The Version of Record of this manuscript has been published and is available in the International Journal of Human-Computer Interaction. Date of publication pf the electronic version: 25th October 2018. DOI: 10.1080/10447318.2018.1541546

Please cite as:
Chittaro L., Sioni R. (2019) Selecting Menu Items in Mobile Head-Mounted Displays: Effects of Selection Technique and Active Area, *International Journal of Human–Computer Interaction*, 35, pp. 1501-1516.

Selecting Menu Items in Mobile Head-Mounted Displays:
Effects of Selection Technique and Active Area

Luca Chittaro[1] and Riccardo Sioni[1]*

[1]Human–Computer Interaction Lab (HCI Lab), Department of Mathematics, Computer Science, and Physics, University of Udine, Udine, Italy

*To whom correspondence should be addressed:
riccardo.sioni@uniud.it
Department of Mathematics, Computer Science, and Physics
Via delle Scienze 206, 33100, Udine, Italy.
Telephone +39-0432-55-8413

Abstract

Head-mounted displays (HMDs) are increasingly available to users after the launch of new-generation consumer devices. Moreover, mobile HMDs such as Samsung Gear VR and Google Daydream View allow users to experience VR through a smartphone, without requiring connection to a PC. Commercial applications for mobile HMDs exploit different techniques to perform menu selection tasks. This paper contrasts the two most used techniques, i.e., dwell-based and touchpad-based selection, which were not experimentally compared before. We consider different versions of a menu pointing and selection task in which participants interacted with a Samsung Gear VR. Results show that participants were slower with the dwell-based technique rather than the touchpad-based technique. However, the dwell-based technique led to fewer errors and was perceived as more usable, more comfortable and less fatiguing than the touchpad-based technique. We also evaluated two different active areas for the selection, discussing the results.

Keywords: Mobile HMDs, virtual reality, menus, selection techniques, user studies

Introduction

Head-mounted displays (HMDs) for virtual reality (VR) are available to a growing number of users, thanks to the availability of new-generation consumer devices such as the Oculus Rift, HTC Vive, PlayStation VR, Samsung Gear VR, Google Cardboard and Google Daydream View. Devices like the Rift and the Vive emphasize performance, and require connection to powerful PCs for providing users with complex, high quality graphics, and 6 DOF tracking. On the other hand, HMDs such as the Cardboard, the Daydream View, and the Gear VR (*mobile HMDs,* in the following) trade some graphic performance and tracking capabilities for greater convenience and lower cost, allowing users to experience VR anywhere, using their own smartphone. These features may have contributed to the success of mobile HMDs: in 2017, Samsung alone shipped 3.7 million units of its mobile HMD. For comparison, the most successful non-mobile HMD, Sony PlayStation VR, has shipped 1.7 million units in 2017 (SuperData Research, 2018).

The Gear VR includes a small touch-sensitive surface and a physical button on its right side to interact with mobile VR applications, i.e., applications designed for mobile HMDs. A similar solution is adopted also by Google Cardboard, and Cardboard-compatible devices. Commercial mobile VR applications have come up with different ways of exploiting built-in touchpads and buttons to perform selection tasks. In some applications, users can point at a virtual object by moving the head to position an on-screen cursor over the object, and then select it by pressing the button on the HMD or tapping on the built-in touchpad with one hand. Other applications have done away with the physical controls, keeping user’s hands completely free, and allowing users to move their head to position the cursor and then dwell for a specific amount of time over the object to trigger the selection. Hands-free interaction with VR applications can be considered an especially compelling alternative to traditional controls (e.g., joypad, or mouse and keyboard) in case of upper limb motor disabilities as well as situations in which users’ hands are continuously occupied with other tasks (Zhai, Morimoto, & Ihde, 1999).

Our study aims at contrasting two different menu selection techniques currently employed in mobile HMDs, i.e., dwell-based selection and touchpad-based selection. To the best of our knowledge, these selection techniques for mobile HMDs have not been experimentally compared yet for menu interaction. Our study considers different versions of a pointing and selection task in which participants were asked to interact with menus in stereoscopic VR on the Gear VR. Moreover, we took into consideration two types of active area for the menu items, reflecting alternatives that are found in existing mobile HMD applications. The task we selected is representative of the actions users are required to perform when interacting with a wide range of mobile VR applications: from basic interactions with movie player interfaces to menu selection in interactive VR experiences. We evaluated the two selection techniques, as well as the two types of active area, in terms of usability, user performance, and comfort.

The paper is organized as follows. First, we introduce dwell-based selection research, and introduce the techniques used in mobile HMDs. Then, we describe the interfaces we developed and the details of the user study, presenting the results and their discussion. Finally, we outline conclusions and future work.

Related Work

Dwell-Based Selection and the Midas Touch Problem

During interaction with an immersive VR experience, the most common option to trigger a selection is to press a button (Argelaguet & Andujar, 2013). However, other interaction techniques have been suggested; for example, selection could be triggered by *dwelling* on the target (Jacob, 1991; Steed et al., 2016). In dwell-based selection, the selection is issued if the user looks at the desired object for a given time. By using gaze or head tracking information for both pointing and selecting an object, there is no need to click the mouse button or, more generally, issue an explicit command to select (Jacob, 1991). However, a dwell time is required, otherwise this approach can become unusable: without a dwell time, commands would be issued everywhere a user looks at, making inadvertent selections highly likely. This well-known issue is called the *Midas touch problem* (Jacob, 1991; Ohno, 1998). Dwell time helps in preventing unwanted selections, reducing the impact of the Midas touch problem (Jacob, 1991; Zander, Gaertner, Kothe, & Vilimek, 2011). However, it also reduces the speed advantage of using eye or head movements for input rather than traditional selection devices like mouse or keyboard (Jacob, 1991), and may annoy and demotivate experienced users (Zander et al., 2011). Another limitation of dwell-based selection discussed by Zander et al. (2011) is that the system cannot know whether the user is dwelling on a GUI widget to trigger a selection or for other reasons. For example, (s)he may be having difficulties in reading the widget description or in understanding the meaning of its icon, or (s)he may be reflecting about the corresponding action. Users' intentions (and thus an optimal dwell time) cannot be inferred from the duration of users' fixations (Zander et al., 2011). Different dwell times have been suggested in the literature. In his proposal of a general model of selection in VR, Steed (2006) suggests the use of dwelling as a feasible selection technique, suggesting a dwell time in the order of a second. As reported by Müller-Tomfelde (2007), studies in the HCI literature proposed dwell times ranging from 300 ms to about 2 s; Müller-Tomfelde points out that the adjustment of the dwell time is usually based on ad-hoc experiences of the developer.

Note that, when head movement instead of gaze is used for dwell-based selection, eye movement is decoupled from head movement and does not affect selection. In other words, using head movement for dwell-based selection allows users to freely look around the display, while controlling head orientation to avoid unwanted selections (Bates & Istance, 2003), possibly mitigating the Midas touch problem. Pointing with head movements leads to a higher performance than gaze-based pointing (Bates & Istance, 2003; Qian & Teather, 2017), and reduces the error rate in point-and-select tasks (Jalaliniya, Mardanbeigi, Pederson, & Hansen, 2014; Qian & Teather, 2017). Furthermore, novice users prefer pointing and selecting using head movement rather than gaze (Bates & Istance, 2003). For these reasons, recent research has started focusing on the design of interaction techniques that could exploit natural eye movements (Piumsomboon, Lee, Linderman, & Billinghurst, 2017) to propose alternatives to dwell-based selection techniques that could be effective.

Manual Selection and Dwell-Based Selection in the VR Literature

Dwell-based and manual selection techniques have been generally compared in the literature in the context of tasks carried out on PC screens that involve eye-gaze selection of 2D objects (Bohan & Chaparro, 1998; Miniotas, 2000; Sibert & Jacob, 2000; Ware & Mikaelian, 1987), or gaze-based interaction with on-screen keyboards to perform text entry (Bee & André, 2008; Hansen, Johansen, Hansen, Itoh, & Mashino, 2003; Huckauf & Urbina, 2008). When considering these selection tasks, studies with HMDs are only a few (e.g., Yu et al., 2017) and, to the best of our knowledge, the VR literature has not yet contrasted manual selection and dwell-based selection with menu tasks using a mobile HMD.

Table 1 summarizes research studies that focused on pointing and selection tasks with HMDs. The table focuses on studies that discuss dwell-based or manual-based selection, and studies that focus on participants' interaction with GUI widgets. The following subsections discuss these studies in more detail.

