The Effect of Stereo Display Deficiencies on Virtual Hand Pointing

Mayra Donaji Barrera Machuca
School of Interactive Arts + Technology
Simon Fraser University, Vancouver, Canada
mbarrera@sfu.ca

Wolfgang Stuerzlinger
School of Interactive Arts + Technology
Simon Fraser University, Vancouver, Canada
w.s@sfu.ca

ABSTRACT

The limitations of stereo display systems affect depth perception, e.g., due to the vergence-accommodation conflict or diplopia. We performed three studies to understand how stereo display deficiencies impact 3D pointing for targets in front of a screen and close to the user, i.e., in peripersonal space. Our first two experiments compare movements with and without a change in visual depth for virtual respectively physical targets. Results indicate that selecting targets along the depth axis is slower and has less throughput for virtual targets, while physical pointing demonstrates the opposite result. We then propose a new 3D extension for Fitts' law that models the effect of stereo display deficiencies. Next, our third experiment verifies the model and measures more broadly how the change in visual depth between targets affects pointing performance in peripersonal space and confirms significant effects on time and throughput. Finally, we discuss implications for 3D user interface design.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; Interaction techniques; Pointing.

KEYWORDS

3D Pointing, Cursor, Selection, Fitt's law

ACM Reference Format:
Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. 2019. The Effect of Stereo Display Deficiencies on Virtual Hand Pointing. In *Proceedings of CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019).* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300437

1 INTRODUCTION

With the broad availability of 3D displays as well as cheaper virtual reality (VR) and augmented reality (AR) devices, applications increasingly allow users to directly manipulate virtual 3D objects [33, 41], as such interaction is easy to use and requires little training [9]. Examples include using a controller to select an option in a floating menu or to pick up 3D objects in a virtual environment (VE) with their virtual hand. Most such systems use stereo display, i.e., show two different images to the users' eyes from viewpoints that correspond to the eye positions in a human head, to enable better spatial perception. Current stereo displays do not render spatial cues perfectly. For objects in peripersonal space, the human vision system can face depth perception issues, e.g., such as the vergence-accommodation conflict [25], diplopia [11], age-related near field vision problems [63] and personal stereo deficiencies. Little previous work has studied if stereo display deficiencies affect 3D pointing within arms' reach. If people cannot perceive depth accurately, they might face difficulties pointing to targets, as they may not be able to judge quickly if the cursor is where the target is.

Here, we focus only on virtual hand/wand pointing, i.e., selection of 3D targets within arms' reach, and ignore ray-based 3D pointing techniques [4, 9]. Our primary goal is to investigate how 2D Fitts' law [18, 19], could be extended to 3D model selection with current stereo display systems. Fitts' law predicts the movement time (MT) for how quickly people can point to a target on a screen:

$$MT = a + b * log_2(D/W + 1) = a + b * ID \qquad (1)$$

Where D and W are the target distance respectively size, while a and b are empirically derived via linear regression, the logarithmic term in Fitts' law is known as the index of difficulty (ID) and indicates the overall pointing task difficulty. We hypothesize that stereo display deficiencies affect selection more with an increasing change of visual depth between targets, due to the associated depth perception issues. Thus, movements directly towards or away from the viewer should have worse performance than lateral, i.e., side-to-side movements. Previously proposed 3D selection models [13, 42] ignore the effect of visual depth changes. A 3D Fitts' law formulation that captures the effect of common stereo

---

CHI 2019, May 4-9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300437

display shortcomings for interaction within arms' reach will aid user interface designers to create 3D user interfaces with better performance and improve user experience.

As with real-world reaching movements, all targets in peripersonal space, up to 70 cm from the user, require users to accurately estimate target depth for a successful selection. Our first two experiments involve only four different movement directions, two with a pronounced change in visual depth and two without. The first study investigates these movements in a stereo display system, and the second in a real-world setting. Based on a comparison of the results of both studies, we propose a new model for 3D pointing that depends on the depth change between targets. Our third study looks at a bigger range of movement directions and is used to verify the new model.

Through modelling how stereo display deficiencies affect performance our work extends previous work on ray-based pointing [59] to virtual hand techniques. We also extend the work by Lin and Woldegiorgis [35], as we compare the differences between physical and virtual targeting, by analyzing pure depth motions in isolation, and by modelling how the change in visual depth between targets affects performance in peripersonal space. Our contributions are:

- A comparison between virtual 3D target selections with the same or different visual depths in peripersonal space in a stereo display system. We show that movements with a change in visual depth suffer in time and throughput.
- A comparison of selecting physical objects with the same or different visual depths in peripersonal space. We show that lateral movements suffer regarding time and throughput. For the subset of conditions matching experiment 1, we identify that users are faster and exhibit higher throughout with physical targets.
- A new 3D Fitts' law model for 3D selection with virtual hands techniques that takes the effect of stereo display deficiencies into account.
- A study of targeting performance for 3D objects in peripersonal space with varying visual depth between targets. We identify that the predictive power of our new model surpasses previous ones.
- Recommendations for 3D user interface design.

2 MOTIVATION & HYPOTHESES

Eye-hand coordination research has found that vision impairments can affect visually guided motions [17] and that human lateral target discrimination is better than depth discrimination [60]. Thus, our goal is to identify if and how stereo display deficiencies affect pointing performance. We also aim to extend Fitts' law to stereo 3D displays. Previous work indicated that, compared with physical targeting, stereo displays affect both distance perception [34, 51] and pointing performance with ray-casting [20, 35] negatively. No existing model predicts how such deficiencies affect target selection with the hand/wand, i.e., direct 3D interaction, with stereo displays. Modelling this effect is important because stereo display systems are frequently used to display 3D scenes and are central to many VR and AR applications [27, 41, 45]. Most interactions in such systems let users directly interact with 3D objects. Thus, a better understanding and a model of how stereo displays affects 3D selection are needed.

H1 - The effect of a change in visual depth on selection

We hypothesize that a change in depth between targets negatively impacts selection performance in peripersonal space. Based on our hypothesis, we expect that pointing tasks with targets at different visual depths will exhibit different movement times and throughput than targets at the same visual depth. Standard Fitts' law depends only on D and W and does not include a dependency on visual depth. This hypothesis is motivated by the previous work discussed above and especially that such an effect was identified for ray-based pointing [59]. Compared to ray-based pointing, virtual hand selection has a bigger dependency on eye-hand coordination, as the user needs to place their hand within the object, simulating real-world interaction [8, 50].

H2 - The effect of stereo deficiencies on target selection

Previous work has shown that eye-hand coordination in stereo displays is affected by the target plane location [38, 60]. As we believe that the deficiencies of stereo displays affect 3D pointing performance negatively, we thus also hypothesize that the difference between physical and virtual targeting will increase with a change in visual depth. We expect that the difference between physical lateral and depth movements to be smaller than for virtual ones.

Model for the effect of stereo display deficiencies on pointing

We aim to predict the effect of depth depiction deficiencies of stereo displays on 3D pointing, i.e., how a change of visual depth between targets affects pointing time. We model this through an additional parameter in Fitts' law that depends on the change of target distance. Our motivation for an additional parameter is the asymmetry of human vision regarding acuity. With typical values for depth acuity, depth discrimination varies between 0.2 and 1 mm at distances between 30 and 70 cm, but visual/lateral acuity is smaller, e.g., 0.15 mm at 50 cm. This difference alone motivates the addition of a different, depth-change-dependent term for a 3D pointing model.

3 RELATED WORK

Virtual hand/wand techniques for 3D selection have been widely explored [7, 11, 48]. Such techniques require the user to intersect the target with their hand or wand and apply only to targets that are within arms' reach (in peripersonal space) and outside the screen (for screen-based systems). To successfully select a target at such relatively close distances, users need to accurately perceive the position of the target and then move their hand there. We first review relevant work on depth perception and then hand movements.

Depth Perception in Peri-personal Space in VR

In their review, Kenyon and Ellis [30] examine depth perception limitations in VEs. They identified that visual acuity and display resolution should match to provide a faithful image and that (visible) depth quantization should be avoided. Renner et al. [51] also reviewed previous work on human depth perception in VEs and identified a mean underestimation, 74% of the actual distance. This difference was independent of the VR display system but could be a consequence of each individuals' vision system [24] and age [63]. Nonpictorial depth perception at distances less than 2 m is mostly based on stereopsis, motion parallax, and convergence and accommodation [10, 16, 47, 51]. Vergence is the simultaneous rotational movement of the eyes when there is a change of the target distance, while accommodation is the change in the (eye) lens curvature to focus on objects at various distances [51]. For physical targets, vergence and accommodation are coupled.

