Understanding the Effect of Latency on User Performance of Target Selection in Virtual Reality

Yushi Wei, Rongkai Shi, Kemu Xu, Jialin Wang, BoYu Gao, Pan Hui, *Fellow, IEEE*, Lingyun Yu, and Hai-Ning Liang, *Member, IEEE*

***Abstract*—High latency is often introduced due to limited computational capabilities and high hardware demands. It has proven to significantly impair user performance in target selection, a fundamental interaction task. Existing research has established that latency negatively impacts selection times and success rates in 2D interactive systems; however, the underlying behavioral mechanisms remain unclear. This article investigates the effects of latency on selection times, success rates, and endpoint distributions in Virtual Reality (VR) with controller-based raycasting and bare-hand direct touch—the two most common selection methods. Our results from a user study (N = 31) revealed distinct patterns between the two methods, leading to two novel mathematical models that account for latency, target width, and movement amplitude. These two models were validated via a new dataset collected from a second user study (N = 16) and were demonstrated to outperform the existing models. Our findings provide actionable recommendations to mitigate the negative impacts of latency and improve user experience in VR interface design.**

***Index Terms*—Virtual reality, human performance modeling, endpoint, head-mounted display, latency.**

I. Introduction

With the increasing complexity of functionality and the continued improvement of high-precision graphics in current virtual reality (VR) systems, it is common to encounter situations of high latency in virtual environments (VEs) due to the still relatively limited computational capability and the high demand for hardware performance [37], [55]. High latency has been directly demonstrated to affect user performance across various tasks. Among them, target selection is often taken as a representative case, since it constitutes a fundamental component of interaction: it shapes the interaction flow, represents user intent, and precedes subsequent actions such as object manipulation or command execution [68]. Specifically, numerous empirical studies in 2D interactive environments have confirmed the negative effects of latency on human behavior and performance. For instance, previous research has shown that high latency can significantly increase users' selection time, known as movement time (MT) in Fitts' Law [18], [34]. Additionally, it has been observed that higher latency leads to lower selection success rates in both pointing and temporal selection tasks [11], [45]. While these prior studies have identified the lowered selection performance caused by the latency in 2D scenarios, it remains unknown how and why latency affects users' selection behavior and performance, particularly in more complex 3D VEs where high latency is more likely to occur.

Nowadays, behavior models have become one of the effective tools for understanding user behavior and interaction patterns in interactive tasks, particularly target selection tasks. Among these models, the endpoint distribution has exhibited its powerful capability in interpreting and explaining user behavior and performance. Prior work has leveraged the endpoint distribution model to understand the effect caused by varying factors, such as target width and movement amplitude in static selection [8], [70] and movement speed in moving target pointing [22]. Moreover, beyond explaining behavioral patterns, the model has also proved its strong applicability in predicting selection success rates and assisting users in improving selection efficiency [9], [10], [59], [70]. Therefore, this approach could be utilized to understand the effect and compensate for the potential negative influences of latency in target selection tasks in VR.

In this paper, we aim to bridge the gap in our understanding of **how latency affects user selection behavior and performance in VR environments**. To do so, we conducted a user study to collect behavioral data for selection tasks in VR (Section III). In the user study, we investigated the selection performance of the two most used interaction techniques: *controller-based raycasting* and *bare-hand touch*. These two interaction techniques were representatives of two input modalities (controller and bare-hand interaction) with their best-suited interaction metaphors (raycasting and touch) for selection [33].

Received 22 November 2024; revised 9 October 2025; accepted 5 November 2025. Date of publication 21 November 2025; date of current version 6 February 2026. This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFC3307602, in part by the National Natural Science Foundation of China under Grant 62272396 and Grant 62372212, and in part by the Guangdong Provincial Talent Program under Grant 2023JC10X009. Recommended for acceptance by A. Simeone. *(Corresponding author: Hai-Ning Liang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Human Subjects Panel of the Hong Kong University of Science and Technology (Guangzhou) under Application No. HKUST(GZ)-HSP-2024-0055.

Yushi Wei, Rongkai Shi, Jialin Wang, Pan Hui, and Hai-Ning Liang are with the Computational Media and Arts Thrust in the Information Hub, Hong Kong University of Science and Technology, Guangzhou 511458, China (e-mail: hainingliang@hkust-gz.edu.cn).

Kemu Xu was with the Department of Computing, School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China. She is now with the School of Informatics, University of Edinburgh, EH8 9YL Edinburgh, U.K.

BoYu Gao is with the College of Cyber Security and Guangdong Institute of Smart Education, Jinan University, Guangzhou 510632, China.

Lingyun Yu is with the Department of Computing, School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China.

Digital Object Identifier 10.1109/TVCG.2025.3634367

After this, we conducted a second user study aimed at further verifying the correctness of the observed behavioral patterns and the proposed models (see Section V). We leveraged the models to predict success rates. The results demonstrate that both models outperformed baseline models derived from prior work across all metrics. Finally, we provided a summary of findings and a list of design recommendations for VR designers (see Section VI). We hope the findings and design recommendations derived from this research could help designers understand the effects of latency, create more effective interfaces, minimize the negative impacts of latency, predict selection performance, and guide future VR design.

In short, the main contributions of this paper include:

- *Empirical Analysis and Interpretation of Latency Effects:* Conducting a comprehensive user study to empirically analyze and interpret the impact of latency on user behavior and performance of target selection with controller-based raycasting and bare-hand touch in VEs (Section III).
- *Behavioral Modeling:* Development of two novel mathematical models to explain how the latency, target width, and movement amplitude affect endpoint distributions, movement times, and success rates in VR-based selection tasks (Section IV).
- *Prediction and Assistance:* Applying the two proposed models to predict success rates and improve users' selection performance (Section V).
- *Design Recommendations:* Providing an actionable set of design recommendations based on the findings to help VR designers understand the effects of latency and improve the effectiveness of VR interfaces (Section VI).

II. Related Work

A. Effect of Latency

End-to-end latency (or delay/lag) has been demonstrated to significantly influence users' interaction performance and behaviors in different types of systems, such as desktop computers [53], tablets [66], and VR head-mounted displays (HMDs) [37], [55]. In addition, various sources in these systems can cause latency, including but not limited to hardware sensor demands [46], data calculation processing [56], graphic rendering [4], refresh rate from the display [43], synchronization lag time across devices [64], and network performance [17]. Accordingly, latency is typically considered uncontrollable, inevitable/unavoidable, and inconsistent across different devices and platforms [30].

While reducing system latency poses challenges and complexities, existing studies strive to comprehend and minimize its impact to enhance user experiences with interactive systems. Within VR systems, Ware and Balakrishnan [57] were among the first to empirically demonstrate that high latency is a main factor leading to unsatisfying experience. The following works have further confirmed that elevated latency levels could directly induce and exacerbate cybersickness symptoms [55], [58]. This significant effect on subjective experience can be partially attributed to people's heightened sensitivity to latency, with the shortest detectable visual displays being 2.83 ms [42]. Therefore, given the adverse effects caused by latency, researchers have identified several thresholds, above which users are prone to abandoning their tasks. These thresholds vary depending on interactive systems and the given tasks, such as 180 ms for playing first-person shooter games on a PC [6], 600 ms for tapping on a touchscreen [5], and 350 ms for dragging visual elements on a tablet [32].

Besides the subjective experience, latency can affect user performance when interacting with systems—even in the essential pointing selection tasks. The impact of the latency in pointing selection tasks can be interpreted via Fitts' Law [15], which demonstrated that decreased target size (typically denoted as target width W) and increased distance to the target (i.e., movement amplitude A) correspond to extended time used for selection (i.e., movement time MT) [26], [34]. Two extended Fitts' Law models have been proposed to explain how movement time is affected by latency in selection tasks. These models empirically demonstrated a linear relationship, showing that increased latency prolongs total movement time [18], [35]. Beyond movement time, Pavlovych and Stuerzlinger [45] further showed that incremental latency directly undermines selection accuracy when targeting static objects in 2D scenarios. Their finding was further confirmed in subsequent studies involving 2D temporal moving target selection tasks across different devices and contexts, such as casual 2D tapping games on touchscreen devices [30], first-person shooter games on PC [11], [12], [13], [14]. However, the impact of latency in 3D scenarios is still unknown, as it is underexplored.

As demands for more complex functionality and advanced computer graphics grow, the built-in computing power of current VR HMDs remains inferior to desktop systems, resulting in frequent latency issues during VR use. Therefore, it is important to gain a deeper understanding of the impact of latency on fundamental selection tasks and identify ways to improve user experience and performance in VR applications.

B. Endpoint Distribution Modeling

The term "endpoint" refers to the position (of the cursor or pointing device) after each pointing selection action. Empirical evidence has demonstrated that the distribution of endpoints within 1D pointing selection tasks follows a Gaussian distribution [60]:

$$X \sim N(\mu, \Sigma). \tag{1}$$

Where μ and Σ represent the means and standard deviations of the distribution. Moreover, this finding was subsequently extended to 2D selection scenarios by Grossman and Balakrishnan [16], where the distribution was shown to follow a bivariate Gaussian distribution:

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \tag{2}$$

This distribution combines two variables, X and Y, each independently following a Gaussian distribution [27]. In this context, μ denotes the mean vector, and Σ represents the covariance matrix. Specifically, μ_x and μ_y are the means, σ_x and σ_y are the standard deviations for X and Y, respectively, and ρ indicates their correlation.

In selection tasks, the distribution can be decomposed into different components. One reflects the relative component, such as speed-accuracy tradeoff, while the other represents the absolute components, such as absolute precision, which is influenced by human motor control and hardware limitations [8], [70]. The speed-accuracy tradeoff can be understood as the interaction strategy employed by users, which determines the balance between selection precision and the amount of time they need to spend on the selection task [15], [71]. Furthermore, similar to the factors in Fitts' Law (see our discussion in the previous subsection: Section II-A), target width (W) and movement amplitude (A) have also been identified as two factors significantly influencing the deviation in endpoint distribution [8], [70]. Bi and Zhai [8] illustrated the strong positive linear relationship between target width and the standard deviation Σ of the distribution in 2D finger-pointing tasks. This finding was further confirmed in selection tasks in VR HMDs by Yu et al. [70]. In contrast to the consistent trend in target width across different selection scenarios, the effects of movement amplitude on endpoint distribution depend on the interactivity of tasks, such as the target state (moving versus static) and selection scenarios (2D versus 3D) [16], [22], [70].

The endpoint distribution model has been extended and validated to cover more complex scenarios, such as 3D multi-modal selection [59], and 1D and 2D moving target selection [20], [21], [22], [24]. For instance, Wei et al. [59] demonstrated that eye and accompanying head endpoints followed a dual-distribution pattern in gaze-based selection in mixed reality HMDs. They further drew inspiration from the work of Bi and Zhai [9], who first proposed combining Bayes' theorem with the endpoint distribution to enhance selection accuracy, and introduced a new model that integrates eye and head distributions as a novel multi-modal approach to enhance gaze-based selection accuracy in 3D environments. Additionally, Huang et al. [22] extended the model to 2D moving target selection, achieving improvements of 4% in selection speed and 41% in accuracy. Apart from enhancing selection efficiency, these works elucidated the impact of various factors on people's behaviors, encompassing but not limited to target width [59], movement amplitude [70], target moving speed [22], and target shape [23].

