A New Experimental Paradigm For The Assessment Of User Behavior In Multimodal Interaction

Nikola Bubalo[1], Felix Schüssel[2], Frank Honold[2], Michael Weber[2], and Anke Huckauf[1]

[1] Department of General Psychology, Ulm University, Germany
[2] Institute of Media Informatics, Ulm University, Germany
{nikola.bubalo,felix.schuessel,frank.honold, michael.weber,anke.huckauf}@uni-ulm.de

Abstract. Adaptive multimodal interaction requires assessing the user behavior. However, it is still unclear which variables exactly have to be sensed in order to optimally adapt the system behavior. In the current paper, we report a paradigm allowing to independently examine effects of task difficulty, a users previous experience and his success criteria on various important indicators of the choice of modality, effectivity, efficiency, workload, and user experience. First data show that the paradigm is suitable for the induction and examination of these factors.

1 Introduction

Adaptive multimodal interfaces (MMI) in human computer interaction (HCI) are intended to facilitate the interaction between users and technical systems. To this end, adaptive MMI are expected to be as flexible as possible to accommodate the widest range of users in the most possible range of settings. Recent adaptive MMI in HCI research are, in their design, focusing on browsing through and sorting of databases or websites. Respective scenarios can be traced back to certain cognitive tasks [4] which are similar to those in Bolts Put-that-there setup [3]. One key aspect are adaptive multimodal inputs which have already been investigated for some time [7, 8]. Even with this focus on multimodal input huge differences in results remain [14],[2]. This could be due to the fact that users themselves learn and change their behavior and preferences based on their experience [6]. Consequently, for an adaptive multimodal system to accustom the UI more efficiently to the user, it needs to monitor and predict the behavior of the user constantly. There is still a lack of knowledge about which determinants affect multimodal interaction behavior to which extend, and how these determinants interact with each other. Combination of speech and touch interfaces are still far from common which impedes studies with large subject numbers. As an alternative approach to the existing setups, we developed an experimental design which enables researchers to control a vast range of factors, mix them in the intended proportions and measure their effects on user behavior.

2 Testbed

We decided to use speech and touch input, which are common and frequent input modalities in HCI [5], as well as multimodal feedback (visual & auditory). In order to mimic respective processes from real life, where users have to search for the right button or piece of information on screen, users are presented with a matrix of colored geometric shapes (e.g. circles, squares, triangles) on a touch screen, which could be either of red, blue or green color (see figure 1). We designed a conjunction search task along those two feature dimensions (i.e. color and shape) in order to avoid pop-up effects by preventing top-down advance cuing of the target object [10]. Both feature dimensions manipulate the workload necessary to complete a task, which is supported by the notion that colors are best coded verbally while positional information is best coded spatially which can be based on various psychological models [15], [12]. The interactive system accepts touch and speech inputs while expecting two inputs to complete one interaction trial. Users detect the target object, which is a unique combination of color and shape (e.g. the single green triangle in the trial). Targets can appear at a random position in a matrix of distractors. A correct answer consists of indicating the position and the color of the target. All objects are labeled with increasing numbers for easier verbal reference.

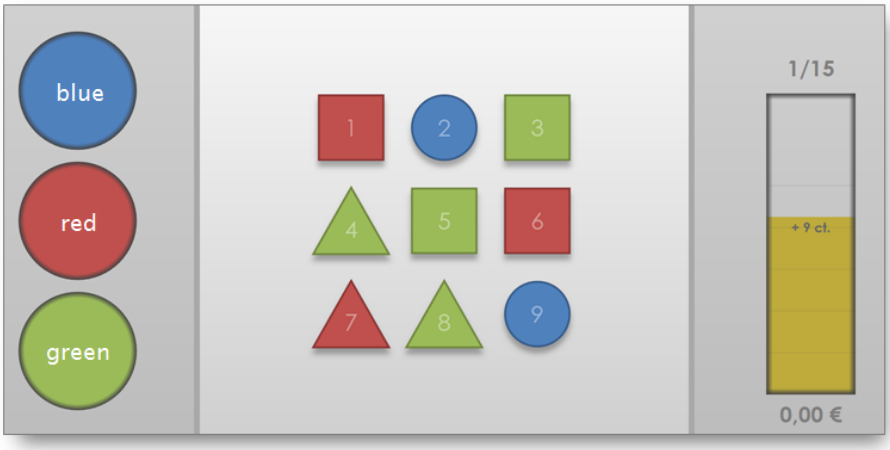

Fig. 1. Screenshot of a 3x3 matrix. In this example, the target (red object at position 7) had to be identified via two interactions. One interaction is used to select the color, the other is used to reference the position. Each interaction could be performed either via touch or via speech. No temporal order was given in execution of these two interactions.

