In-Depth Analysis of Multimodal Interaction: An Explorative Paradigm

Felix Schüssel[1(✉)], Frank Honold[1], Nikola Bubalo[2], Anke Huckauf[2], Harald Traue[3], and Dilana Hazer-Rau[3]

[1] Institute of Media Informatics, Ulm University, Ulm, Germany
{felix.schuessel,frank.honold}@uni-ulm.de
[2] Department of General Psychology, Ulm University, Ulm, Germany
{nikola.bubalo,anke.huckauf}@uni-ulm.de
[3] Section of Medical Psychology, Ulm University, Ulm, Germany
{harald.traue,dilana.hazer}@uni-ulm.de
https://www.uni-ulm.de/in/mi.html
https://www.uni-ulm.de/en/in/psy-paed
http://www.emotion-lab.org

Abstract. Understanding the way people interact with multimodal systems is essential for their design and requires extensive empirical research. While approaches to design such systems have been explored from a technical perspective, the generic principles that drive the way users interact with them are largely unknown. Literature describes many findings, most of them specific to certain domains and sometimes even contradicting each other, and thus can hardly be generalized. In this article, we introduce an experimental setup that – despite being rather abstract – remains generic and allows in-depth exploration of various aspects with potential influence on users' way of interaction. We describe the gamified task of our setup and present different variations for empirical research targeting specific research questions. Applying the experimental paradigm offers the chance for new in-depth insights into the general principles and influencing factors of multimodal interaction, which could in turn be transferred to many real-world applications.

Keywords: Multimodal interaction · Experimental paradigm · Empirical research · Interaction histories · Pressure of time and success · Cognitive load

1 Introduction

Multimodal interaction has been a topic of research for some while now. There has been a lot of progress concerning how to model and process multimodal inputs. Still, little is known about the generic principles that apply, e.g. the choice of modalities, the temporal relations of multimodal inputs, and what may be an even more important factor, the contextual parameters that influence multimodal interaction. To tackle these questions, we have designed an abstract, but still generic, experimental paradigm that allows the exploration of these questions

M. Kurosu (Ed.): HCI 2016, Part II, LNCS 9732, pp. 233–240, 2016.
DOI: 10.1007/978-3-319-39516-6_22

in a flexible but controlled manner. Based on the developed paradigm, individual applications are generated and applied in different experimental setups.

2 Related Work

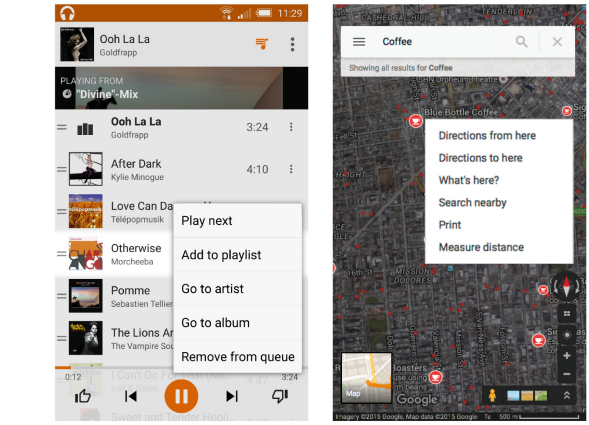
There have been a number of approaches on how to design multimodal interfaces and on how to model multimodal inputs from a system's perspective (see [5] for an overview). A more generic perspective on multimodal interaction is examined by Turk [13]. Two of the open challenges stated therein are a thorough understanding of the issues relating to the cognitive load of users, and the development of better guidance and best practices for the design and evaluation of multimodal systems (ibid.). Tackling these challenges requires empirical evidence. Accordingly, there has been a lot of empirical research in multimodal interaction, mostly specific to a certain domain, including map interactions [4,7–9], augmented reality [6], image manipulation [2], and music players [3].

