Evaluation of Text Selection Techniques in Virtual Reality Head-Mounted Displays

Wenge Xu*
DMT Lab, Birmingham City University
Birmingham, UK

Xuanru Meng,† Kangyou Yu‡
Xi'an Jiaotong-Liverpool University
Suzhou, China

Sayan Sarcar§
DMT Lab, Birmingham City University
Birmingham, UK

Hai-Ning Liang¶
Xi'an Jiaotong-Liverpool University
Suzhou, China
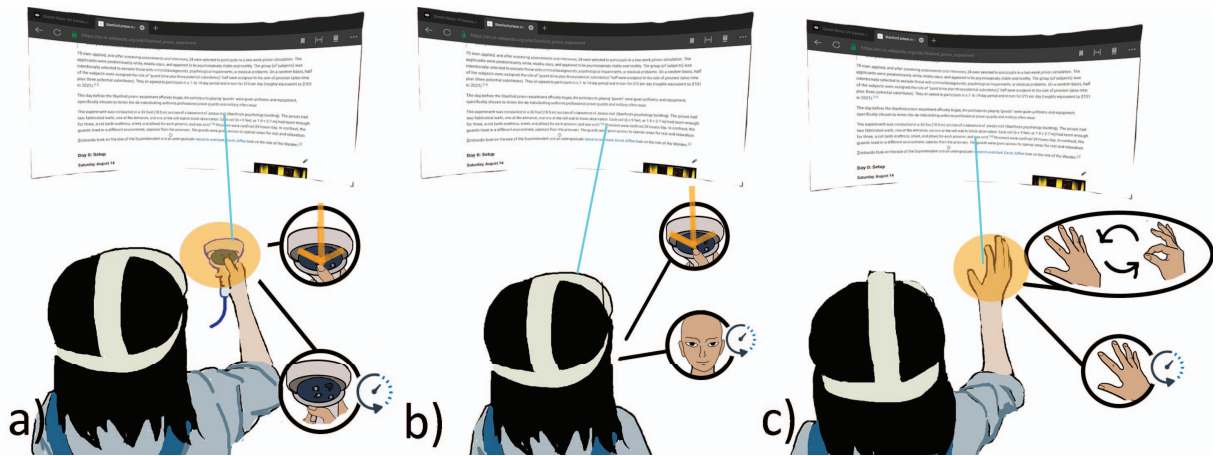
Figure 1: The six text selection techniques explored in this research, grouped by three pointing mechanisms: (a) *Controller*-based: Controller+Click and Controller+Dwell, (b) *Head*-based: Head+Click and Head+Dwell, and (c) *Hand*-based: Hand+Pinch and Hand+Dwell.

ABSTRACT

Text selection is an essential activity in interactive systems, including virtual reality (VR) head-mounted displays (HMDs). It is useful for: sharing information across apps or platforms, highlighting and making notes while reading articles, and text editing tasks. Despite its usefulness, the space of text selection interaction is underexplored in VR HMDs. In this research, we performed a user study with 24 participants to investigate the performance and user preference of six text selection techniques (Controller+Dwell, Controller+Click, Head+Dwell, Head+Click, Hand+Dwell, Hand+Pinch). Results reveal that Head+Click is ranked first since it has excellent speed-accuracy performance (2nd fastest task completion speed with 3rd lowest total error rate), provides the best user experience, and produces a very low workload—followed by Controller+Click, which has the fastest speed and comparable experience with Head+Click, but much higher total error rate. Other methods can also be useful depending on the goals of the system or the users. As a first systematic evaluation of pointing×selection techniques for text selection in VR, the results of this work provide a strong foundation for further research in this area of growing importance to the future of VR to help it become a more ubiquitous and pervasive platform.

*e-mail: wenge.xu@bcu.ac.uk

†e-mail: xuanru.meng18@student.xjtlu.edu.cn

‡e-mail: kangyou.yu18@student.xjtlu.edu.cn

§e-mail: sayan.sarcar@bcu.ac.uk

¶e-mail: haining.liang@xjtlu.edu.cn (*corresponding author*)

Keywords: Text Selection, Virtual Reality, Pointing Methods, Selection Mechanisms, User Study

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—interaction techniques; Human-centered computing—Interaction design—Empirical studies in interaction design

1 INTRODUCTION

Text selection is essential when reading text content such as newspapers, magazines, and academic papers to highlight important elements for later reference or to copy/cut and transfer the content to another document, application, or platform. Text selection has been well-researched for PCs (e.g., using the mouse [4, 27]) and, to a lesser extent, for touchscreen devices (e.g., using a touch-based text handler in smartphones or tablets). More recently, there have been some initial explorations with augmented reality (AR) head-mounted displays (HMDs) (e.g., using a smartphone for text selection in AR HMDs [7]). However, these methods are not suitable for virtual reality (VR) HMDs because (1) traditional input devices like the mouse require a flat surface to operate on [42], which is not accessible or natural to VR users, (2) using a touchscreen device or smartphone as a dedicated input device for VR HMDs is not practical and may not lead to good performance and usability, (3) these external devices are not visible for VR users.

Today's VR HMDs enable users to chat with friends (e.g., via VR Facebook Messenger), watch movies/TV shows, read text-based content (e.g., news, blogs, papers), and even work in a virtual office (e.g., Horizon Workrooms by Facebook) [13, 35]. These activities typically involve selecting text fragments frequently. For instance, people use text selection for (1) sharing information across plat-

forms or applications (newspaper apps, Tripadvisor, IMDB) with their friends when exchanging instant text messages, (2) searching a movie review on IMDB before watching it, (3) making notes while reading academic papers or books, and (4) text editing [7]. Despite the increasing need for text selection tasks in VR HMDs, to our knowledge, there has been no study that has explored text selection techniques for these devices, especially based on common pointing and selection mechanisms available in current commercial VR HMDs.

There is some research on standard mid-air pointing tasks (e.g., object acquisition task [3, 22, 56] and virtual keyboard-based text entry task [41, 48]). However, text selection is very different from these tasks because (1) it requires users to hold the selection mechanism until they finish hovering over the target sentence (more precise control); (2) the density of text information is always high during the task: letters and sentences tend to be close to one another; (3) the size of the letters are very small for selection compared to objects in other typical selection tasks. Nevertheless, like other mid-air pointing tasks, text selection in VR HMDs requires (1) a mechanism for the identification of the objects to be selected, and (2) some signal or command to indicate their selection [32].

At present, *Controller*-based pointing (see Fig. 1a) [19,38,54] and *Head*-based pointing (see Fig. 1b) [50, 52] are the most commonly used pointing methods in commercial VR HMDs for identifying a target object that needs to be selected. Controller-based pointing has become the dominant interaction method because it (1) offers the highest throughput and best accuracy, and is often preferred by most users [20], (2) does not cause serious hand/arm fatigue, (3) provides rich functionalities when operating with the pointing device's built-in buttons. Head-based pointing (1) is important when controllers cannot be tracked, are not around, or has no power [29, 50], and (2) represents a fitting alternative for people with hand mobility or stability issues. With recent developments of high accuracy and low latency hand tracking techniques, a third pointing method—*Hand*-based (palm) ray-casting pointing has gained rapid attention and has been integrated into several VR HMDs (such as the Quest and HTC VIVE series). For instance, Quest 2 renders a ray cast from the point between the user's thumb and index finger to the virtual environment to serve as a pointing mechanism (see Fig. 1c).