Hong and Kang [26] investigated stereoscopic fusion for objects at different visual depths, i.e., distances from the screen. They found that for close locations, the view direction of both eyes intersects outside the range of the depth-of-field, which affects accommodation. This effect implies potential fatigue due to a vergence-accommodation conflict. Suryakumar et al. [56] identified that the vergence angle affects vergence time. When vergence and focal distances differ, perceived depth is less accurate [26, 30]. Hoffman et al. [25] also established that differences between the focal and the vergence distance could reduce stereo-acuity and cause visual fatigue.

Multiple studies established that depth cue conflicts affect depth perception [6, 30, 46, 48]. Specifically, in stereo display systems, the eyes need to focus on the display at a fixed distance, whereas in the real world they need to (con)verge at different distances to correctly perceive visual depth. Swan et al. [57] examined how humans perceive and estimate target depth in VEs in comparison to physical ones, and found that users overestimate distances in AR. They attributed this to the eye vergence angle. The specific distance cues that cause this effect have not been identified [34].

3D Pointing in Stereo Display Systems

Moving the hand or input device to the position of a 3D target is called 3D pointing. 3D selection then involves clicking a button to select said target. Previous work [34, 40] found that stereo displays are beneficial for depth-related tasks in the near-field. However, compared to the real world, distance perception is compressed in a stereoscopic view.

Depth cue conflicts affect pointing performance with virtual ray techniques [31, 32, 58]. Teather and Stuerzlinger [58] showed that varying target depth affects performance. Janzen et al. [29] found that performance for target depths between 110 and 330 cm is affected and identified an effect of the user's distance to the screen. Lin and Woldegiorgis [35] studied the effect of depth cue conflicts on virtual hand pointing and found that overestimation decreased with distance from the user. Still, they studied only distances beyond 65 cm, i.e., outside peripersonal space. Barrera and Stuerzlinger [5] found that lateral and depth movements were different when selecting nearby targets but they did not compare their results with physical pointing.

In eye-hand coordination research, Tramper and Gielen [60] investigated coordination differences in the frontal and depth planes for tracking and tracing tasks. For example, for tracking gaze leads finger position in the frontal plane but lags behind the finger in the depth plane. They found that the different lead times reflect differences in the dynamics of visuomotor control between coupled eye movements in the same lateral direction and opposite eye movements for depth accommodation.

Previous work has also studied the biomechanics of 3D pointing. The plane of shoulder exertion affects the muscles used [3, 39, 55]. Others [43, 52] observed that hand movements that cross the vertical midline of the body are more complex than those that do not. Lubos et al. [36] identified that visual perception has a larger effect on selection than motor actions.

Fitts Law and 3D Pointing

Fitts' law [19] is a widely used model for performance in pointing tasks [54] and predicts movement time (MT). A refined version, ISO 9241-411 [28], combines speed and accuracy into throughput to make the measurement less dependent on user strategies. Throughput can be used to assess performance differences between 3D pointing conditions [37]. Yet, the original Fitts' law formulation does not always correctly describe the data [29, 53]. Thus, researchers have proposed two-part models, such as the Shannon-Welford formulation [53], which includes a second logarithmic term to reflect a second process.

2D Fitts' law has been used to measure the difficulty of 3D pointing tasks [23, 32, 49]. However, the traditional 2D formulation of Fitts' law does not accurately predict 3D movement times or throughput in stereo displays and previous work adapted it. For example, Murata and Iwase [42] incorporated a parameter for the movement direction on a vertical plane facing the user, i.e., only when all targets have the same visual depth. Cha and Myung [13] added inclination and azimuth angles to Fitts' law. However, their work covers only forward motions with an azimuth between 30° and 60°, and -30° and -60° with physical targets. We extend their work to covers straightforward and backwards motions and a broader range of directions. Also, Cha and Myung [13] did not incorporate the effect of stereo display deficiencies in their work. While the angle between the 2D movement direction and the target orientation on a screen has been used to model 2D pointing movements [2, 22, 31], where the depth of the target is irrelevant, here we model the effect of the change in visual depth between 3D targets independent of target orientation. A model for ray-based pointing in 3D volumetric displays incorporates the angle between the movement vector and the target width vector [21].

4 USER STUDY 1

This study aims to establish a baseline for virtual 3D target selection at different visual depth using virtual hand techniques, to understand how the visual depth and the distance from the screen affect 3D selection.

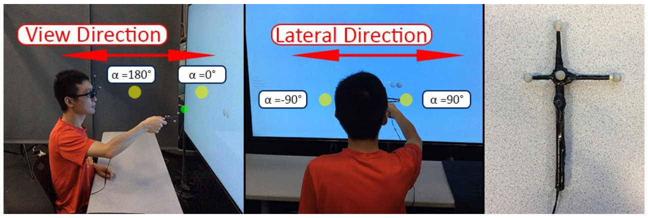

Figure 1: Side and back view of the experimental setting for study 1 with a close-up of the input device. Users select virtual targets in front of the screen by moving the wand to them and clicking a button. Movements were either along the view direction or lateral. The yellow spheres show exemplary targets positions, and the smaller green one the virtual cursor above the wand.

Methodology

Participants: We recruited 12 paid participants from the community (50% female). 42% of them were between 20 and 24 years old, 34% 25-29 years, 16% less than 19 and 8% 35-39 years old. All participants measured normal when tested for stereo viewing capability and used their dominant hand for the task. All of them were familiar with interactive 3D content, typically through 3D video games, 67% played between 0-5 hours/week, 8% 5-10 hours, 8% 10-20 hours, and 17% played more than 20 hours. A quarter was familiar with 3D CAD systems.

Apparatus: We used a 3.60 GHz Windows PC, Unity3D, a Nvidia GTX970, and an 85" stereo Samsung TV at 3840x2160. The 3D scene consisted of an open VE with no additional pictorial depth cues to ensure that external cues, such as shadows, could not be used to estimate target positions. We used eight 250Hz OptiTrack cameras calibrated to sub-millimetre precision for 3D tracking of the head and handheld wand (Figure 1). The virtual cursor, shown as a green 1 cm sphere, followed the 3D wand position. We placed this cursor 2 cm above the wand tip to avoid diplopia with the real wand. The task required users to intersect targets with the virtual cursor and to select them with a wand button click. We did not adjust for individual interocular distance (IPD) but did take motion parallax through head tracking into account. The stereo glasses afford a relatively wide FOV (approx. 120°horizontal). End-to-end system latency was ~140 ms.

We used the TV's 120Hz stereo capabilities with the included active shutter 3D glasses to show the targets outside the screen. Targets were two yellow spheres placed at specific distances from the user. All targets were displayed at eye height for each participant (Figure 1): a pair centred in front of the viewer, but with different visual depths ($\alpha = 0^\circ$, 180°), and a pair in the lateral direction with the same visual depths but different positions ($\alpha = -90^\circ$, 90°). Target depth was measured relative to the screen. When intersected by the virtual cursor, targets highlighted. Users only interacted with a single pair of targets at any given time. The current target was visible, while the inactive target was invisible. Upon selection, the two targets alternated. Users performed a constant number of selections for a given target separation, but depending on the movement direction, the visual depth changed or stayed the same.

Procedure: First, participants were tested to see if they could merge stereo targets correctly. Then participants were seated at the middle of the screen behind the table to keep their body parallel to the TV (Figure 1). Participants sat 75 cm away from the screen in a school chair (no swivel or casters). They used their dominant hand to perform the task. Target distances were between 40 and 70 cm from the user, i.e., between 5 and 35 cm from the screen. Participants were instructed only to move their arm while keeping their head and body in (about) the same position to keep the view direction mostly constant. The visual target height was matched to each participants' eye-level to eliminate the effect of vertical disparity. Then, they were instructed on the task and encouraged to practice until they felt comfortable with it. During training, the target shape was different from the one in the actual experiment.

