2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 978-1-6654-5325-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISMAR55827.2022.00056

Comparing the Fidelity of Contemporary Pointing with Controller Interactions on Performance of Personal Space Target Selection

Sabarish V. Babu*
Clemson University

Hsiao-Chuan Huang†
National Yang Ming Chiao Tung University

Robert J. Teather‡
Carleton University

Jung-Hong Chuang§
National Yang Ming Chiao Tung University

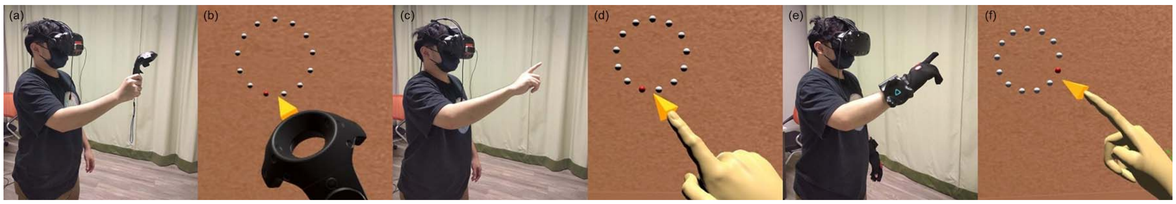

Figure 1: Left to right (a) shows participant in Controller condition and (b) Fitts' law selection in first person perspective. Similarly, (c) and (d) are images of Point-LF condition, and (e) and (f) are images of Point-HF condition.

Abstract

The goal of this research is to provide much needed empirical data on how the fidelity of popular hand gesture tracked based pointing metaphors versus commodity controller based input affects the efficiency and speed-accuracy tradeoff in users' spatial selection in personal space interactions in VR. We conduct two experiments in which participants select spherical targets arranged in a circle in personal space, or near-field within their maximum arms reach distance, in VR. Both experiments required participants to select the targets with either a VR controller or with their dominant hand's index finger, which was tracked with one of two popular contemporary tracking methods. In the first experiment, the targets are arranged in a flat circle in accordance with the ISO 9241-9 Fitts' law standard, and the simulation selected random combinations of 3 target amplitudes and 3 target widths. Targets were placed centered around the users' eye level, and the arrangement was placed at either 60%, 75%, or 90% depth plane of the users' maximum arm's reach. In experiment 2, the targets varied in depth randomly from one depth plane to another within the same configuration of 13 targets within a trial set, which resembled button selection task in hierarchical menus in differing depth planes in the near-field. The study was conducted using the HTC Vive head-mounted display, and used either a VR controller (HTC Vive), low-fidelity virtual pointing (Leap Motion), or a high-fidelity virtual pointing (tracked VR glove) conditions. Our results revealed that low-fidelity pointing performed worse than both high-fidelity pointing and the VR controller. Overall, target selection performance was found to be worse in depth planes closer to the maximum arms reach, as compared to middle and nearer distances.

Index Terms: Human-centered computing—Virtual reality; Computing methodologies—Perception

*e-mail: sbabu@clemson.edu
†e-mail: udjjyelm633@gmail.com
‡e-mail: RobTeather@cunet.carleton.ca
§e-mail: jhchuang@cs.nctu.edu.tw

1 Introduction

Virtual reality (VR) is steadily becoming more widely accessible as low-cost hardware such as head-mounted displays, motion trackers and hand tracking systems becomes more widespread. This trend toward low-cost consumer VR systems has led to several applications in VR ranging from motor skills training, rehabilitation, and others [15, 48]. These simulations often require the user to move the controller or virtual hand to select a target, and interact with menus or other objects within the users' near-field or personal space. A wide variety of low-cost input devices and techniques have been implemented in VR systems to facilitate near-field target selection [1, 41, 53].

Several gesture based input methods have been developed to facilitate natural interaction in near-field VR applications for selecting buttons on spatially located menu planes or objects in personal space interactions [25]. These include low cost infrared depth camera based input devices such as the Leap Motion [54] and Intel's RealSense [46], to mid-to-higher end commodity devices such as the Noitom Hi5 glove [40], CyberGlove III [51] and Vicon [56] that typically costs thousands of dollars. The low-end commodity gesture tracking systems typically have high latency and low framerate, poor gesture recognition capabilities that is partially due to computer vision based tracking systems that often require line-of-sight, may not be robust to occlusion, and providing a lower fidelity of interaction [27]. Whereas high-end consumer gesture tracking systems typically have low latency, high framerate, and have systems that track fine finger motions and are robust to occlusions, providing a high fidelity of interaction [27, 36]. On the other hand, VR controller based spatial selection and manipulation has become commonplace and is considered a best existing method of interaction in most commercial VR systems [2, 25, 45]. One reason for this is that in popular VR systems, such as the HTC Vive or Oculus Rift, the controller's position and orientation are tracked using recent advances in high fidelity tracking systems that are used to track both the users' head and controller pose. Therefore, it becomes important to compare and contrast the efficacy and effectiveness of the interaction and system fidelity (McMahan et al. [35]) of commodity gesture based input devices with the standard controller based input device on spatial selection performance.

A popular method for evaluating the efficacy and effectiveness of specific input modes is Fitts' law, and while it has been initially designed for 1D stylus motions [18] and later 2D interactions [30], it

978-1-6654-5325-7/22/$31.00 ©2022 IEEE
DOI 10.1109/ISMAR55827.2022.00056

has been widely used in recent studies for 3D input devices [1,2,53]. Different factors that have been studied with Fitts' law include controller types [45,53], haptic and pseudo-haptic feedback [7,52], using gesture and mouse interfaces [9], spatial offsets [19,23], and avatar presence [47]. Researchers have evaluated the efficacy of spatial selection of popular commodity depth camera based pointing to controller based input, and they found that controller based spatial selection was more efficient as compared to low fidelity pointing [2]. However, it is unclear how the fidelity of commodity gesture based pointing interaction techniques as compared to popular controller based input affects the performance of spatial selection in personal space interactions in VR. Thus, our study greatly extends the prior research and focuses on empirically evaluating how the system and interaction fidelity of gesture based pointing input techniques as compared to controller based input differently affects personal space or near-field spatial selection in VR.

In this paper, we present the results of two experiments using a selection task to evaluate fine-motor target selection within personal space or the distance within the users' maximum arms reach, as defined by Cutting and Vishton [12]. We evaluated common 3D input methods, specifically comparing controller-based selection to two fidelity levels of hand gesture based (i.e., without a controller) selection in VR. We also studied how selection efficiency and the accuracy-speed tradeoff is affected by selection target depth, within the users' maximum arm's reach in personal space or near-field interactions in VR.

In our first experiment, we compared controller and pointing interactions in a task closely resembling the ISO 9241-9 standard of a circular arrangement of 13 targets in a plane. We positioned the targets in a planar configuration in 3D space. Such a task replicates selecting menu widgets such as buttons in a GUI plane that is placed as a menu object in personal space. Such tasks are common in modern VR systems (e.g., the Oculus Home). We centered this target arrangement around the users' optical or central viewing axis at their eye height, and presented the targets at depth planes of either 60%, 75%, or 90% of the user's maximum arm's reach, to assess the influence of near-field target (or menu object) depth in such tasks.

Similar to the first experiment, the second experiment used a circular configuration of 13 targets centered around the users' viewing axis. However, the main difference was that rather than arranging all targets in a single depth plane, we positioned each individual target at a random depth plane at either 60%, 75%, or 90% of the user's maximum arm's reach. This made it difficult to predict at which depth the next target would be. As a result, the user would have to dynamically adjust their end-effector depth from one selection to the next along the viewing axis. This second experiment allowed us to evaluate the selection of menu widgets and buttons in a manner similar to a hierarchical menu GUI system, such that the buttons may be arranged at different depth planes along the viewing direction. This scenario also typifies VR personal space selection in which users' may not know what target they will acquire next, as the next target may appear at an arbitrary depth in the near-field.