In conclusion, endpoint distribution models can help interface designers and interaction technique developers effectively understand the impact of potential factors on user behavior patterns. In this work, we utilized endpoint distribution modeling to investigate the impact of latency on people's selection strategies and mitigate potential adverse effects on actual selection tasks in VR HMDs.

C. Hand-Based Selection in Virtual Reality

Pointing selection is fundamental in VEs for users to interact with objects within the scene. A growing number of studies have focused on this task and investigated various input modalities, such as hand [33], [58], [69], gaze [25], [59], or head [40], [47]. Among these input modalities, hand-based input modalities are more widely used in daily interactions and experimental settings than others [33], [38], [69].

Hand-based input can be further divided into two types: controller-based and bare-hand-based [33], [48]. For controller-based input, users hold the controller and use its buttons to select or manipulate objects. In contrast, bare-hand-based input enables users to interact directly with objects in VEs, with their hand movements and gestures tracked by either the HMD's built-in hand-tracking cameras or external tracking devices [41], [51]. These two hand-based input modalities are commonly used for selection tasks with two frequent interaction metaphors: indirect *raycasting* and direct *touch* [28], [39]. The raycasting technique emits a virtual ray from the user's hand or controller. The object intersected by the ray would be selected when a selection is triggered [7]. On the other hand, touch allows users to select objects directly via their hands or fingers [44].

Previous studies have illustrated discrepancies in terms of users' objective performance and subjective preference between the two hand-based input modalities (controllers versus bare hands) and between the two interaction metaphors (raycasting versus touch) [33], [54], [72]. While bare-hand interaction often provides a more intuitive, natural, enjoyable, and immersive experience [3], [31], [36], [50], [52], the controller typically outperformed bare-hand interaction in accuracy and speed [63]. More relevant to our work, Luong et al. [33] demonstrated that controller-based raycasting offers greater accuracy and speed as well as reduced physical exertion compared to other combinations, including controller-based touch, bare-hand touch, and bare-hand raycasting. Moreover, when selecting objects using bare hands, direct touch is preferable to raycasting.

To comprehend the impact of latency and better reflect real-world application scenarios in immersive VEs, we chose two typical combinations in our study following the recommendations made by Luong et al. [33]: controller-based selection with raycasting and bare-hand selection with touch. We refer to the two combinations as the two **interaction techniques** hereafter.

III. User Study 1: Data Collection and Behaviors Understanding

The goal of this user study is threefold: (1) to understand how latency affects user behaviors and performance in pointing selection tasks in VR HMDs, specifically regarding their endpoint distribution with the mean vector μ and the covariance matrix Σ, selection time, and accuracy; (2) to verify the potential relationship between latency, target width, and movement amplitude on the selection task; and (3) to examine whether the interaction techniques (controller-based raycasting and bare-hand touch) would lead to different effects of latency on selection behaviors and performance.

A. Hypotheses

Based on the synthesis of related work discussed in Section II, the following four hypotheses (**H**) were formulated before the study:

- **H1**. *The endpoints remain a bivariate Gaussian distribution under varying latency levels.* While Pavlovych and

Sturzlinger have indicated that latency can significantly influence users' selection performance (movement time and accuracy) in a 2D environment [45], we still posit that the introduced latency would not alter the endpoints to deviate from the bivariate Gaussian distribution in 3D VEs provided by VR HMDs, as in other interactive systems [8], [21], [59], [70].

- **H2**. *Latency significantly impacts the endpoint distribution, and different latency levels can have different effects on the mean vector μ and the covariance matrix σ.* If latency is observed in the selection tasks, it will inevitably impact users' selection experience and cause additional noise to the device control [8], [45]. Such an impact would be reflected in the distribution of endpoints.
- **H3**. *The introduced latency prolongs selection times and reduces accuracy.* Latency has long been recognized as a key factor that degrades user selection performance, negatively impacting both the selection time [26], [34] and the accuracy [45] in 2D environments. We hypothesize that such negative influences of latency on selection time and accuracy would remain in immersive VR environments.
- **H4**. *The latency has stronger effects on user performance and endpoint distribution when using bare-hand touch compared to controller-based raycasting.* We hypothesized stronger latency effects for bare-hand touch than for controller-based raycasting, given the differences in interactivity between these two techniques as discussed in Section II-C.

B. Participants and Apparatus

Thirty-one participants (16 male and 15 female), whose ages ranged from 19 to 36 ($mean = 22.24$, $s.d. = 3.4$), were recruited from a local university. All participants reported having normal or corrected-to-normal vision and confirmed clear visibility of all objects within the testing scene. A 7-point Likert scale was used to assess their familiarity with VR devices and interactions, and higher scores indicated a greater level of familiarization. Five participants (16.1%) considered themselves to be highly familiar with VR devices (scoring 6 or 7), while six participants (19.3%) reported low familiarity (scoring between 1 and 2). Over half of the participants (20 participants, 64.5%) showed moderate familiarity with VR devices (with scores of 3 to 5).

A Meta Quest 3 VR headset (OS version 67.0) was employed in this study, which has a 2064 $\times$ 2208 screen resolution per eye and a refresh rate of 120 hz. Additionally, it features a field-of-view (FoV) of 110° horizontally and 96° vertically. To better simulate a real application scenario, the compatible Quest 3 controller and built-in hand-tracking camera were used to enable participants to select targets using raycasting and touch, respectively. The experimental program was developed using the Unity3D game engine (version 2022.3.0f1) with C# and integrated with the Oculus Integration SDK (version 57.0.2). The program was deployed on a high-performance PC equipped with an Intel Core i9 processor, NVIDIA RTX 3090 graphics card, and 64GB DDR5 RAM. The headset and the PC were connected via a 3-meter USB-C cable compatible with the USB 4.0 protocol.

C. Latency: Definition, Metrics, and Control

Latency mainly comes from the input device and the executable program when interacting with virtual elements in VR HMDs. That said, latency came from hand-based input devices (controller or hand-tracking modules) and the VE run via the Unity3D engine in our case, respectively. With respect to hand-based input devices, prior studies have reported variability in latency across different inputs, yet typically within low ranges; for example, on the Quest 2, hand-tracking latency averaged ~45 ms, while controllers ranged from 25-40 ms [2]. Such latencies primarily involve the hardware response and signal transmission of the tracking module, and are relatively constant and controllable as they are mainly influenced by hardware performance. However, different headsets exhibit varying latencies, and accurately evaluating and comparing these values across diverse hardware configurations presents significant challenges [29], [66]. Furthermore, including these latencies in our study would complicate the replication process. Therefore, we have chosen not to consider the latency associated with the input device from our aggregate latency measurements.

In our experimental setup, we utilized two metrics to measure latency: the app tracking to mid-photon latency (ATML) and the timewarp to mid-photon latency (TML).[1] ATML measures the time delay from when the application receives tracking data (accounting only for the head movement tracking data) to when the image is displayed on the screen. This process includes four steps: acquiring the data, processing it, rendering the image, and finally displaying it. On the other hand, TML measures the time delay from when the timewarp processing is completed to when the image is ultimately displayed on the screen. Briefly, ATML captures the aggregate latency experienced under experimental conditions, encompassing all stages from data acquisition to display, while TML limits the estimation of latency to the inherent latency in the current VEs, excluding the additional latency introduced by our experimental program. Notably, neither ATML nor TML takes into account the latency of tracking systems. This distinction allows a clear differentiation between the base latency and any extra delays caused by our testing setup.

To control latency, we employed a time-shifted rendering method in which each frame was generated from a system state offset by the experimental condition, rather than from the current state. As a result, all visual elements in the scene were uniformly delayed, producing controlled increments of motion-to-photon latency. This manipulation was applied consistently across both interaction techniques (controller-based and bare-hand) and was conducted in a pure VR setting. Additionally, as detailed in Section II-A, the actual latency experienced within application scenarios is not solely dependent on the input device but is also influenced by various other factors. These include, but are not limited to, network conditions, device performance,

[1]Please refer to https://developers.meta.com/horizon/documentation/native/pc/dg-hud/ for more technical details

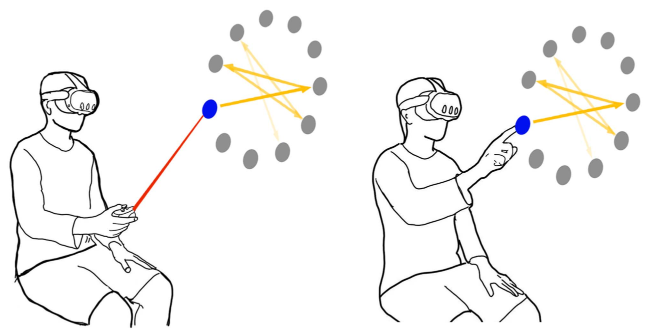


Fig. 1. An example of the experimental task in the user study. LEFT: participants use the controller-based raycasting (in red) to select the target (the blue sphere). RIGHT: participants use the bare-hand touch to select the target. The yellow arrows indicate the selection sequence in each condition. They are only for illustration and were not shown to participants.

and the required interaction tasks. This complexity underscores the challenge of maintaining constant latency across different usage conditions. Therefore, the following three methods were implemented in our experimental setup to minimize additional latency and prevent accidental fluctuations: (1) the experimental program was deployed on a high-performance PC and streamlined to remove all non-essential elements in the experimental scenarios, thus minimizing external latency due to hardware performance limitations; (2) the program was designed without internet functionality, eliminating any latency that could arise from network issues; and (3) the experimenter monitored two aforementioned metrics to ensure accurate latency and address unforeseen situations throughout the study.

D. Selection Task and Techniques

The task adopted the ISO9241-9 Fitts' ring design [1]. In this task, 11 grey virtual spheres were positioned like a ring in front of the participant at eye level. Each ring represented one target width $\times$ movement amplitude condition with 11 repetitive trials. All spheres were positioned at a predetermined depth of 10 m for the indirect raycasting technique and 0.35 m for the direct touch.

Each condition started by selecting the 'beginning' sphere at the center of the Fitts' ring. Participants needed to use the given techniques to select the target, which turned blue when each trial was initiated. Once a selection was completed, a short beep sound effect was provided, and the selected target would return to grey. The next target would be the one in the opposite position, and the alternating sequence proceeded clockwise (see Fig. 1) to ensure consistent movement amplitude between targets in two adjacent trials. One exception is the amplitude between the 'beginning' sphere and the first target in the first trial. This trial was removed from our data collection, leaving 10 recorded trials for each condition.

We had two interaction techniques in this user study: controller-based raycasting and bare-hand touch. For controller-based raycasting, an infinite red ray was used as a visual indicator to aid participants during selection. Once participants have aimed at the target, they can press the trigger button to confirm the selection (see Fig. 1 LEFT). With bare-hand touch, participants used their index fingers to 'poke' the target (see Fig. 1 RIGHT). In alignment with the setting described by Loung et al. [33], confirming the selection via bare-hand touch was not triggered by gestures but by the index fingertip's location if the depth over the target was more than 0.35 m. For both techniques, the physical controllers and hands were not visually represented; interaction was conveyed solely through their virtual counterparts, a controller model with a ray indicator and a semi-transparent virtual hand for bare-hand touch, implemented using the Oculus Integration SDK without any additional motion filtering beyond the default system processing.[2]

Once the selection was activated, the endpoint was recorded using a spherical coordinate system regardless of whether the target was successfully selected. The coordinate system was introduced with two axes: the x-axis aligns in the direction of movement (e.g., from the prior target toward the subsequent target), while the y-axis is perpendicular to it. The origin, $(0, 0)$, was located at the center of the target [70].