The task was a 3D version of an ISO 9241-411 task [28], with targets positioned along a single axis. Users had to select the two 3D targets reciprocally. At the beginning of each task, participants saw a yellow (target) sphere floating in front of them and the green sphere for the virtual cursor at the wand. Participants were asked to alternate between selecting the targets as quickly and accurately as possible. We emphasized that movements have to be continuous between targets until they saw a resting prompt, which occurred between changes of movement axes. If they missed a target, participants had to continue to the next one. If a part of the virtual cursor was inside the target when participants clicked the wand button, we recorded a successful selection; otherwise, a miss. They had to select a specific number of targets (11) for each combination of size and distance, i.e., a set. The distance between targets and their size changed randomly, selected without replacement from the available options. Once they did all sets for a movement direction, the task changed to the other direction, counterbalanced across all participants.

Design

The study used a 4x3x3 within-subjects design. The independent variables were movement direction (α = -90°- 90°, and 0°-180°), target separations (10, 20, and 30 cm) and sizes (1.5, 2.5, and 3.5 cm). Overall, participants saw the targets at seven different visual depths. Average target size is approximately constant between changes in movement direction. Dependent variables were movement time (milliseconds, ms), error rate (percentage of missed targets), and throughput (bits per second, bps). Per target ID 11 trials were recorded. Across all distances and sizes, we used 9 distinct IDs from 1.94 to 4.39 bits. Each participant completed 3 repetitions for each ID, for a total of 594 trials (3 x 11 x 9 x 2). Across all participants, we recorded a total of 7128 trials.

Results

The results were analyzed using repeated measures ANOVA with $\alpha = 0.05$. Initially, we attempted to remove outliers. Yet, a 3σ criterion removed mostly data for view direction movements. Because the targets in the view direction are likely more affected by stereo deficiencies, we decided only to exclude double-clicks (2.4% of the data). As the data was not normally distributed we tried a log transform, but this did not lead to a normal distribution. We believe that the shape of the distribution of our data might be a side effect of the stereo display deficiencies. Thus, we picked the Aligned Rank Transform (ART) [62] to transform our data before the ANOVA. We followed ART instructions do compare the means. For cases with interactions we used an interaction contrast as recommended by ART to compare the means. For cases without interactions between factors, the means were analyzed using the Tukey Kramer test. Statistical results are reported in Table 1, with *** for $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and $n.s.$ not significant

Movement time: Overall, there was a significant main effect of direction on movement time (MT), (Figure 2.1.1). Average MTs for movements in the lateral direction were significantly faster than for the view direction. A post-hoc test found two different groups: α = (0°, 180°) and (90°, −90°).

Error Rate: There was a significant main effect between movement directions on error rate (Figure 2.1.2). The error rate for movements in the right lateral direction was significantly lower than for left or back and forth movements.

Throughput: (Effective) throughput was computed using the ISO 9241-411 method adapted for 3D motions. There was a significant effect of movement direction on throughput. All directions were significantly different (Figure 2.1.3).

Movement Path: We also analyzed the movement paths using target re-entry events (Figure 2.1.4), speed, ballistic and correction times. Both ballistic and correction times were calculated using Nieuwenhuizen's method [44]. There was a significant main effect of movement direction on all measures (Table 1). Both movement directions are significantly different for target-reentry and ballistic time, with lateral movements (α = 90°, −90°) being "superior" in all measures than in the view direction (α = 0°, 180°). For speed and correction time (α = 0°, 90°, −90°) were different from (α = 180°).

Table 1: User study 1 statistical results.

	Mov. Direction		ID		Mov. Direction X ID	
Measure	$F(3, 33)$	p	$F(8, 88)$	p	$F(24, 264)$	p
Movement Time	65.5	***	57.9	***	2.6	n.s.
Error Rate	4.9	**	3.2	**	1.7	n.s.
Throughput	28.7	***	5.2	***	1.3	n.s.
Target Re-entry	272.2	***	16.6	***	2.5	n.s.
Speed	11.35	***	346.4	***	3.1	*
Ballistic Time	33.9	***	27.9	***	2.1	n.s.
Correction Time	10.2	*	49.6	***	1.7	n.s.

Discussion

Lateral movements had noticeable better performance than those in the view direction. For lateral movements, MT vs. ID has the same slope, while for movements in the view direction it has not (Figure 3a). Using effective ID (IDe's) yields the same result, which provides further support. This effect is also visible in the throughput measure, where movements in the view direction have consistently about 20% fewer bps than lateral ones. When analyzing the movement paths, movements in the lateral direction follow previously identified patterns [44]. Yet, the post-hoc test on correction time shows that movements towards the user in the view direction have a much-extended correction phase, which could be interpreted as evidence for participants not being able

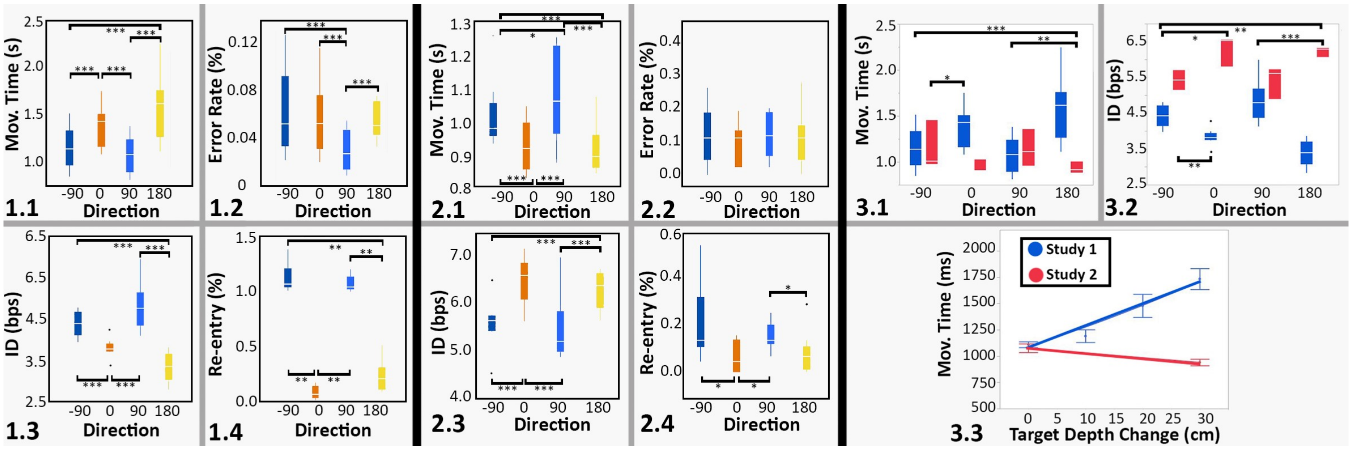

Figure 2: Study 1 and 2 results and comparison, with 95% CIs. (1) Movement Time, (2) Error Rate, (3) Throughput and (4) Target Re-entry. Comparison between matching conditions from study 1 and 2 are shown in (3.1) movement time, (3.2) throughput, and (3.3) movement time vs target depth change.

to perceive depth well, likely due to stereo display deficiencies (Figure 3b). Observations during the experiment also confirm that for movements in the view direction, participants sometimes made gross depth-estimation errors. Only after they identified their error, they started a second submovement to reach the correct position. We identified such behaviours in the data, as for depth movements 15% of the correction phases had a high speed (more than 20% of maximum). In contrast, only 6% of the lateral movements had such high-speed corrections. This effect is also evident in the Tukey-Kramer result for target re-entry, where all movement directions are in different groups. These findings support hypothesis H1, as selecting targets with similar IDs but in different movement directions exhibit different performance, and the gap increases with higher IDs.

Our results establish that a change in visual depth between targets affects user performance, as performance for movements in the view direction was worse than for lateral movements. Our results not only quantify this effect concerning movement time but also regarding error rate and throughput. Kopper et al.'s work on virtual ray techniques [31] showed that the target's angular size affects performance. Yet, while targets at different visual depths have different perceptual size, our observed differences (25% slower for depth movements) are well beyond any effect that can be explained through visual size differences alone (9%). On the other hand, the difference could be explained through human vision characteristics. Suryakumar et al. [56] found that (in the absence of accommodative cues) "pure" vergence times increase notably with vergence angle. For our vergence angles (0.6°, 1.3°, and 2°) they measured roughly 200 ms to 350 ms to verge in ideal conditions. In our study, we observe an average difference of 380 ms between movement directions across all IDs, which is at the upper end of this range. While this is not conclusive, their results can explain a substantial part of the effect. Overall, these results support our hypothesis H1, that stereo display deficiencies affect selection performance.

