Pointing at 3D Targets in a Stereo Head-Tracked Virtual Environment

Robert J. Teather* Wolfgang Stuerzlinger†

Dept. of Computer Science & Engineering
York University, Canada

ABSTRACT

We present three experiments that systematically examine pointing tasks in fish tank VR using the ISO 9241-9 standard. All experiments used a tracked stylus for a both direct touch and ray-based technique. Mouse-based techniques were also studied. Our goal was to investigate means of comparing 2D and 3D pointing techniques. The first experiment used a 2D task constrained to the display surface, allowing direct validation against other 2D studies. The second experiment used targets stereoscopically presented above and parallel to the display, i.e., the same task, but without tactile feedback afforded by the screen. The third experiment used targets varying in all three dimensions. Results of these studies suggest that the conventional 2D formulation of Fitts' law works well for planar pointing tasks even without tactile feedback, and with stereo display. Fully 3D motions using the ray and mouse based techniques are less well modeled.

KEYWORDS: Pointing, Fitts' law, ISO 9241-9, virtual reality.

INDEX TERMS: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – virtual reality. H.5.2 [Information Interfaces and Presentation]: User Interfaces – input devices, interaction styles.

1 INTRODUCTION

Rapid aimed target pointing is a fundamental task in user interfaces. It is the basis of direct manipulation, and required for selecting objects and activating subsequent operations. The WIMP interface paradigm (Windows, Icons, Menus and Pointing device) popular in desktop computing exemplifies this, as many operations are accessible by pointing at interface widgets. Pointing tasks in 2D interfaces have received a great deal of attention from human-computer interaction researchers, and are well understood and modeled by Fitts' law [6, 14].

Three-dimensional pointing, or target selection, is less well understood. The pointing task itself is more complex, as the cursor is usually controlled by *at least* 3 degrees of freedom (DOF), i.e., the position in the *x*, *y*, and *z* directions. Some techniques (e.g., ray-casting) also require the control of the pointing device orientation, for up to 6 DOF's. In contrast, 2D pointing requires only control of 2 DOF's – position in the *x* and *y* directions.

Several researchers have developed 3D extensions to Fitts' law by adding extra terms to the model, which (potentially artificially) improve its predictive capabilities. These attempts also use subtly different performance measures so it is hard to directly compare results across studies. This raises another question: Is 3D pointing kinematically different from 2D tasks? Or can technical issues, such as input device technology, explain the difference?

We present three studies on 3D pointing in a fish tank virtual environment using a 3D version of the ISO 9241-9 task [9]. ISO 9241-9 describes a tapping task, based on Fitts' law and is extensively used for evaluating 2D pointing devices, such as mice, pens/styli, and touch screens. It characterizes device performance using throughput, a single measure that combines speed and accuracy and thus accounts for different user strategies and device characteristics, enabling direct comparison between studies [14].

* email: rteather@cse.yorku.ca
† email: wolfgang@cse.yorku.ca

1.1 Motivation

ISO 9241-9 is widely used in 2D pointing research as it allows *direct* comparison *between* studies. There is currently no such standard for 3D interfaces. Using a standard highlights the benefits (and pitfalls) of 3D technology with consistency and may enable direct comparison with 2D devices. Our main goal was to determine how well the standard evaluates 3D pointing. We first examined situations where 2D and 3D pointing tasks are directly comparable. Our first study used planar movements between targets displayed at the screen surface, where one would expect that stereo would have no effect. The second study used the same task, but target circles were stereoscopically displayed at different heights parallel to the display. This was intended to determine which pointing techniques were affected by stereo display or tactile feedback. The third study used movement between targets at different heights, allowing us to investigate the difference between 2D and 3D motions in isolation from other factors.

Our second goal was to test exemplary 2D and 3D pointing techniques using a well-understood method. The mouse was included as a benchmark. According to numerous ISO 9241-9 studies its pointing throughput is around 3.5–4.5 bits per second [20]. Our data matches this, which validates our results and methodology against other work. We chose techniques based on ray casting and the "virtual hand" metaphor (using a tracked stylus) as archetypical 3D pointing techniques. Our reasons for selecting these can be found in Section 3. Although most of the *relative* differences of the examined factors/techniques are established by previous work, we present the first standard-based comparison that enables us to characterize the differences in a more *absolute* sense, and in comparison with 2D pointing.

2 POINTING TASKS

Pointing has received attention from the VR community [2-5, 7, 11, 13, 17, 19, 21, 22, 24], but is not characterized as well as in 2D user interfaces. Most VR object selection/manipulation techniques follow two paradigms: ray/occlusion-based techniques or virtual hand/volume metaphors. Ray techniques cast a virtual ray from the user's hand, finger, or cursor and enable selection/manipulation of objects hit by the ray. Virtual hand metaphors require intersection of the 3D position of the hand or cursor with objects. Both paradigms can potentially be modeled by Fitts' law. Our virtual "hand" technique uses a tracked stylus as a depth cursor to approximate of the actual hand position. This corresponds directly to the task that Fitts used originally [6, 14].

2.1 Fitts' Law

Fitts' law [6] predicts rapid aimed movements and is given by:

$$MT = a + b \cdot ID, \quad \text{where} \quad ID = \log_2\left(\frac{A}{W}+1\right) \qquad (1)$$

MT is movement time, and *a* and *b* are empirically determined constants for a given pointing technique. *ID* is the index of difficulty (in bits). A is movement amplitude (distance to the target), and W is the target width. ID represents the overall task difficulty based on the distance to, and size of, the target. Hence, smaller farther targets are harder to hit than closer larger targets.

Fitts' law can be used as a predictive model via linear regression of measured movement times onto ID. Although individual Fitts' pointing tasks are simple, most complex pointing tasks are made up of multiple such motions. Thus, Fitts' law is a basis for the prediction of complex tasks as well.

