Following closely? The effects of viewing conditions on gaze versus mouse transfer in remote cooperation

Romy Mueller, Jens R. Helmert, Sebastian Pannasch, & Boris M. Velichkovsky
Applied Cognitive Research Unit, Technische Universitaet Dresden, Germany
romy.mueller@psychologie.tu-dresden.de

Abstract. Gaze transfer can support reference resolution and improve performance in remote cooperation, as compared to verbal communication alone. However, recent research lacks the critical comparison to mouse pointing transfer. In the present experiment, both gaze and mouse transfer were used in a remote cooperative situation. One partner (*assistant*) had to move a window to continuously open up parts of the display that were relevant to the other partner (*searcher*) for completing the task. Either the searcher's mouse or gaze position was transmitted as a means of guidance for the assistant's actions. Additionally, the assistant could either see the positions of the objects relevant for the task (*objects*) or just the searcher's cursor on a grey screen (*grid*). In the objects condition, performance was similar for mouse and gaze transfer, whereas in the grid condition we observed shorter solution times for mouse as compared to gaze transfer. Cross recurrence analyses quantifying the window and cursor alignment suggest that in the latter condition, assistants were more confidently following the searcher's mouse compared to his gaze cursor. The results will be discussed with regard to their implications for the application of gaze transfer in remote cooperation.

Gaze versus mouse transfer

When people have to cooperate remotely, gaze transfer in the form of a cursor superimposed on the visual material has been found to improve performance in spatial tasks (Velichkovsky, 1995). This effect is presumably due to the supporting function of gaze in ambiguity resolution: a direct grounding of verbal references in the shared workspace becomes possible, reducing the need for effortful verbal explanations.

Although beneficial, when compared to purely verbal interactions, gaze is not the only way to disambiguate object reference in remote cooperation. By transferring computer mouse positions, interlocutors are provided with a very easy and direct method to indicate object locations in such settings. In previous research investigating similarities and differences between both forms of cursor transfer, no differences in task performance were found between them (Velichkovsky, 1995, Mueller, Helmert, Pannasch & Velichkovsky, submitted).

Despite the lack of performance differences, gaze and mouse transfer seem to have differential effects on the cooperation process. Mueller et al. (submitted) showed that transferring one partner's gaze to the other resulted in difficulties to interpret the communicative intention (or lack thereof) on a moment-by-moment basis when solving a cooperative puzzle task. However, at the same time the partner who saw the gaze cursor was strongly drawn to it. Consequently, gaze transfer required more effortful verbal disambiguation. The authors argued that solving puzzles strongly relied on intentional pointing; a task characteristic that makes the mouse most suitable. Thus, when comparing gaze and mouse transfer, the task characteristics are important and must be taken into account.

In the present study, we investigated the differences between both forms of cursor transfer in a task where information about a person's attention and search process is crucial. In this way, viewing a gaze cursor could exhibit potential benefits over purely intentional mouse pointing. Within this framework, 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 it is referring to.

We asked pairs of participants to solve a joint path-selection task with a strong spatial component on different processing levels (colour differentiation, form identification, calculation). One partner (*searcher*) was provided with the full visual information necessary to solve the task, but saw only a section of the display, while the other partner (*assistant*) had to move the searchers viewing window to reveal the respective display areas that were 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) or had no task-relevant visual information.

We chose this simplistic task to create a situation with maximal clarity of the component actions required of both partners to perform the task. In this way, first, the assistant's way of helping the searcher is straightforward. Second, at any given moment the assistant knows what component actions (in terms of the three steps of task solution, see procedure) the searcher is engaged in. In this way, the searcher's gaze or mouse parameters in relation to the objects (if visible) should be easily interpretable as an indication of what he is doing. The difficulty of inferring his state or his needs is decreased to a minimum, and information about his focus of attention should have maximal potential to be useful.

Method

Subjects

Forty-eight subjects aged 18-51 years (M: 23.9, SD: 6.1, 32 females) participated in the experiment. They were invited in pairs and assigned to one of the two experimental roles (searcher or assistant), resulting in a total of 24 pairs.

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 using the EYELINK CL infrared eye tracking system (SR Research Ltd., Ontario, Canada) in the remote recording mode.