Together, our experiments attempt to robustly investigate how the fidelity of pointing based interaction and a controller-based interaction affects selection performance in both the lateral side-to-side direction and forward-backward (viewing) direction movement axes of the 3D space. We designed these experiments to fill a void in the literature relating to selection efficiency in low-fidelity and high-fidelity pointing versus controller interactions in personal space or near-field interactions.

2 Related Work

2.1 Brief Background on Fitts' Law

Fitts' law is a predictive model stating that the movement time (MT) for an aimed motion is linearly related to the Index of Difficulty (ID), which in turn is based on the target size (known as *width* or W) and the distance between the current cursor position and the target (known as *amplitude* or A). This relationship is described by the following equations:

$$MT = a + b * ID \tag{1}$$

where

$$ID = \log_2(\frac{A}{W} + 1) \tag{2}$$

in Equation 1, a and b are constants determined through linear regression, and this can extend from 1D and 2D interactions [18,30].

2.1.1 ISO 9241-9 Standard

The ISO 9241-9 standard was proposed to standardize the evaluation of pointing devices, adjusting W and A for the accuracy of the aimed movements [38]. These adjustments also adjust the value of ID to what is called the *effective* Index of Difficulty, or ID_e, depicted as:

$$ID_e = \log_2(\frac{A_e}{W_e} + 1) \tag{3}$$

The variable W_e, the effective width, is calculated as

$$W_e = 4.133 * SD_x \tag{4}$$

where SD_x is the standard deviation of the displacements from the target center. To derive W_e, this error is multiplied by 4.133, which corresponds to a z score of ± 2.066, which accounts for 96% of the selections. Thus, this accuracy adjust normalizes the experiment error rate to 4%, facilitating comparison between experiments with varying error rates [50]. Similarly, A_e is the average of how far the users actually moved the cursor rather than the distance that was presented to them. The ISO 9241-9 standard most commonly implements a reciprocal tapping task that involves moving the selector to different targets that are laid out in a circular arrangement of different amplitudes and target widths.

This extension to ID is in turn used to calculate pointing throughput (TP). Throughput combines speed (via movement time) and accuracy (via the accuracy adjustment of W_e) to produce a metric of task performance that is stable despite speed-accuracy trade-offs commonly observed in such tasks [31]. Throughput is calculated as:

$$TP = \frac{ID_e}{MT} \tag{5}$$

This helps to facilitate comparison between different studies rather than simply comparing movement time, and thus it can be used to more easily compare and contrast different interaction methods.

2.2 Fitts' law in Virtual Reality

Fitts' law is a model of paychomotor behavior in rapid aimed movements [18]. Derived from Shannon's Theorem, this principle states that the time to reach a target is linearly related to the difficulty of the task, which is itself dependent on the distance to the target and the size of the target [30]. This was initially used in one-dimensional movements, specifically evaluating reaction time within two targets [29]. MacKenzie et al. then helped to extend Fitts' law to two-dimensional interfaces. They list six 2D input modes: mouse, trackball, joystick, touchpad, helmet-mounted sight, and eye tracking [30]. Fitts' law was then used as a basis for these modes in order to evaluate the tools consistently, as the differences between them could make it difficult to compare them accurately by conventional means. Since that original study, the results and the findings of Fitts' original study have been widely applied to many experiments and interfaces, including video game controllers [38], eye-tracker selection [44], and mobile device tilt control [32].

It is worth noting that several researchers have applied Fitts' law to 3D selection of targets in mid-air. To evaluate the benefits of tactile feedback, for example, previous researchers have compared mid-air selection and selection against a hard surface. Teather and Stuerzlinger [53] compared 2D interfaces with solid surface and 3D interfaces with mid-air targets using Fitts' law. They found that the efficacy of spatial selections using input devices in 3D interactions in VR environments conforms to the Fitts' law model. In addition, Batmaz et al. [6] compared mid-air interactions with passive haptics with VR controllers and hand tracking. They found that the mid-air interactions were faster without passive haptics, potentially due to the vibration feedback that was provided.

2.3 Hand Gesture and Controller Input Metaphors

As VR evolves, so too is the way that users interact with their environments. The most common form of user input in VR is the motion tracked controllers provided with devices like the Oculus and HTC Vive headsets [2, 25, 45]. Pham et al. compared user performance in pointing using with a mouse versus a Vive controller or a 3D pen. They found that user performance was best with the 3D pen, showing that a precision grip offered the best pointing performance [41].

LaViola et al. divided the process of selecting object into six basic modes: grasping, pointing, surface, indirect, bimanual, and hybrid [24, 25]. Their particular focus was in grasping metaphors, which were divided further into hand-based and finger-based grasping. The former uses a single point tracked by a glove or a hand-held controller, while the latter tracks the individual's fingers separately. In addition, the Leap Motion gesture tracking system is a common low-cost method of tracking a user's hands and fingers [54]. Performance in gesture based selection was similar with this method to selection using the Microsoft Kinect [11]. Other studies have revealed that 2D selection with such devices offers lower performance than a mouse [8]. Similarly, earlier results from Johnsgard revealed that the mouse offered more efficient selection than using tracked VR glove-based input [22].

One must also consider the level of both visual and interaction fidelity offered by a given tracking technology. Li et al. evaluated how participants completed 3D pointing tasks with different offset conditions, and report that participants were the most efficient without an offset applied to their end effectors [26]. Hand models and avatars are another aspect of tracking users' end effectors. Pusch et al. conducted a study on hand visual fidelity, comparing a static 3D hand model to a realistically tracked 3D hand model [43]. Although the realistic hand model was preferred, it did not improve objective performance. Conversely, Ebrahimi et al. compared different levels of visual fidelity of self-avatar hands on near-field distance estimation, concluding that as the visual fidelity of the self-avatar increased so did the participants' spatial perception [17]. Batmaz and Stuerzlinger studied how movements were effected by rotational jitter, and they found that jitter noticeably affected error rate, throughput, and movement time at different thresholds [5].

2.4 Personal Space or Near-Field Interactions

Interaction spaces are divided into three areas: near-field or personal space (within participants' maximum arms reach or slightly beyond), action space (between 6 to about 30 m), and vista space (beyond 30 m) [12]. Near-field interactions require accurate depth perception in order to carry out fine motor actions. In personal space, monocular and binocular perceptual cues, stereoscopic viewing, retinal disparity and motion parallax are important factor in accurately presenting depth. Research shows that stereoscopic is most effective in a near-field space for motor actions in VR [21, 34].

In a related study, Machuca and Stuerzlinger studied virtual selection efficiency using a stylus as compared to a virtual pointing tasks in a large-screen stereoscopic display [3]. They found participants' selection performance was more erratic when moving along the viewing axis, but this discrepancy did not take place when in a real-world environment. This study also found that movement time was strongly influenced by the change in target depth. Lubos et al. analyzed user performance in different spatial locations relative to the user's viewpoint [28]. They found that the error rate was highest along the viewing axis as compared to movement axis within perpendicular menu plane. Finally, Batmaz et al. also showed that participants were more efficient along the lateral axis than they were in moving front-to-back along the viewing axis in near-field spatial selection in VR and AR [4]. These studies were all conducted using controller based input. Our work extends this research in systematically evaluating the effect of interaction fidelity of pointing metaphors as compared to controller based input on near-field spatial selection efficiency in VR.