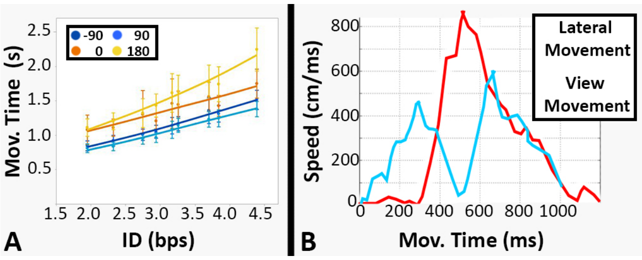

Figure 3: (a) MT vs. ID, (b) Typical lateral movement and exemplary second sub-movement in the view direction.

5 USER STUDY 2

Previous studies have identified that distance estimation for physical targets is better than for virtual targets [16, 34]. Thus, the objective of this study is to establish a baseline for physical 3D target selection at different visual depths in different movement directions. We also wanted to identify if the effects found in study 1 generalize to all 3D selections or are related to stereo display deficiencies. For this, we presented real targets at the same movement directions as study 1: a lateral pair ($\alpha = -90°, 90°$) and one in the view direction ($\alpha = 0°, 180°$). Due to mechanical restrictions, all conditions in this study had the same separation distance (30 cm) between targets, which matched one of the distances used in study 1. In effect, this study replicates a subset of study 1 in a physical setting.

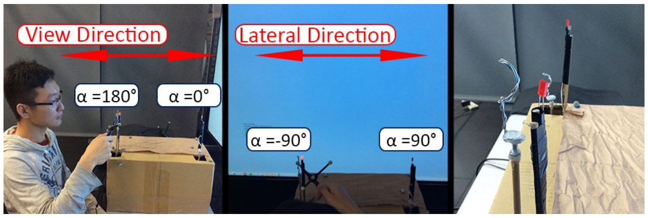

Figure 4: Side and back view of the physical prototype with a close-up of the targets. Users select targets by moving the top of the wand into the sphere indicated by the wire half-circle and clicking a button. Movements were either along the view direction (α = 0°, 180°) or lateral ($\alpha = -90°$, 90°).

Methodology

Participants: We recruited 12 new paid participants from the university community (59% female). 59% of the participants were between 20 and 24 years old, 17% 25-29 years, 8% less than 19, 8% 30-34 years, and 8% 35-39 years old. All participants measured normal when tested for their stereo viewing capability and used their dominant hand. All participants were familiar with interactive 3D content, most of them in the form of 3D video games, 75% played between 0-5 hours/week, 17% 5-10 hours, and 8% 10-20 hours/week. 17% of participants were familiar with 3D CAD systems.

Apparatus: We built a custom physical prototype (Figure 4) for the experiment, which mimics a subset of study 1 target sets. The prototype used two Behringer X32 motorized faders controlled by an Arduino Uno, with a top speed of 0.43 mm/ms. A core goal was to avoid obscuring one target with another, especially for movements in the view direction. The current target was made visible by being raised, while the inactive target was hidden by being lowered into the box at top speed. Upon selection, the two targets alternated. The distance between the two targets was 30 cm, and target size changed depending on the condition. The target shape was a half-circle made with wire that users could move through with the wand. We chose an open shape, as pilots with solid spheres revealed contact-based pointing strategies that are not comparable to the movements in study 1. To highlight selection, a LED turned on when the pointing device was inside the target sphere implied by the half-circular wire. We used the same tracking cameras and wand pointing device as in study 1, and selection was again indicated by pressing the wand button.

Procedure: The procedure was the same as in study 1, with matching target distances and relative positions. We again adapted the vertical position of the apparatus so that targets were at the participant's eye-level. The task was also the same as in study 1, with the exception that the distance between targets did not change. The experimenter physically changed the size of the targets, depending on the condition. The target size was selected without replacement from the available options and the first movement direction was counterbalanced across participants. At the beginning of each trial, participants saw a curved wire which indicated the shape of the target sphere. Participants were asked to select the target as quickly and accurately as possible by moving the wand top inside that sphere. The tip of the wand was a sphere of 1 cm, similar to the virtual cursor. Instructions again emphasized that the movement had to be continuous from target to target until they saw a 60 second resting prompt. The movement direction was changed after the whole set by rotating the apparatus to enable participants to perform trials with the other direction.

Design

The study used a 2x3 within-subjects design. The independent variables were movement direction ($\alpha = -90°$ - 90°, and α = 0° - 180°) and target size (1.5, 2.5, and 3.5 cm). Overall, participants saw targets at three different visual depths. The dependent variables were movement time (ms), error rate (percentage of targets missed), and throughput (bps). We recorded 11 trials per target ID. We used 3 distinct IDs between 3.38 and 4.52 bits. Each participant completed 3 repetitions of each ID, for a total of 198 trials (3 x 11 x 2 x 3). Across all participants, we recorded a total of 2376 trials.

Results

Data was not normally distributed. We did not remove outliers in *MT* and error distance, as this would have removed mostly data in the view direction. Thus, we only removed double-clicks (5.4%). We excluded one participant due to mechanical problems. After being ranked by ART, the results were analyzed using repeated measures ANOVA with α = 0.05. We used the same procedure as study 1 to compare means. Statistical results are reported in Table 2, with *** for $p < 0.001$, * $p < 0.05$, and *n.s.* not significant.

Movement time: There was a significant main effect of movement direction on *MT* (Figure 2.2.1). Both lateral movements were significantly slower than those in the view direction. A post-hoc comparison found a significant difference in times between α = (0°, 180°) and the other 2 directions.

Error Rate: There was no significant main effect between movement types on error rate (Figure 2.2.2).

Throughput: There was a significant effect of movement direction on throughput (Figure 2.2.3). Throughput for the lateral direction was significantly lower than for the view direction. A post-hoc comparison found a significant difference between two groups: α = (0°, 180°) vs. $\alpha = (-90°, 90°)$.

Movement Path: There was a significant effect of movement direction on target re-entry (Figure 2.2.4) and correction time. Speed and ballistic time were not significant. For correction time view direction movements (α = 0°, 180°) were statistically different from lateral ones ($\alpha = -90°, 90°$). For target re-entry ($\alpha = 0°$) is different from ($\alpha = -90°, 90°$), and ($\alpha = 180°$) is different from ($\alpha = 90°$).

Table 2: User study 2 statistical results.

	Mov. Direction		ID		Mov. Dir X ID	
Measure	$F(3,30)$	p	$F(2,20)$	p	$F(6,60)$	p
Movement Time	15.4	***	29.3	***	3.5	*
Error Rate	1.58	n.s.	14.3	***	1.03	n.s.
Throughput	14.5	***	4.1	*	0.19	n.s.
Target Re-entry	4.7	*	2.4	n.s.	1.2	n.s.
Speed	0.75	n.s.	2.4	n.s.	1.0	n.s.
Ballistic Time	0.6	n.s.	9.4	**	0.4	n.s.
Correction Time	15.8	***	62.5	***	3.6	*

Discussions

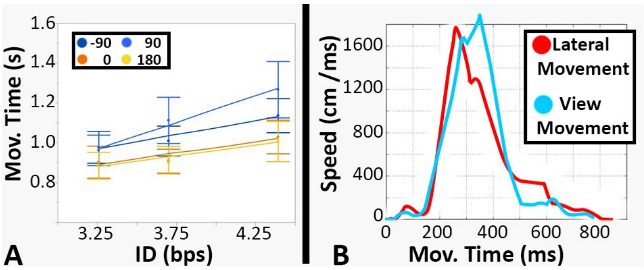

Figure 5: (a) Fitts' law Model for movement directions. (b) Examples of a single path movements for the same participant in the same condition: movement in the lateral direction and in the view direction

Lateral movements had noticeably worse performance than in the view direction (Figure 5a). The same analysis using effective ID's (IDe's) yields the same result, which supports this finding. This effect is also visible in the throughput measure, where lateral movements have consistently about 14% fewer bps than movements in the view direction. Analyzing the movement paths (Figure 5b), we found that movements both in the view and in the lateral direction follow standard patterns [44]. Also, only 2% of both lateral and depth movements had correction phases with speed higher than 20% of the maximum. We see this as evidence that the absence of stereo display deficiencies makes it easier to select real targets, regardless of the movement direction. The Tukey-Kramer test for re-entry showed that movements away from the user are different from all the other movements, but that movements towards the user are similar to left movements. These findings support hypothesis H1, as the effect of movement direction is less pronounced for real targets, where there are no stereo deficiencies. These findings also support H2, as the difference between movement directions is smaller than in experiment 1. Biomechanical factors can explain the asymmetry between lateral movements for selecting a target that is about −60° vs. 60° relative to the view direction [55] and by the hemispheric asymmetry for right and left side movements [61]. Yet, this effect is small (and not large enough to include in our model).

