Influence of User Characteristics on the Usage of Gesture and Speech in a Smart Office Environment

Stefan Schaffer, Julia Seebode, Ina Wechsung, Florian Metze, and Christine Kühnel

Quality and Usability Lab, TU Berlin,
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
stefan.schaffer@zmms.tu-berlin.de,julia.seebode@zmms.tu-berlin.de,
ina.wechsung@telekom.de,fmetze@cs.cmu.edu,christine.kuehnel@telekom.de
http://www.qu.tlabs.tu-berlin.de

Abstract. In this paper we present the results and analysis of an evaluation study conducted on the gesture and speech based interface to an office information system. The presented analysis focuses on differences in subjective ratings between expert users and non-expert users as well as male and female users. Participants performed several tasks with three versions of the system. Each time, they were asked to rate the system afterwards by filling out two different questionnaires. The results show no impact of prior experience with the system on user ratings. But ratings are affected by domain expertise in speech-dialog-systems and gesture controlled systems. Hardly any gender differences are found, with the sole exception of ratings of the hedonic quality - stimulation.

Key words: usability studies, multimodal dialogue, haptic interfaces

1 Introduction

Since the first steps taken in the sixties, gesture control via touch screen has more and more become a part of everyday life, being for example the sole interface to many ticket machines and to Personal Digital Assistants (PDAs). Such devices with a gesture controlled interface can be used for an increasing variety of tasks. Interacting with those applications via stylus or finger is widely spread nowadays, but the combination with speech or other input modalities might still improve the interfaces usability as multimodal systems are often considered to be more flexible and intuitive [1].

To assess the quality of a system a number of now well established questionnaires has been developed [2]. During a system evaluation those questionnaires are typically given to a group of participants who have been using the system in question. Here, it must be taken into account that the perceived quality of a multimodal system is not only affected by the systems characteristics but also by the context of usage and by user characteristics. The ratings can thus be

influenced by individual differences between the users that may have an impact on system usage and should not be neglected. One example is that the cognitive processes carried out when participants answer a question depend on individual characteristics such as memory capacity [3]. If user differences are systematic so-called user groups can be defined and characterized. This might lead to the result that a certain interface is well-suited for one user group but must be improved to cater for other groups.

In the here reported study we focused on the impact of user characteristics such as gender aspects and differences between expert and novice user on the perceived quality of the system. According to [4] gender differences in regard to attitude are expected as well as differences between expert and novice users (cf. [5]).

In the following section the system used for the study is described and a short overview on participants and procedure is given. Results and analysis are presented in Section 3 and discussed in Section 4. The paper closes with a view on future work.

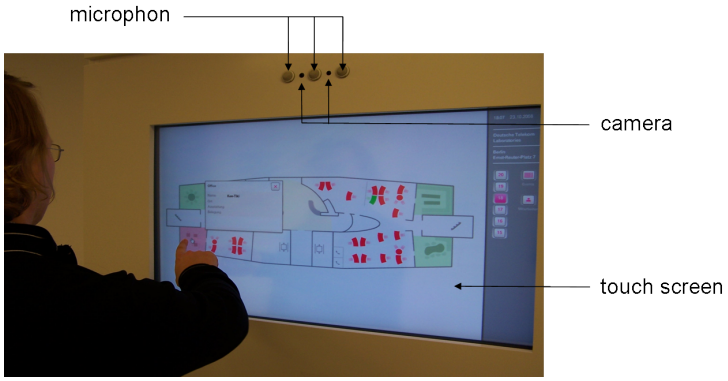

Fig. 1. User interacting with the system

2 Method

2.1 Participants and material

In this study we evaluated an office information and room management system, called *attentive displays*, which was originally operable via gesture input and has been enhanced with the possibility of speech input (cf. Figure 1). The systems output is given via the graphical user interface. Three different implementations of the system were compared, each offering different input possibilities to the user: GESTURE, SPEECH and MULTIMODAL (gesture and speech). The input modality was varied as within factor.

Thirty-six German speakers (17 male and 19 female) aged between 21 and 39 (M=27.19, SD=3.45) took part in an one hour experiment. Half of the participants were employees at Deutsche Telekom Laboratories (intern) and are therefore considered as experienced users concerning the usage of the *attentive displays*. The other half consisted of students who were paid to take part in the experiment (extern) and had no prior experience with the system. As the *attentive displays* originally offered gesture input only, the experienced users were accustomed to the information system itself, namely its capabilities concerning the given tasks (see below). They had not, however, any experience in interacting with the system via speech, but only in using the gesture input to control the system. The non-experienced (extern) users in turn had never in any form been exposed to the system before.

Two questionnaires were used to collect user ratings regarding the perceived quality of the system: a modified SASSI questionnaire [6] and the AttrakDiff [7].

