

The Role of Spoken Feedback in Experiencing Multimodal Interfaces as Human-like

Pernilla Qvarfordt

Arne Jönsson

Nils Dahlbäck

Department of Computer and Information Science

Linköping University

SE-581 83 Linköping, Sweden

+46 13 28 10 00

{perqv, arnjo, nilda}@ida.liu.se

ABSTRACT

If user interfaces should be made human-like vs. tool-like has been debated in the HCI field, and this debate affects the development of multimodal interfaces. However, little empirical study has been done to support either view so far. Even if there is evidence that humans interpret media as other humans, this does not mean that humans experience the interfaces as human-like. We studied how people experience a multimodal timetable system with varying degree of human-like spoken feedback in a Wizard-of-Oz study. The results showed that users' views and preferences lean significantly towards anthropomorphism after actually experiencing the multimodal timetable system. The more human-like the spoken feedback is the more participants preferred the system to be human-like. The results also showed that the users experience matched their preferences. This shows that in order to appreciate a human-like interface, the users have to experience it.

Categories and Subject Descriptors

H.5.2 [Information interfaces and Presentation]: User Interfaces – *voice I/O, evaluation/methodology, natural language, graphical user interfaces*

General Terms

Experimentation, Human Factors

Keywords

anthropomorphism, multimodal interaction, spoken feedback, Wizard of Oz

1. INTRODUCTION

Traditionally, in human-computer interaction the computer system has been regarded as a tool. In recent years this view have been challenged by the development of multimodal interfaces and life-like characters, see e.g. Cassell [3], Cassell et al. [4] Cohen [6], Lester et al. [15], Horvitz [11], Oviatt [17], Oviatt et al [18] and Takechui and Nagao [24]. These kinds of user interfaces tries to mimic the way humans communicate with each other to make the

interaction easier and more natural (Maybury [16]). To further put weight behind the argument that human-human communication can be used as model for human-computer interaction is the work of Reeves and Nass [19] often cited. Reeves and Nass argue that humans seem to have one basic way of seeing the world; this way is based on how we see other humans. Therefore it is possible that we also have one basic way to interact with the world; that is how we interact with other humans.

Human-like computer interfaces are sometimes called anthropomorphic interfaces. Anthropomorphism is used ambiguously in Human-Computer Interaction and in Multimodal Interaction. In these areas it is more common that the word is used for interface agents, see e.g. Heckman and Wobbrock [10], i.e. anthropomorphism is a characteristic of the user interface. Anthropomorphism originally means that users attribute human characteristics to an artefact, e.g. a computer system. These two meanings are related; for example if the computer interface has an animated agent present, the users are likely to attribute human characteristics to it (Schaumburg [21]). However, it is not necessary that an interface agent is present to make the user experience it as human-like, a famous example is ELIZA [26], which appeared to be human-like by some of its users when it was first developed. To attribute human characteristics to a computer interface is therefore not limited to interfaces that shows a human face. Instead a wide range of interface features can in some manner make the user experience the interface as human-like. In order to avoid the ambiguity of the word anthropomorphism we choose to use the word human-like in this paper. This also stress that an interface can be more or less like a human. Anthropomorphism is not absolute.

When designing multimodal interaction, it is appropriate to make conscious decisions whether the computer interface should be human-like or tool-like. In order to make this kind of decision, knowledge must be gained about what makes the user experience one interface as human-like and another as tool-like. This paper investigates how spoken feedback influences the users' experience of a multimodal interface as being human-like or tool-like.

2. Related research

Much research has been carried out on both evaluating multimodal interaction and the use of life-like characters. Some of these systems are more tool-like, for example the pen-based systems used by Oviatt [17] and Suhm et al. [23]. These tool-like multimodal systems have been liked by the users in the evaluation studies. Life-like characters, which are human-like, have not showed overall benefit that its promoters hoped for (see Burgoon et al. [2] and Dehn and van Mulken [8]). One reason for these

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICMI '03, November 5–7, 2003, Vancouver, British Columbia, Canada.
Copyright 2003 ACM 1-58113-621-8/03/0011...\$5.00.

results can be that more human-like interfaces face different challenges than tool-like interfaces. Some examples of these challenges are: consistency between the characters verbal and visual cues are important for the users' positive experience of a system (Isbister and Nass[12]), the facial expression influenced the users' experience and interaction with the computer system (Walker et al. [25]), and the presence of an animated character can make some users more anxious when using the computer system Rickenberg and Reeves [20]. All these results are interesting, but they do not give an answer to when a computer interface is experienced as more human-like or tool-like.

