

Multimodal Interaction: Intuitive, Robust, and Preferred?

Anja B. Naumann¹, Ina Wechsung¹, and Jörn Hurtienne²

¹ Deutsche Telekom Laboratories, Berlin Institute of Technology, Ernst-Reuter-Platz 7,
10587 Berlin, Germany

² Chair of Human-Machine Systems, Berlin Institute of Technology, Franklinstr. 28-29,
10587 Berlin, Germany
anja.naumann,ina.wechsung@telekom.de; hurtienne@acm.org

Abstract. We investigated if and under which conditions multimodal interfaces (*touch, speech, motion control*) fulfil the expectation of being superior to unimodal interfaces. The results show that the possibility of multimodal interaction with a handheld mobile device turned out to be more intuitive, more robust, and more preferred than the interaction with the individual modalities *speech* and *motion control*. However, it was not clearly superior to *touch*.

Keywords: multimodal interfaces, intuitive use, usability.

1 Introduction

Multimodal interfaces allow for using multiple communication channels simultaneously or one after another [1] which is assumed to be analogue to natural human-to-human communication [2]. This led to the expectation of multimodal systems enabling flexible, robust, efficient, and intuitive interaction [1]. However, there are results both supporting this assumption [e.g. 3] and showing the contrary [e.g. 4]. In this study we investigate if multimodal interaction really is better than interacting with single modalities in a handheld mobile device (*touch, speech, motion control*) and how different modality conditions support intuitive interaction (defined as the users' effective and subconscious application of prior knowledge [5]) and robust interaction (i.e. fast and error free interaction). In addition, we want to find out if multimodality is utilized fully and if modalities which are preferred are also used more often.

2 Method

30 participants (15 male/15 female) aged between 22 and 78 years took part in the study. The device used for testing was a smartphone (HTC Touch Diamond; IBM embedded ViaVoice) controllable via *motion* (tilt and twist), *speech* and *touch* input. System output was graphical for all modalities. For motion control, additional tactile

feedback and for speech control, additional auditive feedback was given. The application tested was a multimodal mailbox system capable of handling speech-, e-mail- and fax-messages, call forwarding, and notifications of mailbox messages.

For measuring the consequences of intuitive use according to [5] we used a questionnaire with the subscales: *Perceived Cognitive Load* (6 items, e.g. “The use of the system was not complicated.”), *Perceived Task Completion*: (6 items, e.g. “I was able to reach all my goals with the system.”), *Perceived Error Rate*: (3 items, e.g. “No problems occurred while using the system.”), *Perceived Effort of Learning*: (6 items, e.g. “I immediately knew how to operate the system.”), *Familiarity*: (6 items, e.g. “The operation of the system always met my expectations.”), *Global*: (1 item: “The use of the system was intuitive.”). Items were answered on a five-point Likert scale (0= strongly disagree, 4= strongly agree). All results show scale means. Higher values indicate a higher probability for intuitive use of the system tested. For measuring the robustness, performance data (successful task completion and aborts of task execution) were logged while the participant was dealing with the tasks. Also, the task duration and the chosen modality were recorded. The participants had to execute 4 blocks of tasks with a total of 14 tasks (get messages, reply to them, forward, and sort messages as well as changing notification options). The tasks were chosen to equally favour the different modalities. If the task goal was not achieved within three trials, task execution was aborted and the next task started. First, participants were asked to solve all tasks with a given modality. Afterwards, participants evaluated the interaction via the questionnaire. This was repeated for all three modalities, the sequence of the modalities was balanced between the participants. In the final block, participants were free to choose the modalities they used for solving the task. In this condition it was always possible to switch or to combine modalities. Again, the participants evaluated the interaction after solving all tasks in this condition.

3 Results

Questionnaire for measuring intuitive use: An ANOVA showed significant differences between the four test blocks. For most subscales, the lowest value and thus the worst rating was given to *motion control*. The best rating was always given to the *multimodal* condition (see Figure 1). As post hoc tests show, the *multimodal* test block was better rated than the interaction in all three unimodal blocks on the scales *Perceived Cognitive Load*, $F(3, 81)=12.27$, $p=.000$, part. $\eta^2=.31$, *Perceived Effort of Learning*, $F(2.37, 61.74)=10.16$, $p=.000$, part. $\eta^2=.25$, and *Global*, $F(2.20, 57.16)=16.34$, $p=.000$, part. $\eta^2=.39$. Regarding the scales *Perceived Task Completion*, $F(3, 81)=21.59$, $p=.000$, part. $\eta^2=.44$, *Perceived Error Rate*, $F(3, 78)=14.95$, $p=.000$, part. $\eta^2=.37$, and *Familiarity*, $F(3, 78)=22.81$, $p=.000$, part. $\eta^2=.47$, the test blocks *speech* and *motion control* were rated worse than the *touch* and the *multimodal* test condition. *Touch* was rated as good as *multimodal*.

Robustness (successful first trials and aborts): The participants had more successful first trials in the *multimodal* and in the *touch* condition, $F(3, 81)=26.58$, $p=.000$, part. $\eta^2=.50$. In both conditions, more than 60 percent of the tasks were solved in the first trial (see Figure 2).