E. Design and Procedure

The experiment used a $3 \times 3 \times 4 \times 2$ within-subjects design with four independent variables: Target Width W ($1^\circ, 2.5^\circ$, and 4°), Movement Amplitude A ($15^\circ, 35^\circ$, and 55°), Latency L (50 ms, 110 ms, 170 ms, and 230 ms), and Interaction Techniques IT (controller-based raycasting and bare-hand touch, short as CR and BT hereafter).

The values of W and A were determined by our pilot study and represented in visual angles. We utilized target width to simulate varying levels of selection difficulty. A W of approximately 2.5° has been empirically shown to provide a suitable size for accurate selection in VR HMDs with over 80% success rate [59], [70]. Thus, we introduced one smaller (1°) and one larger (4°) width to simulate more challenging and easier scenarios. The movement amplitudes were set to reflect three distinct scenarios: (1) the Fitts' ring with an A of 15° was fully within the FoV, thus eliminating the need for head movement to compensate for the FoV constraint; (2) the Fitts' ring with an A of 35° was at the edge of the FoV requiring minimal head movement; and (3) only part of the Fitts' ring with an A of 55° was visible within the FoV, necessitating significant head movement. Four latency conditions were adopted for our experiment. The minimal latency was defined as 50 ms, comprising a $\sim$35 ms baseline measured in our experimental scene plus an additional 15 ms manually introduced. This 50 ms baseline was chosen to approximate real application scenarios with richer interactive elements. Building on this value, we then increased latency in fixed 60 ms steps to systematically evaluate its effects.

The entire experiment was divided into two sessions by the IT. The order of the two sessions alternated between participants. Within each session, the order of L was counterbalanced using the Latin Square method. For each L, the nine combinations of $W \times A$ were presented in a random order to minimize any learning effects. In total, 22,320 endpoints (=3 $W \times 3$ $A \times$

[2]Controllers: https://developers.meta.com/horizon/design/controllers/; Hands: https://developers.meta.com/horizon/design/hand-representation/.

4 $L \times 2\ IT \times 31$ participants $\times$ 10 repetitions) were collected from the user study.

The total duration of the experiment was approximately forty minutes for each participant. At the beginning of the experiment, participants were asked to complete a questionnaire to collect their demographic information and were briefly introduced to the experiment. Following this, they wore the VR HMD and confirmed that they were in a comfortable position and could clearly see all objects in the VE. After this, practice trials lasting about five minutes were conducted. This practice was designed to help them become familiar with the devices, the task, and the two interaction techniques. After participants got familiar with the task and interaction techniques, they could start the formal trials. Participants were required to rest between the two conditions to mitigate any potential fatigue caused by the lengthy and repetitive selection. They could resume the next condition whenever they felt prepared to continue, ensuring their performance was not adversely affected by fatigue or other external factors. Meanwhile, given that latency may potentially induce cybersickness, participants were informed that they could withdraw from the experiment at any time if they experienced any discomfort. No participant reported such symptoms, and none chose to discontinue the study—that is, all participants were able to complete it. Finally, to more accurately reflect real application scenarios, participants were encouraged to select targets naturally, using their own strategies and balancing speed with accuracy according to their personal judgment [65].

F. Measurements, Data Pre-Process and Analysis Methods

In this study, we measured seven dependent variables, including μ_x, μ_y, σ_x, σ_y, ρ, movement time, and success rate. The variables μ_x, μ_y, σ_x, σ_y, and ρ were components of the bivariate Gaussian distribution. Movement time refers to the time taken to select a target from the start position to the end. The success rate was calculated based on the endpoint of each selection; a selection was considered successful if the endpoint was located within the target boundary. Additionally, a short interview was conducted after each IT session to understand the participants' selection experience and strategy.

Regarding the data pre-processing, we applied two procedures to exclude outliers from our dataset. First, outliers were identified as measurements on the two axes or time in a single trial with more than three standard deviations from the average results for each condition. Second, we excluded trials where participants inadvertently activated the confirmation mechanism twice in rapid succession. The following trial after such a case was also removed because the actual movement amplitude for this trial might not strictly follow the condition. In total, 896 trials, constituting about 4% of the dataset, were removed, including 417 trials from the CR condition and 477 trials from the BT condition.

Matlab (R2023a) and SPSS (version 29.0) were utilized for data analysis. In Matlab, we first performed the Kolmogorov-Smirnov tests via the `kstest()` function to assess the normality of endpoint distributions along the x- and y-axes. Subsequently, we used maximum likelihood estimation via `mle()` to estimate the parameters μ and σ. Correlation coefficients (ρ) were derived using `corrcoef()`. The average values of movement time and success rate were calculated for each condition ($1W \times 1A \times 1L \times 1IT$). After these processes, repeated-measures ANOVA (RM-ANOVA) was conducted in SPSS to explore the relationships between dependent and independent factors. The Greenhouse-Geisser correction was applied to adjust degrees of freedom when the assumption of sphericity was not met, and Bonferroni corrections were applied for post-hoc pairwise comparisons.

TABLE I
SIGNIFICANT RESULTS OF RM-ANOVAS FOR THE CR-BASED ENDPOINTS ($\alpha = 0.05$). DV REPRESENTS THE DEPENDENT VARIABLE.

Factor	*DV*	df_{effect}	df_{error}	F	p	η_p^2
W	μ_x	2	60	11.468	0.002	0.214
A	μ_x	2	60	8.115	0.004	0.233
$L \times A$	μ_x	6	180	4.315	0.011	0.126
W	μ_y	1.854	55.616	3.424	0.043	0.102
A	μ_y	2	60	7.377	0.001	0.197
W	σ_x	1.988	59.633	115.513	<0.001	0.794
A	σ_x	2	60	23.677	<0.001	0.441
W	σ_y	1.906	57.189	141.959	<0.001	0.826
A	σ_y	2	60	41.882	<0.001	0.583

G. Controller-Based Raycasting Endpoints

1) Normality Tests and Significance Tests: The results from the Kolmogorov-Smirnov tests did not reject the normality hypothesis. Thus, consistent with findings from prior research [59], [70], it was determined that all sets of selection endpoints using controller-based raycasting adhere to a Gaussian distribution. Moreover, the results of the RM-ANOVA revealed significant main effects of the factors and an interaction effect on our endpoints-related dependent variables, as summarized in Table I.

2) Regression: According to the main effects previously outlined, a pronounced linear relationship was observed between the factors and the dependent variables. Regarding the linear regression, the relationships identified were as follows: $\mu_x = -0.0414W - 0.2504 (R^2 = 0.9702)$, $\mu_x = -0.0081A - 0.0582 (R^2 = 0.9968)$, $\mu_y = 0.0114W + 0.058 (R^2 = 0.6921)$, $\mu_y = 0.0028A - 0.0176 (R^2 = 0.9734)$, $\sigma_x = 0.1089W + 0.5056 (R^2 = 0.9998)$, $\sigma_x = 0.0052A + 0.5868 (R^2 = 0.9994)$, $\sigma_y = 0.0839W + 0.4269 (R^2 = 0.9995)$ and $\sigma_y = 0.0045A + 0.4713 (R^2 = 0.9976)$.

Furthermore, curvilinear regression was conducted in response to the observed interaction effect between μ_x and $L \times A$ (see Fig. 2). The formula derived was $\mu_x = a + bL + cA + dL \cdot A$, where the constants are as follows: $a = 0.1469, b = -0.0017, c = -0.0162$ and $d = 0.00006$ ($\mu_x = 0.1469 - 0.0017L - 0.0162A + 0.00006L \cdot A$). To mitigate potential overfitting from the added coefficients, we evaluated the 95% confidence intervals (CI) for each, noting that only the intercept value crossed zero ($a = [-0.0430, 0.3369], b = [-0.00292, -0.00048], c = [-0.0214, -0.01119]$, and $d = [0.000027, 0.000093]$). The same pattern was observed within the significance testing

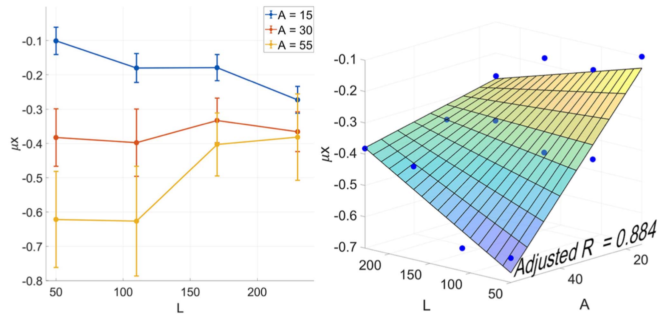


Fig. 2. LEFT: The line plot showing the relationship between $L \times A$ and μ_x in the CR condition. RIGHT: Fitting results for the curvilinear regression between $L \times A$ and μ_x in the CR condition. Error bars represent standard errors.

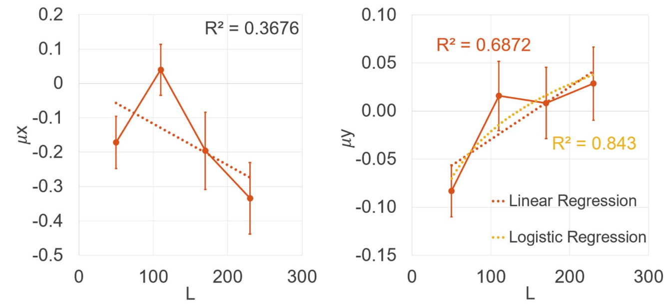


Fig. 3. LEFT: The results of the linear regression between μ_x and L in the BT condition. RIGHT: the results of linear and logistic regression between μ_y and L in the BT condition. Error bars represent standard errors.

TABLE II
SIGNIFICANT RESULTS OF RM-ANOVAS FOR THE BT-BASED ENDPOINTS ($\alpha = 0.05$). DV REPRESENTS THE DEPENDENT VARIABLE.

Factor	*DV*	df_{effect}	df_{error}	*F*	*p*	η_p^2
L	μ_x	2.939	88.162	3.566	0.018	0.106
A	μ_x	2	60	31.282	<0.001	0.510
L	μ_y	2.938	88.143	2.746	0.049	0.084
L	σ_x	3	90	16.876	<0.001	0.360
A	σ_x	2	60	28.424	<0.001	0.487
L	σ_y	2.654	79.63	5.567	0.002	0.157
A	σ_y	1.874	56.234	53.308	<0.001	0.640

of each coefficient, revealing that while the intercept a was not significant ($p = 0.112$), coefficients b ($p = 0.012$), c ($p < 0.001$), and d ($p = 0.002$) all demonstrated strong statistical significance.

H. Bare-Hand-Based Touch Endpoints

1) Normality Tests and Significance Tests: Aligning with the results on CR-based selection, the Kolmogorov-Smirnov tests did not reject the hypothesis of normality when the interaction technique was the bare-hand touch (BT). All sets of endpoints are normally distributed. Notably, we found that latency had more significant effects on endpoint distribution when using BT. Table II illustrates the factors that had a statistically significant main effect on the endpoints-related dependent variables. Specifically, the factor of L had a significant main effect on μ_x, μ_y, σ_x, and σ_y; while A had a significant main effect on μ_x, σ_x, and σ_y.