Previous attempts to make a distinction between human-like and tool-like interfaces have been tried both by analytical means and by empirical means. Frohlich [9] identified two ways to make input to an interface, either by *actions* or by *language*. *Actions* correspond to Direct Manipulation, and *language* corresponds to interacting with a computer system through language, i.e. command language, natural language or field fill-in. If this distinction should be used to distinguish between tool-like systems and human-like systems, one issue would arise as problematic. The category of language is probably too wide. It is not likely that users think the system is human-like when they fill in a form or select an item in a menu. In the latter case, it is also possible to ask what the difference is between pushing a button, selecting an item in a menu, or uttering the command. In all three cases the effect can be the same, for example saving a file.

Laurel [14] has suggested that it is how the user is treated in the interaction that makes the user experience the interface as anthropomorphic or not. If the interaction is more about interacting with the tool than with the task, then the user will experience the interaction as passing through a hidden intermediary rather than interacting directly with the system. She explains the intermediary as an "ill-formed presence or persona" [p.75] that does not belong to the context, and takes care of what the user wants to do. In this example the user gives a command as input. It could be written, spoken, or chosen in a menu. The effect is, according to Laurel, the same. The system takes care of the command, executes it, and reports what it has done.

To further complicate the issue between tool-like interfaces and anthropomorphic interfaces, Reeves and Nass [19] have shown that humans respond in the same way to various media as to other humans. To human perception, human beings and media seem to be treated in the same way. This could imply that even if a computer system is designed to be tool-like, the users perceive it as if it was anthropomorphic. However, this does not mean that the user thinks of, or experiences the system as a human. Jönsson and Dahlbäck [13] and Shechtman and Horowitz [22] have independently shown that humans communicate very differently with other humans compared to communicating with computers. Humans tend to make assumptions about their partner and adjust their language to fit the partner. To use the same strategies to interpret humans and computer systems is not the same as to think of the system as human.

This discussion shows that it is hard to draw a line separating anthropomorphic from non-anthropomorphic systems. It seems that it comes down, not to analytical properties in user interfaces, but to how users experience the interaction, and so far this issue has not been sufficiently investigated in an empirical manner.

3. DESIGN OF THE STUDY

The purpose of this study was to investigate if users of a multimodal timetable information system think of it as human-like or tool-like, and if spoken feedback influences this experience. Three different conditions were designed to test this. The difference between the conditions were: if spoken feedback was present or not, and what type of spoken feedback that was given to the users. Users interacted in the same manner in all three conditions; that is by speech and pointing on a touch screen.

Without spoken feedback—only the graphical feedback was used.

Limited spoken feedback—in addition to the graphical feedback functions, some limited spoken feedback was used. The spoken feedback in this condition mainly reported what the system was doing, e.g. "Searching the timetable", if there were any errors, and if new information became visible on the screen. The spoken feedback could also give help messages.

Complete spoken feedback—in this condition the same spoken and graphical feedback were used as in the limited spoken feedback condition. However, in this condition the system could also ask for missing information, i.e. clarifying questions, and take initiative if the user was inactive for a longer time.

The main difference between the limited spoken feedback condition and the complete spoken feedback condition is that the system takes initiative in the latter condition. The initiatives are taken if the user pauses. The initiatives are supposed to help the users get started again. Figure 1 shows an example of a transcribed dialogue from the limited feedback condition.

System:	Välkommen till Malin PQL <i>Welcome to Malin PQL</i>	opening
User 7:	Ja:g vi:ll åka: buss frå:n Mö:rtstige:n ti:ll Si:mhalle:n * <i>! wa:nt to take a bus fro:m Mö:rtstige:n to: the in:door swimming poo:l *</i>	
System:	Vilken dag vill du åka? <i>What day would you like to go?</i>	clarification
User 7:	Sö:nda: . och jag vill vara framme: klockan fyra eller sexton <i>Su:nday . and I want to be there: at four o'clock or sixteen</i>	
System	Letar i tidtabellen <i>Searching the timetable</i>	process handling

Figure 1. Transcription of a spoken dialogue in the complete feedback condition. (.) denote full stop and (:) prolonged vowel.

Note the stars (*) in Figure 1. At this point the user makes a break, and the system takes initiative and asks for more information. In the limited feedback condition, in Figure 2, the user makes a break at the same location. However, here the system does not take any initiative when the user makes a pause. Instead it waits for more input.

In the limited spoken feedback condition the system made partial openings. This means that the system says "Welcome to Malin PQL", but does not take initiative to ask for information, by saying for example "From where do you want to go?" as it could do in the complete feedback condition. In the dialogue in Figure 2, the user is fast in making his first utterance, and the system adjusts to this behaviour.

