2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)

DepthMove: Leveraging Head Motions in the Depth Dimension to Interact with Virtual Reality Head-Worn Displays

Difeng Yu, Hai-Ning Liang†, Xueshi Lu, Tianyu Zhang, Wenge Xu
Xi'an Jiaotong-Liverpool University

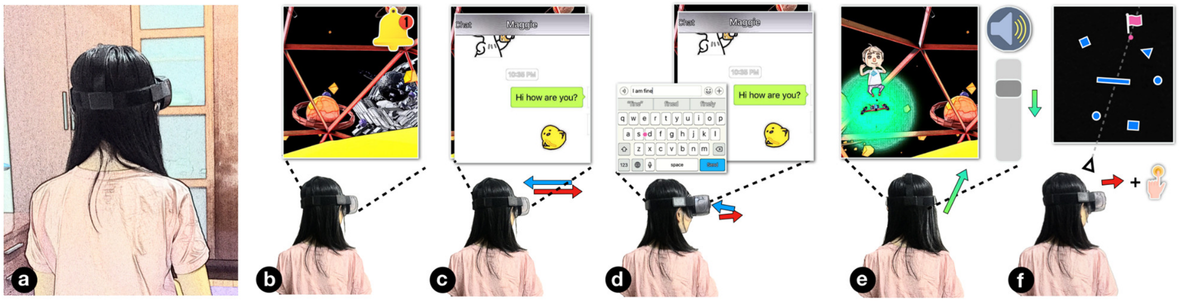

Fig.1. An application scenario for DepthMove. (a-b) A user is watching a movie using a virtual reality head-worn display. She then performs a series of movements towards the depth dimension to (c) switch between interfaces, (d) enter some short burst of texts in an instant messenger app, and (e) turn down the volume of the video player. (f) An example of DepthMove working to complement hand-based input to allow selection of fully occluded targets.

ABSTRACT

Head-based interactions are very handy for virtual reality (VR) head-worn display (HWD) systems. A useful head-based interaction technique could help users to interact with VR environments in a hands-free manner (i.e., without the need of a hand-held device). Moreover, it can sometimes be seamlessly integrated with other input modalities to provide richer interaction possibilities. This paper explores the potential of a new approach that we call DepthMove to allow interactions that are based on head motions along the depth dimension. With DepthMove, a user can interact with a VR system proactively by moving the head perpendicular to the VR HWD forward or backward. We use two user studies to investigate, model, and optimize DepthMove by taking into consideration user performance, subjective response, and social acceptability. The results allow us to determine the optimal and comfortable DepthMove range. We also distill recommendations that can be used to guide the design of interfaces that use DepthMove for efficient and accurate interaction in VR HWD systems. A third study is conducted to demonstrate the usefulness of DepthMove relative to other techniques in four application scenarios.

Keywords: Virtual reality, target selection, head-based interaction, hands-free interaction, head-worn displays, 3D position tracking

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction Styles

1 INTRODUCTION

Head-based interactions are indispensable for virtual reality (VR) head-worn display (HWD) systems. When a user's hands are occupied, or other ancillary devices (e.g., handheld controllers) are not available or not easily accessible, efficient head-based interaction techniques could be an alternative hands-free input approach for the user to interact with the VR system. Fig. 1a-b present an example scenario which calls for a feasible head-based interaction approach. A user is watching a movie in an HWD-based VR environment but suddenly gets interrupted by a message from an instant messaging app. When watching movies, the user does not usually carry a controller, and it might be inconvenient for her to take down the VR headset to find one. In this situation, a head-based approach to switch to the messaging interface and send a quick reply could become very practical.

Two input mechanisms could be suitable for the above scenario: speech control and gaze control. Although speech control mechanisms have been shown to be efficient in many situations (e.g. [55, 56]), they could also be disturbing to others in a shared environment [57]. Gaze control has shown to be inaccurate at times [36, 58] and having gaze trackers in the VR HWD increases their cost.

For most current HWD headsets, it may not be necessary to have additional peripheral devices for interaction since the headsets have a number of sensors that can capture head and even body movements [32, 64]. These movements can be translated into commands for the HWD without needing a handheld device. Current head-based interaction techniques for HWD are based mainly on the use of the dwell-based approach [38], which requires users to focus on a target for a certain period of time (dwell time) in order to select and then interact with it. However, it is difficult to determine the dwell time [58] since longer dwell times slow down performance, while shorter dwell times might cause unintentional selections and errors. Moreover, because of the pre-defined dwell time, users are always "pushed" to select a target and quickly move to the next one. The user needs to be very focused and act carefully to avoid making unwanted false selections—this process can make interaction stressful and tiring.

This paper explores a new head-based interaction approach we name DepthMove for VR HWD. It allows the user to select and interact with objects by making depth dimension movements—that is,

---

* Difeng Yu, Hai-Ning Liang, Xueshi Lu, Tianyu Zhang, and Wenge Xu are with the Dept. of Computer Science and Software Engineering at Xi'an Jiaotong-Liverpool University, Suzhou, China. Emails: {Difeng.Yu14; Xueshi.Lu17; Tianyu.Zhang16}@student.xjtlu.edu.cn; {HaiNing.Ling; Wenge.Xu}@xjtlu.edu.cn.
† Corresponding Author: Hai-Ning Liang; Email:haining.liang@xjtlu.edu.cn

978-1-7281-4765-9/19/$31.00 ©2019 IEEE
DOI 10.1109/ISMAR.2019.00-20

moving the head perpendicular to the VR HWD forward or backward. With DepthMove, a user can not only interact with VR in a hands-free manner, which is suitable for the above example scenario, but also use DepthMove as an additional complementary channel to other input modalities, such as using a hand controller. It can, therefore, offer richer interaction possibilities for coping with complex scenarios like selecting fully occluded targets (see Fig 1f; more on this later).

Two user studies are conducted to evaluate DepthMove on selection tasks. In the first study, we use a 1D selection task to investigate the feasibility of DepthMove. We then use the users' performance data and subjective responses to model the movement time and extrapolate the desirable, optimal movement range. In the second study, we further explore the desired features (including target sizes, placement directions, distances, and cursor gain ratios) of 3D flat user interfaces (UI), like a 2D control panel located in 3D, when using DepthMove. From the results, we are able to distill several recommendations for designing UI that can take advantage of DepthMove. In a third follow-up study, we compare DepthMove with other interaction techniques in four practical scenarios to demonstrate its potential use.

The contributions of this work include: (1) the DepthMove approach for head-based interactions in VR HWD; (2) an in-depth evaluation of DepthMove; (3) ten design recommendations for using DepthMove; and (4) general interaction scenarios and specific applications of DepthMove in various types of interactions.

2 BACKGROUND AND RELATED WORK

In this section, we present related work with respect to interactions based on HWD, depth dimension movements, and perception of depth in VR HWD.

2.1 Interactions based on HWD

Head-based interactions have been extensively explored in HWD [24, 60] and other display types [25, 26, 27]. We present these head-based interactions in HWD from the perspective of selection and manipulation tasks [7].

Head-based selection techniques with HWD have been proposed from the early days [30, 31]. Typically, these techniques use a virtual ray that emanates from the tracked head position in the direction towards which the head is facing. The user is able to select the closest object that is intersected by the ray when a confirmation is issued, like pressing a button on a controller [36], tilting the head [37], or dwelling on the target [38]. Some refinement techniques (secondary refinement based on the primary selection) [32] and pointing facilitators [33] have also been explored to enhance the selection performance within the context of head-based selection in HWD. The recent work of Pinpointing [36] demonstrates that head movements are significantly more accurate than eye-based techniques. Moreover, head movement as a refinement technique is generally faster than hand gestures and handheld device refinements with AR HWD [36]. This indicates the potential of using head motions as an accurate and relatively fast selection approach. In addition, head-based selection has also been applied for text entry [34] and for users with motor disabilities [35].

Head-based manipulation techniques have also been studied. Most of them leverage the flexibility and variety of head rotational movements. For example, head rotation (orientation) has been used in the HWD-based interface Gluey [39] to help with content migration. Other activities include rotating the camera around a specific anchor [40], performing scaling-liked tasks [41], interacting with maps (such as panning and zooming) [42], and switching to a new state [43]. Apart from selection and manipulation, head-based techniques have also been proposed for navigation tasks [7, 44, 45].

According to our review, although a large number of techniques have been proposed leveraging the flexibility and variety of head orientations (yaw, pitch, and roll), interactions based on head translations along the 3D space (forward/backward, up/down, and left/right) remain largely underexplored and underutilized. This is not an efficient use of current HWD because head position tracking is supported by many commercial 6DoF HWD including both stationary (like Oculus RIFT and HTC VIVE) and standalone systems (such as Oculus Quest and Microsoft HoloLens). Moreover, from our perspective, these translational movements have great potential to issue interactive commands that are fast and precise. Therefore, this research explores the usage of the head translational movements, specifically depth dimension movements (forward or backward), for interacting with objects in VR environments.

2.2 Depth Dimension Movements

According to our above review, very few studies have been conducted to assess the strength and use of head translational movements in the depth dimension within HCI. One of the reasons might be because the human neck muscles have only limited bandwidth [32]. Furthermore, depth axis movements have thought to be the most difficult [18, 46, 47].

