Effects of Visual Conflicts on 3D Selection Task Performance in Stereoscopic Display Environments

Gerd Bruder[*], Frank Steinicke[†]
Department of Computer Science
University of Würzburg

Wolfgang Stürzlinger[‡]
Department of Computer Science and Engineering
York University

ABSTRACT

Mid-air direct-touch interaction in stereoscopic display environments poses challenges to the design of 3D user interfaces. Not only is passive haptic feedback usually absent when selecting a virtual object displayed with positive or negative parallax relative to a display surface, but such setups also suffer from inherent visual conflicts, such as vergence/accommodation mismatches and double vision. In particular, if the user tries to select a virtual object with a finger or input device, either the virtual object or the user's finger will appear blurred, resulting in an ambiguity for selections that may significantly impact the user's performance.

In this paper we evaluate the effect of visual conflicts for mid-air 3D selection performance within arm's reach on a stereoscopic table with a Fitts' Law experiment. We compare three different techniques with different levels of visual conflicts for selecting a virtual object: real hand, virtual offset cursor, and virtual offset hand. Our results show that the error rate is highest for the real hand condition and less for the virtual offset-based techniques. However, our results indicate that selections with the real hand resulted in the highest effective throughput of all conditions. This suggests that virtual offset-based techniques do not improve overall performance.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input Devices and Strategies, Evaluation / Methodology.

1 INTRODUCTION

Stereoscopic display poses challenges to the design of user interfaces for semi-immersive virtual environments (VEs). Standard (multi-)touch technologies can provide intuitive and natural interaction with objects displayed with zero parallax on a display surface. But, sensing human gestures and postures in "mid-air" above the surface introduces challenges to the design of high-performance interaction techniques for selection and manipulation [1, 16]. Both the increase in the degrees-of-freedom that have to be controlled as well as the absence of passive haptic feedback and resulting interpenetration and occlusion issues can reduce performance when touching virtual objects displayed with negative parallax, i. e., in front of a display surface [1]. Objects displayed behind the surface with positive parallax cannot be reached by direct interaction.

Using *direct input* in such setups suffers from inherent visual conflicts [15]. In particular, when a user tries to touch a virtual object that is displayed with negative parallax, either the virtual object or the user's finger will appear blurred. While visual distance cues from the convergence angle to the virtual object and to the user's finger may indicate that they are at the same position, the focus distance is usually adjusted to the user's real finger, with the more distant display surface being out of focus. This results in the finger appearing sharp, whereas the virtual object appears blurred (see Figure 1(a)). However, such relative differences in blur between objects are used by the perceptual system to judge interrelations and relative distances between objects [5]. This visual problem thus results in an ambiguity when trying to select a virtual object using a finger or input device and may significantly impact user performance.

A possible solution to these visual conflicts comes from research on direct interaction versus interaction at a distance [2, 11, 12]. A simple solution is to place a virtual cursor with a short offset next to the user's tracked finger and then to use this *offset cursor* to select other virtual objects. Since both the offset cursor and the virtual objects are displayed stereoscopically, there is then no more mismatch in blur between real and virtual object (see Figure 1(b)). While this reduces visual conflicts, it is not clear whether this results in improved overall selection task performance. In particular, decoupling the motor and visual space during natural hand interaction may degrade performance due to the kinematics of point and grasp gestures in 3D space and the underlying cognitive functions [9, 17]. Moreover, using a virtual offset cursor does not eliminate all visual conflicts in stereoscopic display environments. With increased parallax and distance from the display surface, the vergence-accommodation conflict increases, which has been found to affect size and distance judgments, as well as judgments of interrelations between displayed virtual objects [8]. To account for this limitation, multiple approaches have been proposed to better support spatial perception in stereoscopic setups. For example, research suggests that spatial judgments benefit from familiar size cues being provided to the user, such as using a *virtual hand* for interaction with virtual objects [15, 18] (see Figure 1(c)).

