Brain activity and eye-movements: Multimodal interaction in human-machine systems

Jeronimo Dzaack, Thorsten Zander, Roman Vilimek, Sandra Trösterer und Matthias Rötting

Keywords: Multimodal interaction, tool, EEG, eye-movements, human-machine systems

Zusammenfassung

Dieser Beitrag beschreibt den Ansatz der Multimodalen Mensch-Maschine-Interaktion. Es wird eine Schnittstellen-Definition und Realisierung vorgestellt, die eine Vielzahl von simultanen Eingabemöglichkeiten verarbeitet. Anhand einer konkreten Studie, die Augenbewegungen und Hirnsignale als verschiedene Eingabemodalitäten nutzt, wird der Nutzen von Multimodaler Interaktion belegt.

Abstract

This Paper describes the approach of multimodal human-machine interaction. It presents a definition and implementation of an interface framework, which allows for processing multiple instances of input. In a study on the combination of eye movements and brain signals as input, the use of multimodal interaction is proven.

Introduction

The physical exchange of information between a computing system and a human (e.g. light or sound) can be described as human-computer interaction. This interaction is often limited to a small set of modalities and devices (e.g. mouse, keypad, display). But the interaction of humans with an environment or with other humans involve not only a small set of modalities, but all human senses to perceive, interact and act in everyday situations. This multimodality and the freedom of choosing a suitable modality or device are missing in almost all computing systems.

In recent research several approaches exist to overcome these shortfalls by integrating more modalities and devices into computing systems, e.g. gesture, eye movements or speech. However, most of these approaches utilize only one and the same modality for input and output (unimodal interactive systems, Bernsen, 2008, p. 8). To enable a natural and seamless interaction of humans with computing systems, new interaction concepts and techniques need to be developed and evaluated. One of the key aspects is the use of at least two different modalities for input and output (multi-modal interactive systems, Bernsen, 2008, p. 8). This enables the user to access complex functionalities as well as information contents in an easier way than traditional systems and takes into account the users' native expectations and behavior. The combined use of eye movements to navigate on a computer display and brain activity to select targets is an example of multimodal human-computer interaction.

There are a number of challenges and possibilities connected with the development and evaluation of systems that provide multimodal interaction strategies. The enhancement of traditional computing systems by additional modalities and devices aims at the improvement of the information bandwidth between the user and the system to provide an interaction setting that is comparable to human means of communication. Thus, for the development and evaluation of multimodal computing systems not only theoretical and engineering points of view need to be considered but also human factors (Maybury, 1991).An important improvement of multimodal

interactive systems is the free combination and use of modalities by the user. To enable this freedom and to meet the users' expectations several design principles need to be met (for an overview see Raman, 2003 and Reeves et al., 2004): (1) multiple modalities need to be synchronized to allow redundant and parallel interaction, (2) multiple modalities should share a common interaction state to enable e.g. switching of modalities and multi-device interaction, (3) multimodal systems should degrade gracefully to enable the use of supplementary and complementary modalities in the case of changing capabilities of the user (e.g. change from mobile to stationary use of a computing system), (4) effects and behavior of multimodal systems should be predictable by the user in order to choose the appropriate modality in a given context and (5) multimodal computing systems should adapt their functionality to the user's state and environment (e.g. detecting the need for hands-free operation).

Due to this and the increasing complexity of computing systems and new technologies, there is a need for new concepts that support the development and evaluation of multimodal human-machine interaction.

iCOMMIC

To support the development and evaluation of multimodal human-computer interaction we engineered the integrated controller for multimodal interaction (iCOMMIC). iCOMMIC is an integrative software framework for multimodal human-computer interaction. It enables (1) to integrate different input and output modalities, (2) to combine these modalities in any way and (3) to record user data in an appropriate way for psychological investigations. It allows interaction designers to connect e.g. the output of an eye tracker to the cursor or an external light bulb to signal a specific behavior as well as a gesture input to a language generation system.

iCOMMIC was developed in a user centered, parallel and iterative process model (Timpe & Kolrep, 2002). For this purpose we conducted interviews with experts from the field of human-computer interaction and with future users to survey requirements regarding the development and evaluation of multimodal computing systems (i.e. we conducted an initial task analysis). Based on the requirements we designed a conceptual framework for iCOMMIC, followed by implementing the tool and conducting verification and validation studies.

