The Analysis of Eye Movements in the Context of Cognitive Technical Systems: Three Critical Issues

Sebastian Pannasch[1,2], Jens R. Helmert[2], Romy Müller[2], and Boris M. Velichkovsky[2,3]

[1] Brain Research Unit, Aalto University School of Science, Espoo, Finland
[2] Applied Cognitive Research/Psychology III, Technische Universitaet Dresden, Germany
[3] Department of Cognitive Studies, Kurchatov Institute, Moscow, Russian Federation
{pannasch,helmert,mueller,velich}@applied-cognition.org

Abstract. Understanding mechanisms of attention is important in the context of research and application. Eye tracking is a promising method to approach this question, especially for the development of future cognitive technical systems. Based on three examples, we discuss aspects of eye gaze behaviour which are relevant for research and application. First, we demonstrate the omnipresent influence of sudden auditory and visual events on the duration of fixations. Second, we show that the correspondence between gaze direction and attention allocation is determined by characteristics of the task. Third, we explore how eye movements can be used for information transmission in remote collaboration by comparing it with verbal interaction and the mouse cursor. Analysing eye tracking in the context of future applications reveals a great potential but requires solid knowledge of the various facets of gaze behavior.

Keywords: eye movements, attention, fixation duration, remote collaboration.

1 Introduction

Action and interaction with objects and other persons in the environment requires attention. The definition, understanding, and measurement of attention is one of the central research topics in psychology and cognitive science [1]. This interest is motivated not only by fundamental research questions but also by the increasing complexity of our (technical) environments. Currently, it becomes more and more challenging for users of technical devices to monitor and organize their interaction with them, as these requirements strongly increase mental load. Future developments, therefore, should build on solid knowledge about perception, attention and information processing to directly incorporate these processes into the design of attention- and intention-sensitive interfaces.

To approach this problem, several behavioural and psychophysiological measurement techniques have been employed in the past. Within the methodological arsenal, eye tracking and the analysis of human eye movements are most appropriate to investigate attention and provide attention-based support. This argument is founded on two main advantages of eye tracking. First, it is assumed that the direction of the

A. Esposito et al. (Eds.): Cognitive Behavioural Systems 2011, LNCS 7403, pp. 19–34, 2012.

eyes corresponds to the allocation of visual attention, thus measuring gaze behaviour can provide insights about mental processing [2]. Second, video-based eye tracking is a non-invasive, general tool, which can be applied to almost all everyday life situations [3].

In humans, as in all higher primates, vision is the dominant sensory modality and the important role of eye movements for visual processing has been repeatedly emphasized [e.g. 4]. During visual perception, information is sampled from the environment via 'active vision' [5]. Saccades—fast ballistic movements—redirect the foveal region of the eyes from one fixation point to another. During saccades, the intake and processing of visual information is largely suppressed and therefore limited to the periods of fixations [6]. This interplay of fixations and saccades is essential, as highest visual acuity is limited to the small foveal region; outside this high-resolution area, vision becomes blurred and the perception of colour is reduced. Eye movement behaviour in many everyday situations, such as reading text or inspecting images, can be described as an alternation between fixations and saccades.

Investigating human eye movement behaviour generates a quantity of rich data and therefore allows for an analysis of various parameters [for reviews 7, 8]. Traditionally, these analyses have largely relied on *when*, *where* and *how* information is gathered from the visual environment. Here, we will focus on the first two aspects and additionally consider a specific feature of gaze behaviour in *social* interaction. First, considering the *when* aspect is of importance with regard to the level of information processing. Particularly, we will examine the duration of fixations in free visual exploration. Second, regarding the *where* characteristic of gaze behaviour, it is usually assumed that the direction of the eyes allow for accurate estimations of the ongoing focus of interest and processing at any given time. Third, when communicating with other people, gaze behaviour has a particular function in *social* interaction; here we will investigate its contribution when direct communication is impaired in situations of remote collaboration. The selected characteristics provide representative examples about the functional importance of eye movements when trying to understand mechanisms of attention and information processing.

In our opinion, eye tracking will play an important role in the development of attentive interfaces. First versions of technical systems based on the analysis of eye gaze behaviour have already emerged [e.g. 9]. However, it should be mentioned here that this perspective is not new, and in each decade since the 1950s, the discussion about the use of eye tracking for solving new problems has persistently returned [10]. While confident of the potential of eye tracking, we are nonetheless well aware of the potential risks for failure when eye movements are assumed to serve as a simple attention pointer. In fact, successfully implementing attention-sensitive devices requires a deep understanding of the underlying mechanisms of gaze control as well as a careful interpretation of the resulting behaviour.

In the following sections we will present recent results from three different domains of eye movement research and thereby highlight potentials and pitfalls in understanding the complex control mechanisms of eye movements and discuss their significance for the development of cognitive technical systems.

2 Sensitivity of Fixations to Distraction

Visual fixations represent the time intervals dedicated to visual information uptake and processing. Their durations are often considered as reflecting the entropy of the fixated information: longer fixation durations are associated with the processing of demanding information and higher task complexity [11]. It is generally assumed that the employment of more cognitive efforts is expressed in longer fixations, for instance when eye movements are analysed in the context of reading [12] or scene perception [13]. However, this hypothesis raises the question if and to what extent also other factors can modulate the duration of fixations, as this would make it difficult to attribute these temporal variations to the ongoing information processing. In fact, it has been shown that sudden changes in the environment lead to a robust prolongation of the fixation duration [e.g. 14].