Chen et al. [27] compared the joystick-based input devices and head-controlled steering metaphors in navigation tasks in a CAVE VR system. In this system, users are able to move in the virtual world through body motions (e.g., step forward to move forward in the VE). Chen et al. have suggested that the head-controlled paradigms led to improved user performance, a greater sense of immersion, and a lower occurrence of cybersickness. However, their work is not explicitly aimed at exploring the head translational movements, but its focus is to compare two navigation techniques using other body parts. Other researchers have also explored directional movement for navigation [62, 63] and menu selection [64], but they have not focused on depth dimension movements. Recent work [60] explored the design space of head gestures and found through a user elicitation study that stretching the neck forward and backward could be useful for zooming in and out commands.

Apart from head motions, researchers have investigated depth dimension movements using the hand to a limited extent. In both [46] (with a handheld tracker) and [47] (device-free), the researchers have used hand movements to control a 3D cursor. They have argued that moving forward and backward (in the depth dimension) is slower than moving laterally for target selection. Grossman and Balakrishnan [48] have proposed four ways to help users select overlapping targets in 3D volumetric displays. Their Depth Ray technique, which allows users to select targets by controlling how the depth marker is shown, performs particularly well. Our review seems to show that hand depth movements offer some benefits (e.g., selecting occluded targets) but might exhibit slow performance. In contrast, as argued in [36], head movements could be faster than hand gestures in some cases and, as such, if we use head motions, it may be possible to improve performance and user experience.

In this research, rather than considering only the head depth movements using the neck, we also allow users to make use of their whole body to "help" the tracked HWD move towards the depth dimension. That is, as long as the HWD is moving along the depth dimension, irrespective of whether the user is using the neck or the whole body, this movement is considered as the depth movement (DepthMove). We hypothesized that this could potentially overcome the limitations of neck motions alone, thereby increasing the bandwidth of the depth dimension movements and making these motions more comfortable to perform. In addition, we argue that DepthMove could offer an extra dimension of interaction, which could lead to new possibilities for interacting with VR systems.

2.3 Perception of Depth in VR HWD

Perceiving depth information can be challenging in a 3D environment [18]. In order to interact with objects in different depths, it is important to provide proper visual cues to enhance a viewer's depth perception, which past literature (cf. [1, 6, 7, 8, 9, 10, 11]) have explored.

In addition to providing depth cues for visible objects, some techniques have the potential to help the user infer depth information of fully occluded targets [19, 20]. One way is to project the shadow of the objects on a mesh to provide depth indication [21, 22]. However, these projected shadows take up some extra space in the user's view to enhance the user's perception. Another way is to make the targets semi-transparent [23], but the opacity (or transparency) of these objects could be difficult to determine. Also, this method could be ineffective when many targets are occluded by each other.

Teather and colleagues have conducted a series of studies on pointing targets in 3D environments with different depths [12, 13, 14, 15, 16, 17] and have provided evidence that selecting targets located in different depth leads to different selection difficulties because targets located far away are visually smaller—this is also well-supported by Fitts' law [3]. They also found that by providing different depth cues (e.g., texture, highlighting, "support cylinders" as in [13]), the performance could be different. They argue that researchers should carefully consider the use of depth cues when conducting selection experiments.

Since the aim of this paper is to evaluate the potential of using DepthMove, instead of exploring proper aids to support movements along the depth dimension, depth cues are carefully selected and controlled in our user studies—we describe this later in the paper.

3 DEPTHMOVE

From the user's point of view, the basic directions that he or she can move are up/down, left/right, and forward/backward. Movements in the forward/backward direction are described as depth dimension movements. Specifically, in this paper, we define the movements perpendicular to the VR HWD screen as *DepthMove*.

To provide visual aids for DepthMove, we have included a *depth cursor* into the virtual environment (VE). The user needs first to calibrate the initial HWD position when starting DepthMove to let the program log the current head position. The distance between the depth cursor and the user is set to a pre-defined constant value in this process. After this, when a user performs DepthMove, the cursor goes deeper into (or moves back from) the VE (see Fig. 2a). When the user turns, the cursor also follows the user's rotation but does not go into the VE further or away from it (see Fig. 2b). Movements perpendicular to the depth dimension do not affect the depth of the cursor. Overall, it looks like the cursor is glued to the center of the user's view. When the user is performing DepthMove, the cursor goes forward or backward according to the user's movement. The cursor's gain ratio (i.e., the relative movement of the tracked HWD position to the movement of the depth cursor) can be easily changed.

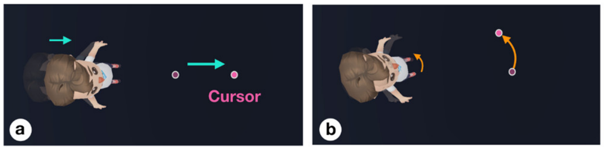

Fig. 2. An illustration of the depth cursor movement relative to the user's movement. (a) When a user performs forward DepthMove, the cursor goes deeper into the VE; (b) when the user turns, the cursor also follows the user's rotation but does not go further into (or away from) the VE.

4 STUDY 1 – 1D MOVEMENT TASK IN VR HMD

In this first study, we investigated DepthMove through a 1D task; in other words, only the movements parallel to the z-axis were examined. We were interested in assessing the potential of using DepthMove in fully immersive VR environments and how it could be used for interaction scenarios. We also aimed to explore the basic features of DepthMove and determine the movement range that could be used in real applications. Specifically, the experiment was designed to achieve five aims:

A1. Modeling the movement time of DepthMove.
A2. Examining the movement trajectories of DepthMove.
A3. Evaluating the general comfort level, motion sickness, and social acceptability of DepthMove.
A4. Finding the comfortable movement range and desired target width of DepthMove with relatively shorter movement time, higher accuracy, and better subjective responses.
A5. Identifying how to use DepthMove for interactions in 3D VE.

4.1 Participants

We recruited 18 participants (5 females, 13 males) between the ages of 17 to 28 years (mean = 21.1) from a local university campus. According to the pre-experiment questionnaire, all of them had at least some VR experience before. They all had no problems moving their head and body back and forth and had normal or corrected-to-normal vision.

4.2 Apparatus and Materials

The experiment was conducted on an Intel Core i7 processor PC with a dedicated NVIDIA GTX 1080 Ti graphics card. The software was written in C# using the Unity3D platform. The Oculus RIFT CV1, a commercial HWD VR device, was used to immerse the user into the 3D VE. A pair of Oculus Touch controllers were used to allow the users to interact with the VE wirelessly.

In this study, the participants were asked to use the index trigger of the Oculus Touch to confirm the selection and use another two buttons to adjust the calibration and to proceed from one scene to another (more on this later).

4.3 Task, Experimental Environment and Setting

The experimental task was designed as a series of 1D selection tasks between two target pairs [1, 2, 3] to assess DepthMove. In this way, we were able to both identify the comfortable depth movement range and desired target width. The data would also allow us to model movements along the depth dimension.

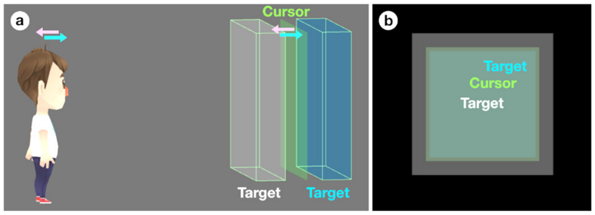

Fig. 3. (a) A possible placement of the targets which was abandoned; (b) two targets and the cursor are occluded with each other from the user's perspective.

As mentioned in Section 2.3, to assess depth movements in 1D, depth cues must be carefully selected and provided. Because our goal was to evaluate DepthMove, the visual cues should be clear enough so that they would not affect the movement performance. The most intuitive way was to place the targets in different depths, and the user would then control the cursor plane to move along the depth dimension and select a target (see Fig. 3a). In this case, the two targets would occlude each other at some point (see Fig. 3b). We could either (1) project the shadow of the targets (like in [21, 22]) or (2) make the targets semi-transparent (as in [23]) to provide the depth cue information visually. However, both approaches may not be useful because far away targets would remain small, thus making them difficult to select and potentially leading to longer selection time and higher errors [3]. According to our pilot studies, we found that the size difference of the targets could influence the study results. To ensure that targets placed in different depths look visually the same, we decided to place the targets on a single flat plane with the same depth. A user's depth motion was mapped to the left and right motion of the selection cursor on the same flat plane.

As shown in Fig. 4a, two identical vertical bars with variable width (W) and fixed height were displayed as targets. The distance between the two targets was set as the other variable distance D. The cursor was a thin vertical line with the same height as the targets. There was a calibration clock placed above one target (see Fig. 4b). It was designed to ensure that the user did not deviate from the effective direction of depth movements (the Z-axis). The clock pointer would turn red when the deviation exceeds the threshold α (in this case it was 10°). All objects were placed on the same vertical 2D plane, which was 10 meters away from the user's position. All objects were placed within the field of view (FoV) of the Oculus RIFT CV1 when the user was looking forward towards the front direction. The background color of the VE was black to maximize contrast.

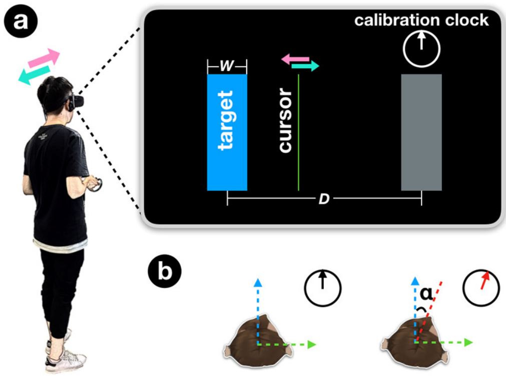

Fig. 4. (a) An illustration of a participant performing the selection task in an HWD-based VE; (b) the calibration clock tells users the forward (effective) direction of DepthMove. The clock pointer turns red when the deviation exceeds the threshold α.