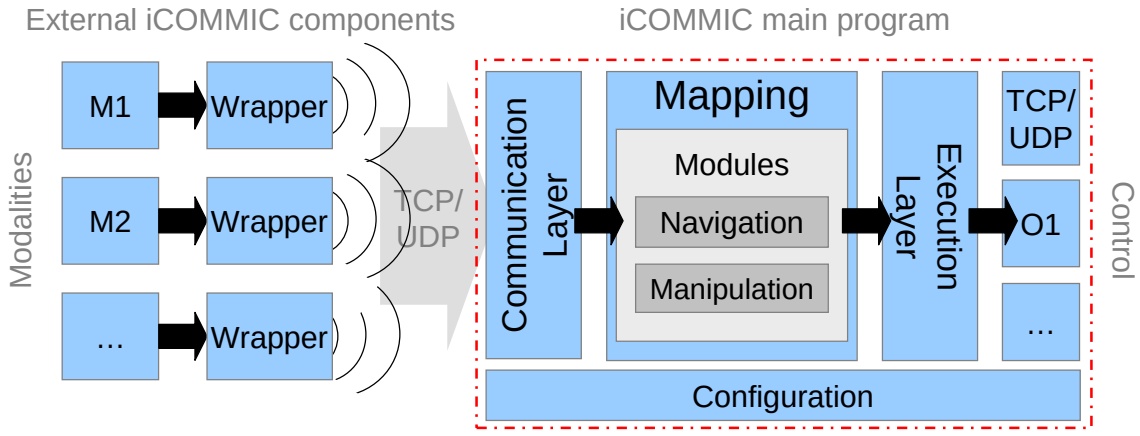

Fig. 1: Data flow in iCOMMIC: connection of different modalities (M) (left side), mapping of modalities to the output devices (O) of the computer (middle), and execution of control commands (right side).

Design and Functionality

iCOMMIC is designed to allow a user (1) to interact with a computing system in a multimodal way, (2) to freely choose the number of modalities, (3) to combine them and (4) to switch between their use. Additionally, an experimenter is enabled to design experimental settings in the context of multimodal user interfaces and to analyze the users' behavior based on the logged data. For this purpose we designed iCOMMIC in three independent components: (1) capturing and connection of different modalities, (2) mapping of modalities to devices and (3) execution of control commands by the computing system (see Fig. 1).

The first component allows the integration of different modalities into iCOMMIC (e.g. hand-position by mouse, eye-position by eye-tracker, gesture by gesture recognition). This is done by wrapping the output signals of the devices by a standardized data format and sending this data to the main application, using a network protocol. For each input device a wrapper can be implemented individually and in a customized manner. The advantage of the wrapper concept is that different interaction techniques and filter functions can already be realized in the wrapper in a preprocessing phase. It would be possible e.g. to implement different filter functions for gaze data that differentiate between intended blinks or normal eye lid activity or between saccades and fixations. The incoming data can be interpreted by the second component and, according to predefined rules by the user, mapped to existing devices of the computing system. For example, a user can use the data gained from an eye tracker to navigate on the screen, but use a mouse click or a gesture to manipulate the element, which a user looks at. The last component ensures the execution of the command in the computing system. It transfers the control commands of the mapping process to a command that is executable by the particular computer device.

Evaluation and Correction

The verification of iCOMMIC was organized in three evaluation levels that build upon each other to verify that (1) all requirements are met by iCOMMIC and that (2) the software is correctly engineered. During the first phase of evaluation, the internal software modules (i.e. classes) were tested in a module test to show that the basal functionality is realized. Subsequently the interaction between the software modules was tested in an integration test. For this purpose logic arrays were combined to form system blocks. In the last step these system blocks were integrated into a system and evaluated as a whole (i.e. system test).

For each evaluation level, several use cases and requirements were defined. All identified errors were corrected by readjusting the software and the framework to the results of the verification if necessary. After completing the verification, we found that iCOMMIC was developed in a correct manner and met all given requirements.

Regarding user aspects, we conducted a validation study integrated into an experimental study to validate that the right system was engineered for the original purpose (i.e. conducting experimental studies in the field of multimodal computing systems). In this experimental study we found that all requirements regarding the user – as found in the initial task analysis - were met by iCOMMIC (Dzaack et al. 2009).

BC(eye)

This study is aimed to provide an example that the combination of different modalities allows for combining the benefits of both and eliminating their flaws. We will define an application of multimodal interaction to solve the Midas-Touch problem (Jacob, Legett, Meyers and Pausch 1993). With the idea of "eyes as output" Richard Bolt introduced eye-gaze input to facilitate human-computer interaction already in 1982. However, whereas moving the mouse cursor with eye movements is quite intuitive, it is not that easy to find a good mechanism for performing

the click operation (Midas-Touch Problem). Most solutions are based on dwell times, i.e. the user has to fixate an item for a pre-defined period of time in order to activate it. This technique has to face the inherent problem of finding the optimal dwell time. If the dwell time is too short, click events will be carried out unintentionally leading to errors. If it is too long, fewer errors will be made but more experienced users will get annoyed.