In this paper we compare selection task performance on a stereoscopic table with a Fitts' Law experiment testing three input techniques with different levels of visual conflicts. We use metrics of movement time, error rates, error distances, and resulting effective throughput as overall performance indicator. We compare selecting virtual objects with three approaches:

1. Direct input with the tip of the user's index finger of the dominant hand (see Figure 1(a)).
2. Distant input with a virtual offset cursor at 10cm next to the tracked fingertip (see Figure 1(b)).
3. Distant input with a full virtual hand representation with the same offset as the offset cursor (see Figure 1(c)).

The results of our experiment provide guidelines for the choice of input techniques for interaction with graphical elements in stereoscopic display environments.

2 RELATED WORK

Selection behavior and performance with direct or indirect input techniques have been the focus of several areas of previous work.

---

[*] e-mail: gerd.bruder@uni-wuerzburg.de
[†] e-mail: frank.steinicke@uni-wuerzburg.de
[‡] e-mail: wolfgang@cse.yorku.ca

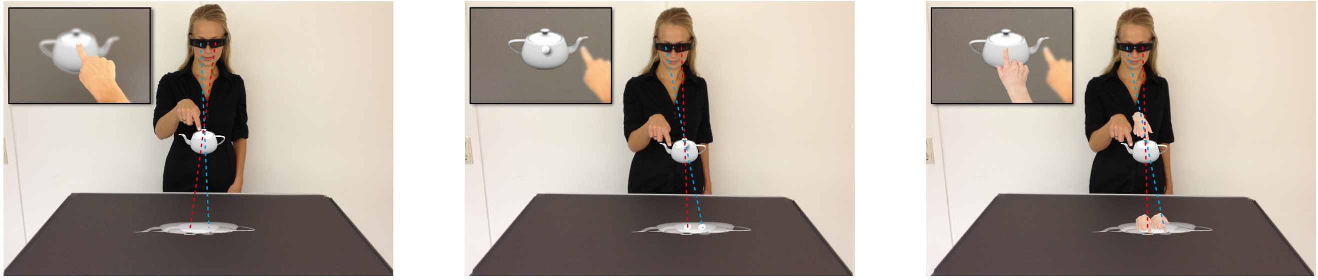

(a) (b) (c)

Figure 1: Illustration of visual conflicts during 3D selection of stereoscopically displayed objects: (a) The user is focused on her finger, with the virtual object appearing blurred. (b) Displaying a virtual offset cursor (white marker) at a fixed distance from the real fingertip reduces visual conflicts. (c) A virtual offset hand cursor provides familiar and additional size and distance cues for selection.

2.1 Direct Selection of Virtual Objects

Direct selection of stereoscopically displayed objects in the 3D space in front of a display is enabled by tracking technologies, such as optical marker systems, the Leap Motion, or the Microsoft Kinect. The evaluation of the kinematics and user behavior when selecting virtual objects by hand gestures or movements of the user's arm can be divided roughly in two phases [7]: a *ballistic phase* in which the user's attention is focused on the object to be selected and the hand is brought in the proximity of the goal through proprioceptive motor control, as well as a *correction phase* that incorporates visual feedback to incrementally reduce the distance between the hand and goal.

MacKenzie et al. [9] showed that Fitts' Law holds for the kinematics of arm movements, i. e., greater precision in the correction phase is accompanied by earlier deceleration of arm movements. When touching an intangible object in 3D space, the missing passive haptic feedback and visual conflicts cause an ambiguity in depth perception and object interrelations, leading to confusion and a significant number of overshoot errors [1]. Recent work has tried to reduce such conflicts by moving tangible surfaces into place when a user tries to interact with a virtual object [1].