Reference	Pointing technique	Selection technique	Selection target	VR device	Task(s)	Independent variable(s)	Dependent variable(s)	Main findings
Frees et al. (2007)	Tracking of hand-held 6-DoF stylus	Button press	3D objects	HMD	T1: translate an object to match another object's position. T2: rotate an object to match another object's rotation. T3: translate and rotate an object to match another object's position and rotation.	Interaction technique (direct, i.e., object follows stylus movements; *PRISM*, i.e., object manipulation speed adjusted according to stylus speed); error tolerance (easy, medium, hard)	Manipulation completed per trial; error rate; task completion time; user feedback.	More manipulation completions, less errors and faster task completion with PRISM than direct interaction. PRISM interaction less fatiguing than direct interaction.
Lin et al. (2009)	Tracking of medium finger	Button press	3D objects	HMD; 3D TV; projector display	Select objects as quickly as possible in a specific order.	Display device (3D TV, HMD, projection display); target position (along x, y, z axis); Fitts' Index of Difficulty (2.0, 3.0, 4.0)	Movement time; throughput; simulation sickness.	Highest performance with projection display; worst results in terms of movement time, throughput, and sickness with HMD.
Lubos et al. (2014)	Tracking of index finger	Button press	3D objects	HMD	Select objects as quickly as possible in a specific order.	Ring position compared to participants' celiac plexus (top-left, top-center, top-right, middle-left, middle-center, middle-right, bottom-left, bottom-center, bottom-right); spheres' distance from participants' starting position (close, middle, far); sphere radius (two radii calculated to obtain a Fitts' Index of Difficulty of 3.0 and 4.0 respectively for each starting position)	Movement time; error rate; error distance; throughput.	Higher movement time with rings in bottom positions than those in top positions. Higher movement time with higher index of difficulty (i.e., combination of sphere radius and distance). Higher error rate with higher index of difficulty. Lower throughput with rings located in bottom positions compared to rings in other positions.
Cournia et al. (2003)	Gaze tracking; tracking of hand-held 6-DoF mouse	Button press	3D objects	HMD	Each object, when selected, fades to reveal a 2D label with a letter; participants have to locate the object containing a target letter as quickly as possible.	Pointing technique (gaze, mouse); distance (arm's length, beyond arm's reach)	Task completion time	For distant objects, longer time to complete task with gaze than 6-DoF mouse.
Tanriverdi & Jacob (2000)	Gaze tracking; tracking of index finger	Dwelling	3D objects	HMD	Each object, when selected, fades to reveal a 2D label with a letter; participants have to locate the two objects containing a target letter as quickly as possible.	Pointing technique (gaze tracking, finger tracking); distance (within arm's reach, 5-15 inches beyond arm's reach)	Task completion time; number of selected objects which position was correctly recalled; participants' satisfaction.	Shorter completion time with gaze-based than hand-based interaction. More difficulty in recalling location of selected object with gaze-based than hand-based interaction. No difference in satisfaction.

Steed et al. (2016)	Head tracking	Dwelling	GUI widgets	HMD	Fill a questionnaire	-	-	Dwell-based questionnaire filling criticized probably because of lack of audio/visual feedback of selection.
Bowman & Wingrave (2001)	-	Finger pinching	GUI widgets	HMD	Change an object shape, color and texture to match a target object.	Menu type (items displayed on fingers, floating pull-down menu, virtual tablet); experience with 3D environment (no experience, some experience); trials considered (all 30 trials, last 10 trials)	Task completion time; subjective comfort rate	Faster interaction with tablet menu than with other menu types. Differences in completion time smaller for experienced participants than for inexperienced ones. In last 10 trials, experienced participants faster with TULIP than with floating menu. Floating menu more fatiguing than other menu types.
Dias et al. (2016)	-	Gripping gesture, button press	3D objects, GUI widgets	HMD	T1: perform menu selection (preliminary study). T2: translate, rotate and scale to match another object's position, rotation, and scale.	Menu selection method (grasping gesture, button presses on a controller); 3D manipulation method (hand gestures, button presses on a controller)	Informal feedback; time to complete task; accuracy	Menu interaction through gestures slower and more error-prone than button-based interaction, but preferred over other methods. Object translation faster with button presses than gestures; rotation faster with gestures than button presses. No differences in manipulation accuracy.
Davis et al. (2016)	-	Finger gestures	GUI widgets	HMD	-	-	Informal feedback from participants	Hand gestures quick to learn and use to perform actions onscreen, discern icons and their meanings, and navigate menus.
Kim et al. (2000)	Tracking of 6-DoF mouse	Button press, finger pinching	GUI widgets	HMD	T1: create a 3D object and view help information. T2: create a 3D object, select it, and change its color.	Interaction method (6-DoF mouse tracking (pointing) and button press (selection), pinching (selection))	Task completion time, pointing completion time, selection completion time, number of errors	Selection faster than pointing with 3D mouse tracking with button press; pointing faster than selection with hand movements and gestures
Yu et al. (2017)	Head tracking	Dwelling, button press, gesture	Keys on a virtual keyboard	HMD	Enter text as fast as possible	Key selection technique (dwell-based, manual, gesture-based)	Text entry rate; subjective assessment of speed, accuracy, distance, fatigue, learnability, preference	Text entry rate higher with manual selection than dwell-based selection; text entry rate higher with gesture-based selection than manual selection; participants found the level of fatigue acceptable for all selection techniques

Table 1. Studies that employ HMDs in pointing and selecting tasks and tasks that require users to interact with GUI widgets.

Pointing and selection techniques. Many studies in the literature implemented VR interaction techniques that exploit hand or finger gestures (e.g., Lubos, Bruder, & Steinicke, 2014) for pointing, often tracking a hand-held device (e.g., Frees, Kessler, & Kay, 2007). Selection is usually triggered by button presses (e.g., Lin, Sun, Chen, & Cheng, 2009) or by dwelling (Tanriverdi & Jacob, 2000). To the best of our knowledge, no study has explored the use of built-in buttons or touchpads of mobile HMDs for pointing or selection. Most of the literature focused its attention on gaze-based pointing (e.g., Cournia, Smith, & Duchowski, 2003; Tanriverdi & Jacob, 2000).Consumer HMDs (e.g., Rift, Vive, Gear VR, Cardboard, Daydream View) do not include gaze detection, and thus research work that exploits instead head tracking for pointing and selection is particularly relevant to our study. Recently, Yu et al. (2017) compared participants' performance in a text entry task on a virtual keyboard displayed by a Gear VR, in which participants could point at keys by controlling a cursor through head movements. In comparing efficiency of a dwell-based technique (with a dwell time of 800 ms) and a manual selection technique (requiring participants to press a button on a Bluetooth keyboard connected to the HMD), the study reports a greater efficiency, in terms of words per minute, of the manual selection technique compared to the dwell-based technique. Steed et al. (2016) exploited dwell-based selection through head tracking during a study that immersed participants in VR while wearing a Gear VR or a Cardboard. Dwelling was not the focus of the evaluation, and was used only to allow participants to fill pre- and post-test questionnaires directly within the VR application by pointing at an answer for 1.5 s. However, from the participants' verbal feedback collected in the study, the dwell-based selection implemented for questionnaire filling attracted some criticism, likely due to the lack of visual feedback about the progression of dwell. Indeed, sharp and clear feedback seems to be essential in dwell-based interaction, especially with short dwell times (Marajanta, Aula, & Räihä, 2014).

Menu interaction in VR. Most research work on menu interaction in VR focused on selection tasks based on hand and finger gestures (e.g., Bowman & Wingrave, 2001; Dias, Pinto, Eliseu, & Santos, 2016). These studies generally focused on user interaction with radial menus (also called *pie menus*) and similar menu layouts. Bowman and Wingrave (2001) proposed an interaction technique based on finger pinching, i.e., the detection of touches between the user's thumb and the other fingers. The authors compared different menu styles: (i) a collection of menu items that are displayed on participants' fingers; (ii) a pull-down menu floating mid-air; (iii) a menu placed over a virtual tablet which position and rotation corresponded to a physical tabled held by participants in the real world. The third option led to faster interactions compared to the other two options. Dias et al. (2016) exploited instead gripping gestures and focused not only on menu interaction, but also on navigation and 3D object manipulation inside an immersive 3D environment presented through a HMDs for desktop PCs. With regards to navigation and object manipulation, the gesture-based technique performed better, while the controller-based interaction reportedly allowed for a faster interaction with the radial menus.

Despite the large body of research on radial menus in VR, Davis, Gabbard, Bowman, and Gracanin (2016) observed that such design, while ideal for small data sets and icon, is usually replaced by more traditional layouts in support of text-heavy and larger menu data sets.

Manual Selection and Dwell-Based Selection in Mobile VR Applications

In a wide range of mobile VR applications, users are often required to interact with one or more menus: from basic operations of movie player interfaces to action selection in interactive VR experiences and video games. Manual selection is a commonly used technique in mobile VR applications. It requires users to explicitly confirm the selection, e.g., by pressing a button or by tapping on a touch-sensitive surface. For example, the Gear VR version of *Oculus Home* (the main menu displayed when the user puts on the HMD, see Figure 1), and *Universal Menu* (a control panel displaying the various functions of the HMD like general settings and user profile management), require users to interact with GUI widgets by moving their head to place a cursor (a white dot at the center of users' view, Figure 1) over the widgets, and selecting them by tapping on the touchpad. The same technique is employed also for many applications such as *Samsung Video*, a video player for Gear VR.

Dwell-based selection is exploited by mobile applications such as *Titans of Space* (DrashVR, 2016), *Land's End* (Ustwo Games, 2015), and *Cineveo* (Mindprobe, 2016). The Gear VR version of Titans of Space allows users to choose how to perform selection of GUI widgets before starting a virtual visit of the solar system. Users can choose to move their head to dwell on a GUI widget, press a button on a joypad (optional external peripheral), or tap on the built-in touchpad. In the case of dwell-based selection, the application provides visual feedback by flashing the dwelled GUI widget at an increasing rate until selection is performed after 1 s.

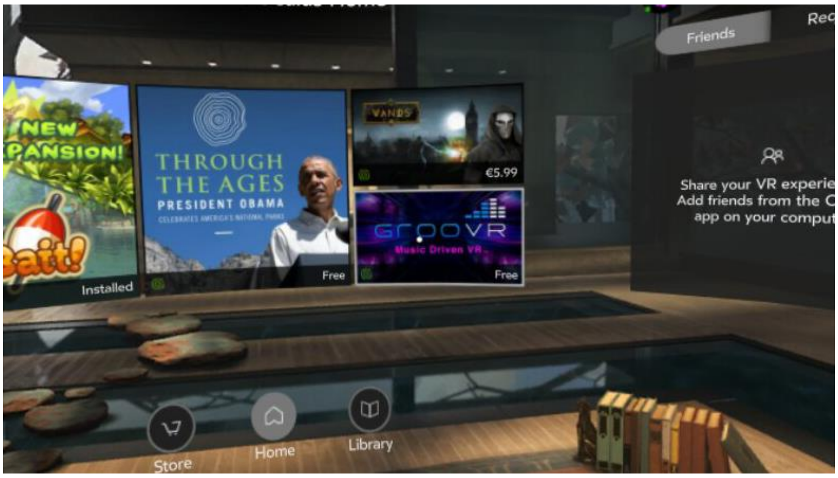


Figure 1. A screenshot from Oculus Home in the Gear VR. A white dot (see the center of the picture) is used as a cursor, and indicates users where they are currently looking.