3 Experiment Design and System Description

3.1 Research Questions and Hypotheses

The main contribution of this study was to examine the efficiency of spatial selection in personal space or near-field in two levels of fidelity of gesture based pointing as opposed to controller based interaction. Our research questions were as follows: 1. *To what extent is the efficiency of selection measured by the linear relationship between movement time and effective index of difficulty different between low-fidelity and high-fidelity pointing metaphors and controller input in personal space?* 2. *To what extent is the accuracy versus speed trade-off, as defined by Fitts' law Throughput, different between low-fidelity pointing, high-fidelity pointing and controller input in personal space spatial selection in VR?* To address these questions, we propose the following hypotheses:

H1: With regards to condition, the regression profiles (slopes and intercepts) and throughput of the participants' selection in the controller condition are expected to be different than either of the pointing conditions.

H2: With regards to condition, the regression profiles and throughput of the participants' selection in the low-fidelity condition are expected to be different than that of the high-fidelity pointing condition.

H3: With regards to personal space interactions, the regression profiles and throughput of the participants' selection of targets in the 60%, 75%, and 90% depth planes of the participants' maximum arm's reach are expected to be different.

H4: With regards to personal space interactions, the regression profiles and throughput of the participants' selection of targets *between* depth planes are expected to be different from the regression profiles and throughput *within* the same depth plane.

H1 is based on the expectation that participants' performance modeled by linear regression is expected to be characteristically different between the conditions of the experiment. **H2** is based on the expectation that the fidelity of gesture based tracking for pointing based input is expected to affect the participants' performance. **H3** is based on the expectation that the based on the biomechanics of reaching behaviors, participants are expected to easily reach targets in the middle of their personal space as compared to reaching to targets close to the critical reach boundary [14]. **H4** is relevant to experiment 2, where participants' target selection efficiency is expected to be different when participants are selecting targets within the same depth plane as compared to between different depth planes. Thus, we want to rigorously evaluate the efficiency and the speed-accuracy trade-off of two levels of fidelity of pointing based methods against controller based interaction on near-field spatial selection in VR, using the Fitts' law ISO 9241-9 framework.

3.2 Participants

We recruited 60 participants in Experiment 1 and 60 participants in Experiment 2. The participants who took part in experiment 1 and 2 were distinct and there was no overlap in participants whatsoever. Participants ages ranged from 19 to 28 years, and they were from diverse backgrounds and majors. All participants had 20/20 vision or corrected with contact lenses. All had normal motor functions, and had little to no experience with VR. As a between subjects experiment, participants were randomly assigned to one of the three conditions (Controller, Point-LF, or Point-HF).

3.3 Apparatus and Interaction Scenario

Our experiment simulation used a PC with an NVIDIA GeForce GTX 1080Ti graphics card. We used an HTC Vive VR headset with a refresh rate if 90 Hz and a horizontal field of view of 110 degrees. For user input, we implemented three conditions: the *Controller* condition used the Vive controller as the input device, the Low-Fidelity Pointing (*Point-LF*) condition used a Leap Motion gesture tracking device [54], while the High-Fidelity Pointing (*Point-HF*) condition used a Noitom Hi5 VR Glove [40] (See Figure 1). The simulation was developed using the Unity game engine [55].

Our conditions represent very popular interaction metaphors and techniques for the direct selection and manipulation of targets in the near-field [25]. Users can use tools like controllers or articulated virtual hand representations to point to targets in menu planes and select them in personal space interactions. Gesture tracked pointing interaction metaphors afford natural direct fine-motor interaction with targets, as compared to using a hand held tool to select targets in the Controller condition. However, the system fidelity of tracking varies a lot in gesture based input devices. Devices like the Leap Motion [54] use depth camera based tracking to users' hands, but these systems require line of sight with the participants hands and fingers and suffer from low latency and low frame rate issues [37]. Whereas glove based systems are more robust to tracking of hand and finger poses, and are more robust in terms of latency and framerate [27]. Therefore, it is important to evaluate how the fidelity of popular hand gesture based interaction metaphors affect the efficiency of near-field target selection as compared to controller based technique in VR.

The first step of the experiment procedure was a calibration phase. The HTC Vive system, the Hi5 Glove, and the Leap Motion input device's native tracking system's innate calibration routines were run. Then, in both pointing conditions, we first ensured that the self-avatar hand was the same size and scale as the participants' physical hand, by measuring the dominant and non-dominant hands and scaling the virtual hand's length, width and thickness to match that of the participant's physical hand. Thus, we ensured the virtual hand (including fingers and palm) in both pointing fidelity conditions were exactly the same and closely matched the dimensions of the participants' hand.

In the Point-LF condition, we measured the exact position and orientation of the Leap Motion device's origin to the HMD's Vive tracking system origin, and applied a transform in the scene graph that reported the Leap Motion tracker's hand position and orientation. In the Point-HF condition, the tracking system of the Hi5 Glove directly uses the HTC Vive tracker as part of its innate tracking, so the position and orientation of the virtual hand was accurately tracked by the Hi5 glove's tracking system. We found that the position and orientation calibration of the virtual hand was not necessary with the Hi5 glove.

Finally, in order to verify the size, scale, and pose of the virtual hand matched that of the real hand in both pointing conditions, we used visuo-haptic feedback method by running a Vive controller through the silhouette of the palm, back of the hand and fingers to insure that the physical and visual location of controller on the hand was as accurate as possible. When we noticed an offset, then the transform described above was adjusted to eliminate any mismatch in the pose between the virtual and the physical hand. For the controller condition, we ensured that the virtual Vive controller was in the same physical position on the dominant hand of the participant, using visuo-haptic feedback by using a second controller and running it on the surface of controller in the participant's hand.

In the interest of space, we refer the reader to the technical specifications of the different components of our experiment simulation HTC Vive [20], Noitom Hi5VRGlove [40] and Leap Motion [54] to the references provided. We conducted a system evaluation of latency and framerate in all three conditions (Controller, Point-HF, and Point-LF) using Niehorster et al.'s method [39]. The evaluation consisted of a study to measure 10 samples of latency and framerate for the different conditions for simple translational and rotational movements. The method involves using a high framerate camera, such as a Go Pro sports camera, with a capture framerate of 240FPS that was attached to a stand to capture the real controller or hand as well as the virtual representation of the same through a monocular view port of the HMD (with the lens removed). For measuring translation latency, the hand or controller was moved in a straight line multiple times capturing several trial data. For measuring rotational latency, the hand or controller was rotated about the vertical axis multiple times. Footage was captured for each of these movements, and analyzed using video editing software (Adobe Premiere). Using the known framerate of the high-speed camera, we calculated the latency from the number of frames it took for the time of movement/rotation of the physical hand or controller to the corresponding movement/rotation of the virtual counterpart over multiple trials. Our analysis revealed that the mean framerate for the different conditions were as follows: Point-HF (70.5Hz), Point-LF (62Hz), Controller (73Hz). The mean end-to-end latency of the conditions were as follows: Controller (Pos. lag and Ori. Lag = 33.33ms), Point-LF (Pos. lag = 70.83ms, Ori. Lag=87.5ms), Point-HF (Pos. lag = 29.2ms, Ori. lag = 37.5ms).

We used spherical targets for the selection tasks in a manner similar to prior Fitts' law research [2,4,5,28,53]. The selected target was rendered as red, and all the others were rendered as white. In the Point-HF and Point-LF conditions, participants used their index fingers on the dominant hand, and in the Controller condition they used the Controller on the dominant hand to select the target.

3.4 Procedure

In each experiment, after the informed consent process, the participants were randomly assigned to one of 3 conditions (Controller, Point-LF, or Point-HF). They completed a demographics questionnaire and were led to the experiment simulation. There, the participants inter-pupillary distance (IPD) and maximum arms reach were measured and used to calibrate the HMD/stereo graphics pipeline and target presentation distance, as described above.

