Influence of user's mental model on natural gaze behavior during human-computer interaction

Thomas Bader[1,2]
thomas.bader@kit.edu

[1]Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT),
Germany

Jürgen Beyerer[1,2]
juergen.beyerer@iosb.fraunhofer.de

[2]Fraunhofer IOSB
Intitut für Optronik, Systemtechnik und
Bildauswertung
Germany

ABSTRACT

Natural gaze behavior during human-computer interaction provides valuable information about user's cognitive processes and intentions. Including it as an additional input modality therefore provides great potential to improve human-computer interaction. However, the relations between natural gaze behavior and underlying cognitive processes still is unexplored to a large extend. In this paper we identify and characterize major factors influencing natural gaze behavior during human-computer interaction with a focus on the role of user's mental model about the interactive system in that context. In a user study we investigate how natural gaze behavior can be influenced by interaction design and point out implications for usage of gaze as additional modality in gaze-based interfaces.

Author Keywords

gaze based interaction, natural gaze behavior, multimodal interfaces

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces

INTRODUCTION

In general there are two ways to incorporate eye gaze as an input modality into multimodal human-computer interfaces. The first way is to force the user to consciously look at certain locations in order to trigger actions. One example for such approaches is eye typing, which has been studied for decades [9]. Eye gaze is used directly as pointing device and actions are mostly triggered by dwell times, which determine how long a certain object needs to be looked at until it is activated (e.g., a key on a virtual keyboard). The biggest advantages of such approaches are, that they are easy

2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction. February 13, 2011, Palo Alto, California, USA.

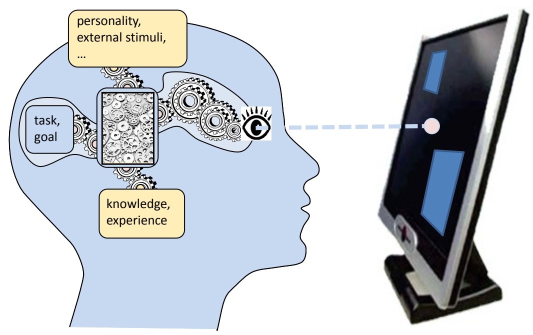
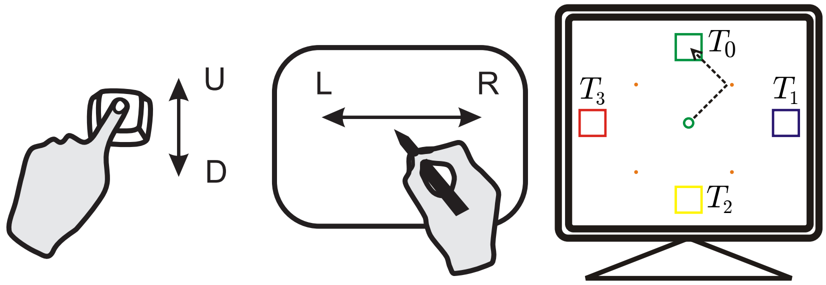


Figure 1. Dependency of natural gaze behavior

and straightforward to implement and do not require analysis of complex gaze behavior. Especially for people with severe disabilities such input techniques often provide the only way for interacting with visual interfaces. However, for most people conscious and direct usage of gaze as input modality is very unnatural and hence requires training and/or induces cognitive workload[5].

The second way to use eye gaze as input modality is to interpret natural gaze behavior during human-computer interaction, while using another modality as primary input modality. Promising examples for such interaction techniques are presented in [4] and [12]. In both approaches *natural* gaze behavior is analyzed and the user is not forced to diverge from that natural behavior for interaction purposes. *iDict* [4] analyzes the duration of fixations while the user reads a text in a foreign language and automatically provides a translation of the fixated word if a longer fixation is detected. In the approach "Manual And Gaze Input Cascaded (MAGIC) Pointing"[12] the mouse pointer is placed close to the currently fixated object in order to eliminate a large portion of the cursor movement. Both approaches do not use gaze directly as pointing or input device, but interpret gaze data in the context of the task (reading, pointing).

In general, the second approach has the advantage that valuable information contained in natural gaze behavior can be used for improving human-computer interaction. Addition-



(a) Key (b) Pen (c) Task

Figure 2. Input devices and task

ally, the user has not to consciously diverge from natural gaze behavior.