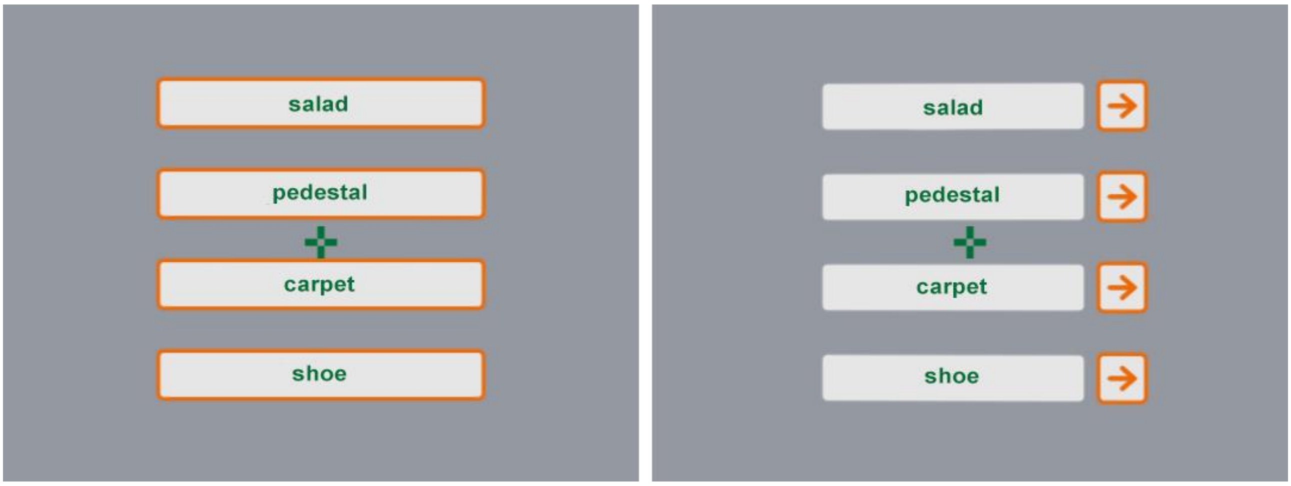
Land's End exploits dwelling to allow users navigate a 3D environment. More specifically, at any moment users can see one or more small white circles near them; after dwelling on one of circles for about 1 s, the circle is selected and the user's avatar moves to the location of the circle through an animation. Visual feedback is provided to inform users when a circle is pointed at as well as when it is selected. Steed and colleagues (2016) indicate Land's End as an example of application with a good feedback about dwelling.

Cineveo for Cardboard is a stereoscopic movie player that immerses users inside a 3D virtual cinema. Dwell-based selection is used to interact with the GUI (e.g., to play and pause a video), and users receive visual feedback in the form of a small annulus around the cursor that gets filled up during dwelling.

Cmoar VR Cinema (Cmoar, 2016) for Cardboard is another movie player that immerses users in a virtual cinema, and exploits dwell-based selection for interacting with the GUI. Unlike Cineveo and Titans of Space, in each GUI widget the visual component (the icon, which is usually the selection target, i.e., the active area of the widget) and the textual component (the label) are separated, and users have to dwell on the visual component to trigger a selection. The visual decoupling of icon and label is also employed by the Universal Menu of the Gear VR: the functions of the HMD are initially shown as rounded icons; once one of them is pointed by users, it is highlighted, and a textual label is shown alongside it to provide further information. As in Cmoar VR Cinema, only the icon component of the GUI widget is a possible target for selection.

Evaluated Techniques and Task

The interface developed for the present study contained a set of four menu items arranged vertically in front of users on an invisible plane which was 1 m distant from the camera and orthogonal to the line of sight (Figure 2). A green crosshair was displayed at the center of the screen to indicate where participants were pointing. The interface was stereoscopic to exploit participants' depth perception: as described in the following, a menu item can be highlighted by having it come closer to the participant. Each item was composed by a textual label and an image. The image was the target the user had to point before performing selection. Participants were asked to choose only one of the four items by pointing and selecting it. Initially, the interface presented a practice session, introduced by a brief textual explanation, in which participants were asked to perform five pointing and selection trials. More specifically, for each trial in the practice session, they had to select the item showing the text "select me"; the wrong items showed the text "do not select me".



a b

Figure 2. The two menu item designs employed in our study. The active area for pointing and selection is respectively (a) the item text and (b) the arrow icon on the right.

After completing the practice session, participants were asked to perform 60 pointing and selection trials (task session). During this session, participants were asked to select the item containing the name of a food (e.g., bread, tomato, meat…), ignoring the non-food items (e.g., chair, phone, shoe…). For each trial in both practice and task sessions, the position of the correct item among the four displayed was assigned randomly. Also, during the task session, the text displayed by each item was randomly selected from a set of edible or non-edible items.

Four versions of the interface were developed for the present study. They differ in (i) the selection technique, and (ii) the design of the menu items. More specifically, in two versions of the interface the active area included the item text, and thus the active area and the item coincided (Figure 2a). In the other two versions of the interface, active areas were displayed as squared icons showing an arrow pointing to the right, with the item text placed to the left of it (Figure 2b): the textual component and the active area of the item were separated. Figure 2a and 2b show how the active areas in the latter two versions of the interface were smaller than the active areas in the former ones.

The action that participants had to carry out to perform pointing was the same across the four versions of the interface: to point at an active area of an item, participants had to move their head to position the active area under the crosshair. For each active area, two versions of the interface exploited two different techniques that participants could use to select an item after pointing at its active area (i.e., the item text or the arrow icon): touching the HMD touchpad (touchpad-based selection, see Figure 3) or dwelling on the active area for 1 s (dwell-based selection). The length of the dwell time was chosen on the basis of dwell times used in existing Gear VR applications (see the Related Work section).

Figure 3. A user is touching the Gear VR touchpad with his index finger.

The interface provided visual and audio feedback for pointing and selection. When an active area (button or icon) was pointed at with the crosshair, the corresponding item moved 2 cm towards the participant. Such movement was chosen over other possible visual feedback (e.g., change of the size or color of the item) to exploit participants' depth perception.

When using the touchpad-based selection technique, participants could select an item by touching the HMD touchpad when the crosshair was over the corresponding active area.

When using dwell-based selection, feedback about dwell time was given in the form of a green annulus displayed around the crosshair. When participants were not pointing at any active area, the annulus was completely translucent. When participants were dwelling on an active area, a sector of the annulus was opaque. The sector angle increased continuously with dwelling (Figure 4). For reference, when participants started dwelling on an active area, the sector angle was 0°. After dwelling for 0.5 s on the same active area, the sector angle was 180°, i.e., the left half of the annulus was opaque, and the right half was translucent. After 1 s of dwelling on the same active area, the sector angle was 360°, i.e., the annulus was completely opaque. When the annulus was completely opaque, the item was selected, and the annulus became translucent again. If participants moved the crosshair outside the active area before the annulus was completely opaque, the selection operation was cancelled, and the annulus became completely translucent. When using touchpad-based selection, the annulus was not displayed.

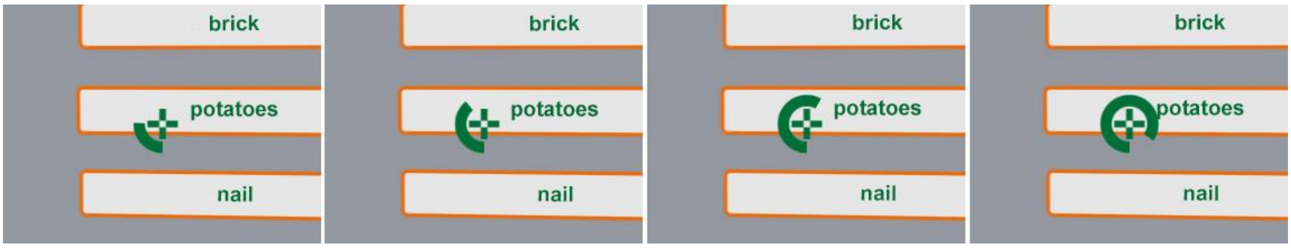


Figure 4. From left to right, the annulus sector angle increased as the participant kept dwelling on a menu item.

Regardless of the selection technique used by participants, audio feedback was the same. Participants could hear two different click sounds: one when the crosshair overlapped an active area, and another one when selection was triggered.

The task described in this section is different from the one proposed by Soukoreff and MacKenzie (2004). In their work, they propose a standardized experimental design, based on Fitts'

Law, to evaluate the performance of pointing and selection tasks. We focused instead on designing an interface that, as anticipated in the introduction, is representative of basic operations of menu selection in mobile VR applications.

Experimental Evaluation

The user evaluation followed a 2 by 2 full-factorial within-subjects design, with technique (*dwell-based selection* or *touchpad-based selection*, DW selection or TP selection for short) and active area (*item text* or *arrow icon*, TXT area and ARW area in the following) as the independent variables.