There are three primary selection mechanisms in VR HMDs to allow users to express selection—(1) *Dwell*: using a dwell time on the target, (2) *Click*: clicking a button on the controller, and (3) *Pinch*: using one hand to perform a pinch gesture [33, 49]. In this research, we conduct a study with 24 participants to evaluate the performance and usability of 6 text selection techniques (i.e., Controller+Dwell, Controller+Click, Head+Dwell, Head+Click, Hand+Dwell, Hand+Pinch) [1] for VR HMDs based on the 3 standard pointing methods (Controller, Head, Hand) and the 3 primary selection mechanisms (Dwell, Click, Pinch). Our results suggest that Head+Click and Controller+Click are the leading candidates where Head+Click should be considered first since it has excellent performance (2nd in speed with a relatively low total error rate), provides the best user experience, and produces a very low workload. Controller+Click has the best speed and a comparable workload as Head+Click, but because it has slightly lower user experience scores and a worse total error rate, it should be considered as a second option.

The main contributions of this work include: (1) *A first evaluation* of six standard text selection techniques for VR HMDs regarding performance and user preference; and (2) a set of usage recommendations that are derived from our experiment results and observations during the experiment.

---

[1] The reasons for excluding the other 3 possible combinations are provided in Sect. 3.4.

2 RELATED WORK

2.1 Text Selection in AR/VR HMDs

To the best of our knowledge, text selection has only been done for AR HMDs. For instance, EYEditor [11] uses a ring mouse for cursor navigation and text selection, where a button is used for placing the cursor before and after the text fragment to be selected while the selection is made via a touchpad. Lee et al. [26] have employed a force-sensitive smartphone as their input device, where users exert a force on a thumb-sized circular button to select the desired text fragment. Similarly, Darbar et al. [7] have explored the use of a smartphone as the input mechanism for text selection in AR HMDs. They found that continuous touch is more efficient than discrete touch, spatial movement, and ray casting. This prior research focusing on AR all employed an external input device (such as a ring mouse or smartphone) for accomplishing text selection, which is not feasible or practical for commercial VR HMDs.

Text selection in VR differs from AR in several aspects. (1) *VR is less problematic and complex*: text selection in VR is less of a problematic task than in AR, as there are no layer interference, color blending, and layout foreground-background issues [48]. As such, text can be presented in a legible and readable manner to users [23]. (2) *VR has more input possibilities to choose from*: the primary pointing methods such as *Controller*-based pointing, *Head*-based pointing, and *Hand*-based pointing are more accessible for VR users than for AR users. For instance, today's VR HMDs often come with a dedicated controller for input, while this is not always the case for AR HMDs (in fact, only the Magic Leap 1 comes with a hand-held controller). Finally, (3) *VR as the future work/office space*: based on current developments of AR/VR technology, text selection in text editing and document preparation scenarios is more practical and feasible in VR for possible future office settings (e.g., VR office environments based on immersive HMDs [13, 35]). Given the above, this study explores accessible interaction techniques that are available in commercial VR HMDs for text selection.

Our work mainly focuses on user-driven text selection, where users have to indicate both the start and the end of the text fragment. We did not focus on automatic text selection, such as the double-click technique for word selection and the triple-click technique for the whole paragraph text selection derived from desktop computers with a mouse. Instead, we decided to explore user-driven text selection because it is a more suitable starting point for text selection studies, as it covers all types of text selection levels (e.g., character, word, sub-word [7]). On the other hand, automatic text selections are limited to word-, sentence-, or paragraph-level selection. This is in line with early text entry studies where character-level input is first explored then the swipe-/gestural-based (i.e., word-level) input followed after.

2.2 Mid-air Pointing-based Interaction in HMDs

There are three main interaction approaches in VR HMDs to allow users to identify virtual objects—*Controller*-based, *Head*-based, and *Hand*-based. Below, we present a review of these pointing methods and potential selection mechanisms that can be used to complete a mid-air interaction [32].

Controller-based. Like the mouse in desktop systems, the controller has become a standard input device for most HMDs (e.g., Quest series, HTC VIVE series, Magic Leap series) [19, 38, 54]. In this method, a ray is cast from the controller to the virtual environment to serve as a pointing mechanism. The end of the ray is akin to a cursor to assist in the object's identification. Users need to point to the target object by rotating or moving the controller mid-air and, once the cursor is on the desired object, the selection is often achieved by clicking a button on the controller [19]. When clicking the button becomes a problem (e.g., for very small objects or an object located in very close proximity to other objects), a dwell time can be used as the selection mechanism [48].

Head-based. Head-based interaction has been actively studied in HMDs [1, 6, 19, 28, 49]. It has been widely adopted as a standard way for pointing at virtual objects without using hands or hand-held pointing devices [24]. Selection for head-based pointing often relies on a dwell time [29, 33, 52], a second modality such as clicking a button on a controller [24, 52], or making a gesture using bare hands [33, 49]. One prior study in AR that empirically compares selection methods shows that Dwell causes a lower error rate than clicking a button on a keyboard and a hand-held button device, voice (verbally confirming) [30]. A recent study in VR also shows that Dwell leads to a significantly lower error rate than button clicking and hand gesture (i.e., pinch) [33].

Hand-based. Hand-based interaction is now possible in several VR HMDs (e.g., Quest series, HTC VIVE series). To select a nearby object [32, 54], users first hover their hand over the target object and then make a selection by performing a hand gesture—e.g., a pinch in the Oculus Quest 2. For a distant object (e.g., keys on a keyboard or items on a menu), users use hand-based pointing to point at the object followed by a hand gesture for selection. For instance, the Quest 2 renders a ray cast from the point between the user's thumb and index finger to the virtual environment, and the selection of the object can be made via a pinch gesture. In scenarios when performing gestures can lead to errors, users can use dwell to prevent these errors.

Although these three mid-air interaction methods have been widely studied in HMD systems [33, 48, 49, 52], their performance and usability for text selection tasks are underexplored. Due to the increasing importance of these tasks, understanding how these methods affect user performance and usability is crucial to the development of efficient and usable techniques for text selection. Therefore, this research aims to fill this gap and has conducted the first systematic study of these pointing mechanisms in combination with their suitable selection mechanisms for text selection in VR HMDs.

There are some other novel selection mechanisms such as eye blinking–intentionally blinking the eye(s) [28, 29], voice input–verbally issuing a command [51], and sound volume–increasing the sound volume [16] can also be used. They have been explored in the context of hands-free interaction in a recent work [31].

3 EVALUATED METHODS

This section describes how the final six combinations from the three Pointing Methods (Controller, Head, Hand) and three Input Mechanisms (Dwell, Pinch, Click) are implemented and operationalized in our study and why Hand+Click, Head+Pinch, and Controller+Pinch are not considered. Table 1 summarizes the advantages and disadvantages of our six text selection methods. All six techniques were developed and run in Unity3D (Unity v2019.4.10f1) with default pointing methods and input functions provided by the Oculus Unity Plugin (OVRPlugin v1.61.0).

3.1 Controller-based Pointing

A ray is rendered from the controller to the virtual environment to serve as a pointing mechanism.

Controller+Dwell (ConD). The user starts the selection of the desired text fragment by dwelling (staying on an area for 1s) at the beginning of the first letter and ends the selection by dwelling again at the end of the text. The dwell duration was decided through a series of pilot studies testing a range of thresholds from 400ms to 1s (from text entry studies [14, 28, 36, 50]).

Controller+Click (ConC). The user needs to press and hold the main button in the controller at the beginning of the target letter to start the selection and then release the button after the cursor reaches the end of the target text to complete the selection.

