Int. J. Human–Computer Studies 163 (2022) 102817

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Understanding user performance of acquiring targets with motion-in-depth in virtual reality

Jin Huang [a,b], John J. Dudley [c], Stephen Uzor [c], Dong Wu [b], Per Ola Kristensson [c], Feng Tian [*,a,b]

[a] Institute of Software Chinese Academy of Sciences, No.4 South Fourth Street, Zhong Guan Cun, Beijing 100190, China
[b] University of Chinese Academy of Sciences, No.19A Yuquan Road, Shijingshan District, Beijing 100049, China
[c] University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom

ARTICLE INFO

Keywords:
Motion-in-depth
User performance
Moving target acquisition
Virtual reality

ABSTRACT

Virtual reality (VR) user interfaces contain numerous dynamic interactive tasks, among which acquiring moving targets is a common basic one. Previous studies have investigated user performance in moving target acquisition in desktop and touchscreen settings. However, these findings are not directly transferable to VR where targets and user input have complete freedom in three dimensions. This paper concentrates on motion-in-depth, that is, where a target predominantly exhibits approaching or receding movement as opposed to lateral motion across the user's field of view. We report on two studies investigating how various factors including texture, shadow, alignment, moving speed and moving direction affect: 1) perception accuracy of 3D targets with motion-in-depth, and 2) user performance, which we define as the combination of movement time (*MT*) and error rate (*ER*), in a target acquisition task involving motion-in-depth. Our data reveal a number of empirical results that are distinct from the depth perception of static targets and the user performance of 1D/2D target acquisition. We found that *MT* and *ER* when acquiring targets with motion-in-depth have strong regularities as the data showed good fits with Jagacinski's model for movement time estimation and a Ternary-Gaussian model for error rate prediction. We conclude with implications derived from this study for future designs.

1. Introduction

Interactive dynamic content is ubiquitous in modern computing applications such as games, real-time simulations and data visualizations. Dynamic content is particularly prevalent in virtual reality (VR) given the opportunity this environment affords for natural, embodied interaction with objects in three dimensions. One of the most common and fundamental interaction tasks encountered in such scenarios is the acquisition of moving targets, such as hitting a virtual tennis ball in a game. Despite its prevalence, acquiring moving targets is still very challenging for most users due to the level of sensory-motor coordination required (Brenner and Smeets, 1996; Franklin and Wolpert, 2011). In VR, the higher degree of freedom of motion makes it more difficult for users to locate and select moving targets. A better understanding of the factors influencing moving target acquisition in VR, and how these factors affect user performance, can help drive improvements in interaction design.

The essential difference between interactions in VR and other lower-dimensional (1D/2D) settings lies in the extra degree of freedom in depth. This addition of depth not only broadens the design space for user interfaces, but also invokes a series of specific research topics, such as the effect of depth on user performance (Janzen et al., 2016; Teather and Stuerzlinger, 2013) and perceiving (Hubona et al., 1999) or reaching (Batmaz et al., 2019) objects in depth. Previous studies (e.g., (Batmaz et al., 2019; Hubona et al., 1999; Janzen et al., 2016)) indicate that the perception and behavioral patterns of users in the depth dimension are different from those in other dimensions; this motivates our investigation of user performance in acquiring moving targets specifically in the depth dimension, or what we term 'motion-in-depth' in this paper.

Compared to the extensive studies on static target pointing, there is far less existing work on assisting moving target pointing in VR. Recent works in moving target acquisition offer good explanations and models for user pointing behaviors in moving targets (Hasan et al., 2011; Huang et al., 2018; Lee et al., 2018). However, these studies have been conducted only in 1D or 2D spaces and with traditional input devices (e.g., mouse and touch screen); therefore, the results from these studies cannot be directly transferred and generalized to target motion in the depth dimension.

* Corresponding author.
E-mail address: tianfeng@iscas.ac.cn (F. Tian).

https://doi.org/10.1016/j.ijhcs.2022.102817
Received 7 April 2021; Received in revised form 11 March 2022; Accepted 15 March 2022
Available online 24 March 2022
1071-5819/

To gain an understanding of the factors affecting user performance in VR-based target acquisition with motion-in-depth, we conducted two user studies. The first study investigates the influence of speed, moving direction, texture, shadow and alignment on *perception accuracy* of objects with motion-in-depth. The second study explores how the aforementioned factors affect *user performance*, which we define as the combination of movement time (*MT*) and error rate (*ER*), in a target acquisition task involving motion-in-depth. We found that target speed has the greatest impact on users perception, followed by shadow and direction movement (i.e. approaching or receding). We discovered empirical evidence that differs from previous research on target selection in conventional user interfaces, such as the influence of target speed on *MT* depending on the target's moving direction and initial distance having a significant impact on selection *ER*. Data of *MT* and the *ER* of motion-in-depth showed good fits with Jagacinski's model and a Ternary-Gaussian model, implying strong lawful regularities of *MT* and *ER* in this task. Based on the empirical evidence from these two studies, we derive several important implications for the design of applications involving targets with motion-in-depth.

In summary, this paper makes the following three contributions:

a) An empirical investigation of the effects of five design factors: 1) texture; 2) shadow; 3) alignment; 4) moving speed; and 5) moving direction on perception accuracy of motion-in-depth.
b) An empirical investigation and model analysis of the effects of the above mentioned five design factors and two task variables: 1) initial distance; and 2) target size on *MT* and *ER* for targets with motion-in-depth.
c) A set of design guidelines for applications in virtual environments to improve the design of dynamic content with motion-in-depth.

2. Related work

We categorize the related work on user performance in target acquisition with motion-in-depth into three areas: target acquisition in VR, perception and user performance in the depth dimension, and descriptive models for moving target acquisition.

2.1. Target acquisition in VR

Target Acquisition is a basic operation in HCI (Argelaguet and Andujar, 2013). In 2D space, target acquisition has been extensively researched, resulting in well-established selection paradigms, such as mouse (Dragicevic, 2004), pen (Accot and Zhai, 2002) and touch input (Luo and Vogel, 2014). There has also been much work on selection techniques in 3D space, which includes volumetric displays (Grossman and Balakrishnan, 2006), as well as VR (Huang et al., 2019b; Tu et al., 2019) and augmented reality (AR) (Wolf et al., 2018). Compared to 2D, interaction in 3D environments is more challenging due to the additional third dimension (Hinckley et al., 1994). Consequently, several attempts have been made to develop techniques to enable more efficient interaction in 3D spaces. For instance, Feiner (2003) developed a technique, called the Flexible Pointer, which allows users to select objects that are partially or fully obscured by other objects in 3D space. Poupyrev et al. (1996) developed the go-go cursor, an extendable projected cursor designed to facilitate reaching of both near and far objects based on the user's arm extension. In the commercial VR space, target acquisition is primarily achieved using two main and extensively-studied methods – raycasting and virtual hand.

Raycasting is a method where a ray is projected from an input device (either a controller or a hand) towards an interface. Targets are acquired at the point the projected ray intersects with an interface element, e.g. a button. To leverage simpler interactions in 3D space, researchers have exploited the use of 2D image planes and the raycasting technique (Lubos et al., 2014; Tu et al., 2019). For instance, Ramcharitar and Teather (2018) developed the EZCursorVR technique, which shows a 2D cursor on a plane in front of the user in VR, to make depth-based selections beyond the image plane possible. Tu et al. (2019) compared pointing to a crossing technique, both of which used raycasting as a foundation for target acquisition in VR. The raycasting method is widely used in commercial VR applications to support the selection of distant targets; the interaction is typically implemented as a laser-like pointer projected from the VR controller or hand into the virtual environment. The virtual hand technique places a cursor at the position of the user's hand in 3D space. In its simplest form, it is represented by a point cursor placed at the user's hand or a hand-held wand—for instance, (Barrera Machuca and Stuerzlinger, 2019; Batmaz et al., 2019). The benefit of the virtual hand method, versus the raycasting technique is that more natural interactions can be enabled in 3D space, with the user able to move the hand and acquire targets in three dimensions. Barrera Machuca and Stuerzlinger (2019) highlighted a variation of Fitts' Law to predict movement depth using a virtual hand-based cursor. Batmaz et al. (2019) also implemented a similar virtual hand technique to explore the effect of target depth on target acquisition in VR and AR.

While raycasting is mostly used in static VR UIs (e.g. home menus for Oculus and HTC Vive), virtual hand interactions are increasingly used in both static and dynamic interaction scenarios to promote natural gestures, especially with the capabilities for hand tracking in modern VR headsets (e.g. Oculus Quest). In moving target selection such as required in modern VR games, virtual hand selection using a VR controller is often the dominant method. This is especially true when movement is in the depth dimension—for instance, using the sword in Beat Saber to hit blocks moving towards the player, or using a bat in a table tennis VR game. Using a raycasting method for target acquisition in these types of interactions would be both cumbersome and unnatural. Therefore, since our work is concerned with moving target selection in the depth dimension, we primarily use the virtual hand method, instead of raycasting, for target acquisition.

2.2. Perception and user performance in the depth dimension

The human visual system is incredibly complex, with depth perception specifically affected by a wide range of factors. In this section, we review those factors most relevant to the investigation of moving target acquisition in VR.

VR headsets present different images to each eye to produce the illusion of depth in the virtual environment. Visual cues in VR differs from physical environments. The presentation of different images to both eyes causes a dissociation of accommodation and vergence in stereoscopic displays, whereas the curvature of the lenses accommodates to the distance of the display (Maruhn et al., 2019). Subjects' estimation of egocentric distances (i.e., the distance from oneself to an object) in VR is consistently lower than that in physical environments (Lin and Woldegiorgis, 2017; Renner et al., 2013). These works highlight the need of conducting empirical evaluation for depth perception in VR.

Wann et al. (1995) succinctly describe several factors that make depth perception in VR fundamentally distinct from normal viewing. The key difference is the fact that in VR, each eye is actually viewing a two-dimensional image source. As a results, users are not presented with the full complement of visual cues that are present in the real world.

Renner et al. (2013) provides a highly informative survey of the factors influencing egocentric distance perception in virtual environments. Among the many factors that can influence depth perception, perhaps the most relevant in a VR context are: movement, shadow and texture.

The influence of the motion of objects (Sauer et al., 2001; Wanger et al., 1992) and the motion of the viewer (Smets, 1992) on depth perception has been widely investigated. Sauer et al. (2001) examined the impact that object self-rotation can have on the perception of object size at different depths, with a rotating cylinder being judged to have greater object self-depth than the equivalent object without rotation. Smets (1992) examined the role of observer movement in depth perception for remote teleoperation. In the apparatus developed by Smets (1992), head movements of the viewer produced corresponding camera movements and this 'depth television' (video stream on a 2D display) enabled superior telemanipulation performance.

Object texture can provide useful cues about object shape and orientation that facilitates size and depth perception (Todd and Akerstrom, 1987; Young et al., 1993). The perceived depth of an object in the foreground can also be influenced by the texture of the background (O'Brien and Johnston, 2000).

Shadow has been found to play a key role in depth perception (Puerta, 1989). Yonas (1979) makes the distinction between two types of object shadow: attached and cast. An attached shadow is shadow present on the object itself and provides information about its shape and form. A cast shadow is a shadow appearing in the environment due to the presence of the object and a corresponding light source. As Hubona et al. (1999) note, "cast shadows influence perceived object size, elevation and depth." Shadows can also provide useful cues about object shape (Wanger, 1992). Mamassian et al. (1998) demonstrated light source motion (more readily altered in a virtual environment) can distinctly impact depth perception.