To ensure that differences in hand gesture did not influence the results, all participants in the Pointing condition were instructed to point using their index finger straightened out with the other fingers folded inwards towards the palm of the hand. Also, in order to ensure that targets were not rendered outside of the critical reach boundary (or the maximum reach without shoulder rotation), participants were told not to rotate or extend their shoulders in providing the maximum arms reach measurement at the start of the study. This ensured that targets were not rendered in the range between critical reach boundary (maximum arms reach distance without shoulder motion) and absolute reach boundary (maximum arms reach distance with shoulder motion), which requires additional motor effort [10,33].

Participants were then given 2 practice trials, after which the participants started selection of targets presented in each trial configuration as described in the section below in VR.

3.5 Trials and Target Presentation

Our experiment included the following independent variables, as illustrated in figure 2: *Interaction Metaphor:* Controller, Point-LF, Point-HF, *Target depth:* 60%, 75%, or 90% of maximum arm's reach, *Target distance:* 5 cm, 9 cm, and 18 cm, *Target width:* 0.65 cm, 0.85 cm, 1.05 cm.

Interaction metaphor was assigned between-subjects, while all other independent variables were assigned within-subjects. We

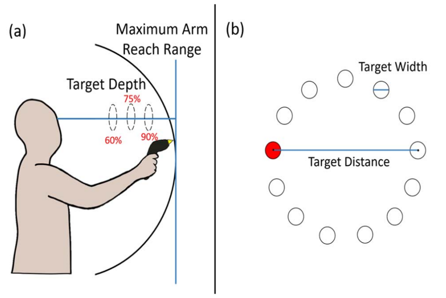

Figure 2: Target presentation depth (a), distance and width (b).

arranged targets in a circle, in accordance with the ISO 9241-9 standard [50]. The widths and distances were chosen to match those used by Teather and Stuerzlinger, who compared selection performance with targets in 2D or 3D in the near-field [53]. Participants completed 13 selection trials in each condition, and completed two sets of all conditions. Thus, in total, each participant performed 702 trials (2 sets x 3 depth planes x 3 target distances x 3 target widths x 13 trials).

In Experiment 1, we presented each arrangement of 13 targets in each combination of distance, width and depth, selected once in random order (i.e., a total of 27 combinations). Experiment 2 employed a similar design, with each arrangement of targets being selected in a random combination of widths and distances. However, the main differences is that in a given arrangement, each target was placed at a random depth that varied from one target to the next, so each target in a set of 13 could be within 60%, 75%, or 90% of the maximum arm's reach. As a result, this task simulated selecting targets randomly within and between depth planes and both in the Fitts' law movement axis as well as the viewing/depth axis, as often done in selection of targets in hierarchical 3D menus where the targets may vary in different depth planes [13, 42]. In experiment 2, the movement axis (amplitude) for each trial was defined as the center-to-center axis between the start and destination target (either along the viewing or depth axis, or between targets on the same depth plane), based on which Ae was calculated as the participant's movement during selection projected along that axis.

4 Experiment 1 Results

4.1 Multiple Regression Analysis

We conducted a multiple regression analysis to evaluate how effective index of difficulty of the targets, target depth, and the interaction metaphor conditions predict the efficiency of movement time to the targets. We first conducted an analysis of standardized on the data to identify any outliers, using which data that were beyond $\pm$ 3.0 of the standardized residuals were removed. The final standardized residual minimum was -1.50 to +1.50. Tests to see if the data met the assumptions of collinearity indicated that multicollinearity was not a concern (Effective Index of Difficulty, Tolerance = 1.0, VIF = 1.0; Conditions, Tolerance = 1.0, VIF = 1.0; Target Depth, Tolerance = 1.0, VIF = 1.0). The data met the assumptions of independent errors (Durbin-Watson value = 1.47). The histogram of standardized residuals indicated that the data contained approximately normally distributed errors, as did the P-P plot of standardized residuals, which showed that the data were close to a linear regression profile. The scatterplot of standardized residuals showed that the data met the assumptions of homogeneity of variance and linearity, as well as the data met the assumptions of non-zero variance.

We conducted a multiple regression to examine if effective index of difficulty, condition (gesture vs. controller), depth of target by arms reach (60%, 75%, and 90% of maximum arms reach) predicted movement time. We found a significant regression equation F(3, 1559) = 382.57, $p < 0.001$, with an $R^2 = 0.43$. Participants' predicted movement time was equal to $-203ms + 199.42 \times ID_e - 16.74 \times Condition + 417.58 \times TargetDepth$; where effective index of difficulty (ID_e) was measured in bits, condition 1 was controller, 2 was gesture condition, 3 was glove condition, and depth of target ($TargetDepth$) was coded as 60% = 0.60, 75% = 0.75 and 90% = 0.90. Movement time increased by 199.42ms for every bit of effective index of difficulty, 417.58ms for an adjacent difference in target depth within maximum arms reach, and a difference in -16.74ms was calculated on average between two adjacent conditions. All three independent variables, effective index of difficulty ($p< 0.001$), condition (p = 0.0054) and depth of target ($p < 0.001$) were significant predictors of movement time.

In order to evaluate the significant interaction effects, the continuous independent variable of ID_e were mean centered in order to eliminate multicollinearity effects and then the interaction terms were computed such as Effective ID x Target Depth, Effective ID x Condition, Target Depth x Condition, Effective ID x Target Depth x Condition, which were added to the model in a hierarchical multiple regression. The regression model with the interaction variables was found to be significant, F(7, 1559) = 165.88, $p < 0.001$, with an R^2 = 0.41 (with the change in R^2 of 0.003). The model did not reveal any significant interaction effects.

By condition, the linear regression equation for the controller condition (R^2 = 0.395) is movement time is equal to $21.85ms + 207 \times ID_e$, the linear regression equation for the low fidelity pointing condition (R^2 = 0.43) is movement time is equal to $95.26ms + 212 \times ID_e$, and the linear regression equation for the high fidelity pointing condition (R^2 = 0.375) is movement time is equal to $120ms + 178 \times ID_e$ (See Figure 3(a)). By target depth, the linear regression equation for targets at 60% of the participants' maximum arms reach (R^2 = 0.41) is movement time equals to $76.26ms + 187 \times ID_e$, the linear regression equation for targets at 75% of the participants' maximum arms reach (R^2 = 0.43) is movement time equals to $36.15ms + 205 \times ID_e$, and the linear regression equation for targets at 90% of the participants' maximum arms reach (R^2 = 0.37) is movement time equals to $117ms + 206 \times ID_e$ (See Figure 3(b)).

4.2 Comparative Analysis of Regression Coefficients and Throughput

In order to evaluate if there are any systematic statistical differences in R^2, Slopes and Intercepts between conditions, random depth in the relationship between index of difficulty and movement time in experiment 1, we subjected these variables extracted from each participant's data to a 2 (condition) x 3 (target depth) repeated measures ANOVA analysis with condition and random target depth as independent variables. The analysis of slopes revealed a significant main effect of condition F(2, 186) = 4.88, p = 0.009, η^2 = 0.10. The mean slopes in the Point-HF condition (M=182.71, SD=66.35) was significantly lower than the mean slope for the Point-LF condition (M=212.67, SD=76.11) p = 0.007, and the mean slope for the Controller condition was in the middle (M=205.76, SD=66.83).

