Pointing and Visual Feedback for Spatial Interaction in Large-Screen Display Environments

Barry A. Po, Brian D. Fisher, and Kellogg S. Booth

Department of Computer Science, University of British Columbia
201-2366 Main Mall
Vancouver, British Columbia, Canada
{po, fisher, ksbooth}@cs.ubc.ca

Abstract. The two visual systems hypothesis in neuroscience suggests that pointing without visual feedback may be less affected by spatial visual illusions than cognitive interactions such as judged target location. Our study examined predictions of this theory for target localization on a large-screen display. We contrasted pointing interactions under varying levels of visual feedback with location judgments of targets that were surrounded by an offset frame. As predicted by the theory, the frame led to systematic errors in verbal report of target location but not in pointing without visual feedback for some participants. We also found that pointing with visual feedback produced a similar level of error as location judgments, while temporally lagged visual feedback appeared to reduce these errors somewhat. This suggests that pointing without visual feedback may be a useful interaction technique in situations described by the two visual systems literature, especially with large-screen displays and immersive environments.

1 Introduction and Background

Two-dimensional pointing plays a central role in our day-to-day interaction with desktop user interfaces. Three-dimensional pointing enables spatial interaction with objects in non-desktop interfaces such as large-scale collaborative workspaces. Designers rely on pointing as an interaction technique because it is commonly believed that the usability of complex applications is enhanced by employing interactions that mirror familiar gestures.

Common intuition tells us that visual feedback helps make pointing an effective method of interaction. Current design practices reflect a belief that pointing becomes unreliable when it is uncoupled from visual feedback in display environments, especially when there are multiple potential targets. However, studies from psychology tell us that our intuition is not always correct, and in particular that visual feedback can degrade pointing performance if certain circumstances exist.

Our research has examined the influence of varying levels of visual feedback on target selection in a large-screen graphical display environment. The first

step in our research was a pilot study that successfully reproduced the previous laboratory findings of Bridgeman et al. [2], but in a setting more representative of collaborative and immersive display environments. This demonstrated that a displaced frame surrounding a target will cause some participants to make systematic errors in verbally reported location judgments but not in pointing without visual feedback (also known as "open-loop" pointing in experimental psychology). The classic work on this is known as the *Induced Roelofs Effect* [2, 18], part of the theoretical framework for our research.

After our pilot study, we designed a larger study with four conditions, the first two the same as in Bridgeman's work, in order to examine the impact of multiple visual representations on two additional interaction conditions: pointing with a fast-response visible cursor and pointing with a slow-response (temporally lagged) visible cursor. These conditions were identified as pointing with visual feedback and pointing with lagged visual feedback respectively.

We tested for the presence of the Induced Roelofs Effect and found differences across the four conditions consistent with the theoretical predictions. In the remainder of this section, we provide background material on the problems that we are addressing, related work, and the theoretical framework for our research. We then describe our study and results, followed by a closing discussion and future work, with implications for the design and implementation of spatial interaction techniques in large-screen display environments.

1.1   Feedback in Spatial Interaction

Contrary to our intuition, there is considerable evidence to suggest that visual feedback may not positively affect user performance in all display environments and interactive situations. A study of object selection and positioning by Poupyrev et al. [17] indicates that visual feedback does not always improve user performance, particularly in situations where users are in close proximity to the display. Hayne et al. [6] report that the use of visual feedback has potential drawbacks in shared display systems, where visual feedback may take the form of multiple cursors that compete for valuable display space and obscure informative parts of the display. Wolf and Rhyne [22] further observe that maintaining real-time visual feedback significantly increases the workload of network and processor resources.

Even if we believe that these particular disadvantages are insignificant compared to the apparent advantage of visual feedback, there are other potential concerns that should give us pause when we consider the use of visual feedback in large-screen display environments. In instances where a large group of users are concurrently interacting with a large-scale display, it is possible that the presence of a multiplicity of cursors may make it difficult for users to keep track of a particular user's cursor. Additionally, the presence of one or more visible cursors may draw attention away from other, more important display items, especially when the cursors move or change state.

Thus, it seems that the presence of visual feedback of pointer position has potential drawbacks, and that removing that feedback might alleviate those con-

cerns. Unfortunately, it is not known how the removal of visual feedback would affect overall user performance in large-screen display environments. While most users achieve similar and adequate levels of performance when pointing with visual feedback, it is reasonable to expect higher error and greater individual differences in the accuracy and precision of pointing without visual feedback. Thus, the current practice of providing visual feedback at least appears to help equalize performance across a larger user population with a wide range of pointing abilities.

1.2   Voice as a Substitute for Pointing

Given the difficulties that surround pointing and other gestural interactions, we might attempt to sidestep these kinds of issues altogether by using alternative methods of interaction. One alternative that has been proposed for advanced applications is voice command, using voice recognition technology to support system interaction. For many years, researchers have worked toward introducing reliable voice input technology into user interfaces [9]. There is a substantial body of literature that documents advancements in speech and voice recognition technology, and a variety of user studies that report enhanced performance of voice-activated interfaces relative to, and in conjunction with, conventional input devices such as keyboards and mice [3,16].

