Interacting with Computers Advance Access published December 7, 2016

doi:10.1093/iwc/iww035

3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays

KASIM ÖZACAR[1,2,*], JUAN DAVID HINCAPIÉ-RAMOS[3], KAZUKI TAKASHIMA[1] AND YOSHIFUMI KITAMURA[1]

[1]*Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan*
[2]*Computer Engineering Department, Karabuk University, Karabuk, Turkey*
[3]*Lenovo R&T, Beijing, China*
[*]*Corresponding author: kozacar@riec.tohoku.ac.jp, ksmzcr@gmail.com*

We conducted a user study evaluating five selection techniques for augmented reality in optical see-through head-mounted displays (OST-HMDs). The techniques we studied aim at supporting mobile usage scenarios where the devices do not need external tracking tools or special environments, and therefore we selected techniques that rely solely on tracking technologies built into conventional commercially available OST-HMDs [i.e. gesture trackers, gaze tracking and inertial measurement units (IMUs)]. While two techniques are based on raycasting using built-in IMU sensing, three techniques are based on a hand-controlled 3D cursor using gestural tracking. We compared these techniques in an experiment with 12 participants. Our results show that raycasting using head orientation (i.e. IMU on the headset) was the fastest, fatigueless and the most preferable technique to select spatially arranged objects. We discuss the implications of our findings for design of interaction techniques in mobile OST-HMDs.

RESEARCH HIGHLIGHTS

- An exploration of 3D selection techniques which rely solely on built-in sensors in OST-HMDs for mobile augmented reality usage scenarios.
- A user study evaluating five techniques for 3D object selection within arm's reach.
- A discussion of the implications of the experimental results for design of selection techniques in mobile OST-HMDs.

Keywords: human computer interaction (HCI); mixed/augmented reality; ubiquitous and mobile devices; interaction techniques; pointing; usability testing

Editorial Board Member: Dyi-Yih Michael Lin

Received 28 April 2016; Revised 11 October 2016; Editorial Decision 12 October 2016; Accepted 18 October 2016

1. INTRODUCTION

Optical see-through head-mounted displays (OST-HMDs) offer augmented reality (AR) experiences by superimposing a virtual information layer on top of the real world. Unlike earlier AR systems which relied on complex infrastructures such as external 3D positional tracking systems, new commercially available AR devices are equipped with built-in control systems such as hand-held controllers, gesture systems and simultaneous location and mapping (SLAM), which allow users to interact with the virtual content in any environments and without external and complex infrastructures. These new built-in tracking capabilities have lowered the cost and complexity of using OST-HMDs, and have the potential to make these devices and applications a reality for everyday usages and activities.

However, despite of potential of the mobile AR devices their input technology is still limited, especially for the wide range of interactions required for spatially located 3D objects. 3D interaction techniques developed for other environments such as 3D desktop displays and immersive 3D systems do not directly translate to OST-HMDs because of limitations in display field of view (FOV), and the quality and tracking capacity of the embedded sensors. Moreover, the research about how the existing 3D selection techniques apply to the new AR devices is very limited. Despite some researchers having explored mid-air selection techniques (Argelaguet and Andujar, 2013; Grossman and Balakrishnan, 2004; Grossman and Balakrishnan, 2006; Vogel and Balakrishnan, 2005), and 3D techniques with the built-in sensors and hand-held controllers of the devices [e.g. GyroWand (Hincapie-Ramos *et al.*, 2015) and Myopoint (Haque *et al.*, 2015)], no research has compared these selection techniques and their relative advantages for OST-HMDs. Our work investigates how different existing 3D interaction techniques apply to the unique circumstances of OST-HMDs in mobile scenario. Particularly important is to understand the capabilities of current built-in display and sensing technologies on these devices, and to determine how they influence selection performance and user preference.

This paper explores how 3D selection techniques compare in mobile OST-HMDs through a user study. The study compares five existing techniques that could work on OST-HMDs in mobile setting with embedded motion and optical sensors: GyroWand (Hincapie-Ramos *et al.*, 2015), Head cursor, gesture-enabled dwell, flick and bimanual selection techniques. GyroWand is a ray-casting technique using the inertial measurement unit (IMU) of a handheld controller. Head cursor is another raycasting technique based on head rotation sensed by a built-in IMU within the OST-HMD. Dwell, flick and bimanual selection are hand-based techniques with different confirmation mechanisms that enable users to control a 3D cursor. We studied how these five techniques perform in simple spatial object selection tasks. While the five techniques do not cover all aspects of spatial object selection, they involve different modalities, degree-of-freedoms (DOFs) and limb movement ranges. We believe that exploratory a comparison offers valuable design insights for interaction techniques for arm's reach object selection in mobile AR settings for OST-HMDs.

In our comparative study, we explored how well users were able to acquire a set of arbitrary objects spatially located within arm's reach. We collected data about the task completion time, accuracy and arm fatigue. We also examined two types of reference frames, body and world, that could characterize the selection techniques and their performances. The body reference enables content to be always at the same location in relation to the user's body, while the world reference enables content to remain fixed in space regardless user's location.

Our work presents the following contributions:

- An exploration of several 3D selection techniques for within-arms reach spatial object selection in mobile OST-HMDs.
- A discussion of the best selection technique for spatial object selection in mobile AR. Our user study shows that head-directed 2D cursor based on raycasting outperforms the other selection techniques in terms of speed and user preference in both reference frames: body and world.
- A discussion of the techniques and their major design characteristics specific for mobile OST-HMDs.

2. RELATED WORK

Latest OST-HMDs require a set of optical and movement sensors aimed at providing spatial tracking of user's actions and input without relying on external infrastructures. In this section, we review these built-in tracking technologies in OST-HMD, their use for spatial selection techniques as well as comparative studies on 3D selection techniques.

2.1. Sensing technologies of OST-HMDs

An important approach to track motions of OST-HMDs relies on using cameras embedded on the device to determine the position of the viewer in relation to the real world (see Microsoft HoloLens, 2015). This approach uses image processing techniques such as feature extraction and matching to build a model of the world. Implementation of these technique can be found on commercial SDKs such as (Meaio, 2015) or Vuforia for Smart EyeWear (Vuforia, 2015); SLAM using depth sensors (RGBDSLAM 2015) and traditional low-end RGB cameras (Ventura *et al.*, 2014). In some cases, the camera information can be enriched with rotational and inertial data as provided by IMUs (Zhou *et al.*, 2014). Besides using cameras for the purpose of positioning the OST-HMD in the real world, the camera can also be utilized to track users' hands gestures and movements for mid-air interactions. The optical tracking capabilities (i.e. camera) of OST-HMDs are nonetheless limited and influence the designs of interaction techniques, which can be realized in OST-HMDs. In the case of raycasting techniques, such limited tracking supports only raycasting from user's head (Tanriverdi and Jacob, 2000) and from the head through the hand (Pierce *et al.*, 1997). Another approach is using motion sensors that are not located on the head piece, instead located on a hand-held controller (Epson, 2015) in which an IMU in the controller allows user to control the ray with the device orientation (Hincapie-Ramos *et al.*, 2015). For OST-HMDs without a hand-held controller, the IMU data could be provided by external wireless hand-held controllers such as wands, rings (Ens, 2016) or even smartphones.

2.2. 3D selection techniques

2.2.1. *Selection with virtual raycasting*

Raycasting has been widely studied for 3D user interfaces in virtual reality (Argelaguet and Andujar, 2013; Bowman *et al.*, 2004;

Olwal and Feiner, 2003). In its default implementation, a virtual ray is casted from a point of origin, usually a tracked hand or hand-held controller, along a given pointing direction. When a trigger event is issued, the ray intersects several virtual objects—typically the object closest to the ray is selected. Raycasting is a technique for distant pointing, which has limitations for object manipulations such as translation and rotation. Further, raycasting often requires disambiguation mechanism among possible multiple targets particularly for densely populated virtual environments.

