The Effect of Visual Depth on the Vergence-Accommodation Conflict on 3D Selection Performance within Virtual Reality Headsets

Mohammad Raihanul Bashar[1*], Mayra Donaji Barrera Machuca[2], Wolfgang Stuerzlinger[3], Anil Ufuk Batmaz[1]

[1*]Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. W., Montreal, H3G 1M8, Quebec, Canada.

[2]Faculty of Computer Science, Dalhousie University, 6050 University Avenue, Halifax, B3H 4R2, Nova Scotia, Canada.

[3]School of Interactive Arts & Technology, Simon Fraser University, 8888 University Dr W, Burnaby,, Burnaby, V5A 1S6, British Columbia, Canada.

*Corresponding author(s). E-mail(s): mohammadraihanul.bashar@mail.concordia.ca;
Contributing authors: mbarrera@dal.ca; w.s@sfu.ca; ufuk.batmaz@concordia.ca;

Abstract

Prior studies have shown that the vergence-accommodation conflict (VAC) negatively affects the interaction performance in Virtual Reality (VR) and Augmented Reality (AR) systems, particularly as object depth increases. This paper examines user selection performance across six different visual depths. Through a study closely resembling prior research, eighteen participants participated in an ISO 9241:411 task with six different depth distances. We observed that with higher depth values, selection times increased while the throughput performance of the participants decreased. Based on this finding, we propose a Fitts' law model based on the focal distances, which models pointing times in VR and AR systems with substantially higher accuracy. We hope that our findings aid developers in creating 3D user interfaces for VR and AR that offer better performance and an improved user experience.

Keywords: Virtual Reality, Vergence-Accommodation Conflict, Stereo Deficiencies, Distal Pointing, Ray Casting, Selection, Fitts' Law

1 Introduction

Various industries, including education, engineering, arts, healthcare, and entertainment, are increasingly adopting Virtual Reality (VR) and Augmented Reality (AR) systems. This adoption is facilitated by the introduction of self-contained, wireless, and lightweight head-mounted displays (HMDs) such as the Quest 2/3/Pro and Pico 4 for VR and the HoloLens 2 and Magic Leap 2 for AR. Other HMDs, like the XR-4 Varjo and the Pimax 4k, even support 4k resolution with minimal latency in rendering and tracking [1], alongside wide fields of view [2]. All these devices are capable of delivering high-definition graphics at high frame rates and offer an adjustable interpupillary distance (IPD), i.e., the distance between the center of the pupils of the eyes. Yet, regardless of these technological advancements, studies indicate that current VR and AR HMDs still struggle to present all depth cues accurately, as identified in previous work [3–10]. Depth perception, the ability to judge spatial distances and object placement in a three-dimensional space, relies on a combination of monocular (e.g., size, motion parallax, accommodation, or texture gradient) and binocular cues (e.g., stereopsis or vergence). While modern HMDs simulate some of these cues effectively, the lack of accurate representation of accommodation cues can lead to perceptual inaccuracies, which impact pointing performance [11].

One critical issue that affects current stereo displays is the vergence-accommodation conflict (VAC), a mismatch between the vergence (the inward or outward turning of the eyes to align on an object) and accommodation (the adjustment of the eye's lens to optically focus on the object by changing its shape). This conflict arises because most stereo displays render content on a fixed focal plane, requiring the eyes to focus at a static depth while vergence cues indicate different spatial distances. This mismatch disrupts natural depth perception, leading to degraded performance in spatial tasks and discomfort for users [4, 12, 13]. The VAC can significantly affect behavioral performance, particularly in 3D selection tasks, as users experience increased cognitive and perceptual strain when attempting to combine conflicting depth cues. Studies have shown that the VAC hampers the users' ability to quickly and accurately perceive an object's spatial position, and as a result, users experience difficulties while interacting with targets positioned at varying distances [11, 13].

Several studies have examined how stereo displays with a single focal plane degrade interaction performance. For instance, Batmaz et al.[13] built a multifocal display to compare user performance under both multifocal and singlefocal conditions. Their results confirmed that the VAC impairs speed and precision, although the 3D pointing task they used involved only very limited variations in visual depth. To explore the effect of the VAC further, a new experimental methodology to specifically isolate the effect of the VAC on 3D selection within a singlefocal display system using raycasting for targets beyond arm's reach was introduced [14]. According to their results, the varying VAC conditions significantly increased the time required for selection and decreased throughput performance.

One common factor in all this past work [11, 13–15] is that the user performance is affected by the target visual depth, e.g., the distance that the target is away from the user. Yet, previously proposed 3D models for selection [11, 16, 17] ignore the effect of the target visual depth. For example, Murata and Iwase [16] included 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 [17] added inclination and azimuth angles to Fitts' law, but their formulation mixed target visual depth and target position in front of the user. Only Barrera and Stuerzlinger [11] added a parameter to the Shannon formulation of Fitts' Law [18] representing the change in target depth. However, their linear model does not account for the nonlinear optical properties of accommodation (optical focus) in the human eye when target distances vary [19]. Thus, this work aims to identify a 3D Fitts' Law formulation that captures the effect of the target visual depth more accurately. Such a formulation can aid 3D user interface developers in designing better systems that improve the user experience in VR and AR.

In this paper, we ran a user study with 18 participants to examine the impact of the VAC on 3D selection performance for targets placed at varying visual depths with the raycasting interaction method. We aimed to better identify the relationship between visual depth and user performance to propose a new 3D Fitts' law formulation that models movement time (MT) more accurately. Our results show a relationship between visual depth and user selection performance in MT and throughput (THP). Then, based on the insight that pointing at targets with varying depth likely depends on the change in diopters needed to focus on said targets in the real world, we propose a new 3D Fitts' law formulation that uses the variation in diopters (ViD) to better model the data. Building on the insights embedded in this formulation, the design of future 3D user interfaces can benefit from a more accurate human movement model based on visual depth.

2 Related Work

This section discusses previous work on depth perception in virtual environments (VE). Then, we review previous studies on 3D selection for far-away targets, e.g., distal pointing. Finally, we present past work on modeling 3D target selection.

2.1 Physiological Constraints on Depth Perception in HMDs

Depth perception is the ability of a person to identify an object's position in space using different pictorial and non-pictorial cues [20]. In this paper, we focus on non-pictorial depth cues, such as stereopsis, motion parallax, convergence, and accommodation, that humans utilize when pointing at targets [4, 21–23]. Specifically, the issue caused by the fixed focal distance of the stereo display systems used in VR and AR HMDs to show 3D content is called the vergence-accommodation conflict (VAC). The VAC affects the performance of the human visual system [24, 25] by creating depth perception issues [22, 26] and visual fatigue due to focal and vergence differences [27]. Moreover, it also increases the user's cognitive load [28].

The *zone of comfort* [29] refers to the range in terms of accommodation within which the human eye can maintain clear vision without significant effort. This range typically spans about ± 0.5 diopters (D) around the current focal plane [30, 31]. A *diopter* is the unit of measurement used to express the optical power, and it is measured as the reciprocal of focal length in meters. In this paper, D represents the eye's optical focus that changes via the eye's lens accommodation. The eyes can focus comfortably within this zone without triggering visual fatigue or discomfort. Studies [32, 33] suggest that maintaining accommodation demands within this ± 0.5 D range can help minimize visual discomfort and fatigue, especially in prolonged viewing scenarios. Situations disrupting the accommodation-vergence relationship, such as in stereoscopic displays with fixed focal planes or 3D imagery that exceeds the comfort zone, lead to visual strain [34, 35].

The young human eye can shift its focus from distant objects (infinity) to as close as 6.5 cm from the eye [36]. This substantial change in focal power, approximately 15 D, occurs because the eye muscles contract, allowing the lens to thicken and increase its focusing power. This adjustment can happen in as little as 224 ± 30 *milliseconds* in bright light, but can also take substantially longer. Further, the amplitude of accommodation declines with age [37, 38]. Therefore, VR systems that demand frequent changes in perceived depth must consider these biological constraints or risk inducing excessive eye strain and diminishing user comfort. By understanding and mitigating the VAC, designers of VR/AR HMDs can more effectively align the display's focal requirements with the human visual processes. Properly handling stereopsis, motion parallax, convergence, and accommodation supports more accurate depth perception, reduces visual discomfort, and enhances overall immersion in 3D interactive environments [39, 40]. Our study investigates one of these depth perception challenges by analyzing the VAC's impact on 3D selection performance, providing insights into how the user experience can be improved in VR/AR HMDs.

2.2 Impact of Visual Depth on Distal 3D Pointing

When selecting 3D objects beyond arm's reach, one of the most frequently used interaction techniques is **raycasting** [41, 42]. Raycasting allows users to point to a ray from an input device, such as a controller, and confirm selections, e.g., with a button press, making it effective for remote targets. Raycasting mimics real-world pointing behavior, such as using a laser pointer, making it intuitive for first-time users [43]. It is commonly used in VR applications, which makes it familiar to many users. Compared to virtual hand manipulation, raycasting enables faster object selection while also reducing fatigue associated with mid-air hand-reaching techniques [43].

