Acta Psychologica 152 (2014) 19–28

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Limitations of gaze transfer: Without visual context, eye movements do not to help to coordinate joint action, whereas mouse movements do

Romy Müller *, Jens R. Helmert, Sebastian Pannasch

Department of Psychology, Engineering Psychology and Applied Cognitive Research, Technische Universitaet Dresden, Germany

ARTICLE INFO

Article history:
Received 7 November 2013
Received in revised form 10 July 2014
Accepted 17 July 2014
Available online xxxx

PsycINFO classifications:
2700
3040

Keywords:
Gaze transfer
Remote cooperation
Visual attention
Grounding

ABSTRACT

Remote cooperation can be improved by transferring the gaze of one participant to the other. However, based on a partner's gaze, an interpretation of his communicative intention can be difficult. Thus, gaze transfer has been inferior to mouse transfer in remote spatial referencing tasks where locations had to be pointed out explicitly. Given that eye movements serve as an indicator of visual attention, it remains to be investigated whether gaze and mouse transfer differentially affect the coordination of joint action when the situation demands an understanding of the partner's search strategies. In the present study, a gaze or mouse cursor was transferred from a searcher to an assistant in a hierarchical decision task. The assistant could use this cursor to guide his movement of a window which continuously opened up the display parts the searcher needed to find the right solution. In this context, we investigated how the ease of using gaze transfer depended on whether a link could be established between the partner's eye movements and the objects he was looking at. Therefore, in addition to the searcher's cursor, the assistant either saw the positions of these objects or only a grey background. When the objects were visible, performance and the number of spoken words were similar for gaze and mouse transfer. However, without them, gaze transfer resulted in longer solution times and more verbal effort as participants relied more strongly on speech to coordinate the window movement. Moreover, an analysis of the spatio-temporal coupling of the transmitted cursor and the window indicated that when no visual object information was available, assistants confidently followed the searcher's mouse but not his gaze cursor. Once again, the results highlight the importance of carefully considering task characteristics when applying gaze transfer in remote cooperation.

1. Introduction

1.1. Benefits and difficulties of gaze transfer

Remote cooperation poses a challenge in coordinating joint action, due to a lack of some nonverbal cues that are typically present during natural communication. When interacting face-to-face, eye movements play a significant role in avoiding misunderstandings. They are closely linked to processes of visual attention (Just & Carpenter, 1976) and therefore can assist in establishing a joint focus of attention (Bruner, 1981) and inferring the object of a partner's referring expressions (Hanna & Brennan, 2007). Trying to emulate this natural function of the eyes, one approach to resolving ambiguities in remote cooperation is to superimpose a person's gaze on the partner's screen as a cursor. Such gaze transfer has been shown to improve performance compared to purely verbal interaction during joint visual search (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; Neider, Chen, Dickinson, Brennan, & Zelinsky, 2010) and cooperative problem solving (Velichkovsky, 1995). These gaze benefits have been explained by the potential of eye movement information to support a smoother coordination between the partners, enabling them to avoid redundant search (Brennan et al., 2008) or to use brief deictic verbal references instead of elaborate object descriptions (Neider et al., 2010; Velichkovsky, 1995).

It seems quite intuitive to assume that eye movements as an indicator of a person's visual attention should make it easier to understand what that person is doing or trying to communicate, and indeed, eye movements have been labelled as a “window into mind” (Velichkovsky & Hansen, 1996). However, there is recent evidence suggesting that gaze transfer is not as unproblematic as previous studies might suggest (Müller, Helmert, Pannasch, & Velichkovsky, 2013). In a joint puzzle task, the transfer of a gaze cursor from an expert to a novice was compared not only with purely verbal interaction but also with simple mouse pointing. Although there were no differences in overall performance between both types of cursor transfer, the mouse clearly outperformed gaze with regard to the cooperative process: Reactions to the gaze cursor were slower overall, and participants were especially hesitant to react to it in situations where this reaction posed the risk of

* Corresponding author at: Department of Psychology, Engineering Psychology and Applied Cognitive Research, Technische Universitaet Dresden, Helmholtzstraße 10, 01069 Dresden, Germany. Tel.: +49 351 463 35330.
E-mail address: romy.mueller@tu-dresden.de (R. Müller).

http://dx.doi.org/10.1016/j.actpsy.2014.07.013

an error. Along the same lines, using gaze increased the effort participants incurred while verbally referring to puzzle pieces. Taken together, these results are indicative of uncertainties about the gaze cursor's communicative function: It appears to be hard to determine whether a particular gaze is intended to be an instruction or merely a part of the person's search process.[1]

However, it would be premature to conclude that gaze transfer was only distracting without adding any value, because a puzzle tasks might not be optimally suited for the benefits of gaze transfer to emerge. There are at least two reasons for this. First, solving puzzles does not necessarily require a partner, which might have reduced participants' motivation to pay close attention to that partner's search process. Second, information about the search process is of limited value when the pieces are randomly distributed and only the correctly identified target piece needs to be communicated. In this setting, gaze is reduced to an intentional pointer. This inevitably raises the question in what way the abundance of information contained in eye movements could have led to any benefits at all. At the same time, these considerations raise the possibility that gaze transfer might be more helpful in settings in which the value of the information contained in it goes beyond that of explicit instructing.

1.2. Harnessing the potentials of eye movements in cooperative tasks

In order to conduct a more appropriate comparison of gaze and mouse transfer, a genuinely cooperative task is required in which two interdependent partners need to be informed about each other's ongoing activity and visual attention. To construct such a task, we looked at real-world cooperative settings to extract their underlying mechanisms and apply them in a controlled laboratory experiment. One outstanding feature characterizing many cooperative settings is that participants assume asymmetric roles and perform complementary actions, with one of them creating the conditions for the other one to act. For example, consider a driver of a lifting ramp who moves that ramp to different parts of a building, enabling a window cleaner standing on the ramp to clean the windows in different areas. In more general terms, an assistant is providing the framework for his partner to work on the details of the joint task. Although the assistant does not necessarily possess the abilities to work on these details by himself, he certainly has to take his partner's activities into consideration in order to align his own actions to them and provide effective support.

We adopted the general idea of an assistant controlling his partner's workspace as it occurs in the lifting ramp scenario sketched above. In our computerized, abstract version of it, an assistant was in charge of moving his partner's field of view with his mouse while the rest of the screen was occluded, similar to the *moving window paradigm* (McConkie & Rayner, 1975). Thus, in our joint moving window (JMW) task, the partner (henceforth called *searcher*) was only able to see and act in a small, rectangular frame, while the rest of the display was covered by a black mask. Under these conditions, it is crucial that the assistant moves the window according to the searcher's needs. In the real world counterpart of the window cleaner on the lifting ramp, this knowledge may stem from the driver observing what parts of the building are currently being worked on and which are finished already, perhaps even inferring how long an ongoing action will take. In the JMW task, this is where gaze transfer comes into play. Eye movement parameters, especially fixation durations, are task-dependent (Land & Tatler, 2009; Rayner, 1998) and indicative of a person's mental processing activities (Velichkovsky, 2002). Thus, if the cognitive and spatio-temporal requirements vary over the course of a task, the searcher's eye movements presumably can be used as a cue to his ongoing activities. Therefore, we provided the assistant with a depiction of the searcher's gaze cursor.