However, natural gaze behavior is highly complex and many different influencing factors have to be considered for appropriate interpretation (see Figure 1). Therefore, a thorough understanding of natural gaze behavior during human-computer interaction is necessary in order to incorporate it as input modality in intelligent user interfaces. It has been shown that the task and the experience of users are key factors influencing natural gaze behavior (e.g., in [6, 8]).

Numerous studies of natural gaze behavior and hand-eye coordination during manipulative activities in natural environments like block-copying [10], basic object manipulation [6], driving [7] and playing cricket [8] revealed gaze shifts and fixations to be commonly proactive (eye-movements occured previous to movements of the manipulated object or the manipulator). In addition, a detailed study on hand-eye coordination during an object manipulation task [6] revealed, that subjects almost exclusively fixated landmarks critical for the control of the task and never the moving object or hand. Such landmarks could be obstacles or objects in general that are critical for the completion of the task, like in [8] where batsmen concentrated on the ball, and not on their hands or the bat. These studies show, that natural gaze behavior is complex and determined by many different parameters (e.g., position of obstacles in [6] or previous experience of a person [8]).

Gaze behavior was also studied in various tasks related to HCI. In [11] results of a study on hand-eye coordination during a pointing task with different indirect input devices are described. The main finding of the study is that users used a variety of different hand-eye coordination patterns while moving the cursor to a target on the screen. Also in [1], where natural gaze behavior was investigated during a direct manipulation task at a large tabletop display, many different gaze behaviors were observed. Other studies from the field of psychology and physiology, e.g. [3, 2] investigated differences in gaze behavior during action execution and observation. They distinguished three different gaze behaviors, namely proactive, reactive and tracking gaze behavior [3].

In all of the above studies on natural gaze behavior, numerous different gaze patterns were observed during task execution and were described informally. However, an understanding of the reasons why a person looks at a certain location in a certain situation is necessary to judge the usefulness of natural gaze behavior for HCI and to integrate gaze with other modalities, respectively.

In this paper we report about a study in which we tried to characterize different influences on natural gaze behavior during an object manipulation task. Additionally, we point out their implications for designing gaze-based multimodal interaction techniques for future intelligent user interfaces.

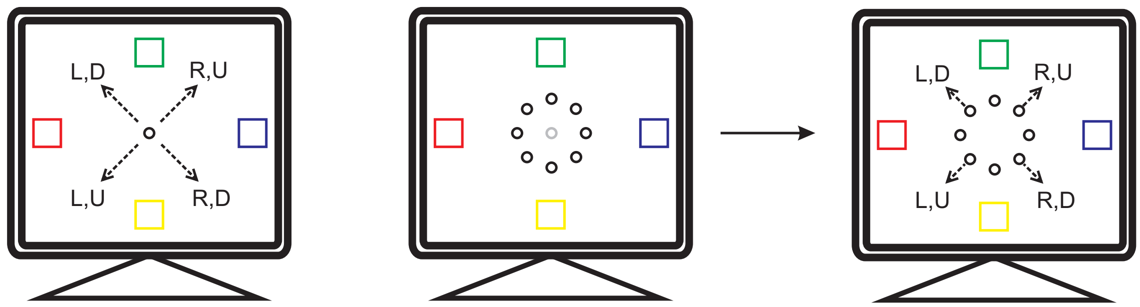
USER STUDY

Task and Apparatus

The task to be solved by participants is designed based upon a basic object manipulation task as it is common in many GUIs. The visual representation of an object has to be moved from one location to another on a display. However, in order to being able to investigate effects of user's mental model on natural gaze behavior in a controlled way, we designed the mapping between input and system reaction in an unusual way not expected by the users. This ensures that all users have the same level of knowledge about the system at the beginning of the experiment and can be considered as novice users. Additionally, we are able to monitor changes in natural gaze behavior with increasing knowledge about the system.

As input devices we use one single key of a keyboard (Figure 2(a)) and a pen tablet, while only horizontal movements of the pen on the tablet are interpreted by the system (Figure 2(b)). The task is illustrated in Figure 2(c). A colored point which initially is displayed at the center of the display is to be moved to one of the four squares $T_0, ..., T_3$ with the same color. Note that the labels $T_0, ..., T_3$ shown in Figure 2(c) were not displayed to the user during the experiment and only serve as reference for the respective target areas within this paper.