3.2 Hand-based Pointing

A ray is rendered from the point between the user's thumb and index finger to the virtual environment to serve as a pointing mechanism. For this research, we upgraded the tracking rate to 60Hz to improve the hand tracking performance (i.e., lower latency for hand-based pointing and higher accuracy for pinch gesture recognition). In addition, we controlled the lighting of the physical environment to ensure the hand tracking was reliable and consistent during the study.

Hand+Dwell (HandD). HandD is analogous to ConD, but the pointing is done through Hand-based pointing.

Hand+Pinch (HandP). The user needs to use their hand to point at the beginning of the target text, perform the "pinch and drag" gesture to start the selection, and then release the pinch gesture at the end of the target text fragment to end the selection.

3.3 Head-based Pointing

A ray is extended from the HMD towards the viewing direction into the virtual environment.

Head+Dwell (HeadD). HeadD is analogous to ConD, with the pointing achieved by Head-based pointing.

Head+Click (HeadC). Selection is analogous to ConC, but the pointing is done through Head-based pointing.

3.4 Other Possible Combinations that Are Not Used

We did not include Hand+Click and Controller+Pinch because the existing literature suggests that one-handed interaction is often preferred by VR users over using two hands for interacting with objects in VR [34]. This is especially the case for tasks that can be done with one hand, as one-handed interaction is less physically and mentally demanding [45] and it offers numerous benefits (e.g., allowing the other hand to hold other items or performing other tasks [10, 21]). As such, this research has focused primarily on using one-hand.

Head+Pinch is excluded because: (1) our pilot study suggests that users are more likely to make errors as they frequently move their hand unintentionally outside the ideal hand interaction area of the device when their hand is used as a peripheral input modality [47], (2) it leads to higher fatigue compared to Head+Click, which is already included as one of the evaluated techniques, and (3) it is disliked by users when it is used as a selection mechanism compared to Click and Dwell [33].

3.5 Testbed Environment

Figure 2 shows the test environment. An 'instruction panel' is located on the left side where the participant could see the text that needed to be selected. The 'interaction panel' is located at the center, where the participant needed to use each technique to select text fragments. In addition, two buttons are provided: 'Delete' for deleting the wrong selection and 'Next' for moving to the next trial.

The following parameters are set based on the recommendations from previous studies and then further tested and agreed upon by 5 users from a pilot study. We controlled the length of the materials to be between 9-12 lines, with each line having around 40 characters with spaces [44], in both panels. The plane is set as 2.6m which was tested and agreed by target users, 2.6m also falls in the recommended reading distance suggested by [8]. For the text style, we used Sans-serif Arial with a light color [8]. Angular size was set as 1.8°, which was within the recommended range suggested by [8].

Visual support and feedback were provided in three ways: (1) The end of the ray is akin to a cursor, (2) changing the color of the selected text to yellow, (3) changing the color of the cursor when a selection was started/stopped, (4) a visual indicator is provided for showing dwell progress. We did not visualize the ray because users from our pilot studies suggest the cursor alone is more effective than a combination of visualizing the ray and the cursor in helping them understand where they are pointing.

Table 1: Summary of the advantages and disadvantages of the six text selection methods used in our study.

| Technique	Advantages	Disadvantages
Controller+Dwell	(1) easy to control, (2) minimal hand fatigue, (3) less error-prone [33]	(1) not device-free, (2) slow due to dwell time and making users feel 'pushed' [29]
Controller+Click	(1) easy to control, (2) minimal hand fatigue, (3) user familiarity (mouse-like interaction), (4) tactile feedback	(1) not device-free
Hand+Dwell	(1) device-free, (2) less error-prone [33], (3) natural interaction	(1) slow due to dwell time and making users feel 'pushed' [29], (2) hand fatigue, (3) limited hand-interaction area as a result of a limited hand tracking area that can be captured via the front-facing cameras [47]
Hand+Pinch	(1) device-free, (2) natural interaction	(1) hand fatigue, (2) limited hand-interaction area [47]
Head+Dwell	(1) device-free, (2) hands-free, (3) less error-prone [33]	(1) slow due to dwell time and making users feel 'pushed' [29], (2) potential cybersickness [53]
Head+Click	(1) hand mobility does not affect the selection, (2) minimal hand fatigue, (3) tactile feedback	(1) potential cybersickness [53], (2) not device-free
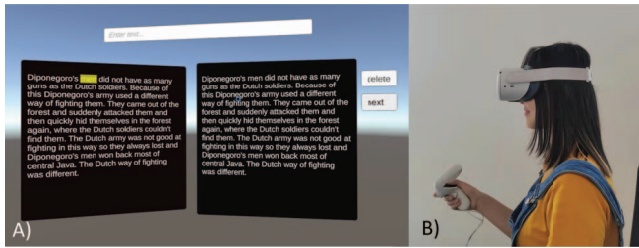
Figure 2: A) We used the same experiment environment for all conditions. An 'instruction panel' is located on the left side, slightly tilted towards the user. The 'interaction panel' is located in the center, slightly tilted towards the user. B) A picture of the experiment setting showing how a participant is performing a text selection task.

4 EXPERIMENT

4.1 Participants and Apparatus

We recruited 24 unpaid participants (12 males; 12 females) between the ages of 18-30 (M=21, SD=2.06) from a local university campus through a database of participants. They all had no issues reading the content clearly within the VR environment and no difficulties moving their arms, hand, and heads. 23 participants were right-handed. 17 participants had previous experience with VR HMDs, but only 3 were regular VR users (weekly) and none of them had interacted with the Oculus Quest 2—the device used in this experiment. Participants were seated on the office chair during the experiment and were allowed to rest their hands on the chair handles if needed (see Fig. 2B).

4.2 Design

The experiment followed a one-way within-subjects design with Interaction Techniques (HandD, HandP, HeadD, HeadC, ConD, and ConC) as the independent variable. For each condition, participants needed to complete 3 training trials (1 short, 1 medium, and 1 long text fragments) and 27 trials (9 short, 9 medium, and 9 long text fragments following a randomized order of appearance) which were randomly sampled from a corpus of standardized English reading assessment [37]. Each selection target would only appear once in a specific condition. The order of the conditions was counterbalanced across participants to avoid learning effects with Sentence Lengths also randomized. Excluding the training texts, we collected 3888 trials (24 participants × 6 interaction techniques × 27 texts). Although the local area had no local COVID-19 cases for 12 months before the experiment, we cleaned and sanitized the device before and after each participant's turn and followed extra safety measures

to ensure the safety of the participants and researchers (e.g., wearing a mask and staying at a safe distance).

For assessing the text selection performance and experience, we collected the following measurements:

- *Objective*: (1) Task-completion time: in line with the previous studies in pointing experiments, we only report task completion time where there were no errors during the trials to handle the speed-accuracy trade-off. The task completion time for each trial is defined as the time from when the cursor first hovers over the first target letter to the time they complete the correct selection. By recording in this way, the time spent on clicking the button, performing the pinch, and dwelling are included for analysis. (2) Total error rate: (the number of wrong sentences + the number of deletions) / total number of attempts, (3) Not corrected error rate: the number of wrong sentences / total number of sentences.

- *Subjective*: NASA-TLX questionnaire [15] to measure workload, User Experience Questionnaire (UEQ) [25] to measure user experience, and comments on advantages and disadvantages of each technique plus users' ranking of the text selection techniques.

4.3 Procedure

Before the experiment began, participants were told the goal of the research and the experimental procedure. Then, they needed to fill out a demographic questionnaire (e.g., age, gender, and experience with VR) and signed the consent to participate in the experiment. In addition, they were told to finish the trials as fast and as accurately as they could. Error correction was allowed by using the delete button in the VR scene.