Material and Measures

The four versions of the interface were run on a Samsung Gear VR containing a Samsung Galaxy S6.

To collect participants' demographic data and their subjective opinions, we employed the following questionnaires:

- *Demographic questionnaire*. We recorded participants' age and gender, and we interviewed them about how many hours they had previously used HMDs.
- *Simulator Sickness Questionnaire*. We employed four essential items (1, 3, 4 and 8) from the Simulator Sickness Questionnaire (SSQ; Kennedy, Lane, Berbaum, & Lilienthal, 1993): (i) general discomfort, (ii) headache, (iii) eye strain, and (iv) nausea. Participants rated how much each symptom affected them at that moment on a 4-point scale (1 = "none", 2 = "slight", 3 = "moderate", 4 = "severe"). We did not use the other 12 SSQ items because we wanted to minimize the negative influence of pre-test compilation of SSQ that was reported on participants' responses to post-test SSQ (Young, Adelstein, & Ellis, 2006).
- *System Usability Scale*. The System Usability Scale (SUS; Brooke, 1996) is a reliable, 10-item usability scale that can be used for global assessments of systems usability (Brooke, 1996). Items are rated on a 5-level Likert scale (0 = "strongly disagree", 4 = "strongly agree"). Since the second, fourth, sixth, eighth and tenth items contain negative statements, their scores must be reversed. Then, the usability score is obtained by multiplying by 2.5 the sum of participants' ratings, and thus it is in the 0-100 range.
- *Device Assessment Questionnaire*. The Device Assessment Questionnaire (DAQ; Douglas, Kirkpatrick, & MacKenzie, 1999) is a 13-item questionnaire covering issues of physical operation, fatigue and comfort, speed and accuracy, and overall usability. Each item is rated from low to high on a 5-level scale. Anchors are different for each item; for example, anchors for item 1 ("The force required for actuation was…") are 1 = "too low" and 5 = "too high", and anchors for item 12 ("General comfort") are 1 = "very uncomfortable" and 5 = "very comfortable".
- *Ranking*. At the end of the evaluation, participants were asked to rank from 1 (best) to 4 (worst) the four versions of the interface (ties were not allowed) according to their subjective preference.
- *Logging*. During the task, the interface recorded the following data:
    - Time to reach active areas, i.e., how much time participants spent reaching active areas during each trial. More precisely, the time to reach active areas was the amount of time during which the crosshair was not over any active area;
    - Selection time, i.e., how much time participants spent hovering over active areas during each trial. More precisely, selection time was the amount of time during which the crosshair was over one of the active areas;
    - Number of errors, i.e., the number of times participants selected the wrong item during the 60 trials. This number was in the 0-60 range.

The following measures were collected for each condition and user:

- Difference between pre-test and post-test scores of general discomfort, headache, eye strain, and nausea (from the selected items of SSQ);
- Usability score (from SUS);
- Required force score, smoothness score, mental effort score, physical effort score, ease of accuracy score, speed score, finger fatigue score, wrist fatigue score, arm fatigue score, shoulder fatigue score, neck fatigue score, general comfort score, and ease of use score (from DAQ);
- Ranking of each experimental condition;
- Mean time to reach active areas, mean selection time, total task time (as the sum of total time to reach active areas and total selection time, both gathered from logs), number of errors (from logs).

Participants

The within-subjects study involved a sample of 24 participants (13 M, 11 F) recruited among graduate and undergraduate students at our university. Participants were volunteers who received no compensation. Mean age was 25.71 (SD = 3.66), and mean number of hours of previous HMD use was 1.50 (SD = 2.06).

Procedure

Participants were clearly informed that the collected experimental data was going to be analyzed anonymously for research purposes, then demographic data were collected. The order of conditions was counterbalanced using a Latin Square design to prevent order effects. Before each condition, the experimenter briefly described the associated technique and active area. Then, participants filled the four items from the SSQ and wore the Gear VR. Participants were instructed to carry out the task as fast as possible.

For each experimental condition, after carrying out the five trials of the practice session, participants were asked if they had understood how to point and select. All participants responded affirmatively in all experimental conditions. Then, after carrying out the 60 trials of the task session, participants took off the HMD, and filled again the four items from the SSQ, as well as the SUS and the DAQ.

After trying all four conditions, participants were asked to rank them. Finally, participants were debriefed about the experiment, and thanked for their participation.

Results

To analyze questionnaire data, we employed the nonparametric procedure described by Wobbrock, Findlater, Gergle, and Higgins (2011), which involves first an aligned rank transformation using *ARTool* (Wobbrock et al., 2011), followed by a two-factor repeated measures analysis of variance (RM ANOVA) on the transformed data. When we observed a significant interaction, we analyzed simple effects by carrying out multiple comparisons on non-transformed data with Wilcoxon signed-rank tests with Bonferroni correction, testing each hypothesis at a statistical significance level of $0.05/4 = 0.0125$. When we observed a significant, non-disordinal interaction, we also interpreted the main effects as suggested by Cohen (2013). To analyze data logged by the interface (time to reach active areas, selection time, and number of errors), we carried out two-factor RM ANOVAs.

In the following sections, the four experimental conditions are indicated as DW-TXT (*dwell-based selection technique* with *item text active area*), DW-ARW (*dwell-based selection technique* with *arrow icon active area*), TP-TXT (*touchpad-based selection technique* with *item text active area*), and TP-ARW (*touchpad-based selection technique* with *arrow icon active area*).

Sickness

Figure 5 shows the mean values of the difference between pre-test and post-test scores of general discomfort, headache, eye strain, and nausea. Table 2 shows the results of the statistical analysis. The analyses revealed no main effects and no significant interaction of technique and active area for the difference between pre-test and post-test scores of the four sickness-related variables.

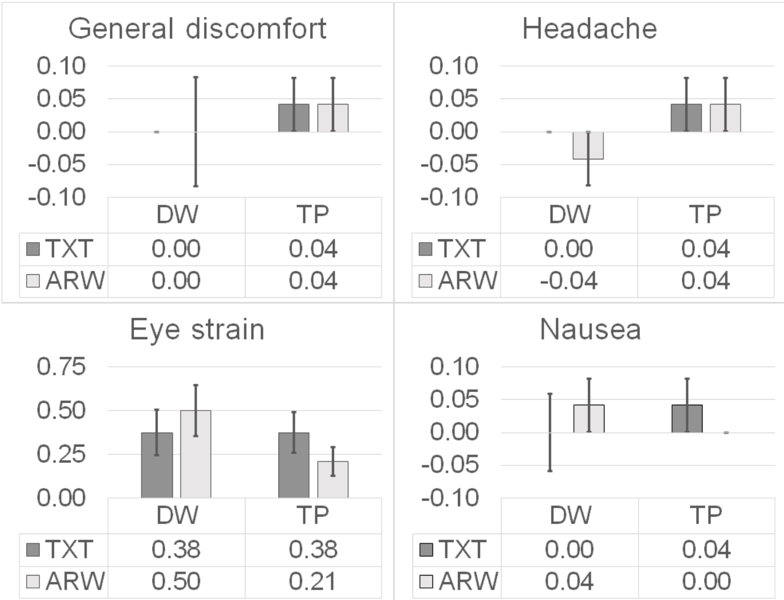


Figure 5. Mean values of the difference between pre-test and post-test scores of general discomfort, headache, eye strain, and nausea (from the selected items of SSQ). Error bars indicate ±SE. *DW*: dwell-based selection; *TP*: touchpad-based selection; *TXT*: item text active area; *ARW*: arrow icon active area.

SSQ item	Main effect of technique			Main effect of active area			Interaction			Post-hoc simple effects (p)			
	F(1,23)	p	η_p^2	F(1,23)	p	η_p^2	F(1,23)	p	η_p^2	DW-TXT vs. DW-ARW	TP-TXT vs. TP-ARW	DW-TXT vs. TP-TXT	DW-ARW vs. TP-ARW
General discomfort	0.32	0.58	0.01	0.71	0.41	0.03	0.71	0.41	0.03	-	-	-	-
Headache	3.44	0.08	0.13	0.13	0.72	0.01	0.13	0.72	0.01	-	-	-	-
Eye strain	0.38	0.54	0.02	0.13	0.73	0.01	0.69	0.41	0.03	-	-	-	-
Nausea	0.12	0.73	0.01	0.12	0.73	0.01	0.99	0.33	0.04	-	-	-	-

Table 2. Results of RM ANOVA for mean values of the difference between pre-test and post-test scores of general discomfort, headache, eye strain, and nausea. Statistical significance is highlighted by an asterisk (alpha level for post-hoc analyses is 0.0125). *DW-TXT*: dwell-based selection technique with item text active area; *DW-ARW*: dwell-based selection technique with arrow icon active area; *TP-TXT*: touchpad-based selection technique with item text active area; *TP-ARW*: touchpad-based selection technique with arrow icon active area.

Usability

Figure 6 shows mean values of usability, and Table 3 shows the results of the statistical analysis. The analysis showed that DW selection was perceived as more usable than TP selection when used with the ARW area.

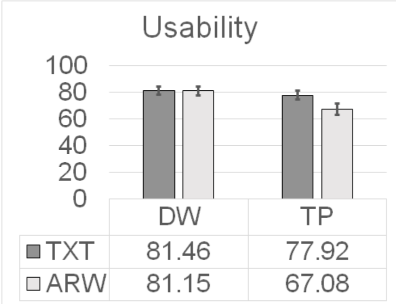


Figure 6. Mean value of SUS usability score. Error bars indicate ±SE. *DW*: dwell-based selection; *TP*: touchpad-based selection; *TXT*: item text active area; *ARW*: arrow icon active area.