A timer, indicating the remaining time left and the points to be won in that trial, can be shown on the side of the screen corresponding to the user's

dominant hand. On the other side of the screen, three round buttons in the three afore-mentioned colors are depicted and labeled with their corresponding color names. All objects are comfortably reachable. If not instructed otherwise, the task can be completed either exclusively by touch (touching the object and the corresponding color button), exclusively through speech (naming the number of the object and its color), or a combination of those modalities (touch object and name color or vice versa). The modality choices, reaction times and errors of the users can be recorded. Furthermore, the order, duration and temporal relationships of individual inputs can be recorded as well.

The optional inclusion of an induction phase, facilitates the investigation of previous user experience on behavior. The cognitive load of the interface is kept as low as possible by only displaying the absolutely essential information and the appliance of gestalt laws in the design. Our testbed enables the manipulation and examination of the effects of cognitive load on user behavior and performance through task difficulty, which can be varied by varying the number of distractors and thus the matrix size (e.g. 3x3, 4x4, 5x5).The testbed enables the gradual manipulation and measurement of user engagement with the system. Furthermore the combination of clearly defined previous user experience with the concept of an ideal modality choice for this task facilitates the opportunity to explore the trade-off mechanics between those two factors.

3 Study

The experiment consisted of two blocks of which the first block had to be solved with a specific modality combination (e.g. color via touch + position via speech). It was aimed to induce a specific interaction history with the system. This induction block entailed 90 trials, segmented into three equal parts with increasing task difficulty (30 trials for each difficulty). The second block of the experiment consisted of 45 trials with a balanced but randomized sequence of different levels of difficulty. The user was free to choose which modalities to use in whichever temporal order (free choice block). We made sure that the experiment and the informed consent were in accordance to the WMA Declaration of Helsinki [1]. In total, 42 volunteers, 33 male and 9 female, took part in the experiment. After signing an informed consent each user was introduced to the system and the game mechanics before the experimental block started. They were equally distributed into the four induction conditions and compensated with money to which the individually achieved amount in the game was added. 11.9% of the users were left handed. The average age of the users was 27.17 years (SD = 9.18). 95.2% of the users had prior experience with multi-touch interfaces like smartphones or tablets. After both the induction block and the free choice block users were asked to fill out the NASA TLX questionnaire, which measures workload. At the end of the experiment, each user was asked for the preferred kind of interaction, could give feedback on the experiment, and was payed.

4 Results

Over all in the free choice block, a preference for *color speech - position touch* was observed with 51% of all choices, whereas the opposite, *touch color - speech position* was rarely chosen with 3.3%. The exclusive interaction modes are comparable in frequency: *exclusive touch* 23.9% of cases, and *exclusive speech* 21.8% of cases. Taking into account the induced user interaction history, the picture changes remarkably and varies between interaction history groups. *Exclusive touch* was used 99.2% more often when induced in advance, *touch color - speech position* 236% and *color speech - position touch* 45.5%. That is, in all but the *exclusive speech* condition (-4.1%) users employed the induced interaction mode more frequently than the other groups of users.

The workload ratings after the induction block were not significantly different between the four interaction mode groups ($F(3) < 1$). Consequently, we have to proceed with the assumption that the kind of interaction modes used did not affect the subjective workload. Looking at performance measures divided by task difficulty however unfolds another picture. With rising difficulty users made more mistakes and took longer to complete a trial. Interestingly none of the users with previous experience in *exclusive speech* and *exclusive touch* chose to use *color touch - position speech* and users of the *color speech - position touch* group tried it a few times but did not use it in the easy condition. Additionally, using *color speech - position touch* resulted frequently in the lowest error rates. Both observations supports the concept of an ideal task specific modality (i.e. *color speech - position touch*).

The overall error rates (i.e., independently of user interaction history) of users show that *color speech - position touch* holds the lowest error rate with 4.75%, followed by *touch color - speech position* with a 9.2% error rate. Using *exclusive touch* resulted in a 9.1% error rate, while *exclusive speech* holds the most errors with 17%. Taking the interaction history into account, users made significantly less errors when using the familiar interaction mode ($F(3) = 26.6, p < 0,001$). In case of *color speech - position touch* however, the untrained subjects performed better than those who had trained it ($t = 3.15, df = 40, p = 0.003$). This could be explained by the fact that users in general made the least errors in this condition, and thus training had no improving effect.