Understanding the underlying mechanisms of this change-related prolongation is an interesting research endeavour in itself, but it also turns out to be of particular importance when analysing eye movements in applied contexts. For instance, it has been demonstrated that the change of a traffic light (i.e. from green to red) results in a pronounced prolongation of the respective fixation [15]; based on such a feature, one could think of fixation-based hazard recognition.

Therefore, it is necessary to understand what processes take place within a single fixation and how they contribute to its duration. In the present experiment, we examined if periods of higher or lower sensitivity to distraction within a fixation can be identified. Recently, it has been reported that fixations can be influenced by the appearance of visual, acoustic and haptic events [16]. To further investigate this phenomenon, we presented visual and auditory distractors.

2.1 Methods

Subjects. Seventeen students (10 females) of the Technische Universität Dresden with a mean age of 23.4 years (range 20-30 years) took part in this experiment. All subjects reported normal or corrected-to-normal vision, normal hearing and received course credit for participation. The study was conducted in conformity with the declaration of Helsinki.

Apparatus. Participants were seated in a dimly illuminated, sound-attenuated room. Eye movements were sampled monocularly at 250 Hz using the SR EyeLink I infrared eye tracking system with on-line detection of saccades and fixations and a spatial accuracy of better than 0.5°. Stimuli were shown using a CRT display (19-inch Samtron 98 PDF) at 800 by 600 pixels at a refresh rate of 100 Hz. Viewed from a distance of 80 cm, the screen subtended a visual angle of 27.1° horizontally and 20.5° vertically.

Stimuli. Ten digitized pieces of fine art by European seventeenth to nineteenth-century painters served as stimulus material. Visual and auditory distractors were presented to systematically investigate influences of gaze-contingent distractions. Visual distractors were implemented as colour inversion of an image segment with a size of 50 by 50 pixels, always appearing 50 pixels to the left of the ongoing fixation. Auditory distractors consisted of a 900 Hz sinusoidal tone, presented at a sound pressure level of 70 dB via PC-loudspeakers on both sides of the screen.

Procedure. Subjects were informed that the purpose of the study was to investigate eye movement patterns in art perception and were asked to study the images in order to be prepared to answer subsequent questions regarding the image content. They were aware of the presentation of distractors but instructed to ignore them. The experiment was run in two consecutive blocks of varying distractor modality (visual or auditory), each containing five pictures. The order of blocks was counterbalanced across subjects. A 9-point calibration and validation was performed before the start of each block. Before each trial, a drift correction was performed. Distractor presentation always began after an initial period of 20 s of scene inspection in order to allow subjects firstly to explore each image without disturbance. Once all 21 distractors (see below) were shown, the image was replaced by five questions which had to be answered by clicking 'yes' or 'no' on-screen buttons using the mouse. The total duration of the experiment was about 40 min.

Distractors were presented at every fifth fixation during a trial. This presentation interval was selected according to previous work [16] and warranted enough unaffected fixations in between, serving as baseline. Distractors were triggered by the fixation onset with a latency of 50, 100, 150, 200, 250, 300 or 350 ms and presented with a duration of 75 ms. For each onset delay, three distractors were shown in a randomized order, resulting in a total of 21 distractors per image. If a fixation was terminated before reaching the onset latency, the program waited for the next suitable fixation. The image presentation lasted until all 21 distractors were presented (65 seconds on average).

2.2 Results

Fixations around eyeblinks and outside the presentation screen were removed. Further processing included only distracted fixations and the two adjacent non-distracted fixations. The non-distracted fixations served as baseline. To assure comparability of the baseline and the distractor condition, fixations of shorter duration than the respective distractor latency (see above) were excluded, resulting in a total of 25686 (82%) valid fixations. *Eta*-squared values are reported as estimates of the effect size [17].

To investigate the effects of the visual and auditory distractors, fixation durations of the baseline and the distractor condition were compared. Medians of fixation duration were applied to a 2 (modality: visual, auditory) × 2 (fixation type: distracted, baseline) repeated measures analysis of variance (ANOVA) and revealed significant main effects for fixation type, $F(1,16) = 223.96$, $p < .001$, $\eta^2 = .68$, but not for modality, $F < 1$. Furthermore, we found a significant interaction for modality × fixation type, $F(1,16) = 5.95$, $p = .027$, $\eta^2 = .004$. Regarding the main effect of fixation type, fixation durations were longer when affected by a distractor presentation (*M*s: 315 vs. 250 ms). The interaction was based on the slightly stronger influence of visual distractors

(*M*s: 318 vs. 311 ms), while the average fixation duration in both baseline conditions was similar (*M*s: 248 vs. 251 ms).

In order to examine if the appearance of a distraction at various latencies within a fixation induces differential effects, we calculated the differences between distracted and baseline fixations, for each modality and latency. The obtained difference values were applied to a 2 (modality: visual, auditory) × 7 (latency: 50, 100, 150, 200, 250, 300, 350) repeated measures ANOVA and revealed a significant main effect for modality, $F(1,16) = 20.05$, $p < .001$, $\eta^2 = .084$, but not for latency, $F(6,96) = 1.69$, $p = .132$. No interaction effect was found, $F < 1$. The obtained main effect for modality is based on a stronger general influence of visual distraction, evidenced in larger difference values between baseline and distractor fixations (*M*s: 46 vs. 13 ms).