The complexity of the visual system highlights the need to perform empirical evaluations to understand how these factors specifically affect a given task. Even more confounding, is the fact that depth perception has also been shown to be influenced by induced psychological factors. For example, increased walking effort due to wearing a heavy backpack can lead to overestimation of distance (Proffitt et al., 2003). Therefore, in this paper, we take a structured approach to investigating the most relevant factors hypothesised to dictate perception and performance in VR moving target acquisition.

2.3. Descriptive models for moving target acquisition

Modeling user performance in HCI is a common method used to represent the underlying mechanisms behind an interaction phenomenon. *MT* and *ER* are the two most important user performance metrics in moving target acquisition. As a result, there are mainly two types of models in the literature to interpret *MT* and *ER* respectively.

One of the most well-known *MT* models is Jagacinski's model (Jagacinski et al., 1980). Dating back to Jagacinski et al.'s original work, they extended Fitts' law (Fitts, 1954) to moving target acquisition by interpreting the effect of target speed using a modified index of difficulty (*ID*). Similar to the Fitts' *ID*, the modified index of difficulty reflects the lengthening of *MT* arising from the interaction of target width and speed by including a term for the speed-to-width ratio. Another well-known approach in *MT* estimation comes from Hoffmann (1991), who proposed two alternative models for *MT* prediction by applying the linear control model theory to infer the index of difficulty in moving target acquisition. A steady-state position error which reduces the effective target width was used to simulate the human response of acquiring moving targets, and finally to lead to an *MT* estimation. By considering the effect of target motion on human temporal accuracy, Tresilian (2005); Tresilian and Lonergan (2002) found different effects of target speed in different participants' target acquisition strategies such as pursuit and hitting. Tresilian et al. suggested estimating *MT* with different functions of target speed instead of treating the effects of target speed by a fixed function. This paper adopts Jagacinski's model to investigate *MT* performance in motion-in-depth due to its wide acceptance and high interpretability.

ER models for moving target acquisition predict how likely the user is to acquire the target under specific task conditions. There are two typical types of errors in acquiring moving targets. The first type of error occurs in the spatial domain. A spatial error is counted when the spatial location of the user's selection endpoint falls outside the physical boundary of the target (Huang et al., 2018; Wobbrock et al., 2008). Unlike the errors in the spatial domain, the second type of error happens in the temporal domain. An error in the temporal domain occurs if the user initiated response (Lee and Oulasvirta, 2016) (also known as time-to-contact (Rolin et al., 2018)) is outside of a limited time window. To model *ER* in the spatial domain, one of the most common methods is to predict the distribution of spatial endpoints of users. A recent approach for modeling the spatial errors in moving target acquisition is the Ternary-Gaussian model (Huang et al., 2020; 2018; 2019a). The Ternary-Gaussian model assumes the endpoint distribution is a Gaussian distribution consisting of three Gaussian components, generated from uncertainties of the input device, target size, and target speed. Given the distribution of spatial endpoint, errors can be obtained through integrating the probability distribution function within the selection region of the targets.

Similarly, to model *ER* in the temporal domain, we can estimate the distribution of temporal endpoints. Lee and Oulasvirta (2016) derive a model to predict *ER* in a temporal pointing task by assuming that users have an implicit temporal point of aim within the target time window, that is, the point in time at which they intend the input event to be registered. Based on this approach, various models have been proposed to predict *ER* in moving target acquisition with multiple cues (Lee et al., 2018), motion delay (Lee et al., 2019) and click planning (Park and Lee, 2020). In this paper, we focus particularly on the spatial pointing task, thus the Ternary-Gaussian model is used to analyze the spatial errors in the task.

3. Motivation and research questions

As mentioned earlier, acquiring targets with motion-in-depth is an important and relatively understudied problem. The way in which modern VR displays create the illusion of 3D means that the scale of the virtual space far from our eyes is visually much smaller than the nearby space. The position of an object moving in the depth dimension may only be perceived by variation in size under certain viewing conditions. This difficulty in perceiving the object's position is very likely to result in varied user performance when attempting to reach the object. Furthermore, various aspects of the moving target acquisition task such as initial distance, target size, speed, and moving direction may have completely different effects on user performance when compared to 1D/2D scenarios in which the visual size and speed of the target do not change over time. This largely unexplored landscape is the fundamental motivation for conducting this study.

In order to better explain user performance in acquiring targets with motion-in-depth, we divided the research into two studies. The first study focused on perception accuracy of 3D targets with motion-in-depth, while the second study investigated the user performance (i.e., the results of human actions) in acquiring such targets. Moving target selection is commonly thought of as a ballistic motion that consists of two processes: planning and execution (Brenner and Smeets, 1996; Smeets and Brenner, 1995). The planning process entails primarily perceiving information about the target location and speed, which is used as the primary guiding basis for the human body to take action in the following stage. As a result, perception accuracy for target motion could greatly affect the final user performance of the task. Observations of perception accuracy from the first study can not only strengthen our final conclusion on user performance, but also provide a more thorough explanation of potential findings.

In order to make this study more practical for HCI designers, we focus on the impact of associated VR design factors on perception accuracy and user performance rather than the underlying mechanism of the human sensory-motor system. Based on a synthesis of previous work on VR target selection (Batmaz et al., 2019), depth perception (Wanger et al., 1992), and moving target acquisition (Huang et al., 2018), we choose to study the following five design factors that we consider to be most worthy of investigation in VR: texture, shadow, alignment, moving speed and moving direction. Texture and shadow (including lighting effect) are two of the most widely mentioned influencing factors in the existing literature. Texture reveals spatial relations by amplifying the information that one surface occludes another (Gibson, 1950), whereas shadow, which is caused by one object obstructing the light falling on another, is a powerful source of information for spatial position (Yonas, 1979). Alignment, moving speed, and moving direction are three factors derived from the target's spatial movement; they yield relative displacement between the target and the environment, as well as the relative displacement between the target and the user, which is also considered a source of information for spatial relations.

The literature also identifies many other factors that could influence depth perception and user performance in target acquisition. However, in our studies we choose to omit factors that have already been established as industry standards and implanted into the hardware and software systems of VR devices. Such factors, as summarized in Wanger's study (Wanger et al., 1992), include convergence and accommodation, binocular disparity, perspective, etc. What we are more interested in is the factors that designers can control more directly, and how to use these cues to deliver interactions and interfaces that meet specified design purposes.

With the aforementioned motivation, this paper aims to answer the following three research questions (RQs):

RQ1) Do the design factors of Texture, Shadow, Alignment, Moving Speed and Moving Direction affect perception accuracy of targets with motion-in-depth, and which factors have the greatest impact?
RQ2) How do these factors affect user performance in terms of *MT* and *ER* in acquisition of targets with motion-in-depth?
RQ3) Does the acquisition of targets with motion-in-depth possess lawful regularities?

In the following sections, we first introduce the definition of motion-in-depth and the corresponding testing environment in VR, and then report on two studies to answer these questions. Specifically, Study 1 was conducted to investigate RQ1, and Study 2 was designed for answering RQ2 and RQ3.

4. Definition and the testing environment of motion-in-depth

In this paper motion-in-depth refers to the movement of an object approaching or receding from a user. Correspondingly, the perception of motion-in-depth means the awareness of the approaching or receding motion of an object and the dynamically changing location of the object. Acquiring an object with motion-in-depth is then defined as using any available pointing technique to reach an object with approaching or receding motion. In VR, we construct this motion pattern with interactive content (e.g., a ball) that only moves in the depth dimension which is perpendicular to the user's viewing plane (see Fig. 1(a)). Targets and cursors used in this study are fixed in the horizontal (x) and vertical (y) dimensions. As in any other target acquisition task, *MT* and *ER* are the two most appropriate user performance metrics in the task of acquiring objects with motion-in-depth.

The five factors the ways we manipulated them in the VR environment are as follow.

- *Texture*: There are many kinds of texture in VR. We adopted a general representative texture, a checkerboard pattern, previously introduced by Wanger et al. (1992). Two conditions were introduced for texture: texture-on and texture-off. For the texture-on condition, objects were textured with a checkerboard pattern, which provides information of surface orientation and distance: both cues relevant to perceiving depth motion. For the texture-off condition, materials on the objects were removed, resulting in them being rendered in a flat gray color (see Fig. 1(b) left).
- *Shadow*: As mentioned in Yonas (1979), there are two main types of shadow in computer graphics systems: attached shadow and cast shadow. Shadows are generated by light sources. Different positions and numbers of light sources produce different shadow effects (Hubona et al., 1999). For simplicity, we used a single light source generating both attached and cast shadow when shadow is enabled (see Fig. 1(c) right), and we remove all shadows otherwise (see Fig. 1 (c) left).
- *Alignment*: In Wanger's study (Wanger et al., 1992), an important factor "viewpoint motion" was investigated, which was manipulated by allowing the participants to move their viewpoint along a horizontal axis to watch the target from the left or right side. We included a similar factor called *Alignment* in this study by allowing participants' different views of the target motion. The effect of Alignment was examined at two possible levels: collinear—the target motion coincides with the participants' viewing direction (see Fig. 1(d) left); parallel—the target moves along a path parallel to the participants' viewing direction, and the path is offset by a certain distance to the right or left from the participants' viewing direction (see Fig. 1(d) right).

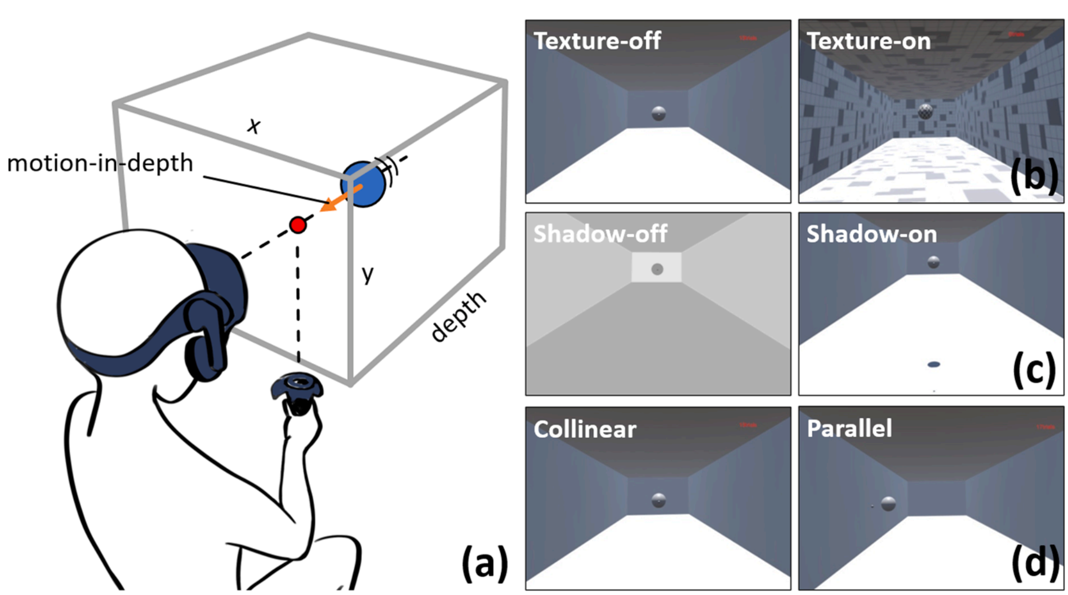

Fig. 1. VR configuration for studying moving target acquisition in depth. a) Schematic diagram of acquisition of targets with motion-in-depth; b-d) Conditions of the design factors investigated in this study.