2) Regression: Linear regression was conducted to interpret the relationships between the factors and the dependent variables. The results indicated a weak linear relationship between L and μ_x ($\mu_x = -0.0012L + 0.0039$ with $R^2 = 0.3676$) and a moderate linear relationship between L and μ_y ($\mu_y = 0.0005L - 0.0839$ with $R^2 = 0.6827$) (see Fig. 3). All other factors demonstrated strong linear relationships: $\mu_x = -0.0133A + 0.3237 (R^2 = 0.9972)$, $\sigma_x = 0.0031L + 0.959 (R^2 = 0.9477)$, $\sigma_x = 0.0108A + 0.9916 (R^2 = 0.9999)$, $\sigma_y = 0.0015L + 1.0448$ $(R^2 = 0.8835)$, and $\sigma_y = 0.0078A + 0.9665 (R^2 = 0.9955)$.

Linear regression lacks sufficient capability to explain the influence of L on μ_x and μ_y. Therefore, we employed additional methods to interpret their effects better. For μ_x, we first introduced more coefficients into the regression function, such as polynomial regression, to achieve better fitting results. Although the R^2 (the coefficient of determination) was improved to exceed 0.99, the increased number of polynomials and extra coefficients risked overfitting, making it challenging to explain the effect of the factors on μ_x through the coefficient values. Thus, given that a main effect was also demonstrated between A and μ_x, and previous works have shown that the effect on μ_x can be aggregated [21], [59], [70], it is feasible to conduct nonlinear regression to obtain fitted coefficients after aggregating the factors, rather than using the coefficients obtained from empirical studies [65]. Therefore, the nonlinear regression was performed, and the expression combining A and L can be written as $\mu_x = a + bA + cL$. Consequently, our fitted results indicated an R^2 of 0.731. However, it should be noted that the outcome of c was different from that of a and b. The 95% CI for c crossed zero ($a = [0.10345, 0.74347]$, $b = [-0.01889, -0.00626]$, and $c = [-0.00276, 0.00034]$), and the significance test showed that c was statistically non-significant ($a : p = 0.015, b : p = 0.001$, and $c : p = 0.111$). Additionally, the weaker η_p^2 values for L compared to factor A also confirmed this pattern (see Table II). Therefore, for simplicity and based on the aforementioned results, we can recognize that the effect of L on μ_x is negligible. Additionally, for μ_y, we adopted logistic regression to fit the values of μ_y under different L. This method demonstrated better performance ($\mu_y = 0.0706 \ln(L) - 0.3461$, $R = 0.843$) and higher interpretability, indicating that the effect of L on μ_y gradually decreases as L increases (see Fig. 3 RIGHT).

I. Movement Time and Success Rate

We first examined the general discrepancy between the two interaction techniques IT in terms of movement time and success rate. The results of RM-ANOVA indicated that IT led to significant differences in both movement time ($F_{1,1115} = 167.421, p < 0.001, \eta_p^2 = 0.131$) and success rate ($F_{1,1115} = 1100.988, p < 0.001, \eta_p^2 = 0.497$). In the following subsections, we compared CR and BT in each condition ($1A \times 1W \times 1L$). The results are summarized in Fig. 4 TOP.

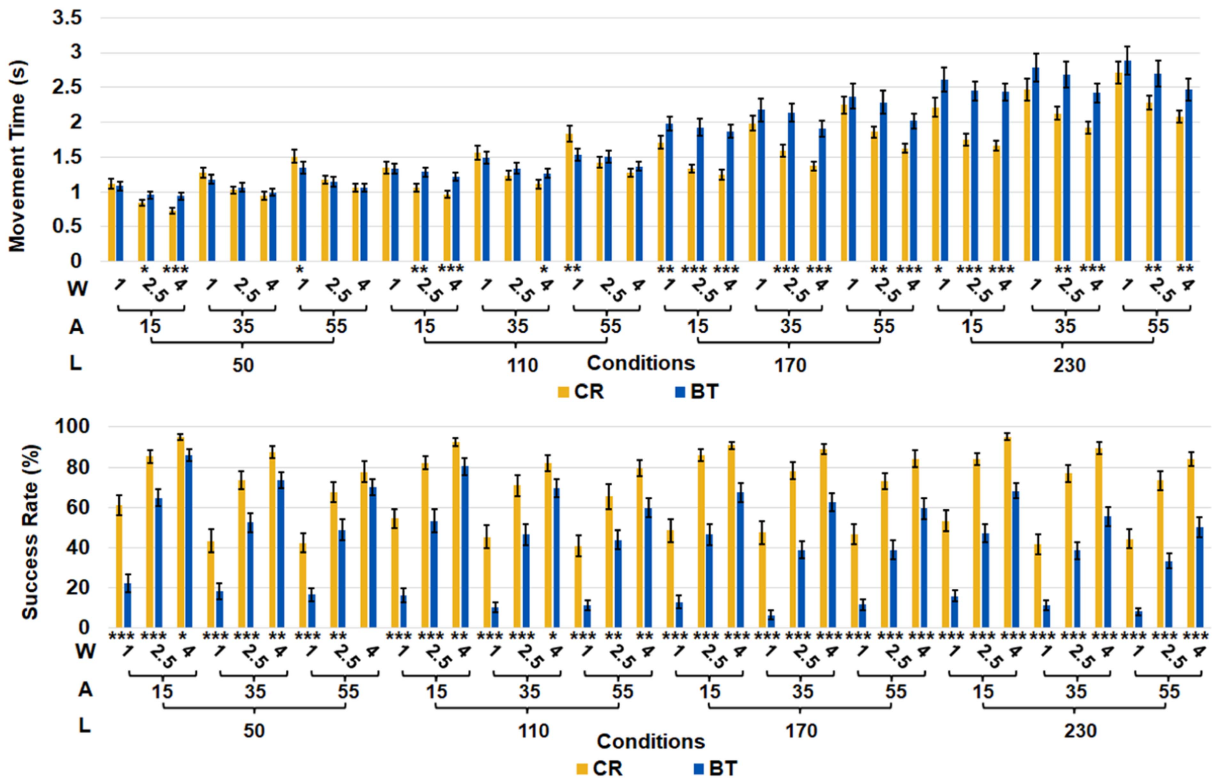


Fig. 4. TOP: Movement time of CR and BT conditions. BOTTOM: The success rate of CR and BT conditions. The '$*$,' '$**$,' and '$***$' indicate $p < 0.05$, $p < 0.01$ and $p < 0.001$ between the two interaction techniques conditions, respectively. Error bars represent standard errors.

1) Movement Time With CR: The RM-ANOVA showed significant main effects of L ($F_{3,90} = 200.584, p < 0.001, \eta_p^2 = 0.870$), A ($F_{2,60} = 233.446, p < 0.001, \eta_p^2 = 0.886$) and W ($F_{2,60} = 71.299, p < 0.001, \eta_p^2 = 0.704$) on the movement time. Three significant interaction effects were found: $L \times A$ ($F_{6,180} = 4.405, p < 0.001, \eta_p^2 = 0.128$), $L \times W$ ($F_{6,180} = 2.791, p = 0.013, \eta_p^2 = 0.085$), and $A \times W$ ($F_{3.666,109.977} = 4.066, p = 0.005, \eta_p^2 = 0.119$). Post-hoc pairwise comparison analysis was conducted to investigate the effect of L on movement time further. The results showed significant differences between each of the four conditions (all $p < 0.001$), with a higher latency condition leading to a longer movement time. Additionally, a strong linear relationship was found between L and movement time MT, expressed as $MT = 0.0059L + 0.7242$ ($R^2 = 0.9786$).

2) Movement Time With BT: Similar to the trends observed for CR, we identified similar trends in the BT condition. We found significant main effects of L ($F_{3,90} = 123.547, p < 0.001, \eta_p^2 = 0.805$), A ($F_{2,60} = 21.993, p < 0.001, \eta_p^2 = 0.423$), and W ($F_{6,180} = 19.644, p < 0.001, \eta_p^2 = 0.396$) on movement time. The result of the linear regression between L and MT was formulated as $MT = 0.0088L + 0.5553$ ($R^2 = 0.9771$). The effect of L was further verified by the pairwise comparison analysis, which showed significant differences between each two L levels (all $p < 0.001$). Finally, the only interaction effect on movement time that emerged was $A \times W$ ($F_{3.512,105.347} = 4.855, p = 0.002, \eta_p^2 = 0.139$).

3) Success Rate With CR: L did not exhibit a significant main effect on the success rate in the CR condition. The significant main effects of A ($F_{2,60} = 45.239, p < 0.001, \eta_p^2 = 0.601$) and W ($F_{2,60} = 157.131, p < 0.001, \eta_p^2 = 0.840$) were observed, where a larger target width and smaller movement amplitude resulted in a higher success rate. We also found a significant interaction effect of $L \times A$ ($F_{4.594,137.807} = 2.908, p = 0.019, \eta_p^2 = 0.088$) on success rate.

4) Success Rate With BT: In contrast to the CR condition, the results of RM-ANOVA tests showed a significant main effect of L on the success rate in the BT condition ($F_{3,90} = 21.506, p < 0.001, \eta_p^2 = 0.418$). There was also a significant interaction effect of $L \times W$ ($F_{5.051,151.532} = 2.788, p = 0.019, \eta_p^2 = 0.085$) on the success rate. Regarding L, a clear negative correlation was observed, where higher L resulted in a lower success rate. This relationship can be expressed through linear regression as $successrate = -0.0008L + 0.5301$ ($R^2 = 0.9437$). The results of post-hoc pairwise comparisons showed significant differences when L was 50 ms compared to 110 ms ($\Delta = 6.9\%$, $p = 0.019$), 170 ms ($\Delta = 12\%$, $p = 0.024$), and 230 ms ($\Delta = 13.8\%$, $p = 0.019$). At 110 ms, substantial increases were evident relative to 170 ms ($\Delta = 5.1\%, p = 0.035$) and 230 ms ($\Delta = 7\%$, $p = 0.019$). No significant effects were observed between 170 ms and 230 ms. Furthermore, concerning factors A ($F_{1.781,53.442} = 39.485, p < 0.001, \eta_p^2 = 0.568$) and W ($F_{1.759,52.773} = 291.581, p < 0.001, \eta_p^2 = 0.907$), significant main effects on the success rate were witnessed. When taking into account the interactions among these factors, $A \times W$

($F_{4,120} = 3.836, p = 0.012, \eta_p^2 = 0.113$) also demonstrated a significant influence on success rate.

J. Discussion

1) Responses to Hypotheses: The results from Sections III-G1 and III-H1 indicated that the endpoints in each condition followed a normal distribution on the x- and y-axis, respectively. Moreover, our results further revealed that no factors led to a statistically significant main effect on ρ, and the value of ρ was close to 0. This suggests that the variables X and Y do not influence each other's probability of occurrence, indicating their independence. Thus, **H1** is supported, indicating that the endpoint distribution under different levels of latency is likely to remain a bivariate Gaussian distribution.

H2 can be confirmed based on the results from Sections III-G and III-H, where the mean vector μ and the covariance matrix Σ were significantly influenced by latency. Furthermore, we examined the effect of latency L on the movement time and success rates, which revealed significant main effects on both. This result supports our **H3**. Moreover, based on the conclusions from **H2** and **H3**, **H4** was further supported. The results revealed distinct patterns regarding the effect of latency across the two interaction techniques. Specifically, latency significantly impacted the endpoint distribution, movement time, and success rate *more* when the interaction technique was *bare-hand touch* than the controller-based raycasting.

To gain further insight into the relationship between the factors and the dependent variables, we elaborate on these relationships and discuss the underlying reasons for these effects in the following sections.