Raycasting implementations vary in various aspects including how the ray is controlled. Controlling the ray requires a point of origin and a direction. These two values can be provided by tracking the position and orientation of a controller or wand (Grossman and Balakrishnan, 2006; Olwal and Feiner, 2003), the user hands (Bowman *et al.*, 2001; META, 2015), the head (Tanriverdi and Jacob, 2000) or a mix of hands and head (Jota, 2010; Pierce *et al.*, 1997). However, the most common approach relies on tracking a hand-held controller with a button for selection confirmation. Another important variation is the shape of the ray: it can be an actual line in 3D space (Argelaguet and Andujar, 2013) or it can have an aperture angle effectively configuring a cone (Forsberg *et al.*, 1996). Using a cone improves the selection of smaller targets and compensates for hand trembling and jitter, while increasing the need to disambiguate between possible targets.

2.2.2. *Selection with hand-controlled 3D cursor*

Previous studies showed that direct mid-air selection using the hands, in which visual and motor spaces are closely coupled, provides higher performance than indirect or remote selection techniques (Djajadiningrat, 1998; Lemmerman and LaViola, 2007; Wang and MacKenzie, 1999). However, designing mechanisms for selection confirmation is critical since it significantly effects the performance of entire selection task. Therefore, numerous work has considerable attentions on this mechanism using typically physical button or touchpad, or voice command, dwell time and quick hand gestures.

One approach uses a dwell time as the confirmation mechanism, where a click event is recognized if an input device (e.g. cursor) is kept in a steady state for a certain time threshold, even though it introduces a constant lag.

Using quick hand gestures is another approach to confirm a selection. Grossman and Balakrishnan (2006) proposed thumb trigger gesture where thumb finger moves in and out toward the index finger in mid-air object selection in a volumetric display. Vogel and Balakrishnan (2005) compared this with their AirTap technique, which is similar to our index finger movement when clicking a mouse for large-size 2D display interaction. Thumb trigger gesture has advantage of kinesthetic feedback since it still touches the side of hand. According to Wang and MacKenzie (2000), the absence of these physical surfaces significantly degrades performance of object selection. Nevertheless, Vogel and Balakrishnan (2005)'s results showed that there is no significant difference in trial performance time or error rate between these two different click gestures.

2.3. A brief survey of selection techniques

Table 1 summarizes main elements of existing studies of comparisons on 3D selection techniques. These studies, alongside with the one presented in this paper, can help readers understand the tradeoffs of existing selection techniques. Unlike others, our study focuses on the use of the selection techniques in the particular case of mobile OST-HMDs. Table 1 presents our classification based on main characteristics (i.e. display type and input type), performance and usability of selection techniques. As general goals of interaction techniques, in this survey, we focused on speed, accuracy and fatigue as the parameters of effectiveness and efficiency.

Lee *et al.* (2003) compared four selection techniques such as head-directed, hand-directed and head and head-directed raycasting, and image plane techniques for a large 2D display setting. Their results showed that image-plane technique outperformed the others in terms of speed, accuracy and hand fatigue. Besides, the head-directed raycasting performed the worst performance in terms of speed, accuracy and hand fatigue.

Tanriverdi and Jacob (2000) compared eye movement-based 2D cursor, which uses dwell-time selection confirmation for a focused object, to conventional 3D cursor for arm's reach and remote selection tasks. Results showed that eye movement selection was faster than hand-based pointing for only the remote objects, which required to user to move forward to select them. Also, users did not spend extra physical effort to select the objects with eye-movement-based interaction. Ware and Lowther (1997) compared one-eyed image-plane using 2D cursor with the virtual hand using 3D cursor for VR HMD. Image-plane technique was faster than virtual hand. Grossman and Balakrishnan (2006) confirmed that raycasting is faster than virtual hand for volumetric displays, where the objects are within arm's reach.

While a set of comparative studies have been done as shown above; they are generally focused on large 2D display, stereoscopic displays, VR HMDs or volumetric displays. None of the studies focused on arm's reach selection in OST-HMD settings. For recent OST-HMDs, it is therefore important to revisit the primary and fundamental questions to find what technique is the most suitable to select spatial object in arm's reach for speed, accuracy and usability in mobile OST-HMDs scenario.

3. SELECTION TECHNIQUES FOR MOBILE OST-HMDs

We implemented and compared five techniques that can be realized using sensors (image and movement sensors) embedded

Table 1. Summary of the comparison of selection techniques.

Compared techniques	Cursor type	Target range	Display		Evaluation
			Type	Dimension	
Head, hand-directed, head–hand directed ray cursor, and image-plane (Lee *et al.*, 2003)	Ray, 2D cursor	Remote	Large display	2D	*Image-plane*: The fastest, the most accurate, no fatigue *Head–hand directed ray cursor*: The slowest, the least accurate, a lot of fatigue
Eye-directed image-plane, hand-based selection (Tanriverdi and Jacob, 2000)	Ray, 3D cursor	Arm-reach and remote	VR HMD	3D	*Eye-directed image-plane (remote):* The fastest, no fatigue *Eye-directed image plane*: No fatigue
Image-plane, hand-based selection (Ware and Lowther, 1997)	3D cursor, 2D cursor	Remote	Stereoscopic display	3D	*Image-plane*: The fastest
Hand-directed ray, hand-based selection (Grossman and Balakrishnan, 2006)	Ray, 3D cursor	Arm-reach	Volumetric display	3D	*Hand-directed ray* : The fastest

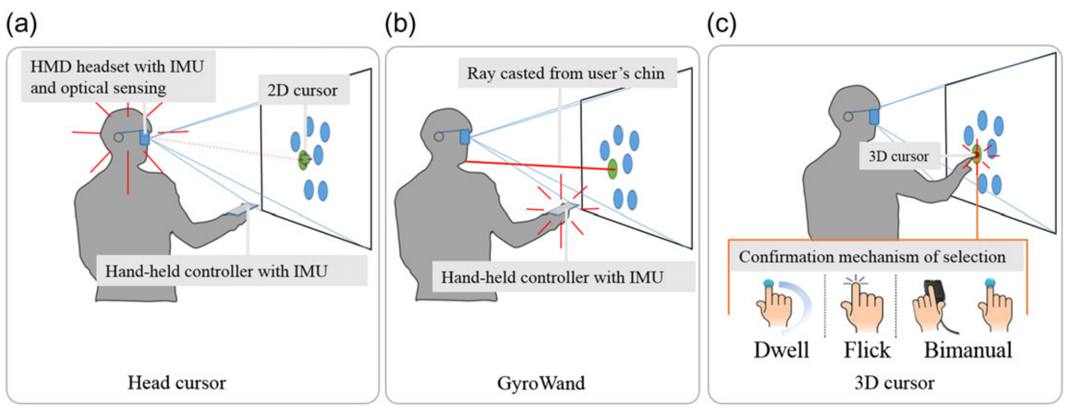

Figure 1. Interaction techniques for object selection within arm's reach on OST-HMD. (**a**) Head cursor: user looks straightforward and stationary cursor appears at the center screen. User rotates his/her head to hover a target with the cursor. The ray direction is controlled using the IMU on the head piece. (**b**) GyroWand: the ray origin is the user's chin. Direction of the ray cursor is controlled using the IMU on the hand-held controller. (**c**) 3D cursor: hand-movement-based techniques using optical sensing with different selection confirmation mechanisms: Dwell, Flick and Bimanual.

in a mobile OST-HMD. Our goal is to understand how efficient they are for object selection within arm's reach in mobile AR settings. To find out suitable interaction techniques, we pick up and explore five potential interaction techniques.