2.3 Discussion

In accordance with previous findings [16], we observed event-related prolongations of fixations for visual as well as for auditory distraction. This result is important for the interpretation of fixation durations in applied contexts: Something as ordinary as a ringing phone might be responsible for a prolonged fixation. Thus, the fixation duration represents a highly sensitive parameter, reflecting internal processing mechanisms as well as reacting to external events. Incorporating the fixation duration in attention-sensitive interfaces therefore requires considering this interaction. The analysis of the difference values provides clear evidence that the appearance of a distracting event at any time within a fixation evokes a similar prolongation effect.

Furthermore, the influence of visual distractors was stronger than that of auditory distractors, which corresponds to earlier reports [16, 18]. However, based on the current findings, it remains to be determined whether this difference results from different processing mechanisms or to a lack of comparability between the two types of distractors. Although both distracting events were shown within the same experimental paradigm, we cannot be sure that a colour inversion of 50 x 50 pixels represents a comparable event to a 900 Hz sinusoidal tone of 70 dB.

3 Focus of Attention

In general, it is assumed that visual attention is allocated to the position of the current fixation. Reportability of fixational content is often considered a measure of attention allocation. While in most cases a perfect fit of gaze position and attentional direction can be found, there is evidence that subjects report contents ahead [19] as well as behind [20] the position of the current fixation. It therefore is of interest to understand if the relationship between fixation position and attention allocation changes according to particular requirements (for instance with respect to the task at hand) or if the above mentioned controversial results are rather based on differences in the paradigms.

Support for the attention-ahead-of-fixation assumption is mainly found in laboratory settings using so-called ‘fixate-and-jump’ paradigms. In such settings, saccades are programmed due to arbitrary commands, as criticized by Fischer [21]. In complex, everyday tasks under rather natural settings, it has been shown that subjects report the content of the current or previous fixation. Notwithstanding, to employ eye movement analysis in applied domains, these temporal characteristics are essential and require a deeper understanding. To contribute to this discussion and to allow for a precise investigation of attention allocation, we analysed the subjective focus of visual attention in a continuous paradigm using different task instructions. The current experiment is based on hierarchical approaches of attention, assuming that attention operates on different levels [e.g. 1, 11, 22].

3.1 Method

Subjects. Eighteen students (9 females) of the Technische Universität Dresden with a mean age of 23.4 years (range 20-31 years) took part in this experiment. All subjects reported normal or corrected-to-normal vision, normal hearing and received course credit for participation in the study conducted in conformity with the declaration of Helsinki.

Apparatus. The same apparatus as in the first experiment was used.

Stimuli. A total of 121 black and white pictograms served as stimuli and were shown with a size of 2° x 2° of visual angle. Within each trial, six pictograms were presented in a circular array with a diameter of 13.3° of visual angle (Figure 1C).

Procedure. All subjects completed three blocks of 60 trials. Within one block all trials were of the same task condition. Three different tasks were employed. In the *position* condition (Figure 1A), the task was to click on the empty position where subjects felt they were looking at the time of the trial end. In the *content* condition, subjects had to choose the pictogram they thought they had inspected during the beep. Therefore, the test screen contained three pictograms: the actual fixated one, the previously fixated one and the next one in the array (Figure 1B). In the condition *position and content*, the trial screen remained unchanged.

A 9-point calibration and validation was performed before the start of the experiment. Each trial started with a drift correction at one of the six locations where the pictograms were shown. Among the trials, the drift correction location was

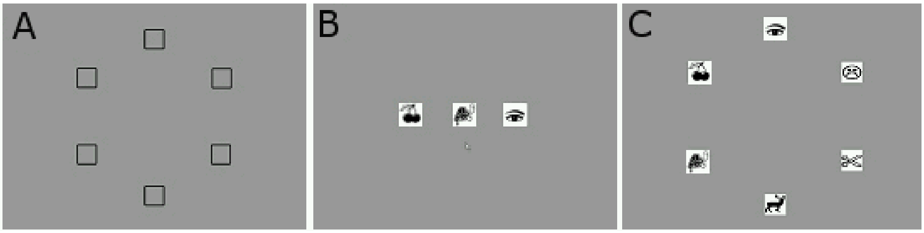

Fig. 1. Initial arrangement of pictograms during the trials (C). Test screens for the different tasks: Position (A), Content (B), and Position and Content (C).

randomly selected but counterbalanced across the block. After the onset of the pictograms, subjects had to scan the display in a clockwise manner, starting from the drift correction position. As soon as a predefined pictogram was fixated, a countdown started (400-600 ms, steps of 50 ms). Once the end of the countdown was reached, a beep was provided to signal the end of presentation. This countdown procedure and time ensured that the trial end in the majority of cases appeared around the start and end of a fixation. Subsequently, the test screens were presented. After the judgments were made, a new trial was initiated. The presentation order of blocks was counterbalanced across subjects; at the end of each block, subjects had a break of five minutes. An experimental session lasted approximately 40 minutes.

3.2 Results

Before the statistical analysis, some trials were rejected due to invalid recording. Furthermore, trials were excluded in which the last fixation position did not correspond to the position of a pictogram. Subsequent to this preprocessing, 2380 valid trials remained (74% of all trials). In these valid trials, participants had chosen the pictogram they felt they last fixated: the pictogram previous to the one the eye fixated during the signal tone (*previous*); the actual pictogram the eye fixated during the signal tone (*actual*), or the next pictogram (*next*). As another factor in the analyses, the viewing times of the last pictogram were considered. They were divided into three categories based on tertiles, with the same number of cases in each category. This resulted in three viewing time conditions; the respective median values and ranges are shown in Table 1.