We examined throughput, which is calculated as the effective index of difficulty divided by movement time. Throughput is a measure of performance and a tradeoff between speed and accuracy, and is measured in bits per second. The throughput values were subjected to a 3 (condition) x 3 (target depth) mixed model ANOVA analysis. The ANOVA analysis revealed a significant main effect of condition F(2, 186) = 18.6, $p < 0.001$, η^2 = 0.17, a main effect of target distance F(2, 186) = 13.0, $p < 0.001$, η^2 = 0.13 (see Figure 3(c)). Post-hoc pairwise comparisons using Tukey's HSD method on conditions revealed that the throughput in the Point-LF condition (M=4.82b/s, SD=0.60) was significantly lower than the Controller condition (M=4.82b/s, SD=0.75) $p < 0.001$, as well as the Point-HF condition (M=4.96b/s, SD=0.68) $p < 0.001$. Post-hoc pairwise comparisons using Bonferroni method on the target depth revealed that the throughput in the selection performance at 90% of

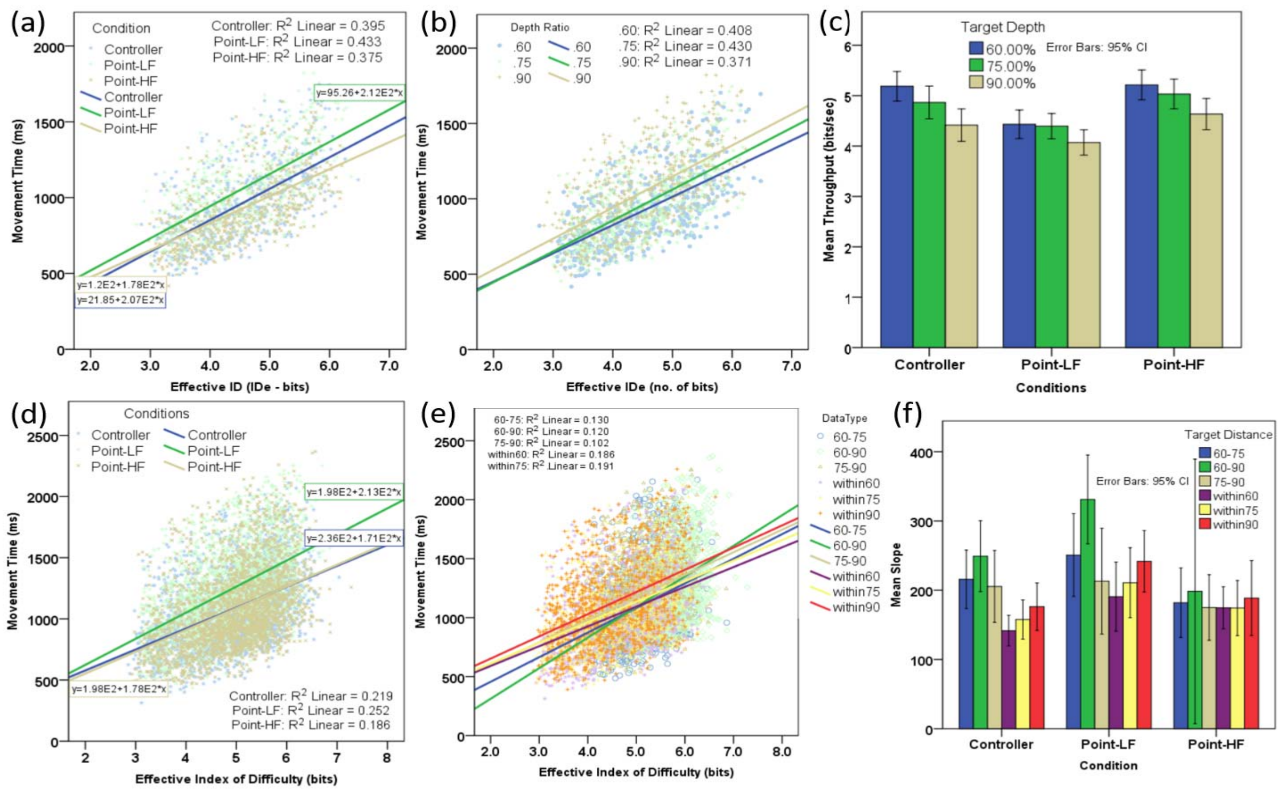

Figure 3: (a) Expt 1: Movement time (MT) vs. Effective ID (ID_e) graph showing the linear regression profiles of the different conditions. (b) Expt 1: Movement time (MT) vs. Effective ID (ID_e) graph showing the linear regression profiles of the different depth of target presentation. (c) Expt 1: Graph of mean Throughput (bits/sec) by the different conditions and target depth. (d) Expt 2: Movement time (MT) vs. Effective ID (ID_e) graph showing the linear regression profiles of the different conditions. (e) Expt 2: Movement time (MT) vs. Effective ID (ID_e) graph showing the linear regression profiles of the performance in the different target depth. (f) Expt 2: Graph of mean slope of participant's regression profiles of their performance by presented target distance.

the maximum arms reach (M=4.37b/s, SD=0.67) was significantly lower than selection performance at 60% of the maximum arms reach (M=4.94b/s, SD=0.72) $p < 0.001$, as well as 75% of the maximum arms reach (M=4.76b/s, SD=0.68) $p = 0.002$.

5 Experiment 2 Results

5.1 Multiple Regression Analysis

In this experiment, participants randomly reached to targets between 60%, 75%, 90% of their maximum arm reach within the same reciprocal tapping target configuration. In a manner similar to experiment 1, the selection data from the participants between the Controller, Point-LF and Point-HF selection metaphors were collected, and an analysis of standard residuals was carried out on the data to identify any outliers. After removing outliers, the range of the standard residuals were from a min of -1.5 to max +1.5. Tests to see if the data met the assumptions of collinearity indicated that multicollinearity was not a concern (Effective Index of Difficulty, Tolerance = 0.722, VIF = 1.38; Condition, Tolerance = 0.721, VIF = 1.38; Target Distance, Tolerance = 0.996, VIF = 1.00). The data met the assumption of independent errors (Durbin-Watson value = 1.23). We also verified that the data met the assumptions of normality, homogeneity of variance and linearity.

A multiple regression was conducted to examine if effective index of difficulty, condition (Controller vs. Point-LF vs. Point-HF), random depth of target by arms reach (60%, 75%, and 90% of maximum arms reach) predicted movement time. A significant regression equation was found F(3, 6680) = 719.39 $p < 0.001$, with an R^2 = 0.25. Participants' predicted movement time is equal to $406.37ms + 174.92 \times ID_e – 85.46 \times Condition + 10.35 \times TargetDepth$, where effective index of difficult is measured in bits, condition 1 is Controller, 2 is Point-LF and 3 is Point-HF, and depth of target presented is encoded as within 60% = 1, within 75% = 2, within 90% = 3, between 60% and 75% = 4, between 75% and 90% = 5, and between 60% and 90% = 6. Movement time increased by 174.92ms for every bit of effective index of difficulty, 10.35ms for a difference in target depth within maximum arms reach, and movement time decreased by -85.46ms for a difference between conditions. Three independent variables, effective index of difficulty ($p < 0.001$), condition ($p < 0.001$), and target depth ($p = 0.008$) were all significant predictors of movement time.

In order to evaluate the significant interaction effects, the independent variables were mean centered and the interaction terms centered IDe x Target Depth, centered IDe x Condition, Target Depth x Condition, and centered IDe x Target Depth x Condition were derived and added to the model. The hierarchical regression model with mean centered independent variables and interaction terms were found to be significant F(7, 6688) = 310.52, $p < 0.001$, with an R^2 = 0.26 (with the change in R^2 of 0.01). In additional to the significant independent variables, the second level model found a significant condition by effective index of difficulty interaction ($p = 0.040$).

