., and Rich, C.: Lessons Learned in Building Spoken Language Collaborative Interface Agents. In *Proceedings of ANLP/NNAACLP 2000 Workshop on Conversational Systems,* Seattle, WA (2000) 1-6
19. Hayes, P. J. and Reddy, D. R.: Steps toward graceful interaction in spoken and written man-machine communication. *International Journal of Man-Machine Studies* 19 (1983) *231–284*
20. Flycht-Eriksson, A.: A Domain Knowledge Manager for Dialogue Systems. In *Proceedings of European Conference on Artificial Intelligence (ECAI'00).* Berlin: Germany (2000)