A participant would perform a sequence of eight reciprocal taps to complete one test condition (a single D and W combination). By moving in the depth dimension backward and forward, the participant was able to control the cursor (here in this 1D task, the depth cursor was a straight line) to move rightward and leftward. The gain ratio G [29, 50] of the moving cursor was set to the fixed value of 18. That is, if the head position moved N along the depth dimension, the cursor would move $18N$ towards the left direction. The participant was asked to move the cursor from the starting target (in grey) to the destination target (in blue) and then tapped the index trigger when the cursor was on the destination target to confirm the current selection. When tapped, the destination target would turn green for a brief time to indicate a correct tap or red to indicate an incorrect tap. When the trial was complete, the destination target would turn grey and the starting target blue, which would then become the new destination target. After eight reciprocal taps, the participants would need to control the cursor to move it back to the start to activate the next test condition. Participants were required to complete the whole task standing in a comfortable and natural position and were allowed to rest at any point if they felt tired. A pilot study showed that participants could quickly get used to this visual mapping.

4.4 Design and Procedure

The study used a 9×3 within-subjects design with two factors: target depth D (Forward: [F6-1]; Backward: [B1-3]; see Table 1) and target width W (0.033m [$W0$], 0.017m [$W1$], 0.011m [$W2$]). The order of depth and width was counterbalanced using the Latin Square approach. For a given D, all three W conditions were presented consecutively. In total, we had 18 (participants) × 9 (target depths) × 3 (target widths) × 8 (repetitions) = 3888 trials.

Table 1. The target depth coding.

	Forward						Backward		
Code	F6	F5	F4	F3	F2	F1	B1	B2	B3
Dis (m)	.200	.167	.133	.100	.067	.033	.033	.067	.100

The experiment lasted approximately 20 minutes for each participant. After filling a pre-experiment questionnaire to gather demographic information, participants were introduced to the apparatus and the selection task. They were asked to complete the task as quickly as possible, with a goal of 90% of accuracy. Next, they were invited to wear the devices and calibrate their head position. We instructed them to stand straight in a comfortable posture and press Button A of the Oculus Touch to set the current head position as the start position (in the start position, the cursor was placed in the middle of the starting target). After this phase, any DepthMove motion that departed from the start position would be reflected by the corresponding cursor movement. Participants could adjust their positions and calibrate again, or they could proceed to the next scene by pressing Button B. They then practiced 6 randomly chosen (D and W) trials. After these initial stages, the participants had to move the cursor back to the start position and pressed Button B to proceed to carry out the selection tasks. They were allowed to take a break anytime they wanted, but none of them did so in the actual experiment.

The experiment ended with participants completing several post-experiment questionnaires to assess the general comfort level, motion sickness, and social acceptability of DepthMove as an interactive method for VR HMD. The questionnaires employed were:

- *Comfort Rating Questionnaire*. It quantifies how the users' comfort levels vary across 9 different depths used in this study with 5-point Likert scale questions.
- *Simulator Sickness Questionnaire (SSQ)* [4]. This is used for assessing simulator sickness caused by DepthMove based on 16 questions rated on 4-point scales. Three distinct symptom clusters (Nausea, Oculomotor, and Disorientation) and the Total Severity scores can be computed from the scales.
- *Social Acceptability Questionnaire*. We used a similar format as [5]. It assesses participants' overall emotion/impression during the task through a 5-point Likert scale (Q1), in front of whom (Q2) and in what locations (Q3) he/she would like or want to use DepthMove, and he/she would accept to move among which of the 9 different depths (Q4).

Participants were also welcomed to leave free-text feedback about the whole experiment.

4.5 Results

The data were first analyzed using repeated-measures ANOVAs (RM-ANOVA) on movement time and error rate. Bonferroni corrections were applied for pair-wise comparisons, and we adjusted degrees of freedom with a Greenhouse-Geisser correction for violation of sphericity. We then averaged the data over all participants [1, 28, 49] in order to conduct a linear regression analysis, which resulted in 27 values for each of the 27 D-W pairs. Movement trajectory was analyzed through plots. We examined the subjective response and motion sickness through the data gathered from the questionnaires.

4.5.1 Movement Time and Error Rate

The RM-ANOVA yielded a significant main effect of W ($F_{1.313,\,22.326} = 37.598$, $p < .001$, $\eta_p^2 = .689$) and D ($F_{5.535,\,94.092} = 40.370$, $p < .001$, $\eta_p^2 = .704$) on movement time. The $W \times D$ interaction effect were at the margin of being statistically significant ($F_{6.414,\,109.038} = 2.018$, $p = .065$, $\eta_p^2 = .106$). According to the pairwise comparisons (see Fig. 5a), we found that, except B1, F2, B2, all other target depths D took participants significantly longer movement time to reach comparing to F1. We also found that participants performed significantly faster with W0 than W1 and W2 (see Fig. 5b). As suggested by Fitts' law [3], longer movement distance and smaller target width would lead to longer movement time, as reflected by the above results.

We found a significant main effect of W ($F_{1.654,\ 28.119}$ = 34.599, p < .001, η_p^2 = .671) and D ($F_{4.555,\ 77.436}$ = 3.065, p = .017, η_p^2= .153) on error rate. There was no interaction effect between W and D ($F_{6.798,\ 115.571}$ = 1.069, p = .387, η_p^2 = .059). According to pairwise comparisons, when compared to F1, no other target depths D had significantly higher error rate. Fig. 5c showed that W0 had a significantly lower error rate than W1 and W2.

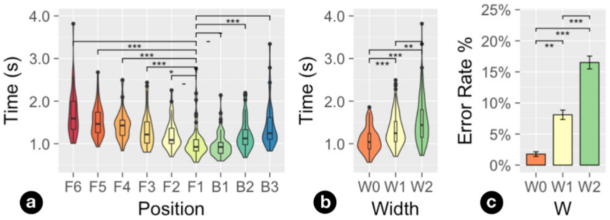

Fig. 5. Plots of movement time with different *D* (a) and *W* (b). Bar plot of error rate regarding *W* (c); error bars indicate the standard error. In (a), statistical significances are marked according to target depth F1; while in (b-c), statistical significance are marked across all conditions (- = p > .05, * = p < .05, ** = p < .01, and *** = p < .001).

4.5.2 Modeling Movement Time

According to Fitts' law [3, 28, 29], it was not difficult to understand the change of movement time by varying D and W. We used *Soukoreff and MacKenzie's* [28] formulation, where movement time (MT) would depend on the ratio of movement amplitude (A) and target width (W),

$$MT = a + b \log(A/W + 1) \qquad (1)$$

Results showed that the model fit the data well with $R^2 = 0.955$ and RMSE (root mean square error) = 0.061. The corresponding a and b of the equation were 0.482 and 0.327.

4.5.3 Movement Trajectory

The DepthMove trajectories were recorded with intervals of 0.1 seconds. According to Fig. 6, participants tended to move a little downwards when performing the forward DepthMove. Also, regardless of the distances they had to move, participants tended to move fast between two targets and reached a peak at the center of the interval. They slowed down after reaching the target.

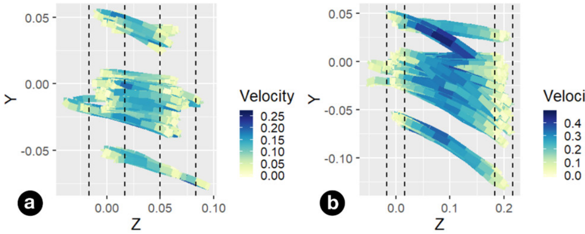

Fig. 6. Trajectory plots with (a) W=W0, D=F2; (b) W=W0, D=F6.

4.5.4 Subjective Response and Motion Sickness

According to the results from the social acceptability questionnaire, participants' overall feelings (Q1) during the task were rated 3.78 out of 5 on average. Most of them were quite positive towards DepthMove; only one participant gave a low rating of 2. Most participants were willing to perform DepthMove alone or in front of people familiar to them (Q2, see Fig. 7a). They preferred private spaces (such as their home and workplace) rather than public areas when performing DepthMove (Q3, see Fig. 7b). They preferred (above 75% acceptance rate) short-range DepthMove within the range of F2, F1, and B1 (Q4, Fig. 7c). In addition, short-range DepthMove within F2, F1, and B1 were also more comfortable according to the results from the comfort rating questionnaire (ratings above 4; Fig. 7d).

With respect to the results from SSQ [4], the average scores of Nausea (N), Oculomotor (O), and Disorientation (D) were 16.43, 29.48, and 45.63, respectively. The Total Severity (TS) was 33.25. This was larger than a high-incidence prone aircraft simulator (18.8) used by Kennedy et al. [4]. In addition, the three individual factors were all larger than the given threshold (14.7, 20.0, 12.4). We should note that Kennedy et al. sampled the subjects from a simulator training community, while the subjects from normal population like our participants would give probably higher ratings. Even if this was the case, the results suggested that a series of 216 trials of DepthMove used in this study had caused a certain degree of sickness, especially for disorientation.

Most participants accepted short-range movements but disliked long-range movements. For example, one participant commented that "*I prefer a short burst of movement using my neck, but not long movements which I had to use my full body to do that*." One other participant said that "*I felt I looked stupid when I was moving forward and backward, especially when I was wearing the helmet and I didn't know what others'* [people's] *responses towards me would be*." While some of them felt somewhat awkward, most felt they could use it well. One commented that "*I didn't care about other people's feelings when I was performing the task*".