In order to examine the interaction effects, we examined them via multiple simple linear regression analysis. By condition, the linear regression equation for the Controller condition is $236ms + 171 \times ID_e$ ($R^2 = 0.22$), the linear regression equation for the Point-LF condition is $198ms + 213 \times ID_e$ ($R^2 = 0.25$), and the linear regression equation for the Point-HF condition is $198ms + 178 \times ID_e$ (R^2 = 0.19) (see Figure 3(d)).

By target distance, participant selection performance within 60% of arms reach is $255ms + 168 \times ID_e$ (R^2 = 0.19), within 75% of max arms reach is $263ms + 174 \times ID_e$ (R^2 = 0.19), and within 90% of max arms reach is $276ms + 188 \times ID_e$ (R^2 = 0.19). For selection performance between 60% and 75% is $39ms + 209 \times ID_e$ (R^2 = 0.13), between 75% and 90% is $215ms + 191 \times ID_e$ (R^2 = 0.10), and between 60% and 90% of max arms reach is $-206ms + 259 \times ID_e$ (R^2 = 0.12) (see Figure 3(e)). As we have seen in experiment 1, quantitatively and qualitatively the regression profiles for 60% and 75% of the max arm's length reaches are similar, but the performance for 90% of max arms reach is different requiring more time at that plane of selection. The performance of random selection between 60% to 75% of max arms reach has a more gradual slope and intercept closer to 0. The slope of the reaches between 60% to 75% and between 75% to 90% are similar, but the intercept for reaches between 75% to 90% of the maximum arms reach is high at 215ms. Whereas, the intercept for the reaches between 60% to 90% has the lowest intercept and the highest slope as compared to any other random depth of target selection. The 75% to 90% reaches for selection has the highest intercept among the random distances of selection, and has a gradual slope similar to performance at other random distances.

5.2 Comparative Analysis of Regression Coefficients and Throughput

In order to evaluate if there are any systematic statistical differences in R^2, Slopes and Intercepts in the participants' performance between conditions and random depth in the relationship between effective index of difficulty and movement time in experiment 2, we subjected these variables extracted from each participant's data to a 3 (condition) x 6 (target depth) mixed model ANOVA analysis with condition as a between-subjects variable and random target depth as a within-subjects variables. The analysis of slopes revealed a significant main effect of condition $F(2, 360) = 5.92$, $p = 0.003$, η^2= 0.04, and a main effect of random target depth $F(2, 360) = 3.10$, $p = 0.01$, η^2= 0.05. Post-hoc pairwise comparisons using Tukey's HSD revealed that mean slopes in the Point-LF condition (M=239.67, SD=130.61) was significantly higher than the Controller condition (M=191, SD=91.16) (see Figure 3(f)). Post-hoc pairwise comparisons using Bonferroni method revealed that the slope for the target selection distance between 60% and 90% of the maximum arms reach was the largest (M=259.50, SD=408.17) and was significant higher than target selection within 60% of the maximum arms reach (M=169.05, SD=78.38) $p = 0.007$ and within 75% of the maximum arms reach (M=180.90, SD=88.12) $p = 0.033$.

A similar analysis of the intercepts of the participants' selection performance also revealed a significant main effect of random target depth $F(5, 360) = 3.24$, $p = 0.007$, η^2= 0.05. Intercepts of the participants' target selection performance of targets between 60% and 90% of the participants' maximum arms reach was different and lower than any other presented distance. Post-hoc pairwise comparisons using Bonferroni method revealed that the mean intercepts of the participants' target selection between 60% and 90% depth of the participants' maximum arms reach (M=-190.23ms, SD=1506) was significantly lower than target selection within 60% of the maximum arms reach (M=254.28ms, SD=368.06) $p = 0.022$, target selection within 75% of the maximum arms reach (M=237.17ms, SD=352.86) $p = 0.034$, and target selection within 90% of the maximum arms reach (M=224.74ms, SD=397.66) $p = 0.045$.

In experiment 2, we also examined throughput, which is calculated as the effective index of difficulty divided by movement time. Throughput is a measure of performance and a tradeoff between speed and accuracy, and is measured in bits per second. The throughput values were subjected to a 3 (condition) x 6 (target depth) mixed model ANOVA analysis. The ANOVA analysis revealed a significant main effect of condition $F(2, 360) = 49.45$, $p < 0.001$, η^2= 0.22, and a main effect of target depth $F(5, 360) = 8.29$, $p < 0.001$, η^2= 0.11 (see Figure 4). Throughput in the Point-LF condition (M=3.97b/s, SD=0.45) was significantly lower than the Controller condition (M=4.63b/s, SD=0.65) $p < 0.001$ and the Point-HF condition (M=4.61, SD=0.70) $p < 0.001$. Interestingly throughput in the random target distance between 60% and 75% (M=4.67b/s, SD=0.72) was the highest, and throughput in the within 90% target distance (M=4.04b/s, SD=0.60) was the lowest. Throughput for target selection between 60% and 75% of the participants' maximum arms reach was significantly higher than target selection between 75% to 90% of the participants' maximum arms reach $p = 0.015$, target selection within 75% of the participants' maximum arms reach (M=4.36b/s, SD=0.61) $p = 0.048$, and target selection within 90% of the participants' maximum arms reach $p < 0.001$. Throughput of trials in the within 90% target distance (M=4.04b/s, SD=0.61) was significantly lower than other target presentation distances ($p < 0.001$) except between 75% and 90% of the participants' maximum arms reach.

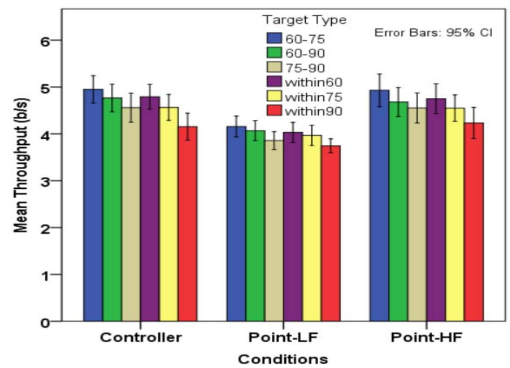

Figure 4: Graph of Throughput (bits/sec) in the different conditions of experiment 2 by different target distances presented as a proportion of the participants' maximum arms reach

6 Discussion

Overall, we found that as effective index of difficulty increases, movement time also linearly increased in all three conditions. For each input condition and for each depth ratio, the movement time had an upward slope with effective index of difficulty. This is important as it shows that regardless of input fidelity and the depth ratio, Fitts' law was able to successfully model the relationship between target amplitude and width.

Our first hypothesis was that *the slopes, intercepts, and throughput of the participants' selection in the controller condition are expected to be different than that of either of the pointing conditions*, and our second hypothesis was that *the slopes, intercepts, and throughput of the participants' selection in the low-fidelity pointing condition are expected to be different than that of the high-fidelity pointing condition*. Both of these hypotheses were partially supported. In the multiple regression analysis for both experiments, the input condition was a significant predictor of movement time. Concerning the first hypothesis, our analysis only revealed that the slope was different between the controller and pointing conditions in Experiment 2, and even then, only the low-fidelity condition had a different slope. This information tells us that as the index of difficulty rose in Experiment 2, the movement time rose faster in the low-fidelity condition than in the controller condition, implying a higher level of performance in the controller condition than in the low-fidelity pointing condition. There was no significant difference in the intercepts between any of the three conditions, which is a novel result.