2.2 ISO 9241-9

ISO 9241-9 [9] employs a standardized pointing task based on Fitts' law. The standard uses *throughput* as a primary measure [6]. Throughput (TP) is defined in bits per second as:

$$TP = \frac{\log_2\left(\frac{A_e}{W_e}+1\right)}{MT}, \quad \text{where} \quad W_e = 4.133 \cdot SD_x \qquad (2)$$

The log term is the *effective* index of difficulty, ID_e, and MT is the measured average movement time for a given condition. The formulation for ID_e is similar to ID in equation (1), but uses the *effective* width and amplitude in place of W and A. This accounts for the task users *actually* performed, as opposed to the task they were presented [14]. SD_x is the standard deviation of the over/under-shoot to the target, projected onto the task axis (the vector between subsequent targets) for a given condition. This assumes that movement endpoints are normally distributed around the target centre and 4.133 (±2.066) standard deviations (i.e., 96%) of clicks hit the target [10]. W_e corrects the miss rate to 4%, allowing comparison between studies with differing error rates [14]. A_e is the average movement distance for a given condition.

Throughput incorporates speed and accuracy into a single measure, and is unaffected by speed-accuracy trade-offs [15][1]. For example, compare a user who works quickly, but misses many targets, with a highly precise user who always hits the target – the second is effectively performing a more difficult task. Alternatively, if every hit is just outside a target, the user is effectively hitting a slightly larger target. Effective measures are computed across both hits and misses to better account for real user behavior, and thus enable more meaningful comparison. Effective measures may also make throughput less sensitive to device characteristics (e.g., device noise). This is desirable in cross-device comparisons.

2.3 Fitts' law Extensions to 3D

Fitts' law was developed for one-dimensional aimed motions but works extremely well for 2D motions [14]. Straightforward extensions to 3D pointing tend not to work as well. Note that adding *any* extra parameter in a regression analysis will *always* improve the correlation [18]. Thus, it is unclear if extra factors improve the model's predictive capabilities in an appropriate way.

For example, Murata and Iwase [16] studied pointing tasks in a 2D vertically oriented plane. This was not a full 3D task as it did not involve hitting targets at varying depths. They derived a directional ID model that incorporates the movement angle to the target. The authors report a higher correlation between MT and their new ID model. They also found that upward movements took longer than downward movements, possibly due to gravity.

Grossman *et al.* [7] used a volumetric display to investigate pointing at 3D targets that varied in height, width, and depth. Their model used the movement direction through the target. However, all targets were positioned on a plane parallel to the ground, effectively a 2D task. Their results may also not extend to other VR systems as volumetric displays provide different (more accurate) depth cues compared to stereo systems. Their 3D model is also inconsistent with the 2D model used by the ISO standard, preventing direct comparison between 2D and 3D pointing.

[1] Alternative proposals to address the speed-accuracy trade-off have no clear advantage over throughput, which is currently the most frequently used measure.

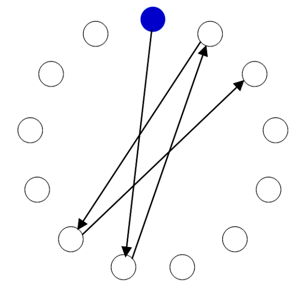

Figure 1. ISO 9241-9 reciprocal tapping task with thirteen targets. Participants click the highlighted target, starting with the top-most one. Targets highlight in the pattern indicated by the arrows.

3 Pointing/Selection Techniques

Many 3D pointing/selection techniques are used in VR systems. These are usually classified roughly into two categories: virtual hand (depth cursor) and ray-based techniques. We implemented several such techniques for our study:

Pen Touch (PT)

This technique uses a tracked stylus and displays a 1 mm diameter cursor (the "virtual pen tip") co-located with the stylus tip. The virtual tip is tested for target intersections. This is representative of depth cursor techniques, or those that require intersection of the hand (representation) with targets [3, 4]. We chose not to use the user's actual hand to ensure consistency with other work that used actual input devices. This technique also simulates pen-based systems [12], if used on a 2D display surface. Hence results can also be compared directly to known results for the stylus in 2D [20]. The technique is also sensitive to the effects of the tactile feedback afforded by the screen, i.e., when performing a pseudo-3D task first *at* and then *above* the display surface.

Pen Ray (PR)

This technique also uses the stylus, but casts a ray from its tip into the scene. The ray visually extends from the stylus tip, and a small sphere is displayed where it intersects the scene. The effect is similar to a laser pointer. This technique is representative of ray-casting techniques commonly used in 3D user interfaces [3, 19], and in game consoles such as the Nintendo *Wii*. It may require up to 6DOF for control, hence we expect it be slower than lower-DOF techniques based on previous work [3, 19, 23]. We believe it will also have worse pointing throughput. We intend to establish a throughput score for this technique through our work.

Mouse Cursor (MC)

This technique uses a mouse controlled 3D cursor that moves in the screen plane. A ray from the dominant eye is cast through the cursor position to determine which target is hit. This technique represents the system cursor used in "non-VR" 3D graphics software, such as games and CAD, and also allows comparison with 2D work [14, 20, 21]. The cursor was displayed as a 1 mm sphere stereoscopically displayed at the screen surface to ensure visual consistency across techniques. Naturally, this technique works only for targets presented in the plane of the display (i.e., at 0 cm "height") as higher targets occlude the cursor.

Floating Cursor (FC)

Both the FC and SC techniques are designed to address the mouse cursor occlusion issue for targets above the display. Instead of

Technique	*Code*	*Control Traits*	*Feedback displayed*
Pen Touch	PT	Direct touch with tip of tracked stylus	1 mm sphere co-located with physical pen-tip
Pen Ray	PR	Ray pointing with tracked stylus	Ray emitted from pen-tip, 1 mm dot at scene intersect
Mouse Cursor	MC	Mouse controlled cursor in screen plane	1 mm sphere rendered in place of system cursor at 0 cm height
Floating Cursor	FC	Mouse-controlled cursor, 8.5 cm from screen	1 mm sphere rendered to one eye, in front of targets
Sliding Cursor	SC	Mouse/head-controlled cursor that slides over scene	1 mm sphere displayed at closest ray/scene intersection

Table 1. Summary of pointing/selection techniques used in the studies.