- *Moving Speed*: Objects move at a specified speed. Previous evidence suggests that moving speed has a significant impact on selection accuracy in moving target acquisition (Huang et al., 2018). We therefore included target speed as a factor in our study and examined it at three levels: slow, medium and fast.
- *Moving Direction*: Objects move in a specified direction. Previous studies showed that participants have different performance in acquiring objects depending on moving direction (Tresilian, 2005). In this paper, moving direction was included as study conditions with two possible levels: approaching and receding.

5. Study 1: perception accuracy of motion-in-depth

In Study 1, we aim to answer whether the design factors of texture, shadow, alignment, moving speed and moving direction affect perception accuracy of motion-in-depth, and which of these have the greatest impact. We adopted the positioning task presented in Wanger's study (Wanger et al., 1992) and modified it to a depth-marking task: Participants controlled a cursor ball to mark the depth of two target balls connected by an imaginary line segment (see Fig. 2(a)). The two main reasons for adopting this two-target-balls task are: 1) the cursor ball used to mark depth is located at the center of the participants' field-of-view, which makes the marking more accurate; and 2) the two-target-balls setting prevents the cursor ball from being obscured by the target. We describe the task in detail below.

5.1. Depth-marking task

Three balls (i.e., two target balls and one cursor ball), which can only move in the depth dimension, were displayed in a virtual room. The two balls form the endpoints of an imaginary line segment whose orientation varied randomly from trial to trial. The cursor ball, aligned vertically and horizontally with the midpoint of the imaginary line segment of the two target balls, was controlled using an HTC Vive controller. As the cursor ball is fixed in the x and y axes, only the z-component of the controller motion was used to adjust the cursor position.

When the task begins, the target balls move toward (approaching) or away (receding) from the participant with a fixed speed. The participant must adjust the depth of the cursor ball to lie at the midpoint of the imaginary line segment joining the two target balls as quickly and as accurately as possible.

5.2. Participants and apparatus

Twelve participants (4 female) were recruited. The age range of the participants was from 25 to 50 years, with a mean age of 38.6 years and a standard deviation of 9.5 years. Arm length of the participants was ranged from 57.5 to 71.9 cm, with a mean of 61.8 cm and a standard deviation of 3.6 cm. All participants were right-handed and two had prior experience with VR applications before the experiment. Participants were screened for stereo vision by oral questioning and each of them was compensated with $20 for their time. The experiment was conducted on a computer, with an Intel Core i7 4 Quad core CPU at 2.6 GHz, a NVIDIA GeForce 1080 GPU, and 8 GB of RAM, running Microsoft Windows 10. We used an HTC Vive Pro-head-mounted display connected to the computer via HDMI. The device features a resolution of 1440×1600 per eye, a 90 Hz refresh rate and a 110° field of view. The experimental environment was developed with Unity 3D in C#.

5.3. Factors

We investigated the five factors outlined in Section 4 at several levels. These are: Speed (slow, medium, fast), Direction (approaching, receding), Alignment (collinear, parallel), Texture (on, off) and Shadow (on, off).

The three moving speeds of the target balls were set as 2 cm/s, 4 cm/s and 8 cm/s respectively. In the collinear condition, the midpoint of the target balls was placed directly in front of the participant's view, and approached or receded from the participant on a path collinear with their viewing direction. In the parallel condition, the midpoint of the target balls was offset by 50 cm to the right or left from the participants' viewing direction and moved along a path parallel to the participants' nominal viewing direction. In the texture-on condition, the objects in the scene including the target balls, the cursor ball, the walls, the ground and the ceiling of the room were mapped with texture in a checkerboard pattern, while in the texture-off condition, all of the objects were displayed in a different shade of grey – the target balls and the cursor ball were darker while other objects were lighter. In the shadow-on condition, a downward directional light was placed at the top of the room, generating both attached and cast shadow for every object in the scene. In the shadow-off condition, the directional light was removed and there were no shadows present.

5.4. Procedure

Participants were seated in the middle of a tracking area in a non-swivel chair. They were asked to maintain their head facing straight ahead in the nominal direction (the same direction of target motion) during the study (see Fig. 2(b)). After putting the headset on, participants were introduced to the task and encouraged to practice until they mastered the task.

To start the test, a participants push a button on the controller and hold the cursor ball in a start area (a region 30 cm in front of participants and at the same height as their eyes). Following a 500–1000 ms delay, the target balls appear in the center of the start area and immediately begin moving toward (approaching) or away (receding) from the

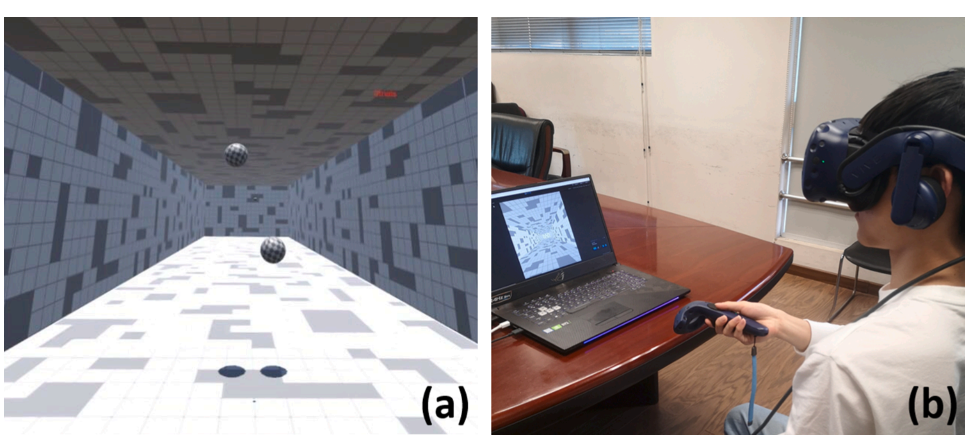

Fig. 2. Task and apparatus in Study 1. a) The two-target-balls task; b) A participant took part in Study 1.

participants with a fixed speed. The participants are asked to move the cursor ball (0.5 cm in size) to track the midpoint of the two target balls (5 cm in size and separated by 20 cm) as quickly and as accurately as possible. The test ends after the participants track the midpoint of the target balls for 3 s. The position and timestamp of the cursor ball and the midpoint of the target balls during the process are recorded using a 100 Hz sampling rate.

Each participant made 6 repeats for each condition in the test, yielding 3 Speed × 2 Direction × 2 Alignment × 2 Texture × 2 Shadow × 6 repeats = 288 trials in total. The order of the 288 trials was randomly arranged for each participant to ensure that a participant could never predict the upcoming condition of the next trial. This avoids the learning effect induced by repeating 6 trials of the same condition. Participants were permitted to take a break of up to 60 s if they felt uncomfortable at any time during the experiment.

5.5. Measures

We used the root mean squared error (RMSE) between the depth of the target balls and the cursor during the tracking process to measure participants' perception accuracy:

$$RMSE = \sqrt{\frac{1}{m}\sum\nolimits_{i=1}^{m}\left(TargetDepth_i - CursorDepth_i\right)^2} \quad (1)$$

where, m is the number of samples collected in one trial. The RMSE provides a non-negative, scale-invariant measure of the errors.

5.6. Results

For our statistical analysis, we first binned our data by different participants in different conditions. A mean value of RMSE of one participant over the 6 repeats in one condition was treated as one sample in the following analysis.

Nonparametric Friedman tests (Wilcoxon, 1992) were used for statistical analysis as one fourth of the conditions (12 in 48) of the perception accuracy data (i.e., RMSE) were not normally distributed. Effect sizes were calculated using Kendall's coefficient of concordance (Kendall's W) (Kendall and Smith, 1939).

Significant main effects of Speed, Direction, Alignment and Shadow were found on RMSE (all $p < 0.001$). Participants demonstrated higher perception accuracy for targets at slower moving speeds, targets that are approaching the participants, targets with motion misaligned to the participants' viewing direction, and targets with shadow. No significant main effect of Texture was found on RMSE ($p = 0.637$). Table 1 shows all statistical results for the factors of Speed, Direction, Alignment, Texture and Shadow. Boxplots in Fig. 3 present RMSE for each condition of the five factors.

To establish a quantitative measure of the contributions of the factors to perception accuracy, we conducted linear regression on our data. In the following analysis, the factor of speed was measured by its physical unit (i.e., centimeter), other factors were measured as follows: for Direction, receding = 0 and approaching = 1; for Alignment, collinear = 0 and parallel = 1; for Texture, texture-off = 0 and texture-on = 1; and for Shadow, shadow-off = 0 and shadow-on = 1.

The linear regression was conducted with the RMSE as a dependent variable and all of the five factors as independent variables. As shown in Table 2, only the relationships between the RMSE and Speed, Shadow and Direction were significant. The unstandardized coefficients of the linear regression revealed that, a 1 cm/s increase in Speed was predicted to result in a 0.259 cm increase in RMSE; enabling shadow (shadow was measured as shadow-off = 0 and shadow-on = 1) was predicted to result in a 0.561 cm decrease in RMSE; changing the moving direction from receding to approaching (direction was measured as receding = 0 and approaching = 1) was predicted to result in a 0.509 cm decrease in RMSE. On the other hand, speed had the biggest absolute standardized coefficient (0.304), followed by Shadow (0.132) and Direction (0.119). These results indicate that within the test scope of each variable described in this paper, the factor of speed has the greatest impact on participants' perception accuracy, followed by Shadow and Direction.

Table 1
Statistical results for the factors of Speed, Direction, Alignment, Texture and Shadow on RMSE. Asterisks represent pairwise significant differences, which are noted as: $p < 0.001$(**) and $p < 0.05$(*). The condition with the lowest RMSE is highlighted in bold for each factor that has a significant effect.

Factors	Conditions	RMSE (SD) cm	χ^2	p	Kendall's W
Speed	**Low (2 cm/s)**	**1.48 (1.46)**	215.930	<0.001**	0.562
	Medium (4 cm/s)	1.95 (1.92)			
	High (8 cm/s)	3.03 (2.56)			
Direction	**Approaching**	**1.90 (1.65)**	18.000	<0.001**	0.063
	Receding	2.41 (2.50)			
Alignment	Collinear	2.23 (2.11)	12.129	<0.001**	0.042
	Parallel	**2.07 (2.14)**			
Shadow	**Shadow-on**	**1.87 (1.71)**	29.389	<0.001**	0.102
	Shadow-off	2.43 (2.45)			
Texture	Texture-on	2.14 (2.00)	0.222	0.637	0.001
	Texture-off	2.16 (2.25)			

5.7. Discussion

Our results suggest that participants have a higher perception accuracy on targets at slower moving speeds in the depth dimension. This is consistent with prior work conducted in 1D and 2D settings (Brenner and Smeets, 1996; Huang et al., 2018; 2019a), The uncertainty in motion estimation is thought to be caused by delay in the human sensory-motor system (Huang et al., 2018; Miall et al., 1993). Our data suggests that this sensory-motor delay also affects motion estimation in the depth dimension.

The study data reveals that the participants could perceive motion-in-depth more precisely when the shadow of the target balls was enabled. Similar results can be found in previous studies on human perception of static targets with depth (Hubona et al., 1999; Wanger et al., 1992). Shadow provides a ground-plane-relative reference for the target distance which can enhance one's perception of motion.

We observed higher perception accuracy when the target balls are moving toward the participants compared to when they are receding. We suggest this observation can be explained by the following characteristics of the acquisition task. When the target balls are moving away from the participants, the distance to the target balls increases with the movement. The movement of objects further from the participants' eyes appears smaller, thereby likely making it harder for the participants to perceive the target motion (i.e., lower perception accuracy). In contrast, when the target balls are moving toward the participants, the distance between the target balls and the participants reduces and the visual motion of objects increases. This likely makes it easier for participants to perceive the target motion (i.e., higher perception accuracy).