Previous work has found that pointing at targets at different visual depths affects performance negatively. For example, Teather and Stuerzlinger [44] showed that varying target depth affects performance. Janzen et al. [45] found that pointing performance for targets at depths between 110 and 330 cm is also affected. They also identified an effect of the user's distance to the screen, which would indicate an issue related to the focal distance. For 3D virtual hand/wand pointing, past work found that lateral and depth movements were different when selecting targets displayed on a large stereo display [11] or AR and VR headsets [46].

Finally, Batmaz et al. [14] identified an effect of the VAC when selecting distal 3D targets with *raycasting* using an HTC Vive with a focal plane at ≈75 cm. Their results showed that for targets on the HMDs' focal plane, i.e., in a condition without the VAC, participants were able to perform better in terms of time, errors, and THP than with a constant or varying VAC. If performance decreases in VAC conditions, then depth must be explicitly considered in a predictive model, rather than treating 3D selection like a 2D task. However, they did not propose a new 3D Fitts' law that accommodates target depth. This raises a question: Can we systematically model these depth effects into a predictive formula? Our work builds on these insights, examining whether diopter changes and the effect of VAC can be modeled and thus predicted for general situations.

2.3 3D Fitts Law

Fitts' Law [47] is a widely used model for predicting performance in pointing tasks by relating MT to the ratio of distance to target width. This ratio is represented by the index of difficulty (ID), which captures how distance and target size combine to influence task complexity. A refined version, ISO 9241-411 [48], combines speed and accuracy into THP to make the measurement less dependent on user strategies. Here, we utilize the angular version of Fitts' Law [18] that is appropriate for distal 3D pointing using raycasting, which effectively considers only the user's wrist rotation to select a target, see Equation 1. To calculate the angular ID (ID_A), α defines the angular distance between targets, and ω the angular target width. The constant k is a relative weight [18], typically set to 1:

$$MT = a + b \cdot \log_2 \left(\frac{\alpha}{\omega^k} + 1 \right) = a + b \cdot ID_A \tag{1}$$

We define THP for angular movements also based on effective measures. See Equation 2, where α_e represents the effective angular distance, i.e., the actual angular movement distance to the target, and ω_e the effective angular target width, the distribution of the angular selection coordinates, calculated as $\omega_e = 4.133 \times SD_x$.

$$ID_e = \log_2 \left(\frac{\alpha_e}{\omega_e^k} + 1 \right) = \log_2 \left(\frac{\alpha_e}{(4.133 \cdot SD_x)^k} + 1 \right) \tag{2}$$

Yet, one key issue of using the traditional Fitts' law formulations, including the angular version of Fitts' Law [18], is that they do not always correctly predict 3D movement times or throughput in stereo displays [11, 49]. Past work has tried to address this issue by proposing novel 3D Fitts' Law variations [11, 16, 17]. Most of these works focus on the difference in arm movement between 2D and 3D target selection. For example, Murata and Iwase [16] and Cha and Myung [17] considered the diagonal arm movement when selecting targets in depth. Yet, past work has found that the biomechanics of the arm movement is not the only influence on 3D target selection and that the VAC also affects it [50]. Finally, Barrera and Stuerzlinger [11] considered the change in target depth, but focused primarily on peri-personal 3D target selection, where depth-related perceptual challenges (e.g., the VAC) are more pronounced. Consequently, their findings may have limited applicability to distal target selection tasks, where the VAC is generally less pronounced, but depth perception becomes less reliable due to reduced disparity cues, and precise motor control is required for accurate selection. Thus, we extend this past work by proposing a 3D Fitts' Law model that considers distal target selection.

3 Motivation & Hypotheses

While previous studies [15, 51] investigated the VAC under varifocal and single-focal conditions, they primarily compared results at only two or three different depth distances. Their findings indicate that user performance, particularly movement time, increases with increasing depth distance, even though the perceived target size is the same at different depth distances [11]. These results suggest that user movement time may vary with different depth distances, raising the question of whether the Fitts' law model applies (robustly) to systems that suffer from the VAC. Thus, in this paper, we investigated the following hypothesis:

- **H1: When targets are displayed with the same perceptual size and angular lateral distance, user movement time varies at different depth distances.** Given that previous work has shown that the VAC significantly affects user performance, we expect that as the depth of targets moves away from the system's focal plane, user performance significantly decreases.
- **H2: A variation of Fitts' law that accounts for changes in visual target depth better models user performance under VAC.** Researchers have demonstrated that Fitts' law is robust under various depth conditions and have used it to model human movement time. We believe that incorporating changes in visual target depth into the Fitts' law formulation will enhance its predictive power and better model user performance across different depth distances.

4 User Study

The objective of this study is to examine the impact of the VAC on 3D selection performance for targets placed at varying visual depths. Specifically, we aimed to explore how the VAC affects pointing and selection times, error rates, and throughput as participants interact with virtual objects positioned at different visual depths. Unlike previous studies that focused on linear distances, we opted to use the change in diopters (the measure of optical focus) as the primary independent variable to model the variation in visual depth.

We based our approach on findings from previous work [34], which showed that the primary factor affecting performance in the presence of a VAC is the additional time it takes participants to resolve the VAC, i.e., the *conflicting* visual depth information. By expressing depth changes in diopters rather than simple linear distances, we seek to isolate the effects of VAC more directly and aim to produce a model that better correlates the effect of visual depth variation with selection performance. We also hypothesized that increasing diopter differences (i.e., moving objects closer to or further from the focal plane) would significantly affect user performance, specifically regarding selection time and throughput.

4.1 Methodology

Participants:

An a priori power analysis was conducted using G*Power (version 3.1.9.7) [52] to determine the minimum sample size needed to test the study's hypothesis. Results indicated that a sample of 12 participants would achieve 80% power to detect a large effect ($\eta^2 = 0.14$), at an $\alpha = .05$ significance level with a repeated-measures ANOVA. To ensure robustness and account for potential dropouts or variability, we thus recruited a total of 18 participants (3 female, 15 male), which is in line with Amini et. al.'s recommendation for Fitts' law studies [53]. Their ages ranged from 20 to 40 years ($\mu = 26.38, \sigma = 4.48$). Participants had diverse backgrounds, including engineering, advertising, and customer service, providing a wide range of perspectives for the evaluation. All participants were right-handed, with 16 reporting right-eye dominance and 2 left-eye dominance. Twelve had normal vision, while the remaining six had a corrected-to-normal vision; none reported color vision deficiencies or visual impairments of binocular vision. Regarding previous VR experience, seven had never used VR, three had experienced it 1–5 times, and eight had used it more than five times.

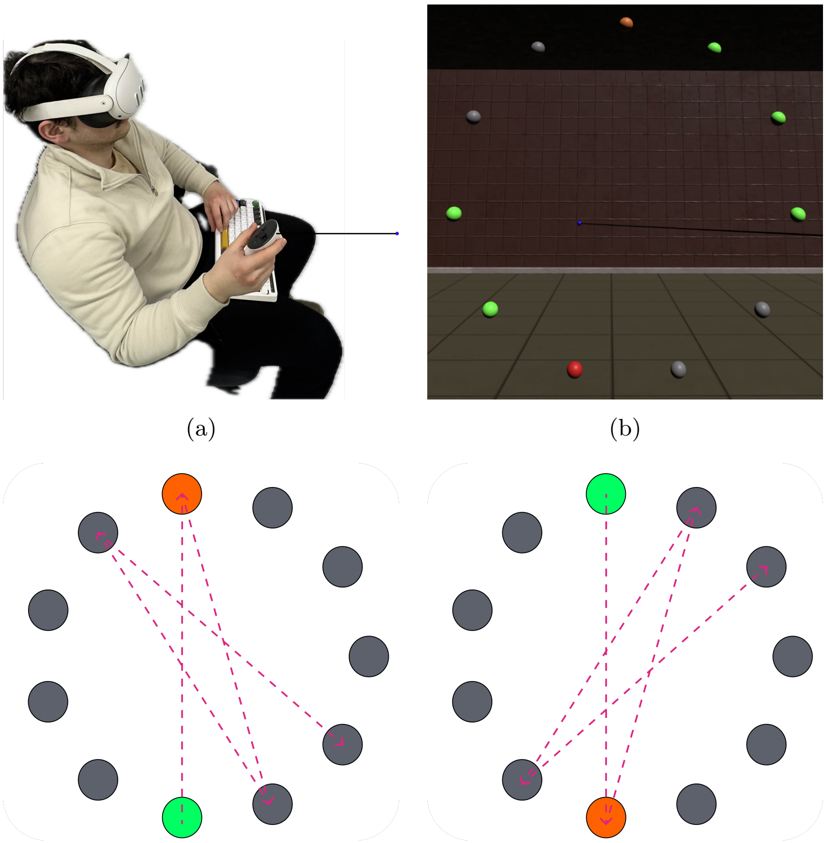
Apparatus:

We ran the experiment on a PC equipped with an Intel® Core™ i9-13900KF processor (3.0 GHz), 32 GB of RAM, and an NVIDIA GeForce RTX 4070 graphics card. We used the Oculus Quest 3 HMD for the VR setup, which has a focal distance of 1.3 m [54], a per-eye resolution of 2064×2208 pixels, and a horizontal and vertical field of view (FOV) of 110° and 96°, respectively. We developed the virtual scene using Unity version 2021.3.24 and the Oculus Unity Integration SDK (v57.0.1-deprecated) and displayed the experiment scene in the HMD via Meta Quest Link.