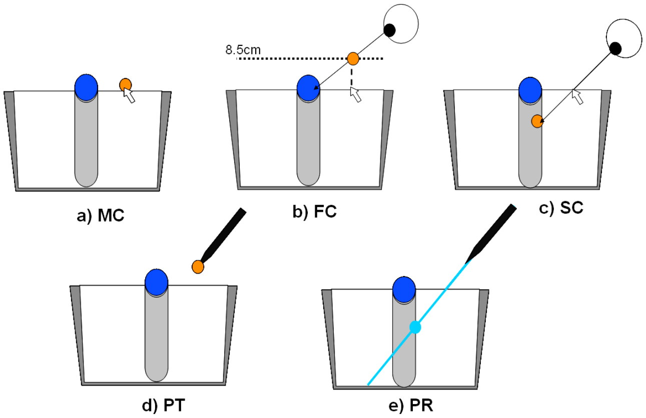

Figure 2. The box represents the scene, and the cylinder represents a target. The system mouse cursor was *never* displayed and is shown only for reference. (a) Mouse cursor, MC. (b) Floating cursor, FC: one-eyed cursor above the system cursor position. (c) Sliding cursor displayed a 3D cursor where the mouse ray intersected the scene, SC. (d) Pen touch, PT. (e) Pen ray, PR.

displaying a cursor in the screen plane, the mouse-controlled FC "floats" in a plane parallel to the display slightly above the "highest" targets. In our experiments, this was 8.5 cm above the screen. This floating cursor is rendered only to the dominant eye, hence stereo depth discrimination is impossible. Previous work found that such a cursor outperformed 3D cursors (like our PT) in a Fitts' study [23] that did not use the ISO standard. We include this technique to compare both against the mouse cursor (MC) and against other 3D pointing techniques.

Sliding Cursor (SC)

The SC technique represents the "depth cursors" sometimes used in games and mouse-controlled 3D graphics systems [1]. SC uses the position of both the system cursor and the eye to compute the position of a 3D cursor in the scene. A ray is cast from the eye through the system cursor (which is not displayed) and the 3D cursor is displayed where the ray intersects the scene. Thus, the cursor slides along the visible scene geometry, and enables 3DOF cursor control with only 2DOF input. This technique is a compromise between 2D and 3D pointing techniques and has not been evaluated previously in a classic pointing experiment.

The visualizations (ray/cursor) were rendered in stereo for all techniques, except the one-eyed floating cursor. The MC condition was included as a benchmark and for external validation, and was thus expected to perform in the range of 3.5 – 4.5 bps [20]. This also enables us to rank the techniques in a 2D task, a constrained 3D task, and finally in a full 3D task.

The FC and MC techniques require only 2DOF control, while the SC and pen techniques require between 2 and 6DOF, due to the tracker (e.g., partial head control for SC). The PT technique requires the user to control all 3 translations simultaneously, whereas all other techniques require only accuracy in 2 dimensions (rotations for PR). Although this makes the PT pointing task somewhat different, there is no way to decouple this within our setup. For the mouse and ray based techniques, there is no way to specify target depth to select occluded targets as both select always the closest target. We do not see this as a limitation, as head-tracking makes it easy to see an occluded target by moving one's head to and then select it. Moreover, due to our target layout, occlusion rarely, if ever, occurred in our experiment.

3.1 System Design Issues

We considered several additional factors. First, we used round targets to ensure there was only a single W (size) parameter. Disks or spheres are both reasonable choices for 3D targets. A pilot study revealed no significant difference for different target types. We decided to use spherical targets as the more natural 3D extension of 2D circles. Target spheres were centered at the midpoint of the top of cylinders, instead of floating in space, as most participants in a pilot found it difficult to determine the actual target depth using only stereo depth cues.

We mounted a button on the side of the stylus as the mean to indicate selection. Otherwise, trials could only end upon successful intersection with the target, i.e., it would be impossible to miss targets. This would drastically influence the selection time distribution and produce implausible throughput scores. We opted for this design as all alternatives, such as using the non-dominant hand, would complicate the task or introduce issues with bimanual task division. However, this button introduces the potential for "wiggle" in the pen tip upon the button press. We did not observe this issue in our setup, likely due to the button placement.

Finally, ray techniques permit several choices of the "cursor" position for distance calculations, namely, the scene intersection point, the target plane intersection, or the closest ray point to the target. For all ray-based techniques, we used the closest ray point to the target for computing effective width. For the cursor-based techniques, we used the cursor position. These distances are always projected onto the task axis for the W_e computation.

4 Methodology

Here we describe the general apparatus and procedure used in all three experiments. Details for each experiment are outlined later.

4.1 Participants

There were 12 different paid participants in each experiment. For Exp. 1, 8 were male (aged 22 to 28, mean 25 years), for Exp. 2, 7 male (aged 20 to 29, mean 24 years), and for Exp. 3, 7 male (aged 21 to 32, mean 27 years). All were students, and were recruited via posters or a web page. Most were non-technical students with little gaming experience. We targeted this population as experience suggests that a technical background or gaming experience can result in unusually high performance levels. All had normal or corrected vision, and could see stereo imagery. This was screened by asking them to measure the height of a stereo target displayed target with a ruler. All were right-handed.

4.2 Apparatus

All studies were conducted on a 3 GHz PC with an NVidia QuadroFX 3400 graphics card. We used a 22" CRT monitor at 800 x 600 resolution and a 120 Hz refresh rate. The stereo display used Stereographics *CrystalEyes* shutter glasses and emitter. A NaturalPoint *OptiTrack* system with five cameras was used for tracking both the head and a 12 cm long stylus. It had a single button, connected to the computer via a re-engineered USB mouse. We positioned the display on its back, rather than upright. The whole setup can be seen in Figure 3.

The equipment was carefully calibrated to (approximately) 1 mm accuracy. Tracker noise was around 1 mm RMS. We also verified that the update rate of the system was indeed 120 Hz and measured the mean end-to-end tracker latency at 63 ms. Mouse latency was around 35 ms. The apparatus was calibrated such that displayed objects could be accurately measured with a physical object, e.g., a ruler. Thus it was possible to line up physical features (e.g., the real pen tip) with features on the image.

Software

The software presented the inside of a 10 cm deep box matching the display size. Target cylinders were arranged along the circumference of a circle at the bottom of the box (Figure 3, top right). The cylinders and box were textured to enhance depth perception. Target spheres were displayed centered at the midpoints of the cylinder tops and appeared at varying heights at or above the screen surface. Targets highlighted red when intersected by the cursor to improve feedback. In Exp. 1, the mono display mode rendered the same image for both eyes; all other experiments used stereo display exclusively. We used quad-buffering and off-axis frustums to provide stereoscopic head-coupled display. A 1 mm diameter sphere depicted the 3D cursor. This was co-located with the tip of the physical pen in the PT technique. The cursor was rendered semi-transparent to provide clear feedback when it intersected a target [24].