6 STUDY 1 & 2 COMPARISION AND 3D FITTS' LAW MODEL

To better understand how stereo display deficiencies affect 3D pointing, we compared the results of studies 1 and 2. Due to mechanical restrictions, study 2 only evaluated a single distance. Thus, we limited study 1 results in the following to the same target separation, 30 cm, as in study 2. Analysis of this limited data from study 1 still shows the same significant results as the full data. Given that the statistical results across matching conditions in the two studies come out in opposite ways and that lateral movements match closely (1123 ms for virtual movements vs 1148 ms for physical movements), we believe we can rule out either latency or resolution as a primary explanation. Also, we ruled out diplopia as a potential issue for targets between 30 and 70 cm from the user through early pilots. If there was a strong effect of the fixed IPD on the depth perception, one could expect a notable difference between study 1 and 2 for the lateral condition. As we did not observe such a difference, we believe that the fixed IPD did not notably bias our results.

Based on the significant effect of depth movements in study 1, we believe that stereo deficiencies, especially the vergence-accommodation conflict, are the likeliest cause for our observations. There is on average a 560 ms difference between the physical and virtual conditions in the view direction. Suryakumar et al. [56] compared pure vergence movements with disparity with vergence-accommodation movements without disparity and identified a trend where pure vergence movements with disparity take 50 ms longer. Our difference is much larger, but together with the fact that depth perception is impacted negatively on stereoscopic 3D displays [26, 30, 60], we see that as an indication that the fundamental interaction (vision-action) loop for pointing is impacted, too. The longer and faster corrections are another indication for this.

When comparing movement times across matching study conditions, study 2 movement times are statistically faster ($\mu = 1010ms$, $\sigma = 600$) than in study 1 ($\mu = 1290ms$, $\sigma = 600$) for both directions (Figure 2.3.1). The speed difference between virtual and physical selection is 19%, which roughly matches the 12% found in previous work [20]. Nieuwenhuize, et al.'s [44] results also support ours. Extending their work, we identify that the difference also exists regarding throughput (physical = $5.85bps$ vs. virtual = $4.12bps$) (Figure 2.3.2).

The difference between movement directions between experiments is also statistically significant ($F_{3,63} = 52.5, p < 0.0001$), which supports our H2.

Extending previous work [26, 55] we showed that stereo display deficiencies affect performance for virtual hand pointing. A potential confound is the difference between full sphere and half circle targets. Also, one could argue that the participants in study 1 practiced more. Yet, even with these potential advantages, study 2 participants were faster and more accurate. We expect that for the same amount of practice this effect would only increase. Still, we acknowledge that physical pointing needs to be explored in more depth for complete proof, especially for different target distances. For the movement paths (Figures 3b and 5b), we found that in stereo display systems only movements in the view direction exhibit longer initial phases followed by longer or faster corrections. For all other movements (including the lateral ones in study 1), movement paths conform to rapid aimed movements. Our results are also supported by previous work that found that movement kinematics in the initial movement phase are similar between physical and virtual targeting in stereo displays [20]. The post-hoc analysis of the movement paths that identifies a significant difference between movement directions further strengthens these results. In summary, our results support our hypotheses as follows:

- **H1**: the change in visual depth between targets with stereo displays affects 3D virtual hand pointing performance negatively. **Supported**
- **H2**: the deficiencies of stereo display systems increase the effect of the change in target depth for virtual hand pointing negatively. **Supported**

Based on these results, we propose a new 3D Fitts' law model for 3D virtual hand selection with stereo displays, where we add a factor depending on the change of target depth (CTD). We based our decision on the linear effect found in study 1 ($r^2 = 0.96$) and by Hong and Kang [26] for peri-personal space pointing. This linear effect is also present in study 2 but is smaller and due to its smaller magnitude could be explained there by biomechanical factors [3, 39, 43, 52] or depth perception issues [14, 47], see Figure 2.3.3. Yet, for study 1, the magnitude of the effect is (much) larger, which points to potential additional causes for stereo displays. Also, when fitting study 1 data with Fitts' law, we see an r^2 of (only) 0.59, $AIC = 243$. Even with the Shannon-Welford model, we see an r^2 of only 0.6, $AIC = 253$. The AIC score [1] has been used to choose between pointing models, with a rejection criterion for choosing among models of 10 pts as established by Burnham et al. [12]. Thus, we hypothesize that adding CTD into Fitts' law to calculate movement time will better predict performance and propose the following model:

$$MT = a + b * ID + c * CTD \tag{2}$$

Where a, b, and c are arbitrary constants to be determined through linear regression and ID, CTD are the index of difficulty and the change in target depth, respectively. CTD is measured in centimetres. $MT = 389.2 + 237.8 * ID + 15.74 * CTD$ models study 1 data well ($r^2 = 0.94$, $AIC = 205$). Consider that, after the multiplications, the magnitude of the last two terms is similar, which explains the improved fit for our new model. We recognize that adding a term (inside or outside of the ID) can create problems with units. Yet, we have not been able to identify a more elegant solution that matches our data.

7 USER STUDY 3

This study measures 3D target selection performance for a more comprehensive range of changes of target depths (CTDs) and movement directions. For this, we arranged targets in a circle along the view direction, which yields different visual depths for each target and lets us measure the effect of the CTD. This placement also requires a different movement direction for each pair of targets, which is a natural extension of the ISO task towards investigations of depth-based movements.

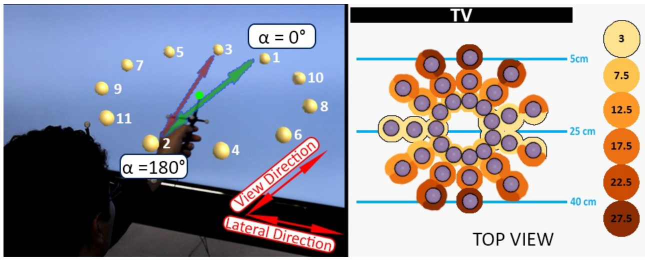

Figure 6: (a) The different target positions in study 3. Movements were back and forth between two opposite targets, as in the ISO methodology (two example target pairs shown with coloured arrows). Participants never saw this view; they only saw the current and the next target, both at eye level. (b) Top view of all targets binned into groups by depth change depending on direction.

Methodology

Participants: We recruited a different set of 12 paid participants from the local community (42% female). 58% were between 20-24 years old, 25% 25-29 years, 8% 19 or younger, and 9% 35-39 years old. All participants performed normally when tested for their stereo viewing capability and used their dominant hand to do the task. As with the other two experiments, all our participants were familiar with interactive 3D content, most of them in the form of 3D video games, 58%

played between 0-5 hours/week, 25% 5-10 hours, and 17% 10-20 hours.

Apparatus: The hardware setup and procedure was identical to study 1. The software was modified to use a circle of 11 target positions, adjusted to be at the eye level of each participant in a plane perpendicular to the screen (Figure 6a). The 11 targets were grouped into pairs, with each being part of two pairs. This placement gives each target a different visual depth, which then requires participants to both adapt to a new target depth and to change the movement direction for each new pair of targets (Figure 6b). Separating the targets into pairs also lets us record separate data for every movement direction and target distance, maintaining consistency with study 1 and 2 and increasing internal validity. All the targets in a circle were invisible, except for the current target and the next target, which was 10% transparent. Based on experience from a pilot, we chose to make the next target partially visible, as this makes it easier to understand which target pair needs to be selected.