The spoken feedback was constructed to avoid formulations that could be considered to be anthropomorphic, e.g. using the word “P”. This kind of formulations can influence the user experience of the system (Brennan and Ohari [1]).

System: Välkommen till Malin PQL **opening**
Welcome to Malin PQL

User 9: m: . okej jag vill åka från Mö:rstigen . äm: till Simmhallen . * öh på söndag . och jag vill vara där: klockan fyra
 m: . okay I want to go from Mö:rstigen . uh to the indoor swimming pool . * uh on sunday . and I want to be there at four o'clock

System: Ett ögonblick **process handling**
one moment, please

System: Den närmaste hållplatsen är med i tidtabellen **direct attention**
The closest bus stop is shown in the timetable

Figure 2. Transcriptions from the limited feedback condition.

3.1 Participants

In this study, data were collected from 30 participants, 16 women and 14 men. The ages of the participants ranged from 19 to 59 years, with a median age of 25 years. The participants had different background in computer literacy, education, and profession. The participants received a ticket to the cinema as compensation for their time.

The participants were randomly assigned to one of the three conditions, ten participants in each group.

3.2 The experimental set-up

The experimental environment consisted of two parts, the timetable information system, and the experimental set-up.

3.2.1 The timetable information system

The timetable information system is implemented in Java, and was fully functioning except for the speech and gesture recognition, which was simulated using a Wizard-of-Oz.

The graphical user interface is presented in Figure 3. At the top of the interface, there are several connected fill-in fields. All these fields need to be filled in by the user in order to get a timetable presented. The timetable is visible to the left under the fill-in fields. A list of alternative bus stops is shown under the timetable. The map to the right can show locations requested by the user. The field for showing a location is under the map.

The fill in field serves as a support to help the users know what to say to the timetable system (See Figure 4). Both the fields for asking for a bus route and asking for a location in the map have a button connected to them. The “show location” button is visible in Figure 4, and the button for searching for a bus route is in Figure 3.

The users have flexibility in how to express locations to the system. The user can ask for a bus stop, a location such as the Cathedral, a street, or an area such as the city centre. When the user gives a location other than a bus stop as the point of arrival or departure the system will calculate which bus stops are closest to the location. If the location is an area, the system will show which bus stops are inside the area. When the user has asked a

question the system will present a list of alternative bus stops under the timetable, as shown in Figure 3. The users can choose whether they are satisfied with the timetables, or whether they want to redo the question with one of the presented alternatives.

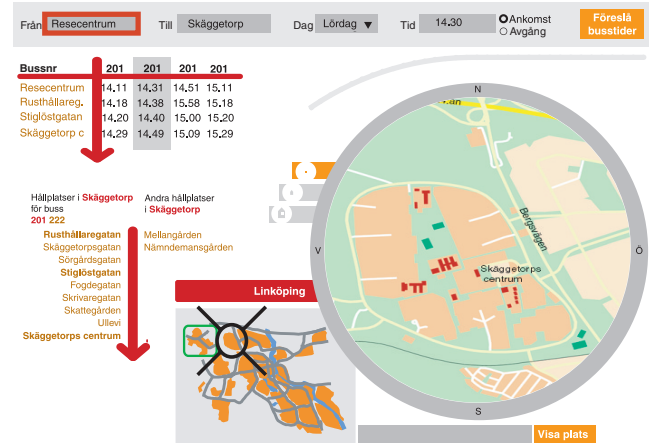


Figure 3. The user interface to the multimodal timetable information system.

Fields for a question on bus routes

Från Till Dag Tid
 From To Day: Saturday Time

Field for question to show location in the map

Don't understand Show location

Figure 4. The fill in fields in the user interface.

The user interface is connected to three different databases; the timetable database and two geographical information systems (GIS). The timetable database used here is the local public transportation company’s (Östgotatrafiken) database¹. In order to allow the participants to give streets, areas, and landmarks as input to the timetable system, a spatial reasoner was implemented in one of the GIS, which figures out which bus stops are located in an area, or are closest to a landmark or street. The spatial reasoner runs in ArcView on a PC connected to the internal network in the Department of Computer and Information Science, Linköping University. The last database is also a GIS and it handles the maps. This GIS, an ArcIMS, run on an Apache server on a UNIX machine. The ArcIMS creates maps in GIF-format that can be displayed in the user interface.

3.2.2 The Wizards’ applications

Two Wizards were needed to control the graphical output and the spoken feedback. In the following the Wizard responsible for the graphical output is called Application Wizard and the Wizard responsible for the spoken feedback is called Feedback Wizard.