The nature of this gaze transfer differs from that of the puzzle task (Müller et al., 2013), where eye movements had been used as a means of intentional communication. The results had indicated that observers cannot easily infer the communicative function of particular gaze instances, corroborating the suggestion from Human Computer Interaction (HCI) research that using isolated fixations as explicit commands can be problematic (Jacob, 1991). Alternatively, gaze transfer can reflect the person's "viewing behaviour" in a broader sense, serving as an indicator of his interest and ongoing activity, which presumably can aid an assistant to find out how to act in the most helpful way (cf. Qvarfordt, Beymer, & Zhai, 2005). When applied in this way, gaze transfer should visualize aspects of the partner's solution process that are not accessible when only looking at his intentional, manual actions (Ballard, Hayhoe, Li, & Whitehead, 1992). Therefore, in the JMW task we used gaze transfer as a byproduct of the actual solution process, directly representing the searcher's purposeful activity. To test whether this can lead to a *specific* benefit beyond that of a mere spatial indicator, we compared gaze transfer with mouse transfer. We did not include a condition without cursor transfer, because there already is plenty of evidence for benefits of gaze transfer over purely verbal interaction (Brennan et al., 2008; Müller et al., 2013; Neider et al., 2010; Velichkovsky, 1995). Thus, repeating this comparison in a task that is even more suited for gaze transfer did not seem particularly interesting.

While the function of gaze transfer in the JMW task differs from that in a puzzle task by not being an explicit instruction, it differs from collaborative search (Brennan et al., 2008; Neider et al., 2010) for the opposite reason. During collaborative search, the partner's gaze can be used as a source of additional information that is monitored peripherally while two people are doing the same thing in parallel. Conversely, the JMW task is a decision making task in which two partners contribute their own specific abilities in order to reach a joint solution. It requires the assistant to understand the searcher's gaze in terms of the underlying cognitive processes and the joint goal. To increase this role of inferring cognitive processes and activities from gaze even more, our task was composed of several component operations that differed in the required level of processing (Craik & Lockhart, 1972) and the corresponding eye movement parameters (Velichkovsky, 2002). Specifically, the searcher had to perform colour discrimination, count the number of objects of different shapes and calculate the sum of numbers. A prestudy confirmed that these subtasks differed in the eye movement parameters they produced, with longer fixations and smaller saccades for the latter subtasks. Thus, in the present paradigm, closely observing the partner's eye movements should be informative about his current activities.

There is a significant boundary condition for the usability of any indicator of task-related mental processing: The solution process itself must be comprehensible. This requires certain knowledge about the necessary actions and the way they relate to task-relevant objects in the environment. In this context, a particularly interesting variable is the visibility of task-relevant objects for the recipient of gaze transfer. In principle, it is possible that gaze can be used to infer the partner's locus of attention in a merely spatial manner, which could be concluded from previous studies using joint visual search (Brennan et al., 2008; Neider et al., 2010). However, when applying gaze transfer to support the small-scale coordination of joint action, the ability to make inferences about a partner's visual attention should depend on knowing what he is attending *to*. Therefore, we varied whether the assistant was provided with partial information about the stimulus material (i.e. the relevant screen areas and object locations), or no information.

[1] Note that there is a confound when ascribing differences between gaze and mouse to the information they transfer: The cursors also differ in terms of their movement profiles. Therefore, one could argue that gaze transfer might not be difficult to interpret per se but simply too fast and variable, and in that way more distracting. These two accounts, i.e. the amount of information versus movement, cannot easily be disentangled, because additional information (e.g. about search processes) necessarily manifests in movement. An option would be to vary the characteristics of the cursor movement somewhat, for example by smoothing gaze or adding random noise to the mouse cursor. Such manipulations are beyond the scope of the present paper, but the issue should be kept in mind when interpreting the results.

1.3. Research questions and hypotheses

The present study aimed at answering three questions. First, will gaze transfer be more beneficial than a merely intentional form of communication to support the coordination of joint action when the partner's visual attention is of direct relevance to the task? Second, what are the boundary conditions for people's ability to use a partner's eye movements to guide their own actions? Specifically, does the ability to understand and apply gaze transfer depend on the observer's abilities to relate it to task-relevant objects? Third, we were interested in the exact ways in which gaze transfer versus intentional mouse communication affected the coordination of joint action in terms of verbal dialogues and the spatio-temporal coupling between the searcher's cursor and the assistant's window movement.

The assumption underlying the present paradigm was the following: If a searcher's eye movements support the assistant in understanding the search process, he should be able to move the window in a way that better enables the searcher to complete his task. Note that this way of evaluating gaze transfer differs from previous studies which directly tested the effects of gaze visualizations on the performance of the observer (Mehta, Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005; Stein & Brennan, 2004; Velichkovsky, 1995). In contrast, the JMW task uses a more indirect approach, testing how gaze transfer can reflect back on the producer himself via the effects it has on an observer's actions, which in turn set the stage for the producer's own performance. In that way, the present paradigm is more similar to investigations of gaze input in HCI research (e.g. Vertegaal, 2008). In these studies, a user's gaze is interpreted by a computer which then responds in a particular way, and ultimately the effectiveness of gaze input is measured in terms of the user's performance in interacting with the computer to solve a particular task. Similarly, in the JMW task there are a number of mutual dependencies, with the searcher's gaze determining the assistant's response and that response influencing the performance of the searcher. This certainly makes the task more complex, but it emphasizes the genuinely *interactive* role of gaze in joint action.

To investigate the effects of gaze transfer on joint action, we analysed solution times, the amount and content of verbal utterances and the coordination of the window movement with the searcher's gaze or mouse cursor. In terms of performance and speech, we hypothesized that if the assistant is able to understand and react to gaze transfer, this will lead to faster solutions and less verbal effort of the searcher (particularly verbal utterances related to the window movement). However, such benefits might be buried in difficulties stemming from the fact that the searcher's gaze is used for an active manipulation of his field of view. Even if the task-relevant objects remain visible within the window, the partly unpredictable movement in the searcher's visual periphery may be disruptive to visual sampling and attentional focusing (Kawahara, Yanase, & Kitazaki, 2012). Therefore, whether the window movement is harmful or not will presumably depend on the assistant's ability of tailoring that movement to the searcher's needs. Searchers might try to support this by making a higher verbal effort to avoid misunderstandings and prevent unwanted window movements in the gaze condition. To test this, we compared the amount of explicit verbal feedback utterances and expected more feedback during gaze as compared to mouse transfer.

Another factor that might put the relative merit of gaze transfer into doubt is the way in which the searcher uses his mouse. It is conceivable that mouse movements can also be applied to indicate search processes, namely by simulating one's own gaze. In this case, gaze benefits might not arise at all, because the information contained in both cursors would be highly similar. To test for such strategies, we analysed the spatio-temporal correspondence between searchers' mouse cursors in the mouse transfer conditions and their (not transmitted) eye movements.

In terms of understanding the influence of gaze transfer on joint action, one of the most interesting aspects is the way in which it affects the small-scale coordination between both partners. For example, does the gaze cursor lead assistants to follow it more or less closely than they follow the mouse cursor? On the one hand, it could be expected that the mouse leads to a closer following, because it serves a purely communicative purpose in the JMW task (i.e. indicating “Move the window here, please!”). On the other hand, assistants might follow the gaze cursor more closely instead, continuously tracking it with their mouse (corresponding to the window centre) in an attempt to always provide optimal viewing conditions for the searcher. Whereas this would certainly be a rather challenging task, it might make it very easy to find the right sections and objects, as our prestudy with a gaze-contingent window suggested. To quantify the spatio-temporal coupling of the transmitted cursor and the window movement, we calculated their distances at different temporal delays.

Our last goal was to test how the visibility of the object locations for the assistant affected all the above parameters. Generally, visualizing task-relevant objects and the partners' manipulations of them can enable a better coordination of joint work (Whittaker, 1995, 2003), which is why we expected overall solutions to be faster with visible objects. Moreover, it can be assumed that object visibility will affect gaze transfer specifically, because it presumably determines the assistant's ability to make sense of the cursor movement. Object information may support a segregation of the continuous eye movement signal into meaningful units, enabling their interpretation as specific, object-directed gazes. This, in turn, could facilitate decisions about how and when to move the window. Conversely, as these cues for sequencing the eye movements are unavailable in the invisible condition, comparable performance should only be possible if assistants were to follow *all* of the gaze cursor's movements. However, this is highly unlikely, given the speed and unpredictability of eye movements. Therefore, we expected difficulties in using gaze transfer without object information.