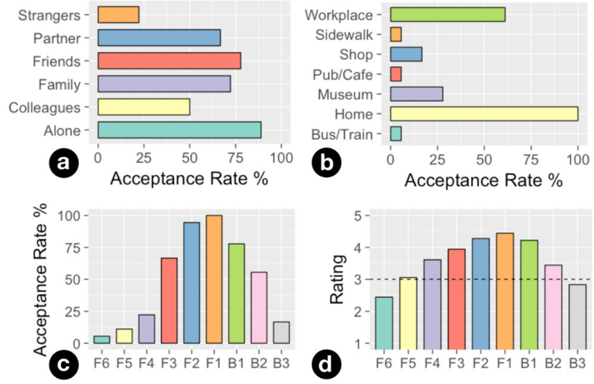

Fig. 7. Acceptance rates for audiences (a), locations (b), and different depths (c) according to results from the social acceptability questionnaire. (d) Average comfort rating of the DepthMove with respect to different depths (dash line indicates neutral responses).

4.6 Discussion

In this section, we summarize the results according to our aims for this experiment [**A1-5**] and discuss each in detail.

The linear regressions analysis was conducted to achieve **A1**. Following prior works [1, 52, 53], we found that the Soukoreff and MacKenzie's formulation could model DepthMove very well (with $R^2 > 0.95$). This suggested that the movement time of DepthMove could be predicted with Fitts' law models.

We drew the trajectory plots to find the general features of DepthMove [**A2**]. Studies of the kinematics of pointing movements revealed that after an initial high-velocity peak, there could be one or more lower velocity, secondary peaks (reacceleration) [21, 54]. However, we did not find any secondary sub-movements with DepthMove. Only one velocity peak was revealed according to the plots (even when varying D and W). We suspected that the human body could not afford the DepthMove to re-accelerate in a short period of time. Moreover, from the trajectory plots, we found that participants could deviate a bit when performing DepthMove without additional visual cues (this was also the reason why we had added the calibration clock). This probably indicates that clear visual cues should be embedded in the real-world use of DepthMove to ensure that the user is moving in the right direction. Otherwise, it can possibly increase body collisions, especially when the user is fully immersed in the VE.

The SSQ results [**A3**] showed that the task caused a certain degree of sickness by reciprocally selecting the two targets in different depths (to be accurate, 216 trails of DepthMove in this study). Therefore, it indicated that DepthMove for long-firing work (such as entering very long texts) might be less feasible. Results from the comfort rating questionnaire and social acceptability questionnaire [**A3**] indicated that users were more accepting of short-range DepthMove (typically within the range of F2, F1, and B1)—short-range movements also were more comfortable than long-range movements. However, according to the comfort rating, a slightly long movement (F5-1, B1-2) would not make them feel uncomfortable. Participants accepted to perform the DepthMove in private places (especially in their home and alone) and in front of people with whom they are familiar—they actually would represent the most common scenarios when using current VR HMD.

According to the experimental results [**A4**], we found that DepthMove between the range of F2, F1, B1, and B2 would be relatively faster—this could also be proved by our Fitts' law model. Also, the error rates among these areas were shown to be similar. By also considering the subjective responses, we were able to determine that the space between F2, F1, B1 as the possible movement range for DepthMove (around 0.067m forward and 0.033m backward from the straight standing pose). In addition, the width of the targets should be as large as possible if space is available, and the selectable objects should not be clustered together.

Based on the above discussion, we identified two ways of using DepthMove for interactions in 3D VE [**A5**]. Because the comfortable movement range seemed a bit short (totally 0.1m) and the accuracy would change significantly when selecting targets with different widths, direct mapping (the user moves *N* along the depth dimension then the cursor moves *gainRatio*×*N*) for DepthMove in large 3D VEs might not be practical, as many objects might be unreachable or hard to select by the depth cursor. Instead, we could

S1. Use continuous-to-discrete target location mapping. For example, the user could move (*0*, *N*] towards the depth dimension to select the first object and (*N*, *2N*] to select the second object in the same pointing direction, no matter what the real distance between the first and second object is in VE. With this mapping, DepthMove also allows users to interact with targets located in 3D VE with ease (when there are not too many objects in the same pointing direction) and even enables interaction with (fully) occluded targets, as demonstrated in Section 7.

S2. Interact with targets located on flat surfaces in 3D VE, such as a flat user interface (UI) placed in the front of the user—these 2D interfaces are still frequently used in 3D VE. This can have potential practical uses since the limited movement range might not be a problem; the interaction effect can be triggered by looking at the UI icon/button and moving towards the depth dimension for a certain range.

The evaluation process of **S1** would be very similar to Study 1 but performing DepthMove along different pointing directions. **S2** could be interesting and useful to explore further. Features such as the size and placement position of the UI icons/buttons are important for 2D flat surfaces. It would also be quite different from Study 1. Therefore, we conducted Study 2 to explore **S2** further.

5 STUDY 2 – INTERACTING WITH FLAT SURFACES

In this second study, we extended the exploration of DepthMove for interacting with flat surfaces through another selection task, which contained the fundamental components of 3D interaction, such as spatial orientation and motion [7]. From the first study, we found DepthMove has the potential for interacting with targets located on flat surfaces in 3D VE, such as a flat user interface (UI) placed in the front of the user. Therefore, this study aimed to investigate what the optimal features of the flat UI (e.g., the size and placement position of the UI icons/buttons) would be when using DepthMove.

We thought of two interaction mechanisms for possible real applications. One was totally hands-free (*Hit*), where the user would only perform DepthMove to interact with targets. The selection would be triggered once the depth cursor collided with the targets. The other mechanism, *Tap*, was for scenarios where users would use devices such as a handheld controller to complement DepthMove. It would require the user to move the depth cursor to the effective range for selecting a target and then click the trigger of the controller to confirm the selection. This could simulate the conditions that DepthMove would be used as an additional input channel to work together with hand-held controllers. We thought these two mechanisms could lead to quite different results because of the dissimilar control mechanisms. In addition, since we were not going to compare these two mechanisms (because they were meant for different use), the two mechanisms were evaluated in separate two phrases with a break in between that was long enough for participants to rest before proceeding. In this study, we aimed to:

A6. Explore the optimal flat UI settings that can leverage DepthMove, including target locations, target sizes, and depth cursor gain ratios for both *Tap* and *Hit* mechanisms.

5.1 Participants, Apparatus, and Materials

Another 18 new participants (2 females, 16 males) between the ages of 18 to 27 years (mean = 21.1) were recruited from the same local university campus. According to the pre-experiment questionnaire, five of them had no prior VR experience, while the others had some limited experience. All participants volunteered for this experiment and had no problem moving their head and body back and forth. They had normal or corrected-to-normal vision. We used the same apparatus and control mechanisms as in the previous study.

5.2 Task, Experimental Environment and Setting

The experimental task was designed as a 3D selection task to explore the optimal features of flat UI using DepthMove. As shown in Fig. 9a, 24 same-size targets (in blue color) were placed in front of the participants. They were arranged in 8 compass directions with 3 layers. All targets were located on the same sphere. To perform the tasks, the participants needed to control the depth cursor (in white) to select the highlighted target (in pink) starting from a fixed location (we called it the center location)—this ensured that participants performed the same distance of depth movements to reach all targets in three different layers. An example task is shown in Fig. 9b. For *Tap*, the participants had to move the depth cursor to the target's effective range (the target would turn yellow in this case) and then press the trigger on the controller to select the target. Pressing the trigger when the depth cursor was not within the effective range would cause a selection error. For *Hit*, the selection would happen when the depth cursor collided with the target object. A selection error would occur when the depth of the cursor exceeded the depth of the target but without hitting it. A short sound would be provided to indicate a successful selection for both mechanisms. The participants had to move the cursor back to the central location and clicked the trigger to start the next trial. Mis-clicking on the center location target would not lead to the next trial.

We varied the target angular sizes (α in Fig. 9a), the angular distances between targets and center (θ in Fig. 9a), target directions (8 compass directions), and cursor gain ratios in this experiment. Note that the effective range of the target was changed according to different cursor gain ratios to make sure the real movement depth would be the same when varying the gain ratio. This led to the same difficulty level for selecting targets with different cursor gains.

We provided two major depth cues in this task. The depth cursor would become smaller when performing forward DepthMove and become larger when performing backward DepthMove—this would provide intuitive real-time feedback for users to know that they were performing DepthMove. Visual highlighting was provided when the cursor was within the effective range of the target [13].

Other settings for this experiment were as follows. The width of the target (effective range) for *Tap* was set to be a constant value 0.0167m in real movement depth (see Fig. 9b). The width of the center location was set to be a constant value 0.067m in real depth movements to make sure the participant started from the same location. The radius of the whole sphere where all targets were located on was set 5m in the virtual space. The central location was set to be 4m away from the participants in the virtual space. The size of the cursor was fixed (with a radius of 0.2m in the virtual space).

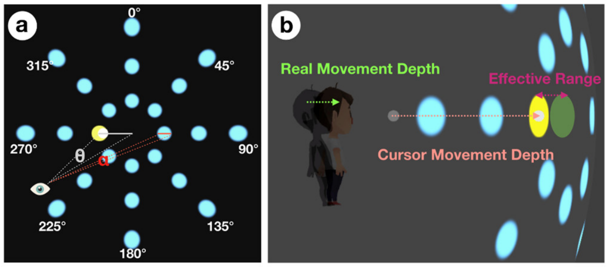

Fig. 9. (a) From a participant's eyes, the angle between targets and the center is marked as θ, and the target size is marked as an angle α. The target selected by DepthMove is highlighted in yellow. (b) An example of a task (Note that only *Tap* has an effective range).