In our study we developed and evaluated a completely different approach: using a EEG-based Brain-Computer Interface (BCI) (Vidal, 1973) to confirm object selections made by eye tracking. Brain-computer interaction and eye-gaze input can be regarded as complementary modalities in the respect that they compensate for each other's disadvantages. The combination overcomes the BCI drawback of having problems differentiating between more than two commands because only one activation thought needs to be tracked reliably. If this activation works properly, a new solution to perform click operations in gaze-based interfaces can be established by providing an explicit, nevertheless not overt visible, command under complete user control.

Experimental Evaluation of BC(eye)

This experiment compares a BCI-based activation of targets in an eye-controlled selection task against two conventional dwell time solutions with different activation latencies. Task difficulty in the selection task was varied by showing either simple visual stimuli with only a few random characters or by presenting more complex visual stimuli featuring a higher number of characters. Two different dwell times, short and long, were chosen for a better representation of the range of typical interaction situations with gaze-controlled applications. Assuming that signal extraction and pattern recognition of current BCIs still need a substantial minimal presence duration of the activation thought and that processing these signals takes additional time, it does not seem very likely that subjects will be able to complete tasks with the BCI faster. The question of interest here is whether they are significantly slower with a BCI than with dwell times. The activation thought via BCI is a conscious, explicit command – in contrast to the implicit commands of dwell time solutions. Thus, the error rate in the BCI condition should be substantially lower, especially for difficult selection tasks.

Tasks

The participants had to perform a search-and-select task. They were presented with stimuli consisting of four characters in the "easy" condition and seven characters in the "difficult" condition. The reference stimulus was displayed in the centre of the screen. Around this item twelve stimuli were shown in a circular arrangement, eleven distractors and one target stimulus, which was identical to the reference stimulus. The radial arrangement of search stimuli ensured a constant spatial distance to the reference stimulus. All search strings consisted of consonants only. The distractors shared a constant amount of characters with the target.

Subjects had to select the target stimulus by either fixating it for the given dwell time or by thinking the activation thought. The selection criterion was that the short version is still well controllable and that the long activation latency is not perceived as slowing down the user. The short dwell time was 1.000 milliseconds, the long dwell time 1.500 milliseconds.

For BCI activation, the participants had to imagine closing both hands to fists and then to turn them against each other like when wringing out a cloth by twisting it tightly. They were told not to involve any overt muscular activity.

Results and Discussion

Time needed for task completion and accuracy (data on errors) were averaged across all subjects for each selection method and level of search difficulty. Trials with errors were not included in the analysis of response time. First, an analysis of variance was conducted on the results (alpha <0.05). In a second step, the data of the easy and difficult condition were pooled

for each selection method. To avoid any problems associated with multiple testing, differences will be regarded as significant with an alpha level of .025 for these comparisons.

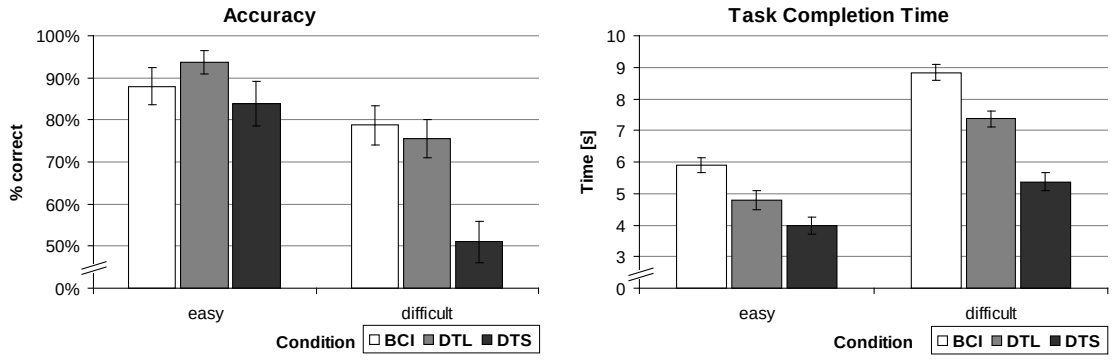

Fig. 2: Percentage of correct selections (left): Brain-Computer Interface (BCI), long dwell times (DTL) and short dwell times (DTS). Task completion times (right): Brain-Computer Interface (BCI), long dwell times (DTL) and short dwell times (DTS).