Results

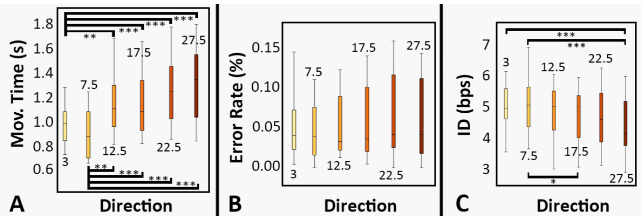

Figure 7: a) Movement Time, (b) Error Rate, (c) Throughput, showing 95% CIs.

As the data was not normally distributed, we ranked data through ART and analyzed the results using repeated measures ANOVA with $\alpha = 0.05$. We excluded double-clicks (2% of the data) but kept MT outliers for consistency with the previous study analyses. Initially, we attempted to analyze the data for each movement direction separately, but the results were not easy to interpret. Yet, as the effect of the change of target depth is our main interest, we transformed our data into CTDs and binned them into buckets of 5 cm, which yielded six distinct ranges. This transformation means that not all IDs were part of a CTD, as a CTD depend on the separation between targets and their angle. Statistical results are reported in Table 3, with *** for $p < 0.001$, * $p < 0.05$, and $n.s.$ not significant.

Movement time: Overall there was a significant main effect of CTD on MT (Figure 7a). Targets with smaller CTD were significantly faster than targets with larger CTD. A post-hoc comparison found a significant difference in MT between the groups for 3 and 7.5 cm from all the other distances.

Error Rate: There was no significant main effect between CTD on error rate (Figure 7b).

Throughput: There was a significant main effect of CTD on throughput. Throughput for targets with smaller CTD was significantly larger. A post-hoc test found two groups: (3, 7.5, 12.5, 17.5, 22.5) and (17.5, 22.5, 27.5 cm) (Figure 7c).

Movement Path: There was a significant effect of CTD on target re-entry, speed, ballistic and correction times. Target re-entry for targets with smaller CTD was significantly higher than for targets with larger CTD. A post-hoc comparison found a difference between 3 cm and the group of all other distances. Both speed and ballistic time had a significant effect. A post-hoc test on both measures yield the same groupings: the two shortest CTDs (3, 7.5 cm) and all the other CTDs. It also grouped the two middle CTDs (12.5, 17.5 cm) and the two larger ones (22.5, 27.5 cm). Finally, correction time for targets with lower CTD was significantly higher. A post-hoc test identified two groups: (3, 7.5cm) and a group of all other distances.

Table 3: User study 3 statistical results.

	CTD		ID	
Measure	$F(5, 55)$	p	$F(8, 88)$	p
Movement Time	17.6	***	37.8	***
Error Rate	0.2	n.s.	5.9	***
Throughput	3.3	**	9	***
Target Re-entry	13.3	***	30.7	***
Speed	91.5	***	135.8	***
Ballistic Time	22.6	***	29.9	***
Correction Time	6.4	***	18.4	***

Discussion

Consistent with study 1 results, study 3 confirms that stereo display deficiencies affect pointing performance negatively. Both studies show that a change in target depth decreases performance. In study 3, we found that movement time was faster for targets with smaller changes in depth (3 and 7.5 cm) compared to larger changes (22.5 and 27.5 cm). Lateral movements also influenced performance, but we attribute this (small effect) again to biomechanics [3, 39]. The overall effect is smaller than in study 1, and we argue that this is due to the gradual change in the target depth of target pairs during the task sequence. Our results extend previous findings that changing target depth reduces performance with ray-pointing [59] towards virtual hand-based pointing.

We fit the data with our proposed model and found $MT = 167.6 + 273.5 * ID + 3.35 * CTD$, with $r^2 = 0.98$, $AIC = 344$. Regression analysis showed that ID and CTD had a significant effect on the predictability of MT ($p < 0.0001$). To evaluate our new model, we compared it against four previously proposed models, see Table 4: Fitts' law, Shannon-Welford, Cha-Myung, and an inclination angle-based model. For the last one, we evaluated only the range 0 and ±90°, which loosely corresponds to the range of directions they examined. We conclude that our new model predicts the data better than any other one. Table 4. Model fits: r^2 (higher is better) and AIC scores (lower is better). The best results are highlighted.

	r^2	AIC
Fitts' Model [18]	0.96	359
Shannon-Welford Model [53]	0.96	361
Inclination angle model [42]	0.92	1268
Cha-Myung model [13]	0.93	2676
Our CTD Model	0.98	344

8 GENERAL DISCUSSION

We analyzed virtual target pointing performance in study 1 and compared movements with a change in visual depth with those with no change. We found that movement time is significantly longer for movements with a change in visual depth. In contrast, the results of study 2 identified the opposite effect in a physical setting (Figure 2.3.1). These results confirm our hypotheses H1 and H2 that stereo display deficiencies affect 3D pointing performance. However, we could not find that the distance to the screen affects pointing performance. Still the vergence-accommodation conflict could be the cause for the results on the largest depth differences. The asymmetry found in lateral movements with no change in depth might be attributable to biomechanical causes. In both study 1 and 2 re-entries for lateral movements were higher than for the depth movements. In conjunction with the movement paths analysis, this led us to believe that our participants used two different strategies for the task. For lateral movements, they used a fast, metronomic movement, with potential overshoot behaviours. In contrast, participants tried to select the target directly for depth movements and were more careful, as is visible in error rates (Figure 2.1.3). This effect was stronger in study 1, where we identified significantly longer and higher-speed correction phases for depth movements. We believe that this is caused by depth perception issues, as also identified by the motion analysis.

We also proposed a new, extended 3D Fitts' law model for virtual hand pointing that includes the effect of stereo deficiencies based on the change of visual depth between targets. Even though the effect is less pronounced for smaller changes in depth, study 3 supports our hypothesis H1. Yet, we found a systematic and mostly linear inverse relationship between target depth and movement time, until about 60 cm. Beyond that performance drops, likely due to biomechanical limitations, such as arm length (Figure 2.3.1). We expect these effects to apply to all directions and changes of target depths in peripersonal space for 3D selection with stereo displays.

Our results imply that it is possible to predict 3D virtual hand pointing performance with a linear model. One could argue that a model inversely proportional to distance might be better, due to the properties of human depth perception [14]. Yet, the limited reach of the human arm restricts the distance range. Still, we see no strong non-linearity in our data, and our new linear model fits the data quite well.

Our work extends Cha and Myung's work [13] not only through replication in common 3D stereo displays but also by investigating "pure" depth movements and by quantifying the effect of target depth changes. Our new extended 3D Fitts' law model predicts the effect of a change in target depth better than other alternatives. Our model could be straightforwardly extended to include the effect of vertical movements [42]. However, vertical movements are independent of depth cue conflicts, and thus there was no strong need to re-study this effect.

One might argue that a new Fitts' law model is not needed for stereo displays, a specific technology, and that it might complicate future research by fragmenting the space [15]. Yet, with the prevalent use of stereoscopic displays in current VR and AR systems, a model is important, also because it enables user interface designers to predict the performance of future 3D interfaces. Overall, we suggest an adaptation of our new extended 3D Fitts' law model for all work that involves modelling 3D pointing performance in stereoscopic display systems with virtual hand selection techniques.

9 RECOMMENDATIONS FOR 3D USER INTERFACE DESIGN

Based on our results, we present several recommendations for 3D user interface design for systems that use stereo displays with virtual hand techniques:

- The "best" working distance for direct interaction in stereo displays is between 50 and 65 cm from the user. In study 1, users selected targets at this distance faster than closer ones.
- Avoid having multiple interactive objects along the same line of sight, as movements to reach them are substantially slower. In study 1, we found that users sometimes perform two correction phases for such movements, which slows them down.
- Strong changes in target depth, of 20 cm or more, should be avoided, as targeting motions are slower. We identified evidence for this effect in study 1 and 3.
- Avoid targets close to the user's eyes, as this position reduces pointing performance and increases the likelihood for motions along the line of sight.
- Targets on the dominant (typically right) side have slightly, but not significantly, better performance than on the other side, based on movement biomechanics. We saw this in study 1 and 2.

Based on our recommendations the best setting of objects for user interfaces in a VE is to have them in a quasi-planar arrangement facing the user about 50-65 cm away. Most interactive objects should be off-center, either on the right or left side. Our results thus further strengthen the design guidelines proposed by Lubos et al. [36]. Finally, important objects should be on the dominant side of the user.