Main effect of technique			Main effect of active area			Interaction			Post-hoc simple effects (p)			
F(1,23)	p	η_p^2	F(1,23)	p	η_p^2	F(1,23)	p	η_p^2	DW-TXT vs. DW-ARW	TP-TXT vs. TP-ARW	DW-TXT vs. TP-TXT	DW-ARW vs. TP-ARW
6.38	<0.05*	0.22	3.07	0.09	0.12	7.38	<0.05*	0.24	0.88	0.02	0.52	<0.01*

Table 3. Results of RM ANOVA and simple effect analyses for usability score. Statistical significance is highlighted by an asterisk (alpha level for post-hoc analyses is 0.0125). *DW-TXT*: dwell-based selection technique with item text active area; *DW-ARW*: dwell-based selection technique with arrow icon active area; *TP-TXT*: touchpad-based selection technique with item text active area; *TP-ARW*: touchpad-based selection technique with arrow icon active area.

Device Assessment

Figure 7 shows mean values of required force, smoothness, mental effort, physical effort, ease of accuracy, speed, finger fatigue, wrist fatigue, arm fatigue, shoulder fatigue, neck fatigue, general comfort, and ease of use. Table 4 reports the results of the analysis carried out on each DAQ item.

DAQ item	Main effect of technique			Main effect of active area			Interaction			Post-hoc simple effects (p)			
	F(1,23)	p	η_p^2	F(1,23)	p	η_p^2	F(1,23)	p	η_p^2	DW-TXT vs. DW-ARW	TP-TXT vs. TP-ARW	DW-TXT vs. TP-TXT	DW-ARW vs. TP-ARW
Required force	6.89	<0.05*	0.23	4.85	<0.05*	0.17	5.42	<0.05*	0.19	0.41	0.05	0.95	0.02
Smoothness	8.30	<0.01*	0.27	0.25	0.62	0.01	<0.01	0.95	<0.01	-	-	-	-
Mental effort	0.58	0.45	0.06	2.09	0.16	0.08	7.16	<0.05*	0.24	0.10	<0.01*	0.05	0.05
Physical effort	14.30	<0.01*	0.38	9.17	<0.01*	0.29	5.80	<0.05*	0.20	1.00	0.03	0.08	<0.01*
Ease of accuracy	1.26	0.27	0.05	1.34	0.26	0.06	3.91	0.06	0.15	-	-	-	-
Speed	0.18	0.67	0.01	5.34	<0.05*	0.19	3.78	0.06	0.14	-	-	-	-
Finger fatigue	20.56	<0.001*	0.47	19.09	<0.001*	0.45	19.09	<0.001*	0.45	1.00	0.19	<0.01*	0.02
Wrist fatigue	16.86	<0.001*	0.42	18.59	<0.001*	0.45	18.59	<0.001*	0.45	1.00	0.71	<0.01*	<0.012*
Arm fatigue	52.54	<0.001*	0.70	12.17	<0.01*	0.35	12.17	<0.01*	0.35	1.00	0.59	<0.001*	<0.001*
Shoulder fatigue	47.00	<0.001*	0.67	11.00	<0.01*	0.32	8.33	<0.01*	0.27	1.00	0.16	<0.01*	<0.001*
Neck fatigue	1.31	0.27	0.05	0.07	0.80	<0.01	0.71	0.41	0.03	-	-	-	-
General comfort	7.05	<0.05*	0.24	21.08	<0.001*	0.48	5.72	<0.05*	0.20	0.83	<0.01*	0.81	<0.01*
Ease of use	1.39	0.25	0.06	0.34	0.57	0.02	6.85	<0.05*	0.23	0.18	0.03	0.35	0.04

Table 4. Results of RM ANOVA and, in case of significant interaction, simple effect analyses for each DAQ item. Statistical significance is highlighted by an asterisk (alpha level for post-hoc analyses is 0.0125). *DW-TXT*: dwell-based selection technique with item text active area; *DW-ARW*: dwell-based selection technique with arrow icon active area; *TP-TXT*: touchpad-based selection technique with item text active area; *TP-ARW*: touchpad-based selection technique with arrow icon active area.

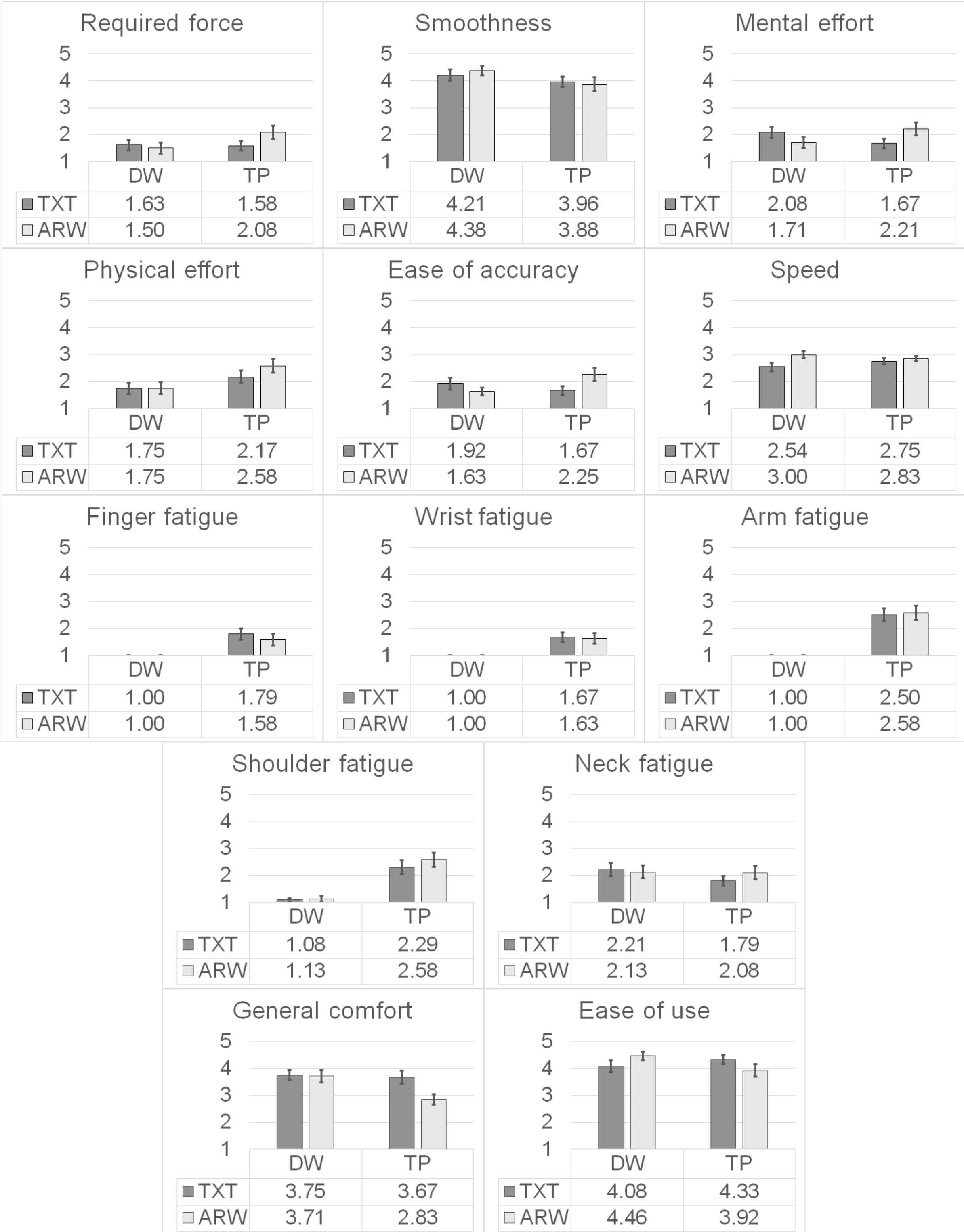


Figure 7. Mean values of required force, smoothness, mental effort, physical effort, ease of accuracy, speed, finger fatigue, wrist fatigue, arm fatigue, shoulder fatigue, neck fatigue, general comfort, and ease of use. Error bars indicate ±SE. *DW*: dwell-based selection; *TP*: touchpad-based selection; *TXT*: item text active area; *ARW*: arrow icon active area.

Results revealed that TP selection required more force than DW selection. Also, the significant interaction indicated that the differences in active area affect DW and TP selection techniques in opposite ways. Pairwise comparisons were not statistically significant, only approaching significance between DW-ARW and TP-ARW conditions, suggesting that TP selection required more force than DW selection when the arrow icon target was used.

The significant main effect of technique observed for smoothness revealed that DW selection was perceived as smoother compared to TP selection.

For mental effort scores, results showed that TP selection required more mental effort when using ARW area than when using TXT area, while no significant differences were observed with DW selection.

For physical effort scores, results indicated that TP selection required more physical effort than DW selection when the ARW area was used. The difference between TP-TXT and TP-ARW conditions only approached significance.

The analyses of finger, wrist, arm and shoulder fatigue scores indicated that TP selection elicited more fatigue than DW selection in the considered body parts when using the TXT area as well as the ARW area (the pairwise comparison between DW-ARW and TP-ARW conditions approached significance for finger fatigue), as also suggested by the main effect of technique observed for the considered scores. For neck fatigue scores, no significant differences could be observed among the experimental conditions.

Finally, the analysis of general comfort scores revealed that participants perceived DW selection as more comfortable than TP selection. In particular, post-hoc analysis revealed that TP-ARW was perceived as less comfortable than both TP-TXT and DW-ARW conditions.

Ranking of Conditions

Figure 8 shows the mean rank for each condition, and Table 5 shows the results of the statistical analysis.