Table 1. Tertile-based categories of viewing times (in ms)

Viewing Time	Minimum	Median	Maximum	N
Short	1	73	139	795
Medium	140	243	400	824
Long	401	498	601	761

The viewing time categories (short, middle, long) served as independent variables for further statistical testing. The dependent variables were probabilities of *previous* and *next* responses. Probabilities of choosing the *actual* position/pictogram were not analysed, as these cases indicated an overlap of eye fixation and perceived focus of visual attention, hence providing no diagnostic information. Two 3 (viewing time) × 3 (task) repeated measures ANOVAs were conducted. For *previous*, significant main effects for task, $F(2, 34) = 3.90$, $p = .030$, $\eta^2 = .02$, and viewing time, $F(2, 34) = 25.23$, $p < .001$, $\eta^2 = .34$, as well as a significant interaction, $F(4, 68) = 6.90$, $p < .001$, $\eta^2 = .05$, were obtained. Concerning task, results show that the highest probability for the *previous* choice was in the *content* condition (11.1%), followed by *position & content* (9.0%) and *position* (5.2%). The factor viewing time clearly shows a decrease

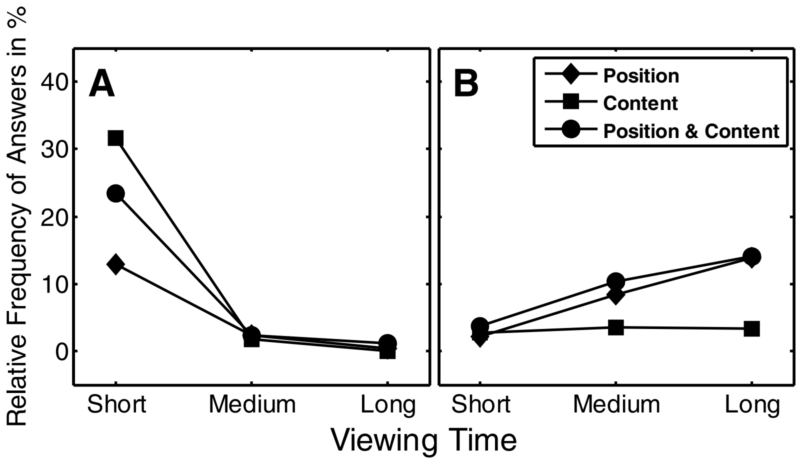

Fig. 2. Interaction of viewing time and task for (A) *previous* and (B) *next* reactions

from *short* (22.6%) to *medium* (2.2%) to *long* (.5%). The interaction dramatically illustrates the specific relationship between viewing time and task condition for the probabilities (Figure 2A). Post hoc analyses reveal significant differences between tasks for short viewing time only, $F(2,34) = 6.08$, $p < .001$, $\eta^2 = .11$.

In the case of *next*, significant main effects were found for task $F(2, 34) = 5.302$, $p = .001$, $\eta^2 = .064$, and viewing time $F(2, 34) = 8.010$, $p < .001$, $\eta^2 = .086$, as well as the interaction of both factors, $F(4, 68) = 5.583$, $p < .001$, $\eta^2 = .037$. With regard to task, the lowest probability was obtained for *content* (3.1%), followed by *position* (8.1%) and *position & content* (9.3%). Looking at viewing time, probabilities increase with viewing time (*short* 2.8%, *medium* 7.4%, and *long* 10.4%). As already described for *previous* reactions, the interaction of both factors shows a systematic pattern (Figure 2B). Here, post hoc analyses revealed significant differences between tasks in the *medium*, $F(2,34) = 4.69$, $p = .016$, $\eta^2 = .084$, and *long* condition, $F(2,34) = 8.35$, $p = .001$, $\eta^2 = .15$, respectively.

3.3 Discussion

The objective of this experiment was to systematically analyze influences of different tasks on the report of the subjective ‘last glance’. Our method permitted the performance of sequences of fixations and saccades—similar to naturalistic gaze behaviour—and furthermore contrasted different visual tasks, namely spatial localisation and identification.

The present results do not support the attention-ahead-of-fixation assumption. In fact, an asynchrony of actual eye position and reported position was only found for short viewing times, and the asynchrony was in direct contrast to studies where eye position was found to lag behind attention [23]. We observed a pronounced dependency of reports on the task at hand. The influence of the task on reporting behaviour also interacted with viewing time. We found a strong tendency to report the pictogram from a previous position. In addition, we discovered a second trend: if the task explicitly involved localisation, the probability of reporting the subsequent fixation position grew with increasing viewing time. This second trend is akin to the results of Fischer [21] as well as to results of the large number of studies using spatial cueing paradigms [19]. However, the trend was by far weaker than that usually reported in the single-saccade spatial cueing experiments. According to our data, slightly more than 10% of fixations were reported as being shifted towards the next spatial location, even with the longest viewing time at an actual position.

4 Gaze Transfer in Remote Collaboration

The final section of this article is dedicated to another domain where eye movements are of central importance: social interaction. Several studies have investigated the role of gaze behaviour in direct social interaction [for review 24] and when interacting with a virtual character [25]. However, in contemporary work life, a major percentage of social interactions take place in the form of remote collaboration, with the partners residing in different locations [26]. It has been demonstrated that transferring the gaze of one partner to the other partner with a cursor superimposed on the visual material can improve the performance in spatial tasks by disambiguating object references [27].