Target diameters and distances were always identical within a single circle of targets. However, cylinder diameters and distance varied *between* target circles. For Exp. 1 and 2, all cylinders in a given trial round were of equal height. and cylinder heights varied between target circles in Exp. 2. In Exp. 3, heights varied between individual targets within a target circle. Target height was measured from the surface of the screen. We refer the reader to the motivation section for the reasons for these design choices.

The software logged movement times and if targets were hit or missed. Cursor and head movements were also logged.

4.3 Procedure

Participants were seated at the display and given 24 to 36 practice trials with each technique to familiarize them with the system, the task, and the techniques. Participants were instructed to "click" each blue highlighted target as quickly and accurately as possible – a standard instruction for Fitts' law experiments, designed to lower result variability. With the PT technique, this meant intersecting the (virtual) pen tip with the target. With the PR technique, they had to intersect the ray with the target. The mouse-based techniques required moving the corresponding cursor to/over the target. Pressing the device button indicated selection and ended the trial. The next target would then highlight regardless if the previous one was hit or missed. The target sequence is shown in Figure 1. Timing started after the first click in each target circle, allowing participants to take breaks as necessary. Participants wore shutter glasses in all conditions.

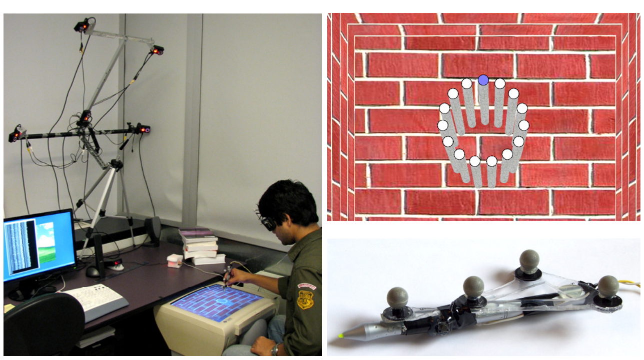

Figure 3 (Left). Experimental setup. (Bottom Right) The tracked stylus. (Top Right) Stereoscopically displayed scene with the target cylinder that extend to or above the screen surface.

4.4 Experiment 1

This first study was intended to establish baseline 2D throughput scores for both the pen- and mouse-based techniques. Targets were only displayed at the screen surface, i.e., at a height of 0 cm.

4.4.1 Design

The experiment used the following four independent variables:

Target Distance: 5, 9, or 18 cm.
Target Size: 0.65, 0.85, or 1.05 cm.
Stereo: on or off.
Technique: PT, PR, MC or FC.

The experiment used a 3×3×2×4 within-subjects design. Target distance and size were chosen randomly (without replacement) for each target circle. Technique was counterbalanced according to a balanced Latin square. Each target circle contained 13 targets, yielding 12 recorded target clicks per circle. Thus, for all 12 participants, a total of 10368 trials were logged. A total of 95 trials (~1%) were dropped as outliers (scores more than three standard deviations from the mean), leaving 10273 recorded trials. The nine combinations of size and distance resulted in nine unique *ID*s ranging from 2.5 to 4.8 bits. The dependent variables were movement time (ms), error rate (missed target percentage), and throughput (bits/second), calculated with Equation (2).

4.4.2 Results

Movement Time

Movement time is the average time (in milliseconds) required to hit targets in a given condition. There was a significant effect for technique on movement time ($F_{3,11} = 60.7$, $p < .0001$). A post-hoc Tukey-Kramer test revealed that PR was significantly slower than the other techniques, ($p < 0.05$), see Figure 4. Stereo also had a significant effect ($F_{1,11} = 10.3$, $p < .01$) on movement time, and slightly decreased movement time on average. There was also a significant interaction effect between technique and stereo ($F_{3,33} = 8.4$, $p < .001$). The pen techniques benefitted more than the mouse. In particular, PR was about 25% faster with stereo. See Figure 4 for an illustration of the results.

Figure 5 shows the regression of movement time on *ID*. Most techniques (especially the mouse techniques, as expected) show fairly high positive correlations between movement time and *ID*.

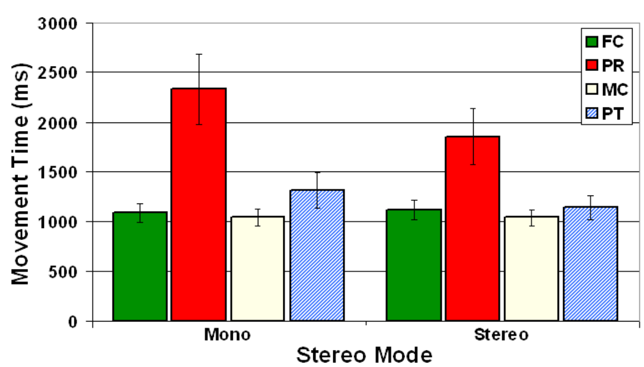

Figure 4. Exp. 1. movement times by technique and stereo mode. Error bars indicate ±1standard error.

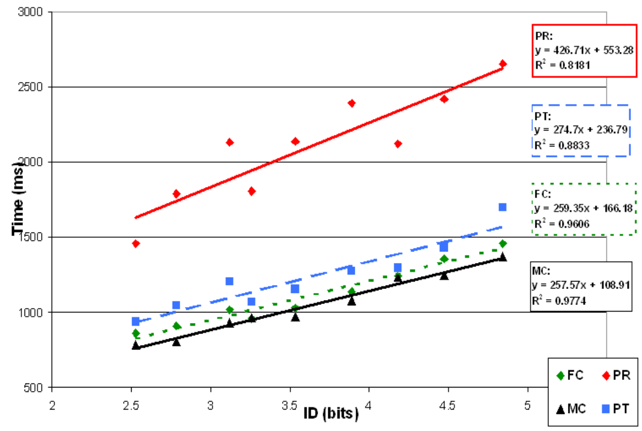

Figure 5. Regression of movement time onto ID for Exp. 1.