Comparing the results of these studies reveals considerable differences. Although the domain and tasks in the work of Oviatt et al. [7–9] and Haas et al. [4] are quite similar, their results are in parts contradictory. While the former reports on users predominantly showing a simultaneous use of modalities, the latter reports on no users showing a simultaneous use of modalities. Similarly, the dependency on task difficulty remains ambiguous. The findings of [2,6] are even more specific to their respective domains. Although these provide some insights, their generalizability and transferability to other applications seems doubtful. Dumas et al. take a broader perspective and present a test bed for the evaluation of fusion engines using a music player as example [3]. They conclude that more work is necessary on fusion engine's adaption to context (i.e. environment and applications), as well as usage patterns and repetitive errors. This shows, that basic research on universal principles, which govern common tasks found in many applications, is still rare.

One aspect of the context is the influence of time pressure and the pressure of success onto the interaction behavior of a user. Getting the right ticket at the ticket vending machine in the train station last minute before the train leaves would be an example for such a situation. Including game elements to the study enables the simulation of such pressures on the user in laboratory settings. Respective gamification methods include feedback [1] on success and time pressure as well as a reward system [12]. These elevate both the intrinsic and extrinsic motivation of the user to complete the given tasks as reasoned by [10] based on the self determination theory.

3 A Visual Search Task for Empirical Research

In search of a task that is common to many different applications, we can identify *operations on objects* as a joint characteristic. Figure 1 shows different application examples.



(a) A music player (b) A map application

(c) A fitness application

Fig. 1. Different applications allowing operations on presented objects.

These kinds of tasks are found throughout many applications and are thus chosen for our research. Empirical research poses additional requirements as well, e.g. tasks must be performed repeatedly without becoming routine or dull, and participants' motivation must be kept high throughout the course of an experiment. In order to remedy these issues, we chose to use a gamified version of the task. In matters of the domain the tasks should take place in, we decided to use abstract representations of objects and operations.

Our solution is a visual search task, where the user has to identify the visually unique object and then specify its location and color (as a replacement for an arbitrary operation). Figure 2 shows a screenshot of the game. In the central area of Fig. 2, objects with differing shapes and colors are presented. In the given example, the green rectangle on position 3 is the unique object to be spotted by the user. The expected input can be provided either by using exclusive touch, exclusive speech, exclusive mouse or a combination of those modalities (e.g. touching the object and naming its color or vice versa).

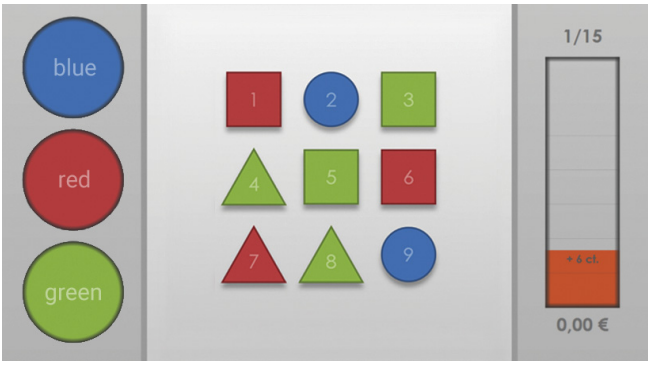


Fig. 2. Screenshot of the game that serves as abstract replacement of operations on objects found in many applications. The user has to spot the single unique object and designate it's location and color. In the above screenshot, the unique object is the red triangle. (Color figure online)

4 Planned Research

The generic design of our setup enables the investigation of isolated factors such as the users previous experience, contextual parameters, and cognitive demand. The following sections provide further details on how the presented experimental paradigm can easily be adjusted to facilitate the respective research. Although they are based on the same paradigm, different setups are used for each focus of research. Where applicable, first results are presented as well.

4.1 A User's Previous Experience: Individual Interaction Histories

In order to investigate the influence of individual user-centered interaction histories, the experimental paradigm is applied as shown in Fig. 3. The applied modalities are speech and touch inputs in any possible multimodal combination as described in Sect. 3. The inclusion of an *induction phase*, which requires users to solve the tasks applying only one of the possible four modality combinations, enables the investigation of the influence of individual interaction histories in the *free interaction phase*. We are particularly interested in the modality preferences of the free interaction phase, depending on the induced modality combination. Is there a favorite modality combination (regarding to error rates) and how long does it take users to apply it when induced otherwise? This could provide insights on the learning behavior of multimodal inputs.