Regardless of the benefits of gaze transfer compared to purely verbal interactions, transferring computer mouse positions provides a rather direct pointing device, also allowing for referential disambiguation. Despite the lack of performance differences between gaze and mouse transfer, both transfer methods have differential effects on the cooperation process. Recent research has revealed that transferring gaze resulted in difficulties in interpreting communicative intention but demonstrated a strong coupling of attention to the transferred cursor [28]. Consequently, gaze transfer required a more effortful verbal disambiguation.

Here, we investigated the effects of gaze and mouse cursor transfer under conditions where the information about a person's attention and search process was crucial. Due to the strong link between attention and eye movements, a gaze cursor could be expected to provide an advantage over purely intentional mouse pointing. Our goal was to determine how the usability of gaze or mouse cursor transfer depends on the partner's ability to link this cursor to the objects in question. Pairs of participants had to solve a joint path-selection task with a strong spatial component on different processing levels (colour differentiation, form identification, calculation).

4.1 Method

Subjects. Forty-eight subjects (32 females) with a mean age of 23.9 years (range 18-51 years) participated in the experiment. They were invited in pairs and assigned to one of two experimental roles (searcher or assistant), resulting in a total of 24 pairs. All subjects reported normal or corrected-to-normal vision and received either course credit or a compensation of €7 for participation in the study conducted in conformity with the declaration of Helsinki.

Apparatus. Both participants were seated in front of their computers in the same room, separated by a portable wall. The computers were connected via Ethernet. Eye movements of the searcher were recorded monocularly at 500 Hz with the SR EyeLink 1000 infrared eye tracking system in the remote recording mode.

Stimuli. Twenty images with a resolution of 1024 by 768 pixels served as stimuli in the experiment. They were composed of a grid of 20 x 20 rectangles, forming three red and three green paths (see Figure 3B). On each path, a variable number of circles and triangles was positioned, containing positive or negative digits; this information was only visible to the searcher within a window of 255 x 190 pixels (1/16 of the screen), while the rest of the screen was covered in black (Figure 3A). For the assistant, the whole screen area was visible. In the condition *objects*, all objects were depicted as circles (Figure 3B), while in the condition *grid*, only a grey background consisting of equidistant vertical and horizontal lines was visible (Figure 3C). The searcher's gaze or mouse position was projected onto the assistant's screen as a tricolour eye-icon.

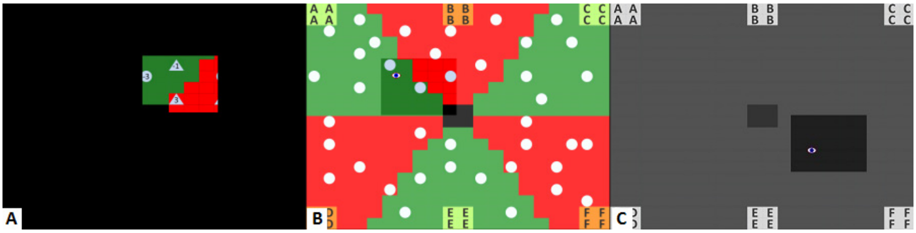

Fig. 3. Stimuli for searcher (A), and for assistant in the conditions objects (B) and grid (C)

Procedure. The experiment consisted of four blocks corresponding to the combinations of the experimental conditions (see below). The basic task in all experimental conditions was the following: In five trials per block, participants had to determine the correct path in a stepwise manner. They first had to select the three red paths, next they had to exclude the path with the least number of circles before finally determining the path which contained the smaller sum of digits. The chosen path had to be selected by the searcher via mouse click on the respective letter target field (see Figure 3B). The form of the paths was identical throughout the whole experiment, only their order changed across trials.

The searcher was provided with the full stimulus information necessary to solve the task, but saw only a section of the display (Figure 3A), while the assistant had to move this viewing window to reveal the respective display areas relevant to the searcher. Either the searcher's gaze or mouse cursor was transferred to the assistant for guidance. Besides the cursor, the assistant either saw the object positions, but not their identity (Figure 3B), or had no task-relevant visual information (Figure 3C). Participants were free to verbally interact in all experimental conditions. We recorded the eye movements of the searcher, the mouse actions of both participants, and their verbal interactions.

4.2 Results

Performance. Mean solutions times were applied to a 2 (cursor: gaze, mouse) × 2 (assistant view: objects, grid) repeated measures ANOVA and revealed main effects for cursor, $F(1,23) = 22.60$, $p < .001$, $\eta^2 = .080$, and assistant view, $F(1,23) = 14.69$, $p < .001$, $\eta^2 = .102$, as well as a significant interaction, $F(1,23) = 11.89$, $p = .002$, $\eta^2 = .041$. Solution times were shorter in mouse than in gaze (*M*s: 78 vs. 102 s) and shorter in objects than in grid (*M*s: 76 vs. 103 s). The interaction results from the fact that the difference between mouse and gaze was only present in grid, $p < .001$, but not in objects, $p = .153$ (see Figure 4A). Mean error rates were 19.4% and showed no differences between the experimental conditions, all $F < 3$, all $p > .1$.