The Application Wizard’s interface was also implemented in Java so that it could be easily integrated with the timetable information system. The Application Wizard’s interface was used to control the graphical feedback on the participants’ utterances to the user and occasionally to control the maps. The Application Wizard’s

¹ Available on the Internet address www.ostgotatrafiken.se

interface was specially designed to speed up the interaction and to give the participant feedback as fast as possible.

The spoken feedback used in the study consisted of pre-synthesised utterances from a concatenation speech synthesizer, T4 (Telia Text till Tal). The utterances were played back to the user via a Macromedia Director application controlled by the Feedback Wizard.

All of the participants' and the Application wizards' actions were recorded in an interaction log. The spoken feedback was as well recorded in a log.

3.2.3 *Physical setting*

The participant and the wizards were located in separate rooms. The participant's microphone was connected to a loudspeaker in the wizards' room. The participant and the wizard had a screen connected to each other, so that the wizard could see what the participant was doing. Each wizard had in addition a screen showing the wizard's application.

The user interface was shown on a flat TFT touch screen, which lay on the participant's desk to make it more convenient for pointing. An experimenter was present in the room with the participants throughout the test, making observations and managing the test. The participant's screen was video filmed by a digital video camera.

3.3 The Wizards

In this study, two different persons acted as Application Wizards mainly due to restrictions in the availability of Wizards. Only one Feedback Wizard was needed.

3.3.1 *Training and instruction*

Before the study the Wizards had extensive training, both with their tools and the scenarios. The Wizards received instructions on how to respond to the user.

3.3.2 *Performance of the Wizards*

The wizards' performance was measured by two measurements, the error rate and the wizards' response time. No differences between the factors Wizard and condition, nor any interaction was found with an analysis of variance.

The overall error rate was 12% on the items entered by the wizard, or 3% of the words uttered by the participants. These error rates include both mistakes from the wizards, i.e. miss recognized words, and dialogue errors, i.e. "Don't understand". The overall error rate did not differ between the conditions.

The Feedback Wizard made 1293 utterances to the participants that were transcribed, 507 in the limited feedback condition and 760 in the complete feedback condition. Only 11 utterances were erroneous, an overall error rate of 0.851%, five for the limited feedback condition and six for the complete feedback condition.

3.4 Material

Two types of material were given to the participants. The first presented tasks to the participants in the form of scenarios. The second was questionnaires used to collect data from the participants. The main source of data in this study came from questionnaires.

3.4.1 *Scenarios*

Five scenarios were developed for the study. The same scenarios were used in all conditions. In addition, two training scenarios were developed; one for the experimenter and one for the participant. The scenarios were developed in accordance with Dahlbäck et al. [7]. The main purpose of the scenarios was to give the participant a context of use.

In one scenario the context was taken from the participants' everyday life, and thus differed for each participant. In the other four scenarios the context was fixed (constrained scenarios). The latter kind of scenarios was the same for all participants. They asked the participants to make a specific trip. However, they were also designed to give the participants freedom to choose how to phrase their points of departure and arrival, by giving street names, landmarks (e.g. the Cathedral), or areas (e.g. the city centre). The scenarios were given both in written and graphical form.

3.4.2 *Questionnaires*

The participants were given questionnaires at three different points during the test session. They were given a background questionnaire, a scenario questionnaire and a post-test questionnaire.

The background questionnaire was divided into two parts. One part was presented verbally to the participants and the other was a written questionnaire. This questionnaire measured participant's knowledge of the town of Linköping, Sweden, their experience of public transportation in Linköping, their background and experience in using computers and multimodal systems.

After each scenario, the participants were asked to fill in a scenario questionnaire. This questionnaire measured perceived efficiency and satisfaction with the solution of the scenario.

The last questionnaire was the post-test questionnaire. This questionnaire aimed at measuring different aspects of the participants' experience.

3.5 Procedure

The session started with a brief introduction to the study, and the experimenter asked the questions in the verbal questionnaire. Then she gave a longer introduction to the study. After that the participants were asked some questions from the background questionnaire. Then the system was introduced and demonstrated for the participant, and after the demonstration the participants also had an opportunity to try the system.

After the introduction, the participant was given the scenarios one at a time, and was instructed to read the scenario and then press a start button in order to start using the system. To finish a scenario the participants pressed a button. Then they were asked to answer the scenario questionnaire. The order of the scenarios was the same for all participants, starting with the constrained scenarios, first the single trip scenarios, then the multiple trips scenarios, and ending with the open scenario. This order of the scenarios was chosen to reflect an increasing level of difficulty.