2.2 Offset-based Selection of Virtual Objects

Using a head-mounted display (HMD), Mine et al. [11] investigated the differences between direct interaction and manipulation at a distance relative to the user's hand. Fixed and variable offsets significantly reduced performance compared to interaction with objects collocated with the user's hand. Since they used a HMD, there were no visual conflicts between the real hand and virtual objects in their evaluation. Poupyrev et al. [13] found no performance difference between direct object manipulation with a virtual hand shown in a HMD and the Go-Go technique using variable offsets within arm's reach. But results for lateral movements suggested that performance improves with direct manipulation. Using a stereoscopic workbench, Paljic et al. [12] placed a virtual crosshair at a short offset next to a tracked stylus held by the user, using this 3D cursor to select virtual objects. They found no differences in selection performance for offset cursors with a distance of 0cm and 20cm, but performance degraded for 40cm and 55cm. Djajadiningrat et al. [2] found no difference in performance for direct and offset-based manipulation in a fishtank VR setup with 0cm and 20cm offsets. The results suggest that humans may achieve optimal performance when visual and motor spaces are superimposed or closely coupled, i. e., when there is a consistent and pervasive illusion in the perceptual and motor systems, with large offsets resulting in degraded performance [17].

2.3 Fitts' Law and Selection

The Fitts' Law [4] empirical model of the tradeoff between speed and accuracy in selection tasks predicts the movement time MT for a given target distance D and size W by $MT = a + b \times \log_2(D/W + 1)$, with empirically derived values a and b. The *index of difficulty (ID)* is given by the log term and indicates overall task difficulty; smaller or farther targets result in increased difficulty. An extension of Fitts' Law is the use of "effective" measures [6]. Using this approach, the error rate is adjusted to the fixed value of 4% by re-sizing targets to their effective width (W_e). Using this, effective throughput incorporates both speed and accuracy into a single measure by "normalizing" the accuracy as effective scores. This throughput is computed as $TP = \log_2(D_e/W_e + 1)/MT$, with D_e the effective distance (average of measured movement distances) and W_e the effective width (standard deviation of error distances multiplied by 4.1333 [10]).

3 EXPERIMENT

We compared the three input techniques for selecting a virtual object in a Fitts' Law task on a stereoscopic tabletop setup with 3D targets shown at different heights above the display surface.

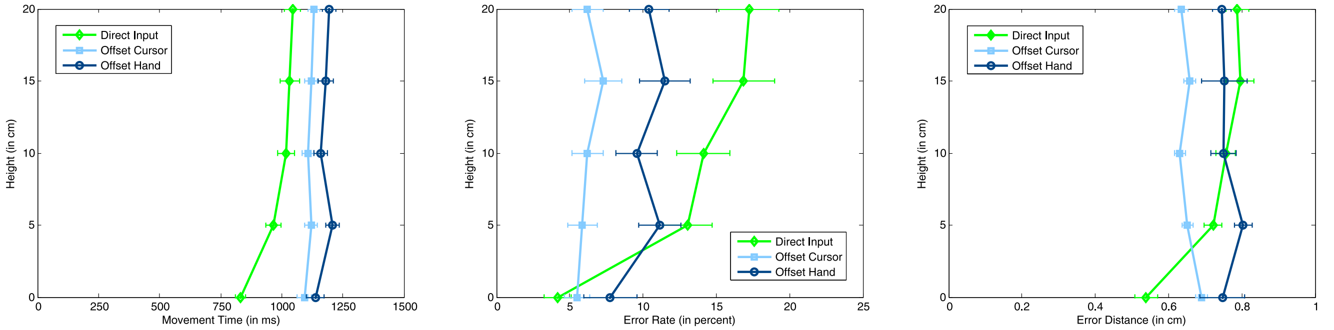
3.1 Participants

7 male and 16 female subjects (ages 18-30, M=21.0) participated in the experiment. Subjects were students of media communication or human-computer-interaction. All subjects received class credit for participating in the experiment. Subjects were right-handed and had normal or corrected to normal vision with interpupillary distances between 5.8cm and 7.3cm (M=6.3cm). Subjects were naïve to the experimental conditions. Subjects were allowed to take a break at any time between experiment trials. The total time per subject including questionnaires, instructions, training, experiment, breaks, and debriefing was about 1.5 hours.

3.2 Apparatus

The 62cm×112cm tabletop setup used in the experiment consists of a stereoscopic back projection screen with a 1280 × 800 Optoma GT720 projector (see Figure 3). Subjects wore active DLP-based shutter glasses at 60Hz per eye. We tracked wireless markers attached to the shutter glasses and another diffused IR LED at the tip of the index finger of a glove worn on the subject's dominant hand using a WorldViz PPT X4 tracking system. The virtual scene was rendered on an Intel Core i7 computer with 3.4GHz processors, 8GB of main memory, and an Nvidia Quadro 4000 graphics card. We measured an end-to-end latency of approximately 65ms between physical movements and a visual response.