2. Methods

2.1. Subjects

Forty-eight students of the Technische Universität Dresden (32 females) in the age range of 18–51 years ($M = 23.9$, $SD = 6.1$) participated in the experiment. They were invited in pairs and randomly assigned to one of the two experimental roles (searcher or assistant), resulting in a total of 24 pairs. All participants had normal or corrected-to-normal vision, spoke fluent German and received course credit or a payment of 5€ per hour.

2.2. Apparatus

Both participants sat in the same room, separated by a portable wall and with their computers being connected via Ethernet. Eye movements of the searcher were recorded monocularly at 500 Hz using the SR EyeLink 1000 infrared eye tracking system (SR Research Ltd., Ontario, Canada) in the remote recording mode.

2.3. Stimuli

Stimuli were presented on two CRT displays (19 inch Samtron 98 PDF) with a resolution of 1024 × 768 px at a refresh rate of 100 Hz. The stimulus set consisted of twenty images which were composed of 20 × 20 rectangles (each with a size of 51 × 38 px), forming six equally sized sections that originated from the centre of the screen (see Fig. 1B). In each image, there were three red and three green sections, with the colour always separating adjacent sections. The order of the section colours was counterbalanced between the images. Each section contained a variable number of 5–10 circles and triangles (with a minimum of 1 and a maximum of 7 objects of the same shape per section). Within

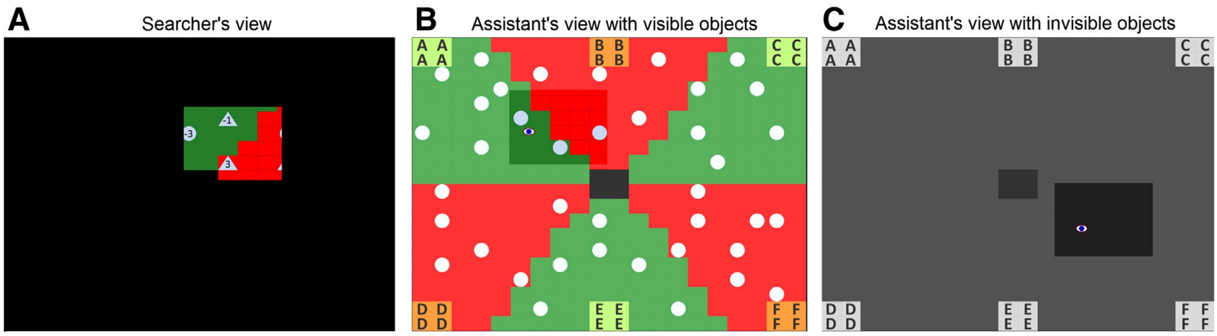

Fig. 1. Stimulus material for searcher (A) and assistant with the objects visible (B) and invisible (C).

the shapes there were positive and negative one-digit numbers (1–9) with a ratio of 3:2. Numbers were chosen randomly, with the restrictions that their sum was between 0 and 20 and that a particular number was not repeated more than three times. In the four corners and in the middle of the upper and lower edges of the screen, there were quadratic solution boxes labelled with the letters A–F, which had to be clicked for selecting the respective section.

The *searcher* had access to all the stimulus information described above, but only within a viewing window of 255 × 190 px (about 7.9 × 5.9° of visual angle, or 1/16 of the screen). The remainder of the screen was covered with a black overlay (see Fig. 1A). In contrast, the visual information for the *assistant* varied between the two object visibility conditions: With the objects *visible*, all objects were depicted as empty circles, but their locations and the section colours were preserved (see Fig. 1B). With the objects *invisible*, the assistant only saw a grey background depicting the grid of 20 × 20 rectangles, as well as the solution boxes, but no objects or sections were presented to him (see Fig. 1C). The assistant had visual access to the entire screen, inside as well as outside the window. To guarantee that he knew the exact location of the window, the outside area was overlaid with a transparent grey layer and the window centre was indicated by his mouse cursor.

Depending on the experimental condition, the searcher's gaze or mouse cursor was projected onto the assistant's screen. In both cursor conditions, the cursor was refreshed every 50 ms, and in the gaze condition the raw signal from the eye tracker was used without filtering or smoothing. Both cursors were presented as a tricolour eye-icon. The same icon was used to indicate the searcher's gaze and mouse in order to keep visual attributes of the cursor constant and prevent confusion for the assistant between his own and the searcher's mouse cursor in the mouse condition. The eye-icon was chosen to make this differentiation very salient for the assistant, but subjects were explicitly instructed in the mouse conditions that the cursor reflected the searcher's mouse, not his gaze.

2.4. Procedure

The experiment consisted of four blocks, each being composed of one practice trial and five experimental trials, with the blocks corresponding to combinations of the experimental conditions (see below). Block order was counterbalanced between participants so that each combination was run once. The order of the stimuli was randomized over the experiment, resulting in a unique mapping of stimuli to blocks for each pair in order to avoid any systematic differences in task difficulty between the experimental conditions. Before each block, a nine-point calibration and validation of the eyetracker was performed and could be repeated between the trials if necessary. To start a trial, both participants had to press the space key.

The basic task in all conditions required participants to determine the correct section in a stepwise manner: First they were instructed to select the three red sections and ignore the green ones, then they had to exclude the one with the least amount of circular objects, and finally they had to decide which one of the remaining two sections had the smaller sum when adding up all numbers contained in the objects. This section had to be selected via a mouse click by the searcher on the respective target field of this section, which automatically terminated the trial and produced a feedback about the correctness of the solution.

At the beginning of each trial, the window centre and the searcher's mouse were positioned in the middle of the screen. The assistant who was in control over the window had to move it across the screen in order to reveal the image parts the searcher needed to fulfill the task. As the window was locked to the assistant's mouse position, each of his mouse movements resulted in a corresponding window movement. Mouse clicks were neither necessary nor were they visualized. No hints or instructions were given about strategies of moving the window or using the cursor transfer. Participants were free to verbally interact in all experimental conditions, and these verbal interactions were recorded. In addition, the searcher's gaze position was transferred to the assistant in the gaze condition and his mouse position was transferred in the mouse condition. In each experimental condition, the eye movements of the searcher as well as the mouse movements of both partners were recorded, regardless of their transmission. In both transfer conditions both partners were informed which cursor the assistant would see. No strategy was instructed or suggested, and both participants were explicitly told that they were free to decide how to use the cursor transfer.

Lastly, the visibility of the objects was varied: With the objects visible, the assistant saw the section colours and object locations but not their shape, whereas in the invisible condition he only saw the grey background (see Fig. 1B and C). As the section layout was identical throughout the experiment, conditions with partial spatial information, showing only colours or objects, were not expected to produce results that greatly differ from the visible or invisible conditions, respectively. Therefore, no such conditions were included.

2.5. Data analysis

Performance and eye movement data were subjected to 2 (*Cursor transfer: gaze, mouse*) × 2 (*Object visibility: visible, invisible*) repeated measures ANOVAs. The analyses of the verbal interactions additionally included the experimental role (*assistant, searcher*) as a between subjects factor. Further analyses will be described in the corresponding sections of the text. Post hoc comparisons were performed with Bonferroni correction.