We found that perception accuracy was significantly improved when the target motion was misaligned with the participants' viewing direction. When the target motion is misaligned from the viewing direction, the target motion in the depth dimension is not completely perpendicular to the participants' retina. Motion of misaligned targets can be perceived not only through the visual change in size of the object but

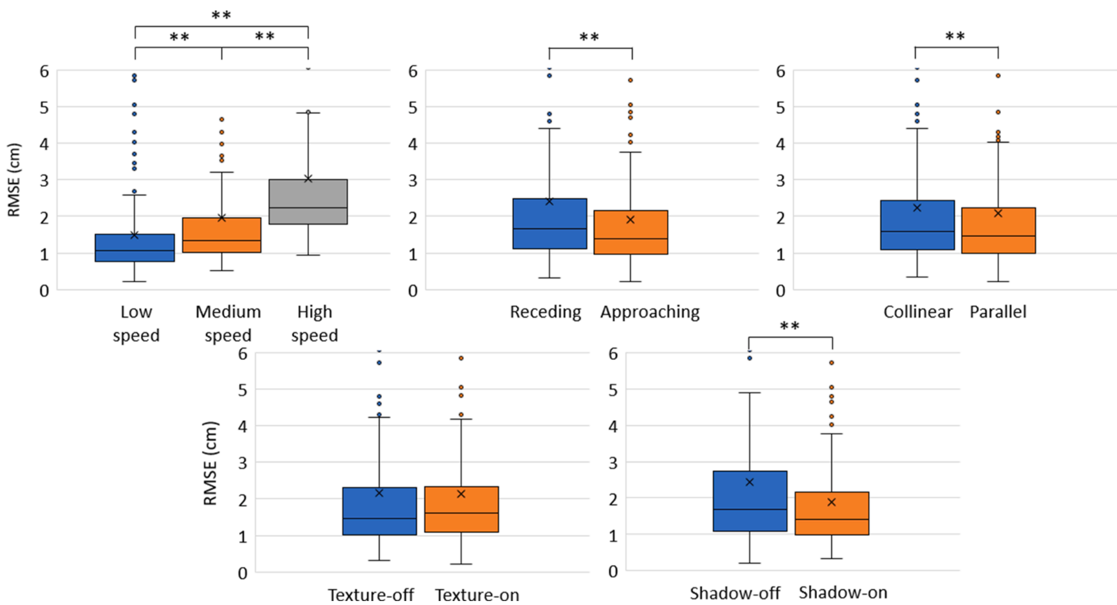

Fig. 3. Boxplot diagrams represent analysis of RMSE for Speed, Direction, Alignment, Texture and Shadow. "x" in the diagrams are mean values for each condition. Asterisks represent pairwise significant differences, which are noted as: $p < 0.001$(**) and $p < 0.05$(*).

Table 2
Coefficients of linear regression for Speed, Direction, Alignment, Texture and Shadow on RMSE. Asterisks represent pairwise significant differences, which are noted as: $p < 0.001$(**) and $p < 0.05$(*).

Factors	Unstandardized coef.	Standardized coef.	*p*
Speed	0.259	0.304	0.001**
Direction	− 0.509	− 0.119	0.002*
Alignment	− 0.156	0.037	0.351
Texture	− 0.023	− 0.006	0.888
Shadow	− 0.561	− 0.132	0.001**

also through the object's relative motion in the *x*-direction (see Fig. 1 (a)). More information about the target location is therefore available to participants this is likely the responsible for improving the participants' perception accuracy.

The linear regression applied on the study data indicates that moving speed had the greatest impact on participants' perception accuracy, followed by shadow and direction. Because the human body has a fixed sensory-motor processing delay of 150 ms (Miall et al., 1993), our brain always receives target depth information after a certain amount of time has passed, and by the time we receive it, the target has already moved a distance proportional to its moving speed.

6. Study 2: user performance of acquiring targets with motion-in-depth

In Study 2, we first investigate how the design factors in VR affect *MT* and *ER* in acquisition of targets with motion-in-depth; and then, we test if the task of acquiring targets with motion-in-depth can be modeled by existing models for moving target acquisition. We fit the *MT* and *ER* models under different conditions of design factors and observe how the factors affect the overall trend of *MT* and *ER*. Study 2 was carried out several weeks after Study 1 was completed. The participants and apparatus were the same as in Study 1. We used the same participant pool in both studies for two reasons: 1) using the same participant pool ensures that the data of Study 1 and Study 2 are comparable by minimizing the impact of individual differences between participants; 2) recruiting the same participants simplifies the organization of the experiments during the COVID-19 pandemic.

6.1. Target acquisition task of motion-in-depth

In contrast to the goals of Study 1 which assessed participants' perception of motion-in-depth, the purpose of Study 2 was to investigate participants' performance when acquiring targets with motion-in-depth. Therefore, we adopted a conventional Fitts-like task paradigm in Study 2: Participants were asked to control a cursor ball to acquire a target ball with motion-in-depth. In the task, the participants dynamically adjust the depth of the cursor ball to reach the target ball and confirm the acquisition with a button press. As in Study 1, the *x* and *y* position of the cursor ball was fixed and aligned with the target ball.

6.2. Factors

We investigated the following factors in Study 2: Initial Distance (near, far); Width (small, medium, large); Speed (slow, medium, fast); Direction (approaching, receding); Alignment (collinear, parallel); Texture (on, off) and Shadow (on, off). In addition to the five factors in Study 1, initial distance and target size were added in Study 2 as they are key factors for a Fitts-like task. All of the factors except the initial distance and target size have the same settings as in Study 1. Initial distance refers to the distance from the starting position of the target ball to the cursor ball in the depth dimension. The two initial distances were set as 10 cm and 20 cm respectively. The three sizes of the target ball were set as 1 cm, 2 cm and 4 cm respectively.

6.3. Procedure

To begin the test, participants push a button on the controller and hold the cursor ball in a start area. After a 500–1000 ms delay, a target ball appears at a certain distance from the center of the start area. The start area is set between the initial position of the target and the participant. Once the target ball appears, it immediately starts moving toward (approaching) or away (receding) from the participant with a fixed speed. The participant must move the cursor ball to reach the target ball and confirm the acquisition with a button press as quickly and accurately as possible. The test ends when the participant confirms the acquisition regardless of whether they hit or miss the target.

Each participant made 6 repeats for each condition in the test, yielding 3 Size × 3 Speed × 2 Direction × 2 Alignment × 2 Texture × 2

Shadow × 6 repeats = 864 trials in total. The order of conditions was randomized. Participants were permitted to take a break of up to 60 s if they felt fatigue in the arm at any time during the experiment.

6.4. Measures

We log participant performance on two measures:

1. Movement time (*MT*): the time duration from the start of the task to the confirmation of the target acquisition.
2. Error rate (*ER*): an error occurs when the button is pressed and the cursor is outside the target. Accordingly, the error rate is defined as the percentage of misses among all acquisitions.

6.5. Results

To conduct the statistical analysis, we binned our data from each participant's performance in each condition generated from the seven factors. For *MT* analysis, the mean *MT* of the 6 trials of one participant in one condition was treated as one sample. For *ER* analysis, the ratio of error trials to the total 6 attempts of one participant in one condition was used as one sample.

All *MT* and the majority of *ER* data were normally distributed, thus we used a multi-way repeated measures ANOVA and post hoc comparisons with Bonferroni adjustment for statistical tests in Study 2. We used the partial eta squared statistic η_p^2 to measure effect size in ANOVA.

6.5.1. Movement time

Significant main effects of Speed, Direction, Width, Initial Distance and Shadow were found on *MT* (all $p < 0.001$). Participants performed faster acquisitions when targets: had higher moving speeds; were larger in size; were approaching; had a closer initial distance: and had shadow. No significant main effects of Alignment ($p = 0.750$) and Texture ($p = 0.063$) were found on *MT*. Table 3 shows all statistical results for the factors of Speed, Direction, Width, Initial Distance, Alignment, Shadow and Texture. Boxplots in Fig. 4 present *MT* for each condition of the seven factors.

Significant interaction effects were found for Speed * Direction ($p < 0.001$), Direction * Width ($p < 0.001$), Direction * Alignment ($p = 0.005$), Direction * Shadow ($p < 0.001$), Width * Shadow ($p = 0.039$), and Initial Distance * Shadow ($p < 0.001$). Other interaction effects were not significant.

6.5.2. Error rate

Significant main effects of Speed, Direction, Width, Initial Distance, Alignment and Shadow were found on *ER* (all $p < 0.001$). Participants demonstrated higher pointing accuracy (i.e., lower *ER*) when targets: had slower moving speeds; were larger in size; were approaching; had a closer initial distance; had motion that was misaligned with viewing direction; and had shadow. No significant main effects of Texture ($p = 0.469$) were found on *ER*. Table 4 shows all statistical results for the factors of Speed, Direction, Width, Initial Distance, Alignment, Shadow and Texture. Boxplots in Fig. 5 present *ER* for each condition of the seven factors.

Significant interaction effects were found for Speed * Width ($p < 0.001$), Speed * Initial Distance ($p = 0.019$), Speed * Shadow ($p = 0.002$), Speed * Direction ($p < 0.001$), Direction * Width ($p = 0.008$), Direction * Shadow ($p = 0.021$), Direction * Initial Distance ($p < 0.001$), Width * Initial Distance ($p = 0.002$), Width * Alignment ($p = 0.024$), Width * Shadow ($p = 0.018$), and Initial Distance * Alignment ($p < 0.001$). Other interaction effects were not significant.

Table 3
Statistical results for the factors of Speed, Direction, Width, Initial Distance, Alignment, Texture and Shadow on *MT*. Asterisks represent pairwise significant differences, which are noted as: $p < 0.001$(**) and $p < 0.05$(*). The condition with the lowest *MT* is highlighted in bold for each factor that has a significant effect.

Factors	Levels	*MT* (SD) s	*p*	*F*	η_p^2
Speed	Low (2 cm/s)	0.995 (0.243)	<0.001**	26.867	0.710
	Medium (4 cm/s)	0.984 (0.249)			
	High (8 cm/s)	**0.960 (0.301)**			
Direction	**Approaching**	**0.888 (0.222)**	<0.001**	311.254	0.966
	Receding	1.071 (0.275)			
Width	Small (1 cm)	1.098 (0.287)	<0.001**	213.998	0.951
	Medium (2 cm)	0.973 (0.247)			
	Large (4 cm)	**0.867 (0.207)**			
Initial Distance	**Near (10 cm)**	**0.884 (0.238)**	<0.001**	551.771	0.980
	Far (20 cm)	1.075 (0.259)			
Alignment	Collinear	0.981 (0.275)	0.750	0.106	0.010
	Parallel	0.978 (0.285)			
Texture	Texture-on	0.986 (0.264)	0.063	4.278	0.280
	Texture-off	0.973 (0.268)			
Shadow	**Shadow-on**	**0.952 (0.239)**	<0.001**	66.113	0.857
	Shadow-off	1.008 (0.288)			

6.6. Lawful regularities

To find out whether the acquisition of targets with motion-in-depth possess lawful regularities, two established models were used to analyze our data: Jagacinski's model (Jagacinski et al., 1980) and the Ternary-Gaussian model (Huang et al., 2018). The first model provides estimates of *MT* while the second model provides estimates of *ER*s for selecting moving targets.