The visual stimulus consisted of a virtual 3D box that matched the dimensions of the tabletop with a depth of 30cm (see Figure 3).



(a) (b) (c)

Figure 2: Results for the three techniques with target object height on the y-axis and (a) movement time, (b) error rate, and (c) error distance, on the x-axis. The error bars show the standard error.

Selection targets were represented by spheres on or above the display surface. Each trial consisted of 11 spheres arranged in a circle. Size, distance, and height of spheres were constant within circles, but varied between trials. Spheres were rendered in white, with the active target highlighted in blue. The targets highlighted in the order specified by the ISO 9241-9 standard [6] (see Figure 3(inset)). The center of the highlighted target sphere indicated the exact 3D position that subjects were instructed to touch with their dominant hand's fingertip. Targets highlighted green when touched, minimizing systematic errors in Fitts' Law experiments [9]. Subjects confirmed selections using a keypad with their non-dominant hand.

3.3 Methods

We used a $3 \times 5 \times 2 \times 2$ within-subjects design with the method of constant stimuli. Target positions and sizes were not related between circles, but presented randomly and uniformly distributed [3]. Independent variables were selection technique (*direct input*, *offset cursor*, and *offset hand*), target height (between 0cm and 20cm in steps of 5cm), as well as target distance (16cm and 25cm) and size (2cm and 3cm). Dependent variables were movement time (between selections), error distance (to target position), error rate (percentage of targets missed), and effective throughput (bits per second). At the beginning of the experiment, we scaled the size of the virtual hand cursor to each subject's real hand size to provide matching familiar size and distance cues.

Trials were divided into three blocks, one for each technique. We randomized their order between subjects. Subjects were positioned standing in an upright posture in front of the tabletop (see Figure 3). Each block started with task descriptions presented via slides on the display surface. Thereafter, subjects completed between 10-15 training trials to minimize training effects.

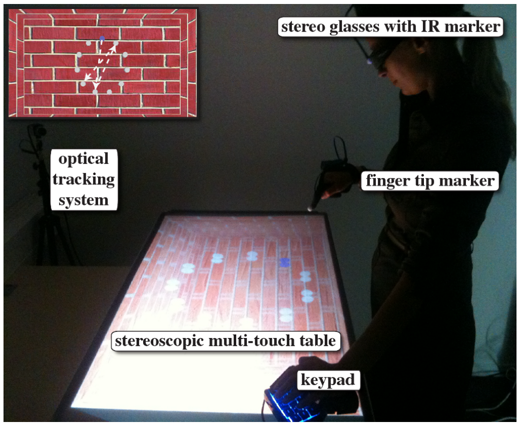


Figure 3: Photo of a subject during the experiment. The inset in the upper left corner illustrates the Fitts' Law selection task.

The subjects were instructed to select the target spheres as quickly and accurately as possible [14]. For this, subjects positioned the tip of the index finger of their dominant hand inside the 3D sphere for the direct input condition, did the same with the virtual cursor, or with the virtual hand's fingertip. All virtual objects were rendered using the standard OpenGL depth test. Subjects received visual feedback through a target turning green, when they correctly positioned the finger or cursor inside the target sphere. Subjects confirmed selections with the keypad with their non-dominant hand. When subjects confirmed a selection while the target sphere was not highlighted, we recorded this as a selection error, and advanced the trial state.

4 RESULTS

Since the results were normally distributed, they were analyzed using a repeated measure ANOVA and Tukey-Kramer multiple comparisons at the 5% significance level (with Bonferonni correction). We had to exclude four subjects from the analysis who showed strong signs of arm fatigue or significantly changed their selection behavior during the experiment.