3. Results

3.1. Performance

To compare task performance between the experimental conditions, we analysed the mean solution times. We did not exclude error trials from this analysis, because error rates were high (19.4%) and the

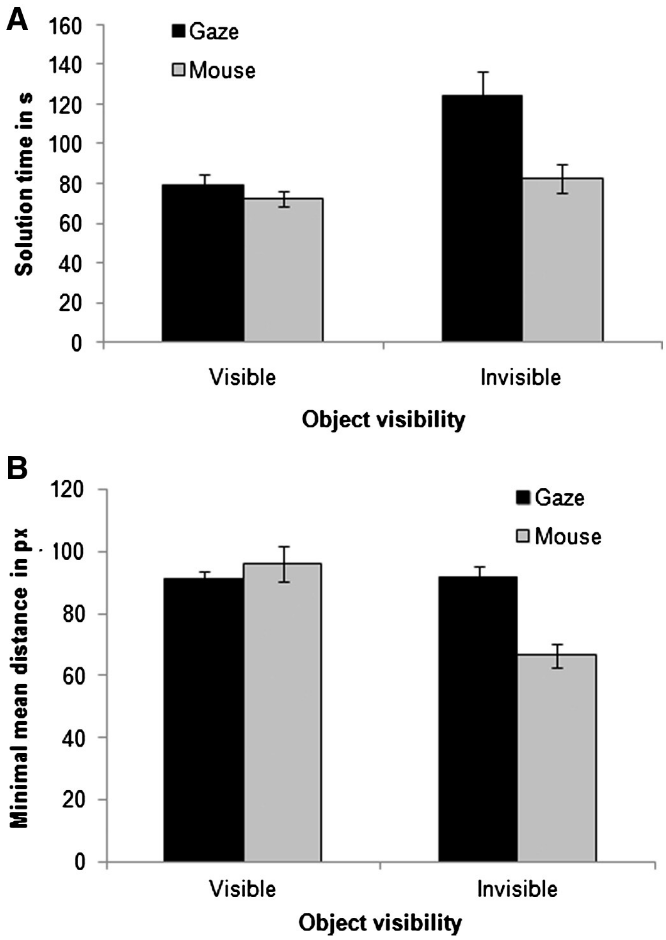

Fig. 2. Mean solution times (A) and minimal mean distances between the window centre and the searcher's cursor (B), depending on cursor transfer and object visibility. Error bars represent standard errors of the mean.

searchers' speech during the experiment suggested that at least in trials in which they verbalized their calculations, errors were almost exclusively due to mistakes in adding up the numbers, which presumably is not related to the process of coordinating joint action. There were main effects of cursor transfer, $F(1,23) = 22.60$, $p < .001$, and object visibility, $F(1,23) = 14.69$, $p < .001$, as well an interaction between both factors, $F(1,23) = 11.89$, $p = .002$. Solution times were shorter in mouse than in gaze (77.5 vs. 101.9 s) and shorter with the objects visible than invisible (75.9 vs. 103.4 s). A significant difference between mouse and gaze was only obtained when the objects were invisible, $p < .001$, but not when they were visible, $p = .153$ (see Fig. 2A).[2]

We did not expect a difference in the error rates in this task as they presumably depend more on the searcher's counting and calculating performance than on the cursor and object visibility. Indeed, error rates (19.4%) did not significantly differ between gaze and mouse, $F(1,23) = 2.58$, $p = .122$. Also, there was no main effect of object visibility and no interaction, both Fs < 1, both ps > .9.

[2] As most people are used to observing another person's mouse but not his gaze, it is possible that the superior performance for mouse transfer was only due to a difference in experience and would disappear with some practice. Therefore, we investigated decreases in solution time between the experimental conditions by including the factor *trial* within a block (1–5) into the ANOVA. There was a weak interaction of cursor and trial, $F(4,92) = 2.53$, $p < .046$, indicating a trend for learning effects in gaze but not in mouse. However, for neither cursor type the solution time decrease across trials was significant, both ps > .05. Critically, the interaction between trial, cursor and object visibility did not approach statistical significance, $F(4,92) = .39$, $p = .818$, making it unlikely that the slow solution times in the gaze/invisible condition would have disappeared with just a little more practice.

3.2. Inter- and intra-individual gaze-mouse coordination

In addition to mean performance, we investigated how the coordination of both partners' actions depended on the type of cursor transfer and object visibility. Initial analyses revealed that the overall amount of window movement did not significantly differ between blocks with gaze and mouse cursor transfer. To determine the interplay between the movement of the transferred cursor and the window, we calculated the positional coupling between the assistant's cursor (corresponding to the window centre) and the searcher's cursor (either representing his gaze or mouse). Therefore, we used the spatial positions of these two cursors at each sampling interval to compute their mean spatial distance over the trial. Then we shifted the time series of the cursor positions against each other in a stepwise manner to compute the mean distance at each temporal delay. Five pairs of subjects were excluded from these analyses as the searchers of these pairs had put aside their mouse in at least one of the mouse conditions at the beginning of the block and were communicating in a purely verbal manner. Obviously, this makes a meaningful analysis of the dependencies between both partners' cursor movements impossible.

The size of the minimal mean distance (i.e. the mean distance at the temporal delay at which the window and the cursor were least far apart) serves to determine how closely the assistant is following the searcher by moving the window to the location of his cursor. An analysis of these distances revealed a main effect of cursor transfer, $F(1,18) = 15.92$, $p < .001$, a main effect of object visibility, $F(1,18) = 16.36$, $p < .001$, and an interaction between them, $F(1,18) = 29.88$, $p < .001$. Overall, distances were higher in gaze than in mouse (91 vs. 81 px), and higher when the objects were visible then when they were invisible (93 vs. 79 px). However, the higher distances for gaze than mouse were restricted to the invisible condition, $p < .001$, while with visible objects distances were similar for gaze and mouse, $p = .252$. On the other hand, only for mouse we were able to observe smaller minimal distances in the invisible case, $p < .001$, whereas in gaze there was no significant difference between visibility conditions, $p = .843$. This result indicates that with the object information available, assistants were following the gaze and the mouse cursor in a comparable manner, whereas in the absence of visual object information, they followed the searchers' mouse cursor more closely, but they did not do so for the gaze cursor (see Fig. 2B).

As outlined in the Introduction, one possibility to provide the assistant with information about one's visual attention in the mouse condition is for the searcher to continuously move the mouse to the positions he is visually inspecting. Such a strategy should be particularly helpful when the objects are invisible. In this case, the assistant is literally "tapping in the dark", and in order to decide where to move the window, he may benefit from any additional information about the searcher's current activity. Indeed, a comparison of the searchers' cumulated mouse paths revealed more extensive mouse movement in the invisible than in the visible condition (11,463 vs. 8912 px per trial), $t(23) = -2.24$, $p = .035$. Moreover, if searchers are tracking their own eyes with their mouse cursor in this condition, this leads to two predictions. First, the minimal mean distance between the searcher's gaze and mouse should be lower. Indeed, this is what we found (90 vs. 101 px for mouse/invisible and mouse/visible, respectively), $t(18) = 2.10$, $p = .050$. Again, the analysis was performed only on the searchers who actually used their mouse in the mouse conditions. Moreover, among the remaining subjects it was only one who caused the effect to be so relatively weak. Although this subject had not completely put aside his mouse, he only moved it casually for a few moments within each trial. When excluding him from the analysis, the difference between the visible and invisible conditions became more reliable, $t(17) = 3.77$, $p = .002$. Second, trying to simulate one's gaze with one's mouse should decrease the temporal delay at which this minimal distance occurs. Again, the data support this prediction: The searcher's mouse movements lagged behind his eyes less in the invisible condition (30 vs. 80 ms), $t(18) = 2.46$, $p = .024$. Thus, when the assistant could not see the objects, searchers kept their mouse cursor closer to where they fixated, both spatially and temporally.