6.6.1. Jagacinski's model analysis

The Jagacinski's model (Jagacinski et al., 1980) for *MT* estimation in moving target acquisition is given by:

$$MT = a + bA + c(V+1)\left(\frac{1}{W} - 1\right) \tag{2}$$

where A is the initial distance, V and W are the target's velocity and width, and a, b, c, are empirically determined constants. By extracting the b from the last two terms and defining a new free parameter $d = c/b$, Jagacinski et al. derived a modified index of difficulty (*ID*) to characterize *MT* in moving target acquisition:

$$ID = A + d(V+1)\left(\frac{1}{W} - 1\right) \tag{3}$$

As there is a free parameter d in the *ID* of this model, the *ID* must be empirically determined for different circumstances. In Jagacinski et al's original work, d was determined by fitting the model to all collected data yielding a value of 1.594. They also mentioned that a constant "1" is subtracted from $1/W$ so that the interaction with velocity will be minimal for the widest target ($1/W = 1.08$ /° for the widest target in their setting). The subtraction of "1" may therefore be considered as a fourth fitting parameter. The model fit our data much better when setting the subtraction constant as a free parameter compared to unity. Therefore, we chose to fit both d and the constant subtracted from $1/W$ in this

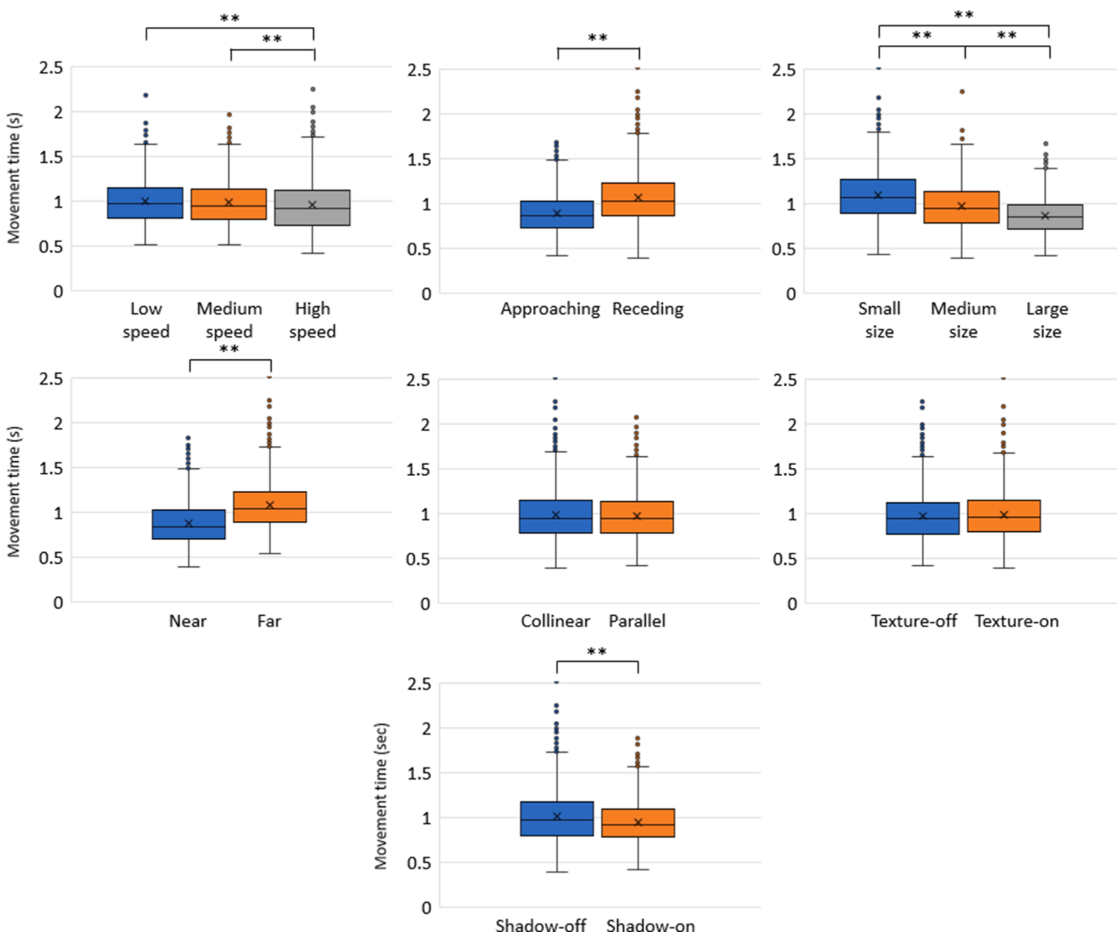

Fig. 4. Boxplots of *MT* for the seven factors of speed, direction, width, initial distance, shadow, texture and alignment. The "x" in the plots indicates the mean value. Asterisks represent pairwise significant differences, which are noted as: $p < 0.001$(**) and $p < 0.05$(*).

study. For clarity, we define the *ID* used in this study as follow:

$$ID = A + d(V+1)\left(\frac{1}{W} - e\right) \quad (4)$$

To ensure a fair comparison between different factors and conditions, we fit d and e in Eq. (4) with all data in this study to obtain a consistent *ID*. To do that, we binned the *MT* data with the *A* (Initial Distance) $\times$ *V* (Speed) $\times$ *W* (Size) conditions, yielding 18 sets of data. We then calculated the mean *MT* in each data set and used them to fit the Jagacinski's model. The result showed a 0.904 R^2 and the constants of d and e were determined to be 2.3320 and 0.7172 respectively.

With the determined *ID*, we fit the Jagacinski's model to our $A \times V \times W$ data sets grouped by different conditions for the factors of Direction, Texture, Shadow, and Alignment, respectively. The fits ranged from 0.746 to 0.918 R^2 as shown in Fig. 6. Strong correlations between the *MT* and *ID* were found when grouping the data by different conditions in Texture, Shadow and Alignment, with R^2 values above 0.848. However, we found relatively low correlations when grouping the data by the two moving directions ($R^2 = 0.769$ for approaching and $R^2 = 0.745$ for receding). In the task of motion-in-depth, there are two opposite influences of target speed on *MT* depending on moving direction: i) when the target was moving away from the participant, increasing speed increases the *MT*; ii) when the target was moving toward the participant, the higher speed leads to shorter *MT* (see Fig. 8). However, we used the same *ID* to fit the data for both moving directions when examining the effect of speed. This led to the poor fits in the two data sets grouped by different moving directions.

In general, the *MT* in each condition could be mostly accounted for by *A* and the *V* to *W* radio in the *ID*. As shown in Fig. 6(a), the two moving directions showed large differences in Jagacinski's model. In the approaching condition, the intercept of the model was only 0.6520 s while it increased to 0.8077 s in the receding condition. In addition, the approaching condition had a lower slope of 0.0179 s/bit compared to 0.0199 s/bit in the receding condition. When shadow was enabled, we obtained a slope of 0.0172 s/bit in Jagacinski's model, which was notably lower than the 0.0206 s/bit obtained in the shadow off condition as shown in Fig. 6(d). The effects of the other two factors on *MT* reflected by Jagacinski's model were small, which is consistent with our statistical analysis (see Fig. 6(b) and (c)).

6.6.2. Ternary-Gaussian model analysis

The second model used to analyze the *ER* data is the Ternary-Gaussian model (Huang et al., 2018), which first models the endpoint distribution of moving targets and then calculates the *ER* via the cumulative distribution functions (CDF). In the Ternary-Gaussian model, the endpoint distribution is assumed to be a random variable X following a Gaussian distribution with mean, μ, and standard deviation, σ:

$$X = N(\mu, \sigma^2) \quad (5)$$

The parameters μ and σ in the Gaussian distribution are formulated as functions of task variables W and V:

$$\mu = a + bW + cV \quad (6)$$

$$\sigma = \sqrt{d + eW^2 + fV^2 + g\frac{V}{W}} \quad (7)$$

Table 4
Statistical results for the factors of Speed, Direction, Width, Initial Distance, Alignment, Texture and Shadow on *ER*. Asterisks represent pairwise significant differences, which are noted as: $p < 0.001$(**) and $p < 0.05$(*). The condition with the lowest *ER* is highlighted in bold for each factor that has a significant effect.

Factors	Levels	*ER* (SD) %	*p*	*F*	η_p^2
Speed	**Low (2 cm/s)**	**43.24 (26.85)**	<0.001**	51.271	0.823
	Medium (4 cm/s)	47.66 (27.04)			
	High (8 cm/s)	54.53 (28.72)			
Direction	**Approaching**	**45.37 (27.32)**	<0.001**	40.018	0.784
	Receding	51.59 (28.19)			
Width	Small (1 cm)	65.65 (23.38)	<0.001**	290.132	0.963
	Medium (2 cm)	46.85 (24.95)			
	Large (4 cm)	**32.93 (25.15)**			
Initial Distance	**Near (10 cm)**	**46.21 (27.79)**	<.001**	69.567	0.863
	Far (20 cm)	50.69 (27.93)			
Alignment	Collinear	52.31 (27.74)	<0.001**	38.550	.778
	Parallel	**44.65 (27.60)**			
Texture	Texture-on	48.78 (27.41)	0.469	0.563	0.049
	Texture-off	48.17 (28.45)			
Shadow	**Shadow-on**	**46.12 (28.19)**	<0.001**	42.758	0.795
	Shadow-off	50.84 (27.48)			

where a, b, c, d, e, f and g are empirically determined constants. Notice that in the Ternary-Gaussian model, the initial distance A was considered to have no significant effect on the endpoint distribution, thus the model does not include the term A (Huang et al., 2018). Given the endpoint distribution, the *ER* can be estimated via the cumulative distribution functions as followed:

$$\begin{aligned} ER(\mu,\sigma) &= 1 - [P(x_1) - P(x_0)] \\ &= 1 - \frac{1}{2}\left[erf\left(\frac{x_1-\mu}{\sigma\sqrt{2}}\right)\right) - erf\left(\frac{x_0-\mu}{\sigma\sqrt{2}}\right))] \end{aligned} \tag{8}$$

where x_0 and x_1 represent the boundaries (in any one dimension) of the target and $erf(x)$ is the error function encountered in integrating the Gaussian distribution. To see how the investigated factors affect the speed-accuracy trade-off in terms of *ER* and endpoint distribution, we fit the Ternary–Gaussian model to $V \times W$ data sets grouped by different Direction, Texture, Shadow, and Alignment conditions, respectively. The model obtained high fits ranging from 0.940 to 0.994 R^2 in our data. The *ER* spectrum predicted by the model are illustrated in Fig. 7.

From the *ER* spectrum, we can clearly see that the *ER*s increased as speed increased and width decreased. As shown in Fig. 7(a), compared with the receding condition, the *ER* in the approaching condition showed a gentler growth curve under the influence of speed, especially for the large targets. In higher speed conditions, approaching targets could have up to 20% lower *ER*s compared to receding ones. In Fig. 7(b), although the *ER*s in the collinear condition were higher (about 10%) as a whole, the *ER* curves of the two conditions show roughly the same trend under the influence of size and speed. In Fig. 7(d), although not very clear, we can see bigger deltas of *ER*s (about 10%) by taking the shadow off in speed = 4 cm/s conditions compared to 2 cm/s and 8 cm/s. It may indicate that shadow has a greater influence on selection accuracy at medium speed. Finally, there was only a slight reduction of the increasing trend of *ER*s by enabling texture in the scene as shown in Fig. 7(c).

6.7. Discussion

In this section we discuss and interpret the findings from Study 2.

6.7.1. Movement time

The results from Study 2 show that targets with higher moving speeds were acquired significantly faster than targets with lower moving speeds. This finding is not consistent with the finding of Jagacinski et al.'s study (Jagacinski et al., 1980) which found that participants were, conversely, slower at capturing faster moving targets.