Additionally, this experimental setup allows for an in-depth analysis of temporal relations of multimodal inputs, particularly with regard to the contradicting findings of the related work concerning the predominance of simultaneous and sequential interaction patterns. Results of a user study with this setup are reported in [11]. It is shown that a classification into simultaneous and sequential users may not be feasible in general. Instead, a more differentiated inspection of individual behavior is proposed and possible uses are discussed (cf. [11]).

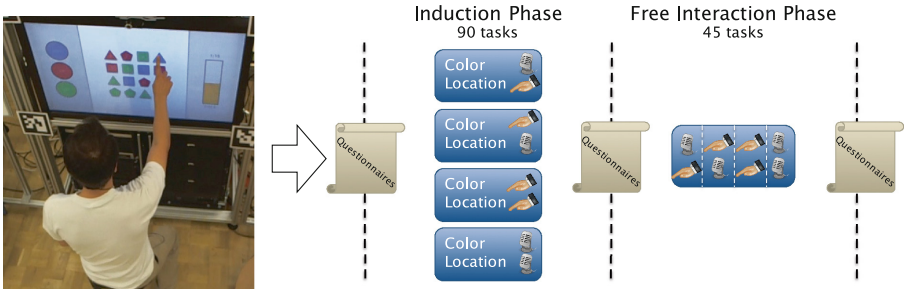


Fig. 3. The experimental procedure to investigate individual interaction histories. The *induction phase* induces a certain modality combination for each subject. In the *free interaction phase*, users can perform inputs in any modality combination.

4.2 Contextual Parameters: Pressure of Success and Time

Regarding contextual parameters, we investigate the influence of pressure to succeed in the task and time pressure, by varying the reward system and the available time to complete a task. These factors are supposed to have a significant influence not only on the error rate, but also on the way people interact with a system. To this end, an experiment was conducted in which we compared two groups of subjects which differed in the amount of auditory and visual feedback given by the system as well as the monetary reward given for the participation. Contrary to the *Feedback* group, the *No-Feedback* group got no auditory or visual feedback whether their input was correct, no timer was presented and consequently no performance dependent monetary rewards were given. Both groups underwent the same experimental procedure (see Fig. 4). Preliminary results indicate that users in the *Feedback* condition try to increase their success by interacting significantly faster than the *No-Feedback* group at the expense of significantly higher error rates. Furthermore, users from the *Feedback* condition chose multimodal interaction (33 % of trials) more often compared to the *No-Feedback* group (29.7 % of trials). Given that the *Feedback* group earns significantly higher monetary rewards, this difference in interaction behavior proves to be an effective way to increase success under pressure.

Regarding the in-depth analysis of the temporal patterns of interaction, temporal interaction parameters like modality overlaps and individual durations are measured under very different contextual conditions while the task is held fixed. We hypothesize that temporal interaction patterns become shorter when the users are under pressure. This could have implications on the fusion of user inputs and their adaption to context within the same application. Preliminary results suggest that the users do indeed act faster in the pressure condition. To be more specific, the temporal overlap of modalities decreases, while the durations of each modality themselves remain almost the same.