User Study Procedure:

The user study was conducted on-site in an indoor laboratory on a university campus. Before starting the experiment, participants completed a questionnaire, giving their consent and demographic details. An experimenter then explained the procedure in detail and assisted participants in putting on and wearing the HMD. We instructed participants to hold the controller with their dominant hand and to keep their hands reasonably close to their shoulders to facilitate the selection of nearby targets Figure 1a. Ray casting was implemented by rendering a virtual line from the center of the controller in the participants' pointing position, with the cursor (a 0.5 cm diameter sphere) placed at the tip of the ray where it intersected with geometry. Participants used their non-dominant hand to press the space bar on the keyboard in front of them to select targets. We choose this selection method to mitigate the

"Heisenberg Effect" [55], i.e., where the mechanical force applied to a selection button can shift the controller and cursor, introducing additional errors.



(a) (b) (c) (d)

Fig. 1: Experimental Setup: a) A photo of a user during the experiment, illustrating also how the cursor/ray is represented in the virtual environment (VE). The space key on the keyboard, placed on the participant's lap, is used to confirm the selection of the target the user aims at. b) The view inside the headset: the current target is highlighted in orange. A green sphere indicates a previously selected (successful) target, and a red one illustrates a previous miss. The image appears slightly shifted due to it being the left-eye camera view in Unity. c) ISO 9241-9 reciprocal selection task. Participants are shown targets arranged in a circular pattern. The dashed lines represent the sequence of target transitions, alternating clockwise. d) This image represents the sequence of targets alternating in an anti-clockwise manner. Targets advance in the pattern indicated by the arrows. Arrows show the ordering for the first four targets.

Participants completed the task in a static virtual environment, similar to the scenes used in previous research [46], but customized for our experimental tasks and design. The experimental setup consisted of a room with a uniform grid texture background as a spatial reference to improve immersion without affecting the main task. During the experiment, participants were primarily cognitively focused on the targets, as selecting the correct target was the main task. Given that their attention was directed towards the targets rather than the environment, any pictorial depth cues present in the scene were unlikely to substantially influence performance. Furthermore, the environment remained fixed and identical for all participants across all conditions, ensuring that no confounding variables were introduced due to variations in background or textures.

The ISO 9241-411 multidirectional selection task is a standardized method for evaluating pointing device performance. Participants are presented with targets arranged in a circular layout. The task requires participants to select highlighted targets one by one. Targets are highlighted in a predetermined sequence, alternating across the circular arrangement, to ensure consistency and increase predictability in target presentation. Figure 1c illustrates a sequence of four targets in a clockwise pattern, showing how they were presented and selected by participants.

Similar to the ISO 9241-411 multidirectional selection task [48], participants were presented with a circle of 11 gray spheres and instructed to select the sphere highlighted in orange: the first target was randomly selected, and subsequent targets alternated in a clockwise or counterclockwise sequence, which was again selected randomly, see Figure 1b. When the virtual cursor approached within 0.5 cm of a sphere, its color changed from orange (or gray, for non-targets) to blue, providing visual feedback by highlighting the intersected sphere [56]. Participants selected the target by pressing the space key on the keyboard with their non-dominant hand. A selection was considered successful, or a "hit," when the controller ray came within 0.5 cm of the highlighted target, causing it to change color from orange to blue (highlighted target, gray for non-targets). This was followed by a space key press, which turned the sphere green. An unsuccessful selection, or "miss," occurred when they selected an incorrect target, which turned the sphere red and was accompanied by an error sound for auditory feedback [57]."

At the start of each round, the system randomly chose the first target (orange) from the 11 spheres. The subsequent target was then selected in either a clockwise or counterclockwise direction, directly across the circle of targets. With 11 targets in total, the next target alternated across the circle until all targets had been either hit or missed.

Participants remained seated throughout the experiment. While we did not restrict their head position, the system used the participants' current head position and rotation to realign the targets after each circle of selections, ensuring they always appeared directly in front of them. This adjustment occurred after every 11 target selections, maintaining consistent alignment with the participant's view, even if they shifted slightly during the experiment. After each round of targets, participants were encouraged to take breaks if needed; however, none chose to do so. The seated position in our study provided a relaxed posture, and we did not observe any significant need for participants to rest their arms.

After the experiment, participants completed a post-experiment survey, where they shared their preferences regarding the experimental conditions and reported their mental and physical fatigue levels. On average, participants completed the experiment in 25 to 30 minutes.

To investigate our hypothesis, we purposefully varied the target depth. Given that the focal distance of the Quest 3 is 133 cm [54], we included targets both in front and behind that distance. Given that optical distances are measured in diopters (D, defined as 100 cm/distance), we used targets that were spaced linearly in terms of optical focus between 1.5 and 0.25 D, which corresponds to a nonlinear spacing in Euclidean distances between 66.6 cm and 400 cm, see Figure 2.

In this work, we used the same angular target sizes and distances as in previous research [58]. However, the focal plane of the HMD used in this study differs from those in earlier studies (133 cm for the Quest 3, whereas previous work had used HTC Vive Pro, which has a focal plane of ≈75 cm [14]), Further, we decided to vary the depth distance with 0.25 D precision, so we had to utilize a different set of target sizes and distances. To isolate the effect of the VAC from perspective-related factors, we adjusted the target sizes and distances so that spheres maintained the same visual angle regardless of distance. This creates the illusion that all targets are uniformly sized, regardless of their actual distances in each condition Figure 2. Exemplary target sizes and distances are shown in Table 1, with the entire table available in the supplementary material.

Table 1: The table below outlines the target sizes and distances used in this study. The two left-most columns represent the angular target sizes and distances, as applied in [58]. The two right-most columns display the corresponding index of difficulty (ID).

Angular Target Size(°)	Angular Target Distance (°)	Target Distance (cm) at 66.66 cm depth	Target Size (cm) at 66.66 cm depth	Target Distance (cm) at 400 cm depth	Target size (cm) at 400 cm depth	Euclidean ID	Angular ID (k=1)
1.45	30.51	18.18	1.82	109.09	10.91	3.46	4.5
2.42	30.51	18.18	3.04	109.09	18.25	2.80	3.8
3.39	30.51	18.18	4.27	109.09	25.63	2.39	3.3
1.49	25.61	15.15	1.83	90.91	10.97	3.21	4.2
2.48	25.61	15.15	3.05	90.91	18.30	2.58	3.5
3.47	25.61	15.15	4.28	90.91	25.66	2.18	3.1
1.42	35.30	21.21	1.83	127.27	10.96	3.66	4.7
2.36	35.30	21.21	3.04	127.27	18.26	2.99	4.0
3.31	35.30	21.21	4.28	127.27	25.69	2.57	3.5

4.2 Experimental Design

We conducted a two-factor within-subjects user study with six depth distances, i.e., **Depth** conditions ($6_{Depth} = 6$ conditions). Participants were asked to select targets under these six depth distances, with each condition counterbalanced using a Latin Square. We collected the following dependent variables:

- **Selection Time:** The duration (in seconds) from the moment the target appears (i.e., highlighted) until the participant selects it by pressing the space bar, then we reset the timer for the next selection. This metric reflects speed in target acquisition.

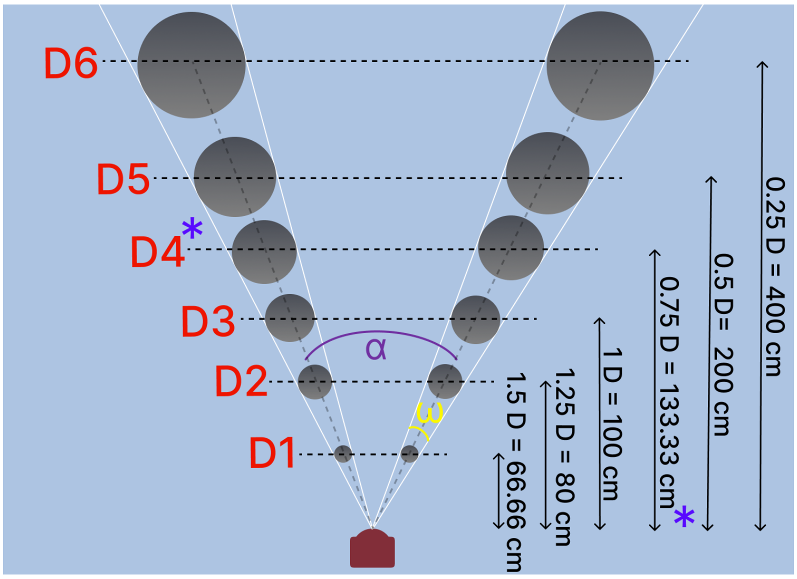


Fig. 2: Illustration of the perceived target sizes. In our experiment, targets always appeared to be the same size, regardless of how far away they were. This figure also illustrates the linear spacing of accommodation distances in terms of diopters for different target distances. The angle α (denoted in purple) represents the angular distance between targets, and ω (shown in yellow) represents the angular size (or width) of the target as perceived by the user [18]. The focal plane of the HMD ($D4$) is indicated with an asterisk (*).