For manipulating the object position we implemented two different interaction techniques. The mapping between inputs and system state transitions (position of the point) is graphically illustrated for the first technique in Figure 3(a). For example, a horizontal movement of the pen to the right



(a) Technique1: movement direction of object given different inputs

(b) Technique2: expanded object state (left) and movement direction of individual representations given different inputs (right)

Figure 3. Mapping of input to system actions for different interaction techniques

(R) causes a movement of the point to the upper right if the key is not pressed (U) and to the lower right if the key is pressed (D). In principle the mapping for the second technique is the same. However, before the object is moved from its initial position, as soon as the pen touches the tablet, its visual representation is split into eight objects arranged on a circle around the initial position, representing possible future object positions (see Figure 3(b) left). This representation in the following is denoted as *expanded state* of the object. In order to avoid hints about the true mapping of inputs to movement directions by this representation, objects are also displayed along directions the object can not be moved to directly (e.g., to the right). However, all eight representations have the same color, namely the color of the target area the object is to be moved into. As soon as the object is in expanded state, a movement of the pen on the tablet leads to a movement of one of the eight object representations into the respective direction, while all other representations are removed. For example, if the pen is moved horizontally to the right (R) and the key is not pressed (U), the object representation in upper right direction is moved to the upper right, while all other objects are faded out (Figure 3(b) right).

In order to move the object from its initial position to the (green) target area T_0 at the top of the display along the path illustrated in Figure 2(c), for both techniques users first would have to move the pen to the right (R) while leaving the key unpressed (U) and, as soon as the little orange help point is reached, press the key (D) and move the pen to the left. An alternative way to solve the task would be to first move the point to the upper left (input: L,D) and then to the upper right (input: R,U). Users were free to chose the way to the respective target areas during the experiment.

In preliminary experiments with Technique1 we observed that experience of users seems to have significant influence on proactivity of gaze behavior. Novice users, for example, mainly directed visual attention towards the initial object position at the beginning of the task. In contrast, expert users predominantly anticipated future object positions. With Technique2 we wanted to investigate whether it is possible to induce more proactive gaze behavior, especially for novice users, by avoiding visual feedback in proximity to the initial object position right before the first object movement. By explicitly presenting possible future object positions to the user we expected gaze movements to be directed more towards those visual targets than towards the initial object position. This would, e.g., allow for robust estimation of users' intention from gaze data.

The size of the display is 33,7 x 27 cm with a resolution of 1280x1024 pixels. Eye-gaze of the users was captured during task execution by a Tobii 1750 tracking device.

Participants

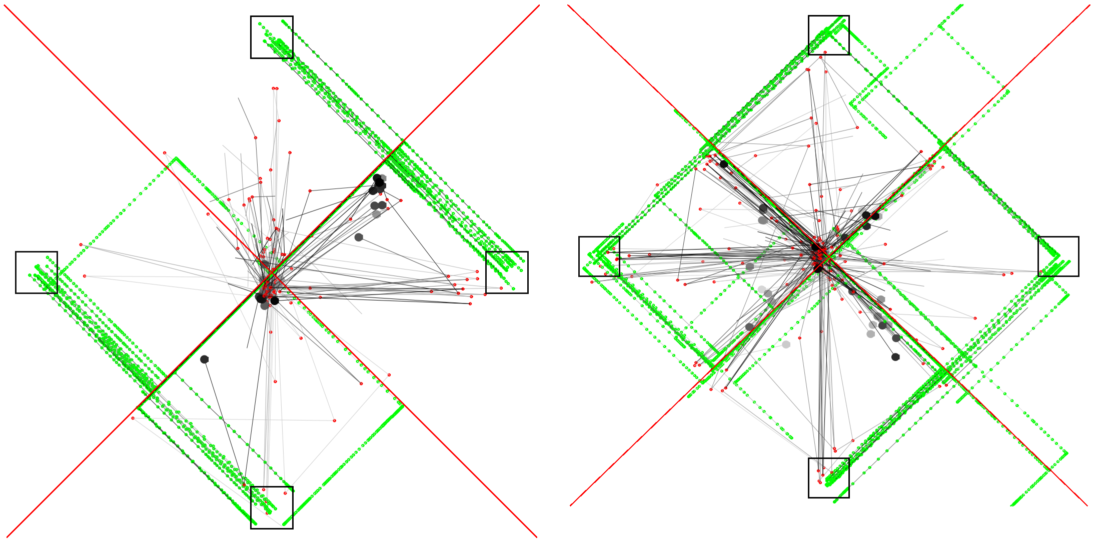
Since we want to investigate effects of mental model building on natural gaze behavior we chose a between-subjects design to avoid any prior knowledge of participants about the task or interaction techniques. We had two groups with 10 participants each. Participants were between 21 and 32 years old and did not know anything about the experiment, except that their gaze is measured.