Stimuli

20 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 1B). For the searcher, a variable number of circles and triangles containing positive or negative digits were located on each path. The searcher had access to the stimulus information only within a viewing window of 255 x 190 pixels (1/16 of the screen), the rest was covered black (see Figure 1A). The assistant saw the whole screen area but either all objects were depicted as circles (*objects*, see Figure 1B), or only a grey background consisting of equidistant vertical and horizontal lines was visible with no further information (*grid*, see Figure 1C). The searcher's gaze or mouse cursor was projected onto the assistant's screen as a tricolor 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 to prevent confusion for the assistant between his own and the searcher's mouse cursor in the mouse condition.

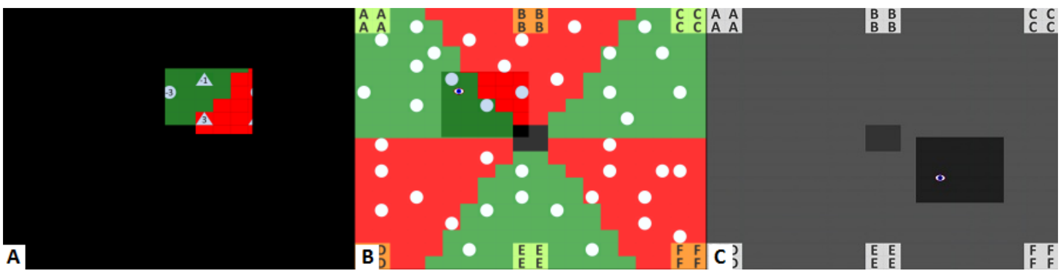

Figure 1. Stimuli for searcher (A) and assistant in both forms of assistant view, objects (B) and grid (C).

Procedure

The experiment consisted of four blocks corresponding to combinations of the experimental conditions. 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 were instructed to first select the three red paths. In the second step, they had to consider all three of them and exclude the one with the least number of circles. In the final step, participants had to determine which one of the two remaining red paths contained the smaller sum of digits. This remaining path had to be selected via mouse click of the searcher on the respective target field (marked with letters, see figure 1 B and C).

The assistant's task was to move a window locked to his own mouse position to reveal parts of the display, so that the searcher could see the stimuli needed for solving the task. Participants were free to verbally interact in all experimental conditions. In *gaze* the searcher's gaze cursor was transferred to the assistant and in *mouse* his mouse cursor was transferred. In *objects* the assistant saw the path colours and object locations but not their form (each objects was depicted as a circle) and in *grid* he saw the grey background with no information about colours or objects. The form of the paths was identical throughout the whole experiment and only their order changed between trials. A condition with partial spatial information showing only colours or objects, respectively, was not expected to produce results that greatly differ from the grid or objects condition, respectively. Thus, no such condition was included.

We recorded the eye movements of the searcher, the mouse actions of both participants, and their verbal interactions.

Results

Performance

To compare task performance between the experimental conditions, we analyzed mean solution times and error rates. For mean solutions times a 2 (*Cursor: gaze, mouse*) x 2 (*Assistant view: objects, grid*) repeated measures ANOVA revealed main effects of cursor, $F(1,23) = 22.60$, $p < .001$, and assistant view, $F(1,23) = 14.69$, $p < .001$, and 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 in objects than in grid (75.9 vs. 103.4 s). The difference between mouse and gaze was only present in grid, $p < .001$, but not in objects, $p = .153$ (see figure 2). Mean error rates were at about 19.4 % and showed no differences between the experimental conditions, all Fs < 3, all ps > .1.

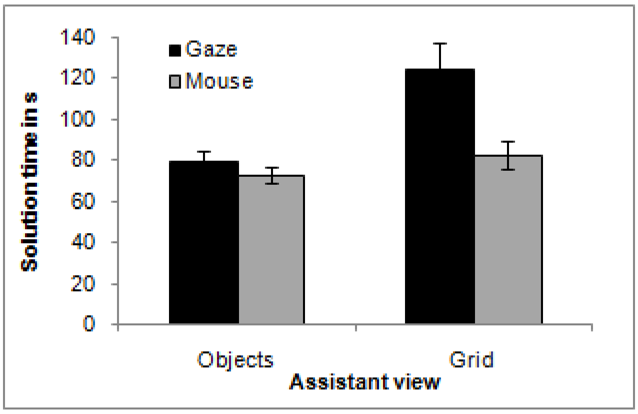

Figure 2. Solution times depending on cursor and assistant view.