Error Rate

Technique had a significant effect on error rate ($F_{3,11} = 11.9$, $p < .0001$). Posthoc analysis indicated that PR had significantly more errors than any of the other three techniques. The error rates were 2.8% for MC, 4.3% for FC, 8.1% for PT, and 13.6% for PR.

Throughput

There was a significant main effect for technique on throughput ($F_{3,11} = 65.4$, $p < .0001$). Consistent with previous work in 2D pointing [20], throughput for the mouse cursor technique was 3.81 bps (*SD* 0.76). Throughput scores are shown in Figure 6. Stereo did not have a significant effect on throughput ($F_{1,11} = 3.66$, $p > .05$), nor was there a significant interaction effect between technique and stereo ($F_{3,33} = 0.77$, ns). The absence of these effects is likely because longer movement times were made up for in accuracy. Even if the raw error rate was higher, low magnitude misses have little impact on $W_{e,}$ and hence throughput.

4.5 Experiment 2

This experiment used targets that were stereoscopically presented at or above the display surface. Target heights were fixed within a target circle, but varied between circles. We included a 0 cm condition to enable comparison with Exp. 1 results. Initially, we also considered 2 cm target heights, but eliminated this condition after pilot study participants were observed resting their hands on the screen surface with PT for such targets. This distorted the data as only the 0 cm height condition should afford tactile feedback. This “cheating” was impossible at 4 cm and higher.

We were also interested in vergence and accommodation conflicts. Our eyes converge (turn inward) so that their gaze crosses at the depth of stimuli. The lenses accommodate (change shape) to focus on the depth of stimuli. Most stereo displays do not replicate this; while the eyes do converge on the perceived object depth, they accommodate to the screen surface, regardless of actual depth. This cue mismatch has negative effects in depth perception experiments [8], and may also affect pointing tasks.

4.5.1 Design

The experiment used the following independent variables:

Target Distance: 5, 9, or 18 cm.
Target Size: 0.65, 0.85, or 1.05 cm.
Target Height: 0, 4, 6 or 8 cm above the display
Technique: PT, PR, FC or SC

The experiment used a 3×3×4×4 within-subjects design. Target distance, diameter and height were chosen randomly (without replacement) for each target circle. Technique was counterbalanced according to a balanced Latin square. Each target circle contained 13 targets, yielding 12 recorded target clicks per circle. Over all 12 participants, 20736 trials were logged. A total of 284 trials (~1.2%) were dropped as outliers, leaving 20452 recorded trials. The same set of *ID*s as in Exp. 1 was used. The dependent variables were movement time (ms), error rate (missed target percentage), and throughput (bits per second).

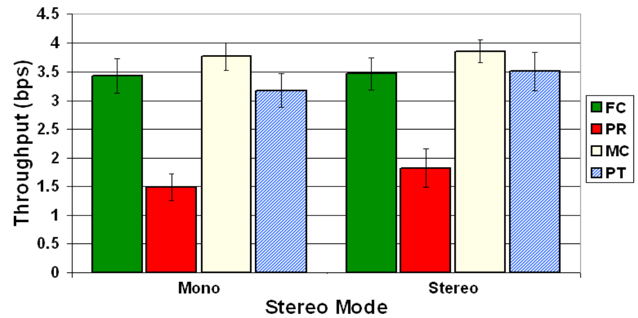

Figure 6. Exp. 1 throughput results by technique and stereo mode. Error bars indicate ±1standard error.

4.5.2 Results

Movement Time

There were significant main effects for technique ($F_{3,11} = 23.5$, $p < .0001$) and target height ($F_{3,11} = 21.5$, $p < .0001$) on movement time. Movement time generally increased with height. This is especially evident in the significant interaction between technique and height ($F_{9,33} = 7.78$, $p < .0001$), see Figure 7. The regression of *MT* on *ID* is shown in Figure 8 and indicates strong correlations, especially for the PR and FC techniques.

Error Rate

Technique had a significant effect on error rate ($F_{3,11} = 18.4$, $p < .0001$), while height alone did not ($F_{3,11} = 0.38$, ns). The mean error rate for each condition was 5.1% for FC, 8.5% for SC, 13.5 for PT, and 16.8% for PR. There was a significant interaction between height and technique ($F_{3,33} = 9.73$, p < .0001). PT had far fewer errors when the targets were at the screen surface, perhaps due to tactile feedback or the absence of stereo conflicts.

Throughput

There was a significant main effect for technique on throughput ($F_{1,11} = 51.0$, $p < .0001$). A Tukey-Kramer test revealed that throughput for the floating cursor was significantly higher than all others, followed by the sliding cursor, and then both pen-based modes in a cluster ($p < .05$), see Figure 9. The FC throughput was 3.65 bps, consistent with 2D mouse pointing throughput scores. There was a significant interaction between technique and target

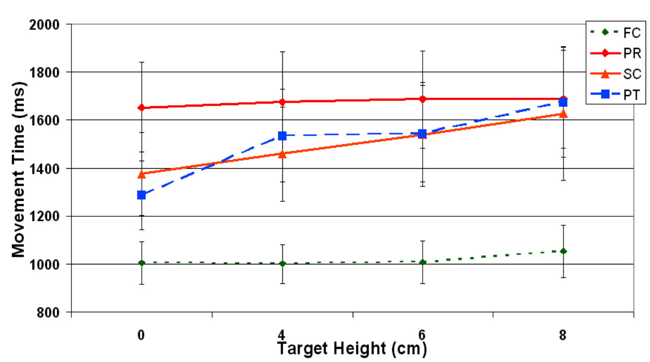

Figure 7. Exp. 2 movement times by target height and technique. Error bars indicate ±1standard error.