3.3. Strategies of coordinating the window movement

The analyses of the minimal mean distances revealed a particularly strong coupling of the searcher's cursor and the window in the mouse/invisible condition. But does that apply for all pairs of subjects or did they use different strategies? Basically, there are two options: Either the assistant can follow the searcher (e.g. by moving the window to the cursor position) or the searcher can follow the assistant (e.g. by searching the area to which the assistant places the window autonomously). Such strategies could be inferred from the distribution of the delay times at which pairs reached their minimal cursor-window distances (see Fig. 3). Positive delays indicate that the assistant is following the searcher, whereas negative delays emerge when the assistant's window movement precedes the searcher's cursor.

As Fig. 3 shows, the distribution of delay times clearly differs between the experimental conditions. While there is a wide distribution for mouse/visible ($SD = 6.61$), the gaze/visible condition ($SD = 4.22$) tends to split into two groups: Some pairs reached their minimal distance at delay times of around 0, whereas others had positive delays, i.e. the cursor movement preceded that of the window. Contrary to the objects visible conditions (Fig. 3A and B), the invisible conditions (Fig. 3C and D) produced *lower* dispersions for mouse ($SD = 2.02$) than gaze ($SD = 3.58$). That is, pairs in the mouse/invisible condition were particularly homogenous in the temporal dynamics of their cursor-window coordination.

Furthermore, it needs to be noted that no pair in either of the invisible conditions had negative delays, i.e. the searcher's cursor movement always preceded that of the window. At first glance this seems to suggest that the gaze cursor was followed just like the mouse when the objects were invisible. However, positive delays do not necessarily indicate following of the cursor per se. A strategy of ignoring the gaze cursor and only following the searcher's verbal instructions will almost inevitably lead to positive delays as well. This is because people typically look at the objects they are talking about (Griffin & Bock, 2000), so that the searcher's gaze can be expected to be located at the place he is asking the assistant to open up. Therefore, delays should not be interpreted in isolation but only in combination with the minimal distance value. Only positive delays combined with low distances are strong evidence for a close following of the cursor itself. This pattern is exactly what can be observed in the mouse/invisible condition (see Fig. 3D).

It is possible that the strategy of using the gaze cursor affects performance. Therefore, we divided subjects into two groups for both gaze transfer conditions according to their cursor usage in the respective condition. We tried different criteria for defining the groups (observation during the experiment, assistants'subjective ratings, delays and distances of the cursor-window coupling). However, when comparing solution times between these strategy groups with t-tests in the four experimental conditions, we found no significant differences, no matter which criterion the grouping was based on, all $ts < 1.17$, all $ps > .2$.

3.4. Speech

Besides the observable activities of moving eyes and mouses, a main aspect of coordinating joint action is the way the partners verbally interact. For all our speech analyses, we excluded task-irrelevant talk as well as the subvocal murmuring some searchers did while adding up the numbers (clearly audible verbalizations of the intermediate results were included). To investigate the overall verbal effort the partners engaged in while solving the task, in the first step we compared the total

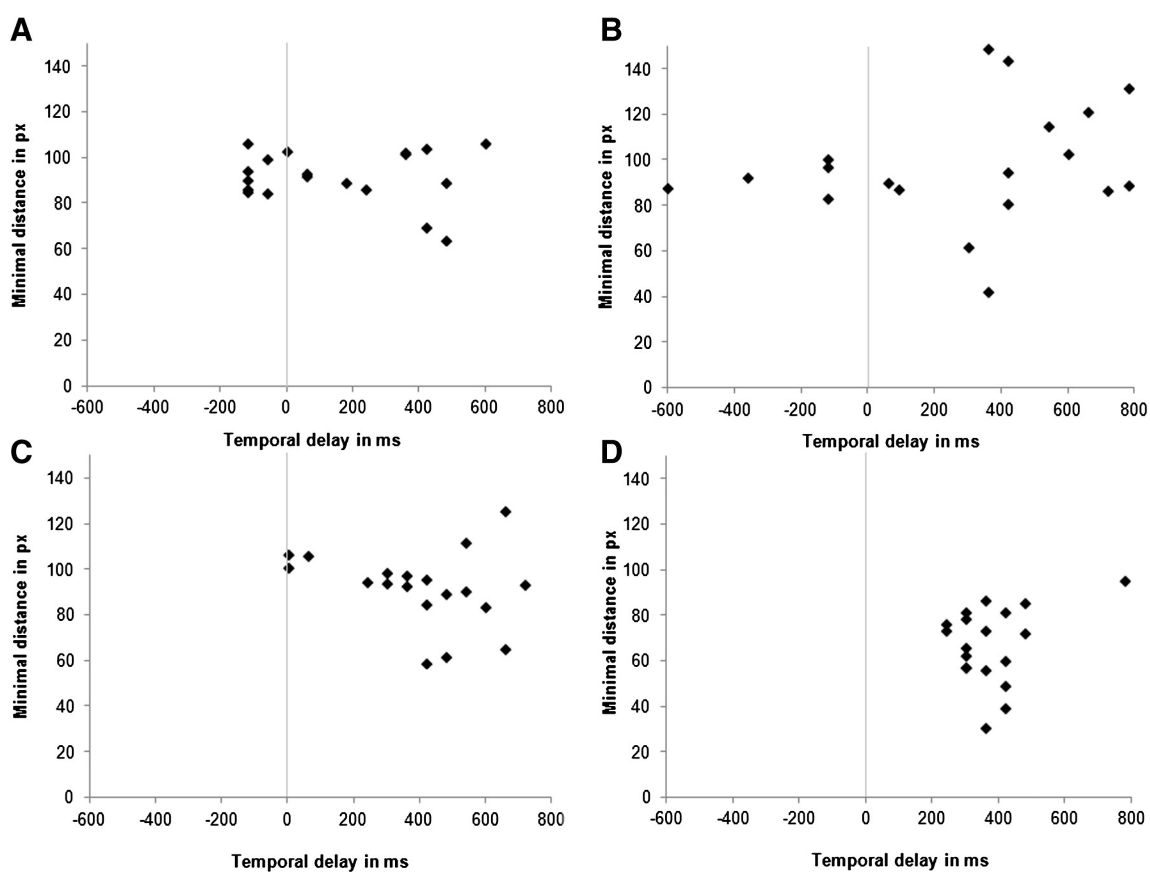

Fig. 3. Minimal mean distances and their corresponding temporal delays in Gaze/Visible (A), Mouse/Visible (B), Gaze/Invisible (C) and Mouse/Invisible (D). Each data point represents one pair of subjects. Positive delays indicate that the movement of the cursor was preceding that of the window (see text for details).

number of words. There were main effects of cursor transfer, $F(1,46) = 28.91$, $p < .001$, object visibility, $F(1,46) = 13.91$, $p < .001$, and experimental role, $F(1,46) = 32.60$, $p < .001$. Furthermore, we observed interactions between cursor transfer and object visibility, $F(1,46) = 15.61$, $p < .001$, cursor transfer and role, $F(1,46) = 15.61$, $p < .001$, as well as object visibility and role, $F(1,46) = 15.99$, $p < .001$. Finally, there was a triple interaction, $F(1,46) = 10.23$, $p = .003$. Pairs spoke more in gaze than in mouse (60.3 vs. 39.8 words), and they spoke more when the objects were invisible than when they were visible (58.5 vs. 41.6 words). However, more words in gaze than in mouse were only uttered with invisible objects, $p < .001$, whereas in the objects' visible condition, the difference was not reliable, $p = .085$. Searchers spoke more than assistants (79.0 vs. 21.1 words) and the interaction of role and cursor transfer showed that the previously mentioned larger word numbers for gaze than mouse were only present for searchers, $p < .001$, but not for assistants, $p = .319$. Also, searchers spoke more with the objects invisible than visible, $p < .001$, whereas assistants' speech did not differ between the object visibility conditions, $p = .850$. The triple interaction can be ascribed to the particularly high word numbers for searchers in the gaze/invisible condition (see Fig. 4A).