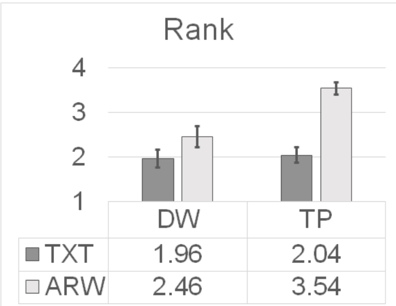


Figure 8. Mean rank for each condition. Ranks range from 1 (best) to 4 (worst). Error bars indicate ±SE. *DW*: dwell-based selection; *TP*: touchpad-based selection; *TXT*: item text active area; *ARW*: arrow icon active area.

Main effect of technique			Main effect of active area			Interaction			Post-hoc simple effects (p)			
F(1,23)	p	η_p^2	F(1,23)	p	η_p^2	F(1,23)	p	η_p^2	DW-TXT vs. DW-ARW	TP-TXT vs. TP-ARW	DW-TXT vs. TP-TXT	DW-ARW vs. TP-ARW
3.58	0.07	0.14	26.83	<0.001*	0.54	8.52	<0.01*	0.27	0.11	<0.001*	0.65	< 0.01*

Table 5. Results of RM ANOVA and simple effect analyses for rank data. Statistical significance is highlighted by an asterisk (alpha level for post-hoc analyses is 0.0125). *DW-TXT*: dwell-based selection technique with item text active area; *DW-ARW*: dwell-based selection technique with arrow icon active area; *TP-TXT*: touchpad-based selection technique with item text active area; *TP-ARW*: touchpad-based selection technique with arrow icon active area.

Results revealed that the TXT area was preferred by participants to the ARW area. Also, the combination of TP selection with the ARW area was considered by participants worse than the same technique with the TXT area, but also worse than DW selection with the ARW area.

Logs

Figure 9 shows mean values of mean selection time, mean time to reach active areas, total task time and number of errors, and Table 6 shows the results of their statistical analyses.

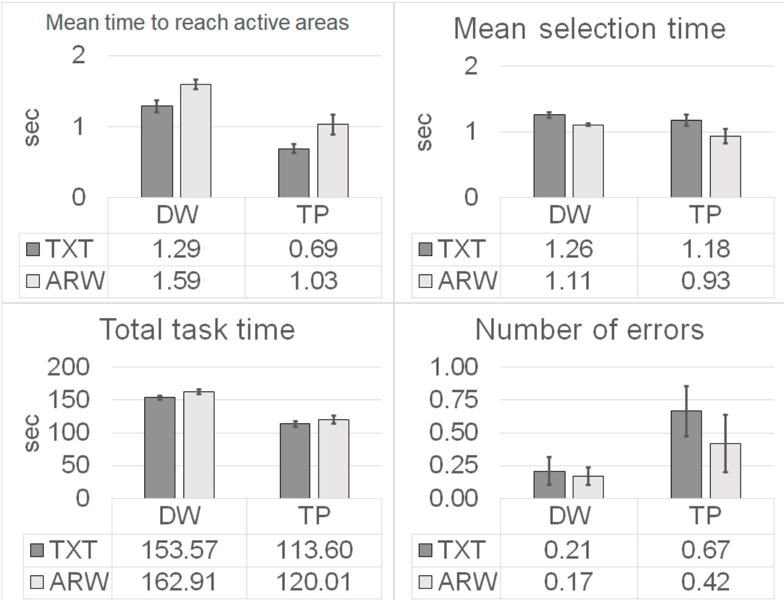


Figure 9. Mean values of mean time to reach active areas, mean selection time, total task time, and number of errors. Error bars indicate ±SE. *DW*: dwell-based selection; *TP*: touchpad-based selection; *TXT*: item text active area; *ARW*: arrow icon active area.

Logs	Main effect of technique			Main effect of active area			Interaction			Post-hoc simple effects (p)			
	$F(1,23)$	p	η_p^2	$F(1,23)$	p	η_p^2	$F(1,23)$	p	η_p^2	DW-TXT vs. DW-ARW	TP-TXT vs. TP-ARW	DW-TXT vs. TP-TXT	DW-ARW vs. TP-ARW
Mean time to reach targets	62.16	<0.001*	0.73	38.69	<0.001*	0.63	<0.01	0.97	<0.01	-	-	-	-
Mean selection time	3.84	0.06	0.14	17.65	<0.001*	0.43	0.25	0.62	0.01	-	-	-	-
Total task time	266.99	<0.001*	0.92	22.31	<0.001*	0.49	0.38	0.55	0.02	-	-	-	-
Number of errors	6.71	<0.05*	0.23	7.39	<0.05*	0.25	4.77	<0.05*	0.17	0.71	0.21	<0.012*	0.30

Table 6. Results of RM ANOVA for mean time to reach targets data, mean selection time data, total task time data, and number of errors data. Statistical significance is highlighted by an asterisk (alpha level for post-hoc analyses is 0.0125). *DW-TXT*: dwell-based selection technique with item text active area; *DW-ARW*: dwell-based selection technique with arrow icon active area; *TP-TXT*: touchpad-based selection technique with item text active area; *TP-ARW*: touchpad-based selection technique with arrow icon active area.

Results revealed that mean time to reach active areas as well as total task time were higher with DW selection than TP selection, regardless of the type of active area. The mean difference between techniques in total task time was 39.97 s (.67 s per trial) and 42.90 s (.72 s per trial) with TXT area and ARW area respectively; the mean difference in time to reach active areas was .60 s and .57 s with TXT area and ARW area respectively. Mean time to reach active areas and total task time were higher with TXT than ARW area regardless of the selection technique. The mean total task time difference between active areas was 9.34 s (.16 s per trial) and 6.41 s (.11 s per trial) with DW and TP selection respectively; the mean difference in mean time to reach active areas between active area types was .31 s and .34 s with DW and TP selection respectively. Mean selection time was higher with TXT than ARW design, regardless of the technique used, with a mean difference of .15 s and .25 s with DW and TP selection respectively. Finally, results showed that participants made more selection errors with TP than DW selection, especially with TXT area.

Discussion

Results showed that participants were slower with DW selection, but they made less selection errors compared to TP selection. Considering subjective data, DW selection was perceived as more usable, comfortable, and generally less fatiguing than TP selection. The analysis of ranking data revealed that TXT area was preferred to ARW area.

Participants' Performance

Total time data showed that participants were significantly slower with DW than TP selection, reinforcing the suggestion of Müller and colleagues (2016) who consider dwell-based approaches generally not suitable for time-critical interactions, e.g., those in the automotive domain. Video games are another example of VR applications in which the use of dwell-based selection must be thoroughly pondered; for example, even a delay of less than a second may disrupt the experience in games that rely on fast interactions, such as time-based puzzle or real-time strategy games. When considering the effects of active area type, we observed that participants were significantly slower with ARW rather than TXT area. This is likely explained by the impact that active area size has on user performance in a pointing and selection task (MacKenzie, 1992): the larger the active area, the lower the time required to select it.

Analyzing participants' performance in more detail, we can explain the significant difference in time to reach active areas between TXT and ARW by considering the difference in spatial relationship between the item text and the associated active area in the two conditions. In both conditions, to decide if an item was the right one to select, participants needed to look at the text, thus hovering the crosshair near it. In the case of TXT, this operation requires hovering the crosshair over the active area, thus increasing selection time rather than time to reach active areas. This was not the case with ARW: when pointing at the text, the crosshair was outside the active area, thus increasing pointing time but not selection time. This hypothesis can also explain why selection time was higher for DW-TXT and TP-TXT conditions. To test this hypothesis, we extracted from the application logs the number of times the crosshair hovered over an active area during each trial. Then, for each participant, we computed the average of this number for each experimental condition. A two-factor RM ANOVA (with selection technique and active area as independent variables) confirmed that participants hovered the crosshair over a higher number of active areas with TXT compared to ARW ($F(1,23)=43.94$, $p<.001$, $\eta_p^2=.66$). Figure 10 shows the mean number of active areas hovered by participants for each condition.

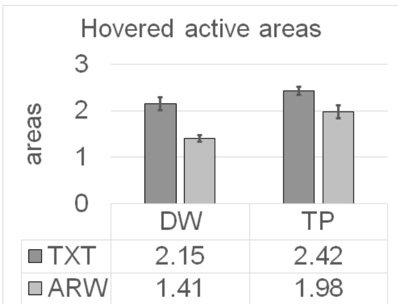


Figure 10. Mean number of active areas hovered by participants for each condition. Error bars indicate ±SE. *DW*: dwell-based selection; *TP*: touchpad-based selection; *TXT*: item text active area; *ARW*: arrow icon active area.

The main effect of selection technique on time to reach active areas is harder to explain, because the action required to perform pointing (i.e., moving the head) was the same across conditions. We can hypothesize that participants were more careful when using DW selection, in order to reduce the probability of inadvertent selections, by avoiding hovering the crosshair over active areas during the pointing phase. Additional analysis of log data seems to support our hypothesis: a two-factor RM ANOVA showed that participants hovered on more active areas with TP than DW selection ($F(1,23)=33.75$, $p<.001$, $\eta_p^2=.60$), resulting in longer pointing time with DW than TP selection (see Figure 10). In the case of TXT, mean time to reach active areas could be higher with DW than TP selection because participants tended to keep the crosshair outside the item for a longer time so that they could read its text without accidentally performing a selection, a precaution that was unnecessary with TP selection. Similarly, in the case of ARW, participants might have tried to avoid inadvertent selections by keeping the crosshair outside active areas as long as possible, even if there was no large cognitive processing involved to comprehend the text to perform the correct selection. The Samsung Gear employed in the study, as well as the other existing consumer HMDs, track only head movements; therefore, eye movement is decoupled from head movement and does not affect pointing and selection. To gather more detailed information about our hypothesis on participants' behavior, an eye-tracking study would be needed.