- **Error Rate:** The number of incorrect selections (misses).
- ID_e: A refined measure of task difficulty based on the actual selection distances and widths (Equation 2).
- (Effective) **Throughput:** A combined measure of speed and accuracy derived from Fitts' law. Based on the *index of difficulty* (ID) and movement time (MT), computed as:
$$THP = \frac{ID_e}{MT} \tag{3}$$
- **Angular Throughput:** Similar to effective throughput but measured in angular terms rather than linear distance. Uses angular distance (α) and angular target width (ω) to calculate the index of difficulty and throughput.
- SD_x: The standard deviation of selected points relative to the target's center along the x-axis (the horizontal direction in the coordinate system). It indicates selection accuracy—lower values indicate that participants consistently aimed closer to the target center.
- **Angular** SD_x: The angular equivalent of SD_x, measuring the angular spread of the selections around the target center. This metric captures depth-independent accuracy by incorporating angular deviations.

To vary the task difficulty, we used 9 ID_As, using all combinations of three **angular target sizes** (3_{ATS}) and three **angular target distances** (3_{ATD}). In total, each participant performed $6_{Depth} \times 9_{ID_A} \times 11$ repetitions = 594 trials. We recorded a total of 10,692 trials across all participants.

In our experiment, targets always appeared to be the same size, regardless of their physical distance from the user. This was achieved by maintaining the angular size of the targets constant across depth conditions. Using Kopper's [18] angular size equation, we calculated the target size as a function of its physical distance and angular size, ensuring that all targets were perceived to have an identical size (Equation 4):

$$Size = 2 \times Distance \times tan(\frac{AngularSize}{2}) \tag{4}$$

5 Results

We pre-processed and plotted the data through JMP and analyzed it using two-way Repeated Measures (RM) ANOVA in SPSS 29.0. We considered data normally distributed if Skewness (S) and Kurtosis (K) values were within ±1 [59, 60]. We first applied a log transformation if the data were not normally distributed. If this still did not yield a normal distribution, we used the Aligned Rank Transform (ART) [61, 62] on the original data before performing ANOVA using ARTool. ARTool automatically checks the correctness of the dataset, and we verified that all data is appropriate for use with ART. We also checked that all the effects except for the effect for which the data were aligned were stripped out. For post-hoc analyses, we used the Bonferroni method (Sidak for ART) and applied Huynh-Feldt correction when $\varepsilon < 0.75$. The figures show the mean in the graphs, and the error bars represent the standard error of the mean. The results are shown in Table 2. For *brevity*, we only provided details of the results related to the VAC.

Table 2: Data Analysis Results

	Depth	ID	Depth * ID
Time	$\boldsymbol{F(5,85) = 5.884, p < 0.001, \eta^2 = 0.257}$	$\boldsymbol{F(8,136) = 74.760, p < 0.001, \eta^2 = 0.815}$	$\boldsymbol{F(9.93,168.85) = 1.783, p < 0.01, \eta^2 = 0.095}$
Error rate	$\boldsymbol{F(5,85) = 2.7991, p < 0.022, \eta^2 = 0.141}$	$\boldsymbol{F(8,136) = 15.235, p < 0.001, \eta^2 = 0.473}$	$\boldsymbol{F(40,680) = 1.986, p < 0.001, \eta^2 = 0.105}$
Throughput	$\boldsymbol{F(5,85) = 7.074, p < 0.001, \eta^2 = 0.294}$	$\boldsymbol{F(8,136) = 11.200, p < 0.001, \eta^2 = 0.397}$	F(32.56,553.43) $= 0.978$, $p < 0.505, \eta^2 = 0.054$
Angular Throughput	$\boldsymbol{F(5,85) = 7.504, p < 0.001, \eta^2 = 0.306}$	$\boldsymbol{F(8,136) = 36.469, p < 0.001, \eta^2 = 0.682}$	F(11.02,187.1) $= 1.039$, $p < 0.414, \eta^2 = 0.058$
ID_e	F(5,85) $= 0.882$, $p < 0.497, \eta^2 = 0.049$	$\boldsymbol{F(8,136) = 44.307, p < 0.001, \eta^2 = 0.723}$	$\boldsymbol{F(36.42,619.09) = 1.967, p < 0.001, \eta^2 = 0.104}$
SD_x	$\boldsymbol{F(5,85) = 259.397, p < 0.001, \eta^2 = 0.938}$	$\boldsymbol{F(8,136) = 3.174, p < 0.001, \eta^2 = 0.661}$	$\boldsymbol{F(38.28,650.69) = 1.967, p < 0.001, \eta^2 = 0.104}$
Angular SD_x	F(5,85) $= 1.215$, $p < .309, \eta^2 = 0.067$	$\boldsymbol{F(8,136) = 31.316, p < 0.001, \eta^2 = 0.648}$	$\boldsymbol{F(32.81,557.79) = 1.855, p < 0.001, \eta^2 = 0.098}$

5.1 Selection Time Results

Time was normally distributed after log transformation ($S = 0.56, K = 0.44$). The ANOVA revealed a significant main effect of **Depth** on selection time, as indicated in Table 2 and visualized in Figure 3. Participants took notably longer to complete selections at higher depth distances (lower diopter values). This finding aligns with prior VR research [14, 46, 63] on user performance outside the HMD's focal plane. Additionally, the **ID** had a significant effect: tasks with higher ID values took more time, consistent with Fitts' law.

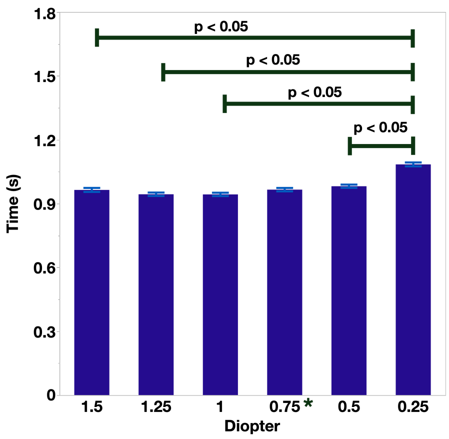


Fig. 3: Selection time across different diopter levels. The graph shows the mean selection time for targets at each depth distance. The error bars, representing the standard error of the mean, are present but not easily distinguishable because the variations in the data were minimal, resulting in very small error bars. This indicates high consistency in participant performance across trials.

5.2 Error Rate Results

We did not observe normality for error rate data ($S = 2.88, K = 6.32$) even after log transformation, so we analyzed it using ART. The results identify that there is a significant effect of **Depth** on error rate, Table 2. Participants made significantly more errors at greater depths (further than the focal plane), see Figure 4. The main effect of **ID** was also significant, Table 2, indicating that tasks with a higher index of difficulty led to higher error rates.

5.3 ID_e Results

The effective index of difficulty dependent variable was normally distributed ($S = -0.05, K = 0.53$). We did not observe any significant effect of Depth on ID_e. However, there was a significant effect of **ID**, Table 2, indicating that tasks with a higher index of difficulty had significantly higher ID_e values. Additionally, the interaction between

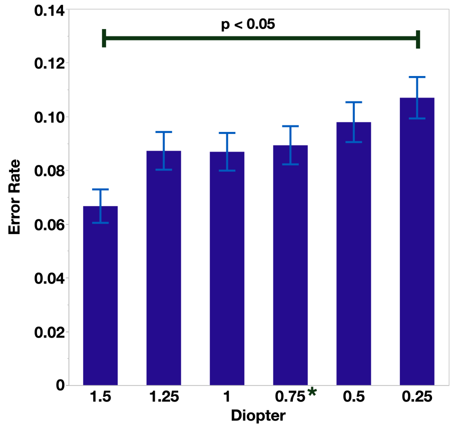


Fig. 4: RM ANOVA results for error rate based on changes in diopters.

Depth and ID was significant, suggesting that the effect of depth on ID_e was influenced by the base difficulty of the task.

5.4 Throughput Results

Throughput was normally distributed ($S = 0.07, K = 0.05$). The analysis revealed a significant main effect of **Depth** on throughput, see Figure 5. This indicates that as the depth increased, the throughput decreased significantly, suggesting that participants performed tasks more slowly and/or less accurately at higher depths. The **ID** results were also significant, Table 2, where lower throughput was observed for a higher index of difficulty.

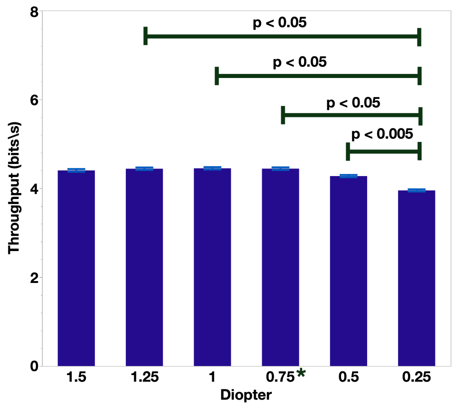


Fig. 5: Throughput across depth conditions. The bars represent the mean throughput values for each depth distance, with error bars indicating the standard error of the mean. Due to the small variability in the data, the error bars are not prominently visible but are present and accounted for in the analysis.

5.5 Angular Throughput Results

Angular throughput was normally distributed ($S = 0.08, K = -0.14$). The results in Figure 6 illustrate that participants' angular throughput significantly decreased at greater **Depths**, Table 2. The results for **ID** were also significant, Table 2, demonstrating that as the index of difficulty increased, angular throughput decreased significantly.