Procedure

The experiment was organized in two phases $A1$ and $A2$ with 40 runs each. Every run consists of moving an object from its initial position at the center of the screen to the respective target area. During both phases of the experiment every color of the object and hence every task occurred 10 times, while the order of tasks was chosen randomly and was the same for all participants. Except the order of tasks there was no difference between phase $A1$ and $A2$.

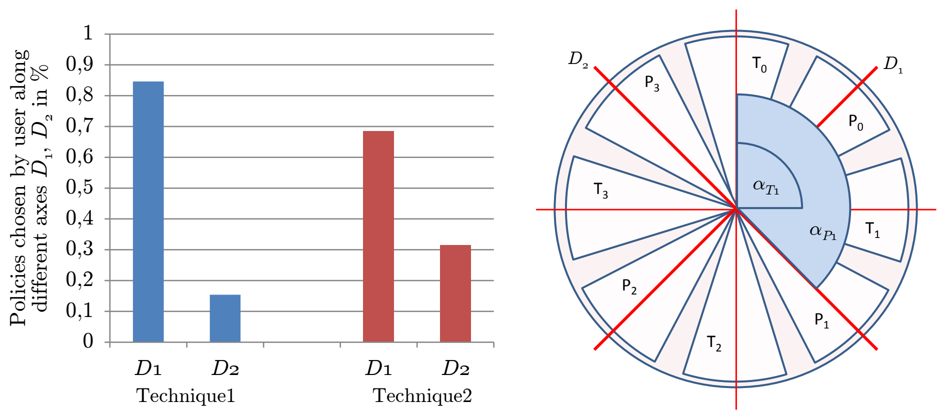
Between the two phases users were asked to fill in a questionnaire in order to capture their mental model. However, in this paper we focus on analysis of objective data only and analysis of subjective data obtained from the questionnaire will be reported in future papers.

In order to allow for a more detailed analysis of the temporal development of objective measures in subsequent sections the two phases are further divided into $A1/1$, $A1/2$, $A2/1$ and $A2/2$ with 20 runs each.



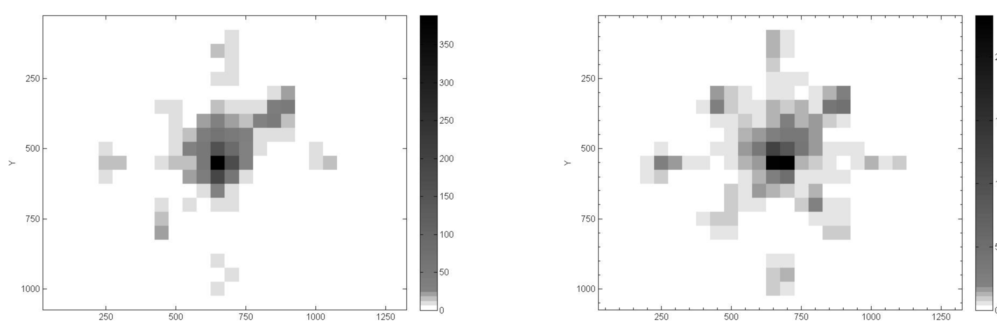
(a) Technique1 (b) Technique2

Figure 4. Data captured for different interaction techniques during phase $A1$ from one user for each technique



(a) Distribution of policies chosen by users (b) Definition of variables

Figure 5. Task solution strategy of users and definition of variables



(a) Pre-object fixations for Technique1 (b) Pre-object fixations for Technique2

Figure 6. Distribution of pre-object fixations

RESULTS

Most interesting from the interaction design perspective are gaze movements which occur before any object movement. In the following we denote such gaze data as *pre-object* gaze data and *pre-object fixations*, respectively. Such data allows for estimating users intentions previous to any input made by the user. Therefore in this work we mainly focus on the analysis of such data.