The techniques are selected from two different approaches. GyroWand and Head cursor are ray based, and use movement sensors (IMU) on the controller and head piece of the OST-HMD, respectively (Fig. 1a and b). Other three techniques (Fig. 1c) are based on 3D cursor controlled by the user'(s) hand. This study focuses on the performance and user's preference of the two ready-to-use raycasting, and three hand-controlled 3D cursor techniques. In order to obtain more reliable data than the one form embedded sensors in the HMD devices available today, we used external trackers to emulate these techniques to validate their performances rather than exploring their real-world issues.

GyroWand and Head cursor use only rotational data to select spatial object: GyroWand uses vertical and horizontal rotations of the hand-held controller; Head cursor uses head rotations. All hand-controlled techniques require data about the location of the hand in space as input, enabling users to acquire spatially located objects through a 3D cursor with different selection confirmation mechanisms.

3.1. Raycasting techniques

Raycasting techniques have previously been studied and found to be efficient for spatial object selection. For mobile OST-HMDs, raycasting techniques can be realized by directing a ray originated at a predefined location in virtual space, and use rotational data acquired from an IMU in the user's hand-held device. However, this naive rotational data typically cannot be directly mapped to ray orientation because the rotational data gathered from IMUs has several well-known problems: noise, sensor drift and axis mapping. Therefore, raycasting techniques interpret IMU rotational data using a state machine and compensate for drift/interference (Hincapie-Ramos *et al*., 2015).

3.1.1. Head cursor

Head cursor is a raycasting technique that enables users to interact with the 2D projection of remote objects on the user's viewing plane as shown in Fig. 1a. The object crossed by the cursor is selected by casting an invisible ray from the eye-point. This technique particularly benefits from the recent available IMU sensor on the head piece of OST-HMD. As the user rotates his/her head, the projected cursor moves on 2D plane. It requires a selection confirmation to make a click on a target for which we used a touchpad on the hand-held controller. When a selection is confirmed by tapping on the controller, the crossed virtual object with the closest 2D projection to the cursor is selected.

3.1.2. GyroWand

We used GyroWand (Hincapie-Ramos *et al*., 2015), a state-of-the-art raycasting technique designed for mobile OST-HMDs. GyroWand introduces fundamental design differences from traditional raycasting. GyroWand does not initiate raycasting from the user's hand, but rather from any spatial coordinate (e.g. chin and shoulder). In this study, GyroWand directs a ray originated at user's chin as on-body location in virtual space using hand-held controller's rotational data acquired from built-in IMU as shown in Fig. 1b.

3.2. Hand-controlled 3D point cursor techniques

We also explored hand-controlled techniques using 3D point cursor (i.e. point cursor techniques), since selecting objects with the fingers seem natural and direct way to interact with objects within arm's reach. As we described earlier, our set-up in the experiment used a set of external 3D motion trackers instead of self-contained sensing capabilities as a way to achieve stability and reliability. Nonetheless, we can expect self-contained gestural tracking to be part of future HMDs as showcased in the upcoming devices such as Atheer One (Atheer, 2015) and the META 2.0 devices (META, 2015). We use a 3D motion tracking environment using an external optical tracking system with infrared cameras and reflective markers on the fingers to avoid technical and methodological constraints caused by implementation.

Moreover, because the way of selection confirmation is critical in this approach, we investigated two types quick finger gestures (such as dwell time and flick) to avoid problem of carrying a hand-held controller, and using both hands as manual confirmation with the controller (flick in Fig. 1c middle). We introduce three techniques for selection confirmation as follows.

3.2.1. Dwell

This technique is one of the most commonly used gestures, which uses dwell time interval as confirmation as shown in Fig. 1c, right. Users' hand/finger remains nearly motionless within a certain predetermined small range of space for a short time. This time interval must be properly considered in advance because it might induce unintended commands if the time interval is too short or too long, an issue known as the Midas touch problem (Robert, 1991). Moreover, it can cause delays of interaction and causes discomfort if the interval is too long. Based on previous studies (Muller, 2007; Špakov and Miniotas 2004), an optimal dwell time of 750 ms is recommended for novice users; therefore, we use this dwell time in this work.

3.2.2. Flick

This technique (Fig. 1c, center) consists of down and up movements of the finger. Since the hand is also pointing, we aimed to design this confirmation mechanism in a way that minimizes hand movement side effects. Because of the interconnectedness of tendons in the hand, movement in a finger affects the others. Therefore, we attached a marker to the index fingertip to detect flick gesture, and a 3-point marker on the hand to track finger position by offsetting its virtual position to index fingertip. Velocity (v) and acceleration of the user's finger provide cues to detect the click gestures; therefore, we have built a recognition state machine similar to Vogel and Balakrishnan (2005) using these parameters as shown in Fig. 2, where finger velocity is marked by v, and D is distance moved in 200 ms; D_{down} and D_{cancel} are thresholds to begin a click down and cancel actions, respectively. D and v are measured along axis of finger movement.

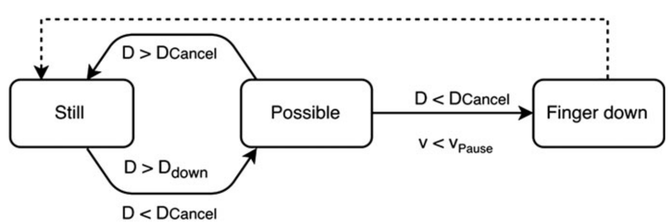

Figure 2. Recognition state machine for flick gesture.

3.2.3. Bimanual: tapping with non-dominant hand

Another possible approach for confirmation is to use dominant hand to move the cursor and the non-dominant hand for confirmation with hand-held device (Fig. 1c, right). A user points at an object with his/her index finger of dominant hand and confirms the selection by pressing a button or tapping a controller with his/her finger of non-dominant hand. Although this approach requires carrying hand-held controller, it is expected to reduce the interaction time because the dwell time is not necessary.

4. EXPERIMENT: SPATIAL 3D OBJECT SELECTION IN MOBILE OST-HMDs

We conduct a controlled user study to explore five techniques that rely solely on tracking technologies built into commercially available OST-HMDs in 3D selection tasks. Our goal of this study is to make an initial assessment of the possible interaction alternatives and that further research is needed to identify the particular aspects in which they differ and the strength of their impact—like the nature of the required movements.

Our experiment aimed at answering the following fundamental questions:

- How fast and accurately can users select a set of arbitrary objects spatially located within arm's reach with interaction techniques?
- How do reference frames affect the task performance?
- How do techniques affect task workload index and user preferences?

We focused on the two reference frames rather than using a fixed content in the user's limited FOV because these reference frames enable users to observe beyond their narrow FOVs. Reference frames also provide users the ability to explore a larger 3D virtual environment that could contain multiple larger objects.

Apparatus: For the mobile AR setting, we used a smart glass, EPSON Moverio BT-200 with a 23° diagonal FOV, a 960 × 540 pixel display resolution and a focal depth of 5 m. Head tracking to simulate reference frames was provided by an external optical tracking system OptiTrack with six cameras. These cameras covered a cubic tracking space of dimensions 1 m (wide) × 1 m (depth) × 2 m (height). The system tracked location and orientation of a participant's smart glasses with a 3-point marker at 60 FPS (frames per second) as user's head movements. That data were transmitted to the smart glasses via UDP over WiFi at a 40 FPS (a 30% reduction due to networking overhead). Participants stood in the center of the tracking cube. The smart glass displayed stereoscopic 3D graphics at 50 FPS. Touch input from users was received through the touchpad of the Moverio BT-200. A participant in the experimental configuration is shown in Fig. 3.