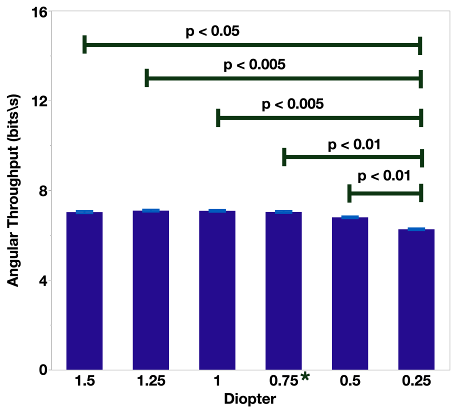


Fig. 6: Angular Throughput (AngTHP) Across different levels of Diopter: Significant Comparisons based on RM ANOVA. Due to the small variability in the data, the error bars are not prominently visible but are present and accounted for in the analysis.

5.6 SD_x Results

The standard deviation along the task axis was normally distributed after log transformation ($S = 0.37, K = -0.40$). We observed a significant difference for the **Depth** conditions Table 2. The results in Figure 7 highlight that the accuracy of the participants was significantly better for lower depth values. **ID** results were also significant, Table 2, indicating it had a substantial impact on the accuracy of the participants. Moreover, the interaction between Depth and ID was also significant, suggesting that the effect of depth on SD_x depends on the index of difficulty.

5.7 Angular SD_x Results

The standard deviation along the task axis was normally distributed after log transformation ($S = 0.12, K = 0.98$). We did not observe any significant effect of **Depth** on Angular SD_x, Table 2, meaning changes in angular distance corresponding to changes in diopters did not significantly affect the user selection accuracy. The result for **ID** was significant, Table 2.

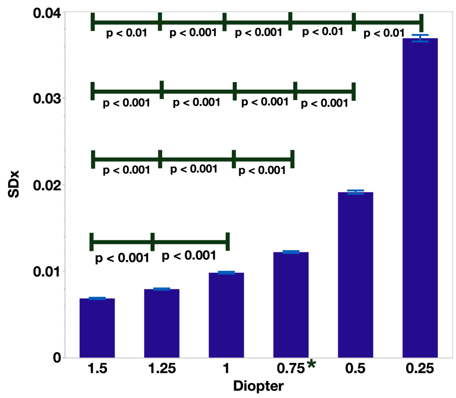


Fig. 7: Participants selection accuracy (SD_x) results along task axis: Significant Comparisons from RM ANOVA

Table 3: Fitts' law analysis for each depth distance

Diopter	a	b	R^2
D_1 (1.50 D = 66.6 cm)	0.35	0.21	0.63
D_2 (1.25 D = 80 cm)	0.26	0.24	0.94
D_3 (1.00 D = 100 cm)	0.35	0.2	0.93
D_4 (0.75 D = 133.3 cm)	0.34	0.22	0.93
D_5 (0.50 D = 200 cm)	0.42	0.2	0.92
D_6 (0.25 D = 400 cm)	0.31	0.27	0.91

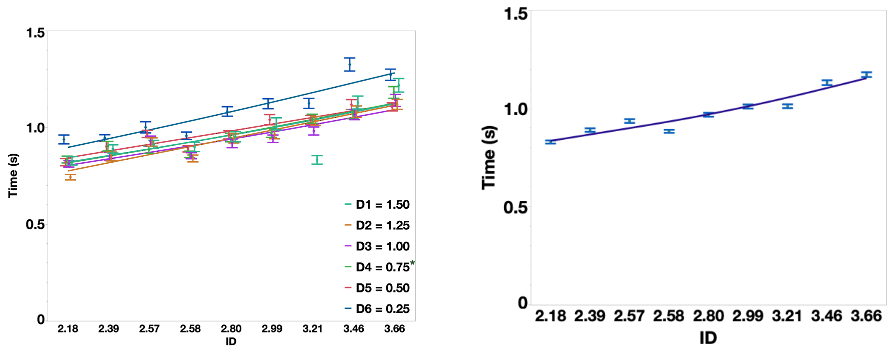
5.8 Fitts' Law Results

When using Fitts' law to model the movement time for the entire experiment (Figure 8b), we identified the following coefficients: $a = 0.34$ and $b = 0.22$ with $R^2 = 0.71$, Figure 8b. Applying Fitts' law across different diopter levels (D1 to D6 in Figure 8a) resulted in the coefficients and fit values shown in Table 3.

Comparing these results across depth values, we observed slight variations in the coefficients. The movement time differences between higher and lower depth conditions remain relatively consistent, as indicated by the stable b-values. However, the variations in a-values suggest that the model explains only a small portion of the variability in movement time, highlighting the need for more complex models or additional factors to explain task performance better. Prior work has extended Fitts' law for 3D virtual environments, showing that interaction mechanics, depth variations, and the capabilities of modern low-cost VR headsets significantly impact movement time and accuracy in selection tasks [64].

The relatively low fit values of $R^2 = 0.71$, $AIC = -136.09$, and $BIC = -130.119$ for the "standard" version of Fitts' law motivated us to further investigate the impact of VAC in the formation of the Fitts' law.

First, we applied Barrera and Stuerzlinger's "Change of Target Depth" (CTD) model to our data [11], as shown in Equation 5.



(a) For different diopters.

(b) For the whole study.

Fig. 8: Fitts' law model for (a) different depth (diopter) levels (Depth conditions correspond to different diopter values (D1 = 1.50, D2 = 1.25, D3 = 1.00, D4 = 0.75*, D5 = 0.50, and D6 = 0.25). The regression lines indicate the trend of increasing movement time with higher ID values for each depth condition. Error bars represent the standard error of the mean, although some may appear small due to minimal variance in the data. The asterisk (*) highlights D4, where the focal plane of the HMD is located, due to the least variation in performance compared to other depths, it is not prominently visible.) and (b) the data for the whole study.

$$MT = a + b * ID + c * CTD \tag{5}$$

In this model, CTD represents the *Euclidean distance change* of the target objects. Using this equation, we get $MT = 0.271 + 0.224 \cdot ID + 0.041 \cdot CTD$, which fits the data with $AIC = -174.076$, $BIC = -166.12$, and $R^2 = 0.868$.

Then, we investigated a new Fitts' law model for 3D target selection in VR with raycasting that incorporates the *V*ariation *i*n *D*iopters, (ViD), as a factor:

$$MT = a + b * ID + c * |ViD| \tag{6}$$

In this model, ViD represents the *change* in diopters of the target objects.

Using this equation, we get $MT = 0.413 + 0.224 \cdot ID - 0.085 \cdot |ViD|$, which fits the data well with $AIC = -68.59$, $BIC = -64.61$, and $R^2 = 0.81$.

When we look at the time results, we observe an increase in $D1$ compared to $D2$ and $D3$. Thus, we realized that it takes longer to execute the task when targets are placed far away from the focal plane. Therefore, instead of a linear model and based on recent insights into the effect of the VAC [65], we propose the following non-linear model:

$$MT = a + b * ID + c * e^{f * |ViD|} \tag{7}$$

Using this equation, we arrive at $MT = 0.309 + 0.224 \cdot ID + 0.553 \cdot e^{-5.688 * |ViD|}$, which fits the data with $AIC = -272.37$, $BIC = -273.44$, and $R^2 = 0.874$. Although the R^2 value of this model provides the highest values, we based our findings on the AIC and BIC values, as R^2 values are not appropriate for comparing non-linear models [66], see also Equation 7. These outcomes identify that the new model is significantly better than the other options.

The AIC score, as introduced by Akaike et al. [67], has been widely used to compare and select pointing models. According to the criteria by Burnham et al. [68], the difference in AIC scores between the two models is significant, which means that the latter model fits the data substantially better than the alternatives, i.e., it is much more likely to be appropriate to explain our data.

Table 4: Comparison of Models, shaded cells highlight the best fit. Note that R^2 should not be used for non-linear models [66] and thus has been formatted in light gray to reflect this.

Model	Formula	Coefficients	R^2	AIC	BIC
Fitts' Model	$a + b * ID$	a= 0.34, b=0.22	0.71	-136.09	-130.119
Barrera and Stuerzlinger [69]	$a + b * ID + c * CTD$	a=0.271, b=0.224, c=0.041	0.86	-174.076	-166.12
Linear model based on diopter change	$a + b * ID + c * ViD$	a=0.413, b=0.224, c=-0.085	0.81	-68.59	-64.61
Proposed model based on [65]	$a + b * ID + c * e^{f*\|ViD\|}$	a=0.309, b=0.224, c=0.553, f=-5.688	0.874	-272.37	-273.44

5.9 Subjective Results

After the experiment, we used a short questionnaire to gather participants' feedback. We asked which experimental condition (in terms of the depth measured in diopters) they preferred and why. Ten participants preferred condition D_3, five preferred D_4, two liked both D_2, and none expressed a preference for conditions D_1, D_5, or D_6. Participants who favored D_3 commented: “I found that the further away the balls were, the harder it was to see where my pointer was going, but the middle distance meant that I could still see the laser. When the balls were too close, it didn't feel as natural to point and press, while the middle distance felt natural”, “Better distance between the pointer and subjects”, “It was balanced depth”, “It was easy to select”, “I had better depth perception and felt more confident selecting smaller balls”, “Not too far, not too close, it is ideal”, and “D3 is the most comfortable because it is not too far and not too close. It is difficult to hit the target accurately if it is too far. The same issue applies when it is too close; it feels like too much movement is involved to hit each target. If it is somewhere in between, it feels just right.” Participants who preferred D_4 remarked: “It felt easier to point and select compared to the very close ones”, “It changes your attention”, and “It was a good balance between too close and too far. Not too much arm or wrist movement required.” Regarding condition D_2, participants noted: “It felt easier to point and select compared to the very close ones,” and “Easier to point at and maintain the aim at the balls.” Finally, we asked participants if they experienced any fatigue after the experiment (on a scale from 1-7, where 1 represents feeling completely normal and 7 represents complete fatigue).