Before each condition started, the corresponding text selection method was explained to the participants, who then had a practice session with three warm-up selections before the experiment stage (27 text selections). After each condition, participants needed to fill out a post-condition questionnaire (NASA-TLX and UEQ). They proceeded to the next condition when they felt rested to avoid any fatigue. Once they completed all conditions, they needed to complete a post-experiment questionnaire and a structured interview. The whole experiment lasted around 70 minutes for each participant.

4.4 Results

We used Shapiro-Wilks tests and Q-Q plots to check if the data had a normal distribution. All tests are with two-tailed p values. Effect sizes were reported whenever feasible (η_p^2). *Performance analysis*. For normally distributed data, we employed two-way repeated-measures ANOVAs with Interaction Techniques (six techniques)

Table 2: Performance data for each Interaction Technique among three Sentence Lengths, mean (SD). The top 3 techniques of each condition are presented by Roman numerals (I: light green ; II: darker light green ; and III: blue-green).

Performance Metrics	Sentence Lengths	ConC	ConD	HandP	HandD	HeadC	HeadD
Task-completion time	Small	I: 1.44 (0.51)	III: 3.08 (1.75)	4.49 (3.79)	3.86 (1.23)	II: 1.69 (0.35)	3.2 (0.87)
	Medium	I: 2.11 (0.65)	III: 3.28 (0.71)	5.20 (2.64)	4.96 (1.65)	II: 2.52 (0.66)	4.38 (1.87)
	Long	I: 2.39 (0.63)	III: 3.71 (0.70)	5.22 (2.25)	5.20 (1.94)	II: 2.92 (0.72)	4.98 (1.47)
TER	Small	25.7% (18.1%)	I: 9.4% (9.3%)	32.9% (19.0%)	25.5% (17.8%)	II: 11.5% (13.8%)	III: 12.3% (13.3%)
	Medium	18.1% (13.3%)	I: 8.3% (13.0%)	28.2% (17.8%)	22.3%(14.4%)	III: 10.8% (10.9%)	II: 8.7% (10.3%)
	Long	20.7% (13.4%)	I: 6.1% (9.9%)	30.7% (15.7%)	21.9% (14.3%)	III: 12.0% (13.1%)	II: 10.8% (9.3%)
NCER	Small	1.6% (4.9%)	II: 1.2% (3.3%)	3.4% (7.7%)	1.9% (4.6%)	I: 0.0% (0.0%)	II: 1.2% (3.4%)
	Medium	III: 1.3% (3.9%)	1.6% (4.5%)	3.7% (8.2%)	1.4% (3.2%)	I: 0.4% (2.0%)	II: 0.9% (4.4%)
	Long	I: 1.6% (3.3%)	2.4% (6.0%)	4.2% (9.4%)	3.6% (7.1%)	II: 2.0% (4.8%)	III: 2.2% (5.8%)

and Sentence Lengths (short, medium, long) as the within-subjects variables. For data that were not normally distributed, we processed the data through Aligned Rank Transform (ART) [46] before using repeated-measure ANOVAs with the transformed data. Bonferroni corrections were used for all pairwise comparisons.

Experience analysis. For normally distributed data, we employed one-way repeated-measures ANOVAs with Interaction Techniques as the within-subjects variable. For data that were not normally distributed, we conducted non-parametric Friedman tests for the ranking. Bonferroni correction was used for pairwise comparisons and Greenhouse-Geisser adjustment was used for degrees of freedom if there were violations of sphericity.

4.4.1 Performance

Task Completion Time. To handle the speed-accuracy trade-off for task completion time, we followed the analysis process from pointing experiments where trials with errors are excluded [43,55,57]. In total, we collected 3888 trials (24 participants × 6 interaction techniques × 27 texts) besides training trials. To analyze selection time, we discarded trials in which participants made a wrong selection (709 error trials, 18.2%), and removed outliers (65 trials, 1.7%), in which the selection time was more than three standard deviations from the mean (mean ± 3std.) in each condition. In total, there were 3179 success trials used for analyzing the task completion time.

There was a statistically significant main effect of Interaction Techniques on task completion time ($F_{5,391} = 145.464, p < .001, \eta_p^2 = .590$). Post-hoc analysis with Bonferroni correction suggested that (1) ConC outperformed all other techniques (all $p < .001$), (2) HeadC outperformed all techniques except ConC (all $p < .001$), (3) ConD outperformed HandD, HandP, HeadD (all $p < .001$). Table 2 shows the mean task completion time among the 6 Interaction Techniques across 3 Sentence Lengths.

There was a significant main effect of Sentence Lengths on Task Completion Time ($F_{2,391} = 68.972, p < .001, \eta_p^2 = .222$). Post-hoc analysis with Bonferroni correction suggested that (1) participants completed Small faster than Medium and Long (both $p < .001$), and (2) participants also completed Medium faster than Long ($p < .001$). We did not find any interaction effect of Sentence Lengths × Interaction Techniques ($F_{10,391} = 0.981, p = .459, \eta_p^2 = .019$).

Total Error Rate (TER). There was a statistically significant main effect of Interaction Techniques on TER ($F_{5,391} = 34.391, p < .001, \eta_p^2 = .270$). Post-hoc analysis with Bonferroni correction suggested that (1) ConC outperformed HandP ($p < .05$), (2) ConD outperformed ConC, HandD, and HandP (all $p < .001$), (3) HeadC outperformed ConC, HandD, and HandP (all $p < .001$), (4) HeadD outperformed ConC, HandD, HandP (all $p < .001$). There was no significant main effect of Sentence Lengths ($F_{5,391} = 2.367, p = .095, \eta_p^2 = .010$) and the interaction effect of Interaction Techniques
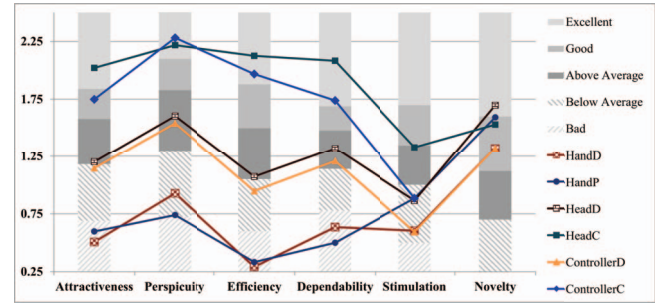

Figure 3: UEQ subscale ratings of the tested methods with respect to comparison benchmarks.

times Sentence Lengths ($F_{10,391} = 0.406, p = .944, \eta_p^2 = .008$). Details of TER data can be found in Table 2.

Not Corrected Error Rate (NCER). Table 2 presents the NCER data for each Interaction Technique. There were significant effects of Interaction Techniques ($F_{5,391} = 8.217, p < .001, \eta_p^2 = .073$), Sentence Lengths ($F_{5,391} = 6.455, p < .01, \eta_p^2 = .024$), and Interaction Techniques *times* Sentence Lengths ($F_{10,391} = 2.524, p < .01, \eta_p^2 = .046$). Post-hoc analysis with Bonferroni corrections was conducted based on the interaction effect, however, we could not find any significant difference.