5.3 Procedure and Design

The study evaluated two input mechanisms (*Tap* and *Hit*) separately. For each mechanism, we employed a 3×3×3×8 within-subjects design with four factors: *angular distance* between targets and the center (in short, target distance) T_{Dis} (θ = 5°, 10°, 15°), *target size* T_S (θ = 1°, 1.5°, 2°), *gain ratio* G (30, 60, 90), and *target direction* T_{Dir} (0°, 45°, 90° 135°, 180°, 225°, 270°, 315°, 360°). These parameters were determined by our pilot study, which aimed to reveal the trends of a parameter's influence on interaction performance but did not give too much workload to each participant. Note that with a fixed R (1m), the actual depth movement distance would be 0.033m, 0.017m, and 0.011m, which corresponded to each G (30, 60, 90). The order of the T_S and G were counterbalanced using the Latin Square approach. T_{Dis} and T_{Dir} were selected randomly (without replacement) for each T_S and G combination. In total, the system recorded 18 (participants) × 2 (mechanisms) × 8 (target directions) × 3 (target sizes) × 3 (gain ratios) × 3 (target distances) = 7776 trials.

The whole study lasted about an hour for each participant. We first invited the participants to fill in a pre-experiment questionnaire to collect their demographic information. They were introduced then to the apparatus, the selection task, and the two selection mechanisms (*Tap* and *Hit*). We instructed them to complete the selection task as quickly as possible with a goal of reaching ~90% accuracy. After, they wore the VR HWD and started to calibrate their head position (similar to the first experiment). They then practiced 12 randomly selected trails and moved onto the actual experiment with *Tap* as the selection mechanism (phase 1). After finishing phase 1, they took a rest for at least ten minutes and until they felt no discomfort. They then started with *Hit* as the selection mechanism (phase 2). After each phase, they were required to fill in the NASA-TLX questionnaire [65] to evaluate their level of workload. Participants were welcomed to leave free-text feedback about the whole experiment. During the experiment, we required them to take a break after finishing 72 trials.

5.4 Results

The data were analyzed using RM-ANOVA on movement time and error rate for each mechanism. We applied Bonferroni corrections for pair-wise comparisons and adjusted degrees of freedom with a Greenhouse-Geisser correction for violation of sphericity. To make this section easier to read, we summarized the detailed RM-ANOVA test results in the Appendix section (before the References section).

For **Tap**, RM-ANOVA yielded a significant main effect of T_{Dis}, T_S, and T_{Dir} on movement time, but not G. When T_{Dis} decreased, and T_S increased, the movement time tended to be shorter. However, pairwise comparisons showed that only T_S = 1° and = 2° had significant difference with respect to movement time (p = .020) but not T_S = 1° and = 1.5° (p=.137) nor T_S = 1.5° and = 2° (p=1.000). T_{Dir} = 90° was the fastest but was not shown to have significant difference with T_{Dir} = 45° (p = .064), 270° (p = 1.000) and 315° (p = .422); while it was considerably faster than all the other directions. There were interaction effects between G × T_{Dis} and T_{Dir} × T_{Dis}. From Fig. 10d, we inferred that when T_{Dis} was small (like 5° and 10°), a larger gain ratio would lead to a shorter movement time; while for larger T_{Dis} (like 15°), an increase of gain ratio did not necessarily reduce the movement time. Fig. 10e shows that when T_{Dis} was larger, the influence of T_{Dir} on movement time became greater. No other interactions were found.

RM-ANOVA showed that T_{Dis} and T_S had a significant main effect on error rate, but not T_{Dir} and G. Pairwise comparisons showed that larger T_S caused a smaller error rate. While T_S = 1° caused a significantly higher error rate than T_S = 1.5° (p = .014) and 2° (p = .010), T_S = 1.5° and 2° showed no significant difference (p = .477). No interaction effect was found.

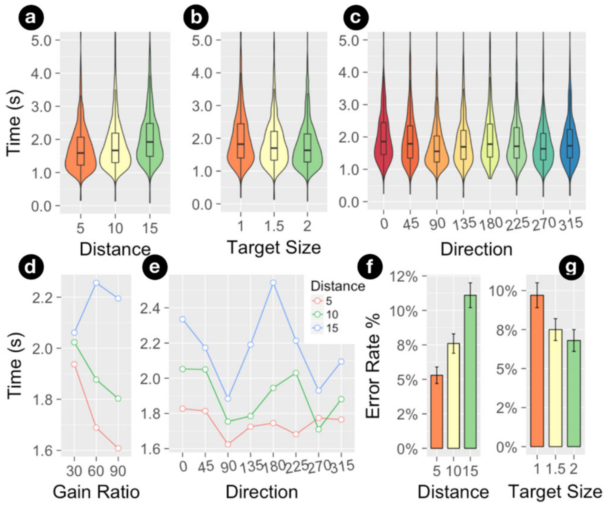

Fig. 10. Significant ANOVA main effects with *Tap*. Plots of movement time in terms of T_{Dis} (a), T_S (b), and T_{Dir} (c). Mean movement time for G × T_{Dis} (d) and T_{Dir} × T_{Dis} (e). Mean error rate in terms of T_{Dis} (f) and T_S (g); error bars indicate the standard error.

For **Hit**, RM-ANOVA indicated a significant main effect of T_{Dis}, T_S, and T_{Dir} on movement time, but not G. Pairwise comparisons showed that larger T_S led to significantly less movement time. While T_{Dis} = 5° induced significantly shorter movement time than T_{Dis} = 10° (p = .006) and 15° (p = .025), T_{Dis} = 10° and 15° showed no significant difference (p = .441). T_{Dir} = 270° was shown to be the fastest; it did not have a significant difference with T_{Dir} = 90° (p = .793), and it was statistically faster than all other directions. There was no interaction effect found across all factors.

RM-ANOVA indicated that T_{Dis}, T_S, T_{Dir}, and G all had a significant main effect on the error rate. Pairwise comparisons suggested that larger G and longer T_{Dis} induced significantly higher error rates. While T_S = 2° caused a significantly lower error rate than T_S =

1° (p = .001) and 1.5° (p = <.001), T_S = 1° and 1.5° showed no significant difference (p = .185). T_{Dir} = 45° and 315° were shown to have the lowest error rate and were significantly lower than all other directions, except T_{Dir} = 0° (p = .360). The ANOVA also showed that $G \times T_{Dis}$, $G \times T_{Dir}$, and $T_{Dis} \times T_{Dir}$ had interaction effects. Fig. 11h shows that when G increased, changing T_{Dis} would have a greater impact on the error rate. A similar effect of G on T_{Dir} was found from Fig. 11i. Fig. 11j shows that when T_{Dis} grew larger, the influence of T_{Dir} on error rate would be greater. No other interactions were found.

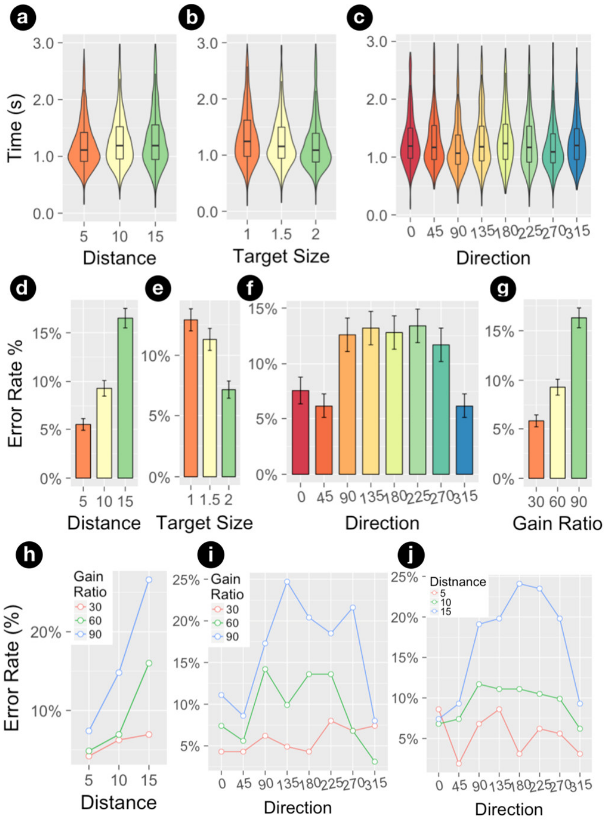

Fig. 11. Significant ANOVA main effects with the ***Hit*** mechanism. Plots of movement time in terms of T_{Dis} (a), T_S (b), and T_{Dir} (c). Mean error regarding T_{Dis} (d), T_S (e), T_{Dir} (f), G (g); error bars indicate the standard error. Mean error rate for $G \times T_{Dis}$ (h), $G \times T_{Dir}$ (i), and $T_{Dir} \times T_{Dis}$ (j).

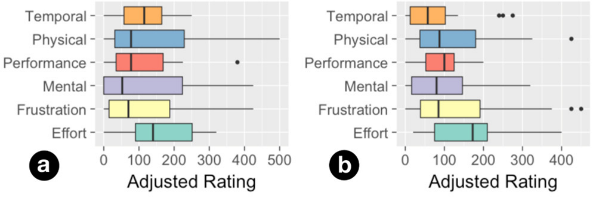

Fig. 12. The boxplot of adjusted rating for the subscales of NASA TLX questionnaire of Tap (a) and Hit (b).

Results from NASA-TLX showed that the weighted rating of *Tap* was 49.53 on average, while *Hit* got 48.73 on average. Friedman test failed to find significant main effect among subscales for both *Tap* ($\chi^2(5)$ = 3.355, p = .645) and *Hit* ($\chi^2(5)$ = 7.213, p = .205). The adjusted rating (= direct rating × weight) for each individual attributed to both techniques is shown in Fig. 12.