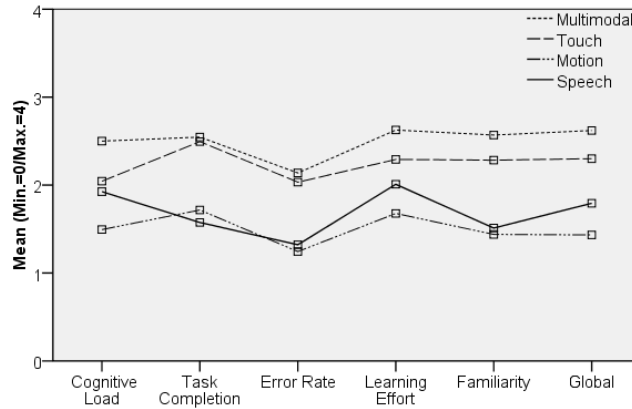


Fig. 1. Results on all subscales of the questionnaire measuring intuitive use by modality

Furthermore, in the *multimodal* condition, less tasks were aborted than in all other conditions, $F(2.25, 60.67)=20.18$, $p=.000$, $\text{part. } \eta^2 = .43$ (see Figure 2).

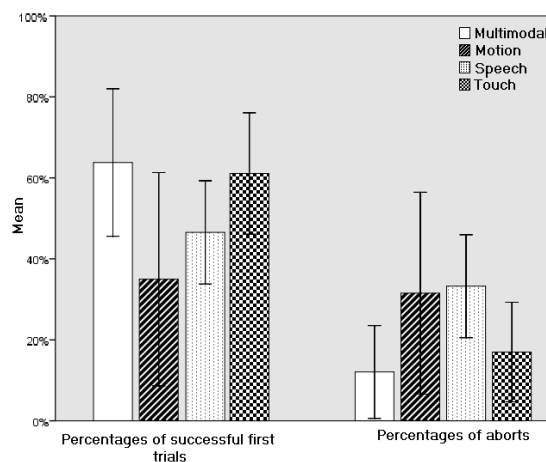


Fig. 2. Percentages of tasks successfully completed in the first trial and percentages of tasks aborted. Error bars display one standard deviation

Task duration: The task duration varied between the four conditions, $F(2.29, 61.94)=76.42$, $p=.000$, $\text{part. } \eta^2 = .74$. Most time was needed for *motion control* and least for the *multimodal* condition.

Frequency of modality usage: In the *multimodal* condition, *touch* was the modality most preferred by the participants, $F(1.35, 36.41)=28.42$, $p=.000$, $\text{part. } \eta^2 = .51$. In 68 % of the tasks, *touch* was the first choice. For 19 % of the tasks, the participants first chose *speech* input. For 7 % of the tasks *motion control* was used. For only 6 % of the tasks a combination of the different modalities was the first choice (i.e. solving different steps of the task with different modalities). In case the

participants were not successful they could try to solve the task for a second and a third time. Also in the second trial, there are differences regarding the preference of modalities, $F(1.76, 47.65)=13.57$, $p=.000$, part. $\eta^2=.33$. Again, most users preferred *touch* (59%), followed by *speech* (20%), a combination of modalities (16%), and *motion control* (5%). In the third trial, 52% of the users preferred *touch*, 21% preferred a combination of modalities, 17% *speech*, and 9% *motion control*, $F(3,69)=5.75$, $p=.001$, part. $\eta^2=.20$.

Modality switches: Only few participants used a flexible interaction strategy. In case they were not successful with a task in the first trial, only 21% of the participants switched to another modality in the second trial, $\chi^2(4, N=147)=323.4$, $p=.000$. Even in the third trial, 73% of the participants kept using the modality they had chosen before. Of the remaining participants most switched to a combination of modalities in both the second (12%) and the third trial (11%).

4 Discussion and Conclusion

Overall, the *multimodal* condition always showed results at least as good as all unimodal conditions regarding performance and subjective ratings. It was rated as more intuitive and less straining than *speech* and *motion control*. Also, the percentage of successful first trials was higher and the percentage of aborts lower than for *speech* and *motion control*. The results show that the possibility of multimodal interaction turned out to be more intuitive, more robust, and more preferred than the interaction in the individual modalities *speech* and *motion control*. However, even if a combination of modalities was possible in the last test block, most participants stayed with *touch*. Thus, for the subjective ratings, there were only small differences shown between the *touch* and the *multimodal* condition. Switching to another modality should pay off the most when errors are made during the first trial solving a task [1]. However, in our study, participants did not switch the modality. They did not utilize the advantage of multimodality for compensating usability problems in other modalities. A long term observation of modality usage and switches is planned in order to get more insight to this question.

References

1. Oviatt, S.: Ten myths of multimodal interaction. *Comm. of the ACM*, 42(11), 74-81 (1999)
2. Chen, F.: Multimodal Interaction Systems. In F. Chen: *Designing Human Interface in Speech Technology*, pp.212-223, Springer, Berlin (2006)
3. Sarter, N.B., & Woods, D.D.: Teamplay with a powerful and independent agent: A full-mission simulation study. *Human Factors*, 42(3), 390-402 (2000)
4. Wechsung, I. & Naumann, A.B.: Evaluation Methods for Multimodal Systems: A Comparison of Standardized Usability Questionnaires. In: *Proceedings of PIT. LNAI*, Vol. 5078, pp. 276-284. Berlin: Springer (2008)
5. Naumann, A., Hurtienne, J., Israel, J.H., Mohs, C., Kindsmüller, M.C., Meyer, H.A. & Husslein, S.: Intuitive Use of User Interfaces: Defining a Vague Concept. In: *Proceedings of HCI 2007*, Vol. 13, LNAI 4562, pp. 128-136, Heidelberg: Springer (2007)