Figure 3. A participant using an OST-HMD with markers for optical 3D motion tracking, having a controller for selection in his hand. Here, he is performing the task with a ray-based technique using IMU.

We implemented techniques for the EPSON Moverio BT-200 in Unity 4.6 for Android 4.0. Tracking data captured by OptiTrack optical tracking systems (six cameras used) were received by a Unity application running on a desktop computer. The data were passed to the Moverio BT-200 via UDP packages using Unity's network capabilities. We used Stereoskopix FOV2GO package to create the stereoscopic rendering in Unity. The package is created by the MxR group at the University of Southern California. All rotational information was calculated using the internal IMU of the Moverio BT-200 hand-held controller and head piece.

Subjects: Twelve participants (four female) volunteered, ages 21–32 (mean = 24.9, SD = 3.56), all right handed. Five participants had previous experience with HMDs, four had experience with virtual reality, three had experience with mid-air interaction and none of them had experience with ray-casting techniques.

Task: Figure 4a and b, respectively, shows objects located in front of the HMD and a pair of a stereoscopic view volume (cube shaped) for target selection tasks. For each session, participants had to select randomly located 16 sphere-shaped targets

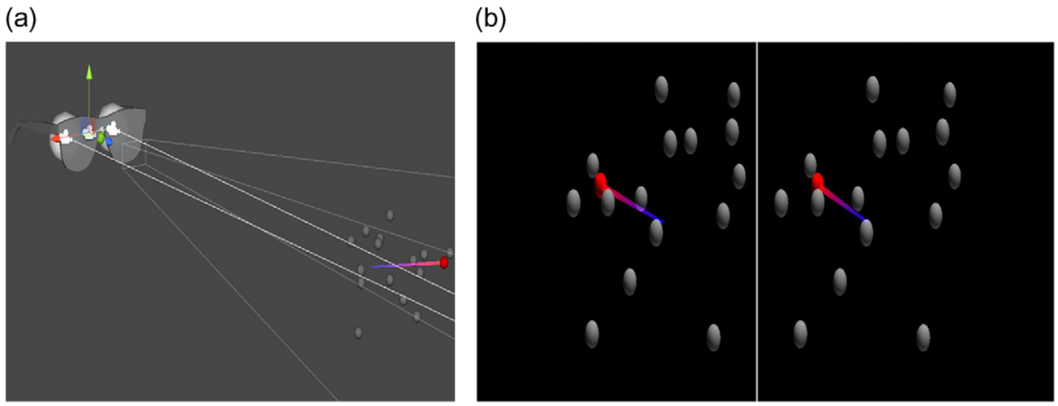

Figure 4. (**a**) Objects of 1 cm width are randomly located 40–65 cm in front of the OST-HMD and (**b**) stereoscopic rendering of the user view.

in the cubic view volume. All targets were initially shown in gray and became highlighted in light green when they encountered a ray or a cursor. The target to be selected next is shown in red. When selection target is crossed by the cursor/ray, its color changes to brighter red. On selection, the system shows the target in magenta and highlights another object in red as the new target. As shown in Fig. 4b, a line colored red-to-blue indicated the location of the next target, guiding the participant when it was out of the FOV. The session ended when the participant completed 16 target selections. Selection error is indicated when a wrong object or no object is selected.

Design: Independent variables were interaction technique and reference frame. We used a 5 × 2 within-subject design to investigate user performance.

We considered five techniques (two raycasting: Head cursor and GyroWand; and three hand-controlled 3D point cursor: Dwell, Bimanual and Flick). We randomly located 3D objects in a virtual cube 50 cm away from the OST-HMD but on different reference frames. While our primary interest is to explore 3D selection techniques for mobile OST-HMD, we were also interested in the effects of reference frame of the technique. Because, for 3D objects within arm's reach, the reference frame makes a difference as objects react differently to user movement. We evaluated two reference frames: body and world (Hincapie *et al.*, 2014). On the body reference frame (see Fig. 5), the content was always at the same location in relation to the body. For example, the content positioned on the body will always be visible when the user turns his head to the right, but not when he looks forward. On the world reference frame (see Fig. 6), the content remains fixed in space irrespective of the user location or gaze direction; we located the objects at the center of the tracking cube and 1.5 m high, participants stood 50 cm away. Participants were asked to hold the controller in their dominant hands (Fig. 3). However, when the confirmation was required in case of using a finger-based technique, they held the controller in their non-dominant hands.

The experimenter demonstrated the task and then asked the participants to train for 10 min with each technique. The

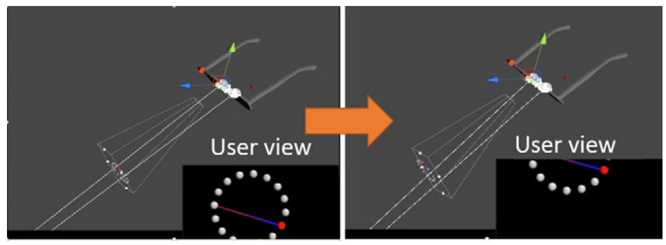

Figure 5. Body reference frame. A user is looking down.

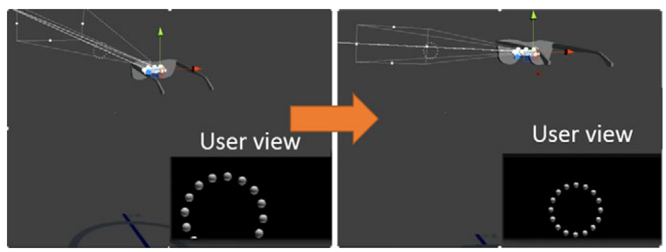

Figure 6. World reference frame. A user is moving away from the targets.

training session consisted of at least 20 selections for each technique. The orders of selection techniques and reference frames were counter-balanced using a Latin-square design. With a total of 5 × 2 = 10 conditions per participant and 16 × 2 = 32 selections per condition, we registered 32 × 10 = 320 selections per participant, 3840 selections in total. All participants completed the experiment in one session lasting ~50 min.

Measures: For each trial, we recorded task completion time, selection errors and cursor arrival times to targets. Completion time is the time elapsed from cursor departure until the object is selected. Selection error is the count that indicates the number of wrong object selections and no object selection. First arrival time is the time elapsed from cursor departure until it first arrives to the target object (before triggering selection), last arrival time is the time elapsed from cursor departure until the last arrival to the target (before triggering selection). A user might arrive at the target object (first arrival), overshoot it, arrive again (last arrival) and then trigger selection. NASA-TLX workload sheets (Hart *et al.*, 1988), which assess subjective task load, and Borg-RPE (Borg, 1998), rating of perceived exertion level, are two widely used measures. Participants filled out Nasa-TLX workload index and Borg-RPE after each condition in total 10 sheets per participant. Participants also ranked each technique after the experiment (see Table 3).

4.1. Results

We removed outliers (5% of the data set) after 3 SD of completion time. We carried out ANOVA and *post hoc* tests with Bonferroni corrections (only for interaction technique) to analyze the measurements. In addition, pooling different error terms would be appropriate if the variance for all the interactions were equal (Winer *et al.*, 1991). For within subject designs, however, the variance is not assumed to be constant over the levels of within subject factors. Therefore, we used unpooled error terms in our study.