After the test the participants were asked to fill in the post-test questionnaire, and they were also free to make any comment in a short interview. The interview was recorded using a minidisk. At the end they were informed that they had participated in a Wizard-of-Oz study, and were asked for permission to use the material for

scientific purposes. They received a ticket to the cinema as a reward for their participation in the test, and were urged not to tell anyone about the study until it was completed. The whole test took from one to two hours.

4. RESULTS AND DISCUSSION

The purpose of this study was to investigate whether spoken feedback made the participants experience the multimodal timetable system as human-like or tool-like, and what the participants prefer after experiencing different kinds of feedback. The analysis of the participants experience was done by asking them if they experienced their interaction as more like asking a human for timetable information, or if it was more like looking up the answer in a paper-based timetable. In order to catch the participants' preferences, they were asked before using the system how they preferred searching for their timetable information, and after they had used the system they were asked how they preferred the system to be; more human-like or more tool-like.

The analysis was done in three steps. The first step was to examine how the participants experienced timetable system, as human-like or tool-like. The second step was to investigate whether the spoken feedback influenced the participants' preferences after having used the system. Finally, the third step was to see if the participants' preference of the system as human-like or tool-like matched with and their experience of the system as human-like and tool-like.

The general performance measurements, task completion time and task success, did not differ for the conditions. This means that the results presented below were only influenced by the manipulation, i.e. the presence and type of the spoken feedback. All participants succeeded in solving their scenarios. The task completion time did not differ between the three conditions, ($F(8, 100)=.550$, ns.). The number of dialogue contributions did not differ between the conditions ($F(2, 26)=1.063$, ns.).

4.1 Experiencing the system as human-like or tool-like

The participants' perception of the interaction differed between the conditions (see Figure 5). Note that the more spoken feedback the participants got, the more of them thought the system was like asking a person than looking up the information in a timetable. When testing for significance with a χ^2 -test, these differences did not show any significance ($\chi^2(2)=2.1$, $p=.350$). However, a χ^2 -test is not sensitive to interactions between the variables. In this case the interaction between the different conditions is prominent; the more spoken feedback there is the more participants experience the system as like asking a person for timetable information.

Also the participants' comments in the interview suggest that they experienced the timetable system differently. Participant 26 added without being asked in the end of the interview the following comment:

I can add one thing; it felt like a tool. (Participant 26, Without spoken feedback, Look up, spoken comment)

A more subtle comment from Participant 10 illustrates the timetable system as human-like. In the interview the participant talks about her being polite to the system.

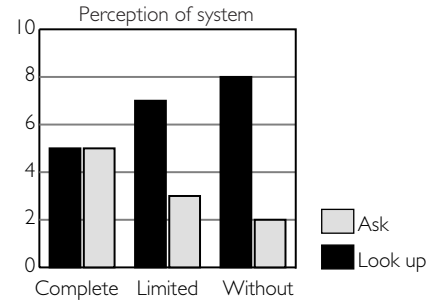


Figure 5. No. of participants experience the system as human-like (ask a person) or tool-like (look up in a timetable)

The thing is, I guess, that you have to realise that it is a machine, and that I don't need to wait and listen to everything. (Participant 10, Full feedback, Ask a person, spoken comment)

The written comments on the question of how the participants experienced the system give some clue as to what the participants thought were the characteristics of tool-like systems and human-like systems. When judging whether the timetable system is more like asking a person or looking up information in a timetable, the amount of spoken feedback mattered for some participants. One participant in the condition without spoken feedback thought that spoken feedback would change the experience of the system.

It should probably talk back in order to be experienced as a person. It was hard to know if it was 'thinking' or if it did not understand what I said. (Participant 18, Without spoken feedback, Look up, written comment)

What Participant 18 wants was not just any spoken feedback, but spoken feedback that support grounding (Clark and Schaeffer [5]), i.e. understanding of what the system is doing and understanding.

Also participants in the full feedback condition thought that more spoken feedback would change their perception. Below is a written comment from one of the participants motivating why she thought that the timetable system was more like using a paper-based timetable.

The system did not give me the answer of what was the most suitable time. Instead I had to look that up myself. (Participant 13, Full feedback, Look up, written comment)

Again it is not just more feedback that Participant 13 misses. Participant 13 wants to have more support in solving the task, and for the system to be more involved in the joint activity.

However, it was clearly not only the spoken feedback that influenced the participants' view of if the system was tool-like or human-like, but also the graphical interface. A human-like system has limited ability to give an overview and more precise answers. This view was apparent across the conditions. Participant 35 and 27 illustrate this view.