In order to understand this contradiction, we examined how *MT* changes with moving speed under different settings of the other six factors. We found that for different moving directions, *MT* had two completely different trends with respect to target speed. As illustrated in Fig. 8, when the target was moving away from the participant, which was similar to the setting in Jagacinski's study, increasing speed increased *MT*. In contrast, when the target was moving toward the participant, the higher speed produced shorter *MT*. This conclusion is supported by previous works from Tresilian (2005) and Tresilian and Lonergan (2002), who found that *MT* is not always positively correlated with target speed. Despite the fact that there are two opposite influences of speed on *MT*, there is still a general trend that faster speed results in shorter *MT* overall. We attribute this to the fact that participants generally had to move faster when acquiring faster moving targets, resulting in an overall shorter *MT*.

Our data also reveals that some aspects of performance for targets motion-in-depth are similar to static targets. Specifically, participants were observed to be faster at acquiring closer initialized and larger targets. Enabling shadow also helped to reduce the time needed to acquire the targets.

6.7.2. Error rate

We found that target Moving Direction had a significant effect on *ER*. In Study 1, participants were found to have lower perception accuracy for receding targets but higher perception accuracy for approaching targets. This reduced perception accuracy for receding targets likely explains the observation that participants made more errors when acquiring receding targets than when acquiring approaching targets.

We found a significant effect for Initial Distance on *ER* when acquiring targets with motion-in-depth. This result appears inconsistent with previous studies in 1D moving target selection (Huang et al., 2018), which have demonstrated that the initial distance of moving targets in the lateral direction has a minimal effect on *ER*s. However, in our motion-in-depth setting involving 3D projection in a VR display, a more distant target is visually smaller. This smaller visual size results in lower accuracy in reaching movements and is likely responsible for the different effect observed between motion-in-depth and 1D/2D tasks (Huang et al., 2018; 2019a) for initial target distance.

The results reveal that Alignment had a significant effect on the pointing *ER*. Setting the target motion to be misaligned from the participants' viewing direction can not only provide a clearer view of the target location but also a clearer view of the boundary of the target. It is likely that this improved observability of the target was responsible for the increased pointing accuracy.

Similar to previous results on static targets, participants demonstrate lower *ER* for targets with larger sizes, and targets with shadow. As observed in the context of moving target acquisition in 1D/2D settings (Huang et al., 2018; 2019a), increasing target speed was found to increase selection *ER*.

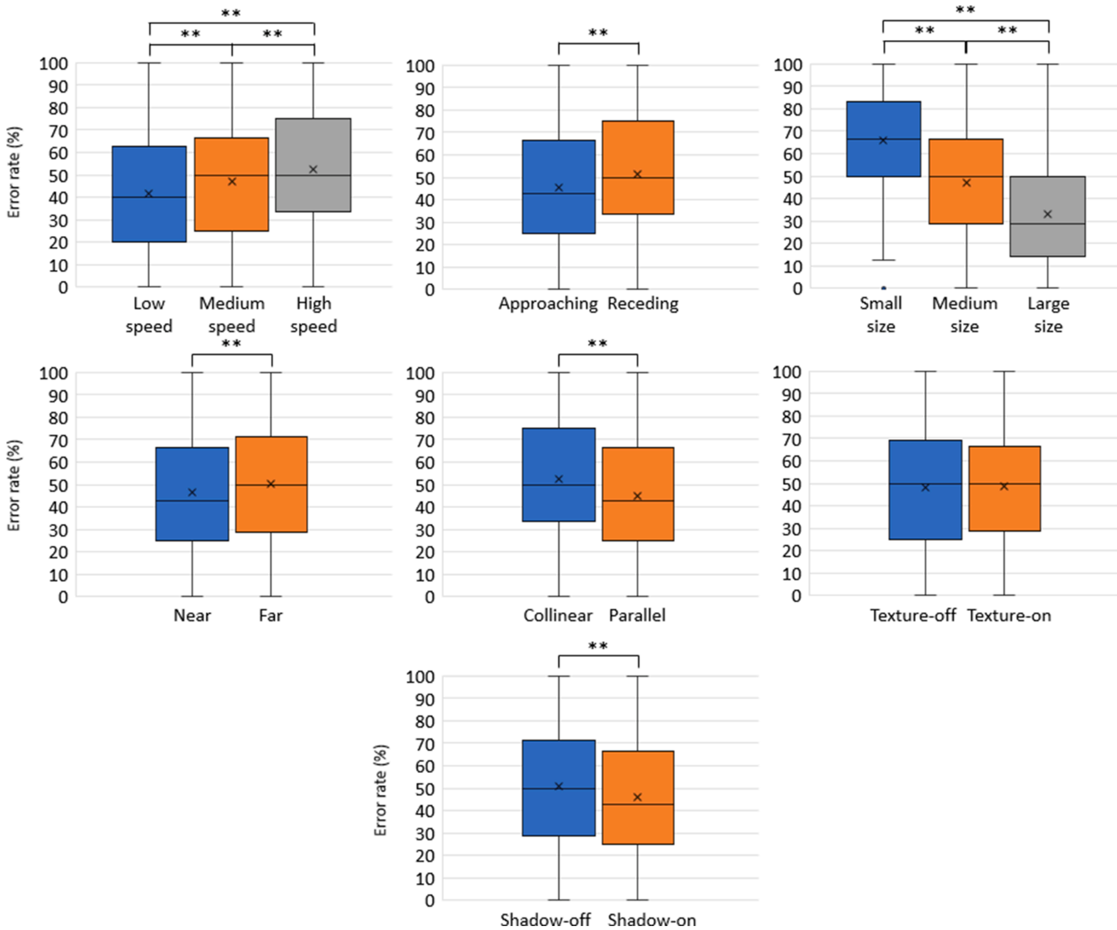

Fig. 5. Boxplots of the *ER* for the seven factors of speed, direction, width, initial distance, shadow, texture and alignment. The "*x*" in the plots indicates the mean value. Asterisks represent pairwise significant differences, which are noted as: $p < 0.001$(**) and $p < 0.05$(*).

6.7.3. Lawful regularities

The data from Study 2 reveals strong lawful regularities in the acquisition of targets with motion-in-depth. Excluding the conditions where the data was grouped according to different moving directions, the *MT* can be largely accounted for by Jagacinski's *ID* showing an overall R^2 of 0.904. High fits were also found when fitting the Ternary-Gaussian model to the *ER* with R^2 ranging from 0.940 to 0.994.

As illustrated by Jagacinski's model, the moving direction and shadow have a large impact on the intercepts and slope of the *MT* function. According to the model, changing the moving direction from receding to approaching reduces the absolute time needed for acquiring targets by approximately 0.15 s and reduces the rate at which *MT* increases with task difficulty by at least 10%. The model also suggests that enabling shadow can reduce the rate at which *MT* increases with task difficulty by more than 16%.

Inspecting the *ER* spectrum predicted by the Ternary-Gaussian model, the *ER* in the approaching condition showed a gentler growth curve under the influence of speed. In higher speed conditions, the model suggests that approaching targets may have up to 20% lower *ER* compared to receding ones. According to the Ternary-Gaussian model, setting the target motion to be misaligned with the user's viewing direction may reduce *ER* by approximately 10%.

7. General discussion

In this section, we review the studies' key findings and derive informative design implications based on those findings. Finally, we describe how our findings can be applied to real-world interaction scenarios.

7.1. Key findings

One of the most interesting findings from this study is that the effect of target speed on *MT* depends on moving direction of the target: if the target moves away from the user, higher speed leads to longer *MT*. Conversely, if the target moves toward the users, higher speed leads to shorter *MT*. This result suggests that the participants adopt different acquisition strategies when dealing with targets in different moving directions. In Study 2, the initial position of the cursor was set between the participants and the initial position of the target. For receding targets, the participants are likely to choose the "pursue and capture" strategy (Jagacinski et al., 1980), while for approaching targets, the participants may choose the "anticipate and strike" strategy (Tresilian and Lonergan, 2002). The difference between the two behavior strategies is likely responsible for the contrasting effect of target speed on *MT*.

Another empirical result we found to be distinct from moving target acquisition in previously examined 1D/2D tasks (Huang et al., 2018; 2019a) refers to the significant effect of initial distance on selection *ER*. Because of the 3D projection, more distant targets are visually smaller, which in turn produces lower accuracy in reaching movements. For 1D/2D target acquisition with movement in the lateral direction, however, the target's visual size does not substantially change during movement. This distinction between the 1D/2D and 3D settings is likely responsible for the different effect of initial distance.

We found that Speed had the greatest impact on participants' perception accuracy, followed by Shadow and Direction. Given the relatively large design range for moving speed, it could be that this is the

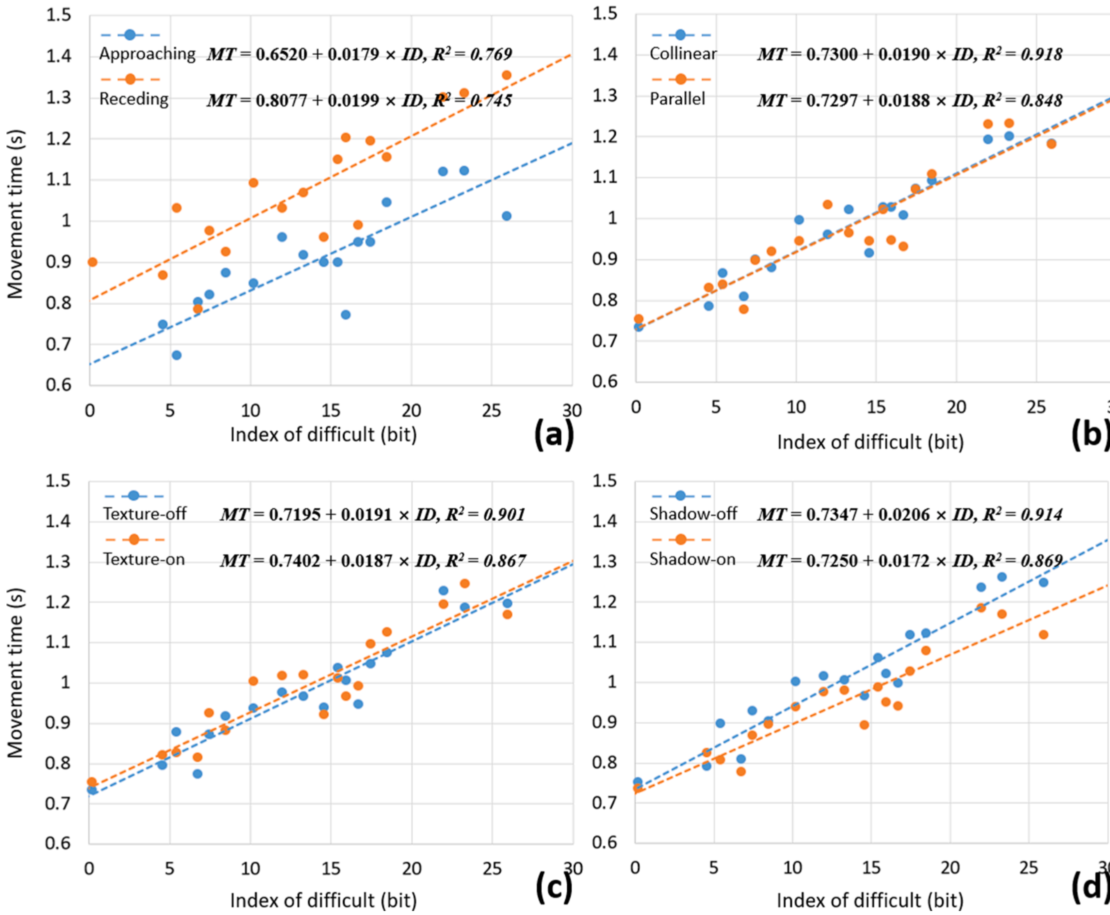

Fig. 6. Jagacinski's models for different conditions for the factors of a) Direction, b) Alignment, c) Texture and d) Shadow.

dominant factor to consider when manipulating perception accuracy in VR applications.