Completion time: As shown in Fig. 7a, we found a main effect for interaction technique ($F_{4,44} = 15.161$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.579$). We did not find main effect for the reference frames ($F_{1,44} = 0.049$, $P = 0.829$, $\eta^2_{\text{partial}} = 0.004$), while there was no significant interaction of technique × reference frame. *Post hoc* tests showed significant differences between all techniques ($P < 0.001$) except, between Bimanual and GyroWand ($P = 1.0$), and between Bimanual and Dwell ($P = 1.0$). Head cursor had the fastest completion time at 2.395 s (SD = 0.939), followed by Bimanual, GyroWand and Dwell at 3.429 s (SD = 1.445), 3.458 s (SD = 1.373) and 3.510 s (SD = 1.295), respectively. Flick was the slowest completion time at 4.015 s (SD = 1.572).

Selection errors: As shown in Fig. 7b, we found a main effect for technique ($F_{4,44} = 11.333$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.398$) and an interaction effect between technique × reference frame ($F_{4,44} = 3.369$, $P = 0.017$, $\eta^2_{\text{partial}} = 0.232$). For Head cursor and GyroWand, world reference frame led to higher selection errors than body reference frame ($F_{1,3575} = 20.532$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.006$) and (($F_{1,3575} = 4.871$, $P = 0.027$, $\eta^2_{\text{partial}} = 0.001$)) . Reference frame had no effect for Bimanual ($P = 0.11$), Dwell ($P = 0.098$) and Flick ($P = 0.209$) techniques.

A change in the selection technique explains 57.9% of the change of selection speed (total time $\eta^2_{\text{partial}} = 0.579$) and 39.8% of the change of selection errors (selection error $\eta^2_{\text{partial}} = 0.398$). Moreover, a change in the reference frame explains 0.4% of the change of selection speed (total time $\eta^2_{\text{partial}} = 0.004$) and 29.4% of the change of selection errors (selection error $\eta^2_{\text{partial}} = 0.294$).

First–last arrival times: Completion time alone does not provide information regarding behavior of ray/cursor movements during the selection task. According to the Woodworth's two-component model (Woodworth, 1899), if a target is not acquired with cursor due to either undershoot or overshoot on the first try, then the user performs iterative slow and more

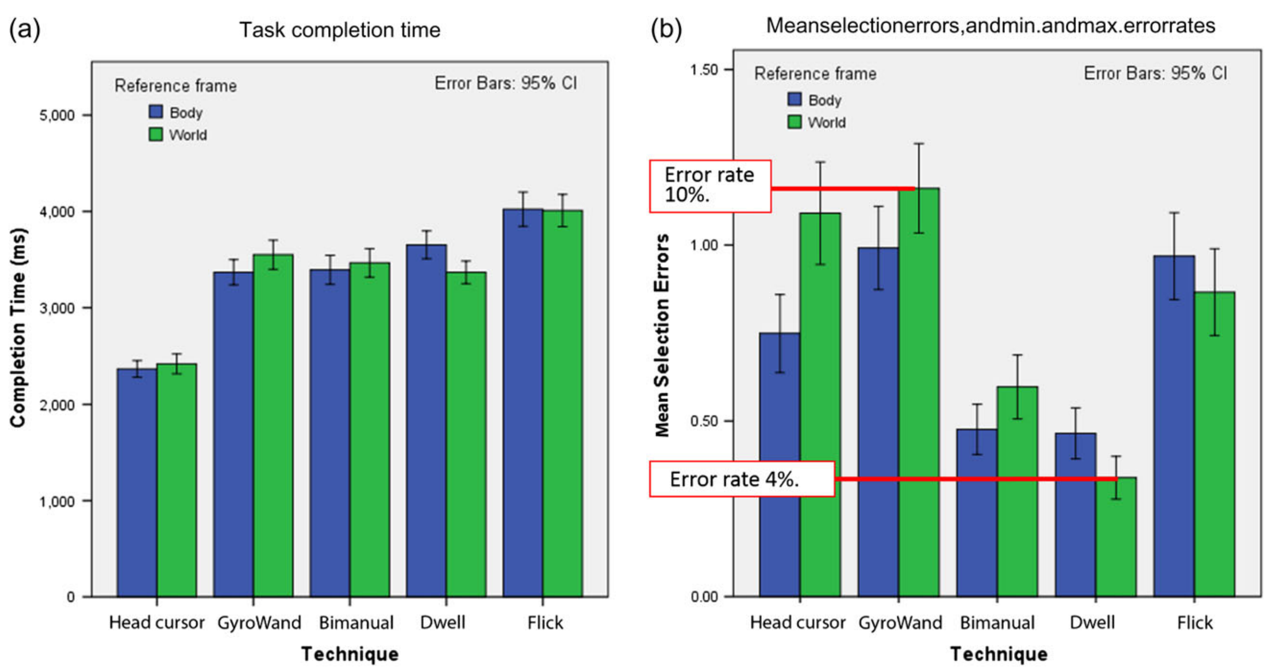

Figure 7. Task completion time and mean selection errors. (**a**) Task completion time and (**b**) mean selection errors, and minimum and maximum error rates.

accurate movements until acquire it, known as corrective phase. The difference between the first and the last arrival times approximates corrective phase in our study. Therefore, we examined this differences to identify behavior of the cursor movements during a selection task for each technique. Figure 8 shows this difference relative to the technique. Results showed a main effect for technique ($F_{4,44} = 9.22$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.456$). *Post hoc* tests showed significant differences between all techniques ($P < 0.001$) except, between Bimanual and Dwell ($P = 0.42$). Bimanual and Dwell were the techniques with the least difference at 0.392 (SD = 0.027) and 0.317 (SD = 0.026), followed by Head cursor at 0.490 (SD = 0.026), GyroWand at 0.762 (SD = 0.027) and Flick at 0.957 (SD = 0.029).

4.1.1. Subjective results

Borg-RPE: Figure 9 shows the results of Borg-RPE. We found a main effect for technique ($F_{4,44} = 25.937$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.702$) and our *post hoc* tests showed significant differences between all techniques ($P < 0.001$) except, between Bimanual and GyroWand ($P = 1.0$), and between Bimanual and Dwell ($P = 1.0$). Head cursor presented smallest physical exertion at 8.50 (SD = 3.35), and followed by GyroWand at 10.792 (SD = 2.55). Results did not show a main effect for reference frame ($P = 0.927$).

Participants' subjective workload assessment and ranking, presented in Tables 2 and 3, explained the quantitative results, clearly showing that participants found the 3D point cursor techniques mentally demanding, tiresome, slow and effortful (see Table 2, shown with asterisk). NASA-TLX shows that mental load, physical load, effort and frustration for ray techniques, Head cursor and GyroWand, are much smaller than for hand-controlled 3D cursor techniques. The main effects of technique on workload index are presented in Table 4.

Assessment of the Head cursor and GyroWand interface was positive with general preference for head cursor (see Table 3). The Head cursor was ranked top by eight participants (*Mean* 1.5), while GyroWand was favored by three participants (*Mean* 2.58). The main effects of technique on ranking are presented in Table 5.

Moreover, some participants stated that they liked the way of controlling the ray with low and small hand movements ('GyroWand is like controlling fancy Lightsaber in Star Wars'), but found it harder to control than Head cursor. The 3D point cursor techniques are physically demanding (Table 2, shown with asterisk); however, provides ease of control (Table 3).

4.2. Discussion

4.2.1. Experimental results

Head cursor outperformed the others in terms of selection speed (mean 2 s). 3D point cursor techniques, except Flick, were more accurate than raycasting techniques. In terms of tradeoff between speed and accuracy, Flick and GyroWand may provide the least benefit since they are both slower and less accurate.