Assuming that the speed and accuracy of voice recognition technology continues to advance, Oviatt and Cohen [15] suggest that vocal interaction may someday be more efficient and accessible. They advocate a mixed approach to interaction, where users can choose between vocal localization and spatial interaction in complex multimodal interfaces. However, aside from the many existing technical concerns, there are also open questions about the inherent effectiveness and reliability of voice-based interfaces. As we will see, verbally reported location judgments can be subject to visual illusions, more than some motor responses, so pointing still seems to be our best choice for interaction in large-screen display environments.

1.3   Pointing for Large-Screen Displays

The use of laser pointers as tools for interaction in large-scale collaborative environments has established pointing as an important research topic, particularly for shared displays and large-scale systems. Kirstein and Muller [10] demonstrate how a camera-based approach can be used to acquire laser projections for display interaction. Oh and Stuerzlinger [13] expand upon this work, describing a system that permits multiple laser pointers to be used in a Single Display Groupware (SDG) environment. In a comparison with other devices for spatial interaction, including touch-screen styli and conventional mice, Myers et al. [12] note that laser pointers present serious technical and user performance challenges.

Limitations with laser pointers, and presumably other similar pointing devices, are most apparent in the difficulties that surround object selection through

button press and the reduction of individual hand jitter [14,21]. Fortunately, continual advances in our understanding of gesture interaction and improvements in physical device design (grip style, form factor, etc.) should eventually make it possible for laser pointers and other pointing devices to overcome many of these physical limitations.

1.4   Cognitive Factors in Spatial Interaction

Formal predictive models of user interaction such as Fitts Law [5] have provided us with an understanding of the spatial and motor factors that limit users' abilities to perform spatial interactions in a variety of situations. We have a comparatively poorer understanding of the cognitive perceptual factors that affect our capacity to execute motor movements in complex display environments. It is clear that the impact of visual feedback on spatial interaction is complex. Previous studies have provided evidence of a strong connection between the perceptual structure of visual information in an interactive display environment and the control structures of the input devices that facilitate interaction [8,20]. Other perceptual phenomena associated with control actions, such as kinesthetic and proprioceptive feedback, also affect spatial interactions in large-screen display environments [1,7].

Research in experimental psychology and functional neuroanatomy provides substantial evidence for cognitive relationships between visual perception and motor action that could affect performance in interactive systems. Central to this is a model of human cognition and motor processes that is known as the *two visual systems hypothesis* [2,11,19]. The hypothesis claims that visually-guided action can be mediated by either or both of two separate mental representations of space: a *cognitive* representation, which generates a world-relative view of visually-perceived objects in surrounding space, and a *sensorimotor* representation, which generates an egocentric view of these same objects. These representations have been isolated in different areas of the brain, primarily through work with brain-damaged patients who lack one of the two systems [11].

The two visual systems hypothesis states that the process of spatial motor interaction draws from two independent mental representations of the visual world, with different performance characteristics. The cognitive representation of space is susceptible to a large set of visual illusions and perceptual phenomena (i.e. the Induced Roelofs Effect, described below) that may bias performance based upon visual representation. In contrast, the sensorimotor representation of space is believed to be less susceptible to these illusory artifacts, but is also thought to have less information overall. The hypothesis further postulates that the two representations of space are informationally encapsulated, and that only one of the representations of space is typically active in configuring the execution of a particular spatial interaction.

Past studies in experimental psychology show that the choice of mental representation that is used in a particular spatial motor interaction, such as pointing, is strongly influenced by the level of visual feedback that is provided to participants [2,11]. These studies also suggest that spatial references to objects through

a non-motor interaction will consistently draw from the cognitive representation of space, with the result that non-motor interactions are highly subject to perceptual biases by a broad range of visual illusions.

Perhaps one of the best examples of how the two visual systems influence task performance is the classic experiment by Milner and Goodale [11] with "blindsight" patients who could do reaching and grasping tasks accurately even in the presence of cues that would ordinarily override the sensorimotor system with the inaccurate cognitive system. Without access to the normally dominant cognitive representation of space, "blindsight" patients rely solely on the sensorimotor representation of space and thus are not affected by these cues.

1.5   The Induced Roelofs Effect

The *Induced Roelofs Effect* [18] is one example of a much larger class of visual display illusions that have been found to selectively bias cognitive representations of visual space. Other illusions such as apparent motion and the motion aftereffect often produce similar effects. Because of the simplicity and effectiveness of the Induced Roelofs Effect we chose to use it as the basis for our current study. This illusion affects the ability of individuals to correctly perceive the spatial positions of presented targets in relatively sparse environments.
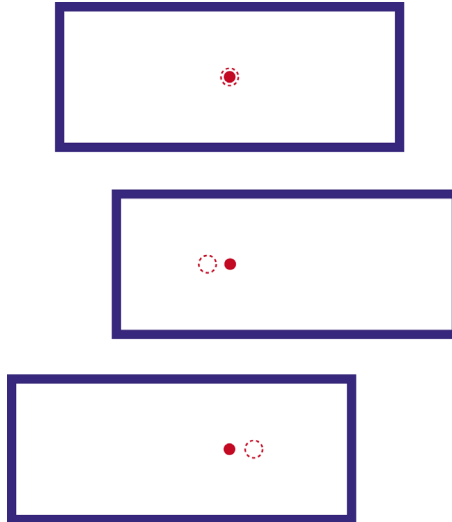

Fig. 1. The Induced Roelofs Effect. When a frame is offset, it appears that a target within the frame is more to the left or right than it really is. Solid circles are actual target positions, while dashed circles are observed reports of target position.