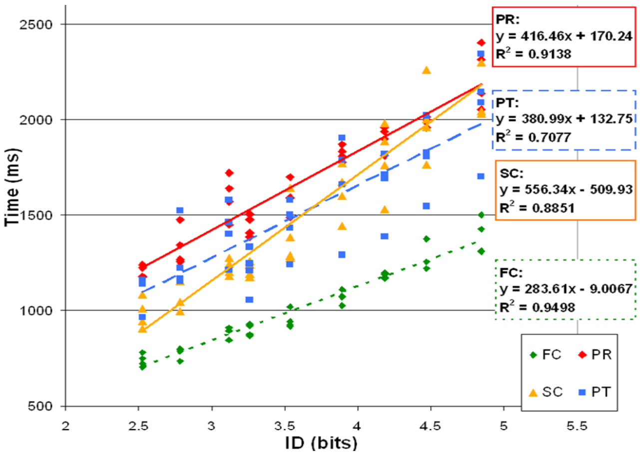

Figure 8. Regression of movement time onto *ID* for Exp. 2. Each *ID* includes one data point for each of the four height conditions.

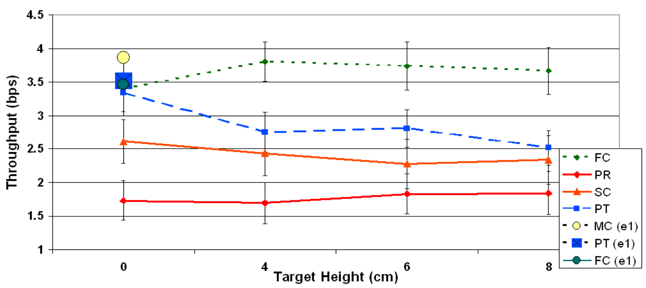

Figure 9. Exp. 2 throughput scores. Error bars indicate ±1standard error. Exp. 1 MC, PT and FC scores are included for reference.

height ($F_{9,33} = 5.56$, $p < .0001$). Throughput for PT was significantly higher for targets *at* the screen surface (3.3 bps, in line with Exp. 1). Again, this is likely due to tactile feedback or the absence of stereo cue conflicts. Figure 9 shows throughput for MC, PT and FC from Exp. 1, as single datapoints for the 0 cm target height. Although MC was excluded from Exp. 2, these are provided for reference. PT and FC performed nearly identically in both studies at 0 cm height. Performance of FC was comparable to MC (within 1 *SE*) in Exp. 1, regardless of target height.

4.6 Experiment 3

Experiment 3 also presented targets above the display. Unlike Exp. 2, target heights varied *within* target circles, necessitating true 3D motions. We kept a circular target arrangement to keep the task simple. The alternative of a spherical arrangement would involve larger numbers of targets and increase the likelihood of confounding target occlusion issues. Hence and also to enable cross-experiment comparisons, we used the same planar task, but angled the plane so that it was no longer parallel to the display. This target arrangement then requires true depth motions.

4.6.1 Apparatus

To set target heights, one target was randomly set to 8 cm, and the following target was set to 0 cm. All other target heights were set according to a sinusoid between these extremes, effectively arranging targets in a randomly oriented "diagonal" plane.

4.6.2 Design

The experiment used the following independent variables:

Target Distance: 5, 9, or 18 cm between cylinder centres.
Target Size: 0.65, 0.85, or 1.05 cm.
Technique: PT, PR, FC or SC

Overall, the experiment used a 3×3×4×4 within-subjects design. Target distance and diameter were chosen randomly (without replacement) for each target circle. Technique was counterbalanced according to a balanced Latin square. Each target circle contained 13 targets, yielding 12 recorded target clicks per circle. For all 12 participants, a total of 20736 trials were logged. A total of 279 trials (~1.3%) were dropped as outliers, leaving 20457 trials for the analysis.

Target distance indicates the distance between the cylinders in the circle. Here, the Euclidean 3D distance between consecutive target spheres was used in the computation of *ID*, which is a more accurate representation of the task performed by participants. Hence, the range of *ID*s is slightly larger than in Exp. 1 and 2, and there are also a larger number of intermediate *ID* values.

The dependent variables were movement time (ms), error rate (missed target percentage), and throughput (bits per second).

4.6.3 Results

Results were analyzed by height difference between consecutive targets. These were binned into 9 groups: ±8, ±6, ±4, ±2 and 0 cm height difference. A negative height difference indicates that the second target was lower than the first, i.e., required movement *into* the scene. The 0 cm bin contains movements (approximately) parallel to the display surface. This is comparable to the 4 cm conditions in Exp. 2.

Movement Time
Both technique ($F_{3,11} = 25.5$, $p < .0001$) and height difference ($F_{3,11} = 8.1$, $p < .0001$) had significant main effects on movement time, see Figure 10. A Tukey-Kramer test revealed that FC was significantly faster than all other techniques. Also, pointing tasks with small depth components (i.e., 0 to 4 cm difference) were significantly faster than those with larger depth components. The regression of movement time onto *ID* is presented in Figure 11. FC still shows a strong correlation between movement time and *ID*. The other techniques (especially PT) all show lower correlations than in the first two experiments.

Error Rate
The error rates were 7.1% for FC, 12.7% for SC, 19.9% for PT and 20.6% for PR. There was a significant main effect for technique on error rate ($F_{3,11} = 11.9$, $p < .0001$). The higher error rates likely reflect increased task difficulty due to varying target heights. There was a significant interaction between technique and height difference ($F_{3,11} = 4.31$, $p < .0001$), with the PT technique becoming significantly worse for larger height differences, while the rest remained fairly consistent.

Throughput
There was a significant main effect for technique on throughput ($F_{3,11} = 20.6$, $p < .0001$). The throughput scores are summarized for each technique in Figure 12. Average throughput scores for 4 cm height for each technique from Exp. 2 are included here for comparison. These scores are quite close to those found in Exp. 3.

5 Discussion

Throughput and Technique Performance
The consistency of the throughput scores reported in our experiments illustrates what is arguably the greatest strength of this measure. It enables direct comparison between our studies, despite varying error rates. Throughput for the mouse techniques was consistent across all three studies. Moreover, the throughput for the mouse cursor in Exp. 1 is consistent with other reports [20]. Similar conditions between experiments showed highly similar throughput scores. Overall, we take this as both internal and external validation of the experimental methodology.