10 CONCLUSIONS

We conducted three studies to investigate the effect of stereo displays on 3D virtual hand pointing. Based on our results, we identify that stereo display deficiencies negatively affect virtual hand pointing, especially for movements in depth. We also show that the change in target depth has a linear relationship with time for 3D pointing tasks with stereo display systems ($r^2 = 0.96$). The results of our third study confirm the effect of the change of target depth on performance with 3D virtual hand pointing techniques. We also introduced a new 3D Fitts' law model for virtual hand techniques with stereo displays, which accounts for the effect of a change in target depth and which explains observations better than previous studies.

Our work successfully models virtual hand target selection performance in stereo displays. While it is the likeliest explanation, we cannot claim that we can prove that the vergence-accommodation conflict is the cause of performance loss for depth pointing. A completely different apparatus, namely a stereo display system that can provide correct vergence-accommodation cues, would be needed to investigate this. We also acknowledge that the integration of depth in the model needs to be revisited for larger distances. We plan to pursue this direction in the future and also to run experiments to investigate the issue in VR and AR headsets.

ACKNOWLEDGMENTS

This research was sponsored in part by CONACYT and NSERC.

REFERENCES

[1] Hirotugu A. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 6 (1974), 716–723. https://doi.org/10.1109/TAC.1974.1100705

[2] Johnny Accot and Shumin Zhai. 2003. Refining Fitts' law models for bivariate pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM Press, New York, New York, USA, 193–200. https://doi.org/10.1145/642611.642646

[3] Nicholas T. Antony and Peter J. Keir. 2010. Effects of posture, movement and hand load on shoulder muscle activity. *Journal of Electromyography and Kinesiology* 20, 2 (Apr 2010), 191–198. https://doi.org/10.1016/j.jelekin.2009.04.010

[4] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (May 2013), 121–136. https://doi.org/10.1016/j.cag.2012.12.003

[5] Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. 2018. Do Stereo Display Deficiencies Affect 3D Pointing? *Extended abstracts of the SIGCHI conference on Human Factors in Computing Systems (CHI '18)* (2018), 1–6. https://doi.org/10.1145/3170427.3188540

[6] Geoffrey P. Bingham, Arthur Bradley, Michael Bailey, and Roy Vinner. 2001. Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments. *Journal of Experimental Psychology: Human Perception and Performance* 27, 6 (2001), 1314–1334. https://doi.org/10.1037//0096-1523.27.6.1314

[7] James Boritz and Kellogg S. Booth. 1997. A study of interactive 3D point location in a computer simulated virtual environment. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '97)*. ACM Press, New York, New York, USA, 181–187. https://doi.org/10.1145/261135.261168

[8] Doug A. Bowman and Larry F. Hodges. 1997. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *Proceedings of the Symposium on Interactive 3D Graphics (SI3D'97)* (1997), 35–ff. https://doi.org/10.1145/253284.253301

[9] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola Jr., and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice* (1st ed.). Addison-Wesley. 512 pages.

[10] James M. Brown and Naomi Weisstein. 1988. A spatial frequency effect on perceived depth. *Perception & Psychophysics* 44, 2 (Mar 1988), 157–166. https://doi.org/10.3758/BF03208708

[11] Gerd Bruder, Frank Steinicke, and Wolfgang Stuerzlinger. 2013. Effects of visual conflicts on 3D selection task performance in stereoscopic display environments. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI'13)*. 115–118. https://doi.org/10.1109/3DUI.2013.6550207

[12] K.P. Burnham and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed)*. Vol. 172. 488 pages. https://doi.org/10.1016/j.ecolmodel.2003.11.004 arXiv:arXiv:1011.1669v3

[13] Yeonjoo Cha and Rohae Myung. 2013. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics* 43, 4 (Jul 2013), 350–355. https://doi.org/10.1016/j.ergon.2013.05.005

[14] James E. Cutting and Peter M. Vishton. 1995. Perceiving Layout and Knowing Distances: The Integration, Relative Potency and Contextual Use of Different Information about Depth. In *Handbook of perception and cognition, Vol 5; Perception of space and motion.* Academic Press, San Diego, California, USA, 69–117.

[15] Heiko Drewes. 2010. Only one Fitts' law formula please!. In *Extended abstracts of the SIGCHI conference on Human Factors in Computing Systems (CHI '10)*. ACM Press, New York, New York, USA, 2813. https://doi.org/10.1145/1753846.1753867

[16] Frank H. Durgin, Dennis R. Proffitt, Thomas J. Olson, and Karen S. Reinke. 1995. Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance* 21, 3 (1995), 679–699. https://doi.org/10.1037/0096-1523.21.3.679

[17] G. N. Dutton, A. Saaed, B. Fahad, R. Fraser, G. McDaid, J. McDade, A. Mackintosh, T. Rane, and K. Spowart. 2004. Association of binocular lower visual field impairment, impaired simultaneous perception, disordered visually guided motion and inaccurate saccades in children with cerebral visual dysfunction-a retrospective observational study. *Eye* 18, 1 (Jan 2004), 27–34. https://doi.org/10.1038/sj.eye.6700541

[18] Paul M. Fitts. 1954. The Information Capacity of the Human Motor System in Controlling the Amplitude of Movements. *Journal of Experimental Psychology* 47, 3 (1954), 381–391.

[19] Paul M. Fitts and James R. Peterson. 1964. Information capacity of discrete motor responses. *Journal of Experimental Psychology* 67, 2 (1964), 103–112. https://doi.org/10.1037/h0045689

[20] Evan D. Graham and Christine L. MacKenzie. 1996. Physical versus virtual pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96)*. ACM Press, New York, New York, USA, 292–299. https://doi.org/10.1145/238386.238532

[21] Tovi Grossman and Ravin Balakrishnan. 2004. Pointing at trivariate targets in 3D environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM Press, New York, New York, USA, 447–454. https://doi.org/10.1145/985692.985749

[22] Tovi Grossman and Ravin Balakrishnan. 2005. A probabilistic approach to modeling two-dimensional pointing. *ACM Transactions on Computer- Human Interaction* 12, 3 (Sep 2005), 435–459. https://doi.org/10.1145/1096737.1096741

[23] Taejin Ha and Woontack Woo. 2010. An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI'10)*. IEEE, 91–98. https://doi.org/10.1109/3DUI.2010.5444713

[24] Robert F. Hess, Long To, Jiawei Zhou, Guangyu Wang, and Jeremy R. Cooperstock. 2015. Stereo Vision: The Haves and Have-Nots. *i-Perception* 6, 3 (Jun 2015), 1–5. https://doi.org/10.1177/2041669515593028

[25] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* 8, 3 (Mar 2008), 33.1–30. https://doi.org/10.1167/8.3.33

[26] Hyungki Hong and Seok Hyon Kang. 2015. Measurement of the lens accommodation in viewing stereoscopic displays. *Journal of the Society for Information Display* 23, 1 (Jan 2015), 19–26. https://doi.org/10.1002/jsid.303

[27] HTC. 2016. VIVE. (2016). https://www.vive.com/ca/

[28] ISO. 2015. ISO 9241-400:2007 Ergonomics of human–system interaction – Part 400: Principles and requirements for physical input devices. (2015).