4.1 Movement Time

Results for movement times are shown in Figure 2(a). We found a significant main effect of touch technique ($F(2,380)=21.70$, $p<.001$). Post-hoc test revealed that subjects required significantly ($p<.001$) less time when using direct input (M=978ms, SD=286ms) in comparison to the offset cursor (M=1115ms, SD=222ms) as well as the offset hand (M=1175ms, SD=262ms). Moreover, subjects required significantly less time ($p<.05$) when using the offset cursor compared to the offset hand.

4.2 Error Rate

Results for error rates are shown in Figure 2(b). We found a significant main effect of technique ($F(2,380)=15.96$, $p<.001$) and object height ($F(4,228)=4.53$, $p<.002$) on the percentage of missed targets. Post-hoc test revealed that subjects made significantly ($p<.001$) fewer errors when using the offset cursor (M=6.2%, SD=9.2%) compared to direct input (M=13.1%, SD=16.0%) and the offset hand (M=10.1%, SD=13.6%). Moreover, subjects made significantly fewer errors ($p<.05$) when using the offset hand compared to direct input. We found a two-way interaction trend between technique and height ($F(8,76)=1.85$, $p<.07$) on error rate. A post-hoc test revealed that subjects made significantly fewer errors when using direct touch for selecting objects displayed at heights of

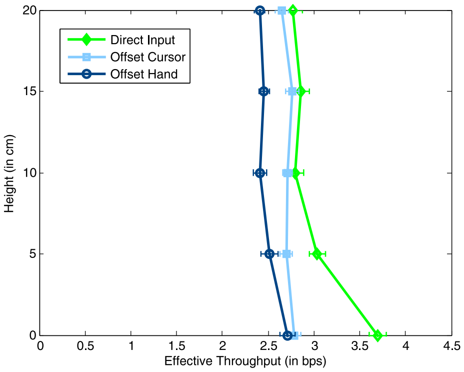


Figure 4: Results for the three techniques with target object height on the y-axis and effective throughput on the x-axis. Throughput combines errors and movement time. Error bars show the standard error.

0cm compared to objects at heights of 15cm ($p<.001$) and heights of 20cm ($p<.005$). For the other techniques we found no significant difference for different heights on error rate.

4.3 Error Distance

Results for error distances between the calibrated center of each sphere and touch positions during selection are shown in Figure 2(c). We found a significant main effect of technique ($F(2,380)$=10.66, $p<.001$) on error distance. Post-hoc test revealed that subjects made significantly ($p<.005$) less errors using the offset cursor (M=0.65cm, SD=0.14cm) compared to direct input (M=0.72cm, SD=0.28cm) ($p<.001$) and the offset hand (M=0.76cm, SD=0.39cm). We found a significant two-way interaction between technique and height ($F(8,76)$=3.51, $p<.002$) on error distance. A post-hoc test showed that subjects were significantly ($p<.001$) more precise when using direct touch for selecting objects displayed at heights of 0cm compared to objects at the other heights. For the offset-based techniques we found no significant difference for different heights on error distance.

4.4 Effective Throughput

Results for effective throughputs are shown in Figure 4. The throughput metric incorporates both speed and accuracy, with higher scores corresponding to better performance. We found a significant main effect of technique ($F(2,380)$=22.60, $p<.001$) and object height ($F(4,228)$=6.74, $p<.001$) on throughput. The average throughput during the experiment was M=3.02bps (SD=0.90bps) for direct input, M=2.72bps (SD=0.58bps) for the offset cursor, and M=2.50bps (SD=0.64bps) for the offset hand. We found a significant two-way interaction between technique and height ($F(8,76)$=2.29, $p<.03$). A post-hoc test revealed that subjects performed significantly ($p<.001$) better when using direct touch for selecting objects displayed at heights of 0cm compared to objects at the other heights. For the offset-based techniques we found no significant difference for different heights on throughput.