Concerning the second hypothesis, the slopes in the two pointing conditions were different in Experiment 1 and 2, as evidenced by Condition being a significant predictor of movement time and by significant difference in slopes in the coefficient analysis. In both experiments, the slope was higher in the low-fidelity pointing condition than it was in the high-fidelity pointing condition, implying that the robustness of hand gesture tracking had a great impact on user performance in selection tasks efficacy. Finally, our analysis showed that throughput was lowest in the low-fidelity pointing condition and that the high-fidelity pointing condition was similar to using a VR controller. Both of these results have the possible implication that with sufficiently robust tracking, pointing based input could yield higher performance in a manner similar to a standard controller. In experiment 2, slopes were different. This may be due to the influence of varying target depth randomly in both the lateral(movement) axis and forward(viewing) axis motion. These results help to explain the findings of Batmaz et al. [6], whose experiment showed superior performance with a controller than with hand tracking when participants were selecting targets presented in a square grid at only a single depth. However, their experiment used the Leap Motion, which is the same tool that we used in our low-fidelity pointing condition. The differences found by that study combined with the findings listed above in our study can indicate that the level of tracking fidelity, potential due to the difference in latency, is the most likely cause of the lowered performance in low-fidelity pointing condition. With a sufficiently robust tracking mechanism, it is possible that hand tracking can yield similar results as a standard consumer-level VR controller in near-field spatial selection.

Our third hypothesis was that *the slopes, intercepts, and throughput of the regression profiles of the participants' selection of targets in depths of 60%, 75%, and 90% of the maximum arm's reach are expected to be different.* This hypothesis was also partially supported. The multiple regression analysis revealed that target depth was a significant predictor of movement time. While there was no difference in the intercepts for the regression models, our analysis revealed that as index of difficulty increased the subsequent increase in movement time was more pronounced when selecting objects at 90% of users' maximum arm's reach than it was at 60% or 75% of maximum arm's reach, regardless of the end-effector representation (pointing or controller). Participants had the lowest throughput at 90% of the arm's reach, and it was significantly different than 60% and 75%. These findings support the notion that performance is superior when the targets are selected towards the middle of the users' personal space, than when the targets are closer to the users' reach envelope or maximum arm's reach in the absence of shoulder motion. These findings support relevant research that shows that near-field depth perception may be poor closer to our maximum reaching boundary [16, 49]. However, the results of this study imply a threshold for how far selection of objects and menu buttons in the near-field can be performed before performance starts to degrade, in common commodity controller based or virtual pointing based interaction metaphor using a glove or a depth camera based input.

Our fourth hypothesis was that *the slopes, intercepts, and throughput of the regression profiles of the participants' selection of targets between depth planes are expected to be different from the profiles within the same depth plane.* This hypothesis was partially supported. The slope was the highest when moving from 60% of maximum arm's reach to 90% of maximum arm's reach. Specifically, it was higher than the conditions within 60% and within 75% of the maximum arm's reach. These results are in line the findings of Machuca and Stuerzlinger [3], who showed that movement time goes up proportionally with the change in target depth in addition to the index of difficulty and that performance is worse when target depth changes rather than when it is constant. However, unlike their work, we found that the average movement time at 90% of the maximum arm's reach was slower than the average movement time moving between 60% and 75% of maximum arm's reach. In addition, at 90% of maximum arm's reach, average throughput was lower than any of the conditions where the user moved from one depth plane to another. This could imply that there may be a limit to how change in target depth can explain movement time, and that the effects of target depth can be diminished when the user's arm is extended closer to the maximum reach or critical boundary without shoulder motion [33].

Regarding system differences between the Point-LF (Leap Motion) vs. Point-HF (Hi5 glove), the reason we used the two different systems is to empirically examine the differences between two popular commodity interaction metaphors that differ in system fidelity (gesture tracking quality, latency and framerate) for direct gesture based input on the efficacy of selection in the natural or native manner in which they are used in VR interaction. Similar empirical evaluations comparing between input devices have been conducted in Fitts' Law research and selection/manipulation studies, i.e. [2, 7, 28, 53]. The purpose of our research is to provide a rich baseline of data on how selection efficiency and throughput (speed-accuracy tradeoff) is affected overall by the fidelity of popular commodity gesture tracked direct interaction metaphors and controller based interaction in near-field target selection in VR.

Limitations: Our experiment compared three methods of common interaction techniques that afforded either a controller or two levels of fidelity of a virtual articulated hand based pointing, on the efficiency of near-field selection. But, there are other specialized interaction metaphors such as pen or stylus based input that we have not yet studied in this paradigm. Our results show that tracking fidelity in pointing metaphors (perhaps the difference in latency) could potentially be the root cause of the difference in performance in personal space selection efficiency in our experiment. However, the current experiment design is unable to tease apart if the result we are seeing is due to inherent differences in the tracking quality or latency.

7 Conclusion and Future Work

Our contribution shows that the interaction and system fidelity of pointing input metaphors significantly impacts personal space target selection efficiency and the speed-accuracy tradeoff. High-fidelity pointing metaphors can perform at similar efficiency and throughput to the standard VR controller in personal space target selection. This important finding suggest that potentially minimizing latency and improving tracking in popular commodity gesture based input systems can greatly enhance the efficiency of near-field spatial selection in VR. Although, on one level our results can seem intuitive, on the other, there was a lack of data on how the fidelity of pointing metaphors affected near-field or personal space target selection efficiency and speed-accuracy tradeoff. Our research provides this much needed data to the research, development and consumer community. Our contribution also shows that regardless of the input method employed or the interaction fidelity, target selection is more efficient at 60% (towards the middle) of the participants' maximum arms reach and at movements between 60% and 75% of the participants' maximum arms reach. This suggests that with regards to menu placement and hierarchical target selection in personal space depth planes, 3D buttons and target selection is more efficient in the middle of the participants' arms reach rather than at the edge. Future work will focus on raycasting vs. pointing vs. controller based target selection in personal space XR interactions.

8 Acknowledgements

The authors would like to thank the participants of our studies for their time and effort. This work was supported in part by the US National Science Foundation (CISE HCC) under Grant No. 2007435.

REFERENCES

[1] F. Argelaguet Sanz and C. Andujar. A survey of 3d object selection techniques for virtual environments. 05 2013.

[2] S. Babu, M.-H. Tsai, T.-W. Hsu, and J.-H. Chuang. An evaluation of the efficiency of popular personal space pointing versus controller based spatial selection in vr. In *ACM Symposium on Applied Perception 2020*, pp. 1–10, 2020.

[3] M. D. Barrera Machuca and W. Stuerzlinger. The effect of stereo display deficiencies on virtual hand pointing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2019.

[4] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr? In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 585–592. IEEE, 2019.

[5] A. U. Batmaz and W. Stuerzlinger. Effects of 3d rotational jitter and selection methods on 3d pointing tasks. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1687–1692. IEEE, 2019.

[6] A. U. Batmaz, X. Sun, D. Taskiran, and W. Stuerzlinger. Hitting the wall: Mid-air interaction for eye-hand coordination. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–5, 2019.

[7] D. Brickler, R. J. Teather, A. T. Duchowski, and S. V. Babu. A fitts' law evaluation of visuo-haptic fidelity and sensory mismatch on user performance in a near-field disc transfer task in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 17(4):1–20, 2020.

[8] M. A. Brown, W. Stuerzlinger, and E. J. de Mendonça Filho. The performance of un-instrumented in-air pointing. In *Graphics Interface*, GI '14, p. 59–66, May 2014. doi: doi/10.5555/2619648.2619659

[9] R. A. Burno, B. Wu, R. Doherty, H. Colett, and R. Elnaggar. Applying fitts' law to gesture based computer interactions. *Procedia Manufacturing*, 3:4342–4349, 2015.

[10] S. K. Card, J. D. Mackinlay, and G. G. Robertson. A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems (TOIS)*, 9(2):99–122, 1991.