[29] Izabelle Janzen, Vasanth K. Rajendran, and Kellogg S. Booth. 2016. The Impact of Target Depth on Pointing Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM Press, New York, New York, USA, 188–199. https://doi.org/10.1145/2858036.2858244

[30] Robert V. Kenyon and Stephen R. Ellis. 2014. Vision, Perception, and Object Manipulation in Virtual Environments. In *Virtual Reality for Physical and Motor Rehabilitation*, Patrice L. Weiss, Emily A. Keshner, and Mindy F. Levin (Eds.). Springer New York, New York, NY, 47–70. https://doi.org/10.1007/978-1-4939-0968-1

[31] Regis Kopper, Doug A. Bowman, Mara G. Silva, and Ryan P. McMahan. 2010. A human motor behavior model for distal pointing tasks. *International Journal of Human - Computer Studies* 68, 10 (Oct 2010), 603–615. https://doi.org/10.1016/j.ijhcs.2010.05.001

[32] Georgios Kouroupetroglou, Alexandros Pino, Athanasios Balmpakakis, Dimitrios Chalastanis, Vasileios Golematis, Nikolaos Ioannou, and Ioannis Koutsoumpas. 2012. Using Wiimote for 2D and 3D Pointing Tasks: Gesture Performance Evaluation. In *Proceedings of the International Conference on Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*. Lecture Notes in Computer Science, Vol. 7206. Springer Berlin Heidelberg, Berlin, Heidelberg, 13–23. https://doi.org/10.1007/978-3-642-34182-3_2

[33] Leap Motion. 2016. Orion. (2016). https://www.leapmotion.com/

[34] Chiuhsiang J. Lin and Bereket H. Woldegiorgis. 2015. Interaction and visual performance in stereoscopic displays: A review. *Journal of the Society for Information Display* 23, 7 (Jul 2015), 319–332. https://doi.org/10.1002/jsid.378

[35] Chiuhsiang J. Lin and Bereket H. Woldegiorgis. 2017. Egocentric distance perception and performance of direct pointing in stereoscopic displays. *Applied Ergonomics* 64 (Oct 2017), 66–74. https://doi.org/10.1016/j.apergo.2017.05.007

[36] Paul Lubos, Gerd Bruder, and Frank Steinicke. 2014. Analysis of direct selection in head-mounted display environments. *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI'14)* (2014), 11–18. https://doi.org/10.1109/3DUI.2014.6798834

[37] I. Scott MacKenzie. 2015. Fitts' Throughput and the Remarkable Case of Touch-Based Target Selection. In *Proceedings of the International Conference on Human-Computer Interaction (HCII'15) (Lecture Notes in Computer Science)*, Masaaki Kurosu (Ed.), Vol. LNCS 9170. Springer International Publishing, Switzerland, 238–249. https://doi.org/10.1007/978-3-319-20916-6_23

[38] Guido Maiello, Miyoung Kwon, and Peter J Bex. 2018. Three-dimensional binocular eyeâĂŞhand coordination in normal vision and with simulated visual impairment. *Experimental Brain Research* 236, 3 (Mar 2018), 691–709. https://doi.org/10.1007/s00221-017-5160-8

[39] Alison McDonald, Bryan R. Picco, Alicia L. Belbeck, Amy Y. Chow, and Clark R. Dickerson. 2012. Spatial dependency of shoulder muscle demands in horizontal pushing and pulling. *Applied Ergonomics* 43, 6 (Nov 2012), 971–978. https://doi.org/10.1016/j.apergo.2012.01.005

[40] John P. McIntire, Paul R. Havig, and Eric E. Geiselman. 2014. Stereoscopic 3D displays and human performance: A comprehensive review. *Displays* 35, 1 (Jan 2014), 18–26. https://doi.org/10.1016/j.displa.2013.10.004

[41] MetaVision. 2012. MetaVision desktop. (2012). http://www.metavision.com/

[42] Atsuo Murata and Hirokazu Iwase. 2001. Extending Fitts' law to a three-dimensional pointing task. *Human Movement Science* 20, 6 (Dec 2001), 791–805. https://doi.org/10.1016/S0167-9457(01)00058-6

[43] Kephart Newell C. 1962. *The Slow Learner in the Classroom.* C. E. Merrill Books, Columbus. 292 pages.

[44] Karin Nieuwenhuizen, Jean-Bernard Martens, Lei Liu, and Robert van Liere. 2009. Insights from Dividing 3D Goal-Directed Movements into Meaningful Phases. *IEEE Computer Graphics and Applications* 29, 6 (Nov 2009), 44–53. https://doi.org/10.1109/MCG.2009.121

[45] Oculus. 2016. Rift. (2016). https://www.oculus.com/

[46] Robert Patterson. 2007. Human Factors of 3-D Displays. *Journal of the Society for Information Display* 15, 11 (2007), 861–871.

[47] Robert Patterson and Wayne L. Martin. 1992. Human stereopsis. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 34, 6 (Dec 1992), 669–92. https://doi.org/10.1177/001872089203400603

[48] Max Pfeiffer and Wolfgang Stuerzlinger. 2015. 3D virtual hand pointing with EMS and vibration feedback. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI'15)*. IEEE, New York, New York, USA, 117–120. https://doi.org/10.1109/3DUI.2015.7131735

[49] Alexandros Pino, Evangelos Tzemis, Nikolaos Ioannou, and Georgios Kouroupetroglou. 2013. Using Kinect for 2D and 3D Pointing Tasks: Performance Evaluation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8007 LNCS. Springer-Verlag Berlin Heidelberg, 358–367. https://doi.org/10.1007/978-3-642-39330-3_38

[50] Ivan Poupyrev, T. Ichikawa, S. Weghorst, and Mark Billinghurst. 1998. Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. *Computer Graphics Forum* 17, 3 (Aug 1998), 41–52. https://doi.org/10.1111/1467-8659.00252

[51] Rebekka S. Renner, Boris M. Velichkovsky, and Jens R. Helmert. 2013. The perception of egocentric distances in virtual environments - A review. *Comput. Surveys* 46, 2 (Nov 2013), 1–40. https://doi.org/10.1145/2543581.2543590 arXiv:arXiv:1502.07526v1

[52] W. N. Schofield. 1976. Do children find movements which cross the body midline difficult? *Quarterly Journal of Experimental Psychology* 28, 4 (1976), 571–582. https://doi.org/10.1080/14640747608400584

[53] Garth B. D. Shoemaker, Takayuki Tsukitani, Yoshifumi Kitamura, and Kellogg S. Booth. 2012. Two-Part Models Capture the Impact of Gain on Pointing Performance. *ACM Transactions on Computer- Human Interaction* 19, 4 (Dec 2012), 1–34. https://doi.org/10.1145/2395131.2395135

[54] R. William Soukoreff and I. Scott MacKenzie. 2004. 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 (Dec 2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001

[55] Helmut Strasser and Karl-Werner Müller. 1999. Favorable movements of the hand-arm system in the horizontal plane assessed by electromyographic investigations and subjective rating. *International Journal of Industrial Ergonomics* 23, 4 (Mar 1999), 339–347. https://doi.org/10.1016/S0169-8141(98)00050-X

[56] Rajaraman Suryakumar, Jason P. Meyers, Elizabeth L. Irving, and William R. Bobier. 2007. Vergence accommodation and monocular closed loop blur accommodation have similar dynamic characteristics. *Vision Research* 47, 3 (2007), 327–337. https://doi.org/10.1016/j.visres.2006.10.006

[57] J. Edward Swan, Gurjot Singh, and Stephen R. Ellis. 2015. Matching and Reaching Depth Judgments with Real and Augmented Reality Targets. *IEEE Transactions on Visualization and Computer Graphics* 21, 11 (Nov 2015), 1289–1298. https://doi.org/10.1109/TVCG.2015.2459895

[58] Robert J. Teather and Wolfgang Stuerzlinger. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI'11)*. IEEE, 87–94. https://doi.org/10.1109/3DUI.2011.5759222

[59] Robert J. Teather and Wolfgang Stuerzlinger. 2013. Pointing at 3d target projections with one-eyed and stereo cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM Press, New York, New York, USA, 159. https://doi.org/10.1145/2470654.2470677

[60] Julian J. Tramper and Stan Gielen. 2011. Visuomotor coordination is different for different directions in three-dimensional space. *The Journal of Neuroscience* 31, 21 (2011), 7857–7866. https://doi.org/10.1523/JNEUROSCI.0486-11.2011

[61] Jean-Luc Velay, Virginie Daffaure, Nathalie Raphael, and Simone Benoit-Dubrocard. 2001. Hemispheric Asymmetry and Interhemispheric Transfer in Pointing Depend on the Spatial Components of the Movement. *Cortex* 37, 1 (Jan 2001), 75–90. https://doi.org/10.1016/S0010-9452(08)70559-8

[62] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 143. https://doi.org/10.1145/1978942.1978963

[63] Charles M. Zaroff, Magosha Knutelska, and Thomas E. Frumkes. 2003. Variation in Stereoacuity: Normative Description, Fixation Disparity, and the Roles of Aging and Gender. *Investigative Ophthalmology & Visual Science* 44, 2 (Feb 2003), 891. https://doi.org/10.1167/iovs.02-0361