The illusion can be described as a systematic bias in the perceived location of targets that are presented within a rectangular frame that is offset to the left

or to the right of a target. When a target and rectangular frame are presented simultaneously, and the frame is offset to the left, there is a tendency to perceive the presented target to be further to the right than it actually is. Likewise, when the frame is offset to the right, there is a tendency to perceive the presented target as being further to the left than it actually is. When the frame is centered, there is no effect. Figure 1 illustrates this by showing the relationship between targets (solid circles) and their perceived locations (dashed circles) in the presence of offset frames.

Because the basic characteristics of the Induced Roelofs Effect are present in many graphical applications (either explicit frames, or display objects that may be perceived as implicit frames), it is possible that this visual illusion may exert some influence in large-screen display environments where the position of the frame varies over time. Related illusions such as apparent motion and the motion after-effect can also be created when virtual camera motion creates large visual flow fields (e.g. in a flythrough with an off-axis point of interest or a long pan).

From the two visual systems hypothesis, we can predict that a cognitive visual system response measure such as verbal report of target location will be more sensitive to the Induced Roelofs Effect than will a sensorimotor visual system response measure such as pointing with no cursor and an unseen hand. The first two conditions in our study test this prediction and essentially replicate Bridgeman et al. [2], but in a setting more typical of large-screen collaborative environments.

1.6   Predictions for HCI

The classic empirical studies of the two visual systems hypothesis in psychology laboratories were conducted in environments devoid of visual feedback to more clearly differentiate between the two systems and demonstrate their autonomy. In applying this theory to HCI we focus instead on the impact of multiple visual representations on the performance of tasks that are characteristic of large display systems, such as pointing with a cursor that may or may not lag the pointer position in time.

Thus, in our extension of earlier psychology studies we predict that pointing will draw from the cognitive representation (and hence exhibit the illusion) only if sufficient visual feedback is provided, in our case by a visible cursor. This prediction is counterintuitive, as it postulates poorer performance with visual feedback present than with it absent. A lagged cursor can be considered an intermediate condition, since the correspondence between pointer motion and cursor motion is less convincing. This leads to our second counterintuitive prediction, that a cursor with a large temporal lag may in fact improve user performance relative to more responsive feedback. We note that these predictions are true only under a specific set of display conditions that are described by the two visual systems hypothesis.

If our mapping of the two visual systems hypothesis onto HCI is correct, then the choice of interaction method and the perceptual feedback that is pro-

vided to users are influential components in determining which of the two visual representations is used. If this is borne out by experimental evidence, we may wish to choose interaction techniques that selectively utilize the sensorimotor representation in conditions where visual illusions are likely. This should lead to more precise performance in some cases, despite the lower overall accuracy of the sensorimotor system. We might then investigate ways to improve the accuracy of that system. These will be discussed later in this paper. First, we present our experimental study.

2   Method

Our study tested our predictions with a within-subjects methodology by using a large-scale display environment that is typical of collaborative situations and interactive large-screen display environments. We used a simple target acquisition task to provide a situation where vocal localization and spatial motor interaction might be equally feasible methods of interaction.

Participants completed four blocks of trials that required them to acquire targets through either voice selection or motor pointing in the presence of a centered or an offset frame. The voice selection versus pointing without visual feedback conditions replicate Bridgeman et al. [2] in our experimental setup, while the pointing with visual feedback and lagged visual feedback conditions extend the methodology to more familiar interactive systems. We are currently unaware of any studies, from experimental psychology or elsewhere, that have looked at pointing with rendered visual feedback in the context of this cognitive-motor model.

2.1   Design

Each participant attended a single, individual session between one and a half hours to two hours in duration. There were four blocks of 54 trials each, with each block characterized by a different method of interaction that changed how a participant indicated the perceived position of a displayed target. The first block (i.e. condition) did not involve pointing.

 – A simple voice protocol for selecting specific targets was used. Since this technique was considered inherently cognitive, interaction of this kind was called cognitive report. All of the participants started their sessions with this method of interaction.

The other three conditions used continuous spatial pointing with varying degrees of visual feedback.

 – A spatial pointing measure without visual feedback of any kind was called *pointing without visual feedback*. The lack of visual feedback eliminated any possibility of making cognitive corrections to the pointing gestures that were being made, effectively breaking the relationship between action and visual perception of that action.