4.4.2 User Preference

UEQ. For the average UEQ scores, HeadC was rated the best (M=1.89, SD=0.77) and HandD (M=0.72, SD=1.26) the worst. ANOVA tests showed a significant main effect of Interaction Techniques ($F_{3.388,77.935} = 8.622, p < .001, \eta_p^2 = .273$) on the UEQ mean score. Post-hoc pairwise comparisons suggested that (1) HeadC had a significantly higher average UEQ score than HandD ($p < .01$), HandP ($p < .001$), and ConD ($p < .01$), (2) ConC had a significantly higher average UEQ score than HandD ($p < .01$) and HandP ($p < .05$).

Regarding the UEQ subscales, ANOVA tests yielded significant main effects of Interaction Techniques on all UEQ subscales. Table 3 shows the detailed results of the ANOVA tests. Details of each UEQ subscale score can be found in Figure 3. HeadC was rated above average to excellent (mainly good to excellent except for only one participant who rated it above average), ConC was rated below average to excellent (mainly good to excellent except for one participant who rated it below average). At the same time, ConD and HeadD were comparable with the benchmark, and HandD, and HandP had worse scores than the benchmark.

Workload. For overall task workload, ConC was rated the best (M=23.64, SD=19.32) and HandP (M=54.43, SD=20.72) the

135

Table 3: ANOVA test results for the UEQ subscales. Significant results where $p < .05$ are highlighted with light green and $p < .001$ with dark green.

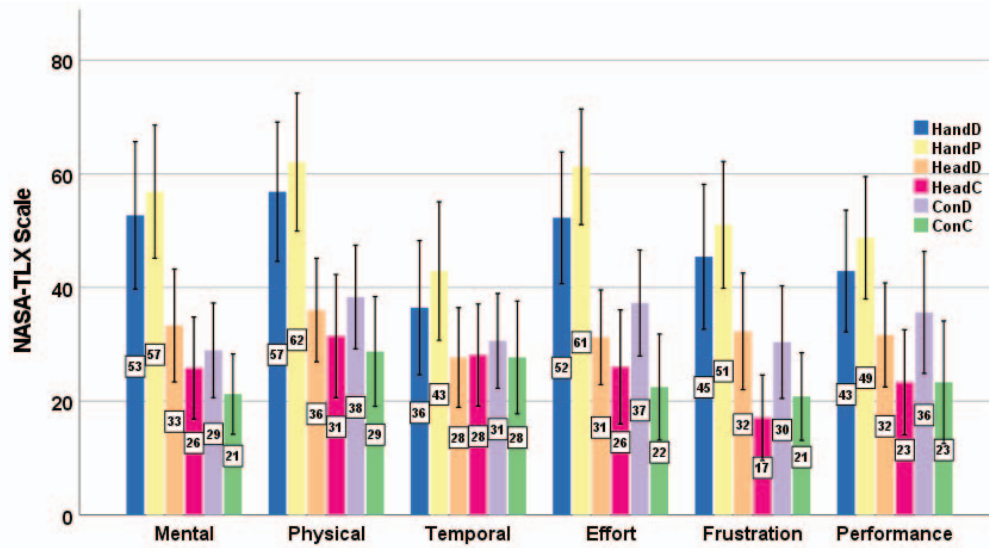|	Interaction Technique	Post-hoc
Attractiveness	$F_{3.307,76.066} = 8.405, p < .001, \eta_p^2 = .268$	HandD<HeadC ($p < .01$), HandD<ConC ($p < .01$), HandP<HeadC ($p < .01$), HandP<ConC ($p < .05$), ConD<HeadC ($p < .01$)
Perspicuity	$F_{5,115} = 12.631, p < .001, \eta_p^2 = .354$	HandD<HeadC ($p < .01$), HandD<ConC ($p < .001$), HandP<HeadC ($p < .001$), HandP<ConC ($p < .001$), HeadD<HeadC ($p < .05$), ConD<HeadC ($p < .05$), ConD<ConC ($p < .05$)
Efficiency	$F_{5,115} = 15.801, p < .001, \eta_p^2 = .407$	HandD<HeadC ($p < .001$), HandD<ConC ($p < .001$), HandP<HeadC ($p < .001$), HandP<ConC ($p < .001$), HeadD<HeadC ($p < .05$), ConD<HeadC ($p < .01$) ConD<ConC ($p < .05$)
Dependability	$F_{3.728,85.746} = 11.077, p < .001, \eta_p^2 = .325$	HandD<HeadC ($p < .01$), HandD<ConC ($p < .01$), HandP<HeadC ($p < .001$), HandP<ConC ($p < .01$), ConD<HeadC ($p < .01$)
Stimulation	$F_{3.132,72.036} = 2.817, p < .05, \eta_p^2 = .109$	HeadD<HeadC ($p < .05$), ConD<HeadC ($p < .01$)
Novelty	$F_{3.525,81.073} = 5.835, p < .001, \eta_p^2 = .202$	ConC<HeadD ($p < .01$), ConC<HeadC ($p < .05$)

Figure 4: The mean responses for the 6 components of the NASA TLX questionnaire. Error bars indicate 95% confidence interval.

worst.An ANOVA test showed a significant main effect of Interaction Techniques ($F_{3.168,72.871} = 12.352, p < .001, \eta_p^2 = .349$) on the overall task workload ratings. Post-hoc pairwise comparisons indicated that (1) users experienced less workload when use HeadC than HandD ($p < .05$) and HandP ($p < .001$), (2) users experienced less workload in ConC than HandD ($p < .01$) and HandP ($p < .001$), (3) users experienced more workload in HandP than HeadD ($p < .01$) and ConD ($p < .05$). Regarding each NASA-TLX workload subscale, ANOVA tests yielded significant main effects of Interaction Techniques among all subscales. Details of the results of the ANOVA tests can be seen in Table 4 and of the workload subscales in Fig. 4.

Ranking. The ranking of conditions shows a preference for HeadC (11 ranked it first while 9 ranked it second) and ConC (11 ranked it first and 7 ranked it second) among the techniques. They were followed by ConD (10 ranked it third while 8 ranked it fourth) and HeadD (5 votes for third and 8 for fourth) as their backup. Hand-based techniques were generally the worst, being either in the fifth place (HandD with 14 votes and HandP with 5 votes) or the final option (HandP with 13 votes and HandD with 6 votes).

4.4.3 Qualitative Feedback

In general, most participants stated positive comments for HeadC: "*simple/easy*" (N=7), "*fast*" (N=6), "*accurate*" (N=9), and "*creative/innovative and practical*" (N=3), or simply stated, "*the best*" (N=2). The only negative comments were "*slightly dizzy because of the head movements*" (N=2). Similarly, most participants left positive feedback for ConC: "*easy/simple and practical*" (N= 17). The reasons given were that "*it just like how we interact with PC with a mouse*" (N=5). However, we also observed negative comments "*tiredness/mobility of the hand lead to selection error during the selection*" (N=11), with others stating that this technique was "*somewhat boring*" (N=3).

In general, participants felt positive about HeadD and ConD. They were perceived to be "*fast, easy, easy to learn*" (N=11), "*creative*" (N=3). The disadvantage of these methods was related to dwell-based selection "*difficult to hold the pose for dwell*" (N=10), "*often suffer the last sec movement which made me have to dwell again*" for Controller users (N=10), and "*dizzy/eye tiredness*" for Head users (N=2).

Although HandD and HandP were perceived to be "*cre-*

Authorized licensed use limited to: Korea Advanced Inst of Science & Tech - KAIST. Downloaded on August 28,2023 at 02:18:52 UTC from IEEE Xplore. Restrictions apply.

Table 4: ANOVA test results for NASA-TLX subscales. Significant results where $p < .05$ are shown in light green and $p < .001$ in dark green.

