Modeling of Modality Selection in Multimodal Human-Computer Interaction

Stefan Schaffer (stefan.schaffer@zmms.tu-berlin.de)
Research Training Group prometei, Berlin Institute of Technology, Germany

Keywords: strategy selection; multimodal user behavior.

Research Interests

The main interest of my research is the development of model-based methods for simulation and automated usability prediction of multimodal interfaces. In particular I want to investigate how modality choice of users can be predicted and simulated by a computational model that estimates the quality of multimodal interfaces. Therefore I am also interested in exploring rules and cognitive processes that impact users' modality selection in multimodal human-computer interaction.

Previous Multimodal HCI Related Work

During my master thesis at Deutsche Telekom Laboratories (T-Labs) I worked on the integration of a speech recognition module in so-called Attentive Displays. The Attentive Displays are an interactive wall-mounted information system for employee and room search in a smart office environment. Originally the system was controlled with a touch screen only. To enhance the input facility I embedded a speech interface. Thereby the system input was altered from unimodal to multimodal.

Several studies have shown multimodal interfaces to be more robust, efficient and flexible than unimodal systems (e.g. Oviatt, 2003). As a part of my master thesis a user study with 36 participants and six tasks was conducted to investigate the effect of multimodality on user behaviour and the perceived quality of the system. In contrast to the assumption that the multimodal system is judged best, the evaluation revealed the perceived quality of both the touch screen and the multimodal version of the system were rated equally. The distinct malfunction of the speech recognition module in the multimodal setup could be a reason for this result. While accomplishing the tasks with the multimodal system it could also be observed that users switched from speech to touch input, after experiencing repeated speech recognition errors. Otherwise speech was the preferred input modality for tasks that were solvable with less interaction steps via speech (Metze et al., 2009).

Current Research Work

Currently I am working towards my PhD, where I am developing user models for the simulation of interaction between users and multimodal dialogue systems. Thereby the modality choice of users has to be simulated in each interaction step. A literature research within HCI related topics exploring user behavior and our previous work show that miscellaneous factors like e.g. expertise (Kamm et al., 2008; Seebode, 2009), task and efficiency (Naumann, 2008) and task success (Wechsung et al., submitted) influence modality selection. According to these findings the Attentive Display user study indicates that efficiency of interaction and system errors affect user behavior (Metze et al., 2009). Usage of shortcuts via speech and modality switch after repeated malfunction of the speech recognition module was observed in the study. Users appear to prefer more efficient interaction strategies.

In the following two subsections I give a closer description of two of my current research projects.

Project 1: Modeling Efficiency-Guided Multimodal Strategy Selection

In order to build a model for selected Attentive Display tasks, human data about modality usage, recorded during the experiment, serves as a target value. Currently the employee search task including a shortcut via speech input is modeled with the cognitive architecture ACT-R (Anderson & Lebiere, 1998). To search the employee "Patrick" by means of touch screen or speech input the following interaction steps have to be fulfilled:

[pre 1] Search button "SEARCH" [post 2]
[pre 2] Press button "SEARCH" [post 3]
[pre 2] Speak "SEARCH PATRICK" [post 11]
[pre 3] Search button "P" [post 4]
[pre 4] Press button "P" [post 5]
[pre 5] Search button "A" [post 6]
[pre 6] Press button "A" [post 7]
[pre 7] Search "T" [post 8]
[pre 8] Press button "T" [post 9]
[pre 9] Search button "PATRICK" [post 10]
[pre 10] Press the button "PATRICK" [post 11]
[pre 11] Search goal cue [post end]

This simple task analysis is implemented in the ACT-R model as instructions in declarative memory. The model also provides a couple of production rules for retrieving the instructions, searching in the interface, pressing buttons and speaking commands. The structure of the model is similar to the model presented by Taatgen et al. (2006) where declarative instruction chunks are associated through pre- and postconditions. This makes it easy to reuse instructions which are used for speech and touch interaction (e.g. production [pre 1] and [pre 11]). Additionally this representation features a practical flexibility which can be used for simulating multimodal interaction with ACT-R. If two chunks with the same precondition are added to declarative memory, different interaction strategies can be retrieved and different postconditions can be set. Thereby it has to be taken into account that chunks where the precondition occurs twice are chosen randomly. Hence the model does not reproduce human behavior. One possibility of solving this problem is to use the ACT-R inherent mechanisms production compilation (Taatgen & Lee, 2003) and utility learning. The production compilation mechanism combines two production rules into one new rule and substitutes retrievals from declarative memory directly into the new rule. Thus specialized productions for speech and touch interaction are created. Utility learning rewards all rules which are involved in reaching the goal. The total reward is a stated value and spreads over the involved rules. Consequently the reward per rule is lower if more rules were involved. By means of these mechanisms it should be possible to let ACT-R learn the utilities of new production rules during an initial training. After the training the strategy involving less production rules should have a higher utility. Hence the more efficient modality should be used with a higher probability.