Averaged over induction modes, task tompletion times (TCTs) did differ significantly between interaction modes ($F = 52.8, p < 0.001$). *Exclusive touch* took the longest on average and significantly longer than the other three interaction modes ($p < 0.001$). Again, taking the induction into account, users performed significantly faster with the familiar interaction mode than with an unfamiliar one. In case of *touch color - speech position* users who were not familiar with this interaction mode performed faster than those who did.

Our results show that users were significantly faster when the familiar modality was used, except for the *color speech - position touch* condition.

5 Discussion

The study demonstrates how this experimental setup can be used to examine various potential determinants in an adaptive MMI. The study shows that previous user experience does have a strong systematic effect on user behavior and performance. For example, the induced interaction experiences in this study and its effect on modality choice does interact with the tendency to use the optimal modality to solve the visual search tasks [9, 13]. The second major determinant of user behavior with MMI is the difficulty of the task to be completed, ergo the cognitive load of the interaction. The higher the cognitive demand of the interaction the more likely users will switch from unimodal to multimodal interaction [11]. Consequently, the detailed examination of this determinant, and its effect on the timing and frequency of strategy changes by the user, are essential to the development of adaptive MMI. One example for an investigation would be the controlled manipulation of mental workload in the present study. Another very important determinant of user behavior is the focus of his engagement. With our paradigm user engagement can be directed, enhanced or diminished in a controlled way. The choice of modality, input speed and accuracy of a user are effective measures, which come included with every adaptive MMI by design and need no additional sensors. Within our paradigm they can be used to detect subtle effects from factors and their interactions due to the control it provides over determinants. The clarity and flexibility of this paradigm facilitates the discovery and modeling of laws and principles governing multimodal interaction. These would then be incorporated into adaptive algorithms and validated in more ecologically valid setups.

References

1. Association, W.M., et al.: World medical association declaration of helsinki. ethical principles for medical research involving human subjects. Bulletin of the World Health Organization 79(4), 373 (2001)
2. Bellik, Y., Rebaï, I., Machrouh, E., Barzaj, Y., Jacquet, C., Pruvost, G., Sansonnet, J.P.: Multimodal interaction within ambient environments: an exploratory study. In: Human-Computer Interaction–INTERACT 2009, pp. 89–92. Springer (2009)
3. Bolt, R.A.: Put-that-there: Voice and gesture at the graphics interface, vol. 14. ACM (1980)
4. Carter, S., Mankoff, J., Klemmer, S.R., Matthews, T.: Exiting the cleanroom: On ecological validity and ubiquitous computing. Human–Computer Interaction 23(1), 47–99 (2008)
5. Cohen, P.R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., Clow, J.: Quickset: Multimodal interaction for distributed applications. In: Proceedings of the fifth ACM international conference on Multimedia. pp. 31–40. ACM (1997)
6. Domjan, M.: The principles of learning and behavior. Cengage Learning (2014)
7. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: A survey. Computer vision and image understanding 108(1), 116–134 (2007)

8. Lalanne, D., Nigay, L., Robinson, P., Vanderdonckt, J., Ladry, J.F., et al.: Fusion engines for multimodal input: a survey. In: Proceedings of the 2009 international conference on Multimodal interfaces. pp. 153–160. ACM (2009)
9. Mignot, C., Valot, C., Carbonell, N.: An experimental study of future natural multimodal human-computer interaction. In: INTERACT'93 and CHI'93 Conference Companion on Human Factors in Computing Systems. pp. 67–68. ACM (1993)
10. Müller, H.J., Krummenacher, J.: Visual search and selective attention. Visual Cognition 14(4-8), 389–410 (2006)
11. Oviatt, S., Coulston, R., Lunsford, R.: When do we interact multimodally?: cognitive load and multimodal communication patterns. In: Proceedings of the 6th international conference on Multimodal interfaces. pp. 129–136. ACM (2004)
12. Paivio, A.: Mental representations: A dual coding approach. Oxford University Press (1990)
13. Ratzka, A.: Explorative studies on multimodal interaction in a pda-and desktop-based scenario. In: Proceedings of the 10th international conference on Multimodal interfaces. pp. 121–128. ACM (2008)
14. Ren, X., Zhang, G., Dai, G.: An experimental study of input modes for multimodal human-computer interaction. In: Advances in Multimodal InterfacesICMI 2000, pp. 49–56. Springer (2000)
15. Wickens, C.D.: Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society 50(3), 449–455 (2008)