Table 1. Modified and excluded SASSI items

Speech (original)	I sometimes wondered if I was using the right word.
Gesture (modified)	I sometimes wondered if I was using the right button.
Multimodal (modified)	I sometimes wondered if I was carrying out the right action.
Speech (original)	I always knew what to say to the system.
Gesture (modified)	I always knew which button to use.
Multimodal (modified)	I always knew which action to carry out.
Speech (original)	It is clear how to speak with the system.
Gesture (modified)	It is clear how to interact with the system.
Multimodal (modified)	It is clear how to interact with the system.
Excluded	The system is pleasant.
Excluded	The system is friendly.

The modified SASSI questionnaire consists of 32 items, rated on a five-point Likert scale with the possible answers being 'strongly agree', 'agree', 'undecided', 'disagree' and 'strongly disagree'. The items form 6 different scales: accuracy, likeability, cognitive demand, habitability, speed and annoyance. As the SASSI questionnaire has been designed for unimodal speech-dialogue-systems some of its items had to be modified or excluded because they were not appropriate for our system (cf. Table 1).

The AttrakDiff questionnaire consists of 28 semantic-differential items. Every item is rated on a 7-point Likert scale with the poles described by antonyms. Taking the average of 7 items respectively yields 4 scales: pragmatic quality (PQ), hedonic quality - stimulation (HQ-S) and hedonic quality - identity (HQ-I) and attractiveness (ATT). The results obtained with these questionnaires are presented and discussed below in reference to the scales contemplated.

Apart from these two questionnaires, demographic data was assessed at the beginning of the test. Here, also the experience with speech-dialogue-systems

(SDS) and with gesture controlled systems (GCS) were assessed as bipolar variables: 10 users had experience with SDS and 18 had experience with GCS, both groups are considered as *domain experts* regarding the specific input modality.

2.2 Procedure

The six tasks to be accomplished by the participants consisted of changing screens, searching for the workplace of employees, for rooms and for bookings and differed in complexity (as defined by the minimum number of interaction steps necessary to complete the task) shown in Table 2.

Table 2. Description of tasks and required interaction steps.

Task	Description	Minimum # of interaction steps required with	
		speech input	gesture input
T1	Show main screen	1	1
T2	Show 18. floor	1	1
T3	Search for employee	3	6
T4	Search for room	2	- [a]
T5	Show event screen	1	1
T6	Show room for event	1	1

[a] If the room was accidentally booked at the time of the test, the task was solvable with two clicks. Otherwise a search over all rooms was necessary.

The experiment was divided into three blocks (cf. Figure 2). Each block consisted of 6 tasks that had to be carried out with the complying modality (GESTURE and SPEECH whose order was randomized, followed by MULTIMODAL input, where the participants could choose their preferred modality for each interaction step). Every block was followed by the two questionnaires: the modified SASSI [6], adapted to gesture/multimodal input and the AttrakDiff [7].

3 Results and Analysis

Neither the analysis of the SASSI questionnaire nor the analysis of the AttrakDiff questionnaire showed any significant differences between the two user groups of intern and extern participants. The AttrakDiff questionnaire yielded no significant differences between experienced (with SDS or GCS) and not experienced users. However significant differences were observed on the SASSI questionnaire. Participants with experience in speech-dialogue-systems (N=10) rated the GESTURE input significantly better on three scales than those without prior experience (N=26) as can be seen in Table 3.

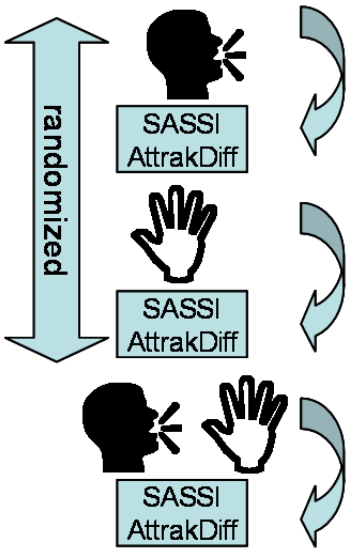

Fig. 2. Test sequence of GESTURE, SPEECH and MULTIMODAL tests

Table 3. Results of the t-test for SDS experts

			SDS experts		non-experts	
SASSI scale	$t(df)$	p	M	SD	M	SD
Accuracy (gesture)	t(33)=3.86	.000	**3.36**	.37	2.81	.40
Likeability (gesture)	t(33)=2.03	.050	**3.35**	.33	3.07	.41
Habitability (gesture)	t(34)=2.24	.032	**3.07**	.63	2.41	.88

The groups with and without experience with gesture input were compared. Again the data assessed with the SASSI questionnaire yielded some significant results. Domain experts regarding gesture controlled systems (N=18) rated the SPEECH input to be less accurate than those without prior experience (N=18). They also rated the SPEECH input to have a (marginally significant) higher cognitive demand and the GESTURE input to be slower (cf. Table 4).