In a timetable one gets a better overview, something that I missed in the program. (Participant 35, Without spoken feedback, Ask a person, written comment)

I think that I get more precise answers from something written such as a timetable. (Participant 27, Full feedback, Look up, written comment)

In the interview Participant 27 further develops what he thinks about the system.

I like this kind of system that is something between looking up in a timetable and asking a person. One can combine the best of both worlds. (Participant 27, Full feedback, Look up, spoken comment)

This indicates that the border between a human-like and a tool-like system is not clear-cut; users can experience a computer system as both human-like and tool-like. Speech input and output gives the users a feeling of asking a person. Graphical overview and precision makes the participants view the system as more like a paper-based timetable. Participant 30 also illustrates this.

Actually a combination. I (voice)ask and get in no uncertain terms. (Participant 30, Limited feedback, Look up, written comment)

The experience of the timetable system as tool-like or human-like influenced to some extent how the participant felt about talking to the system. Independent of the condition, if the participants experienced the system as more like using a paper-based timetable, then they thought that talking to the system was like giving a command. When it comes to the participants who experienced the system as more like asking a person about timetable information, some participants thought it was more like giving a command, and some, more like a conversation or asking someone a question. The confusion about how to categorise the dialogue with the system is illustrated by Participant 33 who in the interview said:

It was more as a command; it was only the first half that was a conversation, which is my part [his contribution]. (Participant 33, Full feedback, Ask a person, spoken comment)

These results are not particularly surprising, since a spoken interaction between two persons can have many forms. Sometimes you ask another person a question, sometimes you give a command and sometimes you just make small talk. The way of determining whether the system is human-like is not dependent on if the users give commands to the multimodal system. However, commands are the only type of spoken interaction you can have with a tool.

4.2 Influence on the attitude

The large majority of the participants stated in the background questionnaire that they preferred looking up information in a timetable to asking a person (see Figure 6). Only two participants stated that they preferred asking a person. However, after the experiment, participants' preference of how the system should be like was strongly influenced by their actual experience. None of the participants in the condition without spoken feedback, but half of the participants in the full feedback condition, preferred the system be more human-like. These preferences were compared, and categorised as *no-influence* and *influence* (see Figure 6 again).

The coded data were analysed using a Chi-square test. The Chi-square analysis showed that the differences were significant over the conditions ($\chi^2(2)=6.477$, $p=.039$). The different conditions were compared using Fisher exact probability test. The largest difference, the difference between the full feedback condition and the condition without spoken feedback, was significant ($p=.0163$).

Few participants mentioned that their anticipation was influenced by the interaction with the timetable system, but those who did gave support to this conclusion independent of how they experienced the system.

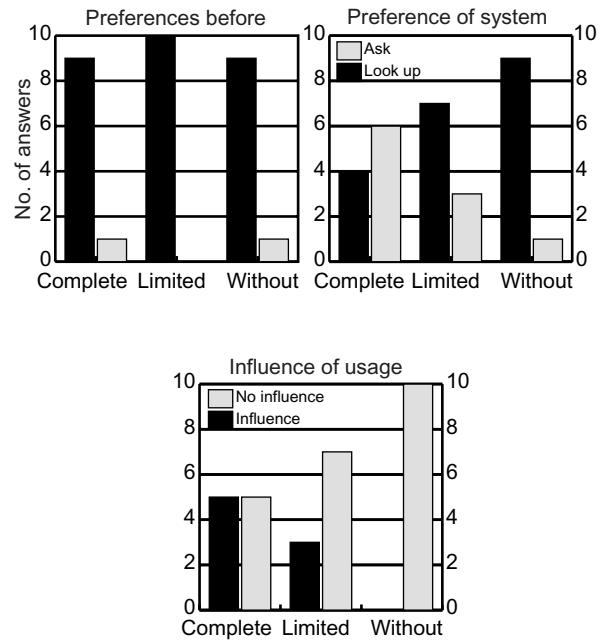


Figure 6. No. of participants preferring getting timetable information by asking a person or looking up in a timetable (before using the system), no. of participants preferring the timetable to be like asking a person or looking up in a timetable (after using the system), and the frequency of influence of the usage.

At the beginning it was a little like this: You are used to searching ordinary bus timetables, like Östgötatrafiken on the Internet. You wanted to use it in the same manner, even though you realise that you could speak naturally. (Participant 4, Without, Look up, spoken comment)

When we talked before about the system it felt like you would type in something now and then. But I used the touch screen very little. When it asks with speech, then you answer with speech. (Participant 8, Full feedback, Ask a person, spoken comment)

These results show that despite the general preference of looking up timetable information in a timetable, the participants who experienced more human-like features in the user interface liked it. Interesting to note is that the participants in the without spoken conditions answered both the questions in the same way.