Window and cursor alignment

To determine the degree of positional coupling of the window centre (corresponding to the assistant's cursor) and the searcher's cursor, a cross recurrence analysis (Marwan & Kurths, 2002) 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.

Peak recurrence rates were subjected to a 2 (*Cursor: gaze, mouse*) x 2 (*Assistant view: objects, grid*) repeated measures ANOVA, revealing main effects of cursor, $F(1,23) = 5.16$, $p = .033$ and assistant view, $F(1,23) = 18.59$, $p < .001$, as well as an interaction between both factors, $F(1,23) = 19.78$, $p < .001$. Stronger coupling was found in mouse than in gaze (32.3 vs. 27.2 %), and there was more coupling in grid than in objects (33.9 vs. 25.6 %). However, recurrence rates were increased in grid as compared to objects only for mouse, $p < .001$, whereas these two assistant view conditions did not differ in gaze, $p = .705$ (see figure 3).

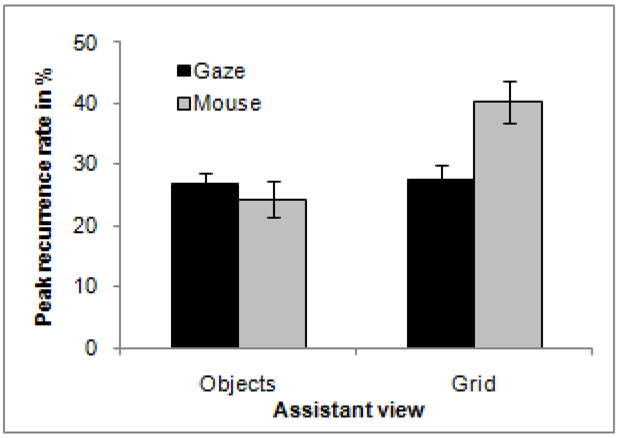

Figure 3. Peak recurrence rates between cursor and window centre, depending on cursor and assistant view.

In addition to peak recurrence, the temporal dynamics of cursor window alignment were investigated. If you consider recurrence rates as a curve that starts ascending at the maximum negative temporal delay (-1000 ms in this study), peaks at a certain delay and then descends until maximum positive delay (1500 ms), the amplitude of the curve is 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, $F(1,23) = 11.95$, $p < .001$, an effect of assistant view, $F(1,23) = 26.48$, $p < .001$, and an interaction between both factors, $F(1,23) = 59.00$, $p < .001$. When using the mouse, higher increases were found 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 %), $p = .444$.

Discussion

Gaze and mouse transfer were compared in remote cooperation. Both forms of transfer were employed to guide an assistant in moving a viewing window to reveal relevant display areas for a searcher. We investigated whether the applicability of gaze versus mouse transfer depended on the assistant's amount of visual information about the task environment.

Solution times were slower for gaze than mouse transfer only when the assistant saw no objects. This result suggests that gaze transfer is not generally ineffective in this type of task. Instead, seeing a partner's gaze position can be as helpful as seeing his mouse when the person to whom the cursor is being transferred is provided with adequate information about the visual task environment. When this visual information was not given, subjects were still able to apply the mouse cursor, whereas performance deteriorated for gaze. How can this effect be accounted for?

We suggest that the interpretability of the gaze and mouse cursor differs, depending on the assistant's viewing conditions. To understand this argument, we first have to consider the type of information transmitted by gaze versus mouse transfer. Eye movements provide a rather direct visualization of visual attention in relation to task-relevant objects, especially in active tasks (Land & Tatler, 2009). Their temporal and spatial parameters are closely related to processing information about these entities (Velichkovsky, 2002). Visual attention usually does not float freely in space. It always implies a relation between a person and the entities that are being attended to. Thus, when an observer does not know what the eye movements transmitted to him are directed at, it presumably is hard him to make sense of them.