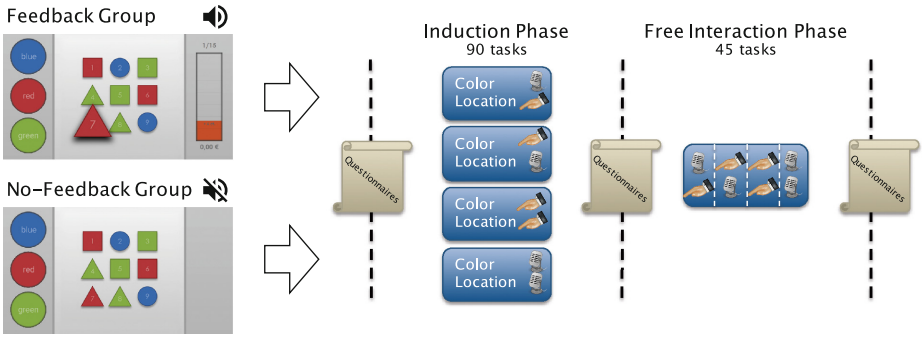


Fig. 4. The experimental procedure to investigate the effects of time pressure and pressure to succeed. One group is set under pressure (*Feedback*), while the other is not (*No-Feedback*). In the *induction phase*, each subject is restricted to use specific modalities. In the *free interaction phase*, users can perform inputs using any modalities. (Color figure online)

4.3 Cognitive Load: Induction of Overload and Underload

Given that users can be overwhelmed by the options and corresponding operations presented to them, diminishing the users' satisfaction with the system in general and thus affecting the user-system interaction, we intend to investigate the effects of cognitive load. Based on the present paradigm, an experiment was conducted to induce cognitive overload and underload in the subjects and investigate their effects by analyzing the users' individual reactions and subjective feedbacks.

The induction of cognitive overload and underload is generated by varying the number of objects and their colors within a task as well as the available given time to solve that task. These variables influence the difficulty of a given task and also affect the user's interaction with the system. Cognitive overload is induced by increasing the task field objects and colors as well as decreasing the available time, while cognitive underload is induced by decreasing the task field objects and colors and increasing the available time. Figure 5 depicts the two variants and the overall experimental procedure.

The interaction modality used during the induction phase can be either speech or mouse and is defined at the beginning of the experiment. Standardized questionnaires are filled by the subjects prior starting the experiment. Further, various kinds of subjective feedbacks including free speech, emotional rating and direct questions as well as baseline breathing phases are also implemented.

In order to enable an easy-to-handle workflow, the course of events within the experiment as well as the used modalities may be completely managed through an external *task set*. Within a task set, the workflow setting of the sequences can be defined individually for every task and every subject, allowing a high flexibility and generalization of the course setup.

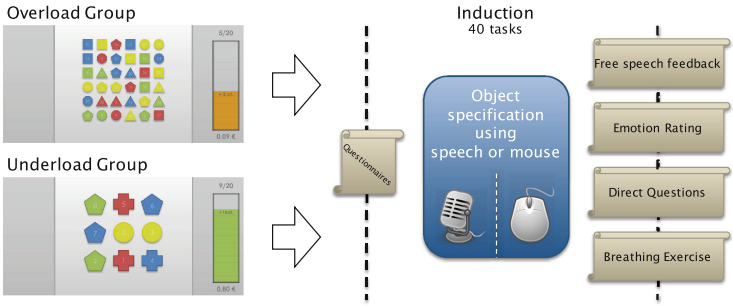


Fig. 5. The experimental procedure to induce and investigate the effects of cognitive load. Both groups (*Overload* and *Underload*) undergo the same procedure, while cognitive load is increased by increasing the task field objects and colors as well as decreasing the time. The modality during the induction phase is set up at the beginning of the experiment. (Color figure online)

5 Conclusion

The presented experimental paradigm enables a controlled investigation of the general laws and principles associated with multimodal interaction and the cognitive load of users, while the results are kept generalizable to a vast number of different implementations of multimodal interaction. Based on the paradigm, we presented three implemented setups covering a broad range of research topics. This includes an investigation of the role of users' previous multimodal interaction experience (their so-called interaction history). The resulting insights into the individuality of multimodal temporal relations will help to improve the fusion of inputs in future systems [11]. The second implementation shows that the influence of contextual parameters like pressure of success and time can be examined by slightly varying the provided feedback. Using fine grained variations of the task's difficulty, one gains control over the amount of cognitive demand imposed on the users, reaching from underchallenged to well overstrained.