– A spatial pointing measure with visual feedback in the form of a small cursor was called *pointing with visual feedback*. This is similar to what is normally provided in typical interactive environments.
– As a final exploratory measure, we tested the effects of spatial pointing in the context of unreliable visual feedback. This was done by inducing a one-half second lag to the same rendered cursor that was used in the previous pointing with visual feedback condition, effectively creating a *pointing with lagged visual feedback* condition. The reason for using such a large amount of lag was not to simulate the interactive lag seen in typical interactive systems, but rather to identify any potential influence of lag on cognitive performance.

To ensure that there would be no unwanted order effects between the pointing without visual feedback and pointing with visual feedback conditions, the order of these two methods of interaction were counterbalanced across participants. All participants ended their sessions with the pointing with lagged visual feedback condition.

2.2   Participants

We recruited fourteen participants between the ages of 17 and 31 to participate. Seven of the participants were male, and seven of the participants were female. All of the participants had normal or corrected-to-normal vision. Twelve of the participants were right-handed, while the remaining two participants were left-handed.

2.3   Apparatus

For each session, participants were seated and centered in front of a three-screen projection display arranged in an angled configuration at a distance of two and a half meters. During this study, only the center screen of the three screens was used. Figure 2 is a photograph of the experimental environment. The active display screen was forward-projected, with a width of 2.75 meters and a height of 2.15 meters. With the exception of illumination from the active screen, all other light sources were extinguished during the sessions. An experimenter was present at all times during all of the sessions.

A PC workstation and a custom software application were used to render trials, to control the flow of each session, and to record quantitative data from participating participants. A six degree-of-freedom Polhemus Fastrak with an attached stylus and a second passive receiver was used to provide spatial interaction for the study. The stylus was used as a pointing implement that was similar in size and shape to a pen or laser pointer. The passive receiver was attached to a plastic helmet as a head-tracking device. To ensure that the performance of the Fastrak would not be impeded by environmental interference, most of the equipment used in the experiment was non-metallic in nature, and all possible sources of metallic interference were kept away from the display environment.
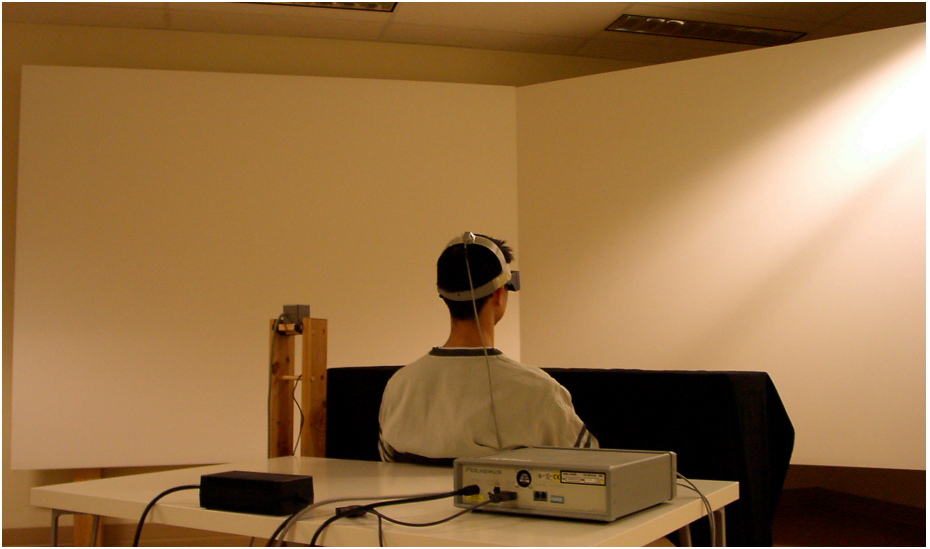
Fig. 2. The room set up that was used during the user study. Participants kept their hands and arms underneath a table set directly in front of them. They also wore a pair of stereo glasses and a head tracker. Spatial interaction was provided by a Polhemus Fastrak and an attached stylus. During the actual sessions, all ambient lighting was extinguished and an experimenter was always present.

In order to limit visual feedback to only the feedback provided on the display, participants were asked to keep their arms and hands out of their own sight at all times by placing them underneath a large, wooden table with a width of 1.2 meters, a length of 0.80 meters, and a height of 0.95 meters. Although no stereoscopic imagery or head-coupled perspectives were used in the present study, participants were required to wear a pair of active stereo glasses and the head-tracked plastic helmet at all times. This particular requirement is intended to support comparison between this study and future planned investigations using stereo and head coupling; similarly, the presence of three screens is for consistency with future studies of wide-angle displays.