Participants reported minimal physical (mean: 2.72, median: 2.5, standard deviation: 1.81) and low mental fatigue (mean: 2.11, median: 1, standard deviation: 1.53). As the overall task was not highly demanding, they reported low levels of mental and physical fatigue. One contributing factor was that with our experimental protocol, participants did not have to hold their arms in mid-air continuously, as they could rest their arms between rounds of selections. Thus, the reported fatigue was minimal, aligning with these findings.

6 Discussion

In this paper, we analyzed the effect of the VAC on user performance using the ISO 9241:411 multidirectional task across six different depth conditions. We aimed to understand how VAC negatively impacts user performance and explore how Fitts' law can be formulated under varying depth conditions. To achieve this, we adjusted target sizes and distances so that they would be perceived to have the same dimensions despite changes in their actual size.

When examining user performance in terms of time, we observed a *significant* increase in movement time for targets that are behind the HMD's focal plane. This finding aligns with previous research on VR interactions at different depths [14, 46, 63], which shows that when objects are not placed at the focal plane of the HMD, user movement time can significantly increase. Further investigation into task execution time revealed that the lowest mean task time observed was 0.943 seconds at D_3, followed closely by 0.96 seconds at D_4, corresponding to the HMD's focal plane, where no vergence-accommodation conflict is present. This $0.017s$ ($17\ ms$) difference could be a result of small variations in human motor behavior or minor fluctuations in depth perception under near-optimal conditions. Although selection times across other depth distances were not statistically significantly different from those at the focal plane (except for D_6), this finding is consistent with previous research [13]. The observed variation in movement time across different depth distances supports our hypothesis **H1**: when targets are perceived to be the same size and at the same lateral distance, user movement time varies with different depth distances.

Our error rate findings echo the same pattern as the one for time. The results showed that participants made significantly fewer errors when interacting with targets closer to the focal plane, irrespective of their position relative to it. This result is consistent with previous research that found increased errors at farther depths due to the VAC [34, 70], where the mismatch between vergence and accommodation at greater distances increases visual strain and reduces interaction accuracy.

When examining the throughput results, we found that the lowest throughput occurred at the lowest diopter value, D_6, i.e., the furthest distance. This suggests that participant throughput performance varies due to the VAC despite perceiving the same target sizes and distances. However, we did not observe any significant difference in the ID_e. A deeper analysis of these results showed that user accuracy, SD_x, also decreased at lower diopters. This was expected, as calculations are based on Euclidean distance. In virtual environments, interaction performance can be influenced by various factors, such as jitter [71, 72], which can negatively impact performance, especially when using raycasting techniques at greater target depths. Even if the selection angle remains the same, the Euclidean selection distance increases with each additional depth, further challenging user performance.

Thus, we also analyzed our results with angular SD_x and angular throughput. The angular throughput results showed that user performance significantly decreased at lower diopter values, i.e., beyond the focal plane. However, unlike the Euclidean SD_x results, we did not find significant differences for angular SD_x. This suggests that users selected targets with consistent angular distances, regardless of changes in depth distance.

The significant decrease in user performance in time and throughput yielded the question of the applicability of the original Fitts' law formulation to studies that are conducted at different target depths. Thus, we decided to investigate how Fitts' law matches the changes in user performance for objects at depths different from the focal plane. Our results showed that Fitts' law has high fit values for the data of each depth distance, except for D_1; however, when we applied Fitts' law to all depth distances as one dataset, we observed that R^2 decreased to 0.71. We initially applied Barrera and Stuerzlinger's Change of Target (CTD) model [69] to our data, and the results showed that this model provided a better fit compared to the conventional Fitts' law model. Building on their model, we proposed a linear model to account for variations in diopters; however, this did not improve the fit. Then, based on previous research on the effects of the VAC in terms of perception time [65], we created a model that uses diopters within an additional exponential term to represent the depth change in the virtual environment. The proposed model increased the R^2 to 0.874, with lower AIC (-272.37) and BIC (-237.44) values compared to the original Fitts' law formulation, which resulted in a (highly) significant improvement in terms of model fitting (also because R^2 values should not be used to compare with a non-linear model [66]). These results support our hypothesis **H2** that a variation of Fitts' law that considers the change in visual target depth models user performance subject to the VAC better.

By explicitly integrating diopters, our improved model ties user 3D selection performance to the more physiologically grounded measure of focal distance. This has immediate benefits for VR/AR interface design, where placing interactive elements closer to the headset's focal plane can mitigate VAC-related performance losses. Additionally, it offers researchers a more accurate means of comparing 3D selection tasks across different immersive systems.

Overall, our results indicate that both $D_5 = 0.5$ D (200 cm) and $D_6 = 0.25$ D (400 cm) show substantially higher variation in terms of user performance, with the farthest distance displaying a clear difference across all measures. This effect is most evident in the throughput measures, which combine speed and accuracy into a single value.

Our participants' subjective comments also align with our quantitative results. As targets were positioned further from the VR headset's focal plane, participants expressed a dislike for such targets, i.e., they found it harder to interact with them. For example, in our user study, participants did not prefer the closest targets (D_1) or the furthest targets (D_6) but preferred those around the focal plane, even though they perceived the targets to be the same sizes and distances.

In this study, we used a VR HMD with a focal plane at 133 cm. Previous studies have used different headsets with varying focal planes, yet our findings align with theirs. For instance, Batmaz et al. [14] used an HTC Vive Pro HMD with a focal plane of 75 cm (D=1.33) and placed targets at 400 cm (D=0.25) to analyze the effects of the VAC at different distances. Their results showed a significant decrease in user performance in terms of time, error rate, throughput, and SD_x when VAC was introduced while users were selecting targets with the raycasting interaction method. Consistent with other VAC studies [13, 63], we observed similar patterns in our study, where user performance decreases when objects are not placed at the focal plane of the headset. However, in the future, our investigation presented here should be extended to the virtual hand interaction methods to evaluate the effect of performance change with different depth distances in VR more comprehensively [13, 14, 50, 63].

When targets were outside the zone of comfort for stereo vision, meaning the difference exceeded +/-0.5 D [30, 31], performance dropped significantly. This is most evident in the significant difference between 0.75 D and 0.25 D, which is consistent with previous work using an HMD with a different focal plane [13]. However, this drop in performance was not observed *in front of* the focal plane, specifically not between 0.75 D and either 1.25 D or 1.5 D. Consequently, we speculate that different factors may influence performance in front of the focal plane.

Our results demonstrated that user performance varies at different depth distances. However, it remains an open research question whether these findings can be applied to other VR and AR HMDs. More importantly, it raises the question of *whether it is possible to map user performance across different headsets to account for the VAC.* Without establishing such a mapping, we believe that replicating a study with the same target sizes and distances at the same target depth but using different headsets might yield different results, particularly if the focal planes of the headsets are different. Our study used only one VR headset with a single focal plane. Although our results align with previous work [13, 46, 50, 63, 69], the robustness of the new Fitts' law formulation must be investigated across different headsets with varying focal planes. Even if the targets are placed at the same focal planes in two different headsets, the results may thus differ due to variations in terms of the VAC. Future studies should explore the proposed model using different VR and AR headsets to verify and validate our findings, ensuring its applicability across diverse display systems.

In our study, we focused on minimizing the influence of background pictorial cues by employing a minimal spatial context for target placement and maintaining a uniform experimental environment. It is worth noting that Fernandes et al. [73] observed that changes in environment or pictorial cues do not significantly influence user selection performance. Their findings suggest that participants rely more on task-relevant depth cues, such as vergence and angular measurements, rather than environmental or pictorial elements. On the ohter hand, Cheng et al. [40] found that the presence of a textured ground surface can influence user target selection performance. Future work could further validate these findings by systematically varying background and environmental conditions in selection tasks to better understand their role in user performance.

We proposed a novel version of Fitts' law based on data collected from our participants, totaling 10,692 data points. Compared to previous studies that have modeled Fitts' law and proposed new models, this dataset is larger. For instance, Barrera and Stuerzlinger collected 7,128 data points in their first user study and 2,376 data points in their second study, which they used to propose a Fitts' law model based on varying target depths [69]. In Cha and Myung's experiment [17], each participant performed 288 trials, with a total of 12 participants, resulting in 3,456 data points. Murata and Iwa [16] had participants perform 128 conditions, with 10 participants, yielding 1,280 data points. Although we collected a large amount of data, future studies should expand the range of ID values and reanalyze the proposed formula to ensure its robustness.