2) Behaviors Interpretation on CR-Based Endpoint Distribution: Our results suggested that the formation of μ_x is significantly affected by the factors W, A, and the interaction factor $L \times A$. Specifically, we identified a strong negative linear relationship between W and μ_x ($R^2 = 0.9702$), as well as between A and μ_x ($R^2 = 0.9968$).

Regarding the effect of W on μ_x, we found that an increase in target size (W) leads users to select targets closer to their starting point. This behavior suggests that users take advantage of larger target sizes to reduce their movement distance, effectively optimizing their effort during selection. This is consistent with previous findings where participants adjust their movement strategies based on target size and required effort [59], [70].

Regarding the main effect of A on μ_x, where larger A values cause the distribution to generally shift toward the start position, the results align with the previous study focusing on head-based pointing selection [70]. Yu et al. deemed the "lazy effect," and the limited A conditions (from 30° to 40° with intervals of 5°) were two causes of this trend. In our experiment, although only three A levels were involved, we considered a more comprehensive situation to represent different levels of visibility caused by the limited FoV, a feature unique to VR devices (see Section III-E). Therefore, the "lazy effect" seems to have had an effect in our case, too.

An interesting pattern was observed in the effect of $L \times A$ on μ_x, as illustrated in Fig. 2 (LEFT). Specifically, with higher latency (L), the performance differences between various movement amplitudes (A) decreased, and improved performance was observed at longer A values. Based on participants' feedback during interviews, we attributed this tendency to an "effort effect." When latency was relatively low, participants preferred to select the target in a natural and comfortable manner. However, under challenging conditions—namely, high latency and large movement amplitudes—they exerted more effort and attention to control their movements and accurately select the target. To further verify this "effort effect." hypothesis, we examined the success rates and identified a similar trend. With increasing L, the long movement amplitude ($A = 55$) showed improved performance, increasing from 62.6% to 68% . In contrast, the short amplitude ($A = 15$) demonstrated reduced success rates, decreasing from 80.4% to 75.2% , while the middle amplitude ($A = 35$) showed a slight improvement, increasing from 68.2% to 69.3% .

In terms of μ_y, we discovered unexpected results that contrasted with previous works [59], [70], which typically set μ_y to zero. In our study, both W and A showed a statistically significant main effect on μ_y. Compared with A ($p = 0.001$), the factor of W presented a weaker effect ($p = 0.043$). We infer two reasons for these contrasts: (1) the insufficient sample size in previous studies, where the participants numbered only 16 to 18 for each study, might have caused the distribution of the endpoints to be more incomplete, and (2) the limitation of the FoV led to new behavioral patterns with A ranging from 15° to 55°, where participants had to move their heads to compensate for the limited FoV, resulting in new cooperation between hand, eye, and head, shaping new effects [19].

Specifically, there was a positive correlation between W and μ_y. When W increases, the center of distribution on the y-axis (μ_y) moves away from the target center. This indicates that a larger target size would reduce selection precision. This effect can be explained by the fact that with smaller targets, participants focus more on precise aiming. On the other hand, another positive correlation was observed between A and μ_y, suggesting that a longer movement amplitude leads to decreased selection precision. This relationship can be attributed to the additional selection noise introduced by longer movements. For instance, when A is small, participants can easily select the target using a straight trajectory. However, as A increases, participants require more coordination between their hands, heads, and eyes, making it difficult to maintain a straight trajectory throughout the entire selection process.

Regarding the endpoints deviation on the x-axis (σ_x) and y-axis (σ_y), we confirmed the same impact of A and W as observed in previous works [8], [16], [70]. The regression analysis demonstrated that larger target width and movement amplitude increased the deviation of endpoints on both axes. However, based on the dual-distribution hypothesis (see Section II-B) and previous findings described in Section II-A, we thought latency L would significantly affect the distribution, particularly Σ, where latency introduces more noise in the absolute precision of the devices and causes more uncertainty in people's selection strategy. Nevertheless, our results did not support this assumption, as L did not show any main effects on either σ_x or σ_y.

To better understand why this pattern emerged, we consulted with participants through our interviews to understand their selection strategies across different latency conditions. Most of them reported that when using CR, the ray was very clear, and it was easy for them to determine the pointing position. Therefore, even with high latency, they could control and adjust the ray until it was precisely aimed at the target. In other words, they preferred to spend more time adjusting the aiming position to ensure higher selection accuracy. These findings were further corroborated by the results discussed in Section III-I, where the success rate was not influenced by latency, but the movement time increased significantly as latency increased.

3) Behaviors Interpretation on BT-Based Endpoints Distribution: We identified more pronounced latency effects in the bare-hand touch (BT) condition. Latency L influenced the entire distribution (μ_x, μ_y, σ_x, and σ_y). Additionally, movement amplitude A also demonstrated strong effects on μ_x, σ_x, and σ_y. However, no main effect was observed for target width W.

First, we examined the effect of A and L on μ_x and μ_y. Similar to the CR condition, a negative correlation between A and μ_x was found in the BT condition. We also attribute this to the "lazy effect". Notably, while latency L showed a main effect on μ_x, our empirical analysis demonstrated that, compared to movement amplitude A, the effect of L is negligible based on the 95% CI values and significance of the coefficients (see Section III-G). Besides, L and μ_y exhibited a logarithmic relationship, where μ_y increased sharply with the initial rise in L before the trend gradually leveled off. It should be noted that when latency is low, μ_y remains negative. As latency increases, μ_y approaches the center of the target. We assume that this phenomenon is similar to the "effort effect," where users tend to put more effort and be more careful in selecting targets when they perceive high latency.

Second, we discerned the impact of L on σ_x and σ_y, which demonstrated a tendency for greater latency to generate a larger distribution on both the x- and y-axis simultaneously. From our interviews, we identified two reasons for this effect: (1) The BT method requires participants to aim and poke the target with their index finger, which is inherently less precise due to the "fat finger" effect [8]. As a result, added latency further diminishes the ability to make precise selections in the BT method more significantly than in the CR method. (2) Participants expected a higher sense of embodiment when using the BT method, which also made them more sensitive to the mismatch between their actual hand movements and the rendered hand movements in the VE. This mismatch reduced their confidence and intuitiveness in controlling their virtual hands for selection compared to the CR method, especially under high-latency conditions.This negative feeling eventually led to reduced precision. With the CR method, participants perceived it as an external device rather than a part or extension of their body and were more accepting of the delays in its movement.

Third, regarding the factor of movement amplitude A, a positive correlation between A and the covariance matrix Σ was observed. This relationship was the same as in prior work; please refer to [70] for more details.

Finally, as for the absence of the main effect for target size W in the BT-based endpoint distribution, we consider it to be due to the "fat finger effect" [8] and the mismatch between actual and rendered movements. Despite designing three representative target sizes, the widths of 1° and 2.5° were still largely obscured by participants' index fingers, hindering precise adjustment to the target's center. Participants often had to rely on intuition to select targets, especially for small or medium-sized targets. Moreover, the mismatch between their real hand movements and the movements rendered in the VE further impeded their ability to make precise selections. As a result, W did not significantly influence the endpoint distribution. Additionally, this "fat finger" effect movement mismatch was also evident in the success rate, as discussed in Section III-I4. The results indicated that W is the most significant factor influencing the selection success rate, with a statistical significance ($p < 0.001$) and the highest effect size ($\eta_p^2 = 0.907$).

4) Effects of Latency on Movement Time and Success Rate: Section III-I highlighted the discrepancy between the two interaction techniques (IT) regarding movement time and success rates. We observed a consistent impact of L on movement time. The results indicate that users spend more time refining and adjusting their selection to mitigate the negative effects caused by high latency. However, the success rate did not exhibit the same tendency, as only the BT was influenced by latency. We attribute this to the inherent differences between the two interaction techniques, such as the absolute pointing precision (the "fat finger" effect) and the subjective experience under high latency (the mismatch issues) between them, as discussed in the aforementioned sections. In brief, consistent with our hypothesis in **H4**, we found that bare-hand touch is more susceptible to latency than controller-based raycasting, with stronger effects observed on both endpoint distribution and success rate. These results highlight the importance of deriving models tailored to individual interaction methods rather than treating them as identical.

K. Summary of Behavioral Patterns

Based on the aforementioned results and discussion, we identified some interesting and important behavioral patterns:

- With increased latency, the effect of movement amplitude on μ_x is diminished when using controller-based raycasting as the interaction technique.
- An increased latency does not influence the standard deviation of the endpoints on both the x- and y-axes when using controller-based raycasting.
- When the interaction technique is bare-hand touch, target width does not affect the distribution of endpoints.
- Latency can significantly affect the entire endpoint distribution when using bare-hand touch.
- Increased latency extends the selection duration regardless of the interaction techniques (either controller-based raycasting or bare-hand touch).
- Increasing latency would lower the success rate when using bare-hand touch for selection, but would not affect the success rate when using controller-based raycasting.

IV. Model Formulation

In this section, we formulated two endpoint distribution models based on the empirical results obtained from the user study.

A. Endpoint Distribution Model for CR-Based Selection

Based on the results discussed in Section III-G, we can first construct a bivariate Gaussian distribution model for the CR-based endpoints:

$$\boldsymbol{\mu} = \begin{bmatrix} g\mathrm{L} + h\mathrm{A} + i\mathrm{LA} + j\mathrm{W} + k \\ l\mathrm{W} + m\mathrm{A} + n \end{bmatrix},$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} (a\mathrm{W} + b\mathrm{A} + c)^2 & 0 \\ 0 & (d\mathrm{W} + e\mathrm{A} + f)^2 \end{bmatrix} \quad (3)$$

Where A, W, and L represent the movement amplitude, target width, and latency, respectively. a, b, c, d, e, f, g, h, i, j, k, l, m, and n are constants. While non-linear regression can provide coefficients that achieve the best-fitting results, it potentially makes it difficult to interpret users' behavioral patterns [67]. Thus, to maintain our models' interpretability regarding users' selection behaviors, we only conducted linear regression to obtain accurate intercept values without revising the coefficients for each factor. Therefore, based on the linear regression of the user study results, the following coefficients were obtained: $a = 0.1089$, $b = 0.0052$, $c = 0.3223$, $d = 0.0839$, $e = 0.0045$, $f = 0.2702$, $g = -0.0017$, $h = -0.0162$, $i = 0.00006$, $j = -0.0414$, $k = 0.7131$, $l = 0.0114$, $m = 0.0028$, and $n = -0.0392$.

B. Endpoint Distribution Model for BT-Based Selection

Given there are differences between the two interaction techniques, another model for BT-based endpoints was proposed using different parameters and coefficients:

$$\boldsymbol{\mu} = \begin{bmatrix} g\mathrm{A} + h \\ i\ln(L) + j \end{bmatrix},$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} (a\mathrm{L} + b\mathrm{A} + c)^2 & 0 \\ 0 & (d\mathrm{L} + e\mathrm{A} + f)^2 \end{bmatrix} \quad (4)$$

Where $a = 0.0031$, $b = 0.0108$, $c = 0.6731$, $d = 0.0015$, $e = 0.0078$, $f = 0.8145$, $g = -0.0133$, $h = 0.0039$, $i = 0.0706$, and $j = -0.3461$.