Table 4. Results of the t-test for GCS experts

			GCS experts		non-experts	
SASSI scale	$t(df)$	p	M	SD	M	SD
Accuracy (speech)	t(34)=2.47	.019	.96	.49	**1.37**	.52
Cog. demand (speech)	t(34)=1.86	.071	1.08	.48	**1.42**	.62
Speed (gesture)	t(34)=2.39	.023	1.78	.26	**2.03**	.36

Concerning gender differences, a t-test carried out on the AttrakDiff data showed significant differences between male and female participants on the AttrakDiff scale HQ-S (cf. Table 5) for all three test conditions.

Table 5. Results of the t-test for the AttrakDiff scale HQ-S

			male		female	
input	$t(df)$	p	M	SD	M	SD
Gesture	t(33)=2.48	.018	.14	.89	**.87**	.85
Speech	t(33)=2.67	.012	.72	.71	**1.29**	.55
Multimodal	t(34)=2.55	.015	.87	.81	**1.49**	.66

The SASSI questionnaire showed no significant differences between male and female participants.

4 Discussion

In accordance with [1] hardly any gender differences are found, with the sole exception of the AttrakDiff scale HQ-S. Apparently, women experience the system as more stimulative, regardless of the input modality. As no significant differences between the two user groups of experienced (intern) and non-experienced (extern) participants could be found the interface design may be considered as intuitive. This might indicate that women set a higher priority on intuitive system design then men.

Concerning the lack of influence of the factor prior experience, two explanations are conceivable. One being that the experts (intern) were experienced only for gesture input and might have a high preference for this input modality, thus forestalling any possible advantages in speech - that they might have compared to the non-experienced users (extern). A second explanation could be that half of the non-experienced had experience with other kinds of systems with gesture input and could thus compensate for some of the experience of the experienced users.

Users, who have expertise with SDS rate the Gesture input to be more accurate, likeable and to have a higher habitability. This might be explained with the low expectations those users have on speech controlled systems.

Users, who have expertise with GCS rate the Speech input to be less accurate and to have a higher cognitive demand. Again, this might be due to the fact, that gesture controlled state-of-the-art systems work reliable. In contrast to this experience, the Speech input may appear even worse.

Furthermore, users who have expertise with GCS rate the Gesture input to be slower than users without experience. This might indicate that the dialogue structure of the *attentive displays* is not yet optimal compared to state-of-the-art systems.

It must be said that the last three findings refer to the analysis of the answer to only two questions in the questionnaire presented at the beginning of the test. This can – at best – be interpreted as a hint towards possible inferences. For further work a more elaborate questionnaire assessing expertise should be used to determine the user expertise in different domains.

5 Future Work

During the conducted usability study the speech recognition module, which was used to extend the *attentive displays*, was affected by a high recognition error rate. One problem was that any acoustic event that occurred in scope of the microphones caused a speech recognition hypothesis. This effect is particularly undesirable if a person that does not want to interact with the system is standing near by. Face detection via cameras is added to the system to handle suchlike situations. Thereby the speech recognition is only started when a face is detected in front of the screen. A second user study with the enhanced system is being analyzed. Additionally, we want to investigate if audiovisual speech recognition, which also can be conducted by the combination of cameras and microphones, enhances the recognition capabilities of the system.

References

1. Jokinen, K., Hurtig, T. 2006. User Expectations and Real Experience on a Multimodal Interactive System. Proceedings of the Interspeech 2006, Pittsburgh, US.
2. Wechsung I. & Naumann, A.: Established Usability Evaluation Methods for Multimodal Systems: A Comparison of Standardized Usability Questionnaires. In Proceedings of PIT 2008. Heidelberg: Springer, in press.
3. Krosnick, J. A.: Survey research. Annual Review of Psychology 50 (1999), pp. 537 - 567
4. Canada, K. and Brusca, F. The technological gender gap: Evidence and recommendations for educators and computer-based instruction designers. Educational Technology Research and Development (1991), 39 (2), pp. 43 - 51
5. Kamm, C., Litman, D., Walker, M. 'From novice to expert: the effect of tutorials on user expertise with spoken dialogue systems.' In Proceedings of the Interspeech 1998, Sydney, Australia
6. Hone, K.S. and Graham, R. Towards a tool for the subjective assessment of speech system interfaces (SASSI). Natural Language Engineering (2000), 6(3/4), pp. 287 - 305
7. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: J. Ziegler & G. Szwillus (eds.) Mensch & Computer 2003. Interaktion in Bewegung, pp. 187–196. B.G. Teubner, Stuttgart