2.4   Procedure

Trials began with the presentation of a small circular target, 0.5 degrees of visual angle in diameter, in one of three positions on the display. Targets were either directly ahead of participants or displaced 1.5 degrees to their left or right. The circular targets were surrounded by a rectangular frame, 21.0 degrees in width, 9.0 degrees in height, and 1.0 degree in line thickness. The frame also appeared in one of three offsets: aligned symmetrically to participants' midline, or offset asymmetrically by 4.0 degrees to the left or to the right. Both target and frame were rendered on a black background. This simultaneous presentation of frame

and target lasted for a period of exactly one second, after which all stimuli were extinguished and the task was performed. Participants were asked to specify the location of the previously presented target either immediately, or after a four second delay. Figure 3 illustrates the five targets used in practice trials (color is not accurate in the figure).
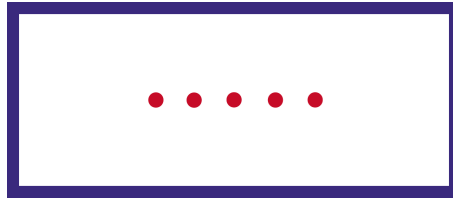

Fig. 3. The five target positions used in our study. Targets were 0.5 degrees in visual angle and were spaced apart by 1.5 degrees from one another. These targets were surrounded by a rectangular frame, 21.0 degrees in width, 9.0 degrees in height, and 1.0 degree in line width. Only one of the five targets was displayed for any given trial.

These conditions resulted in eighteen trial types: three target positions by three frame offsets at two response delays. All trial types were repeated three times in random order for a total of 54 trials per block.

Each block of 54 experimental trials was preceded by a set of practice trials that gave participants an opportunity to familiarize themselves with the possible responses for each condition. Practice trials had five possible targets, the three used in the experimental trials and two additional locations 1.5 degrees to the left of the left target (Far Left), and 1.5 degrees to the right of the right target (Far Right). The far left and far right responses were provided to give participants practice responding to experimental targets that might appear to be at a greater eccentricity than their actual positions as a result of the Induced Roelofs Effect (see below). Practice trials began with a set of randomized target-only presentations, followed by successive sets of randomized target presentations with a frame fixed in the non-offset position. Participants were provided with a minimum of fifteen practice trials for each interaction technique.

Cognitive Report Condition. Cognitive report interaction was simulated in a Wizard of Oz fashion by the experimenter. After presentation of a particular target and frame offset, participants would respond with one of five verbal responses: "Far Left," "Left," "Center," "Right," and "Far Right." Upon hearing a participant's response, the experimenter recorded the response using the experiment software.

The Far Left and Far Right responses were provided to allow participants to respond naturally in conditions where the Induced Roelofs Effect might make it appear as if the targets were at a greater eccentricity than their actual positions (i.e. when target and frame offsets were in opposing directions).

Pointing Conditions. Pointing was performed with the Fastrak stylus. When asked to respond after the presentation of particular target and frame offsets, participants would respond by making an aiming response. When participants were satisfied with their final aiming position, they were asked to hold the stylus position until an audible indication was provided after a period of approximately two seconds. Such a "point-and-dwell" mechanism has been previously suggested to reduce artifacts that result from button presses related to object selection with laser pointers and other pointing implements (the "pen-drop" phenomenon [14]).

3   Results

Participant data that were collected during the study were analyzed in two ways. We explain each analysis before presenting the data. We initially performed a 4 (report) x 3 (target) x 3 (offset) x 2 (response delay) analysis of variance on the data for each subject. No effects were found for response delay, which was a condition included for compatibility with Bridgeman's earlier study [2]. We therefore collapsed the response delay conditions, resulting in six trials for every cell instead of just three.

3.1   Primary Analysis

Two-way ANOVAs tested participant response against independent factors of target position and frame offset and a single dependent factor of participant response. These analyses were run for each participant for the collapsed response delay and each interaction technique to enable inter-subject comparison of the presence or absence of an effect.

For cognitive report, participant responses were one of the five previously described verbal choices. For the pointing measure, participant responses were recorded as the position on the screen where a line projected along the axis of the pointing device would intersect the screen. We note that this measure was some transformation of the intended position, as participants typically point in a body-centric frame (e.g. shoulder or head-centered) rather than an artifact-centric one. In an interactive system it is possible to rescale these measures to translate between user and system coordinates. Only the horizontal (x) coordinate was used in the analysis.

The two-way ANOVAs enabled us to statistically test for the presence or absence of the Induced Roelofs Effect in each participant and condition. In this analysis, presence of the visual illusion would be reflected in a main effect for frame offset that was not complicated by a significant interaction with target position. Presence of a main effect for frame offset in most subjects would support our hypothesis.

3.2   Secondary Analysis

We performed an alternate analysis that set an absolute threshold criterion for the number of trials in which participants were deemed to exhibit the Induced Roelofs Effect. This was done as follows.

For the cognitive report condition, the categorical data was used directly. For each of the three pointing conditions, a separate analysis was made for each participant of the errors in target location for trials that involved offset frames. This was performed in two stages. The first stage, repeated for each of the three pointing conditions, considered only the trials for which the frame was not offset. The mean and standard deviation of a participant's response (the x-coordinate of where the participant pointed on the screen) were computed for each of the three presented target positions (recall that "Far Left" and "Far Right" were never presented in experimental trials).

The second stage used these values to categorize the remaining trails (those with offset frames) as being either a Roelofs trial or a non-Roelofs trial. An offset trial was considered to be Roelofs if the participant's response (x-coordinate) was more than one standard deviation to the left (when the frame is offset right - but to the right when the frame is offset left) of the mean non-offset response for that participant for that target position in that condition. A similar categorization was used for trials without offset frames, which were considered to be non-Roelofs if the response was not within one standard deviation of the mean.