With GyroWand, virtual objects are selected by pointing directly to themselves with a casted ray; however, with Head cursor technique they are selected by pointing to their screen

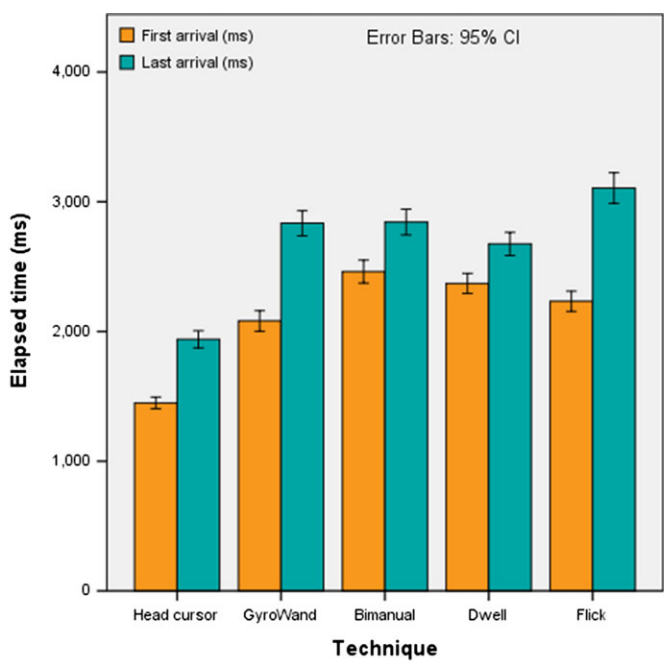

Figure 8. First–last arrival time to target.

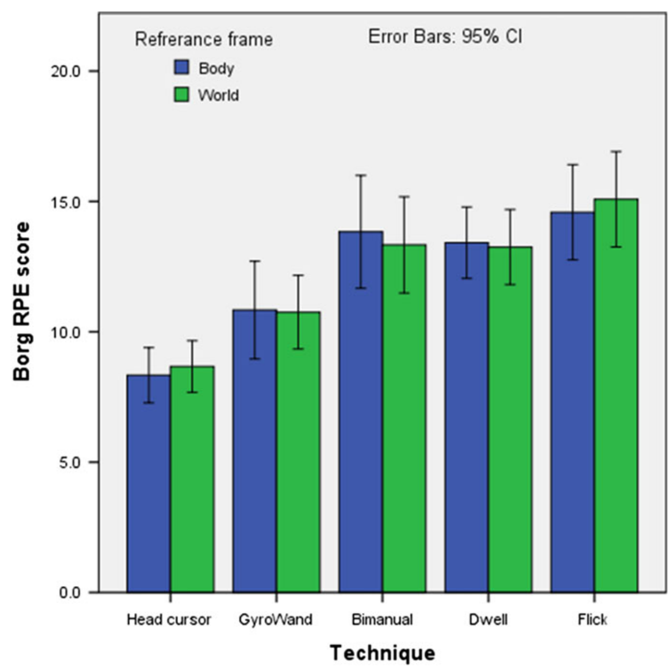

Figure 9. Borg-RPE score.

Table 2. Mean (and standard deviation) subjective preference assessments for each technique.

Nasa-TLX (1–21)	Head cursor	Gyrowand	Dwell	Bimanual	Flick
Mental load (low–high)	4.75 (3.34)	6.79 (4.32)	7.45* (4.2)	7.33* (5.4)	9.04* (5.0)
Physical load (low–high)	4.79 (4.1)	7.08 (4.9)	12.04* (4.6)	13.3* (5.09)	14.29* (4.3)
Temporal load (low–high)	4.29 (2.5)	6.08 (3.41)	7* (3.5)	8.62* (4.37)	10.20* (4.7)
Performance (perfect–failure)	4.04 (2.62)	6.12 (4.53)	6.04* (3.9)	7.75* (4.95)	9.41* (5.0)
Effort (low–high)	4.83 (4.04)	7.04 (4.5)	9.79* (4.7)	10.87* (5.10)	12.16* (4.8)
Frustration (low–high)	4.87 (3.8)	6.70 (4.8)	7.04* (5.3)	7.83* (5.49)	9.125* (5.9)

Table 3. Mean (and standard deviation) of ranking for each technique.

Questionnaire (1–5)	Head cursor	Gyrowand	Dwell	Bimanual	Flick
Overall preference (high–low)	1.5 (0.6)	2.58 (1.44)	3.25 (0.9)	3.5 (1.11)	4.16 (1.14)
Easy-to-understand (disagree–agree)	4.25 (1.01)	4 (0.91)	4 (0.91)	4.5 (0.64)	2.75 (1.65)
Well assisted (disagree–agree)	4.16 (1.28)	3.75 (0.92)	3.25 (1.01)	3.5 (0.76)	2.66 (1.31)
Easy-to-control (disagree–agree)	4.33 (1.1)	3.22 (1.17)	3.75 (0.92)	3.66 (1.02)	2.5 (1.38)

Table 4. Main effects of techniques on Nasa-TLX workload index.

Main effects of technique on	Significance difference between
Mental demand ($F_{4,103} = 10.845$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.296$)	Bimanual and Head cursor ($P = 0.001$)
	Dwell and Head cursor ($P = 0.004$)
	Bimanual and Head cursor ($P = 0.002$)
	Head cursor and Dwell ($P = 0.001$)
	Flick and GyroWand ($P = 0.01$)
	Head cursor and Flick ($P < 0.001$)
Physical demand ($F_{4,103} = 43.907$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.630$)	Bimanual and GyroWand ($P < 0.001$)
	Head cursor and Bimanual ($P < 0.001$)
	Dwell and GyroWand ($P < 0.001$)
	Head cursor and Dwell ($P < 0.001$)
	Head cursor and Flick ($P < 0.001$)
	GyroWand and Flick ($P < 0.001$)
Temporal demand ($F_{4,103} = 19.182$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.427$)	Bimanual and GyroWand ($P = 0.008$)
	Bimanual and Head cursor ($P < 0.001$)
	Dwell and Head cursor ($P = 0.004$)
	Flick and Dwell ($P < 0.001$)
	Flick and GyroWand ($P < 0.001$)
	Flick and Head cursor ($P < 0.001$)
Performance ($F_{4,103} = 11.836$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.315$)	Flick and Dwell ($P = 0.001$)
	Flick and GyroWand ($P = 0.01$)
	Head cursor and Bimanual ($P < 0.001$)
	Head cursor and Flick ($P < 0.001$)
Effort ($F_{4,103} = 11.836$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.457$)	GyroWand and Bimanual ($P < 0.001$)
	GyroWand and Dwell ($P < 0.001$)
	GyroWand and Flick ($P < 0.001$)
	Head cursor and Bimanual ($P < 0.001$)
	Head cursor and Flick ($P < 0.001$)
	Head cursor and Dwell ($P < 0.001$)
Frustration ($F_{4,103} = 5.012$, $P = 0.001$, $\eta^2_{\text{partial}} = 0.163$)	Head cursor and Bimanual ($P = 0.034$)
	Head cursor and Flick ($P < 0.001$)

Table 5. Main effects of techniques on ranking.