Acquisition of targets with motion-in-depth shows strong lawful regularities. *MT* in this task can be largely accounted for by Jagacinski's *ID* showing an overall R^2 of 0.904. Meanwhile, the *ER*s observed in Study 2 fit well with the Ternary-Gaussian model with R^2 ranging from 0.940 to 0.994.

7.2. Design implications

The above findings could not only enrich the literature on moving target acquisition in the context of VR interaction, but also provide recommendations for VR interface design. We present the following implications and takeaways that could be helpful for creating VR user interfaces involving targets with motion-in-depth.

1. Moving speed is the most effective design factor influencing the user's perception accuracy of targets with motion-in-depth. For VR applications that require users to make an accurate judgment on the location of targets with motion-in-depth, designers should consider using technical means to reduce the moving speed of the targets.
2. Moving direction of target in depth dimension (approaching/receding) significantly affects pointing accuracy. For VR applications that contain targets requiring motion-in-depth, designers should give priority to the use of approaching movement, as it may reduce *ER* by up to 20%.
3. In the condition that a target is moving towards the user, designers can consider increasing moving speed of the target which will shorten the *MT*.
4. Initial distance in the acquisition of targets with motion-in-depth is much more important than in conventional user interfaces. Designers should be aware of the effects of initial distance on moving target acquisition.
5. Enabling shadow improves the user's perception accuracy, reduces *MT* and *ER*. VR designers should consider enabling shadow where possible to enhance the user experience of acquiring targets with motion-in-depth.
6. Making target motion misaligned with the user's viewing direction can steadily reduce *ER* by about 10%. If the application scenario requires high pointing accuracy, designers should consider setting the target motion to be misaligned with the user's viewing direction. However, such a design decision may not help to reduce the time needed to select the target.
7. Acquisition of targets with motion-in-depth shows strong lawful regularities. *MT* and *ER* in acquiring targets with motion-in-depth can be largely accounted for by Jagacinski's model and by the Ternary-Gaussian model, respectively. Designers can consider using these models to assist with user interface design.

We should mention that the aforementioned implications can not only work in targets with motion purely in the depth dimension. These implications could also be applied in content with combined 3D motion in VR.

7.3. Applications

The findings of this study may inform the development of VR applications in terms of two alternative design goals: 1) improving interaction efficiency; and 2) increasing the challenge of specified tasks.

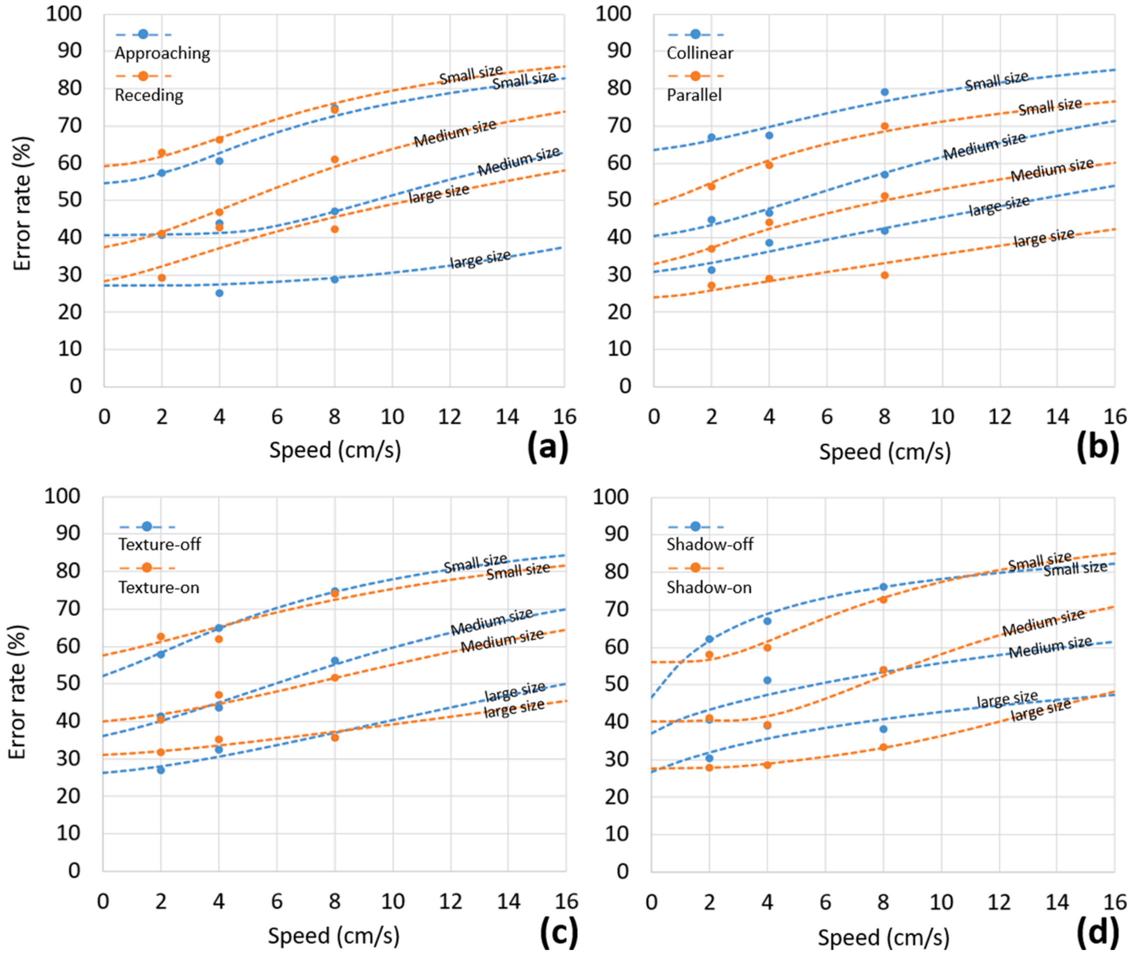

Fig. 7. *ER* spectrum for different conditions predicted by the Ternary-Gaussian model. Dash lines are predicted *ER*s while circles mark the actual data.

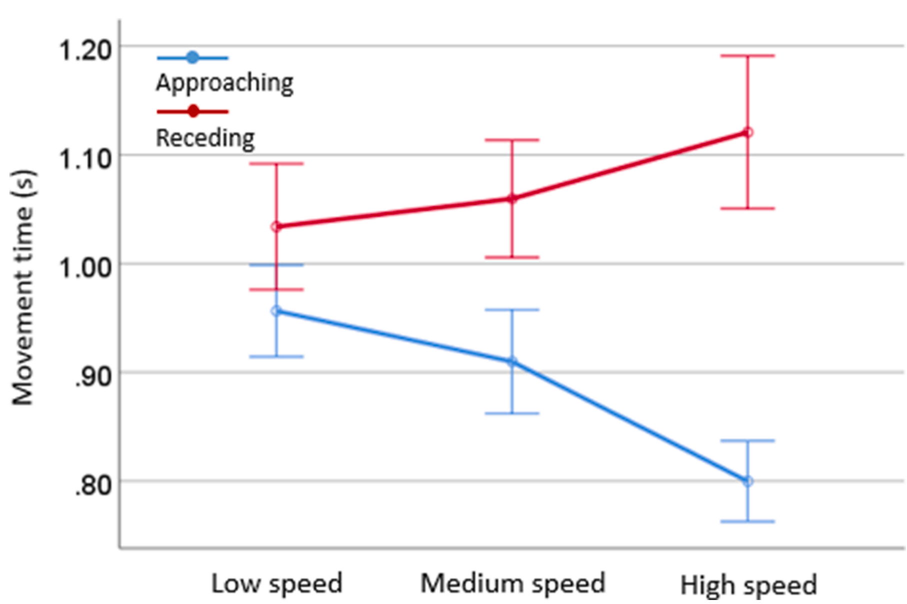

Fig. 8. Interaction between Speed and Direction on *MT*. Error bars indicate the 95% confidence interval.

For the first design goal, our findings may help to improve the design of objects that are presented in VR, such as for moving objects presented in sporting games or animated elements presented in simulation or data visualization applications. As an example, consider a VR application for the dynamic simulation or real-time monitoring of an urban environment (Ketzler et al., 2020; Rudskoy et al., 2021). Such an application may simulate all vehicle movements in the city and allow users to observe and interact with each vehicle in real time. Users may wish to observe the traffic flow on a highway in detail and obtain vehicle information at any time. Based on the findings of this study, it would be sensible to constrain the observation point of the user such that: i) they are above or on the side of the highway rather than collinear with the highway; and ii) the traffic flow move towards the user rather than away. Applying these constraints in combination with choosing an appropriate speed for traffic flow simulation, would improve the user's ability to clearly observe the dynamic traffic flow, as well as their ability to quickly and accurately select any of the vehicles. Similar scenarios can be found in many other fields, such as surgery training (Kühnapfel et al., 2000), astronomy education (Chen et al., 2007), and manufacturing simulations (Ong and Mannan, 2004).

The second design goal is most relevant to the design of VR games. Designers could use the findings presented as a basis for modulating the challenge level in game play. For example, in a VR baseball game like *Everyday Baseball VR*,[1] perhaps the most obvious way to increase the challenge is to speed up the movement of the baseball or reduce the size of the ball. However, our findings also reveal that animating the ball's approach such that it is collinear with the user's viewing direction would likely make it more difficult to perceive or hit than offsetting the ball by a certain distance from the user. Our findings could inspire game designers to create a scene with weakened depth clues, such as less obvious shadows caused by dim lighting, thereby making the game play more challenging. Similar scenarios can be found in the popular VR game *Beat Saber*,[2] in which players must hit the blocks that are constantly flying towards them. Aside from speeding up the blocks, our findings suggest the another way to make the game more difficult is to generate some blocks that move away from the player or fly directly towards the player along the player's viewing direction.

7.4. Limitations and future work

This work did not cover many other potential factors that could affect motion-in-depth, such as blur (Held et al., 2010), depth of field (Zhang et al., 2014) and occlusion (Zhai et al., 1996). These factors affect participants' depth perception since they convey information for spatial relations by different means. Given that the previous studies examining these factors were carried out on static objects only, it would be interesting to investigate what effect these factors might have on the perception and acquisition of moving targets. It is also worthy to investigate how the effects of blur, depth of field, and occlusion interact with the target's moving speed and moving direction as the status of these factors could change dynamically as the target moves in depth.

In this paper, we have not considered fully-3D-motion that combines motion-in-depth with motion in the x and y dimensions. We see such a study as a challenge due to its potentially large problem space: there is a large number of moving directions, and the conditions generated across moving direction and other factors would be numerous. It is challenging to empirically investigate such a problem using a conventional factor analysis approach. We therefore see the potential for the development of a new descriptive model to appropriately analyze such data.

8. Conclusions

In this work, we first defined the concept of motion-in-depth in VR interfaces. Following this, we investigated the effects of texture, shadow, alignment, moving speed and moving direction on perception accuracy and user performance for content with motion-in-depth in the context of VR. Two studies were conducted to identify how various design factors affect perception accuracy and performance when selecting 3D objects with motion-in-depth in VR. Our results indicate that target speed has the greatest impact on users' perception accuracy, followed by shadow and moving direction. We identified effects of moving speed, moving direction, initial distance, shadow and alignment on *MT* and *ER* that are distinct from previous studies examining conventional user interfaces. Further, the good fits of Jagacinski's model and the Ternary-Gaussian model to the *MT* and *ER* data show that the motion-in-depth task abides by lawful regularities. In summary, this paper advances our understanding of how users perceive and interact with moving targets in VR.