A participant was considered to have exhibited Roelofs behavior if more than 50% of the offset trials were Roelofs and less than 50% of the non-offset trials were non-Roelofs. This criterion can be paraphrased by saying that the participant had to make an error at least half of the time in the predicted direction when the frame was offset and had to be close to the correct position at least half of the time when no error was anticipated. The choice of a 50% threshold and the choice of one standard deviation for categorization is arbitrary, but its application is automatic, involving no subjective judgments by the experimenter.

By setting an arbitrary cutoff for the presence or absence of the Induced Roelofs Effect for particular participants, we have a rough indication of which of the two mental representations of space was probably used for the task. We predicted that cognitive report and pointing with visual feedback would demonstrate an Induced Roelofs Effect due to their common use of the cognitive representation of space. We also predicted that pointing without visual feedback would not demonstrate an effect of the illusion, since lack of visual feedback should motivate use of the sensorimotor representation of space.

3.3   Summary of the Data

All of the participants under all of the different interaction techniques were observed to have highly significant main effects of target position, as would be expected if they were performing the task as instructed. Figure 4 presents a summary of the results of the ANOVAs (the primary analysis) and the proportion of trials where responses were consistent with an Induced Roelofs Effect (the

secondary analysis). The figure gives individual p-values from the ANOVAs and the percentage of trials that were judged to show the effect.

Participant Number	Cognitive Report (Vocal Interaction)		Pointing Without Visual Feedback		Pointing With Visual Feedback		Pointing w/ Lagged Visual Feedback	
1	0.454	5.6%	0.577	30.6%	0.834	27.8%	0.291	27.8%
2	0.465	8.3%	0.821	25.0%	0.838	13.9%	0.359	22.2%
3	0.387	5.6%	0.789	33.3%	0.838	27.8%	0.898	38.9%
4	0.178	16.7%	0.936	33.3%	0.834	30.6%	0.084	36.1%
5	**0.001 ***	**100.0%**	0.923	33.3%	**0.037 ***	**41.7%**	0.961	25.0%
6	**0.001 ***	**75.0%**	**0.011 ***	33.3%	0.488	19.4%	**0.021***	33.3%
7	**0.001 ***	**100.0%**	0.841	27.8%	**0.002 ***	**55.6%**	0.074	27.8%
8	**0.001 ***	**100.0%**	0.084	38.9%	**0.001 ***	**63.9%**	**0.008***	**86.1%**
9	**0.001 ***	**94.4%**	**0.001 ***	44.4%	**0.037 ***	**77.8%**	0.094	41.7%
10	**0.001 ***	**88.9%**	0.126	47.2%	**0.001 ***	**83.3%**	**0.005***	**94.4%**
11	**0.001 ***	**100.0%**	**0.001 ***	38.9%	**0.001 ***	**75.0%**	**0.030***	38.9%
12	**0.001 ***	**83.3%**	0.177	47.2%	**0.001 ***	**61.1%**	**0.021***	**77.8%**
13	**0.001 ***	**88.9%**	0.281	30.6%	**0.028 ***	**52.8%**	0.147	36.1%
14	**0.001 ***	**88.9%**	**0.013 ***	38.9%	0.294	44.4%	**0.002***	**44.4%**
Means		*67.9%*		*35.9%*		*48.2%*		*45.0%*

Fig. 4. A summary of individual participant performance in the study. The p-values with asterisks (*) indicate that a particular participant had an observed main effect of frame position for a particular interaction technique. Corresponding numerical values indicate the percentage of frame-displaced trials that were judged to exhibit the Induced Roelofs Effect for a particular participant with a particular interaction technique. Statistically significant values are highlighted in bold.

In the discussion that follows, we follow a standard practice in experimental psychophysics and report the smallest F-values for significant results, and the largest F-values for non-significant results. For main effects of target position, we have $[F(2, 18) > 4.214, p < 0.032]$, indicating that each participant's trial responses were consistent and reliable across all conditions. The interesting analyses were for frame offset.

Cognitive report was found to be most sensitive to the Induced Roelofs Effect, with ten of the fourteen participants exhibiting significant main effects of frame offset $[F(2, 18) > 4.460, p < 0.027]$. For these participants, the mean magnitude of the effect was measured to be roughly one discrete target position from the actual presented target positions. These results were echoed in the percentages of trials that were judged to exhibit the visual illusion, where these participants were observed to easily surpass the defined criteria.

Those four participants who did not exhibit an Induced Roelofs Effect in cognitive report also consistently exhibited no significant illusory effects in any

of the other remaining conditions. There is a particularly interesting division between those participants who either did or did not exhibit the effect. Participants who exhibited a significant effect quite consistently did so, while those who did not quite consistently did not. These observations suggest that despite expected differences in individual participants susceptibility to the illusion, the observed effects are consistent for each participant, and are not due to noise in participant responses.