5.5 Discussion

In this section, we summarize the findings of the second study.

In terms of **Tap**, both target distance and target size played an important role in movement time and error rate. This was expected as a larger target size and a smaller target distance would intuitively lead to shorter movement time and lower error rate. However, it seems that the target size would not always lead to a significant main effect, as T_S = 1.5° and = 2° did not have significant differences on either selection time or error rate. The results also showed that the movement time was shorter when selecting the targets on the upper and horizontal compass (45°, 90°, 270°, 315°, excluding 0°). The participants tended to select the targets located on the lower part slower. This impact of placement direction would become even stronger when the target distance increased, as indicated by the interaction effect. The gain ratio itself would not induce a significant main effect on movement time and error rate for *Tap*. Nevertheless, it would interact with the target distance factor: when the target distance is short (5° and 10°), increasing the gain ratio would lead to less movement time; while for a longer target distance (15°), a smaller gain ratio might be beneficial.

With respect to **Hit**, the target distance and target size again had a significant effect on movement time and error rate—this was similar to *Tap*. For both factors, the effect would not always be significant: T_{Dis} =10° and 15° performed similarly on movement time; T_S = 1° and 1.5° showed no significant difference on the error rate. *Hit* also tended to favor the upper and the horizontal parts of the compass (as 0° and 90° were faster, 45°, 315°, and 0° were more accurate). In contrast to *Tap*, the gain ratio played an important role for *Hit*—a larger gain ratio tended to cause more errors but would not lead to a significant increase on movement time. This could be explained by the fact that participants would easily miss-tap using a large gain ratio. In addition, the gain ratio had interaction effects with target distance and target direction. A larger gain ratio would make the impact of target distance and direction on error rate to become greater. This shows to us that for a large target distance and for target direction from 90° to 270°, decreasing the gain ratio would be beneficial. Similar to *Tap*, the impact of target placement direction on error rate would become greater when the target distance increased.

6 DEPTHMOVE DESIGN RECOMMENDATIONS

Based on the results of the two studies, we distill the following general recommendations:

R1. Calculate the expected movement time for DepthMove through Fitts' law (e.g., Soukoreff and MacKenzie's formulation [28]).
R2. Provide clear visual cues for the movement direction of DepthMove, or the body deviation could increase collisions and could pose some risks to users who are immersed in VE.
R3. Consider the comfortable movement range of DepthMove at around 0.067m forward and 0.033m backward from the straight standing pose (determined by both experimental performance data and subjective responses).
R4. Minimize prolonged interaction time with DepthMove.
R5. Avoid the need to use DepthMove in public areas.

Below we provide recommendations for the design of flat UI.

R6. If space is available and the selectable objects are not clustered, position the target as close as possible to the center view and as large as possible to ensure fast and accurate DepthMove.
R7. Do not place the target on the lower part of the user's view since it could lead to longer and inaccurate DepthMove, especially when the target distance is relatively far from the center view.
R8. Use DepthMove with a small gain ratio for the target whose distance is relatively far from the center view.
R9. For *Tap*-like mechanisms (DepthMove + trigger confirmation), use DepthMove with a large gain ratio for near targets to shorten the movement time; for *Hit*-like mechanisms (DepthMove only), use a small gain ratio to ensure accurate DepthMove.

7 APPLICATIONS FOR DEPTHMOVE

To ground DepthMove into real applications, we introduce four specific task scenarios to which DepthMove could be applied. We hope that our designs can trigger other future ideas about how to use DepthMove. We further present a simple study to evaluate the effectiveness of DepthMove in completing these tasks when compared with other baseline techniques. We illustrate the benefits of using DepthMove in terms of the different types of interaction.

7.1 Four Task Scenarios

We first demonstrate four task scenarios (based on **S1** and **S2**) to which DepthMove can be applied:

- **T1.** *Screen Switching*. Users sometimes need to handle multiple applications and switch from one display to another [61]. This task requires the user to first interact with the screen icon to evoke the screen list and then switch the currently selected screen (with green outline) to the target screen (as shown in Fig. 13d).
- **T2.** *Volume Scaling*. As shown in Fig. 13d, this task requires the user to interact with the voice icon and then scale the volume display (in blue) in the volume bar to the target volume (in red).
- **T3.** *Target Selection*. The task requires the user to complete the general Fitts circle selection task [28] by continuously selecting the highlighted target (see Fig. 14a).
- **T4.** *Occluded Target Selection*. The task requires the user to select the sphere target, which is fully occluded behind the distractors (as shown in Fig. 14b).

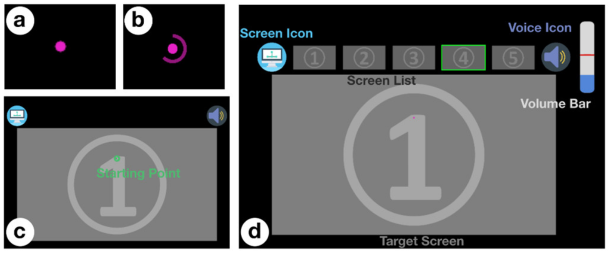

Fig. 13. (a) The appearance of the cursor; (b) the visual indicator of the dwell time for the dwell-based technique; (c) starting scene for T1 and T2; and (d) An introduction of the elements for T1 and T2.

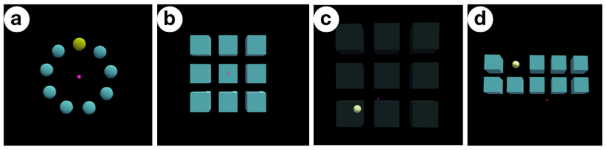

Fig. 14. (a) The target selection task; (b) the occluded target selection task; (c) the user performs a DepthMove to "see" where the target locates; and (d) the Expand technique (the baseline for T4).

7.2 Interaction Mechanisms

In this section, we present the interaction mechanisms we designed for DepthMove to complete the above four tasks. The design was based on the two general 3D VE interaction scenarios [**S1-2**] we discussed in Study 1 and our recommendations for designing DepthMove-type of interfaces. We also present the baseline techniques which are used to compare with DepthMove in the four tasks.

To complete T1 with DepthMove, the user first moves the cursor onto the screen icon (see Fig. 13d) to invoke the screen list and then performs DepthMove to switch between screens while maintaining the cursor on the screen icon. Moving the cursor out of the screen icon would confirm a selection. We used a continuous-to-discrete mapping [**S1**] for screen selection. The movement range was within 0.1m, which followed **R3**. The baseline for this task was the *dwell-based technique*, which required users to focus on a target for a certain period of time (dwell time) in order to make a selection. We set a dwell time of 1s. We were aware that shorter dwell time could lead to a faster speed and applications in text entry usually used a shorter dwell time (like 400ms [34]); however, unlike text entry which requires a continuous input, interacting with UI in daily scenarios using such a small dwell time could easily induce misselections. During our testing stage, we found that the dwell time of 1s could lead to relatively fast input and would not make the novice user feel pushed. To make a selection, the user would first look at the screen icon for a dwell time to open the screen list and then move the cursor onto the target screen in the list for another dwell time.

To carry out T2 with DepthMove, the user would first move the cursor onto the voice icon (see Fig. 13d) to invoke the volume bar and then perform a DepthMove motion to change the volume (volume up when moving deeper). Moving the cursor out of the voice icon would confirm a selection. The movement range design followed **R3**. The dwell-based technique, with a dwell time of 1s was used as a baseline for this task. The user would move the cursor onto the voice icon for a dwell time to evoke the volume bar and then control the cursor on the target volume position (on the volume bar) for another dwell time.

We used *Hit* of DepthMove in Study 2 for T3. The dwell-based technique, with a dwell time of 1s was the baseline. To complete T4 with DepthMove, the user would first perform forward DepthMove (0.033m) to make the depth cursor go "deeper" than the distractors. The distractors were transparent in this case (see Fig. 14c). The user would then move the depth cursor onto the target and click the trigger of the Oculus Touch to make a selection (i.e., Tap). This simulated the condition that DepthMove was used as an additional input channel for selecting fully occluded targets. The baseline technique is the *Expand technique* described in [51]—the user would first click the trigger to reorder the objects on a grid (see Fig. 14d) and then select the target object by pressing the trigger once more.

7.3 Evaluation Study Setup

We recruited another 12 participants (4 females, 8 males) between the age of 18 and 27 years (mean = 21.8) from a local university. They were first introduced to the tasks (T1-T4) and the techniques (DepthMove and the baseline techniques). They then finished a questionnaire to collect their demographic information and proceeded to the four tasks in order. For each task, participants were allowed to practice each technique for as long as they want until they felt they could perform the formal trials. For each trial, the participants needed to click the trigger button of the Oculus Touch on the starting point (Fig. 13c) to be able to start. The participants could carry on the next trial once they finished the current trial correctly; they needed to redo the task if they did it incorrectly, but the time would still be recorded. After the experiment, they were asked to give an overall rating for each technique on the four tasks on a 7-point Likert scale. The whole procedure lasted about 20 minutes.

For each participant, we collected 10 formal trials of T1, 10 of T2, 27 (3 circles) of T3, and another 10 of T4 for both techniques. The order of the techniques was counterbalanced. In total, we gathered 12 (participants) × 2 (techniques) × (10 [T1] + 10 [T2] + 27 [T3] + 10 [T4]) = 1368 timed trials.

7.4 Results and Discussion

In this study, we controlled the errors (participants must finish the trial correctly) and compared the techniques based on the task completion time. We also evaluated subjective feedback.