Main effects of technique on	Significance difference between
Overall preference ($F_{4,55} = 9.544$, $P < 0.001$, $\eta^2_{\text{partial}} = 0.410$)	Bimanual and Head cursor ($P = 0.001$) Dwell and Head cursor ($P = 0.004$) GyroWand and Flick ($P = 0.012$) Flick and Head cursor ($P < 0.001$)
Understandability ($F_{4,55} = 4.461$, $P = 0.03$, $\eta^2_{\text{partial}} = 0.245$)	Bimanual and Flick ($P = 0.03$) Head cursor and Flick ($P = 0.016$)
How assisted ($F_{4,55} = 2.973$, $P = 0.027$, $\eta^2_{\text{partial}} = 0.178$)	Flick and Head cursor ($P = 0.019$)
Controllability ($F_{4,55} = 3.868$, $P < 0.05$, $\eta^2_{\text{partial}} = 0.220$)	Flick and Head cursor ($P = 0.04$)

projection through a 2D cursor. Therefore, while objects in different depths have similar screen projections, GyroWand required more limb orientation than the head controller to select them (see Fig. 10). Another reason, possibly, when using Head cursor in given reference frames, a user rotates his head toward the targets to select them regardless the target remains whether in the FOV or not. However, using GyroWand when the target remains out of the FOV requires user to do a two-step operation; user needs to first rotate his head to spot the target, and then direct the ray by using the controller.

Traditional raycasting has previously been shown to outperform virtual hand using 3D cursor (Bowman *et al*., 1999) for 3D environments such as volumetric display (Grossman and Balakrishnan, 2006) where 3D content was also within arm's reach. However, in our study GyroWand (Raycasting), which is faster than a traditional hand-directed raycasting (Hincapie-Ramos *et al*., 2015), and hand techniques (3D cursor) performed similar speed. While motor and visual space are superimposed in our settings, therefore users can only rely on proprioceptive feedback, study in (Grossman and Balakrishnan, 2006) decoupled these spaces due to using a volumetric display. This decoupling makes the proprioceptive feedback insufficient for users during the task and requires also visual feedback (Argelaguet and Andujar, 2013), which might require additional time, resulting slower performance with 3D cursor.

In our study, the difference between the first and the last arrival times approximates corrective phase of a goal-directed movement. This phase of the movement corrects initial rough movement to select the target with slower velocity (Woodworth, 1899). While both GyroWand and Head cursor are ray based, required times for correction phase are significantly different. We think one reason for this is that input error (e.g. jitter from head/hand or IMU) becomes larger when using GyroWand, because it presents a larger orientation amount for selections (see Fig. 10). To reduce this (if IMU noise induced), while the user is refining the selection with small movements, and therefore generating less errors, the ray apex can be narrower automatically in the corrective phase. However, further study is required to find whether the jitter differences between group of muscles in head

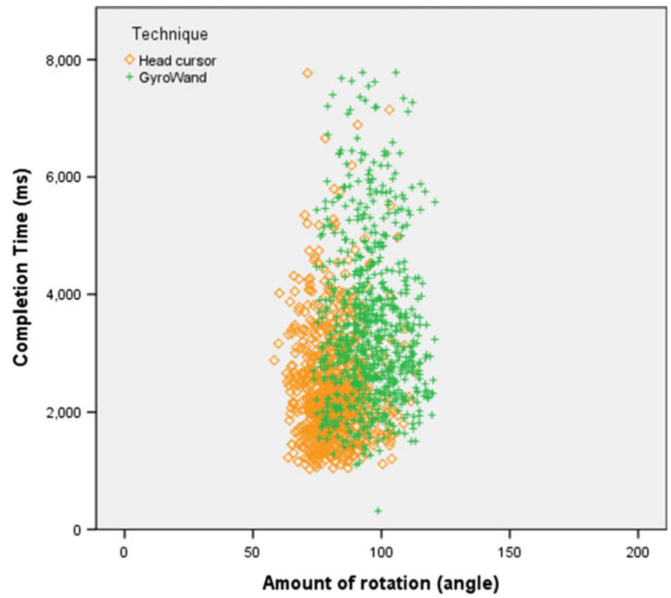

Figure 10. Amount of angular displacements of Head cursor and GyroWand techniques from one selection to the next.

(neck) and hand cause this difference or not. However, for the Flick, we think the participants had difficulties with controlling its mechanism as they mentioned in the questionnaire.

Regardless having experience with HMD, hand-controlled mid-air interaction and virtual reality, 11 participants (91.6%) chose raycasting techniques for 3D selection techniques. This could be because while raycasting techniques require low and small movements of limb. Participants commonly suffered from arm fatigue in the all hand-controlled 3D cursor techniques (see Physical Load in Table 2). Because the repeated use of a mid-air hand is likely to cause fatigue. As we previously stated that hand-controlled 3D cursor is selected because of its naturalness and intuitiveness, which can be a powerful interaction way. However, our results show (despite using high fidelity tracking) that it is not true. Even if we will have an ideal vision-based hand tracker around OST-HMDs in the near future, such a hand-controlled

3D cursor would not be a well-accepted input method due to its fatigue, and social issues in outdoor environments. These problems could be generalized and common discussion point in direct mid-air interaction for mobile setting. Therefore, we can infer that limiting the amount of limb's physical movements is an important consideration when designing interaction techniques. For this reason, IMU-based interaction technique is an alternative that reduces such movements required for interaction as it simply requires minor tilting of the head or hand.

Our study highlights the comparison of selection techniques in terms of speed, accuracy, subjective workload and users' preference as for OST-HMDs in outdoor environments. From our results, we learn that:

(1) Using IMU sensing only is a suitably good approach for object selection in mobile OST-HMD platforms. In the absence of external tracking, as is the case with mobile environments, head-directed 2D cursor is an efficient pointing solution.
(2) Comparing with hand wrist orientations, using head rotations is a good design option to select targets laid out in 3D in OST-HMDs, and is preferable when some objects located out of user's FOV.
(3) Overall, raycasting-based interaction techniques that reduce the amount of limb movements required, are still good to interact the objects within arm's reach with a mobile OST-HMD.

4.3. Design consideration

Participants had difficulty selecting spatial targets with hand-based techniques (3DOFs) within arm's reach, and performed slower selection time than ray-based techniques (2DOFs). However, the translational motions of hand resulted higher accuracy than the rotational motions of the ray-based techniques. Hand-movement-based techniques are generally more accurate than particularly, 2D plane (Head cursor) interface in contrast to findings in Cockburn *et al*. (2011).

Although we explore 3D object selection within user arm's reach, both Head cursor and GyroWand techniques provide selection of targets beyond arm's reach in the reference frames. Since the small movements of these two techniques are less physical demanding and fast, therefore less accurate for small remote targets, they could be probably preferable when interacting with larger targets in relatively large AR environments. Instead of direct hand-controlled 3D point cursor, constrained hand movements, such as arms-down mid-air techniques with higher control-display (CD) gain, using sensors mounted on the thigh, may deliver low arm fatigue and social acceptance issues for hand-controlled 3D cursor techniques (Haque *et al*., 2015). However, the value of CD gain must be sensibly considered because it might increase overshooting, negatively affect accuracy because of the reduced distances in motor space (Liu *et al*., 2015).

5. CONCLUSION

This work presents the results of a formal experiment comparing five selection techniques to interact with virtual spatial objects within arm's reach in OST-HMDs. Selected techniques rely solely on tracking technologies built into conventional commercially available OST-HMDs, enabling interaction in mobile scenario where the devices do not need external tracking environments. We investigated the effectiveness and efficiency of five techniques for arm's reach selection: two are based on raycasting using built-in IMU sensing, and three are based on a hand-controlled 3D point cursor using gestural tracking on an OST-HMD. Results showed that raycasting techniques using head rotation facilitates selection for speed, and ease of use in mobile OST-HMD settings. In addition, hand-controlled 3D point cursor techniques are more accurate; however, presented high levels of arm fatigue in selecting objects. We have learned that reducing amount of limb movements, even for objects within arm's reach, is an important consideration when designing spatial selection techniques for OST-HMDs. We believe that these findings could be useful for designing future interaction techniques in mobile OST-HMDs.