[11] D. Carvalho, M. Bessa, L. Magalhães, and E. Carrapatoso. Performance evaluation of gesture-based interaction between different age groups using fitts' law. In *Proceedings of the XVI international conference on human computer interaction*, pp. 1–7, 2015.

[12] J. E. Cutting and P. M. Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of space and motion*, pp. 69–117. Elsevier, 1995.

[13] M. M. Davis, J. L. Gabbard, D. A. Bowman, and D. Gracanin. Depth-based 3d gesture multi-level radial menu for virtual object manipulation. In *2016 IEEE Virtual Reality (VR)*, pp. 169–170. IEEE, 2016.

[14] B. Day, E. Ebrahimi, L. S. Hartman, C. C. Pagano, A. C. Robb, and S. V. Babu. Examining the effects of altered avatars on perception-action in virtual reality. *Journal of Experimental Psychology: Applied*, 25(1):1, 2019.

[15] P. S. Dukes, A. Hayes, L. F. Hodges, and M. Woodbury. Punching ducks for post-stroke neurorehabilitation: System design and initial exploratory feasibility study. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 47–54. IEEE, 2013.

[16] E. Ebrahimi, B. Altenhoff, L. Hartman, J. A. Jones, S. V. Babu, C. C. Pagano, and T. A. Davis. Effects of visual and proprioceptive information in visuo-motor calibration during a closed-loop physical reach task in immersive virtual environments. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 103–110, 2014.

[17] E. Ebrahimi, L. S. Hartman, A. Robb, C. C. Pagano, and S. V. Babu. Investigating the effects of anthropomorphic fidelity of self-avatars on near field depth perception in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–8. IEEE, 2018.

[18] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

[19] M. J. Fu, A. D. Hershberger, K. Sano, and M. C. Çavuşoğlu. Effect of visuo-haptic co-location on 3d fitts' task performance. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3460–3467. IEEE, 2011.

[20] HTC. Vivehtc vive vr headset, 2021.

[21] N. W. John, N. I. Phillips, L. a. Cenydd, S. R. Pop, D. Coope, I. Kamaly-Asl, C. de Souza, and S. J. Watt. The use of stereoscopy in a neurosurgery training virtual environment. *Presence: Teleoperators and Virtual Environments*, 24(4):289–298, 2016.

[22] T. Johnsgard. Fitts' law with a virtual reality glove and a mouse: Effects of gain. In *Graphics interface*, pp. 8–8. Canadian Information Processing Society, 1994.

[23] L. Kohli, M. C. Whitton, and F. P. Brooks. Redirected touching: The effect of warping space on task performance. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 105–112. IEEE, 2012.

[24] J. J. LaViola Jr. Bringing vr and spatial 3d interaction to the masses through video games. *IEEE Computer Graphics and Applications*, 28(5):10–15, 2008.

[25] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[26] J. Li, I. Cho, and Z. Wartell. Evaluation of cursor offset on 3d selection in vr. In *Proceedings of the Symposium on Spatial User Interaction*, pp. 120–129, 2018.

[27] H. Liu, Z. Zhang, X. Xie, Y. Zhu, Y. Liu, Y. Wang, and S.-C. Zhu. High-fidelity grasping in virtual reality using a glove-based system. In *2019 international conference on robotics and automation (icra)*, pp. 5180–5186. IEEE, 2019.

[28] P. Lubos, G. Bruder, and F. Steinicke. Analysis of direct selection in head-mounted display environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 11–18. IEEE, 2014.

[29] P. M. Fitts and J. R. Peterson. Information capacity of discrete motor responses. *Journal of experimental psychology*, 67:103–12, 03 1964. doi: 10.1037/h0045689

[30] I. S. MacKenzie and W. Buxton. Extending fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 219–226. ACM, 1992.

[31] I. S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1633–1636. ACM, 2008.

[32] I. S. MacKenzie and R. J. Teather. Fittstilt: the application of fitts' law to tilt-based interaction. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pp. 568–577. ACM, 2012.

[33] L. S. Mark, K. Nemeth, D. Gardner, M. J. Dainoff, J. Paasche, M. Duffy, and K. Grandt. Postural dynamics and the preferred critical boundary for visually guided reaching. *Journal of Experimental Psychology: Human Perception and Performance*, 23(5):1365, 1997.

[34] J. P. McIntire, P. R. Havig, and E. E. Geiselman. Stereoscopic 3d displays and human performance: A comprehensive review. *Displays*, 35(1):18–26, 2014.

[35] R. P. McMahan. *Exploring the effects of higher-fidelity display and interaction for virtual reality games*. PhD thesis, Virginia Tech, 2011.

[36] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics*, 18(4):626–633, 2012.

[37] C. Mizera, T. Delrieu, V. Weistroffer, C. Andriot, A. Decatoire, and J.-P. Gazeau. Evaluation of hand-tracking systems in teleoperation and virtual dexterous manipulation. *IEEE Sensors Journal*, 20(3):1642–1655, 2019.

[38] D. Natapov, S. J. Castellucci, and I. S. MacKenzie. Iso 9241-9 evaluation of video game controllers. In *Proceedings of Graphics Interface 2009*, pp. 223–230. Canadian Information Processing Society, 2009.

[39] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017.

[40] Noitom. Hi5vrglove, 2021.

[41] D.-M. Pham and W. Stuerzlinger. Is the pen mightier than the controller? a comparison of input devices for selection in virtual and augmented reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2019.

[42] M. Pourmemar and C. Poullis. Visualizing and interacting with hierarchical menus in immersive augmented reality. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–9, 2019.
[43] A. Pusch, O. Martin, and S. Coquillart. Effects of hand feedback fidelity on near space pointing performance and user acceptance. In *2011 IEEE International Symposium on VR Innovation*, pp. 97–102. IEEE, 2011.
[44] Y. Y. Qian and R. J. Teather. The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 91–98. ACM, 2017.
[45] A. Ramcharitar and R. J. Teather. A fitts' law evaluation of video game controllers: Thumbstick, touchpad and gyrosensor. In *Proceedings of the 2017 chi conference extended abstracts on human factors in computing systems*, pp. 2860–2866. ACM, 2017.
[46] I. Realsense. Intel realsense, 2021.
[47] V. Schwind, J. Leusmann, and N. Henze. Understanding visual-haptic integration of avatar hands using a fitts' law task in virtual reality. In *Proceedings of Mensch und Computer 2019*, pp. 211–222. 2019.
[48] S. Serafin, A. Adjorlu, N. Nilsson, L. Thomsen, and R. Nordahl. Considerations on the use of virtual and augmented reality technologies in music education. In *2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR)*, pp. 1–4. IEEE, 2017.
[49] G. Singh, J. E. Swan, J. A. Jones, and S. R. Ellis. Depth judgments by reaching and matching in near-field augmented reality. In *2012 IEEE Virtual Reality Workshops (VRW)*, pp. 165–166. IEEE, 2012.
[50] 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):751–789, 2004. Fitts' law 50 years later: applications and contributions from human-computer interaction. doi: 10.1016/j.ijhcs.2004.09.001
[51] C. Systems. CuberGlove iii, 2021.
[52] R. J. Teather, D. Natapov, and M. Jenkin. Evaluating haptic feedback in virtual environments using iso 9241–9. In *2010 IEEE Virtual Reality Conference (VR)*, pp. 307–308. IEEE, 2010.
[53] R. J. Teather and W. Stuerzlinger. Pointing at 3d targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 87–94. IEEE, 2011.
[54] Ultraleap. Leapmotiondigital worlds that feel human, 2021.
[55] Unity. Unityreal-time development platform — 3d, 2d vr ar engine, 2021.
[56] Vicon. Viconvicon tracking, 2021.