It was hypothesized that when using gaze, searchers might provide more explicit verbal feedback for the assistants to facilitate their decisions about how and when to move the window. To test this, we calculated the percentage of searchers' feedback utterances (i.e. any utterance providing an explicit confirmation or disconfirmation, e.g. "yes", "okay" or "no") relative to the total amount of their utterances. Percentages were used to differentiate the feedback effects from any overall differences in the amount of speech. There was a main effect of cursor transfer, $F(1,23) = 8.67$, $p = .007$, while object visibility only produced a non-significant trend, $F(1,23) = 4.11$, $p = .055$, and the interaction was not reliable, $F(1,23) = .01$, $p = .917$. As expected, feedback formed a higher proportion of the searchers' utterances during gaze than mouse (21.0 vs. 15.2%). Also, there was a tendency towards more feedback when the objects were invisible (20.2 vs. 16.0%). However, object invisibility did not differentially increase feedback in gaze versus mouse (see Fig. 4B).

In the final step of our speech analyses, we asked how much speech was actually necessary to instruct the assistant to move the window. Therefore, we counted all the searchers' utterances that addressed the positioning of the window (e.g. "Move it here" or "A little more to the right"), and again used their proportion relative to the total amount of a searcher's utterances. Movement-related utterances differed between the two types of cursor transfer, $F(1,23) = 5.78$, $p = .025$, and between the object visibility conditions, $F(1,23) = 8.72$, $p = .007$. These results were qualified by a significant interaction between the two factors, $F(1,23) = 5.10$, $p = .023$. Searchers produced more movement utterances in gaze than in mouse and more with the objects invisible than visible, but it turned out that the difference between gaze and mouse only was present with invisible objects (41.9 vs. 25.6%), $p < .001$, but not in the visible condition (24.0 vs. 22.6%), $p = .803$ (see Fig. 4C).

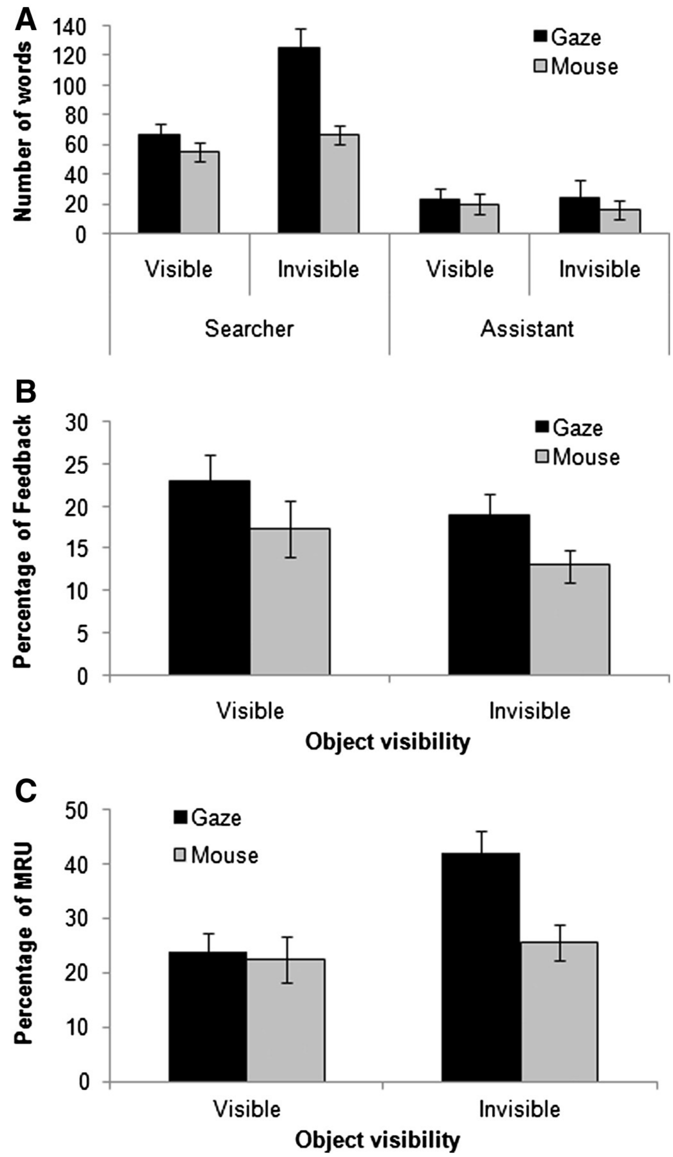

Fig. 4. Number of words for both partners (A), percentage of feedback utterances (B) and percentage of movement-related utterances (MRU, C) depending on cursor transfer and object visibility. Error bars represent standard errors of the mean.

4. Discussion

The present study compared gaze and mouse transfer during remote cooperation in a situation in which knowing about the partner's attention and strategies was potentially useful. It was investigated whether in this context there is a *specific* benefit of gaze transfer that goes beyond mere pointing, and how both modes of interacting affect the coordination joint action. In a joint moving window (JMW) task, an assistant could use a searcher's transmitted gaze or mouse cursor to provide optimal viewing conditions. The assistant himself was endowed with visual information about the object locations, or no visual object information at all.

4.1. Using gaze transfer remains challenging

The results extend our previous findings that mouse transfer outperforms gaze transfer (Müller et al., 2013). In the present study, there was no specific benefit of gaze transfer, indicating that a possible criticism of our previous study (Müller et al., 2013), namely that a spatial referencing task simply is not suited for a fair investigation of the potentials of gaze, is insufficient. Instead, it appears that coordinating joint action with gaze transfer is difficult even when the partner's search and visual attention are directly task-relevant. In more general terms, these findings suggest that knowing more about a partner's task solution processes is not always helpful, at least not without any form of "interpretation support".

4.2. Object visibility has differential effects on gaze and mouse transfer

Despite the lack of a specific gaze benefit, it needs to be noted that gaze also was not inferior to mouse transfer, given that the assistant received it in relation to the objects it was directed at. With visible objects, gaze transfer neither increased solution times nor participants' verbal effort in terms of their overall words or the searcher's utterances related to moving the window. In contrast, when information about the task environment was compromised in the invisible condition, subjects were still able to apply the mouse cursor successfully, but performance severely deteriorated for gaze transfer.

An explanation can be derived from the analysis of the spatio-temporal coupling between the transmitted cursor and the window centre, as reflected in the minimal mean distances between them. This coupling provides insights into the way in which assistants were actually using their partner's gaze or mouse cursor to guide their own actions. An overall stronger coupling for mouse transfer was qualified by an interaction with object visibility, indicating that differences between gaze and mouse only occurred in the invisible condition. Thus, given that the visual task context is provided, gaze and mouse cursors can be used to coordinate joint performance in a similar manner. Conversely, eliminating object visibility resulted in a closer coupling between the mouse cursor and the window centre, whereas this was not the case for the gaze cursor.

Moreover, although there were overall differences in coordination *between* the experimental conditions, we also observed different strategies *within* the conditions. In the gaze/visible condition, there was a tendency for two different ways of coordinating the cursor-window movement: Some assistants followed the searcher, whereas others adopted an autonomous strategy, as reflected in negative delay times of the minimal distance between cursor and window. In the mouse/visible condition, there was a much wider range of delay times. This might indicate that as long as autonomous action is possible (because the assistant possesses the relevant knowledge), mouse transfer allows a flexible choice of a coordination strategy, whereas gaze transfer invites a more homogenous behaviour. A possible reason could be that the fast gaze cursor is hard to ignore and has a high affordance of being used (see below). If this is the case, it does not necessarily pose a problem as we found no effects of strategy on performance. However, it should be considered when applying gaze transfer in joint tasks that require a high degree of autonomy.