5 DISCUSSION AND CONCLUSION

In this paper we evaluated mid-air 3D selection performance within arm's reach for virtual scenes on tabletop setups and discussed the effects of visual conflicts in stereoscopic display. Our results provide interesting guidelines for the choice of input techniques in 3D stereoscopic tabletop setups. In our experiment, direct input provided the highest effective throughput for all tested target heights, suggesting that this technique should be the first choice when developing general-purpose 3D selection user interfaces in such tabletop setups. In this approach, selection times profit from the superimposed visual and motor spaces. The visual conflicts did not degrade performance to a level where offset-based approaches are a viable alternative. However, this guideline is based on the effective throughput metric with the inherent assumption that selection errors have only a moderate severity within the application domain. Based on our results and if precision of selections is the dominant aspect of a particular application, e. g., in cluttered virtual scenes, we suggest using the offset cursor technique. For the offset-based approaches we found significant higher precision and lower selection errors. Finally, although the virtual hand cursor in our experiment was scaled to the size of each subject's real hand to provide familiar size and distance cues (not available with the offset cursor technique), performance was generally reduced. This may be explained by limited visual and interaction fidelity and should be investigated in future work.

REFERENCES

[1] L.-W. Chan, H.-S. Kao, M. Y. Chen, M.-S. Lee, J. Hsu, and Y.-P. Hung. Touching the void: Direct-touch interaction for intangible displays. In *Proc. of ACM CHI*, pages 2625–2634, 2010.

[2] J. P. Djajadiningrat. *Cubby: What You See is Where You Act. Interlacing the display and manipulation spaces*. PhD thesis, Delft University of Technology, 1998.

[3] J. Ferwerda. Psychophysics 101: How to run perception experiments in computer graphics. In *SIGGRAPH Core*, 2008.

[4] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psych.*, 47:381–391, 1954.

[5] R. T. Held, E. A. Cooper, J. F. O'Brien, and M. S. Banks. Using blur to affect perceived distance and size. *ACM TOG*, 29(2):1–16, 2010.

[6] International Organization for Standardization. *ISO/DIS 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices*, 2000.

[7] G. Liu, R. Chua, and J. T. Enns. Attention for perception and action: task interference for action planning, but not for online control. *Exp. Brain Res.*, 185:709–717, 2008.

[8] J. M. Loomis and J. M. Knapp. Visual perception of egocentric distance in real and virtual environments. In L. J. Hettinger and M. W. Haas, editors, *Virtual and adaptive environments*, volume Virtual and adaptive environments. Mahwah, 2003.

[9] C. L. MacKenzie, R. G. Marteniuka, C. Dugasa, D. Liskea, and B. Eickmeiera. Three-dimensional movement trajectories in fitts' task: Implications for control. *Q.J. Exp. Psychology-A*, 39(4):629–647, 1987.

[10] I. S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. In *Proc. of ACM CHI*, pages 1633–1636, 2008.

[11] M. R. Mine, F. P. Brooks, Jr., and C. H. Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proc. of ACM SIGGRAPH*, pages 19–26, 1997.

[12] A. Paljic, S. Coquillart, and J.-M. Burkhardt. A study of distance of manipulation on the responsive workbench. In *Proc. of IPT*, pages 1–8, 2002.

[13] I. Poupyrev, S. Weghorst, M. Billinghurst, and T. Ichikawa. Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. *Computer Graphics Forum*, 17(3):41–52, 1998.

[14] R. J. Teather and W. Stürzlinger. Cursors for 3D pointing. In *Proc. of ACM 3DCHI Workshop*, pages 5–12, 2012.

[15] W. Thompson, R. Fleming, S. Creem-Regehr, and J. K. Stefanucci. *Visual Perception from a Computer Graphics Perspective*. A. K. Peters / CRC Press, 2011.

[16] D. Valkov, F. Steinicke, G. Bruder, and K. H. Hinrichs. 2D Touching of 3D Stereoscopic Objects. In *Proc. of ACM CHI*, pages 1353–1362, 2011.

[17] Y. Wang and C. MacKenzie. Effects of orientation disparity between haptic and graphic displays of objects in virtual environments. In *Proc. of INTERACT*, pages 391–398, 1999.

[18] Y. Yuan and A. Steed. Is the Rubber Hand Illusion induced by immersive virtual reality? In *Proc. of IEEE VR*, pages 95–102, 2010.