The hypothesis that participants were more careful when using DW selection compared to TP selection could also explain the unexpected lack of a significant main effect of technique for mean selection time data. According to the literature (Jacob, 1991), dwelling should have introduced a latency in item selection with DW selection, while TP selection should have allowed participants to perform a selection in a fraction of that time, even taking into account the typical 200 ms reaction time to visual stimuli (Bailey, 1996). However, one must consider that the lack of the Midas touch problem with TP selection may have allowed participants, who were asked to be as quick as possible, to move the crosshair over an active area while processing its text, increasing selection time.

Notably, from the (previously discussed) number of active areas hovered by participants, one can derive the number of times a selection operation was canceled: for each participant in each trial, the number of selection cancellations is the number of times (s)he hovered the crosshair over an active area minus 1 (Figure 11). In other words, every time the participant hovered the crosshair over an active area (s)he also hovered the crosshair outside it except for the last time, when selection was finally successful.

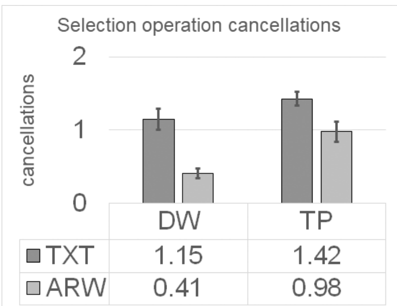


Figure 11. Mean number of selection operations cancelled by participants for each condition. Error bars indicate ±SE. *DW*: dwell-based selection; *TP*: touchpad-based selection; *TXT*: item text active area; *ARW*: arrow icon active area.

One can hypothesize that a higher number of cancellations is correlated to a higher total task time. However, we could observe a significant (negative) correlation only in the DW-ARW condition ($r=-.46$, $p<.05$). Further correlation analyses between cancellations and time to reach active areas and selection time indicated that the higher is the number of selection cancellations, the less is the time spent by participants in reaching active areas (DW-TXT: $r=-.60$, $p<.01$; DW-ARW: $r=-.68$, $p<.001$; TP-TXT: $r=-.58$, $p < .01$, TP-ARW: $r=-.81$, $p<.001$). Furthermore, the higher is the number of selection cancellations, the longer is the selection time (DW-TXT: $r=-.90$, $p<.001$; DW-ARW: $r=-.90$, $p<.001$; TP-TXT: $r=-.68$, $p < .001$, TP-ARW: $r=-.91$, $p<.001$). These correlations suggest that frequent selection cancellations were probably due to participants moving the crosshair in the proximity of active areas (and thus hovering over them multiple times), instead of making deliberately wide head movements that would move the crosshair far from active areas (which would result in a positive correlation between the number of cancellations and time to reach active areas). A two-factor RM ANOVA of the selection cancellation data reported the same results previously discussed for the number of active areas hovered by participants, which was an expected result given the tight relationship between the two variables. For this reason, the same hypothesis we made earlier to explain the results on the number of hovered areas can be applied to explain the significant differences between conditions observed in the selection cancellation data, i.e., participants performed a higher number of selection cancellations with TXT compared to ARW, as well as with TP compared to DW.

Finally, logs showed a higher number of errors when participants were carrying out the task with TP selection compared to DW selection. Such difference was significant with TXT area. It must be noted that the error rates were generally low (from 0.30% in the DW-ARW condition to 1.12% in the TP-TXT condition). The observed result was unexpected, because we believed that the Midas touch problem, even if mitigated by dwell time, should still have caused a high number of unwanted selections. We hypothesize that, in addition to increasing total task time (and pointing time), the greater attention that participants may have paid to pointing and selecting with DW selection could also have helped participants in reducing the number of selection errors. In general, the observed error data suggests that the dwell time we have chosen was more than adequate. Shortening the dwell time may reduce task time at the expense of a higher error rate, as suggested in (Zander et al., 2011). Future studies may explore how different dwell times affect mean selection time and number of errors with DW selection.

Subjective Data

Participants perceived DW selection as more usable and comfortable than TP selection. This result may be explained by the data on perceived higher physical effort and higher finger, wrist, arm and shoulder fatigue with TP selection compared to DW selection, which reflects the fact that TP selection requires participants to use their right arm to select an item, while DW selection allows them to keep the arm in a rest position at all times. Interfaces for VR applications that require a sporadic interaction with the touchpad may be less fatiguing than our task, because they would allow users to rest their arm more frequently, and for more time between interactions. However, one should also consider that there are VR applications that need more frequent pointing and selection operations, e.g., to navigate some immersive 3D experiences (such as Land’s End, described in the Related Work section), users interact with the touchpad with a frequency comparable to the one in our task. As observed by Argelaguet and Andujar (2013), arm fatigue is an important factor that can affect users’ performance in VR tasks. As authors suggest, the design of user interaction with a VR application should take into consideration users’ preferences, physical conditions and desired balance between speed and accuracy.

Finally, the analysis of condition ranking data revealed that participants did not appreciate using ARW area. We expected that such active area would be appreciated more than TXT, especially with DW selection, because the separation of text and active area in items, together with an adequate dwell time, should minimize the risk of accidental selections during the cognitive processing of item text. We can hypothesize that TXT area was preferred over ARW area because bigger active areas are generally faster to point and select (Accot & Zhai, 2003; MacKenzie & Buxton, 1992), and it should be easier for participants to keep the crosshair over the active area of an item. Simple effects analysis also revealed that participants particularly disliked the TP-ARW condition. Indeed, ARW proves useful only in the case of DW selection, as discussed before. This advantage of ARW over TXT area with DW selection comes at the cost of a small head movement, required to move the crosshair over the active area. In the case of TP selection, the separation of text and active area was not needed and, since it still required participants to perform the above mentioned (and, in this case, unnecessary) head movement, it may have affected negatively participants' perception of the condition. Also, these head movements change the touchpad position relative to users' body. Therefore, we can hypothesize that such movements might have made it more difficult for participants to keep their fingers near the touchpad. Since participants cannot see the touchpad while wearing the HMD, this required them to keep physical contact with the right side of the HMD, increasing the risk of accidental touches.

Conclusions

We compared dwell-based and touchpad-based menu selection in mobile HMDs. Participants had to repeatedly point and select one out of four menu items while wearing a Samsung Gear VR. The task selected for the study is representative of the actions users are required to perform when interacting with a wide range of mobile VR applications, from movie players to video games. The dwell-based technique was slower than the touchpad-based technique. However, it led to fewer errors and was perceived as more usable, more comfortable and less fatiguing than the touchpad-based technique. The data collected in our study suggest that the limitations in performance of the proposed dwell-based technique can be attributed to behaviors performed by participants to minimize limitations of dwell techniques, such as the Midas touch problem.

It is important to underline that participants in our study were not frequent HMD users, and thus they represent the large majority of the current population. Our findings suggest that the dwell-based approach is a valid alternative to selection techniques that require to press a button or touch a pad. This is particularly important when considering hands-free uses of HMDs, e.g., in case of upper limb motor disabilities or situations in which users' hands are continuously occupied with other tasks (Zhai et al., 1999).

It must be noted that our evaluation was carried out in the context of a specific device that includes a touch-sensitive surface. Other devices may be equipped with different peripherals; for example, Google Cardboard allows users to interact with applications through a magnetic trigger or, in some versions of the device, a physical button. Furthermore, we only considered two types of active area, which are characterized by different sizes (e.g., see Figure 2). Despite these limitations, our study has practical implications for the design of interfaces for mobile HMDs, because the task we studied is representative of tasks that users have to carry out when interacting with applications for such devices.

An extension of our study could include the evaluation of participants' performance following the ISO 9241-9 standard and Fitts' Law, as in (Lubos et al., 2014; Zhang & MacKenzie, 2007). In particular, by following the recommendations by Soukoreff and MacKenzie (2004) concerning the design of pointing device evaluations, we could improve the comparability and consistency of our study with current as well as forthcoming publications.

It is worth noting that our task used single-word texts to accompany the selectable items, following other studies in the literature (e.g., Bowman & Wingrave, 2001). However, there might be special cases in which an application (e.g., interactive immersive narratives) employs longer (and more complex) text, which can lead to lower user performance, due to the increased time required to read and cognitively process the text displayed by the items. For this reason, future work could thus focus on evaluating the effects of using text with variable length and complexity. This evaluation could also take into consideration different dwell times, because longer and more complex item text may possibly benefit from a longer than 1 s dwell time to avoid the Midas touch problem. Also, as we suggested in the Discussion section, it would be useful to carry out further studies that employ eye-tracking to gather more detailed information on how users behave to avoid inadvertent selections when pointing and selection operations are affected only by head movements, as in the present study.

Finally, although they are not typical of current commercial applications, it would be worthwhile to evaluate alternative item layouts such as radial menus, which have been proposed in the VR literature (e.g., Bowman & Wingrave, 2001).

References

Accot, J., & Zhai, S. (2003). Refining Fitts' law models for bivariate pointing. In *Proceedings of ACM CHI 2003 Conference on Human Factors in Computing Systems* (CHI '03) (pp. 193-200). New York, NY: ACM.

Argelaguet, F., & Andujar, C. (2013). A survey of 3D object selection techniques for virtual environments. *Computer & Graphics, 37*(3), 121-136.

Bailey, R. W. (1996). *Human Performance Engineering: Designing High Quality, Professional User Interfaces for Computer Products, Applications, and Systems* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Bates, R., & Istance, H. O. (2003). Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices. *Universal Access in Information Society, 2*(3), 280-290.