When the objects were invisible, negative cursor-window delays (i.e. autonomous action of the assistant) did not occur at all. Moreover, with invisible objects the delay times were *more* homogenous and the cursor-window distances were smaller for mouse than gaze transfer, indicating a closer following of the cursor. Taken together, the strategy analyses suggest that if visual information is compromised, most receivers of mouse transfer abandon their flexibility and just follow their partner's cursor in an immediate way. No comparable behaviour can be observed when the partner's cursor represents his gaze.

To account for this inconsistency in the assistants' behaviour, we argue that the ability to *interpret* gaze and mouse cursors is differentially affected by the viewing conditions. An explanation might be derived from the nature of the information transmitted via the two types of cursor transfer. Eye movements are a rather direct manifestation of a person's visual attention and information processing activities (Just & Carpenter, 1976; Velichkovsky, 2002). However, this processing occurs in the context of the visual environment: Attention does not float freely in space but is closely linked to the objects a person is interacting with (Land & Tatler, 2009). Thus, gaze information becomes interpretable in the first place because it enables an observer to understand what someone is attending *to*. Consequently, when seeing someone look without knowing what he is looking at, it obviously is hard, if not impossible, to make sense of these gazes. This interpretation is corroborated by some of our subjects' informal reports, expressing that they had tried to ignore the gaze cursor in the invisible condition, because they had not been able to use it. The same conclusion can be drawn from the excessively high word numbers in this condition, reflecting the participants' attempts to coordinate the window movement in a purely verbal manner.

These results fit nicely with other studies showing that the way in which people establish common ground is affected by what they know or believe their partner can see (Hanna, Tanenhaus, & Trueswell, 2003). For example, Richardson, Dale, and Tomlinson (2009) manipulated participants' beliefs regarding the match between their own and their partner's visual input during a discussion. Participants incorporated these beliefs into the way they coordinated their interaction, by compensating an assumed mismatch with more conversational effort as reflected in both their eye movement patterns and speech. However, when people do not possess the relevant knowledge or beliefs about the partner's visual processing (as in our invisible condition), they cannot use this information and are forced to adopt a default strategy of establishing common ground from scratch. This can be done either by using deictic information from the partner's mouse, or high amounts of verbal effort if such tools are unavailable.

4.3. Flexibility is a central difference between gaze and mouse transfer

When using the mouse under impoverished viewing conditions, subjects did not seem to encounter the same difficulties as for gaze. An obvious reason is that mouse cursors can be applied as an explicit pointing device, used for the sole purpose of intentional communication: The assistant knows that whenever the mouse moves, he can simply react, and indeed some of our searchers asked their assistant to "not think, just follow my mouse". Thus, the assistant does not even have to understand why his partner's mouse is moving, he only has to use it as a deictic signal. There is no reason why this ability should be compromised in the absence of visual object information.

The single function of the mouse as a communication device in the present paradigm raises the concern that the ease of reacting to it may not generalize to other situations. In most computer-based tasks, the mouse serves as a tool for object manipulation (e.g. clicking or relocating). One might argue that such a double function of mouse transfer would have increased distraction and uncertainty just like it does for gaze transfer. In the present paradigm, it might have led assistants to erroneously move the window to an area in which the searcher was only intending to manipulate an object, without wanting the assistant to react. However, such a double function of mouse actions should only cause distraction if the objects that a person is manipulating differed from those that he needs to see. This is not supported by the general finding that in active tasks people look at the objects they are acting on (Land & Tatler, 2009). Instead, endowing the mouse with a task-relevant function might even have increased its communicative value, because observing a partner's actions provides an implicit cue for coordinating joint activities (Sebanz, Bekkering, & Knoblich, 2006).

From that perspective, it might even be asked whether the reverse is true and the single function of mouse movements decreased its effectiveness: Should coordinating joint action on the basis of purely intentional signals not slow down the overall interaction? It seems that some of our subjects thought so as well, as five searchers chose to not use the mouse at all, perhaps expecting it to be too cumbersome. As there was no effect of actual mouse usage on performance, this seems to say more about the diversity of people's strategies than about the necessity of mouse transfer. But it illustrates a characteristic of the mouse that presumably makes it so effective: its flexibility. Basically, the mouse can be used in all kinds of ways, and it can even mimic gaze transfer: An analysis of the coupling between searchers' gaze and mouse cursors suggests that they can link their mouse to their own gaze, making their focus of visual attention immediately accessible for an observer even in the absence of actual gaze transfer. In addition, mouse movements can be adjusted to the partner's needs and scaled to the demands of the situation, either performing fast moves of small amplitudes or being reduced to slow and systematic moves. Also, any temporary glitches that would be confusing for the partner (e.g. reflexive movements to distant distractors) can be avoided.

Changing one's gaze to achieve a comparable "audience design" (to borrow a term from linguistics, Bell, 1984) would presumably be very costly. This is because in active tasks, people cannot prevent or control their eye movements without severely compromising their performance (Ballard, Hayhoe, & Pelz, 1995).[3] Thus, when using gaze transfer, people have to find other ways to prevent misunderstandings. There is evidence that in the JMW task, they did so in an explicit, verbal manner, by producing more feedback utterances. Interestingly, this difference between gaze and mouse transfer occurred irrespective of object visibility, suggesting that the increased risk of misinterpreting eye movements is not limited to specific viewing conditions but constitutes a general problem inherent in gaze transfer.

The ways in which subjects used gaze versus mouse transfer can be summarized like this: In a joint task, eye movements inevitably play *different* roles (e.g. information uptake and communication), whereas the mouse can play *any* role a person wants it to play. Thus, the latter can be adjusted to the task at hand, whereas eye movements are more determined by factors outside the user's intention. This makes the mouse a highly flexible communication device, whereas the conditions for successfully using gaze transfer are more limited.

4.4. Difficulties in using gaze transfer, and possible solutions

Many difficulties of gaze transfer originate from the characteristic fast and unsystematic way the eyes move. Such movements can capture an observer's attention, even when they are not relevant to the task (Mulckhuyse, van Zoest, & Theeuwes, 2008; Richardson, 2013). But these characteristics of eye movements also have consequences for an observer's reactions to the cursor, because just following it unselectively would result in inconsistent and jolty movements. When people have to coordinate their actions, this will not only affect the observer of gaze transfer but backfire on the producer as well. For instance, in a situation like the JMW task, each time the assistant reacts to the searcher's gaze in a premature or inappropriate way, the latter one's visual field gets changed.

Therefore, we want to broaden the argument from our previous article (Müller et al., 2013) in which we concluded that gaze transfer is hard to use when it needs to be inferred whether the partner is trying to give instructions or just scanning the display. In the light of the present findings, we suggest that using gaze transfer can produce distortions *whenever* it is possible for an observer to react to it in an immediate way and thereby affect the person whose gaze is transferred. Thus, gaze transfer should probably not be used for moment-by-moment decisions about *whether* and *how* to act. However, we made an interesting observation in the gaze/visible condition: Some pairs chose a strategy in which the assistant decided over the direction of the window movement, and only used the gaze cursor to determine *when* to move, namely when he saw that the searcher had just finished looking at all the objects currently visible. Typically, pairs who adopted this strategy were highly efficient. This also corresponds with the notion that gaze transfer may be helpful in choosing an appropriate timing for topic changes in a simulated tourist guiding task (Qvarfordt et al., 2005).