	Interaction Technique	Post-hoc
Mental	$F_{5,115} = 11.740, p < .001, \eta_p^2 = .338$	HeadC<HandD ($p < .05$), ConD<HandD ($p < .01$), ConC<HandD ($p < .001$), HeadC<HandP ($p < .01$), ConD<HandP ($p < .01$), ConC<HandP ($p < .001$)
Physical	$F_{2.657,61.121} = 7.990, p < .001, \eta_p^2 = .258$	ConC<HandD ($p < .05$), HeadC<HandP ($p < .05$), ConC<HandP ($p < .01$)
Temporal	$F_{5,115} = 2.919, p < .05, \eta_p^2 = .113$	N/A
Effort	$F_{2.739,63.006} = 11.701, p < .001, \eta_p^2 = .337$	ConC<HandD ($p < .05$), HeadD<HandP ($p < .01$), HeadC<HandP ($p < .01$), ConD<HandP ($p < .05$), ConC<HandP ($p < .001$)
Frustration	$F_{3.210,73.833} = 8.730, p < .001, \eta_p^2 = .275$	HeadC<HandD ($p < .05$), ConC<HandD ($p < .05$), HeadC<HandP ($p < .01$), ConC<HandP ($p < .01$)
Performance	$F_{5,115} = 5.474, p < .001, \eta_p^2 = .192$	ConC<HandD ($p < .05$), HeadC<HandP ($p < .05$), ConC<HandP ($p < .05$)

ative/innovative and fun" (N=8), it suffers from "*tiredness*" (N=10) and sometimes are very "*inaccurate (because of the difficulty in holding/performing gestures*" (N=3).

5 DISCUSSION

5.1 Task Performance

In line with a previous study on text entry [41] and supported by the bandwidth of the human muscle groups [5], we found that the default VR Interaction Techniques ConC (i.e., Controller+Click) is the best candidate regarding the task completion speed, followed by HeadC (i.e., Head+Click), and then ConD (i.e., Controller+Dwell). Despite being the fastest option, ConC led to a very high TER (21.5%, in line with previous results in AR HMDs with a similar interaction style [7]) and NCER (1.5%) while HeadC had a much more acceptable TER (11.4%) and NCER (0.8%). Consistent with a previous study that compares selection techniques [9, 33], Dwell-based techniques, as expected, made the least error during selection, ConD had the lowest total error rate while HeadD had the second lowest total error rate.

Hand-based interaction has been implemented in several VR HMDs (e.g., in the VIVE and Quest series). Users can now use it for system control tasks such as changing the home scene, opening/closing an application, adjusting the brightness setting, or entering texts. However, this type of interaction does not seem suitable for long-term fast-paced pointing tasks for selecting small objects. In line with the results from text entry studies (e.g., [48]), we found that hand-based methods (HandD [i.e., Hand+Dwell] and HandP [i.e., Hand+Pinch]) not only had the worst task completion speed but also caused the highest number of errors. A possible reason could be due to their hand's stability; it is generally challenging to hold their hands in mid-air on a consistent and stable basis [48], even if they could use the chair handles or a table to support their hands. Another reason is that the accuracy of hand-based interaction is just not sufficient enough in today's HMDs [39].

In addition, we observed that the Sentence Lengths could affect the task completion time, which is as expected because participants need to move more distance across the paragraphs.

5.2 User Experience

Overall, HeadC and ConC are the leading options based on user experience. HeadC provided a better user experience than HandD, HandP, and ConD (e.g., average UEQ, attractiveness, perspicuity, efficiency, and dependability). In addition, HeadC was rated better than HeadD (i.e., Head+Dwell) and ConD on stimulation. ConC provided a better user experience than HandD and HandP (i.e., average UEQ, attractiveness, perspicuity, efficiency, and dependability). However, this method was considered less novel compared to HeadD

and HeadC because it mimics how the mouse works for desktop and laptop computers.

When we compared these 6 text selection methods with the benchmark scores provided in [40], HeadC and ConC were rated mostly good to excellent, except for the Stimulation subscale, which was only above average for HeadC and below average for ConC. HeadD and ConD were mostly rated with average scores, with hand-based methods rated mostly bad to below average. A possible explanation for why hand-based methods were rated with low user experience scores may be because they could cause hand/arm fatigue, a common issue for mid-air interaction [48, 49]. Another good explanation would be that they are perceptibly slower and had more errors than controller-based and head-based approaches.

5.3 Workload

In line with the UEQ results, ConC and HeadC are the leading options according to the NASA-TLX workload scores. They were comparable across NASA-TLX subscales. ConC had a lower workload than HandD and HandP in all NASA-TLX subscales, except for Temporal. We also found HeadC outperformed HandP in the Mental, Performance, Effort, and Frustration scales. In addition, HeadC outperformed HandD on both Mental and Frustration scales. As mentioned by participants, the workload is high for hand-based methods because the need to hold their hands mid-air on a consistent and stable basis is difficult, which is in line with previous observations [48]. It is worth noting that the Quest 2 already provides an overall stable hand tracking at a 60HZ refreshing rate. We also made sure that the light condition was consistent and in good condition during the experiment to ensure the tracking was stable. Therefore, we believe that the high workload for hand-based methods might be due to arm/hand fatigue [17], which would have made the task more complicated and unnecessarily cumbersome [2].

5.4 Recommendations for Text Selection in VR HMDs

The recommendations derived from our experiment are based on a set of considerations, which include performance, experience, the workload of the interaction technique, the availability of the controller, and users' physical conditions, especially for hand/arm control. They can be divided into two groups based on the goals of the system or the users:

Performance. If a controller is available, users could choose Head+Click since it can lead to the second fastest task completion speed and has an acceptable total error rate. If rapid neck movement is an issue for users to rotate their head for pointing, we suggest using Controller+Dwell and Controller+Click as alternative options. Controller+Dwell should be considered before Controller+Click when accuracy is needed since it has a significantly lower total error rate. If correcting errors during the session is not an issue, users can

use Controller+Click. However, users should be comfortable with a relatively higher total error rate due to hand mobility limitations [17]. Head+Dwell should be considered as the first option in both device-free and hands-free scenarios. As for hands-based interaction techniques (i.e., Hand+Dwell and Hand+Pinch), they should be minimized due to poor performance (in both speed and accuracy).

User Preference. We suggest considering Head+Click and Controller+Click as the top options when a controller is available. However, their usage should be based on the user's preference for usability and workload. If a low workload is preferred, Controller+Click can be chosen since it produces the lowest workload and provides an excellent user experience score. When a good user experience is preferred and users have no neck injuries or concerns about neck fatigue, Head+Click should be considered as the first option because it provides the best user experience score and produces a very low workload.

Head+Dwell should be considered as the first option in device-free scenarios since it provides a better overall user experience and produces a lower workload than hand-based techniques (i.e., Hand+Dwell and Hand+Pinch). Hand+Dwell and Hand+Pinch should be minimized in rapid pointing-based text selection tasks since they generate a high overall workload and low overall user experience, with the potential to lead to the gorilla arm syndrome [17]. However, they can be considered when the user has an issue with rotating their neck.

The above recommendations are derived from our study where the text selection task is assumed to be the primary task. Their inclusion in actual applications should be dependent on the context of use, target users, and availability of equipment.