Future work could study whether the jitter differences between group of muscles in head (neck) and hand cause the difference between the Head cursor and GyroWand techniques.

FUNDING

Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University, and Grants-in-Aid for Scientific Research (KAKENHI) (15H01697).

REFERENCES

Argelaguet, F. and Andujar, C. (2013) A survey of 3D object selection techniques for virtual environments. Comput. Graphics, Vol 37, pp. 121–136. Elsevier.

Atheer One, http://www.atheerair.com/smartglasses. 2015.

Borg, G. (1998) Borg's Perceived Exertion and Pain Scales. Champaign, IL: Human Kinetics.

Bowman, D.A., Johnson, D.B. and Hodges, L.F. (1999) Testbed Evaluation of Virtual Environment Interaction Techniques. In Proceedings of the VRST '99, pp. 26–33. ACM.

Bowman, D.A., Wingrave, C.A., Campbell. J.M, and Ly, V.Q. (2001) Using Pinch Gloves for both Natural and Abstract Interaction Techniques in Virtual Environments, In Proceedings of the HCI'01, pp. 629–633.

Bowman, D.A., Kruijff, E., LaViola, J.J. Jr and Poupyrev, I. (2004) 3D User Interfaces: Theory and Practice. Addison-Wesley.

Cockburn, A., Quinn, P., Gutwin, C., Ramos, G. and Looser, J. (2011) Air pointing: design and evaluation of spatial target acquisition with and without visual feedback. Int. J. Hum.-Comput. Stud., Vol 69, pp. 401–414. Elsevier.

Djajadiningrat, J.P. Cubby: What You See Is Where You Act Interlacing The Display and Manipulation Spaces. Doctoral dissertation, Delft University of Technology, 1998.

Ens, B., Byagowi, A., Han, T., Hincapie-Ramos, J.D. and Irani, P. (2016) Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces. In Proceedings of the SUI '16. ACM.

Forsberg, A., Herndon, K., Zeleznik, R. (1996) Aperture based Selection for Immersive Virtual Environments. In Proceedings of the UIST '96, pp. 95–96. ACM.

Grossman, T. and Balakrishnan, R. (2004) Pointing at Trivariate Targets in 3D Environments. In Proceedings of the CHI'04, pp. 447–454. ACM.

Grossman, T. and Balakrishnan, R. (2006) The Design and Evaluation of Selection Techniques for 3D Volumetric Displays. In Proceedings of the UIST '06, pp. 3–12. ACM.

Haque, F., Nancel, M. and Vogel, D. (2015) Myopoint: Pointing and Clicking Using Forearm Mounted Electromyography and Inertial Motion Sensors. In Proceedings of the CHI '15, pp. 3653–3656. ACM.

Hart, Sandra G. and Lowell, E. (1988) Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research.Adv Psychol, 52, pp. 139–183. North Holland Press, Amsterdam.

Ens, B., Hincapié-Ramos, J.D. and Irani, P. (2014) Ethereal Planes: A Design Framework for 2D Information Space in 3D Mixed Reality Environments. ACM. In Proceedings of the SUI '14, pp. 2–12.

Epson Moverio, http://www.epson.com/moverio. 2015.

Hincapié-Ramos, J.D., Ozacar, K., Irani, P. and Kitamura, Y. (2015) GyroWand: IMU-based Raycasting for Augmented Reality Head-Mounted Displays. In Proceedings of the SUI '15, pp. 89–98. ACM.

Microsoft HoloLens, https://www.microsoft.com/microsoft-hololens/en-us. 2015.

Jota, R., Nacenta, M., Jorge, J., Carpendale, S. and Greenberg, S. (2010) A comparison of Ray Pointing Techniques for Very Large Displays. In Proceedings of the GI'10, pp. 269–276. ACM.

Lee, S., Seo J., Kim GJ, Park C-M, Evaluation of Pointing Techniques for Raycasting Selection in Virtual Environments. In Proceedings of Conference On Virtual Reality and Its Application in Industry, pp. 38–44. 2003.

Lemmerman, D.K. and LaViola, J.J. (2007) Effects of Interaction-Display Offset On User Performance in Surround Screen Virtual Environments. In Proceedings of the VR'07, pp. 303–304. IEEE.

Liu, M., Nancel, M. and Vogel, D. (2015) Gunslinger: Subtle Arms-down Mid-air Interaction. In Proceedings of the UIST'15, pp. 63–71. ACM.

Metaio SDK, http://www.metaio.com/products/sdk/. 2015.

META SpaceGlassess, https://www.getameta.com/. 2015.

Müller, C-Tomfelde. (2007) Dwell-Based Pointing in Applications of Human Computer Interaction. In Proceedings of the INTERACT'07, pp. 560–573. Springer International Publishing.

Olwal, A., Feiner, S. (2003) The Flexible Pointer: An Interaction Technique for Selection in Augmented and Virtual Reality. In Proceedings of the UIST '03, pp. 81–82. ACM.

Pierce, J.S., Forsberg, A., Conway, M.J., Hong, S., Zeleznik, R., Mine M.R. (1997) Image Plane Interaction Techniques in 3D Immersive Environments. In Proceedings of the I3D'97, pp. 39–43. ACM.

RGBDSLAM, https://openslam.org/rgbdslam.html. 2016.

Robert, J.K.J. (1991) The use of eye movements in human–computer interaction techniques: what you look at is what you get. ACM Trans. Inf. Syst.1118

Špakov, O. and Miniotas, D. (2004) On-Line Adjustment of Dwell Time for Target Selection by Gaze. In Proceedings of the NordiCHI '04, pp. 203–206. ACM.

Tanriverdi, V. and Jacob, R.J.K (2000) Interacting with Eye Movements in Virtual Environments. In Proceedings of the CHI '00, pp. 265–272. ACM.

Ventura, J., Arth, C., Reitmayr, G. and Schmalstieg, D. (2014) Global localization from monocular SLAM on a mobile phone. IEEE Trans. Vis. Comput. Graph., 20, 531–539.

Vogel, D. and Balakrishnan, R. (2005) Distant Freehand Pointing and Clicking On Very Large, High Resolution Displays. In Proceedings of the UIST'05, pp. 33–42. ACM.

Vuforia for Smart EyeWear, https://www.qualcomm.com/products/vuforia. 2016.

Wang, Y. and MacKenzie, C.L. (1999) Effects of Orientation Disparity Between Haptic and Graphic Displays of Objects in Virtual Environments. In Proc. INTERACT'99, pp. 391–398. Springer International Publishing.

Wang, Y. and MacKenzie, C. (2000) The Role of Contextual Haptic and Visual Constraints On Object Manipulation in Virtual Environments. In Proceedings of the CHI'00, pp. 532–539. ACM.

Ware, C. and Lowther, K. (1997) Selection using a one-eyed cursor in a Fish Tank VR Environment. ACM Trans. Comput.-Hum. Interact. 309–322.

Winer, B.J., Brown, D. R. and Michels, K. M. (1991) Statistical Principles in Experimental Design (3rd edn), McGraw-Hill, New York.

Woodworth, R. S. (1899) The Accuracy of Voluntary Movement. The Psychological Review: Monograph Supplements, New York, MacMillan.

Zhou, P., Li, M. and Shen, G. (2014) Use it Free: Instantly Knowing Your Phone Attitude. In Proceedings of the MobiCom'14, pp. 605–616. ACM.