In Figure 4 a plot of object- and gaze-data during the first 40 runs is shown for each of the two interaction techniques for one user. Green dots represent object positions, small red dots connected by gray lines are pre-object fixations and larger dots, colored from gray to black, indicate the last pre-object fixation for each run. The red diagonal lines indicate possible movement directions of the object from its initial position and were not shown to the users during the experiments.

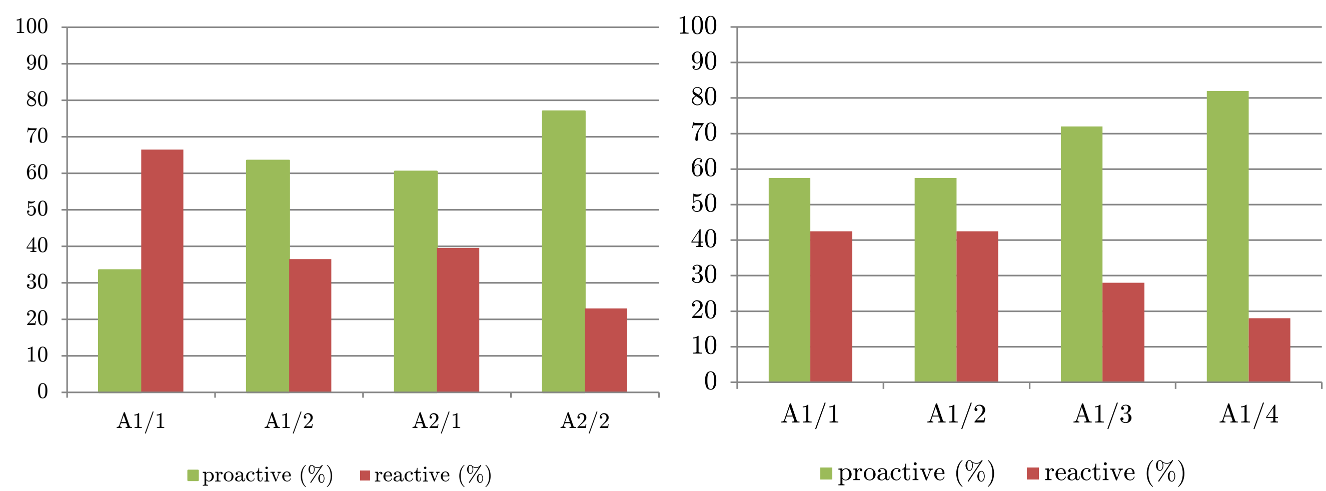
For the first interaction technique two things can be easily seen from Figure 4. First, the preferred policy for solving the task seems to be first moving the object along the diagonal line reaching from the lower left to the upper right (D_1, see Figure 5(b) for definition). This corresponds to an input sequence where the key is not pressed (U) during the first phase. Second, fixations are mainly located at three different positions on the screen. While the last pre-object fixation is either located at the initial position of the object or along the preferred diagonal axis D_1, other fixations also can be observed towards or at the target areas.

Both observations in average can be confirmed for all participants. In Figure 5(a) the distribution of tasks which were solved by moving the object first along the different axes D_1 and D_2 is shown for both interaction techniques. A clear majority of the users first moved the object along D_1 for both interaction techniques. However, the policies with first movement direction along axis D_2 was used more often for Technique2 (31,5 %) compared to Technique1 (15,38 %) .

This difference in interaction behavior also shows an effect on pre-object gaze behavior. Figure 6 shows the distribution of positions of all pre-object fixations for all users and tasks for the two interaction techniques. Note that the color scale at the lower end is not linear in order to improve the visibility of the plot. Both plots show that most pre-object fixations are centered around the initial position of the object. However, also a significant amount of fixations can be observed at different locations on the screen which are related to the task. Except from the initial object position for Technique1 fixations are mainly distributed along axis D_1 or at the target areas. The plot for Technique2 in Figure 6(b) shows also fixations along axis D_2 and in general more proactive fixations. For further task related characterization of fixations we use two features:

- *Distance* d of a fixation from initial object position
- *Direction* α of the vector between fixation and object position