7 Design Recommendations

Recommendation 1

Instead of increasing the size of the objects to compensate for the effect of depth, bring them closer to the focal plane. The results for time, error rate, and throughput indicate that user performance cannot be maintained consistently across different depth distances, even when the perceived target size and distances are the same. For example, when an object, such as a menu in the virtual environment, is moved farther from the user beyond the focal plane, performance decreases despite increasing the object size to account for the distance change. Therefore, we recommend positioning objects closer to the focal plane of the headset, rather than simply bringing them closer to the user.

Recommendation 2

When running a user study, indicate the target depth. These results also raise the question of whether Fitts' law studies in VR can be replicable, even if they use the same perceived target size and target distance. Studies that replicate the previous work on user performance might vary if they do not use the same target depth, even if they use the same headset. Thus, we recommend indicating the target depth in VR publications and future studies to replicate the same target depth with the same headsets (or using the same focal plane).

Recommendation 3

Report the focal plane of the headset. The VAC has detrimental effects on user performance, and one way to improve user interaction performance is by placing targets at the headset's focal plane. However, most manufacturers do not disclose this value or do not provide explicit recommendations to designers, practitioners, and developers for object placement at this optimal distance. Failing to report the focal plane introduces significant challenges for designers and researchers, such as inconsistent performance across applications, difficulty in study replication, increased user discomfort, and reduced interaction performance. Publicly and officially sharing the focal plane of the headset could help the community to create better applications, significantly enhance user interaction performance in virtual environments, and enable a more standardized approach to studying and improving VR/AR interaction.

Recommendation 4

With raycasting, use angular SD_x calculations, not Euclidean SD_x, in ISO 9241:411 tasks if possible. The results based on the Euclidian SD_x and angular SD_x yielded different outcomes. While the results for Euclidean SD_x showed significant differences, we did not observe such results with angular SD_x. This is expected, as rotational pointing movements with raycasting lead to increased Euclidean selection distances from the center of the target as depth increases. To eliminate the depth dependencies in SD_x calculations, we thus recommend using angular SD_x for calculations.

8 Conclusion

In this paper, we analyzed the vergence-accommodation conflict in current stereo display systems across six different depth conditions using an ISO 9241:411 multidirectional selection task. Target sizes were adjusted based on angular measurements to ensure they were perceived to be the same visual size within the virtual environment. The results showed that as target depth increased, task time also increased, while user throughput decreased. The variance in task time motivated us to propose a novel Fitts' law formulation that accounts for depth in virtual environments in terms of diopters. Our results demonstrated that the new model outperforms previous ones. We hope that future studies will build on our findings to better model user movements and actions in virtual environments.

References

[1] Koulieris, G.A., Akşit, K., Stengel, M., Mantiuk, R.K., Mania, K., Richardt, C.: Near-eye display and tracking technologies for virtual and augmented reality. In: Computer Graphics Forum, vol. 38, pp. 493–519 (2019). https://doi.org/10.1111/cgf.13654 . Wiley Online Library

[2] Rakkolainen, I., Turk, M., Höllerer, T.: A compact, wide-fov optical design for head-mounted displays. In: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology. VRST '16, pp. 293–294. Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2993369.2996322

[3] Singh, G., Swan, J.E., Jones, J.A., Ellis, S.R.: Depth judgment measures and occluding surfaces in near-field augmented reality. In: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization, pp. 149–156 (2010). https://doi.org/10.1145/1836248.1836277

[4] Renner, R.S., Velichkovsky, B.M., Helmert, J.R.: The perception of egocentric distances in virtual environments-a review. ACM Computing Surveys (CSUR) **46**(2), 1–40 (2013) https://doi.org/10.1145/2543581.2543590

[5] Lin, C.J., Woldegiorgis, B.H., Caesaron, D.: Distance estimation of near-field visual objects in stereoscopic displays. Journal of the Society for Information

Display **22**(7), 370–379 (2014) https://doi.org/10.1002/jsid.269

[6] Armbrüster, C., Wolter, M., Kuhlen, T., Spijkers, W., Fimm, B.: Depth perception in virtual reality: distance estimations in peri-and extrapersonal space. Cyberpsychology & Behavior **11**(1), 9–15 (2008) https://doi.org/10.1089/cpb.2007.9935

[7] Adams, H., Stefanucci, J., Creem-Regehr, S., Bodenheimer, B.: Depth perception in augmented reality: The effects of display, shadow, and position. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 792–801 (2022). https://doi.org/10.1109/VR51125.2022.00101 . IEEE

[8] Pfeil, K., Masnadi, S., Belga, J., Sera-Josef, J.-V.T., LaViola, J.: Distance perception with a video see-through head-mounted display. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–9 (2021). https://doi.org/10.1145/3411764.3445223

[9] Vienne, C., Masfrand, S., Bourdin, C., Vercher, J.-L.: Depth perception in virtual reality systems: effect of screen distance, environment richness and display factors. IEEE Access **8**, 29099–29110 (2020) https://doi.org/10.1109/ACCESS.2020.2972122

[10] Kelly, J.W., Cherep, L.A., Siegel, Z.D.: Perceived space in the htc vive. ACM Transactions on Applied Perception (TAP) **15**(1), 1–16 (2017) https://doi.org/10.1145/3106155

[11] Barrera Machuca, M.D., Stuerzlinger, W.: 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). https://doi.org/10.1145/3290605.3300437

[12] Banks, M.S., Kim, J., Shibata, T.: Insight into vergence/accommodation mismatch. In: Head-and Helmet-mounted Displays XVIII: Design and Applications, vol. 8735, pp. 59–70 (2013). https://doi.org/10.1117/12.2019866 . SPIE

[13] Batmaz, A.U., Barrera Machuca, M.D., Sun, J., Stuerzlinger, W.: The effect of the vergence-accommodation conflict on virtual hand pointing in immersive displays. In: CHI Conference on Human Factors in Computing Systems. CHI '22. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3491102.3502067

[14] Batmaz, A.U., Hudhud, M.M., Barrera Machuca, M.D., Stuerzlinger, W.: Effect of stereo deficiencies on virtual distal pointing. In: 28th Symposium on Virtual Reality Software and Technology. VRST '22. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3562939.3565621

[15] Batmaz, A.U., Machuca, M.D.B., Pham, D.M., Stuerzlinger, W.: 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 (2019). https://doi.org/10.1109/VR.2019.8797975 . IEEE

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

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

[18] Kopper, R., Bowman, D.A., Silva, M.G., McMahan, R.P.: A human motor behavior model for distal pointing tasks. International journal of human-computer studies **68**(10), 603–615 (2010) https://doi.org/10.1016/j.ijhcs.2010.05.001

[19] Navarro, R.: The optical design of the human eye: a critical review. The Optical Design of the Human Eye: a Critical Review **2**(1), 3–18 (2009) https://doi.org/10.3921/joptom.2009.3

[20] Cutting, J.E., Vishton, P.M.: 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, pp. 69–117. Academic Press, San Diego, California, USA (1995). https://doi.org/10.1016/B978-012240530-3/50005-5

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

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

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

[24] Vienne, C., Sorin, L., Blondé, L., Huynh-Thu, Q., Mamassian, P.: Effect of the accommodation-vergence conflict on vergence eye movements. Vision Research **100**, 124–133 (2014) https://doi.org/10.1016/j.visres.2014.04.017

[25] Fukushima, T., Torii, M., Ukai, K., Wolffsohn, J.S., Gilmartin, B.: The relationship between CA/C ratio and individual differences in dynamic accommodative responses while viewing stereoscopic images. Journal of Vision **9**(13), 21–21 (2009)

https://doi.org/10.1167/9.13.21

[26] Dutton, G.N., Saaed, A., Fahad, B., Fraser, R., McDaid, G., McDade, J., Mackintosh, A., Rane, T., Spowart, K.: 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), 27–34 (2004) https://doi.org/10.1038/sj.eye.6700541

[27] Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. Journal of Vision **8**(3), 33–130 (2008) https://doi.org/10.1167/8.3.33

[28] Daniel, F., Kapoula, Z.: Induced vergence-accommodation conflict reduces cognitive performance in the Stroop test. Scientific Reports **9**(1), 1–13 (2019) https://doi.org/10.1038/s41598-018-37778-y

[29] Percival, A.S.: The relation of convergence to accommodation and its practical bearing. Ophthal. Rev. **11**, 313–328 (1892)

[30] Sheard, C.: The prescription of prisms: As determined by analyses of data on relative amplitudes of convergence and accommodation. Optometry and Vision Science **11**(10), 364–378 (1934)

[31] Banks, M.S., Akeley, K., Hoffman, D.M., Girshick, A.R.: Consequences of incorrect focus cues in stereo displays. Journal of the Society for Information Display **24**(7), 7 (2008)

[32] Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: Predicting visual discomfort with stereo displays. Journal of vision **11**(8), 11–11 (2011) https://doi.org/10.1167/11.8.11

[33] O'Keefe, E., Moffitt, K., Seemiller, E.S., Winterbottom, M., Knapp, J., Hadley, S.: Investigating the use of spectacle lenses to alleviate vergence-accommodation mismatch in a stereoscopic remote vision system. Electronic Imaging **35**, 1–6 (2023) https://doi.org/10.2352/EI.2023.35.2.SDA-390