The accuracy data are summarized in Figure 2 (left) Remarkably, the BCI leads to the best results with 78.7% correct selections, although the difference to the long dwell time, 75.6%, is only marginal. An analysis of the main effects confirms general differences between the activation techniques ($F(2,18) = 12.47$, $p < .001$) and that the difficult search condition leads to more errors ($F(1,9) = 38.37$, $p < .001$).

The pooled BCI accuracy average is 83.3% correct selections; the corresponding values for dwell time long and dwell time short are 84.7% and 67.4%. Pairwise t-tests reveal that the better performance of the BCI compared to “dwell time short” is significant ($t(9) = 3.66$, $p = .005$). The small differences between BCI and “dwell time long” is not reliable ($t(9) = 0.33$, $p = .75$). As expected, the BCI allows users to activate (click) GUI items more precisely than a dwell time solution with short latencies.

Task completion was fastest in both search conditions with short dwell times (easy: 3.98 s; difficult: 5.38 s). Next was dwell time long (4.79 s; 7.37 s), leaving BCI the slowest method of activation (5.90 s; 8.84 s). This general difference between the input methods is statistically confirmed ($F(2,18) = 56.25$, $p < .001$). The results are depicted in Figure 2 (right).

9 of 10 subjects stated that they would prefer the BCI based solution in a productive working environment.

This study shows, that the current state of technology allows performing more accurate activations with BCI than with dwell time solutions with short latencies. For both stimulus complexities the multimodal solution performs as good as the optimal solution for that condition solely basing on eye movements. Hence, it neglects the problems one is confronted in the unimodal approach. Also, it allows the user to work self-paced, not dependent on a given dwell time. Hence, it allows for more natural interaction, as the selection modality is independent from the search modality. The result, that 90% of the users would select the here presented solution shows, that these effects really reach the user.

Discussion and Outlook

Providing multimodal interactive systems has several ad-vantages compared to unimodal interactive systems (e.g. free choice of modalities and their free combination in respect to a contextual setting). With the presented tool iCOMMIC we introduced a tool that supports the development and evaluation of multimodal interaction in computing systems for scientific research. Due to its modular set-up it can be extended easily and enhanced by further input or output

devices. The presented study in the second part showed that multimodal interaction is a powerful technology and allows incorporating different kinds of modalities into human-system interaction.

We believe that multimodal interaction and the technologies presented in this article provide a new base for the realization of new interaction concepts. These new concepts will enable the user to interact with a computing system in a more natural and efficient way, thereby using the advantages of the different modalities a human possesses.

Acknowledgements

We thank Thomas Nicolai for supporting us in programming and conception of iCOMMIC. The BC(eye) study is a cooperation of Siemens AG and TU Berlin. We thank Christian Kothe and Matti Gärtner for support in this study.

References

Bernsen, N. O. (2008). Multimodality theory. In: D. Tzovaras (Ed.), Multimodal user interfaces. *Signals and communication technology* (pp. 5-29). Heidelberg: Springer.

Bolt, R.A.: Eyes at the Interface. *Proceedings of the 1982 Conference on Human Factors in Computing Systems. ACM Press*, New York (1982) 360--362

Dzaack, J., Trösterer, S, Nicolai, T. & Rötting, M. (accepted). iCOMMIC: Multimodal Interaction in Computing Systems. *Proceedings of the 17th World Congress of the International Ergonomics Association*, Peking.

Maybury, M. 1991. Introduction. In: M. Maybury (Ed.) *Intelligent multimedia interfaces* (pp. 18-21). Cambridge, Mass: AAAI Press.

Jacob, R.J.K., Legett, J.J., Myers, B.A., Pausch, R.: Interaction Styles and Input/Output Devices. *Behaviour & Information Technology 12, 69--79 (1993)*

Raman, T. V. (2003). Design Principles for Multimodal Interaction. *MMI Workshop, CHI 2003*. Fort Lauderdale, Florida, USA.

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., Collings, P., Cohen, P., Kraal, B., Martin, J., McTear, M., Raman, T., Stanney, K. M., Su, H., and Wang, Q. Y. (2004). *Communications of the ACM, 47*, 1, 57-59.

Timpe, K.-P. & Kolrep, H. (2002). Das Mensch-Maschine-System als interdisziplinärer Gegenstand. In: K.-P. Timpe, T. Jürgensohn & H. Kolrep (Eds.) *Mensch-Maschine-Systemtechnik – Konzepte, Modellierung, Gestaltung, Evaluation* (pp. 9-40). Düsseldorf: Symposion Publishing.

Vidal, J. J. (1973) Toward direct brain-computer communication. Annual Reviews in Biophysics and Bioengineering, 2:157–180.