5.5 Limitations and Future Work

One limitation of this research is that the study only consists of a single session. It will be useful to perform longer experimental sessions (e.g., 1 or 2 sessions over consecutive 4-5 days like text entry studies [12, 50, 53]). Nevertheless, our study, as the first one to explore text selection in HMDs and to evaluate 6 interaction techniques from 3 common pointing methods (Controller, Head, Hand) and 3 selection mechanisms (Dwell, Click, Pinch), provides a valuable starting pointing for future studies on text selection in VR HMDs.

This work has focused primarily on user-driven text selection where the text is displayed outside users' hand reach area (i.e., far-field interaction). There are situations where (1) the text may be presented closer to users' view and within the reach of their arms, which requires near-field interaction to manipulate the text, or (2) a user wants to select a single word or a whole sentence or paragraph akin to a double-clicking of the mouse in desktop systems (see Sect. 2.1). Because we want to make text selection as fast and usable as possible and in line with how users do this on their desktops, we plan to expand our work and explore new techniques and approaches that are suitable for near-field text manipulation and also can be used for automatic text selection of words, sentences, and paragraphs in the future.

Some design considerations, such as dwell duration, visualization of the ray, and interaction distance, that have been determined via pilot studies could also be revisited to allow their fine-tuning and customization by users in actual application scenarios. In addition, we considered text selection as the primary task during the interaction with VR HMDs and future work will explore techniques to support a seamless experience on text selection to avoid the need for users to switch between techniques and interaction platforms.

We did not consider dual-hand text selection methods, given our focus on available pointing and selection mechanisms of current VR HMDs. However, they are worth exploring in the future because, when designed appropriately to minimize workload and maximize efficiency, they can extend the range of possibilities for users. For instance, researchers could consider: (1) Hand+Hand: users use their dominant hand for pointing and the other hand for selection, or vice-versa; (2) Controller+Controller: One controller is used for pointing and the other one for selection; (3) Hand+Controller: users can use their dominant hand for pointing and the controller for selection; (4) Controller+Hand: users can use a controller with their dominant hand for pointing and their non-dominant hand for selection [18]. This exploration will produce additional results and help develop VR systems that are more usable and will take us one step closer to a practical immersive office of the future.

6 Conclusion

In this work, we have implemented six text selection techniques that resulted from the combination of three pointing methods (Controller, Head, Hand) and three selection mechanisms (Dwell, Click, Pinch) that are widely available and frequently used in commercial virtual reality (VR) head-mounted displays (HMDs). Then, we empirically evaluated these six techniques through a user study with 24 participants to assess their performance and user experience. In general, our results suggest that Head+Click and Controller+Click are the leading candidates among the 6 techniques. Head+Click should be considered as the first option since it has an excellent performance, provides the best user experience, and produces a very low workload—followed by Controller+Click, which has comparable results with Head+Click, except for having a higher total error rate. Controller+Dwell has the lowest total error rate and should be preferred if a low error rate is a plus for the user. Our results also show that current hand-based techniques should be the last possible options for rapid text selection tasks, except when users have difficulties with neck motions.

References

[1] R. Atienza, R. Blonna, M. I. Saludares, J. Casimiro, and V. Fuentes. Interaction techniques using head gaze for virtual reality. In *2016 IEEE Region 10 Symposium (TENSYMP)*, pp. 110–114, 2016. doi: 10.1109/TENCONSpring.2016.7519387

[2] D. A. Bowman, R. P. McMahan, and E. D. Ragan. Questioning naturalism in 3d user interfaces. *Commun. ACM*, 55(9):78–88, Sept. 2012. doi: 10.1145/2330667.2330687

[3] E. Brasier, O. Chapuis, N. Ferey, J. Vezien, and C. Appert. Arpads: Mid-air indirect input for augmented reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 332–343, 2020. doi: 10.1109/ISMAR50242.2020.00060

[4] S. K. CARD, W. K. ENGLISH, and B. J. BURR. Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT. *Ergonomics*, 21(8):601–613, Aug. 1978. Publisher: Taylor & Francis. doi: 10.1080/00140137808931762

[5] S. K. Card, J. D. Mackinlay, and G. G. Robertson. A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst.*, 9(2):99–122, apr 1991. doi: 10.1145/123078.128726

[6] L. Chen, Y. Liu, Y. Li, L. Yu, B. Gao, M. Caon, Y. Yue, and H.-N. Liang. Effect of visual cues on pointing tasks in co-located augmented reality collaboration. In *Symposium on Spatial User Interaction*, SUI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3485279.3485297

[7] R. Darbar, A. Prouzeau, J. Odicio-Vilchez, T. Lainé, and M. Hachet. Exploring Smartphone-enabled Text Selection in AR-HMD. In *Proceedings of Graphics Interface 2021*, GI 2021, pp. 117 – 126. Canadian

Information Processing Society, 2021. ISSN: 0713-5424 event-place: Virtual Event. doi: 10.20380/GI2021.14

[8] T. Dingler, K. Kunze, and B. Outram. VR Reading UIs: Assessing Text Parameters for Reading in VR. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pp. 1–6. Association for Computing Machinery, New York, NY, USA, 2018. event-place: Montreal QC, Canada. doi: 10.1145/3170427.3188695

[9] A. Esteves, Y. Shin, and I. Oakley. Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human-Computer Studies*, 139:102414, 2020. doi: 10.1016/j.ijhcs.2020.102414

[10] S. A. Faleel, M. Gammon, K. Fan, D.-Y. Huang, W. Li, and P. Irani. Hpui: Hand proximate user interfaces for one-handed interactions on head mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021. doi: 10.1109/TVCG.2021.3106493

[11] D. Ghosh, P. S. Foong, S. Zhao, C. Liu, N. Janaka, and V. Erusu. EYEditor: Towards On-the-Go Heads-Up Text Editing Using Voice and Manual Input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376173

[12] J. Gong, Z. Xu, Q. Guo, T. Seyed, X. A. Chen, X. Bi, and X.-D. Yang. *WrisText: One-Handed Text Entry on Smartwatch Using Wrist Gestures*, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2018.

[13] J. Grubert, E. Ofek, M. Pahud, and P. O. Kristensson. The Office of the Future: Virtual, Portable, and Global. *IEEE Computer Graphics and Applications*, 38(6):125–133, Nov. 2018. doi: 10.1109/MCG.2018.2875609

[14] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Text entry in immersive head-mounted display-based virtual reality using standard keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 159–166, 2018. doi: 10.1109/VR.2018.8446059

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

[16] R. Hedeshy, C. Kumar, R. Menges, and S. Staab. Hummer: Text entry by gaze and hum. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445501

[17] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 1063–1072. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2557130

[18] Y.-J. Huang, K.-Y. Liu, S.-S. Lee, and I.-C. Yeh. Evaluation of a hybrid of hand gesture and controller inputs in virtual reality. *International Journal of Human–Computer Interaction*, 37(2):169–180, 2021. doi: 10.1080/10447318.2020.1809248

[19] J. L. Jr, E. Kruijff, R. McMahan, D. Bowman, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley, Boston, 2nd edition ed., Apr. 2017.

[20] W. jun Hou and X. lin Chen. Comparison of eye-based and controller-based selection in virtual reality. *International Journal of Human–Computer Interaction*, 37(5):484–495, 2021. doi: 10.1080/10447318.2020.1826190

[21] A. Karlson, B. Bederson, and J. Contreras-Vidal. Understanding single-handed mobile device interaction. 2006.

[22] M. Krichenbauer, G. Yamamoto, T. Taketom, C. Sandor, and H. Kato. Augmented reality versus virtual reality for 3d object manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1038–1048, 2018. doi: 10.1109/TVCG.2017.2658570

[23] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12. IEEE Computer Society, Seoul, South Korea, Oct. 2010. doi: 10.1109/ISMAR.2010.5643530

[24] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. *Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality*, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2018.