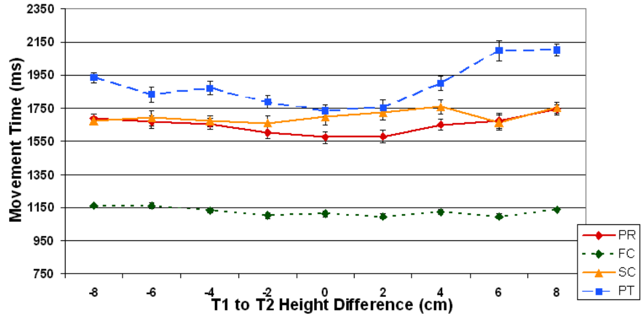

Figure 10. Movement time by technique and target height difference for Exp. 3. Error bars indicate ±1standard error.

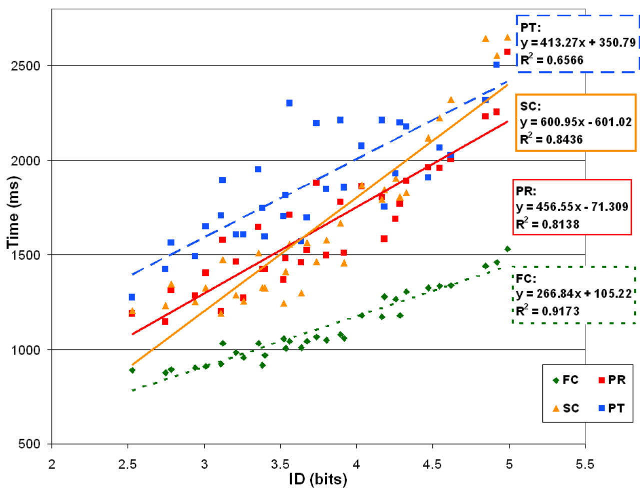

Figure 11. Regression of *MT* onto *ID* for Exp. 3. Note that there are additional *ID* datapoints as the target distance varied more.

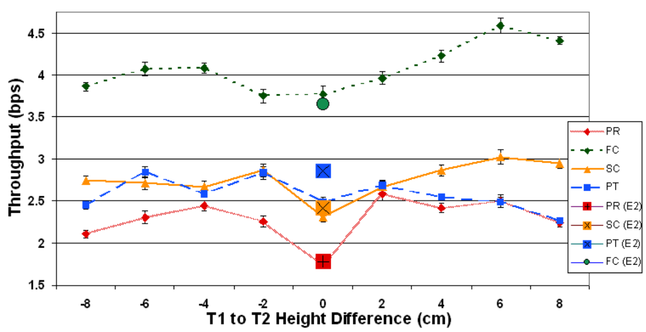

Figure 12. Exp.3 throughput by technique and height difference. Error bars show ±1SE. Marks for throughput scores from Exp. 2 are included for reference (for the 4 cm height difference only).

The natural progression of experiments demonstrated several important results. First, mouse-based pointing throughput is generally unaffected by target depth, regardless of stereo cue conflicts, *if* all targets are at the same depth. If target depths vary there may be an effect due to perspective, which alters the perceived target size. Higher targets are closer to the viewer, and thus appear (slightly, ~10%) larger. This may make them easier to hit. Conversely, downward motions fare worse, as lower targets appear smaller. Since FC is a 2DOF technique, the “correct” target size would be its 2D projection as seen from the current eye position. However, we were unable to verify this, as we did not observe sufficiently strong trends in the data for Exp. 3 for target height differences. We suspect that to be due to the limited depth range (max. 8 cm). This issue affects *only* mouse-based techniques, as hit testing for the PT and PR techniques is performed in 3D motor space. Using an average head position, we computed an estimate of the visual target size at both the screen surface and the maximum height, and computed the *IDs* for the perceived and actual size. Ultimately, for same size targets presented at the maximum height difference, task difficulty decreases between 6 and 11% (smaller targets are more affected). As we see effects that are substantially larger than this, we believe that perceived size is not a *sufficient* explanation.

The FC technique with 2D input outperformed all 3D techniques in Exp. 2 and 3, and had performance comparable to the mouse cursor in Exp. 1. A partial explanation is the influence of the number of DOF on pointing task difficulty [23]. We speculate that stereo cue conflicts may have impacted performance of the 3D-based pointing techniques, as did the absence of tactile feedback in Exp. 2 and 3. We do not believe that user familiarity with the mouse was a notable factor, as the comparatively unfamiliar PT technique performed *as well as* the mouse for targets displayed at the screen surface.

The sliding cursor (SC) performance was on par with the pen techniques. One issue was the added effect of tracker noise and latency, since the cursor position was computed via a ray cast from the eye, which amplifies noise. Since the targets were small (0.65-1.05 cm) even a little tracker jitter may affect performance. This was also observed as an issue in the experiment. Moreover, some participants resorted to sliding the cursor up the fronts of target pillars, even though this was unnecessary. As SC used the eye-ray scene intersection the cursor was partially controlled by the head as well. This too may have affected performance.

The PR technique was worst overall. One aspect is that PR was susceptible to tracker jitter amplified down the ray (similar to SC). Yet, we also see evidence that PR was worse than SC, perhaps due to the potential for extra degrees of freedom to control the pen. Although the pen only required 2DOF rotations to control, positional control was also possible. The presence of a supporting (and thus jitter dampening) surface also helps SC. For a discussion of the effects of technology, see below. Although our results only apply *directly* to fish-tank VR systems, we expect that the effect of jitter will be even more pronounced in larger VR systems due to larger pointing distances, assuming comparable target sizes. Of course, larger targets are possible in a larger display, which may also reduce task difficulty.

We believe that the absence of tactile feedback and presence of stereo conflicts both impacted the PT technique. However, we cannot verify which effect is stronger. Our apparatus currently cannot produce a condition with tactile feedback and without stereo conflicts, or vice versa. Consequently, we can only examine both factors together and report their combined effect.

Finally, we previously reported [21] that a mouse-based system with 35 ms and a 3D tracker with 75 ms end-to-end latency result in a 15% difference in throughput. Cross-referencing our latency measurements (35 ms and 63 ms) with these results, the maximum difference due to latency should be less than 15% here. We thus hypothesize that in Exp. 1 PT may perform similar to the mouse, if a lower latency tracking system is used. However, we can also state that even with a low-latency tracker, PR is extremely unlikely to reach the performance level of the other techniques, as the differences are much larger than 15%. In Exp. 2 and 3, the differences between FC and the other techniques in the above-screen conditions were also larger than 15%. Consequently, we doubt this result would change with a low-latency tracker.