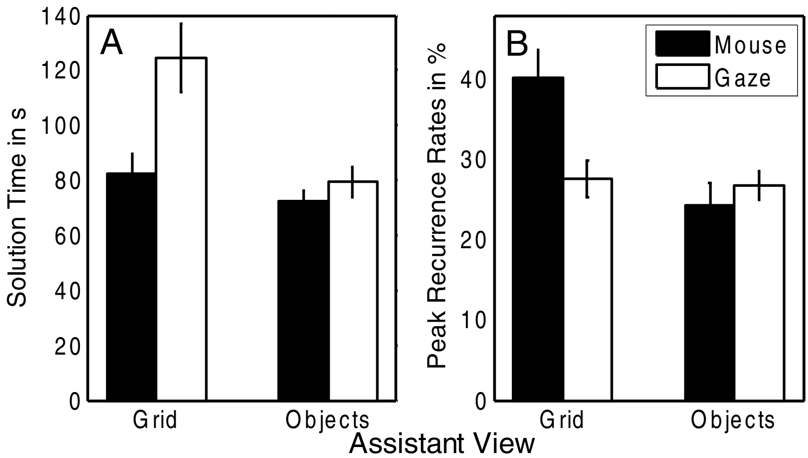

Fig. 4. Solution times (A) and peak recurrence rates between cursor and window centre (B) for the investigated cursor conditions.

Window and Cursor Alignment. In order to examine the positional coupling between the searcher's and the assistant's cursor (with the latter one corresponding to the window centre), a cross recurrence analysis [29] was conducted. This analysis provides cursor recurrence rates as the percentage of samples where the position of searcher and assistant cursor were located at about the same position, qualified by all cursor samples and at different temporal delays between both time series. Peak recurrence rates were subjected to a 2 (cursor: gaze, mouse) × 2 (assistant view: objects, grid) repeated measures ANOVA, revealing main effects for cursor, $F(1,23) = 5.16$, $p = .033$, $\eta^2 = .034$, and assistant view, $F(1,23) = 18.59$, $p < .001$, $\eta^2 = .089$, as well as an interaction, $F(1,23) = 19.78$, $p < .001$, $\eta^2 = .073$. The coupling was stronger in mouse than in gaze (32.3 vs. 27.2%) and stronger in grid than in objects (33.9 vs. 25.6%). However, grid recurrence rates were larger for mouse than for gaze, $p < .001$, whereas no such difference was observed in objects, $p = .705$ (see Figure 4B).

In addition to peak recurrence, the temporal dynamics of the cursor window alignment were investigated. When considering recurrence rates as a curve ascending from a maximum negative temporal delay (-1000 ms in this study), peaking at a certain delay and then descending until maximum positive delay (1500 ms), the resulting amplitude of this curve represents a measure for the degree to which recurrence rates depend on temporal delay. Thus, it provides a measure of how tightly two cursors are coupled. There was an effect of cursor, $F(1,23) = 11.95$, $p = .002$, $\eta^2 = .043$, an effect of assistant view, $F(1,23) = 26.48$, $p < .001$, $\eta^2 = .084$, and an interaction between both factors, $F(1,23) = 59.00$, $p < .001$, $\eta^2 = .109$. When using the mouse, the increase was higher in grid than in objects (15.9 vs. 6.2%), $p < .001$. However, when gaze was used, there was no difference between grid and objects (7.5 and 8.1%), $p = .444$.

4.3 Discussion

Transferring gaze without further visual information about the task environment resulted in longer solution times compared with mouse transfer. Similar solution times were found for gaze and mouse transfer when adequate visual information was available. In this case, seeing a partner's gaze position can be as helpful as seeing their mouse. Without the required visual information, subjects were still able to efficiently use the mouse but not the gaze cursor. How can this effect be accounted for?

We think that the interpretability of the different cursors is directly related to the information they transmit. This information differs between gaze and mouse. Eye movements provide a rather direct visualization of visual attention in relation to task-relevant objects, especially in active tasks [30]. Their temporal and spatial parameters are closely related to processing information about these objects [11]. Visual attention usually does not float freely in space, but always implies a relation between a person and the entities that are being attended to. Thus, transmitting eye movements without an appropriate framework makes it difficult to interpret the gaze behaviour. In contrast, using the mouse as an intentional device for communication allows solely employing it to give messages to the partner. Thus, the partner knows that whatever the mouse does, he can simply react. In fact, several searchers instructed their assistants to "don't think, just follow my cursor". Although the assistants do not understand why a certain mouse movement is executed, they can be sure that it is produced as a deictic sign. In this case, simply following the cursor is a suitable strategy.

Our cross recurrence data support this interpretation. For mouse transfer, the coupling between the searcher's cursor and the window centre increased when no objects were available. Thus, the assistant relied more strongly on the searcher's guidance. Such an increased coupling was not found for gaze transfer. Additionally, the change of recurrence rate over different temporal delays was more than 2.5 times higher for grid than objects in the mouse condition but recurrence rate was similar for gaze in both viewing conditions. Thus, impoverished viewing conditions resulted in closer coupling of attention to mouse movements but not to gaze behaviour.

What is the benefit of mouse movements, and why can they be used more reliably than gaze, at least in the grid condition? While gaze is too fast and unpredictable to be followed unselectively, mouse movements can be adjusted to the requirements by performing slow and systematic moves. Thus, people can follow the mouse without necessarily understanding the meaning of the individual moves.