4.3 Match between the participants' experience of the system and their preference

So far we have looked at if the participants experience the system as human-like or tool-like, if this experience differs between the different conditions, and if the spoken feedback influenced the participants' attitudes. The last thing to look into is the participants' experience of the timetable system as human-like or tool-like matched with their preferences. For this analysis the data from the questionnaire was categorised as match or non-match, and tested for a general difference using χ^2 -test.

Figure 7 shows firstly the perception of the system, secondly the preference of system character, and thirdly whether the participants' perception and their preference matched. The participants' experience of and preference of the system's to be tool-like or human-like match well for all conditions. The χ^2 -test

confirmed that there were no significant differences between the conditions ($\chi^2(2)=1.25, p=.535$).

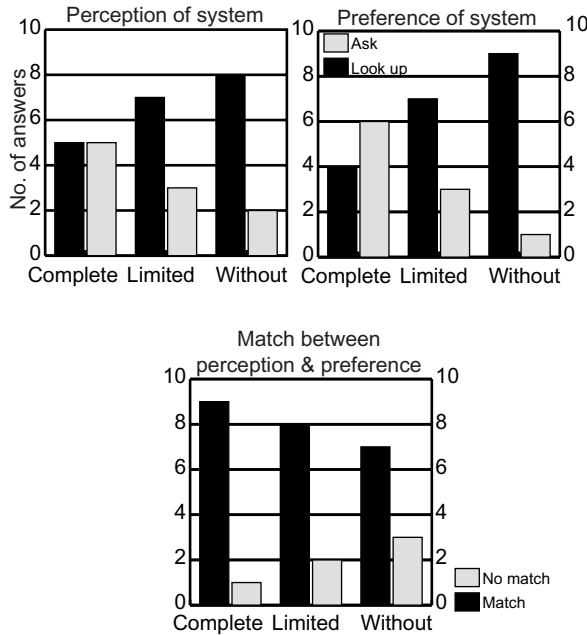


Figure 7. Frequency of experience of and preference for the system being more like asking a person or looking up information a timetable, and frequency of match between these two options.

The participants' comments on these questions show that they not only experience the system in a particular way, but that they also impose a character on the system that fit their preferences. The comments from Participants 4 and 27 illustrate this.

You always wish that a computer would work like a human being. Preferably, it should be a mind reader :-) (Participant 14, Limited feedback, Ask a person, written comment)

I think I like the precise information that you can get if it is less like a conversation than a conversation is. (Participant 27, Full feedback, Look up, spoken comment)

From another angle, Participant 5, illustrates what she did when the character of the system did not match her preference.

I think that I realised that I was thinking about whom was behind the system, someone like Östgötatrafiken, to create some physical person to better handle it in my head.

You tried to think about someone?

Yes, right, like an organisation or a person. Something that was more. I can't accept that it [her speech] just went into a machine. (Participant 5, Limited, Look up, spoken comment)

If users of a computer system not only experience the system as tool-like or human-like, but also have preferences regarding what the system should be like, then this has implications for design. A system should be flexible enough to meet the users' expectations of it as human-like or tool-like on the one hand, and on the other hand be allowed to change the users' expectations, since users' idea of what constitutes a good design solution are very vague before they start using the system.

5. CONCLUSIONS

The participants experienced the multimodal timetable system as either tool-like or human-like, and their experience was influenced by the amount of spoken feedback presented. The more spoken feedback there was, the more the participants experienced the system as like asking a human being for timetable information. The comments from the participants confirm that spoken information is important if they are to experience the system as human-like. To experience the multimodal timetable system as a tool, overview and precision were important factors.

If the participants experienced the system as tool-like or human-like also influenced how the participants described their interaction with the system. Participants that thought the system acted like a tool said they gave commands to the system. Participants that thought the system acted human-like had a more scattered view on their input; they said it either was like asking questions, like a conversation, or like giving commands.

The results from this study also show that, despite a general preference towards tools-like interaction, when the participants experienced a good alternative, they could change their view to prefer a more human-like interaction.

6. ACKNOWLEDGEMENT

This work has been supported by the Graduate School in Human-Machine Interaction, the Swedish Agency for Innovation Systems (VINNOVA) and the Centre for Industrial Information Technology (CENIIT). Special thanks to the wizards Petronella Andersson, Martin Wiman and Behrang Gilanpour.