Modeling 3D Pointing with Fitts’ Law

Exp. 1 demonstrated strong correlations between *MT* and *ID*. For all techniques, the model explains over 80% of the variability. The mouse-based techniques show the highest correlations, at a level consistent with other Fitts' law studies. Correlations were also consistently high across all *three* studies for the mouse-based techniques. The addition of a target height factor does not seem to weaken the predictive capabilities of Fitts' law for these techniques within the range of motions evaluated.

The correlation between *MT* and *ID* for the PT technique was much worse in Exp. 2 and 3 compared to the first. The 2D Fitts' law formulation did not model these 3D pointing motions well. This is likely due to differences in pointing strategies and either stereo cue conflicts and/or the absence of tactile feedback. Without tactile feedback, participants had to rely on imperfect stereo depth cues to determine the correct target height. This is consistent with observations during Exp. 2: participants had a harder time hitting targets displayed above the screen with PT than at the screen, and resorted to "homing" motions, effectively searching for the right height. While in Exp. 2, participants could "discover" the correct height once per target circle, this was impossible in Exp. 3, where participants had to constantly adjust to changing heights. This made the task even harder and we frequently observed "homing motions". We have not yet analyzed our motion logs to fully support this "homing" motion hypothesis.

Most of the investigated techniques are well explained by the 2D formulation of Fitts' law or its direct extension to 3D. Hence, we did not yet investigate other 3D versions of Fitts' law, as our data does not support the use of these 3D versions. Also, as PT performed worst overall, we see little practical value for 3D UI designers in 3D versions of Fitts' law.

6 Conclusions

We conducted a series of experiments replicating the 2D pointing task prescribed by ISO 9241-9 in a fish-tank VR system using both a 3D tracked stylus and a mouse as input devices. Results indicate that 3D pointing performance degrades when targets are displayed stereoscopically above the screen for 3D techniques, but *not* for 2D techniques. Pointing motions *at or parallel to* the surface of the screen are well-modeled using the 2D formulation of Fitts' law for most techniques. Simply using the Euclidean 3D distance rather than 2D distance into account seems to predict 3D motions sufficiently well for most interaction techniques within the investigated range of motions.

6.1 Future Work

We logged motion trails of all techniques studied and plan to investigate this data in the future to better explain the results. We are also planning to investigate pointing in other VR systems, e.g., with haptic devices to attempt to split apart the accommodation-vergence mismatch and tactile feedback issue.

6.2 Acknowledgements

Thanks to the reviewers for helping us to improve the paper. Thanks to NSERC for funding this research.

References

[1] E. Bier. Skitters and jacks: interactive 3D positioning tools. *Interactive 3D graphics workshop 1986*, pp. 183-196.

[2] J. Boritz and K.S. Booth. A study of interactive 3D point location in a computer simulated virtual environment. *VRST 1997*, pp. 181-187.

[3] D.A. Bowman, D.B. Johnson, and L.F. Hodges. Testbed evaluation of virtual environment interaction techniques. *VRST 1999*, pp. 26-33.

[4] D.A. Bowman and L.F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *SI3D 1997*, pp. 35-38.

[5] K. Chun, B. Verplank, F. Barbagli, K. Salisbury. Evaluating haptics and 3D stereo displays using Fitts' law. *HAVE 2004*, pp. 53-58.

[6] P.M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. In *Journal of Experimental Psychology*, *47* (6), 1954, pp. 381-391.

[7] T. Grossman and R. Balakrishnan. Pointing at trivariate targets in 3D environments. *CHI 2004*, pp. 447-454.

[8] D.M. Hoffman, A.R. Girshick, K. Akeley and M.S. Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. In *Journal of Vision*, *8* (3), 2008, pp. 1-30.

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

[10] R.J. Jagacinski and J.M. Flach. Information theory and Fitts' law, in *Control theory for humans: quantitative approaches to modeling performance*, Lawrence Earlbaum Associates, 2003.

[11] R. Jota, M.A. Nacenta, J.A. Jorge, S. Carpendale and S. Greenberg. A comparison of ray pointing techniques for very large displays. *Graphics Interface 2010*, pp. 269-276.

[12] R.W. Lindeman, J.L. Sibert, and J.K. Hahn, Hand-held windows: towards effective 2D interaction in immersive virtual environments. *IEEE VR 1999*, pp. 205-212.

[13] L. Liu, R.v. Liere, K. Nieuwenhuizen, and J.-B. Martens. Comparing aimed movements in the real world and in virtual reality. *IEEE VR 2009*, pp. 219-222.

[14] I.S. MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, *7*, 1992, pp. 91-139.

[15] I.S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. *CHI 2008*, pp. 1633-1636.

[16] A. Murata and H. Iwase. Extending Fitts' law to a three-dimensional pointing task. *Human Movement Science*, *20* (6), 2001, pp. 791-805.

[17] K. Nieuwenhuizen, K, L. Liu, R.v. Liere, and J.-B. Martens. Insights from dividing 3D goal-directed movements into meaningful phases. *IEEE Computer Graphics and Applications*, *29* (6), 2009, pp. 44-53.

[18] R.R. Pagano. Chapter 6: Correlation in *Understanding statistics in the behavioural sciences*. Wadsworth Publishing, 2006.

[19] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. *Eurographics 1998*, pp. 41-52.

[20] R.W. Soukoreff and I.S. MacKenzie. Towards a standard for pointing device evaluation: perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies 61*(6), 2004, pp. 751-789.

[21] R.J. Teather, A. Pavlovych, W. Stuerzlinger and I.S. MacKenzie. Effects of tracking technology, latency, and spatial jitter on object movement. *IEEE 3D User Interfaces 2009*, pp. 43-50.

[22] M.P. Vijay and A. Steed. Profiling the behaviour of 3D selection tasks on movement time when using natural haptic pointing gestures. *VRST 2009*, pp. 79-82.

[23] C. Ware and K. Lowther. Selection using a one-eyed cursor in a fish tank VR environment. *ACM TOCHI*, *4* (4), 1997, pp. 309-322.

[24] S. Zhai, W. Buxton, and P. Milgram. The "silk cursor": investigating transparency for 3D target acquisition. *CHI 1994*, pp. 459 - 464.