Since gaze transfer cannot be controlled as easily by the user, future research will focus on finding ways of adjusting gaze so that it will provide appropriate support. More refined visualization techniques appear to be a fruitful approach to address this issue. For example, smoothing the transferred gaze positions has been shown to increase subjective cursor control in a human computer interaction setting [31] and might improve its usability in cooperative situations as well.

5 General Discussion

The current work addressed three important issues regarding the analysis and interpretation of eye movements in the context of cognitive research and application.

In the first investigation, we demonstrated the susceptibility of visual fixation behaviour to the appearance of irrelevant visual or auditory distractors. According to our results, fixations are sensitive to multimodal distractions throughout their complete time course. Two main conclusions can be drawn from this study. First, when using fixation duration as a measure of information processing, a careful consideration of ongoing activities in the environment is required in order to avoid confounding artefacts in the measurement. Second, this finding should not be understood as a general rejection of the use of fixation durations. Rather we want to emphasize that using this parameter can provide helpful insights into ongoing processing mechanisms. To elaborate on this, Velichkovsky and colleagues [15] took advantage of the high sensitivity of fixations to sudden changes in the environment by suggesting a model for hazard detection, based on the instantaneous increase of fixation durations. Furthermore, it has been suggested that this particular feature of visual fixations can be used as a probe, providing access to different processing modes simply by a systematic presentation of such distractors [32].

The second experiment was concerned with the focus of visual attention by differentiating between the physical location of the eye and the subjective impression of what has been sampled from a visual display. Two main findings can be identified. A correspondence between the direction of the eye and the allocation of attention cannot always be assumed. In contrast to previous studies [19], our analysis is based on a continuous visual task, which can be understood as a viewing task close to natural gaze behaviour. The other key finding is related to the influence of the task. By differentiating between identification and localisation, we discovered a systematic relationship between the position of the eye on the display and the allocation of visual attention. More precisely, we found that in a task that requires identification—which is usually a more demanding and slower process than localisation—there is a higher probability that processing lags behind the actual position of the eye, whereas the opposite is the case when only localisation is required. Thus, when employing gaze behaviour in the dialogue with technical devices, for instance in the context of attentive interfaces, a careful consideration of the task is required. Inferring the allocation of attention solely on the basis of gaze direction can be misleading because visual attention can be ahead or behind the position of the eye. Even if these mismatches in synchronization might have a magnitude of several milliseconds only, they need to be considered for an optimal design of such attentive interfaces.

The final study investigated the contribution of gaze in interactive settings where direct communication is impaired due to spatial separation. Employing gaze transfer in remote collaboration has been a question of interest for research as well as for application. Research questions about gaze transfer are mainly concerned with the problem of how much information can be provided by transferring eye movements and what inferences are possible on the side of the receiver [28]. The inability to find advantages of attention transfer over purely intentional forms of spatial referencing poses serious questions about the information required in the process of establishing a shared understanding [33]. From a more applied perspective, the efficient transfer of knowledge (i.e. expertise) across long distances is of particular importance. The conclusion from the present experiment is—similar to our second study—that task characteristics have to be taken into account when applying gaze transfer. It is possible to follow a mouse cursor almost blindly but when using gaze cursors it is of paramount importance that the recipient perceives them in relation to the corresponding environment. This evokes two further questions: When applying gaze transfer in a more complex task than the one we used here, for instance when specific skills or expertise have to be communicated, which cursors (gaze or mouse) would allow for a better knowledge transmission? It might be the case that seeing someone's gaze behaviour always provokes at least a minimum of interpretation activity, compared with a more mechanistic following of the mouse movement. Thus, the first case is preferable when knowledge transfer is intended. Another important issue is the workload on the side of the expert/transmitter. It can be expected that gaze transfer is less of a burden than requiring explicit mouse movements all the time. Finally, one can even think of reversing the whole paradigm: Why not transfer the novice's gaze to the expert? Making the gaze behaviour of the novice available should allow the expert to easily identify critical instances where support is required.

On the basis of three concrete examples, the research presented here illustrates the potential contribution of human eye movements to understanding mechanisms of information processing. Together with the advantages, we highlighted possible pitfalls in the interpretation and application of eye gaze analysis. Only a careful implementation—considering the many facets of gaze behavior—will allow using the potential of eye movements when developing cognitive technical systems.

Acknowledgments. We are grateful to Cathy Nangini for valuable suggestions. This research was supported by the FP7-PEOPLE-2009-IEF program #254638 to SP, a grant of the TU Dresden Centre for Continuing Education to RM, and the Russian Foundation for Basic Research (#09-06-12003) to BMV.