Certainly, there are conditions under which gaze transfer is more feasible than mouse transfer (e.g. when the hands are used otherwise). Therefore, future research will have to ameliorate its shortcomings, and one fruitful approach is to develop more usable visualizations of a person's gaze. First, in order to render the transmitted cursor less confusing and achieve some degree of permanence, its movement could be presented in the context of the immediately preceding eye movements, for instance as a decaying trace (Mehta et al., 2005). Similarly, smoothing the cursor movement could reduce its speed and positional variability (Helmert, Pannasch & Velichkovsky, 2008). Such a manipulation would also make the visual appearance of eye and mouse movements more similar, reducing the impact of different movement parameters as a potential confound in comparing both types of transfer. A second problem of current visualizations lies in the undifferentiated depiction of the transferred gaze: A highly salient cursor is constantly jumping about. Instead, the cursor's visual appearance could be linked to the function of the respective eye movements. This should be possible to the extent to which this function can be inferred on the basis of eye movement parameters per se (Velichkovsky, 2002), their co-occurrence with other outputs such as speech (Kaur et al., 2003), or their appearance in the context of specific activities (e.g. refixations when reading). A third approach starts from the premise that gaze transfer visualizes a person's attention to objects, which implies that a direct highlighting of these attended objects might be preferable to an indirect spatial indicator such as a cursor. Presumably, such highlighting would facilitate the matching between the gaze and its target in the visual scene, and current work is comparing these direct object visualizations with other forms of gaze transfer.

5. Conclusion

The present study emphasizes three main considerations about gaze transfer. First, using a partner's gaze to coordinate joint action is challenging, not only when it functions as an explicit command but even when the partner's search process is directly task-relevant. Second, whereas mouse cursors can be used in a multitude of ways and even be followed unselectively, when using gaze cursors it is of paramount importance that the recipient perceives them in relation to the environment they are corresponding to. Thus, caution is required to ensure that for an observer of gaze transfer it is directly visible and understandable what his partner is looking at and why. Third, in highly interactive tasks there probably should be a barrier between gaze transfer and the ability to react to it. That is, reactions to gaze should not need to be immediate but only be required after a certain amount of time and observation of the partner's viewing behaviour. If gaze transfer is to be used in settings in which the meaning of gaze has to be inferred from particular cursor movements, technical measures will have to be taken to minimize uncertainty and distraction. In conclusion, our results again highlight the need to take task characteristics into account when applying gaze transfer for remote cooperation.

Acknowledgements

We thank Gernot Pascher for programming the experiment, as well as Bettina Goetze and Cynthia Poenicke for their assistance in data collection. Also, we are grateful to Ben Tatler and two anonymous reviewers for very helpful comments on an earlier version of the manuscript. This research was supported by a grant of the TU Dresden Centre for Continuing Education to RM.

References

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand–eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London, Series B*, *337*, 331–339. http://dx.doi.org/10.1098/rstb.1992.0111.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. http://dx.doi.org/10.1162/jocn.1995.7.1.66.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. http://dx.doi.org/10.1017/S004740450001037X.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477. http://dx.doi.org/10.1016/j.cognition.2007.05.012.

Bruner, J. (1981). The pragmatics of acquisition. In W. Deutsch (Ed.), *The child's construction of language* (pp. 42–64). New York, NY: Academic Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework of memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. http://dx.doi.org/10.1016/S0022-5371(72)80001-X.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. http://dx.doi.org/10.1111/1467-9280.00255.

[3] In the JMW task, subjects refrained from controlling their eye movements. When comparing fixation durations and saccadic amplitudes between gaze and mouse transfer, we did not find any differences. This contrasts with previous studies using gaze transfer in a more intentionally communicative way (Müller et al., 2013; Velichkovsky, 1995) and suggests that when a task leaves people a chance to avoid controlling their gaze, this is what they do.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615. http://dx.doi.org/10.1016/j.jml.2007.01.008.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61. http://dx.doi.org/10.1016/S0749-596X(03)00022-6.

Helmert, J. R., Pannasch, S., & Velichkovsky, B. (2008). Influences of dwell time and cursor control on the performance in gaze driven typing. *Journal of Eye Movement Research*, *2*(4) 3, 1–8.

Jacob, R. J. K. (1991). The use of eye movements in human–computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems (TOIS)*, *9*(2), 152–169. http://dx.doi.org/10.1145/123078.128728.

Just, M.A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. http://dx.doi.org/10.1016/0010-0285(76)90015-3.

Kaur, M., Tremaine, M., Huang, Wilder, J., Gacovski, Z., Flippo, F., et al. (2003). Where is “it”? Event synchronization in gaze–speech input systems. *Paper presented at the 5th international conference on multimodal interfaces, New York, NY*.

Kawahara, J., Yanase, K., & Kitazaki, M. (2012). Attentional capture by the onset and offset of motion signals outside the spatial focus of attention. *Journal of Vision*, *12*(12:10), 1–13. http://dx.doi.org/10.1167/12.12.10.

Land, M. F., & Tatler, B. W. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. Oxford, UK: Oxford University Press.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(3), 578–586. http://dx.doi.org/10.3758/BF03203972.

Mehta, P., Sadasivan, S., Greenstein, J., Gramopadhye, A. K., & Duchowski, A. T. (2005). Evaluating different display techniques for communicating search strategy training in a collaborative virtual aircraft inspection environment. *Paper presented at the Human Factors and Ergonomics Society annual meeting, Orlando, FL*.

Mulckhuyse, M., van Zoest, W., & Theeuwes, J. (2008). Capture of the eyes by relevant and irrelevant onsets. *Experimental Brain Research*, *186*(2), 225–235. http://dx.doi.org/10.1007/s00221-007-1226-3.

Müller, R., Helmert, J. R., Pannasch, S., & Velichkovsky, B.M. (2013). Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *Quarterly Journal of Experimental Psychology*, *66*(7), 1302–1316. http://dx.doi.org/10.1080/17470218.2012.737813.

Neider, M. B., Chen, X., Dickinson, A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, *17*(5), 718–724. http://dx.doi.org/10.3758/PBR.17.5.718.

Qvarfordt, P., Beymer, D., & Zhai, S. (2005). *RealTourist — A study of augmenting human–human and human–computer dialogue with eye–gaze overlay*. Paper presented at the Human–Computer Interaction, INTERACT '05, Rome, Italy.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.372.

Richardson, D. C. (2013). The reciprocity of attention and joint action. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination and beliefs about visual context. *Cognitive Science*, *33*(8), 1468–1482. http://dx.doi.org/10.1111/j.1551-6709.2009.01057.x.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. http://dx.doi.org/10.1016/j.tics.2005.12.009.

Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. *Paper presented at the 6th international conference on multimodal interfaces, State College, PA*.

Velichkovsky, B.M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition*, *3*(2), 199–222. http://dx.doi.org/10.1075/pc.3.2.02vel.

Velichkovsky, B.M. (2002). Heterarchy of cognition: The depths and the highs of a framework for memory research. *Memory*, *10*(5/6), 405–419. http://dx.doi.org/10.1080/09658210244000234.

Velichkovsky, B.M., & Hansen, J. P. (1996). New technological windows into mind: There is more in eyes and brains for human–computer interaction. In R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, & G. Olson (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 496–503). Montreal, Canada: ACM Press.

Vertegaal, R. (2008). A Fitts' Law comparison of eye tracking and manual input in the selection of visual targets. *Paper presented at the 2008 international conference on multimodal interfaces, Chania, Krete, Greece*.

Whittaker, S. (1995). Rethinking video as a technology for interpersonal communications: Theory and design implications. *International Journal of Human–Computer Studies*, *42*(5), 501–529. http://dx.doi.org/10.1006/ijhc.1995.1022.

Whittaker, S. (2003). Things to talk about when talking about things. *Human–Computer Interaction*, *18*(1&2), 149–170. http://dx.doi.org/10.1207/S15327051HCI1812_6.