[1] https://www.oculus.com/experiences/rift/1884062488298857/

[2] https://www.beatsaber.com/

CRediT authorship contribution statement

Jin Huang: Writing – original draft, Methodology, Data curation. **John J. Dudley:** Writing – review & editing. **Stephen Uzor:** Writing – review & editing. **Dong Wu:** Software. **Per Ola Kristensson:** Supervision, Writing – review & editing. **Feng Tian:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 62172397) and Youth Innovation Promotion Association CAS (Grant no. 2020113).

References

Accot, J., Zhai, S., 2002. More than dotting the i's—Foundations for crossing-based interfaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 73–80.
Argelaguet, F., Andujar, C., 2013. A survey of 3D object selection techniques for virtual environments. Comput. Graph. 37 (3), 121–136.
Barrera Machuca, M.D., Stuerzlinger, W., 2019. The effect of stereo display deficiencies on virtual hand pointing. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–14.
Batmaz, A.U., Machuca, M.D.B., Pham, D.M., Stuerzlinger, W., 2019. Do head-mounted display stereo deficiencies affect 3D pointing tasks in AR and VR? 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, pp. 585–592.
Brenner, E., Smeets, J.B., 1996. Hitting moving targets: co-operative control of 'when' and 'where'. Hum. Mov. Sci. 15 (1), 39–53.
Chen, C.-H., Yang, J.-C., Shen, S., Jeng, M.-C., 2007. A desktop virtual reality earth motion system in astronomy education. J. Educ. Technol. Soc. 10 (3), 289–304.
Dragicevic, P., 2004. Combining crossing-based and paper-based interaction paradigms for dragging and dropping between overlapping windows. Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology, pp. 193–196.
Feiner, A.O.S., 2003. The flexible pointer: an interaction technique for selection in augmented and virtual reality. Proc. UIST, vol. 3, pp. 81–82.
Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. 47 (6), 381.
Franklin, D.W., Wolpert, D.M., 2011. Computational mechanisms of sensorimotor control. Neuron 72 (3), 425–442.
Gibson, J. J., 1950. The Perception of the Visual World.
Grossman, T., Balakrishnan, R., 2006. The design and evaluation of selection techniques for 3D volumetric displays. Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, pp. 3–12.
Hasan, K., Grossman, T., Irani, P., 2011. Comet and target ghost: techniques for selecting moving targets. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 839–848.
Held, R.T., Cooper, E.A., O'Brien, J.F., Banks, M.S., 2010. Using blur to affect perceived distance and size. ACM Trans. Graph. 29 (2), 19:1–19:16. https://doi.org/10.1145/1731047.1731057.
Hinckley, K., Pausch, R., Goble, J.C., Kassell, N.F., 1994. A survey of design issues in spatial input. Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology, pp. 213–222.
Hoffmann, E.R., 1991. Capture of moving targets: a modification of Fitts' law. Ergonomics 34 (2), 211–220.
Huang, J., Tian, F., Fan, X., Tu, H., Zhang, H., Peng, X., Wang, H., 2020. Modeling the endpoint uncertainty in crossing-based moving target selection. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–12.
Huang, J., Tian, F., Fan, X., Zhang, X., Zhai, S., 2018. Understanding the uncertainty in 1D unidirectional moving target selection. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–12.

Huang, J., Tian, F., Li, N., Fan, X., 2019. Modeling the uncertainty in 2D moving target selection. Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, pp. 1031–1043.
Huang, S., Daqing, Q., Jiabin, Y., Huawei, T., 2019. Review of studies on target acquisition in virtual reality based on the crossing paradigm. Virtual Real. Intell. Hardw. 1 (3), 251–264.
Hubona, G.S., Wheeler, P.N., Shirah, G.W., Brandt, M., 1999. The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization. ACM Trans. Computer-Human Interact. (TOCHI) 6 (3), 214–242.
Jagacinski, R.J., Repperger, D.W., Ward, S.L., Moran, M.S., 1980. A test of Fitts' law with moving targets. Hum. Factors 22 (2), 225–233.
Janzen, I., Rajendran, V.K., Booth, K.S., 2016. Modeling the impact of depth on pointing performance. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 188–199.
Kendall, M.G., Smith, B.B., 1939. The problem of m rankings. Ann. Math. Stat. 10 (3), 275–287.
Ketzler, B., Naserentin, V., Latino, F., Zangelidis, C., Thuvander, L., Logg, A., 2020. Digital twins for cities: a state of the art review. Built Environ. 46 (4), 547–573.
Kühnapfel, U., Cakmak, H.K., Maaß, H., 2000. Endoscopic surgery training using virtual reality and deformable tissue simulation. Comput. Graph. 24 (5), 671–682.
Lee, B., Kim, S., Oulasvirta, A., Lee, J.-I., Park, E., 2018. Moving target selection: a cue integration model. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–12.
Lee, B., Oulasvirta, A., 2016. Modelling error rates in temporal pointing. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 1857–1868.
Lee, I., Kim, S., Lee, B., 2019. Geometrically compensating effect of end-to-end latency in moving-target selection games. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12.
Lin, C.J., Woldegiorgis, B.H., 2017. Egocentric distance perception and performance of direct pointing in stereoscopic displays. Appl. Ergon. 64, 66–74.
Lubos, P., Bruder, G., Steinicke, F., 2014. Analysis of direct selection in head-mounted display environments. 2014 IEEE Symposium on 3D User Interfaces (3DUI). IEEE, pp. 11–18.
Luo, Y., Vogel, D., 2014. Crossing-based selection with direct touch input. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2627–2636.
Mamassian, P., Knill, D.C., Kersten, D., 1998. The perception of cast shadows. Trends Cogn. Sci. 2 (8), 288–295. https://doi.org/10.1016/S1364-6613(98)01204-2. Publisher: Elsevier
Maruhn, P., Schneider, S., Bengler, K., 2019. Measuring egocentric distance perception in virtual reality: influence of methodologies, locomotion and translation gains. PLoS One 14 (10), e0224651. https://doi.org/10.1371/journal.pone.0224651.Publisher: Public Library of Science
Miall, R.C., Weir, D.J., Wolpert, D.M., Stein, J., 1993. Is the cerebellum a smith predictor? J. Mot. Behav. 25 (3), 203–216.
O'Brien, J., Johnston, A., 2000. When texture takes precedence over motion in depth perception. Perception 29 (4), 437–452. https://doi.org/10.1068/p2955.PMID: 10953763, arXiv:
Ong, S.-K., Mannan, M., 2004. Virtual reality simulations and animations in a web-based interactive manufacturing engineering module. Comput. Educ. 43 (4), 361–382.
Park, E., Lee, B., 2020. An intermittent click planning model. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13.
Poupyrev, I., Billinghurst, M., Weghorst, S., Ichikawa, T., 1996. The go-go interaction technique: non-linear mapping for direct manipulation in VR. Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology, pp. 79–80.
Proffitt, D.R., Stefanucci, J., Banton, T., Epstein, W., 2003. The role of effort in perceiving distance. Psychol. Sci. 14 (2), 106–112. https://doi.org/10.1111/1467-9280.t01-1-01427.
Puerta, A.M., 1989. The power of shadows: shadow stereopsis. JOSA A 6 (2), 309–311. https://doi.org/10.1364/JOSAA.6.000309.Publisher: Optical Society of America
Ramcharitar, A., Teather, R.J., 2018. EZCursorVR: 2D selection with virtual reality head-mounted displays. Proceedings of the 44th Graphics Interface Conference. Canadian Human-Computer Communications Society, pp. 123–130.
Renner, R.S., Velichkovsky, B.M., Helmert, J.R., 2013. The perception of egocentric distances in virtual environments—A review. ACM Comput. Surv. 46 (2), 23:1–23:40. https://doi.org/10.1145/2543581.2543590.
Rolin, R.A., Fooken, J., Spering, M., Pai, D.K., 2018. Perception of looming motion in virtual reality egocentric interception tasks. IEEE Trans. Vis. Comput. Graph. 25 (10), 3042–3048.
Rudskoy, A., Ilin, I., Prokhorov, A., 2021. Digital twins in the intelligent transport systems. Transp. Res. Procedia 54, 927–935.
Sauer, C.W., Saidpour, A., Braunstein, M.L., Andersen, G.J., 2001. Perceived depth of 3-D objects in 3-D scenes. Perception 30 (6), 681–692. https://doi.org/10.1068/p3087.
Smeets, J.B., Brenner, E., 1995. The visual guidance of ballistic arm movements. Multisensory Control of Posture. Springer, pp. 191–197.
Smets, G.J.F., 1992. Designing for telepresence: the interdependence of movement and visual perception implemented. IFAC Proc. Vol. 25 (9), 169–175. https://doi.org/10.1016/S1474-6670(17)50188-3.
Teather, R.J., Stuerzlinger, W., 2013. Pointing at 3D target projections with one-eyed and stereo cursors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 159–168.
Todd, J.T., Akerstrom, R.A., 1987. Perception of three-dimensional form from patterns of optical texture. J. Exp. Psychol. 13 (2), 242.
Tresilian, J.R., 2005. Hitting a moving target: perception and action in the timing of rapid interceptions. Percept. Psychophys. 67 (1), 129–149.
Tresilian, J.R., Lonergan, A., 2002. Intercepting a moving target: effects of temporal precision constraints and movement amplitude. Exp. Brain Res. 142 (2), 193–207.
Tu, H., Huang, S., Yuan, J., Ren, X., Tian, F., 2019. Crossing-based selection with virtual reality head-mounted displays. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–14.
Wanger, L., 1992. The effect of shadow quality on the perception of spatial relationships in computer generated imagery. Proceedings of the 1992 Symposium on Interactive 3D Graphics. Association for Computing Machinery, New York, NY, USA, pp. 39–42. https://doi.org/10.1145/147156.147161.
Wanger, L.R., Ferwerda, J.A., Greenberg, D.P., 1992. Perceiving spatial relationships in computer-generated images. IEEE Comput. Graph. Appl. 12 (3), 44–58.
Wann, J.P., Rushton, S., Mon-Williams, M., 1995. Natural problems for stereoscopic depth perception in virtual environments. Vis. Res. 35 (19), 2731–2736. https://doi.org/10.1016/0042-6989(95)00018-U.
Wilcoxon, F., 1992. Individual comparisons by ranking methods. Breakthroughs in Statistics. Springer, pp. 196–202.
Wobbrock, J.O., Cutrell, E., Harada, S., MacKenzie, I.S., 2008. An error model for pointing based on Fitts' law. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1613–1622.
Wolf, D., Dudley, J.J., Kristensson, P.O., 2018. Performance envelopes of in-air direct and smartwatch indirect control for head-mounted augmented reality. 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, pp. 347–354.
Yonas, A., 1979. Attached and cast shadows. Perception and Pictorial Representation. Praeger.
Young, M.J., Landy, M.S., Maloney, L.T., 1993. A perturbation analysis of depth perception from combinations of texture and motion cues. Vis. Res. 33 (18), 2685–2696.
Zhai, S., Buxton, W., Milgram, P., 1996. The partial-occlusion effect: utilizing semitransparency in 3D human-computer interaction. ACM Trans. Comput.-Hum. Interact. 3 (3), 254–284. https://doi.org/10.1145/234526.234532.
Zhang, T., O'hare, L., Hibbard, P.B., Nefs, H.T., Heynderickx, I., 2014. Depth of field affects perceived depth in stereographs. ACM Trans. Appl. Percept. 11 (4), 18:1–18:18. https://doi.org/10.1145/2667227.