How about the mouse in an otherwise identical situation, where there also is no further visual information available for the assistant? When applying the mouse as an intentionally used device for communication, people can decide to use it solely to give messages to the partner, so that he knows that whatever the mouse does, he can simply react. In fact, several searchers instructed the assistant to “not think, just follow my cursor”. In this way, the assistant may not understand why his partner is making a certain mouse movement, but can be sure that the partner is deliberately producing this action as a deictic sign. In this case, it should be a suitable strategy to simply follow the cursor.

Our cross recurrence data support this hypothesis. The analysis revealed that in mouse, the coupling between the searcher’s transferred cursor and the window centre increased when the assistant saw no objects and thus had to rely on the searcher’s guidance. This increased coupling could not be found in gaze. Similarly, the change of recurrence rate over different temporal delays was more than 2.5 times higher for grid than objects in the mouse condition, whereas it did not differ between viewing conditions in gaze. Thus, in the light of impoverished viewing conditions, the assistants closely followed their partner’s mouse movements, but not their gaze. When using the mean distance between searcher and assistant cursor at different time lags instead of recurrence, comparable results were obtained. Additionally, verbal interaction data not reported in this paper also suggest that in grid subjects tried to avoid relying on the gaze cursor: in this condition not only most words were produced, but also the relative number of utterances regarding the positioning of the window were increased.

What is the benefit of mouse movements, and why can they be used more reliably than gaze, at least in the grid condition? Whereas gaze is too fast and unpredictable to automatically follow every positional change, mouse movements can be reduced to comparably slow and systematic moves. Thus, people can follow the mouse even when they don’t understand the component processes their partner is engaged in. But there is more to instructing with the mouse: not only is it slower than the gaze cursor, but more adjustable as well.

In active tasks, people cannot prevent or control their eye movements without this having detrimental effects on performance (Ballard, Hayhoe & Pelz, 1995). In contrast, searchers using the mouse can decide to scale their moves to the needs of the situation, either indicating only a general direction, or producing a large amount of small amplitude moves.

People can even indicate their own eye movements with their mouse, thereby partly simulating the spatio-temporal characteristics of gaze transfer. Indeed, searchers in our study did just that when it was necessary to communicate their search process more precisely in grid: when comparing the searchers’ own gaze and mouse positions in the mouse conditions, higher peak recurrence rates can be found in grid than in objects, $t(23) = -2.13$, $p = .044$. That is, searchers can use their mouse in a way that it resembles their eyes and in this way indicate their current focus of attention. This adjustability of their mouse movements equips them with a tool to engage in a highly precise communication of attentional foci with their partner, without the distracting features implied in actual gaze transfer.

As gaze transfer cannot be controlled as easily by the user himself, future research will focus on finding ways of adjusting it, so that it will provide a more optimal and situationally appropriate support. More refined visualization techniques appear to be a fruitful approach to address this issue. For example, smoothing the transferred gaze positions by averaging over several samples has been shown to increase subjective cursor control in a human computer interaction setting (Helmert, Pannasch & Velichkovsky, 2008) and might improve its usability in cooperative situations as well.

Taken together, our results highlight the importance of taking task characteristics into account when applying gaze transfer. Although mouse cursors can be followed blindly, when using gaze cursors it is of paramount importance that the recipient can perceive them in relation to the environment they are corresponding to. If gaze transfer shall be applied in situations that call for more confident reactions by the perceiver, technical measures will have to be taken to reduce uncertainty and distraction.

Acknowledgments

We thank Gernot Pascher for support in the experimental setup, and Bettina Goetze and Cynthia Poenicke for their assistance in data collection. This research was supported by a grant of the TU Dresden Centre for Continuing Education to RM.

References

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66-80.

Helmert, J. R., Pannasch, S., & Velichkovsky, B. M. (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.

Land, M. F., & Tatler, B. W. (2009). *Looking and Acting: Vision and Eye Movements during Natural Behaviour*: Oxford University Press.

Marwan, N., & Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A, 302*(5-6), 299–307.

Mueller, R., Helmert, J.R., Pannasch, S., & Velichkovsky, B.V. (submitted). Improving remote cooperation in spatial tasks: Benefits and pitfalls of the gaze transfer approach.

Velichkovsky, B. M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition, 3*(2), 199-222.

Velichkovsky, B. M. (2002). Heterarchy of cognition: The depths and the highs of a framework for memory research. *Memory, 10*(5/6), 405-419.