Along d, fixations are classified in *proactive fixations* ($d > r_p$) and *reactive fixations* ($d \leq r_p$). The threshold r_p defines when a fixation is considered to be on the object (reactive) or not (proactive). While reactive fixations indicate attention allocation towards the current state of the object, proactive fixations are induced by mental planing activity for solving the task or anticipation of future system states. The design of the task allows for distinguishing between fixations which are directed towards one of the target areas and fixation induced by anticipation of the first movement direction of the object by evaluating α. We further denote the different target areas as $T_0, ..., T_3$ in clockwise direction, starting from the top. The different policies users can chose to solve a task are denoted by $P_0, ..., P_3$ in clockwise direction according to the first primary movement direction starting from the top. The definition of T_i and P_i are also illustrated in Figure 5(b). For example, if the task of moving the object to the upper target area is solved by moving the object first to the upper right (R,U) and then to the upper left (L,D), this corresponds to



(a) Technique1

(b) Technique2

Figure 7. Development of ratio between proactive and reactive fixations with increasing knowledge about the system

P_0. First moving the object to the upper left and then to the upper right for the same task would be P_3.

Based on these definitions the target of visual attention A indicated by a fixation can categorized as follows:

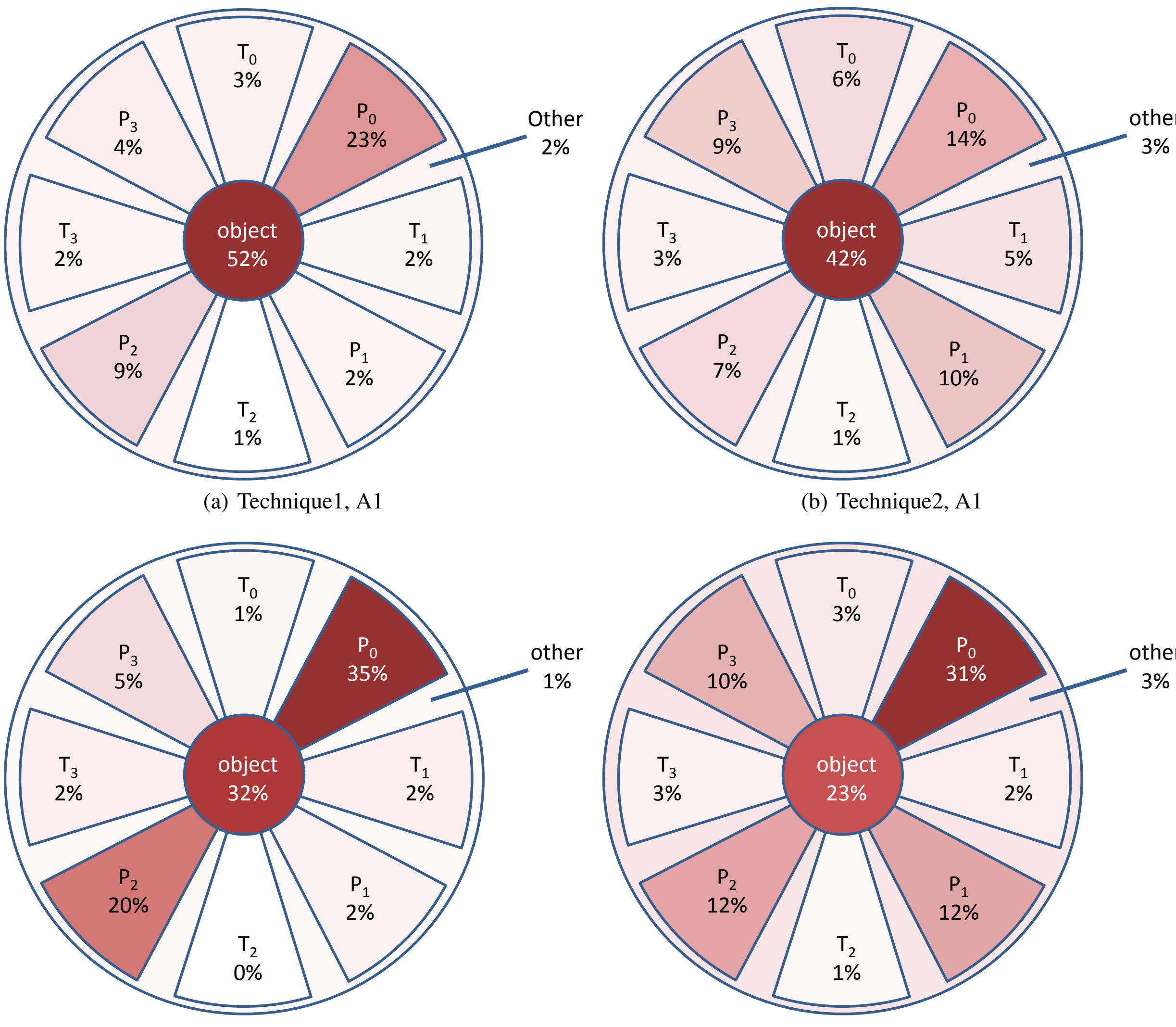
$$A = \begin{cases} T_i \text{ if } |\alpha_{T_i} - \alpha| < \alpha_{max} \\ P_i \text{ if } |\alpha_{P_i} - \alpha| < \alpha_{max} \\ other \end{cases},$$