RM-ANOVA showed that there was a significant main effect of techniques on task completion time for T1 ($F_{1,11}$ = 16.839, p = .002, η_p^2 = .605) and T2 ($F_{1,11}$ = 8.688, p = .013, η_p^2 = .441), but not T3 ($F_{1,11}$ = 2.187, p = .167, η_p^2 = .166) and T4 ($F_{1,11}$ = 4.047, p = .069, η_p^2 = .269). DepthMove was shown to be significantly faster than the baseline technique (dwell-based) for T1 and T2. From Fig. 15a, we can see that the interquartile ranges (IQR) of the baseline technique are much smaller than DepthMove's IQR. Based on our observations, we found that most participants maintained a stable performance using the baseline techniques; when it came to DepthMove, the performance varied among different participants. Some could adapt to DepthMove very quickly and performed in a fast speed, but some others had some slight difficulties using it (this led to longer task completion time and lower ratings). This was due mainly because DepthMove was a new type of interaction and had a learning curve for novice users. Fig. 15a shows that the dwell-based technique led to a lot of outliers in T2, primarily because participants made errors in the first few attempts and corrected them later. From this, we can infer that with a shorter dwell time, the selection time could be shorter but more errors might occur.

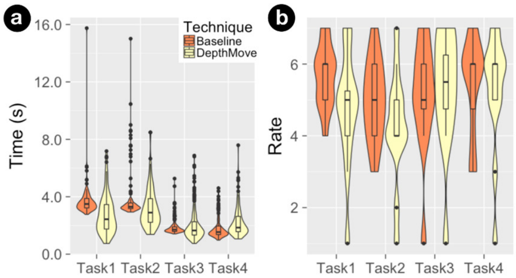

Fig. 15. Plots of task completion time (a) and subjective rating (b) for DepthMove and the baseline techniques across four tasks.

The Wilcoxon signed-rank test showed no significant difference on subjective ratings (Fig. 15b) with DepthMove and the baseline techniques across all tasks (all $p > .1$). Participants described DepthMove as "*exciting*" and "*thrilling*". One commented that "*I enjoy the moment of hitting the target using DepthMove, while the other one* [dwell-based technique] *is boring*". Some found that DepthMove was slightly difficult to use (e.g., "*It was a bit hard for me to control the cursor while performing DepthMove*").

From the results of the study, we have identified some benefits of using DepthMove over the baseline techniques:

B1. DepthMove could be a fast interaction technique for scenarios similar to T1 and T2. Our results show that it is significantly faster than the dwell-based technique with a dwell time of 1 second in these two tasks.

B2. As a hands-free interaction technique, DepthMove allows more proactivity than dwell-based techniques which always "push" the user to move to the next target because of the dwell time. That is, DepthMove allows users to pace themselves which could lead to less stressful interaction.

B3. For the selection of fully occluded targets, DepthMove preserves the original position of the objects while with the Expand technique, users will lose the original context.

B4. For some users, DepthMove might be more interesting to use than the dwell-based technique, as indicated in the comments of participants and our observations.

From the results of the last study, we distill one final recommendation for designing user interfaces with DepthMove:

R10. Depending on the interaction scenario, DepthMove can be the main input method or work to complement another method. As such, a choice can be given to users to select when and how to use DepthMove.

8 LIMITATIONS AND FUTURE WORK

We have identified some limitations of this research and some possibilities for future research. First, our participants were standing during the experiments. Future work could explore other postures like sitting and lying. Second, we did not test the long-term use of DepthMove, which could potentially increase accuracy and performance. Third, we did not explore mobile HWD VR devices (e.g., Samsung Gear VR, and Google Cardboard) which may not have accurate position tracking. In the future, we can apply DepthMove through depth acceleration [66] that can be sensed by the built-in inertial measurement unit (IMU) sensor of smartphones as a cost-efficient alternative. Fourth, we only tested DepthMove in VR, but it can likely be extended to AR and MR systems. Also, future work could explore the possibility of applying DepthMove in various VE. In addition, we are looking forward to seeing more applications and evaluations of DepthMove in various interaction scenarios (like playing a game) either alone as a hands-free technique or as an additional input channel to complement other hand-based mechanisms.

9 CONCLUSION

In this paper, we have presented DepthMove, a new interaction approach for virtual reality (VR) head-worn displays (HWD), either as a hands-free technique or to complement a handheld controller. It allows users to interact with objects by moving the head forward or backward perpendicular to the VR HWD. The first of two studies investigated the feasibility of DepthMove through a 1D selection task. We then modeled user performance and derived the optimal movement range. After that, we proposed two general interaction scenarios for DepthMove. In the second study, we further explored the optimal features of flat 2D interfaces that can work well with DepthMove. From the results, we are able to distill design recommendations for designing DepthMove-based user interfaces. In addition, in a third study, we compared DepthMove with baseline techniques in four task scenarios. The results show that DepthMove has significantly better performance in 2D UI tasks and comparable performance for selection tasks of both occluded and non-occluded targets. All-in-all, DepthMove represents a plausible alternative approach that is efficient and flexible for interacting with content in VR-based virtual environments.

ACKNOWLEDGMENTS

We thank all the participants for the time and the reviewers for their comments and suggestions that have helped improve our paper. This research was supported by AI University Research Centre (AI-URC) at Xi'an Jiaotong-Liverpool University (XJTLU), XJTLU Key Program Special Fund (#KSF-A-03 and #KSF-02), and XJTLU Research Development Fund.

APPENDIX

Summarization of the RM-ANOVA results of Study 2 (adjusted with Greenhouse-Geisser). It shows all significant factors.

Factor	**df**$_{effect}$	**df**$_{error}$	**F**	**p**	η_p^2
Movement time for Tap					
T_{Dis}	1.504	25.569	75.036	<.001	.815
T_S	1.558	26.491	5.694	.013	.251
T_{Dir}	2.951	50.159	7.737	.001	.284
$G \times T_{Dis}$	3.107	52.822	10.183	<.001	.375
$T_{Dir} \times T_{Dis}$	4.598	78.172	2.411	.048	.124
Error Rate for Tap					
T_{Dis}	1.917	32.583	19.427	<.001	.533
T_S	1.887	32.081	5.390	.011	.241
Movement Time for Hit					
T_{Dis}	1.171	19.910	6.327	0.017	.271
T_S	1.647	28.006	17.745	<.001	.511
T_{Dir}	4.070	69.197	4.920	.001	.224
Error Rate for Hit					
T_{Dis}	1.917	32.594	40.298	<.001	.703
T_S	1.660	28.215	12.464	<.001	.423
T_{Dir}	4.636	78.807	5.798	<.001	.254
G	1.697	28.853	18.944	<.001	.527
$G \times T_{Dis}$	2.337	39.726	9.954	<.001	.369
$G \times T_{Dir}$	5.759	97.898	2.253	.047	.117
$T_{Dir} \times T_{Dis}$	7.373	125.171	2.394	.023	.254

REFERENCES

[1] I. Janzen,V. K. Rajendran, and K. S. Booth. "Modeling the impact of depth on pointing performance." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 188-199. ACM, 2016.

[2] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S. Booth. "Two-part models capture the impact of gain on pointing performance." *ACM Transactions on Computer-Human Interaction (TOCHI)* 19, no. 4 (2012): 28.

[3] P. M. Fitts "The information capacity of the human motor system in controlling the amplitude of movement." *Journal of experimental psychology* 47, no. 6 (1954): 381.

[4] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, no. 3 (1993): 203-220.

[5] D. Ahlström, K. Hasan, and P. Irani. "Are you comfortable doing that?: acceptance studies of around-device gestures in and for public settings." In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pp. 193-202. ACM, 2014.

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

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

[8] E. B. Goldstein, and J. Brockmole. *Sensation and perception*. Cengage Learning, 2016.

[9] P. Kellnhofer, P. Didyk, T. Ritschel, B. Masia, K. Myszkowski, and H.-P. Seidel. "Motion parallax in stereo 3D: Model and applications." *ACM Transactions on Graphics (TOG)* 35, no. 6 (2016): 176.

[10] G. Kramida. "Resolving the vergence-accommodation conflict in head-mounted displays." *IEEE Transactions on Visualization & Computer Graphics* 1 (2016): 1-1.

[11] R. J. Teather, and W. Stuerzlinger. "Pointing at 3D targets in a stereo head-tracked virtual environment." In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 87-94. IEEE, 2011.

[12] R. J. Teather and Wolfgang Stuerzlinger. "Pointing at 3d target projections with one-eyed and stereo cursors." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 159-168. ACM, 2013.

[13] R. J. Teather and Wolfgang Stuerzlinger. "Visual aids in 3D point selection experiments." In Proceedings of the 2nd ACM symposium on Spatial user interaction, pp. 127-136. ACM, 2014.

[14] R. J. Teather and Wolfgang Stuerzlinger. "Factors Affecting Mouse-Based 3D Selection in Desktop VR Systems." In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, pp. 10-19. ACM, 2015.

[15] Y. Y. Qian, and R. J. Teather. "The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality." In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI)*, pp. 91-98. ACM, 2017.

[16] R. J. Teather and Wolfgang Stuerzlinger. "Visual aids in 3D point selection experiments." In *Proceedings of the 2nd ACM symposium on Spatial user interaction (SUI)*, pp. 127-136. ACM, 2014.

[17] A. Ramcharitar and R. J. Teather. "EZCursorVR: 2D selection with virtual reality head-mounted displays", *Graphics Interface*. 2018.

[18] J.B.F. Van Erp and A. B. Oving. "Control performance with three translational degrees of freedom." *Human factors* 44, no. 1 (2002): 144-155.