7. REFERENCES

- [1] Brennan, S. E., and Ohaeri, J. O. Effects of message style on users' attribution toward agents. In *Proceedings of the Conference on Human Factors in Computing Systems: CHI'94* (1994), ACM Press, pp. 281–282.
- [2] Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundberg, M., and Allspach, L. Interactivity in human-computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behaviour* 16 (2000), 553–574.
- [3] Cassell, J. Embodied conversational agents, representation and intelligence in user interfaces. *AI Magazine Winter* (2001), 67–83.
- [4] Cassell, J., Bickmore, T., Billinghamurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., and Yan, H. Embodiment in conversational interfaces: Rea. In *Proceeding of the Conference on Human Factors in Computing Systems: CHI'99* (Pittsburgh, 1999), pp. 520–527.
- [5] Clark, H. C., and Schaeffer, E. F. Contributing to discourse. *Cognitive Science* 13 (1989), 259–294.
- [6] Cohen, P. R. The role of natural language in a multimodal interface. In *Proceedings of the Symposium on User Interface Software and Technology: UIST'92* (1992), ACM, pp. 143–149.

- [7] Dahlbäck, N., Jönsson, A., and Ahrenberg, L. Wizard of oz studies - why and how. *Knowledge-Based Systems* 6, 4 (December 1993), 258–266.
- [8] Dehn, D. M., and van Mulken, S. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies* 52 (2000), 1–22.
- [9] Frohlich, D. M. The design space of interface. In *Multimedia: Systems, interaction and Application*, L. Kjelldahl, Ed. Springer-Verlag, Berlin, 1991, pp. 53–69.
- [10] Heckman, C. E., and Wobbrock, J. O. Put your best face forward: Anthropomorphic agents, e-commerce consumers, and the law. In *Proceedings of the International Conference on Autonomous Agents* (2000), pp. 435–442.
- [11] Horvitz, E. Principles for mixed-initiative user interfaces. In *Proceedings of the Conference on Human Factors in Computing Systems: CHI'99* (Pittsburgh, 1999), pp. 159–166.
- [12] Isbister, K., and Nass, C. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies* 53 (2000), 251–267.
- [13] Jönsson, A., and Dahlbäck, N. Talking to a computer is not like talking to your best friend. In *Proceedings of the First Scandinavian Conference on Artificial Intelligence* (Tromsø, Norway, March 9-11, 1988).
- [14] Laurel, B. Interface as mimesis. In *User Centered Systems Design: New Perspectives on Human Computer Interface Design*, D. A. Norman and S. W. Draper, Eds. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986, pp. 87–124.
- [15] Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI '97*, (1997), 359-366.
- [16] Maybury, M. T. Intelligent user interfaces for all. In *User Interfaces for All, Concepts, Methods, and Tools*, C. Stephanidis, Ed. Lawrence Erlbaum Associates, Publishers, Mahwah, NJ, 2001, pp. 65–80.
- [17] Oviatt, S. Multimodal interfaces for dynamic interactive maps. In *Proceedings of the Conference on Human-Factors in Computing Systems: CHI'96* (1996), New York, ACM Press, pp. 95–102.
- [18] Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., and Ferro, D. Designing the user interface for multimodal speech and pen-based gesture application: State-of-the-art systems and future research direction. *Human-Computer Interaction* 15 (2000), 263–322.
- [19] Reeves, B., and Nass, C. *The Media Equation – How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publications and Cambridge University Press, Cambridge, New York, 1996.
- [20] Rickenberg, R., and Reeves, B. The effect of animated characters on anxiety, task performance, and evaluation of user interfaces. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'2000)* (2000), pp. 49–56.
- [21] Schaumburg, H. Computers as tools or as social actors? – The users' perspective on anthropomorphic agent. *International Journal of Cooperative Information Systems* 10, 1, 2 (2001), 217–234.
- [22] Shechtman, N., and Horowitz, L. M. Media inequality in conversation: How people behave differently when interacting with computers and people. In *Proceedings of the Conference on Human Factors in Computing Systems, (CHI'03)* (2003), pp. 281–288.
- [23] Suhm, B., Myers, B., and Waibel, A. Multimodal error correction for speech user interfaces. *ACM Transactions on Computer-Human Interaction* 8, 1 (March 2001), 60–98.
- [24] Takeuchi, A., and Nagao, K. Communicative facial displays as new conversational modality. In *Proceedings of the Conference on Human Factors in Computing Systems (INTERCHI'93)* (1993), pp. 187–193.
- [25] Walker, J. H., Sproull, L., and Subramani, R. Using a human in an interface. In *Proceeding of the Conference on Human Factors in Computing Systems: CHI'94* (1994), pp. 85–91.
- [26] Weizenbaum, J. ELIZA—a computer program for the study of natural language communication between man and machine. *Communication of the ACM* 9, 1 (1966), 36–45.