[34] Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. Journal of vision **8**(3), 33–33 (2008) https://doi.org/10.1167/8.3.33

[35] Lin, C.J., Canny, S.: Effects of virtual target size, position, and parallax on vergence-accommodation conflict as estimated by actual gaze. Scientific Reports **12**(1), 20100 (2022) https://doi.org/10.1089/cpb.2007.9935

[36] O'Leary, D.J., Howell, E.R., et al.: Near visual function in young children. part i: Near point of convergence. part ii: Amplitude of accommodation. part iii: Near

heterophoria. Ophthalmic & physiological optics **20**(3) (2000) https://doi.org/10.1046/j.1475-1313.2000.00498.x

[37] Lockhart, T.E., Shi, W.: Effects of age on dynamic accommodation. Ergonomics **53**(7), 892–903 (2010) https://doi.org/10.1080/00140139.2010.489968

[38] Grzybowski, A., Schachar, R.A., Gaca-Wysocka, M., Schachar, I.H., Kamangar, F., Pierscionek, B.K.: Mechanism of accommodation assessed by change in precisely registered ocular images associated with concurrent change in autorefraction. Graefe's Archive for Clinical and Experimental Ophthalmology **256**, 395–402 (2018) https://doi.org/10.1007/s00417-017-3843-2

[39] Kooi, F.L., Toet, A.: Visual comfort of binocular and 3d displays. Displays **25**(2-3), 99–108 (2004) https://doi.org/10.1016/j.displa.2004.07.004

[40] Cheng, L.-Y., Lin, C.J.: The effects of depth perception viewing on hand–eye coordination in virtual reality environments. Journal of the Society for Information Display **29**(10), 801–817 (2021) https://doi.org/10.1002/jsid.1068

[41] Bowman, D.A., Hodges, L.F.: An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: Proceedings of the 1997 Symposium on Interactive 3D Graphics. I3D '97, p. 35. Association for Computing Machinery, New York, NY, USA (1997). https://doi.org/10.1145/253284.253301

[42] Argelaguet, F., Andujar, C.: A survey of 3d object selection techniques for virtual environments. Computers & Graphics **37**(3), 121–136 (2013) https://doi.org/10.1016/j.cag.2012.12.003

[43] LaViola Jr, J.J., Kruijff, E., McMahan, R.P., Bowman, D., Poupyrev, I.P.: 3D User Interfaces: Theory and Practice. Addison-Wesley Professional, Boston, MA, USA (2017)

[44] Teather, R.J., Stuerzlinger, W.: Pointing at 3D targets in a stereo head-tracked virtual environment. In: 2011 IEEE Symposium on 3D User Interfaces (3DUI), pp. 87–94 (2011). https://doi.org/10.1109/3DUI.2011.5759222

[45] Janzen, I., Rajendran, V.K., Booth, K.S.: Modeling the impact of depth on pointing performance. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, pp. 188–199. ACM, San Jose, CA, USA (2016). https://doi.org/10.1145/2858036.2858244

[46] Batmaz, A.U., Machuca, M.D.B., Pham, D.M., Stuerzlinger, W.: 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 (2019). https://doi.org/10.1109/VR.2019.8797975

[47] Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of experimental psychology **47**(6), 381 (1954) https://doi.org/10.1037/h0055392

[48] ISO: ISO 9241-400:2015 Ergonomics of human-system interaction - Part 411: Evaluation methods for the design of physical input devices (2015)

[49] Stoelen, M.F., Akin, D.L.: Assessment of fitts' law for quantifying combined rotational and translational movements. Human factors **52**(1), 63–77 (2010)

[50] Batmaz, A.U., Hudhud Mughrabi, M., Sarac, M., Barrera Machuca, M.D., Stuerzlinger, W.: Measuring the effect of stereo deficiencies on peripersonal space pointing. In: Conference on Virtual Reality and 3D User Interfaces. VR '23 (2023). https://doi.org/10.1109/VR55154.2023.00063

[51] Batmaz, A.U., Barrera Machuca, M.D., Sun, J., Stuerzlinger, W.: The effect of the vergence-accommodation conflict on virtual hand pointing in immersive displays. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–15 (2022). https://doi.org/10.1145/3491102.3502067

[52] Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A.: G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods **39**(2), 175–191 (2007) https://doi.org/10.3758/BF03193146

[53] Amini, M., Stuerzlinger, W., Teather, R.J., Batmaz, A.U.: A systematic review of fitts' law in 3d extended reality. In: Conference on Human Factors in Computing Systems. CHI '25 (2025). https://doi.org/10.1145//3706598.3713623 . Conditionally accepted

[54] Carmack, J.: x.com — x.com. https://x.com/ID_AA_Carmack/status/1371485209603022853?lang=en. [Accessed 17-09-2024] (15-03-2021)

[55] Bowman, D., Wingrave, C., Campbell, J., Ly, V.Q.: Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments. In: HCI International, pp. 629–633 (2001)

[56] Teather, R.J., Stuerzlinger, W.: Visual aids in 3D point selection experiments. In: Proceedings of the 2nd ACM Symposium on Spatial User Interaction, pp. 127–136 (2014). https://doi.org/10.1145/2659766.2659770 . ACM

[57] Batmaz, A.U., Stuerzlinger, W.: The effect of pitch in auditory error feedback for fitts' tasks in virtual reality training systems. In: 2021 IEEE Virtual Reality and 3D User Interfaces (VR), pp. 85–94 (2021). https://doi.org/10.1109/VR50410.2021.00029 . IEEE

[58] Barrera Machuca, M.D., Stuerzlinger, W.: Do stereo display deficiencies affect

3D pointing? In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. CHI EA '18, pp. 1–6. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3170427.3188540

[59] Hair Jr, J.F., Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate data analysis. Pearson Education Limited (2014)

[60] Mallery, P., George, D.: SPSS for windows step by step: a simple guide and reference. Allyn, Bacon, Boston, (2003)

[61] Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 143–146 (2011). https://doi.org/10.1145/1978942.1978963

[62] Leys, C., Schumann, S.: A nonparametric method to analyze interactions: The adjusted rank transform test. Journal of Experimental Social Psychology **46**(4), 684–688 (2010) https://doi.org/10.1016/j.jesp.2010.02.007

[63] Batmaz, A.U., Turkmen, R., Sarac, M., Barrera Machuca, M.D., Stuerzlinger, W.: Re-investigating the effect of the vergence-accommodation conflict on 3d pointing. In: 29th ACM Symposium on Virtual Reality Software and Technology (VRST 2023), pp. 0–10 (2023). https://doi.org/10.1145/3611659.3615686

[64] Clark, L.D., Bhagat, A.B., Riggs, S.L.: Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. International Journal of Human-Computer Studies **139**, 102413 (2020) https://doi.org/10.1016/j.ijhcs.2020.102413

[65] Spiegel, D.P., Erkelens, I.M.: Vergence-accommodation conflict increases time to focus in augmented reality. Journal of the Society for Information Display **32**(5), 194–205 (2024) https://doi.org/10.1002/jsid.1283

[66] Spiess, A.-N., Neumeyer, N.: An evaluation of r^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a monte carlo approach. BMC pharmacology **10**, 1–11 (2010) https://doi.org/10.1186/1471-2210-10-6

[67] Akaike, H.: A new look at the statistical model identification. IEEE transactions on automatic control **19**(6), 716–723 (1974) https://doi.org/10.1109/TAC.1974.1100705

[68] Jørgensen, S.: Book review: Model selection and multimodel inference. a practical information-theoretic approach, kenneth p. brunham, david r. anderson, springer-verlag, heidelberg. Ecological Modelling **172**(1), 96–97 (2004) https://doi.org/10.1007/b97636

[69] Barrera Machuca, M.D., Stuerzlinger, W., Asente, P.: The effect of spatial ability on immersive 3D drawing. In: Proceedings of the 2019 on Creativity and Cognition, pp. 173–186. ACM, San Diego, CA, USA (2019). https://doi.org/10.1145/3325480.3325489

[70] Paulus, M.M., Straube, A., Eggert, T.: Vergence-accommodation conflict in virtual reality displays induces phoria adaptation. Journal of neurology **264**, 16–17 (2017) https://doi.org/10.1007/s00415-017-8425-z

[71] Batmaz, A.U., Stuerzlinger, W.: The effect of rotational jitter on 3D pointing tasks. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. CHI EA '19. ACM, Glasgow, Scotland, UK (2019). https://doi.org/10.1145/3290607.3312752

[72] Batmaz, A.U., Rajabi Seraji, M., Kneifel, J., Stuerzlinger, W.: No jitter please: Effects of rotational and positional jitter on 3D mid-air interaction. In: Future Technologies Conference. FTC '20, vol. AISC 1289. Springer, Vancouver, Canada (2020). https://doi.org/10.1007/978-3-030-63089-8_52

[73] Fernandes, A.S., Murdison, T.S., Proulx, M.J.: Leveling the playing field: A comparative reevaluation of unmodified eye tracking as an input and interaction modality for vr. IEEE Transactions on Visualization and Computer Graphics **29**(5), 2269–2279 (2023) https://doi.org/10.1109/TVCG.2023.3247058