[19] A. Dey, A. Cunningham, and C. Sandor. "Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments." In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology (VRST)*, pp. 211-218. ACM, 2010.

[20] M. J. Liao and W. W. Johnson. "Characterizing the effects of droplines on target acquisition performance on a 3-D perspective display." *Human factors* 46, no. 3 (2004): 476-496.

[21] Y. H. Lee, S. K. Wu, and Y. P. Liu. "Performance of remote target pointing hand movements in a 3D environment." *Human movement science* 32, no. 3 (2013): 511-526.

[22] G. S. Hubona, P. N. Wheeler, G. W. Shirah, and M. Brandt. "The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization." *ACM Transactions on Computer-Human Interaction (TOCHI)* 6, no. 3 (1999): 214-242.

[23] F. M. Marreiros and Ö. Smedby. "Stereoscopic static depth perception of enclosed 3d objects." In *Proceedings of the ACM Symposium on Applied Perception*, pp. 15-22. ACM, 2013.

[24] Y. C. Tung, C. Y. Hsu, H. Y. Wang, S. Chyou, J. W. Lin, P. J. Wu, A. Valstar, and M. Y. Chen. "User-defined game input for smart glasses in public space." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3327-3336. ACM, 2015.

[25] R. G. Radwin, G. C. Vanderheiden, and M. L. Lin. "A method for evaluating head-controlled computer input devices using Fitts' law." *Human factors* 32, no. 4 (1990): 423-438.

[26] E. D. Ragan, R. Kopper, P. Schuchardt, and D. A. Bowman. "Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task." *IEEE transactions on visualization and computer graphics* 19, no. 5 (2013): 886-896.

[27] W. Chen, A. Plancoulaine, N. Férey, D. Touraine, J. Nelson, and P. Bourdot. "6DoF navigation in virtual worlds: comparison of joystick-based and head-controlled paradigms." In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, pp. 111-114. ACM, 2013.

[28] R. W. Soukoreff, and I. S. MacKenzie. "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI." *International journal of human-computer studies* 61, no. 6 (2004): 751-789.

[29] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S. Booth. "Two-part models capture the impact of gain on pointing performance." *ACM Transactions on Computer-Human Interaction (TOCHI)* 19, no. 4 (2012): 28.

[30] M. R. Mine. "Virtual environment interaction techniques." *UNC Chapel Hill CS Dept*. 1995.

[31] E. LoPresti, D. M. Brienza, J. Angelo, L. Gilbertson, and J. Sakai. "Neck range of motion and use of computer head controls." In *Proceedings of the fourth international ACM conference on Assistive technologies*, pp. 121-128. ACM, 2000.

[32] B. Zhang, Y. H. Chen, C. Tuna, A. Dave, Y. Li, E. Lee, and B. Hartmann. "HOBS: head orientation-based selection in physical spaces." In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pp. 17-25. ACM, 2014.

[33] F. Argelaguet and C. Andujar. "A survey of 3D object selection techniques for virtual environments." *Computers & Graphics* 37, no. 3 (2013): 121-136.

[34] C. Yu, Y. Gu, Z. Yang, X. Yi, H. Luo, and Y. Shi. "Tap, dwell or gesture?: Exploring head-based text entry techniques for hmds." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 4479-4488. ACM, 2017.

[35] A. S. Rodrigues, V. K. da Costa, R. C. Cardoso, M. B. Machado, M. B. Machado, and T. A. Tavares. "Evaluation of a Head-Tracking Pointing Device for Users with Motor Disabilities." In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pp. 156-162. ACM, 2017.

[36] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. "Pinpointing: Precise Head-and Eye-Based Target Selection for Augmented Reality." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 81. ACM, 2018.

[37] O. Špakov, and P. Majaranta. "Enhanced gaze interaction using simple head gestures." In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 705-710. ACM, 2012.

[38] R. J. Jacob. "What you look at is what you get: eye movement-based interaction techniques." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 11-18. ACM, 1990.

[39] M. Serrano, B. Ens, X. D. Yang, and P. Irani. "Gluey: Developing a head-worn display interface to unify the interaction experience in dis-

tributed display environments." In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 161-171. ACM, 2015.

[40] T. Jacob, G. Bailly, E. Lecolinet, G. Casiez, and M. Teyssier. "Desktop Orbital Camera Motions Using Rotational Head Movements." In *Proceedings of the 2016 Symposium on Spatial User Interaction*, pp. 139-148. ACM, 2016.

[41] O. Špakov, P. Isokoski, J. Kangas, J. Rantala, D. Akkil, and R. Raisamo. "Comparison of three implementations of HeadTurn: a multimodal interaction technique with gaze and head turns." In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 289-296. ACM, 2016.

[42] D. Rudi, I. Giannopoulos, P. Kiefer, C. Peier, and M. Raubal. "Interacting with Maps on Optical Head-Mounted Displays." In *Proceedings of the 2016 Symposium on Spatial User Interaction*, pp. 3-12. ACM, 2016.

[43] R. M. Clifford, N. M. B. Tuanquin, and R. W. Lindeman. "Jedi ForceExtension: Telekinesis as a Virtual Reality interaction metaphor." In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 239-240. IEEE, 2017.

[44] S. P. Sargunam, K. R. Moghadam, M. Suhail, and E. D. Ragan. "Guided head rotation and amplified head rotation: Evaluating semi-natural travel and viewing techniques in virtual reality." In *IEEE Virtual Reality (VR)*, pp. 19-28. IEEE, 2017.

[45] E. D. Ragan, S. Scerbo, F. Bacim, and D. A. Bowman. "Amplified head rotation in virtual reality and the effects on 3d search, training transfer, and spatial orientation." *IEEE transactions on visualization and computer graphics* 23, no. 8 (2017): 1880-1895.

[46] T. Grossman and R. Balakrishnan. "Pointing at trivariate targets in 3D environments." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 447-454. ACM, 2004.

[47] Y. H. Lee, S. K. Wu, and Y. P. Liu. "Performance of remote target pointing hand movements in a 3D environment." *Human movement science* 32, no. 3 (2013): 511-526.

[48] T. Grossman and R. Balakrishnan. "The design and evaluation of selection techniques for 3D volumetric displays." In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, pp. 3-12. ACM, 2006.

[49] J. Petford, M. A. Nacenta, and C. Gutwin. "Pointing all around you: selection performance of mouse and ray-cast pointing in full-coverage displays." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 533. ACM, 2018.

[50] G. Wilson, M. McGill, M. Jamieson, J. R. Williamson, and S. A. Brewster. "Object Manipulation in Virtual Reality Under Increasing Levels of Translational Gain." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 99. ACM, 2018.

[51] J. Cashion, C. Wingrave, and J. J. LaViola Jr. "Dense and dynamic 3D selection for game-based virtual environments." *IEEE transactions on visualization and computer graphics* 18, no. 4 (2012): 634-642.

[52] J. Gori, O. Rioul, Y. Guiard, and M. Beaudouin-Lafon. "The Perils of Confounding Factors: How Fitts' Law Experiments can Lead to False Conclusions." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 196. ACM, 2018.

[53] I. S. MacKenzie. "Fitts' law as a research and design tool in human-computer interaction." *Human-computer interaction* 7, no. 1 (1992): 91-139.

[54] S. G. Thompson, D. S. McConnell, J. S. Slocum, and M. Bohan. "Kinematic analysis of multiple constraints on a pointing task." *Human movement science* 26, no. 1 (2007): 11-26.

[55] D. A. Bowman, C. J. Rhoton, and M. S. Pinho. "Text input techniques for immersive virtual environments: An empirical comparison." In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, no. 26, pp. 2154-2158. Sage CA: Los Angeles, CA: SAGE Publications, 2002.

[56] A. Pradhan, K. Mehta, and L. Findlater. "Accessibility Came by Accident: Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 459. ACM, 2018.

[57] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. "Text Entry in Immersive Head-Mounted Display-based Virtual Reality using Standard Keyboards." *arXiv preprint arXiv:1802.00626* (2018).

[58] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris. "Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1118-1130. ACM, 2017.

[59] Pupil Labs. HTC Vive Binocular Add-on, 2018. [Online]. Available: https://pupil-labs.com/store/#vr-ar. [Accessed: 27-Feb-2018].

[60] Y. Yan, C. Yu, X, Yi, and Y. Shi. "HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices." In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, no. 4 (2018): 198.

[61] B. M. Ens, R. Finnegan, and P. P. Irani. "The personal cockpit: a spatial interface for effective task switching on head-worn displays." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3171-3180. ACM, 2014.

[62] V. Interrante, B. Ries, and L. Anderson. "Seven league boots: A new metaphor for augmented locomotion through moderately large scale immersive virtual environments." In *2007 IEEE Symposium on 3D User Interfaces*. IEEE, 2007.

[63] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. "Evaluating display fidelity and interaction fidelity in a virtual reality game." *IEEE transactions on visualization and computer graphics* 18, no. 4 (2012): 626-633.

[64] W. Xu, H. N. Liang, Y. Zhao, D. Yu, and D. Monteiro. "DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 444. ACM, 2019.

[65] S. G. Hart and L. E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." In *Advances in psychology*, vol. 52, pp. 139-183. North-Holland, 1988.

[66] X. Lu, D. Yu, H. N. Liang, X. Feng, and W. Xu. "DepthText: Leveraging Head Movements towards the Depth Dimension for Hands-free Text Entry in Mobile Virtual Reality Systems." In *the 26th IEEE Conference on Virtual Reality and 3D User Interfaces*. IEEE 2019.