Bee, N., & André, E. (2008). Writing with your eye: a dwell time free writing system adapted to the nature of human eye gaze. In *Proceedings of the 4th IEEE Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems* (PIT '08) (pp. 111-122). Berlin, DE: Springer.

Bohan, M., & Chaparro, A. (1998). To click or not to click: a comparison of two target-selection methods for HCI. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (CHI '98) (pp. 219-220). New York, NY: ACM.

Bowman, D. A., & Wingrave, C. A. (2001). Design and evaluation of menu systems for immersive virtual environments. In *Proceedings of IEEE Virtual Reality 2001* (pp. 149-156). New York, NY: IEEE.

Brooke, J. (1996). SUS: a quick and dirty usability scale. *Usability evaluation in industry, 189*(194), 4-7.

Cmoar (2016). *Cmoar VR Cinema*. Available at https://play.google.com/store/apps/details?id=com.Cmoar.CmoarVirtualCinemaPro&hl=it

Cohen, B. H. (2013). *Explaining Psychological Statistics* (4th ed.). New York, NY: John Wiley & Sons.

Cournia, N., Smith, J. D., & Duchowski, A. T. (2003). Gaze- vs. hand-based pointing in virtual environments. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (pp. 772-773). New York, NY: ACM.

Davis, M. M., Gabbard, J. L., Bowman, D. A., & Gracanin, D. (2016). Depth-based 3D gesture multi-level radial menu for virtual object manipulation. In *Proceedings of IEEE Virtual Reality 2016* (IEEE VR '16) (pp. 169-170). New York, NY: IEEE.

Dias, P., Pinto, J., Eliseu, S., & Santos, B. S. (2016). Gesture interactions for virtual immersive environments: navigation, selection and manipulation. In *Proceedings of the 8th Conference on Virtual, Augmented and Mixed Reality* (VAMR '16) (pp. 211-221). Berlin, DE: Springer.

Douglas, S. A., Kirkpatrick, A. E., & MacKenzie, I. S. (1999). Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (CHI '99) (pp. 215-222). New York, NY: ACM.

DrashVR (2016). *Titans of Space*. Available at http://www.titansofspacevr.com/.

Frees, S., Kessler, G. D., & Kay, E. (2007). PRISM interaction for enhancing control in immersive virtual environments. *ACM Transactions on Computer-Human Interaction, 14*(1).

Hansen, J. P., Johansen, A. S., Hansen, D. W., Itoh, K., & Mashino, S. (2003). Command without a click: dwell time typing by mouse and gaze selections. In *Proceedings of the IFIP Conference on Human-Computer Interaction* (INTERACT '03) (pp. 121-128). Amsterdam, NL: IOS Press.

Huckauf, A., & Urbina, M. A. (2008). Gaze with pEYEs: towards an universal input for various applications. In *Proceedings of the 2008 Symposium on Eye-Tracking Research & Applications* (ETRA '08) (pp. 51-54). New York, NY: ACM.

Jacob, R. J. K. (1991). The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems 9*(3), 152-169.

Jalaliniya, S., Mardanbeigi, D., Pederson, T., & Hansen, D. W. (2014). Head and eye movement as pointing modalities for eyewear computers. In *Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops* (BSN Workshops 2014) (pp. 50-53). New York, NY: IEEE.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.

Kim, N., Kim, G. J., Park, C. M., Lee, I., & Lim, S. H. (2000). Multimodal menu presentation and selection in immersive virtual environments. In *Proceedings of IEEE Virtual Reality* (IEEE VR ‘08). New York, NY: IEEE.

Lin, C. J., Sun, T. L., Chen, H. J., & Cheng, P. Y. (2009). Evaluation of visually-controlled task performance in three dimension virtual reality environment. In *Proceedings of Virtual and Mixed Reality 2009* (VMR ’09) (pp. 465-471). Berlin, DE: Springer.

Lubos, P., Bruder, G., & Steinicke, F. (2014). Analysis of direct selection in head-mounted display environments. In *Proceedings of the 2014 IEEE Symposium on 3D User Interfaces* (3DUI ’14) (pp. 11-18). New York, NY: IEEE.

MacKenzie, I. S. (1992). Fitts’ Law as a research and design tool in Human-Computer Interaction. *Human-Computer Interaction, 7*(1), 91-139.

MacKenzie, I. S., & Buxton, W. A. S. (1992). Extending Fitts' law to two-dimensional tasks. In *Proceedings of ACM CHI 1992 Conference on Human Factors in Computing Systems* (CHI ’92) (pp. 219-226). New York, NY: ACM.

Marajanta, P., Aula, A., & Räihä, K. J. (2014). Effects of feedback on eye typing with a short dwell time. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications* (ETRA ’04) (pp. 139-146). New York, NY: ACM.

Mindprobe (2016). *Cineveo*. Available at http://www.mindprobelabs.com/

Miniotas, D. (2000). Application of Fitts’ law to eye gaze interaction. In *Proceedings of ACM CHI 2000 Conference on Human Factors in Computing Systems* (pp. 339-340). New York, NY: ACM.

Müller, M., Günther, T., Kammer, D., Wojdziak, J., Lorenz, S., & Groh, R. (2016). Smart prototyping – improving the evaluation of design concepts using virtual reality. In S. Lackey and R. Schumaker (Eds.), *Virtual, Augmented, and Mixed Reality*, LNCS 9740 (pp. 47-58). Basel, CH: Springer.

Müller-Tomfelde, C. (2007). Dwell-based pointing in applications of human computer interaction. In C. Baranauskas, P. Palanque, J. Abascal and S. D. J. Barbosa (Eds.), *Human-Computer Interaction – INTERACT 2007*, LNCS 4662 (pp. 560-573). Berlin, DE: Springer.

Ohno, T. (1998). Features of eye gaze interface for selection tasks. In *Proceedings of the Third Asian Pacific Computer and Human Interaction* (APCHI ’98) (pp. 176-181). New York, NY: IEEE.

Piumsomboon, T., Lee, G., Linderman, R. W., & Billinghurst, M. (2017). Exploring natural eye-gaze-based interaction for immersive virtual reality. In *Proceedings of the 2017 IEEE Symposium on 3D User Interfaces (3DUI ’17)* (pp. 36-39). New York, NY: IEEE.

Qian, Y., & Teather, R. J. (2017). 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 ’17) (pp. 91-98). New York, NY: ACM.

Sibert, L. E., & Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. In *Proceedings of the ACM conference on Human Factors in Computing Systems* (CHI ’00) (pp. 281-288). New York, NY: ACM.

Soukoreff, R. W., & MacKenzie, S. I. (2004). Towards a standard for pointing device evaluation, perspective on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies, 61*(6), 751-789.

Steed, A. (2006). Towards a general model for selection in virtual environments. In *Proceedings of the 2006 IEEE Symposium on 3D User Interfaces* (3DUI '06) (pp. 103-110). New York, NY: IEEE.

Steed, A., Friston, S., López, M. M., Drummond, J., Pan, Y., & Swapp, D. (2016). An 'in the wild' experiment on presence and embodiment using consumer virtual reality equipment. *IEEE Transactions on Visualization and Computer Graphics, 22*(4), 1406-1414.

SuperData Research (2018). Market Brief — 2017 Digital Games & Interactive Media Year in Review. Retrieved from https://www.superdataresearch.com/market-data/market-brief-year-in-review/

Tanriverdi, V., & Jacob, R. J. K. (2000). Interacting with eye movements in virtual environments. In *Proceedings of ACM CHI 2000 Conference on Human Factors in Computing Systems* (CHI '00) (pp. 265-272). New York, NY: ACM.

Ustwo Games (2015). *Land's End*. Available at http://www.landsendgame.com/

Ware, C., & Mikaelian, H. H. (1987). An evaluation of an eye tracker as a device for computer input. In *Proceedings of the ACM conference on Human Factors in Computing Systems* (CHI '87) (pp. 183-188). New York, NY: ACM.

Wobbrock, J. O., Findlater, L., Gergle, D. and Higgins, J. J. (2011). The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '11) (pp. 143-146). New York, NY: ACM.

Young, S. D., Adelstein, B. D., & Ellis, S. R. (2006). Demand characteristics of a questionnaire used to assess motion sickness in a virtual environment. In *Proceedings of IEEE Virtual Reality* (IEEE VR '06) (pp. 97-102). New York, NY: IEEE.

Yu, C., Gu, Y., Yang, Z., Yi, X., Luo, H., & Shi, Y. (2017). 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* (CHI '17) (pp. 4479-4488). New York, NY: ACM.

Zander, T. O., Gaertner, M., Kothe, C., & Vilimek, R. (2011). Combining eye gaze input with a brain-computer interface for touchless human-computer interaction. *International Journal of Human-Computer Interaction, 27*(1), 38-51.

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (CHI '99) (pp. 246-253). New York, NY: ACM.

Zhang, X., & MacKenzie, I. S. (2007). Evaluating eye tracking with ISO 9241 – Part 9. In J. Jacko (Ed.) Human Computer Interaction, Part III, HCII 2007, LNCS 4522 (pp. 779–788). Berlin, DE: Springer.

Luca Chittaro is full professor of Human Computer Interaction in the Department of Mathematics, Computer Science, and Physics of the University of Udine, Italy, where he heads the HCI Lab (http://hcilab.uniud.it). He has authored or co-authored over 190 international publications, and he is an ACM Distinguished Speaker.

Riccardo Sioni received his PhD in computer science from the University of Udine. He is a postdoctoral research fellow at the HCI Lab in the Department of Mathematics, Computer Science, and Physics of the University of Udine, Italy.