In addition to such flexibility, the presented paradigm has several other advantages over using a specific real-world application for research, such as its easy implementation, the possibility to deploy it on different hardware setups with different modalities, as well as its suitability for lengthy laboratory studies with a lot of repetitions due to its gamified design. Thus, it allows researchers to meet the goal of gaining knowledge of multimodal interaction that is diverse and in-depth, yet still generalizable.

Acknowledgments. This work was supported by the Transregional Collaborative Research Center SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems", which is funded by the German Research Foundation (DFG). It is also supported by a Margarete von Wrangell (MvW) habilitation scholarship funded by the Ministry of Science, Research and the Arts (MWK) of the state of Baden-Württemberg for Dilana Hazer-Rau. Some icons from Fig. 1 were designed by www.Freepik.com.

References

1. Cheong, C., Cheong, F., Filippou, J.: Quick quiz: A gamified approach for enhancing learning. In: PACIS, p. 206 (2013)
2. Dey, P., Madhvanath, S., Ranjan, A., Das, S.: An exploration of gesture-speech multimodal patterns for touch interfaces. In: Proceedings of the 3rd International Conference on Human Computer Interaction, IndiaHCI 2011, pp. 79–83. ACM, New York (2011). http://doi.acm.org/10.1145/2407796.2407808
3. Dumas, B., Ingold, R., Lalanne, D.: Benchmarking fusion engines of multimodal interactive systems. In: Proceedings of the 2009 International Conference on Multimodal Interfaces, ICMI-MLMI 2009, pp. 169–176. ACM, New York (2009)
4. Haas, E.C., Pillalamarri, K.S., Stachowiak, C.C., McCullough, G.: Temporal binding of multimodal controls for dynamic map displays: A systems approach. In: Proceedings of the 13th International Conference on Multimodal Interfaces, ICMI 2011, pp. 409–416. ACM, New York (2011). http://doi.acm.org/10.1145/2070481.2070558
5. Lalanne, D., Nigay, L., Palanque, P., Robinson, P., Vanderdonckt, J., Ladry, J.F.: Fusion engines for multimodal input: a survey. In: Proceedings of the 2009 International Conference on Multimodal Interfaces, ICMI-MLMI 2009, pp. 153–160. ACM, New York (2009). http://doi.acm.org/10.1145/1647314.1647343
6. Lee, M., Billinghurst, M.: A wizard of oz study for an ar multimodal interface. In: Proceedings of the 10th International Conference on Multimodal Interfaces, ICMI 2008, pp. 249–256. ACM, New York (2008). http://doi.acm.org/10.1145/1452392.1452444
7. 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, ICMI 2004, pp. 129–136. ACM, New York (2004)
8. Oviatt, S., Coulston, R., Tomko, S., Xiao, B., Lunsford, R., Wesson, M., Carmichael, L.: Toward a theory of organized multimodal integration patterns during human-computer interaction. In: Proceedings of the 5th International Conference on Multimodal Interfaces, ICMI 2003, pp. 44–51. ACM, New York (2003)
9. Oviatt, S., Lunsford, R., Coulston, R.: Individual differences in multimodal integration patterns: what are they and why do they exist? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2005 pp. 241–249. ACM, New York (2005)
10. Ryan, R.M., Rigby, C.S., Przybylski, A.: The motivational pull of video games: A self-determination theory approach. Motiv. Emot. **30**(4), 344–360 (2006)
11. Schüssel, F., Honold, F., Weber, M., Schmidt, M., Bubalo, N., Huckauf, A.: Multimodal interaction history and its use in error detection and recovery. In: Proceedings of the 16th ACM International Conference on Multimodal Interaction, ICMI 2014, pp. 164–171. ACM, New York (2014)
12. Smith, A.L., Baker, L.: Getting a clue: creating student detectives and dragon slayers in your library. Reference Services Review **39**(4), 628–642 (2011)
13. Turk, M.: Multimodal interaction: A review. Pattern Recogn. Lett. **36**, 189–195 (2014)