C. Overfitting Risk

Our models for CR and BT include 14 and 10 coefficients, respectively. However, we believe that our models do not encounter the risk of overfitting for the following reasons: (1) Each dependent variable (μ_x, μ_y, σ_x, and σ_y) reflects its own attribute, meaning the coefficients for one factor do not influence the performance of others. (2) All coefficients are derived from the empirical results of the user study, ensuring they are interpretable and meaningful (please refer to our discussion in Section III-J). (3) We utilized linear regression to formulate the relationship between independent and dependent variables, as it reveals clearer and interpretable relationships and is generally less prone to overfitting compared to nonlinear models. Therefore, we believe the total number of coefficients does not pose a risk of overfitting.

V. User Study 2: Model Verification and Application

In this section, we conducted a second user study to evaluate the effectiveness, reproducibility, and generalizability of the proposed models. We collected a new dataset that incorporates revised values for factors such as target size and latency. Using this new dataset, we applied the two proposed models to refine the existing selection accuracy prediction model, which previously did not account for latency effects.

A. Participants and Apparatus

Sixteen participants (7 males and 9 females) who did not participate in the first user study were recruited from the same local university. They were aged 19 to 27, with a mean age of 21.25 and a standard deviation of 2.87. We utilized the same measures as in the previous study to assess their familiarity with VR headsets, yielding an average score of 4.38 ($s.d. = 1.77$). All participants reported they could clearly see objects within the scene. The same experimental setup and latency control method from the first study were used.

B. Task, Design, and Procedure

The same task from our first user study was employed. A $3 \times 3 \times 5 \times 2$ within-subjects design was used, incorporating three target widths W (1°, 1.5°, and 2°), three movement amplitudes A (15°, 35°, and 55°), five levels of latency L (ranging from 50 ms to 250 ms, increasing in 50 ms intervals), and two interaction techniques IT (CR and BT). This design allows our experimental conditions to span a wide array of application scenarios, from extreme cases like selecting very small targets with low-performance devices or under high computational load conditions ($W = 1°$, $A = 55°$, and $L = 250$ ms) to everyday user interface interactions ($W = 2°$, $A = 15°$, and $L = 50$ ms). In total, each participant took approximately 55 minutes to complete all conditions. After each block ($1W \times 1A \times 1L \times 1IT$), participants could rest until they felt ready to continue to the next condition. A total of 14,400 endpoints were collected ($3W \times 3A \times 5L \times 2IT \times 16$ participants $\times 10$ repetitions) from this follow-up study.

C. Model Integration and Baselines

1) Model Integration: A target has been selected if the endpoint is within its boundary. This simple and intuitive criterion, often referred to in the literature as the Visual Boundary Criterion (VBC), is most commonly used to determine the success of a target selection task. Based on VBC, we can interpret the probability of successful selection as the probability of the endpoint being situated within the border [10]. The success rate can be calculated through the probability density function:

$$y = \iint_D f\left(\begin{bmatrix} x \\ y \end{bmatrix}, \boldsymbol{\mu}, \Sigma\right) dx, dy \quad (5)$$

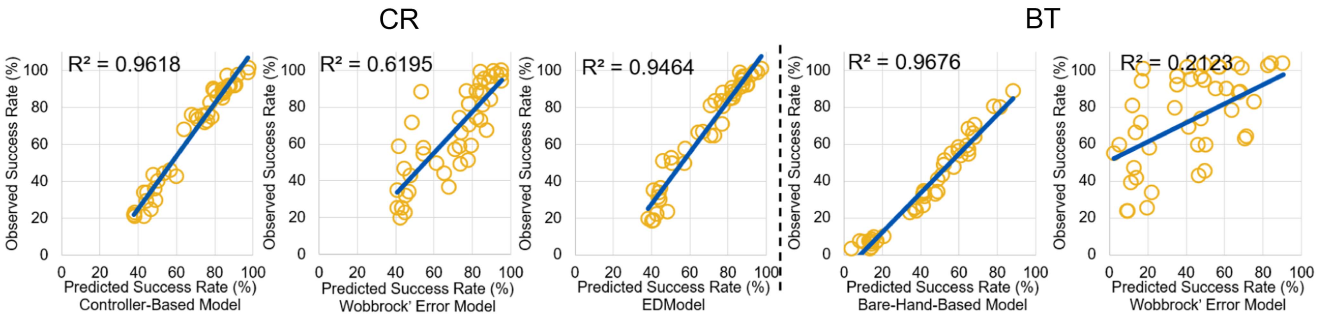


Fig. 5. The visualized fitting results ($n = 45$). From left to right: the first three sub-figures show the fitting results of our proposed model, Wobbrock's error model, and EDModel using the dataset from the CR condition, respectively. The last two sub-figures show the fitting results from our proposed model and Wobbrock's error model using the dataset from the BT condition.

where D refers to the boundary of the target, and μ and Σ represent the mean vector and the covariance matrix. Moreover, given that both of our models demonstrated a correlation ρ of zero, the above equation can be expressed as:

$$y = \iint_D \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}\right) dxdy \tag{6}$$

Thus, we are able to estimate the success rate for each condition ($1L \times 1A \times 1W$) after substituting μ_x, μ_y, σ_x, and σ_y with the parameters from our models.

2) Baselines: To better validate the prediction capability of our models, we selected two models from prior work as our baselines for comparison. For the CR-based selection, we chose the EDModel, which was also proposed based on empirical results and an experimental setup similar to ours [70]. The key difference is that EDModel did not account for the latency effect. The expression of the EDModel is:

$$\boldsymbol{\mu} = \begin{bmatrix} gW + h \\ 0 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} (a\mathrm{A} + b\mathrm{W} + c)^2 & 0 \\ 0 & (d\mathrm{A} + e\mathrm{W} + f)^2 \end{bmatrix} \tag{7}$$

where $a - h$ are empirical values obtained from their user study: $a = 0.0066$, $b = 0.1025$, $c = 0.2663$, $d = 0.0085$, $e = 0.0679$, $f = 0.1437$, $g = -0.1441$, and $h = 0.2649$.

Second, we adopted Wobbrock's error model [61] as the baseline for both interaction techniques (CR and BT). This error rate model was derived from Fitts' Law:

$$P_{\text{error}} = 1 - \text{erf}\left\{\frac{1}{A\sqrt{2}}\left[2.066 \cdot W\left(2^{\frac{MT_e - a'}{b'}} - 1\right)\right]\right\} \tag{8}$$

MT in the equation refers to the movement time, while W and A represent the target width and movement amplitude, respectively. The a and b are empirical parameters derived from Fitts' law. Numerous works have confirmed the predictability of this model in 1D [61], 2D [62], and 3D [70] scenarios.

D. Data Pre-Process and Results

After the experiment, the data were pre-processed using the same approach as in the previous study. Out of a total of 14,400 trials, 925 trials (6.42%) were removed, including 452 trials from the CR condition and 473 trials from the BT condition.

The results, summarized in Fig. 5, show that our models achieved the best performance for both CR ($R^2 = 0.9618$, MAE = 4.95%) and BT ($R^2 = 0.9617$, MAE = 6.25%). In contrast, Wobbrock's error model performed the worst for CR ($R^2 = 0.6195$, MAE = 13.58%) and BT ($R^2 = 0.2123$, MAE = 36.77%). The EDModel showed a slightly worse performance than our models ($R^2 = 0.9464$, MAE = 7.92%). For CR, our models improved R^2 by 34.23% over Wobbrock's model and 1.54% over the EDModel. MAE was improved 8.63% and 2.97% , respectively. For BT, our model significantly improved prediction accuracy, with R^2 increasing by 75.53% and MAE by 30.52% .

E. Discussion

We attribute the discrepancy between our models and the baselines to the model's interpretability of the factors. For instance, Wobbrock's error model considers MT to be a significant factor influencing the success rate. However, as demonstrated in Section III-I4, while latency has a main effect on MT with CR, no statistical main effect on the success rate was observed. This is contrary to Wobbrock's error model. Additionally, Wobbrock's error model needs the coefficients of a and b to be derived from Fitts' Law. However, the original Fitts' Law cannot explain the effect of latency, resulting in biased coefficients (a and b), with $R^2 = 0.2704$ and 0.0312 for CR and BT, respectively. Although refinements to Fitts' Law have been proposed (e.g., [18]), which can enhance the fitness to over 95% across varying latency, they introduce an additional coefficient c that cannot be integrated into Wobbrock's error model. The results from the EDModel further validated our explanation, achieving performance similar to that of our model under the CR condition. These results align with our findings from the user study, where latency did not significantly influence the success rate when using the controller. Thus, even though EDModel did not account for latency effects, it still demonstrated good predictability for success rates. The minor improvement in our model over the EDModel is attributed to our exploration of new effects on μ_x and μ_y, leading to stronger interpretability of users' selection performance and behavioral patterns. Finally, the results for BT-based selection disclosed another reason why Wobbrock's error model could not be utilized in this scenario because Wobbrock's error model considers W as a key factor, but our findings indicate no main effect of W on the distribution of the BT-based endpoints.

In summary, our models achieve the best performance in predicting success rates with both interaction techniques in all metrics. Moreover, we demonstrated the reliability and applicability of our models, which can be valuable tools for game and interface designers to evaluate and improve their designs.

VI. Summary of Findings and Recommendations

In this section, we provide a brief review of our paper along with associated recommendations.

In the first user study (Section III), the behavioral data were collected across various conditions and scenarios, including two interaction techniques (IT), four latency levels (L), three target widths (W), and three movement amplitudes (A). We measured seven dependent variables: μ_x, μ_y, σ_x, σ_y, ρ, movement time, and success rate. After analyzing our collected dataset, several distinct effects and behavior patterns were identified. Specifically, while we confirmed the effects of latency on both interaction techniques, it exhibited a more pronounced impact on controller-based raycasting, significantly affecting μ_y, σ_x, and σ_y. Additionally, an interesting pattern was observed in movement time and success rate when using the controller-based raycasting technique—the increased latency significantly prolonged selection time without compromising accuracy. This pattern did not hold for bare-hand touch selection. The behavior pattern was more intuitive when using the bare-hand touch technique, where increased latency led to longer selection times and reduced success rates.

Although our study design has considered the target widths users will face in daily scenarios (from 1 to 4.5 visual degrees), it would be interesting to introduce larger target widths to analyze their effect. Such investigations would be more worthwhile to discuss in bare-hand touch conditions, as we did not observe target width to have shown any impact on their endpoint distribution. In the future, we aim to expand our participant pool to include a more diverse population to validate these behavioral patterns further.

We formulated two models based on the main effects identified between factors and dependent variables (Section IV). These two models correspond to the two interaction techniques separately, considering their differences and the distinct effects observed. Additionally, a brief discussion was included to address potential over-fitting concerns.

To avoid bias caused by the initial dataset, a second user study was conducted using newly collected data from 16 participants with different latency L and target width W to further validate the applicability and reproducibility of our models (Section V). The two models were utilized to predict success rates by applying the probability density function of the bivariate Gaussian distribution across varying conditions. Fig. 5 illustrates that our models demonstrated the best performance compared to the baselines across all metrics (R^2 and MAE), showcasing their strong ability to predict success rates.

Finally, we propose the following recommendations to aid VR game and interface designers so they can provide enhanced user experience, especially in complex scenarios.

- In virtual environments, the size of the selectable elements should be larger when using bare-hand touch as the interaction technique, even if the size is sufficient for controller-based raycasting selection. Specific values should be determined based on the particular application scenario. Our study results indicated that a target width of 2 visual degrees, which allows for good accuracy (over 75%) with controllers, yields only about a 40% success rate when using bare-hand touch.
- When designing user interfaces for VR HMDs, it is necessary to account for the effects of latency, which can cause significant shifting and scaling of the endpoint distribution, thereby affecting the interaction efficiency and accuracy.
- Raycasting is a better choice for high-latency interaction scenarios, as it could significantly enhance selection accuracy. This recommendation is informed by our observation that although high latency increased selection time for both techniques, severe performance degradation was observed only in the bare-hand touch.