References

1. Cavanagh, P.: Attention routines and the architecture of selection. In: Posner, M.I. (ed.) Cognitive Neuroscience of Attention, pp. 13–28. Guilford Press, New York (2004)
2. Henderson, J.M.: Regarding scenes. Current Directions in Psychological Science 16, 219–222 (2007)
3. Land, M.F.: Eye movements and the control of actions in everyday life. Progress in Retinal and Eye Research 25, 296–324 (2006)
4. Findlay, J.M.: Active vision: Visual activity in everday life. Current Biology 8, R640–R642 (1998)
5. Aloimonos, J., Weiss, I., Bandyopadhyay, A.: Active Vision. International Journal of Computer Vision 1, 333–356 (1987)
6. Matin, E.: Saccadic suppression: A review and an analysis. Psychological Bulletin 81, 899–917 (1974)
7. Kowler, E.: Eye movements: the past 25 years. Vision Research 51, 1457–1483 (2011)
8. Rayner, K.: Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology 62, 1457–1506 (2009)
9. Vertegaal, R., Shell, J.S., Chen, D., Mamuji, A.: Designing for augmented attention: Towards a framework for attentive user interfaces. Computers in Human Behavior 22, 771–789 (2006)
10. Senders, J.W.: Four theoretical and practical questions. Keynote address presented at the Eye Tracking Research and Applications Symposium 2000. In: Duchowski, A.T. (ed.) Proceedings of the Symposium on Eye Tracking Research & Applications, p. 8. ACM Press, Palm Beach Gardens (2000)
11. Velichkovsky, B.M.: Heterarchy of cognition: The depths and the highs of a framework for memory research. Memory 10, 405–419 (2002)
12. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124, 372–422 (1998)
13. Castelhano, M.S., Mack, M.L., Henderson, J.M.: Viewing task influences eye movement control during active scene perception. Journal of Vision 9(6), 1–15 (2009)
14. Reingold, E.M., Stampe, D.M.: Saccadic inhibition and gaze contingent research paradigms. In: Kennedy, A., Radach, R., Heller, D., Pynte, J. (eds.) Reading as a Perceptual Process, pp. 1–26. Elsevier Science Ltd. (2000)
15. Velichkovsky, B.M., Rothert, A., Kopf, M., Dornhoefer, S.M., Joos, M.: Towards an express diagnostics for level of processing and hazard perception. Transportation Research, Part F 5, 145–156 (2002)
16. Pannasch, S., Velichkovsky, B.M.: Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. Visual Cognition 17, 1109–1131 (2009)
17. Levine, T.R., Hullett, C.: Eta-square, partial eta-square, and misreporting of effect size in communication research. Human Communication Research 28, 612–625 (2002)
18. Pannasch, S., Dornhoefer, S.M., Unema, P.J.A., Velichkovsky, B.M.: The omnipresent prolongation of visual fixations: Saccades are inhibited by changes in situation and in subject's activity. Vision Research 41, 3345–3351 (2001)
19. Schneider, W.X., Deubel, H.: Selection-for-perception and selection-for-spatial-motor-action are coupled by visual attention. In: Prinz, W., Hommel, B. (eds.) Attention and Performance XIX: Common Mechanisms in Perception and Action, pp. 609–627. Oxford University Press, Oxford (2002)

20. Tatler, B.W.: Characterising the visual buffer: real-world evidence for overwriting early in each fixation. Perception 30, 993–1006 (2001)
21. Fischer, M.H.: An investigation of attention allocation during sequential eye movement tasks. The Quarterly Journal of Experimental Psychology: Human Experimental Psychology 3, 649–677 (1999)
22. Carr, T.H.: A multilevel approach to visual attention. In: Posner, M.I. (ed.) Cognitive Neuroscience of Attention, pp. 56–70. The Guilford Press, New York (2004)
23. Deubel, H., Irwin, D.E., Schneider, W.X.: The subjective direction of gaze shifts long before the saccade. In: Becker, W., Deubel, H., Mergner, T. (eds.) Current Oculomotor Research: Physiological and Psychological Aspects, pp. 65–70. Plenum, New York (1999)
24. Frischen, A., Bayliss, A.P., Tipper, S.P.: Gaze Cueing of Attention: Visual Attention, Social Cognition, and Individual Differences. Psychological Bulletin 133, 694–724 (2007)
25. Schrammel, F., Pannasch, S., Graupner, S.-T., Mojzisch, A., Velichkovsky, B.M.: Virtual friend or threat? The effects of facial expression and gaze interaction on psychophysiological responses and emotional experience. Psychophysiology 46, 922–931 (2009)
26. Hill, E.J., Ferris, M., Märtinson, V.: Does it matter where you work? A comparison of how three work venues (traditional office, virtual office, and home office) influence aspects of work and personal/family life. Journal of Vocational Behavior 63, 220–241 (2003)
27. Velichkovsky, B.M.: Communicating attention: Gaze position transfer in cooperative problem solving. Pragmatics and Cognition 3, 199–222 (1995)
28. Mueller, R., Helmert, J.R., Pannasch, S., Velichkovsky, B.M.: Improving remote cooperation in spatial tasks: Benefits and pitfalls of the gaze transfer approach. Quarterly Journal of Experimental Psychology (in revision)
29. Marwan, N., Kurths, J.: Nonlinear analysis of bivariate data with cross recurrence plots. Physics Letters A 302, 299–307 (2002)
30. Land, M.F., Tatler, B.W.: Looking and Acting: Vision and Eye Movements during Natural Behaviour. Oxford University Press (2009)
31. Helmert, J.R., Pannasch, S., Velichkovsky, B.M.: Influences of dwell time and cursor control on the performance in gaze driven typing. Journal of Eye Movement Research 2(3), 1–8 (2008)
32. Pannasch, S., Schulz, J., Velichkovsky, B.M.: On the control of visual fixation durations in free viewing of complex images. Attention, Perception, & Psychophysics 73, 1120–1132 (2011)
33. Clark, H.H., Brennan, S.E.: Grounding in communication. In: Resnick, L.B., Levine, J.M., Teasley, S.D. (eds.) Perspectives on socially shared cognition, pp. 127–149. APA Books, Washington, DC (1991)