The aim of this research project is to investigate if ACT-R could be applied directly as a decision mechanism for modality selection in a development environment (the MeMo Workbench), which is based on prior work of the T-Labs (Möller et al., 2006). Furthermore rules for modality selection will be derived.

Project 2: Efficiency-Dependent Thresholds for Modality-Changing

This research project aims to develop a multimodal prototype for purposes of investigating thresholds for modality changing. Users of multimodal systems often have the possibility to choose a specific input modality to perform an interaction step during the processing of a task. Diverse factors influencing modality choice including efficiency-related factors like time to solve the task, interaction steps and cognitive load have to be considered. The objective is to examine whether users change their input modality from touch to speech interaction or vice versa, if the modalities offer different efficiencies. Therefore I propose a task which systematically allows varying the number of interaction steps to solve a goal. Additionally cognitive load should be kept as constant as possible. The task will be integrated in an application on a mobile device.

The findings of this project should be translated into rules which will be used by the MeMo Workbench.

Future Work

In addition to the aforementioned projects further research is required on factors like cognitive load, dual task, experience, system errors and individual user attributes. A detailed factor model describing the effects and relations of the factors to each other should be deployed. Furthermore the developed models should be validated by transferring to other tasks.

My findings about modality selection will be integrated into the MeMo Workbench which so far only facilitates the evaluation of unimodal system models. After the extension MeMo will be validated again. Therefore systems and tasks which have been explored in prior experiments will be modeled with MeMo. The empirical data gathered during the experiments will serve as target values.

Acknowledgments

The research is funded by the German Research Foundation (DFG - 1013 'Prospective Design of Human-Technology Interaction') and supported by Deutsche Telekom Laboratories.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Erlbaum.

Kamm, C., Litman, D., & Walker, M (2008). From novice to expert: the effect of tutorials on user expertise with spoken dialogue systems. *Proccedings of the 9th Annual Conference of theInternational Speech Communication Association*.

Metze, F., Wechsung, I., Schaffer, S., Seebode, J., & Möller, S. (2009). Reliable Evaluation of Multimodal Dialogue Systems. In *Proceedings of the 13th international Conference on Human-Computer interaction*. Springer-Verlag, Berlin, Heidelberg, 75-83.

Möller, S., Englert, R., Engelbrecht, K., Hafner, V., Jameson, A., Oulasvirta, A., Raake, A., Reithinger, N. (2006). MeMo: Towards automatic usability eval-uation of spoken dialogue services by user error simulations. *In Proceedings of INTERSPEECH 2006*. Pittsburgh, PA.

Naumann, A. B., Wechsung, I., & Möller, S. (2008). Factors Influencing Modality Choice in Multimodal Applications. In *Proceedings of the 4th IEEE Tutorial and Research Workshop on Perception and interactive Technologies For Speech-Based Systems: Perception in Multimodal Dialogue Systems*. Springer-Verlag, Berlin, Heidelberg, 37-43.

Oviatt, S. (2003). Multimodal interfaces. In *the Human-Computer interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (pp. 286-304). Hillsdale, NJ: L. Erlbaum Associates.

Seebode, J., Schaffer, S., Wechsung, I. & Metze, F. (2009). Influence of Training on Direct and Indirect Measures for the Evaluation of Multimodal Systems. *10th Annual Conference of the International Speech Communication Association,* (Interspeech 2009).

Taatgen, N.A., Huss, D. & Anderson, J.R. (2006). How Cognitive Models can Inform the Design of Instructions. *Proceedings of the seventh international conference on cognitive modeling* (pp. 304-309). Trieste, Italy: Edizioni Goliardiche.

Taatgen, N.A. & Lee, F.J. (2003). Production Compilation: A simple mechanism to model Complex Skill Acquisition. *Human Factors*, 45(1), 61-76.

Wechsung I., Schaffer S., Schleicher R., Naumann A. & Möller S. (submitted). *The Influence of Expertise and Efficiency on Modality Selection Strategies and Perceived Mental Effort.*