VII. Limitations and Future Work

While our study provides empirical evidence and predictive models on how latency influences target selection in VR, several limitations should be acknowledged and could serve as directions for future work.

First, our experimental design compared controller–raycasting and bare-hand–touch as two representative techniques. This choice, while motivated by prior recommendations, inevitably couples two dimensions—input modality (controller versus hand) and interaction metaphor (ray versus touch). As newer headsets support both metaphors with both input types, in our future work, we plan to decouple these two dimensions and examine them independently.

Second, we instructed participants to perform selections "naturally". This design revealed how participants adapted their strategies to task difficulty, but it also introduced behavioral variability that may confound the interpretation of latency effects. Future work may consider complementing this design with more constrained performance settings and instructions (e.g., "as fast and accurate as possible") to better isolate the influence of latency from task-related strategies.

Third, the experiments were conducted exclusively on a single device (Meta Quest 3) under controlled laboratory conditions. Although this ensured internal validity, the findings may not fully generalize to other headsets or contexts. We plan to extend our validation to different hardware platforms and interaction scenarios to strengthen the external validity of our models, similar to Schneider et al. [49].

Finally, our current analyses focused on speed, accuracy, and endpoint distributions. While these measures capture the core aspects of performance, they do not cover the full spectrum of user experience. We plan to broaden the scope by incorporating additional indicators such as error offsets in unsuccessful trials, error distribution patterns, and subjective measures of workload, comfort, cybersickness, and satisfaction.

VIII. Conclusion

In this paper, we first collected, analyzed, and interpreted users' selection behavior and performance in target selection tasks in VR under different latency, target width, and movement amplitude levels. The selection was completed using controller-based raycasting and bare-hand touch, the two most commonly used selection approaches in VR scenarios. Based on the results, two novel endpoint-based models were proposed, one for each technique. There are two notable differences between these two models and models from prior work: (1) our models incorporate the effect of latency, and (2) we first use endpoint distribution to model and understand users' selection behaviors and performance in VE-based bare-hand pointing tasks. A second user study was conducted to validate our models with a new dataset. In this study, we refined the current model for predicting success rates. The results demonstrated that our models consistently outperformed other models proposed in prior research across all interaction techniques and metrics. Finally, based on our findings, three recommendations were proposed to aid game and interface designers in creating more interesting and efficient gameplay elements and interactive techniques tailored for VR HMDs.

Acknowledgment

We thank the participants for their time and the reviewers for their insightful comments, which helped improve our article.

References

[1] Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9 - Requirements for non-keyboard input devices, International Organization for Standardization, Feb. 2000, pp. 1–57.
[2] D. Abdlkarim et al., "A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the meta quest 2," *Behav. Res. Methods*, vol. 56, no. 2, pp. 1052–1063, 2024.
[3] A. Adkins, L. Lin, A. Normoyle, R. Canales, Y. Ye, and S. Jörg, "Evaluating grasping visualizations and control modes in a VR game," *ACM Trans. Appl. Percept.*, vol. 18, no. 4, Oct. 2021, Art. no. 19, doi: 10.1145/3486582.
[4] T. Aila, V. Miettinen, and P. Nordlund, "Delay streams for graphics hardware," in *Proc. ACM SIGGRAPH Papers*, New York, NY, USA, 2003, pp. 792–800, doi: 10.1145/1201775.882347.
[5] G. Anderson, R. Doherty, and S. Ganapathy, "User perception of touch screen latency," in *Proc. Int. Conf. Des. User Exp. Usability*, Springer, Berlin Heidelberg, 2011, pp. 195–202.
[6] G. Armitage, "An experimental estimation of latency sensitivity in multiplayer quake 3," in *Proc. 11th IEEE Int. Conf. Netw.*, 2003, pp. 137–141, doi: 10.1109/ICON.2003.1266180.
[7] M. Baloup, T. Pietrzak, and G. Casiez, "RayCursor: A 3D pointing facilitation technique based on raycasting," in *Proc. 2019 CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2019, pp. 1–12, doi: 10.1145/3290605.3300331.
[8] X. Bi, Y. Li, and S. Zhai, "FFitts law: Modeling finger touch with Fitts' law," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2013, pp. 1363–1372, doi: 10.1145/2470654.2466180.
[9] X. Bi and S. Zhai, "Bayesian Touch: A statistical criterion of target selection with finger touch," in *Proc. 26th Annu. ACM Symp. User Interface Softw. Technol.*, New York, NY, USA, 2013, pp. 51–60, doi: 10.1145/2501988.2502058.
[10] X. Bi and S. Zhai, "Predicting finger-touch accuracy based on the dual Gaussian distribution model," in *Proc. 29th Annu. Symp. User Interface Softw. Technol.*, New York, NY, USA, 2016, pp. 313–319, doi: 10.1145/2984511.2984546.
[11] M. Claypool, "On models for game input with delay–moving target selection with a mouse," in *Proc. 2016 IEEE Int. Symp. Multimedia*, 2016, pp. 575–582, doi: 10.1109/ISM.2016.0125.
[12] M. Claypool, "Game input with delay–moving target selection with a game controller thumbstick," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 3s, Jun. 2018, Art. no. 57, doi: 10.1145/3187288.
[13] M. Claypool, A. Cockburn, and C. Gutwin, "Game input with delay: Moving target selection parameters," in *Proc. 10th ACM Multimedia Syst. Conf.*, New York, NY, USA, 2019, pp. 25–35, doi: 10.1145/3304109.3306232.
[14] M. Claypool, A. Cockburn, and C. Gutwin, "The impact of motion and delay on selecting game targets with a mouse," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 2s, pp. 73:1–73:24, Apr. 2020, doi: 10.1145/3390464.
[15] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Exp. Psychol.*, vol. 47, no. 6, pp. 381–391, 1954.
[16] T. Grossman and R. Balakrishnan, "A probabilistic approach to modeling two-dimensional pointing," *ACM Trans. Comput.-Hum. Interact.*, vol. 12, no. 3, pp. 435–459, Sep. 2005, doi: 10.1145/1096737.1096741.
[17] C. Guo et al., "Pingmesh: A large-scale system for data center network latency measurement and analysis," in *Proc. 2015 ACM Conf. Special Int. Group Data Commun.*, New York, NY, USA, 2015, pp. 139–152, doi: 10.1145/2785956.2787496.
[18] E. R. Hoffmann, "Fitts' law with transmission delay," *Ergonomics*, vol. 35, no. 1, pp. 37–48, 1992, doi: 10.1080/00140139208967796.
[19] A. Horstmann and K. Hoffmann, "Target selection in eye-hand coordination: Do we reach to where we look or do we look to where we reach?," *Exp. Brain Res.*, vol. 167, no. 2, pp. 187–195, 2005, doi: 10.1007/s00221_005_0038_6.
[20] J. Huang and B. Lee, "Modeling error rates in spatiotemporal moving target selection," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–12, doi: 10.1145/3290607.3313077.
[21] J. Huang, F. Tian, X. Fan, and S. Zhai, "Modeling the endpoint uncertainty in crossing-based moving target selection," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–13, 2019, doi: 10.1145/3355089.3356544.
[22] J. Huang, F. Tian, X. Fan, and S. Zhai, "Modeling the uncertainty in 2D moving target selection," in *Proc. ACM Symp. User Interface Softw. Technol.*, 2019, pp. 1–12, doi: 10.1145/3332165.3347934.
[23] J. Huang, F. Tian, N. Li, and X. Fan, "Modeling the uncertainty in 2D moving target selection," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, New York, NY, USA, 2019, pp. 1031–1043, doi: 10.1145/3332165.3347880.
[24] J. Huang, X. Zhang, F. Tian, X. Fan, and S. Zhai, "Understanding the uncertainty in 1D unidirectional moving target selection," in *Proc. Extended Abstr. 2018 CHI Conf. Hum. Factors Comput. Syst.*, 2018, pp. 1–6, doi: 10.1145/3170427.3188613.
[25] T. Isomoto, T. Ando, B. Shizuki, and S. Takahashi, "Dwell time reduction technique using fitts' law for gaze-based target acquisition," in *Proc. 2018 ACM Symp. Eye Tracking Res. Appl.*, New York, NY, USA, 2018, Art. no. 26, doi: 10.1145/3204493.3204532.
[26] R. Jiang and Z. Gu, "Current theoretical developments and applications of Fitts' law: A literature review," in *Proc. AHFE 2019 Int. Conf. Ergonom. Des.*, Washington DC, USA, Springer, 2020, pp. 753–760.
[27] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions, Volume 1: Models and Applications*, vol. 1, Hoboken, NJ, USA: Wiley, 2004.
[28] J. J. LaViola Jr., E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev, *3D User Interfaces: Theory and Practice*. Boston, MA, USA: Addison-Wesley Professional, 2017.
[29] M. Le Chénéchal and J. Chatel-Goldman, "HTC Vive Pro time performance benchmark for scientific research," in *Proc. Int. Conf. Artif. Reality Telexistence Eurographics Symp. Virtual Environ.*, 2018, pp. 81–84.
[30] I. Lee, S. Kim, and B. Lee, "Geometrically compensating effect of end-to-end latency in moving-target selection games," in *Proc. 2019 CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2019, pp. 1–12, doi: 10.1145/3290605.3300790.
[31] L. Lin et al., "The effect of hand size and interaction modality on the virtual hand illusion," in *Proc. 2019 IEEE Conf. Virtual Reality 3D User Interfaces*, 2019, pp. 510–518, doi: 10.1109/VR.2019.8797787.
[32] X. Liu, Y. Zhang, Z. Yan, and Y. Ge, "Defining 'seamlessly connected': User perceptions of operation latency in cross-device interaction," *Int. J. Hum.- Comput. Stud.*, vol. 177, 2023, Art. no. 103068.
[33] T. Luong, Y. F. Cheng, M. Möbus, A. Fender, and C. Holz, "Controllers or bare hands? A controlled evaluation of input techniques on interaction performance and exertion in virtual reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 11, pp. 4633–4643, Nov. 2023, doi: 10.1109/TVCG.2023.3320211.

[34] I. S. MacKenzie, "Fitts' law as a research and design tool in human-computer interaction," *Hum.- Comput. Interaction*, vol. 7, no. 1, pp. 91–139, 1992.

[35] I. S. MacKenzie and C. Ware, "Lag as a determinant of human performance in interactive systems," in *Proc. INTERACT '93 CHI '93 Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 1993, pp. 488–493, doi: 10.1145/169059.169431.

[36] A. Masurovsky, P. Chojecki, D. Runde, M. Lafci, D. Przewozny, and M. Gaebler, "Controller-free hand tracking for grab-and-place tasks in immersive virtual reality: Design elements and their empirical study," *Multimodal Technol. Interaction*, vol. 4, no. 4, 2020, Art. no. 91, doi: 10.3390/mti4040091.

[37] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks, "Effect of latency on presence in stressful virtual environments," in *Proc. IEEE Virtual Reality*, USA, 2003, pp. 141–148.

[38] X. Meng, W. Xu, and H.-N. Liang, "An exploration of hands-free text selection for virtual reality head-mounted displays," in *Proc. 2022 IEEE Int. Symp. Mixed Augmented Reality*, Los Alamitos, CA, USA, 2022, pp. 74–81, doi: 10.1109/ISMAR55827.2022.00021.