where α_{T_i} and α_{P_i} denote directions of vectors between the initial object position and the corresponding target T_i or first movement direction of policy P_i (see Figure 5(b)).

The thresholds $r_p = 100$ and $\alpha_{max} = 20°$ are chosen based on the analysis of gaze data captured during the experiments.

The development of the ratio of proactive and reactive last pre-object fixations over all phases of the experiment is shown in Figure 7. In average the ratio for Technique1 is 58.625/41.375 (proactive/reactive) and 67.25/32.75 for Technique2. For phase $A1/1$ (first 20 runs) with Technique1 66.5% of all last pre-object fixations are reactive and 33.5% are proactive. In contrast, during phase $A1/1$ with Technique2 57.5% of the fixations are proactive and 42,5% reactive. The plots show both, significant influence of growing experience on the location of the last pre-object fixation and significant differences between the two interaction techniques.

As already mentioned above, we further analyze pre-object proactive fixations regarding the underlying target of visual attention A. Figure 8 shows the distribution of A over all possible targets $T_0, ..., T_3, P_0, ..., P_3$ for all last pre-object fixations. The different areas represent the categories as defined above by r_p and α_{max} and are colored according to the occurrence of fixations within the corresponding area on the screen.



(a) Technique1, A1

(b) Technique2, A1

(c) Technique1, A2

(d) Technique2, A2

Figure 8. Distribution of target of visual attention of last fixation before first object movement for all users and tasks

For both techniques the number of last pre-object fixations which occur on the object are reduced from phase $A1$ to phase $A2$ of the experiment almost to the half. For Technique2 approximately 10% less fixations are made on the object for both of the two phases compared to Technique1. In all plots among all policies $P_0, ..., P_3$ a clear majority of fixations can be found along policy P_0. While for Technique1 proactive fixations are mainly distributed along axis D_1 (policies P_0 and P_2), for Technique2 an almost equal distribution over policies P_1, P_2 and P_3 can be observed. This corresponds to findings illustrated in Figure 5(a), where similar differences in policies chosen by the users for solving the task are depicted.

DISCUSSION

The results in the previous section show that both independent variables we used in our experiment, namely the interaction technique and the experience of users, have significant influence on natural gaze behavior during human-computer interaction.

For both interaction techniques, increasing experience of the user with the system resulted in a highly increased number of proactive fixations with increasing orientation towards policies at the expense of decreasing orientation towards target areas. This development can be explained from an information theoretical perspective. The more knowledge the user has about the dynamics of the system the less new information can be acquired by reactive fixations on the initial object position and by observing the first object movement, respectively. If future expected object positions can be accurately predicted by acquired knowledge, it is more efficient to directly draw visual attention towards expected future object states, e.g., in order to support accurate positioning of the object at the intended target location. The decreasing orientation of visual attention towards target areas can be explained by the same effect. Increasing knowledge of the location of certain target areas decreases the value of directing visual attention towards the target areas.

When comparing gaze data for the different interaction techniques a significantly increased number of proactive fixations and a slight increase in fixations directed towards the target areas can be observed for Technique2. Additionally, while for Technique1 the policies along axis D_1 are predominantly chosen by the users and proactive fixations are mainly distributed along this axis, with Technique2 the policies along axis D_2 are chosen significantly more often and fixations along $P_1, ..., P_3$ are almost equally distributed. Obviously, the different ways how visual feedback is organized for the different interaction techniques not only influences natural gaze behavior, but also human decision processes and task solution strategies.