Of the ten participants who exhibited an Induced Roelofs Effect for cognitive report, six were found to have no effect when asked to respond by pointing without visual feedback $[F(2, 18) < 2.845, p > 0.084]$. This is consistent with our hypothesis that some participants utilize different representations of space for different interaction techniques. The remaining four participants who did show significant effects may have continued to draw from their cognitive representation of space to accomplish the target acquisition task (earlier research also observed some subjects who failed to exhibit any effect [2]). Nevertheless, the percentage of trials that were judged to exhibit the Induced Roelofs Effect demonstrates that there is a marked drop in the number of participant responses that were affected by the visual illusion. In particular, a chi-square test against the trials from cognitive report across all of the participants confirms that there is a highly significant difference $[(1, N = 14) = 12.6, p < 0.001]$ in illusory influence between cognitive report and pointing without visual feedback.

Interestingly, of the six participants who exhibited an Induced Roelofs Effect for cognitive report but not for pointing without visual feedback, all six exhibited a significant effect when asked to respond by pointing with visual feedback $[F(2, 18) > 3.850, p < 0.05]$. This appears to suggest that pointing with visual feedback is more susceptible to illusory effects than pointing without visual feedback, which is entirely consistent with our hypothesis that visual feedback motivates the use of the cognitive representation of space. Once again, these effects are reflected by the percentage of trials that were judged to exhibit the Induced Roelofs Effect, which is noticeably higher than the count observed in pointing without visual feedback. A chi-square test against the trials from vocal interaction across all of the participants found no significant difference between vocal interaction and pointing with visual feedback $[(1, N = 14) = 1.4, p = 0.237]$, suggesting that these two interaction techniques share similar susceptibility to the illusion.

If we distinguish between accuracy and precision of different interaction techniques, another interesting result strongly indicates that pointing without visual feedback is consistent and reliable - it is reasonably precise, although its overall accuracy is generally much poorer than pointing with some level of visual feedback. In an analysis of individual pointing data without visual feedback, it was apparent that most of the errors could be attributed to a consistent bias along the horizontal axis, and that these errors of variance can be characterized as rescaled versions of the variance seen in the corresponding pointing trials with visual feedback. This may reflect participants' choice of a particular body-centered frame for the pointing task.

Our final analysis looked at pointing with lagged visual feedback. We found that only three participants out of the six who showed effects in pointing with and without visual feedback also showed an Induced Roelofs Effect with this interaction technique. A chi-square test against the trials from vocal interaction also demonstrates a significant difference $[(1, N = 14) = 5.6, p = 0.018]$. This suggests that the susceptibility of lagged visual feedback to the Induced Roelofs Effect is intermediate between pointing with purely reliable visual feedback and no visual feedback at all, and it is less susceptible than vocal interaction. However, there are several other factors, especially the lack of counterbalancing between this interaction technique and the other studied techniques, which could have contributed to this finding, and so we cannot reach firm conclusions about interactive lag based on these data.

3.4   Sex Differences

One unexpected finding was a sex difference between participants who exhibited an Induced Roelofs Effect for cognitive report. Only three of the seven male participants demonstrated a significant Induced Roelofs Effect in this condition, while all seven female participants demonstrated the effect. Subsequent t-tests testing gender against the different interaction conditions confirmed this effect for vocal interaction (t=2.828, p=0.030), but were not significant for other interaction techniques. We might interpret such a difference as the result of different levels of contextual influence in the environment. It seems possible that males may be less sensitive than females to context in location judgments under some conditions.

4   Discussion and Future Work

Our results suggest that the selection and implementation of spatial interaction techniques should take into account cognitive factors that may introduce performance errors. Such errors may be unacceptable for safety-critical applications, or where errors of interaction are not easily correctable, or where such errors may go unnoticed.

Visual illusions are an increasing concern when we move away from desktop environments to other, less-constrained environments, such as large-screen displays. The Induced Roelofs Effect is one example of an influential visual illusion, and other illusions that affect motion or size-contrast perception may have stronger effects. Being aware of these cognitive factors may help us design future environments and applications that are robust against such effects.

Our study manipulated frame offset to obtain an Induced Roelofs Effect. We believe that perceptual effects similar to this particular visual illusion are possible in a variety of display situations. For example, an Induced Roelofs Effect might occur when large background objects, perceived as framing a scene, move as a result of changes in rendered data or changes in viewpoint. In collaborative display applications, where interaction is shared by multiple users, instances may

arise where one individual user moves an object that serves as a perceived frame for another user.

In exocentric target acquisition environments, such as advanced air traffic control displays, multiple targets may be grouped within a very small display area, surrounded by contextual cues that might bias perceived spatial position. For these displays, updates may not occur at fully interactive rates, nor necessarily synchronously, and users may attempt to acquire targets by using past spatial perceptions as predictors to improve target tracking, thereby creating an environment that is similar in many key respects to the experimental one we have used. Situations such as these effectively create stimuli whose underlying perceptual effects are similar to those used in this study. Thus, at least where voice command or pointing with visual feedback are concerned, we can expect to see the kinds of errors that occurred in the present study.