[39] D. M. Mifsud, A. S. Williams, F. Ortega, and R. J. Teather, "Augmented reality Fitts' law input comparison between touchpad, pointing gesture, and raycast," in *Proc. 2022 IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, 2022, pp. 590–591, doi: 10.1109/VRW55335.2022.00146.

[40] K. Minakata, J. P. Hansen, I. S. MacKenzie, P. Bækgaard, and V. Rajanna, "Pointing by gaze, head, and foot in a head-mounted display," in *Proc. 11th ACM Symp. Eye Tracking Res. Appl.*, New York, NY, USA, 2019, Art. no. 69, doi: 10.1145/3317956.3318150.

[41] M. A. Moni and A. B. M. S. Ali, "HMM based hand gesture recognition: A review on techniques and approaches," in *Proc. 2nd IEEE Int. Conf. Comput. Sci. Inf. Technol.*, 2009, pp. 433–437, doi: 10.1109/ICCSIT.2009.5234536.

[42] A. Ng, J. Lepinski, D. Wigdor, S. Sanders, and P. Dietz, "Designing for low-latency direct-touch input," in *Proc. 25th Annu. ACM Symp. User Interface Softw. Technol.*, New York, NY, USA, 2012, pp. 453–464, doi: 10.1145/2380116.2380174.

[43] C. A. Ntuen and L. D. Rogers, "Effects of information presentation refresh rate and moving objects on saccade latency," *Int. J. Cogn. Ergonom.*, vol. 5, no. 2, pp. 137–147, 2001, doi: 10.1207/S15327566IJCE05024.

[44] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: A review of techniques," *J. Imag.*, vol. 6, no. 8, pp. 1–29, 2020, doi: 10.3390/jimaging6080073.

[45] A. Pavlovych and W. Stuerzlinger, "The tradeoff between spatial jitter and latency in pointing tasks," in *Proc. 1st ACM SIGCHI Symp. Eng. Interactive Comput. Syst.*, New York, NY, USA, 2009, pp. 187–196, doi: 10.1145/1570433.1570469.

[46] S. Ping, "Delay measurement time synchronization for wireless sensor networks," *Intel Res. Berkeley Lab*, vol. 6, pp. 1–10, 2003.

[47] 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 *Proc. 5th Symp. Spatial User Interaction*, New York, NY, USA, 2017, pp. 91–98, doi: 10.1145/3131277.3132182.

[48] H.-R. Rantamaa, J. Kangas, S. K. Kumar, H. Mehtonen, J. Järnstedt, and R. Raisamo, "Comparison of a VR stylus with a controller, hand tracking, and a mouse for object manipulation and medical marking tasks in virtual reality," *Appl. Sci.*, vol. 13, no. 4, 2023, Art. no. 2251, doi: 10.3390/app13042251.

[49] D. Schneider et al., "Accuracy evaluation of touch tasks in commodity virtual and augmented reality head-mounted displays," in *Proc. 2021 Symp. Spatial User Interaction*, New York, NY, USA, 2021, pp. 1–11, doi: 10.1145/3485279.3485283.

[50] R. Shi et al., "Experimental analysis of freehand multi-object selection techniques in virtual reality head-mounted displays," *Proc. ACM Hum.-Comput. Interact.*, vol. 8, Oct. 2024, Art. no. 529, doi: 10.1145/3698129.

[51] R. Shi, Y. Wei, X. Qin, P. Hui, and H.-N. Liang, "Exploring gaze-assisted and hand-based region selection in augmented reality," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, May 2023, Art. no. 160, doi: 10.1145/3591129.

[52] R. Shi, J. Zhang, Y. Yue, L. Yu, and H.-N. Liang, "Exploration of bare-hand mid-air pointing selection techniques for dense virtual reality environments," in *Proc. Extended Abstr. 2023 CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2023, Art. no. 109, doi: 10.1145/3544549.3585615.

[53] J. Spjut et al., "Latency of 30 ms benefits first person targeting tasks more than refresh rate above 60 Hz," in *Proc. SIGGRAPH Asia Tech. Briefs*, New York, NY, USA, 2019, pp. 110–113, doi: 10.1145/3355088.3365170.

[54] J. Wagner, W. Stuerzlinger, and L. Nedel, "Comparing and combining virtual hand and virtual ray pointer interactions for data manipulation in immersive analytics," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 5, pp. 2513–2523, May 2021, doi: 10.1109/TVCG.2021.3067759.

[55] J. Wang, R. Shi, W. Zheng, W. Xie, D. Kao, and H.-N. Liang, "Effect of frame rate on user experience, performance, and simulator sickness in virtual reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 5, pp. 2478–2488, May 2023, doi: 10.1109/TVCG.2023.3247057.

[56] Y. Wang, M. C. Meyer, and J. Wang, "Real-time delay minimization for data processing in wirelessly networked disaster areas," *IEEE Access*, vol. 7, pp. 2928–2937, 2019, doi: 10.1109/ACCESS.2018.2886075.

[57] C. Ware and R. Balakrishnan, "Reaching for objects in VR displays: Lag and frame rate," *ACM Trans. Comput.-Hum. Interact.*, vol. 1, no. 4, pp. 331–356, Dec. 1994, doi: 10.1145/198425.198426.

[58] Y. Wei et al., "Evaluating and modeling the effect of frame rate on steering performance in virtual reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 31, no. 9, pp. 5447–5461, Sep. 2025.

[59] Y. Wei et al., "Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions," in *Proc. 2023 CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2023, Art. no. 283, doi: 10.1145/3544548.3581042.

[60] A. T. Welford, *Fundamentals of Skill*. London, U.K.: Methuen, 1968.

[61] J. O. Wobbrock, E. Cutrell, S. Harada, and I. S. MacKenzie, "An error model for pointing based on fitts' law," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2008, pp. 1613–1622, doi: 10.1145/1357054.1357306.

[62] J. O. Wobbrock, K. Shinohara, and A. Jansen, "The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2011, pp. 1639–1648, doi: 10.1145/1978942.1979181.

[63] W. Xu, H.-N. Liang, A. He, and Z. Wang, "Pointing and selection methods for text entry in augmented reality head mounted displays," in *Proc. 2019 IEEE Int. Symp. Mixed Augmented Reality*, 2019, pp. 279–288, doi: 10.1109/ISMAR.2019.00026.

[64] Y. Xu et al., "Practical optimizations for lightweight distributed file system on consumer devices," *CCF Trans. High Perform. Comput.*, vol. 4, no. 4, pp. 474–491, 2022.

[65] S. Yamanaka, T. Kinoshita, Y. Oba, R. Tomihari, and H. Miyashita, "Varying subjective speed-accuracy biases to evaluate the generalizability of experimental conclusions on pointing-facilitation techniques," in *Proc. 2023 CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2023, Art. no. 317, doi: 10.1145/3544548.3580740.

[66] S. Yamanaka and W. Stuerzlinger, "The effect of latency on movement time in path-steering," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2024, Art. no. 622, doi: 10.1145/3613904.3642316.

[67] S. Yamanaka and H. Usuba, "Tuning endpoint-variability parameters by observed error rates to obtain better prediction accuracy of pointing misses," in *Proc. 2023 CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2023, Art. no. 579, doi: 10.1145/3544548.3580746.

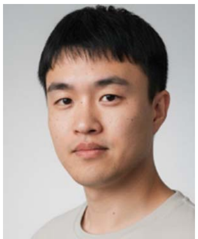
[68] D. Yu, T. Dingler, E. Velloso, and J. Goncalves, "Object selection and manipulation in VR headsets: Research challenges, solutions, and success measurements," *ACM Comput. Surv.*, vol. 57, no. 4, Dec. 2024, Art. no. 98, doi: 10.1145/3706417.

[69] D. Yu et al., "Target selection in head-mounted display virtual reality environments," *J. Univers. Comput. Sci.*, vol. 24, no. 9, pp. 1217–1243, 2018.

[70] D. Yu, H.-N. Liang, X. Lu, K. Fan, and B. Ens, "Modeling endpoint distribution of pointing selection tasks in virtual reality environments," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–13, 2019, doi: 10.1145/3355089.3356544.

[71] S. Zhai, J. Kong, and X. Ren, "Speed-accuracy tradeoff in fitts' law tasks: On the equivalency of actual and nominal pointing precision," *Int. J. Hum.-Comput. Stud.*, vol. 61, no. 6, pp. 823–856, Dec. 2004, doi: 10.1016/j.ijhcs.2004.09.007.

[72] X. Zhou, H. Qin, W. Xiao, L. Jia, and C. Xue, "A comparative usability study of bare hand three-dimensional object selection techniques in virtual environment," *Symmetry*, vol. 12, no. 10, 2020, Art. no. 1723, doi: 10.3390/sym12101723.



Yushi Wei is currently working toward the PhD degree with the Hong Kong University of Science and Technology (Guangzhou), China. His research interests include human-computer interaction, virtual and augmented reality.

Rongkai Shi received the BS and PhD degrees from the University of Liverpool, Liverpool, U.K., and Xi'an Jiaotong-Liverpool University, China. He is currently a postdoctoral researcher with the Hong Kong University of Science and Technology (Guangzhou), China. His research interests include human–computer interaction, extended reality, and interaction design.

Kemu Xu is currently working toward the master's degree with the School of Informatics, University of Edinburgh, U.K. Her research focuses on human-computer interaction, virtual reality, and AI-driven interaction design.

Jialin Wang is a postdoctoral researcher with the Hong Kong University of Science and Technology (Guangzhou), China. His research interests focus on virtual reality, human-computer interaction, visual perception, robotics, and computer graphics.

BoYu Gao is currently an associate professor with the College of Cyber Security, Jinan University, Guangzhou, China, where he leads a VR group for fundamental and applied research, including designing and assessing believable interactions with VR and HCI technologies. He has published more than 50 papers in leading journals and conferences, including *IEEE Transactions on Visualization and Computer Graphics*, IEEE VR, and ACM CHI.

Pan Hui (Fellow, IEEE) received the PhD degree in computer science from the University of Cambridge. He is chair professor of Computational Media and Arts and director of the Center for Metaverse and Computational Creativity with HKUST (Guangzhou). A globally recognized expert in ubiquitous computing and AR/VR, he is a fellow of the Royal Academy of Engineering, and a member of Academia Europaea. He has held prominent roles in both academia and industry across Europe and Asia, including as a distinguished scientist with Deutsche Telekom Laboratories and the Nokia chair in data science with the University of Helsinki.

Lingyun Yu received the PhD degree from the University of Groningen, The Netherlands, in 2013. She is an associate professor with the Department of Computing, Xi'an Jiaotong-Liverpool University, China. After that, she held positions as a lecturer and associate researcher with Hangzhou Dianzi University, China, from 2014 to 2017. From 2016 to 2019, she worked as a post-doctoral fellow with the University of Groningen and the University Medical Center Groningen. Her research focuses on interactive visualization, human-computer interaction, and virtual/augmented reality. She is especially interested in employing immersive visualization and interaction techniques for 3D spatial data.

Hai-Ning Liang (Member, IEEE) received the PhD degree in computer science from the University of Western Ontario, Canada. He is an associate professor in the Computational Media and Arts Thrust with the Hong Kong University of Science and Technology (Guangzhou), China. His research focuses on the design and evaluation of novel interactions and applications for virtual and mixed reality, gaming, and visualization technologies.