For both interaction techniques and independent from experience of users, by far most of the proactive fixations are made along P_0. Participants' gaze behavior seems to be more proactive when moving the object from the left to the right than into the opposite direction. Possible explanations for that bias could be found by further examination of influence of writing direction, handedness or other cultural and individual factors.

For designing interaction based on natural gaze behavior the observations above have different implications. Natural gaze behavior is influenced by many different factors. These factors can either be used for adapting human-computer interaction or they prevent the development of consistent interaction techniques due to their dependency from uncontrollable and varying environmental conditions (e.g., experience of users, different cultural background).

In this user study we identified 4 classes of major factors influencing natural gaze behavior during object manipulation and characterized their influence in proactivity and direction of visual attention:

1. task
2. policy
3. experience of users / state of mental model
4. visual feedback / interaction technique

We further identified phenomenons which probably could be explained by individual differences among users and/or cultural factors (e.g., increased proactivity for P_0).

The first two factors can be used for estimating user's intention from gaze data. Either the goal of the task or the policy chosen by the user to solve the task can be estimated previous to the first object movement and user input, respectively. However, their visibility in gaze data in the form of proactive fixations towards a certain task related location on the display depends to a large extend on the third factor, namely the state of user's mental model. This fact in principal can be used for estimating user's experience and adaptation of interaction. However, if the main goal is to design a consistent gaze-based interaction technique for novice and experienced users the goal would be to minimize the influence of experience on natural gaze behavior. According to the results of our study one option would be to use the fourth factor and to design interaction techniques which reduce this influence as we demonstrated it with Technique2. However, as we showed in the results section there still remain variances in natural gaze behavior which probably can be explained by individual differences among users or cultural factors. These factors have also to be considered when interpreting natural gaze behavior and designing appropriate system reactions.

CONCLUSION

By the experiment described in this paper we were able to identify different factors influencing natural gaze behavior during an object manipulation task and to characterize their influence on proactivity and direction of fixations towards different task-related targets. Additionally, we demonstrated that the influence of individual factors can be changed by interaction design and adjusted visual feedback, respectively.

The results reported in this paper show the variety of information contained in natural gaze behavior. By analyzing natural gaze behavior during human-computer interaction information like user's intention or experience can be inferred which can be used for designing proactive or adaptive intelligent user interfaces.

In future work we plan to further validate the identified dependencies with more complex tasks and to design and evaluate gaze-based multimodal interaction techniques with a focus on multimodal combination of gesture and gaze.

REFERENCES

1. T. Bader, M. Vogelgesang, and E. Klaus. Multimodal integration of natural gaze behavior for intention recognition during object manipulation. In *Proceedings of the 2009 International Conference on Multimodal Interfaces (ICMI-MLMI)*, pages 199–206. ACM, Nov. 2009.

2. J. R. Flanagan and R. S. Johansson. Action plans used in action observation. *Nature*, 424(6950):769–771, Aug 2003.

3. B. Gesierich, A. Bruzzo, G. Ottoboni, and L. Finos. Human gaze behaviour during action execution and observation. *Acta Psychologica*, 128(2):324 – 330, 2008.

4. A. Hyrskykari, P. Majaranta, and K. Räihä. Proactive response to eye movements. In M. Rauterberg, editor, *Human Computer Interaction INTERACT 2003*, pages 129–136. IOS Press, September 2003.

5. R. Jacob and K. Karn. *Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises*, pages 573–605. Elsevier Science, 2003.

6. R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan. Eye-hand coordination in object manipulation. *The Journal of Neuroscience*, 21(17):6917–6932, 2001.

7. M. Land and D. Lee. Where we look when we steer. *Nature*, 369:742–744, 1994.

8. M. F. Land and P. McLeod. From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience*, 3:1340–1345, 2000.

9. P. Majaranta and K.-J. Räihä. Twenty years of eye typing: systems and design issues. In *ETRA '02: Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 15–22, New York, NY, USA, 2002. ACM.

10. J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139:266–277, 2001. 10.1007/s002210100745.

11. B. A. Smith, J. Ho, W. Ark, and S. Zhai. Hand eye coordination patterns in target selection. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, ETRA '00, pages 117–122, New York, NY, USA, 2000. ACM.

12. S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *In CHI99*, pages 246–253. ACM Press, 1999.