Our findings suggest that pointing without visual feedback is a potentially viable technique for spatial interaction with large-screen display environments that should be seriously considered. Pointing without visual feedback solves the problems with multiple cursors that can arise from collaborative interaction with large-scale shared displays. Pointing without visual feedback also appears to minimize the impact of perceptual artifacts and visual illusions on user performance. Compensating for the relatively poor accuracy of pointing without visual feedback is the most significant obstacle to using it as a spatial interaction technique. Our finding that much of the error in pointing without visual feedback has a consistent bias suggests that future work might examine how we can correct for individual differences in order to make pointing without visual feedback a more useful spatial interaction technique.

5   Conclusion

Our comparison of different techniques for spatial interaction with a large-scale interactive display suggests that pointing without visual feedback may be a viable technique that should be considered when developing applications for large-screen display environments. Using the Induced Roelofs Effect as an example of a class of illusions described by the two visual systems hypothesis, we established a differential impact of the illusion on different interaction conditions and for different individuals. Inherently non-motor interactions, such as cognitive report (voice response), were most susceptible to this illusion, while pointing without visual feedback was least susceptible. Several issues remain to be explored, motivating future investigations into interaction techniques that minimize the use of visual feedback in large-screen environments where these kinds of perceptual illusions are likely to occur.

References

1. Balakrishnan, R., Hinckley K.: The Role of Kinesthetic Reference Frames in Two-Handed Input Performance. UIST 1999, ACM Symposium on User Interface Software and Technology (1999) 171–178

2. Bridgeman, B., Peery, S., Anand, S.: Interaction of Cognitive and Sensorimotor Maps of Visual Space. Perception and Psychophysics, Vol.59, No. 3 (1997) 456–469

3. Christian, K., Kules, B., Shneiderman, B., Youssef, A.: A Comparison of Voice Controlled and Mouse Controlled Web Browsing. ASSETS 2000, 72–79

4. Cooper, C.: Visual Dominance and the Control of Action. Proceedings of the 20th Annual Conference of the Cognitive Science Society (1998), 250–255

5. Fitts, P. M.: The Information Capacity of the Human Motor System in Controlling Amplitude of Movement. Journal of Experimental Psychology, Vol. 47 (1956) 381–391

6. Hayne, S., Pendergast, M., Greenberg, S.: Gesturing Through Cursors: Implementing Multiple Pointers in Group Support Systems. Proceedings of the Hawaii International Conference on System Sciences (1993)

7. Hinckley, K., Pausch, R., Proffitt, D.: Attention and Visual Feedback: The Bimanual Frame of Reference. 1997 ACM Symposium on Interactive 3D Graphics (1997) 121–126

8. Jacob, R. J. K., Sibert, L. E., McFarlane, D. C., Mullen, M. P.: Integrality and Separability of Input Devices. ACM Transactions on Computer-Human Interaction, Vol. 1, No. 1 (1994) 3–26

9. Karl, L., Pettey, M., Shneiderman, B.: Speech Activated Versus Mouse-Activated Commands for Word Processing Applications: An Empirical Evaluation. International Journal of Man-Machine Studies, Vol. 39, No. 4, October (1993) 667–687

10. Kirstein, C., Muller, H.: Interaction with a Projection Screen Using a Camera-Tracked Laser Pointer. Proceedings of Multimedia Modeling (1998) 191–192

11. Milner, A. D., Goodale, M. A.: The Visual Brain in Action. Oxford Psychology Series, Vol. 2, (1995) New York: Oxford University Press.

12. Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., Long, A. C.: Interacting at a Distance: Measuring the Performance of Laser Pointers and Other Devices. Proceedings of CHI 2002, 33–40

13. Oh, J., Stuerzlinger, W.: Laser Pointers as Collaborative Pointing Devices. Proceedings of Graphics Interface 2002, 141–149

14. Olsen, D. R., Neilsen, T.: Laser Pointer Interaction. Proceedings of CHI 2002, 17–22

15. Oviatt, S., Cohen, P.: Multimodal Interfaces that Process What Comes Naturally. Communications of the ACM, March, Vol. 43, No. 3 (2001) 45–53

16. Poock, G. K.: Voice Recognition Boosts Command Terminal Throughput. Speech Technology, Vol. 1, No. 2 (1981) 36–39

17. Poupyrev, I., Weghorst, S., Billinghurst, M., Ichikawa, T.: A Framework and Testbed for Studying Manipulation Techniques in Immersive VR. VRST 1997, ACM Symposium on Virtual Reality Software and Technology (1997) 21–28

18. Roelofs, C.: Optische localisation [Optical localization]. Archiv für Augenheilkunde, 109 (1935) 395–415

19. Trevarthen, C. B.: Two Mechanisms of Vision in Primates. Psychologische Forschung, Vol. 31 (1968) 299–337

20. Wang, Y., MacKenzie, C. L., Summers, V. A., Booth, K. S.: The Structure of Object Transportation and Orientation in Human-Computer Interaction. Proceedings of CHI 1998, 312–319

21. Winograd, T., Guimbretiere, F.: Visual Instruments for an Interactive Mural. Abstracts, Proceedings of CHI 1999, 234–235

22. Wolf, C. G., Rhyne, J. R.: Gesturing with Shared Drawing Tools. ACM Conference on Human Factors and Computing Systems (1993)