[25] B. Laugwitz, T. Held, and M. Schrepp. Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger, ed., *HCI and Usability for Education and Work*, vol. 5298, pp. 63–76. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-89350-9_6

[26] L.-H. Lee, Y. Zhu, Y.-P. Yau, T. Braud, X. Su, and P. Hui. One-thumb Text Acquisition on Force-assisted Miniature Interfaces for Mobile Headsets. In *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–10, 2020. doi: 10.1109/PerCom45495.2020.9127378

[27] M. Y. Lin, J. G. Young, and J. T. Dennerlein. Evaluating the effect of four different pointing device designs on upper extremity posture and muscle activity during mousing tasks. *Applied Ergonomics*, 47:259–264, Mar. 2015. doi: 10.1016/j.apergo.2014.10.003

[28] X. Lu, D. Yu, H.-N. Liang, and J. Goncalves. itext: Hands-free text entry on an imaginary keyboard for augmented reality systems. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, p. 815–825. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474788

[29] X. Lu, D. Yu, H.-N. Liang, W. Xu, Y. Chen, X. Li, and K. Hasan. Exploration of Hands-free Text Entry Techniques For Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 344–349. Porto de Galinhas, Brazil, Nov. 2020. ISSN: 1554-7868. doi: 10.1109/ISMAR50242.2020.00061

[30] P. Maier, A. Dey, C. A. L. Waechter, C. Sandor, M. Tönnis, and G. Klinker. An empiric evaluation of confirmation methods for optical see-through head-mounted display calibration. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pp. 267–268, 2011. doi: 10.1109/ISMAR.2011.6143895

[31] X. Meng, W. Xu, and H.-N. Liang. An exploration of hands-free text selection for virtual reality head-mounted displays. In *21st IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, p. 8. IEEE Computer Society, Singapore, 2022.

[32] M. R. Mine. Virtual environment interaction techniques. Technical report, USA, 1995.

[33] A. K. Mutasim, A. U. Batmaz, and W. Stuerzlinger. *Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality*. Association for Computing Machinery, New York, NY, USA, 2021.

[34] V. Nanjappan, H.-N. Liang, F. Lu, K. Papangelis, Y. Yue, and K. L. Man. User-elicited dual-hand interactions for manipulating 3D objects in virtual reality environments. *Human-centric Computing and Information Sciences*, 8(1):31, Dec. 2018. doi: 10.1186/s13673-018-0154-5

[35] E. Ofek, J. Grubert, M. Pahud, M. Phillips, and P. O. Kristensson. Towards a practical virtual office for mobile knowledge workers. August 2020.

[36] M. Porta and M. Turina. Eye-s: A full-screen input modality for pure eye-based communication. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, ETRA '08, p. 27–34. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1344471.1344477

[37] E. Quinn, I. S. P. Nation, and S. Millett. Asian and Pacific speed readings for ESL learners. *ELI Occasional Publication*, 24:74, 2007.

[38] V. K. Rick and R. Poelman. A Survey of Augmented Reality Technologies, Applications and Limitations. *The International Journal of Virtual Reality*, 9(2):1–20, June 2010. doi: 10.20870/IJVR.2010.9.2.2767

[39] D. Schneider, V. Biener, A. Otte, T. Gesslein, P. Gagel, C. Campos, K. Čopič Pucihar, M. Kljun, E. Ofek, M. Pahud, P. O. Kristensson, and J. Grubert. Accuracy evaluation of touch tasks in commodity virtual and augmented reality head-mounted displays. In *Symposium on Spatial User Interaction*, SUI '21, p. 11. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3485279.3485283

[40] M. Schrepp, A. Hinderks, and J. Thomaschewski. Construction of a benchmark for the user experience questionnaire (ueq). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4:40–44,

06 2017. doi: 10.9781/ijimai.2017.445

[41] M. Speicher, A. M. Feit, P. Ziegler, and A. Krüger. Selection-based text entry in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3174221

[42] B. Thomas and W. Piekarski. Glove Based User Interaction Techniques for Augmented Reality in an Outdoor Environment. *Virtual Reality*, 6(3):167–180, Oct. 2002. doi: 10.1007/s100550200017

[43] L. Vanacken, T. Grossman, and K. Coninx. Exploring the effects of environment density and target visibility on object selection in 3d virtual environments. In *2007 IEEE Symposium on 3D User Interfaces*, 2007. doi: 10.1109/3DUI.2007.340783

[44] C. Wei, D. Yu, and T. Dingler. Reading on 3D Surfaces in Virtual Environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 721–728. IEEE, Atlanta, GA, USA, Mar. 2020. ISSN: 2642-5254. doi: 10.1109/VR46266.2020.00095

[45] M. Weise, R. Zender, and U. Lucke. How can i grab that?: Solving issues of interaction in vr by choosing suitable selection and manipulation techniques. *i-com*, 19(2):67–85, 2020. doi: doi:10.1515/icom-2020-0011

[46] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963

[47] W. Xu, H.-N. Liang, Y. Chen, X. Li, and K. Yu. Exploring visual techniques for boundary awareness during interaction in augmented reality head-mounted displays. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 204–211, 2020. doi: 10.1109/VR46266.2020.00039

[48] W. Xu, H.-N. Liang, A. He, and Z. Wang. Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays. In *18th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 419–428. IEEE Computer Society, Beijing, China, 2019. doi: 10.1109/ISMAR.2019.00026

[49] W. Xu, H.-N. Liang, Y. Zhao, D. Yu, and D. Monteiro. DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pp. 1–14. ACM Press, Glasgow, Scotland Uk, 2019. doi: 10.1145/3290605.3300674

[50] W. Xu, H.-N. Liang, Y. Zhao, T. Zhang, D. Yu, and D. Monteiro. RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1991–2001, 2019. doi: 10.1109/TVCG.2019.2898736

[51] N. Yankelovich, G.-A. Levow, and M. Marx. Designing speechacts: Issues in speech user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, p. 369–376. ACM Press/Addison-Wesley Publishing Co., USA, 1995. doi: 10.1145/223904.223952

[52] C. Yu, Y. Gu, Z. Yang, X. Yi, H. Luo, and Y. Shi. *Tap, Dwell or Gesture? Exploring Head-Based Text Entry Techniques for HMDs*, p. 4479–4488. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025964

[53] D. Yu, K. Fan, H. Zhang, D. Monteiro, W. Xu, and H.-N. Liang. PizzaText: Text Entry for Virtual Reality Systems Using Dual Thumbsticks. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2927–2935, Nov. 2018. doi: 10.1109/TVCG.2018.2868581

[54] D. Yu, H.-N. Liang, F. Lu, V. Nanjappan, K. Papangelis, and W. Wang. Target selection in head-mounted display virtual reality environments. *Journal of Universal Computer Science*, 24(9):1271–1243, 2018.

[55] D. Yu, H.-N. Liang, X. Lu, K. Fan, and B. Ens. Modeling endpoint distribution of pointing selection tasks in virtual reality environments. *ACM Trans. Graph.*, 38(6), nov 2019. doi: 10.1145/3355089.3356544

[56] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves. Gaze-supported 3d object manipulation in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery,

New York, NY, USA, 2021. doi: 10.1145/3411764.3445343

[57] D. Yu, Q. Zhou, J. Newn, T. Dingler, E. Velloso, and J. Goncalves. Fully-occluded target selection in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3402–3413, 2020. doi: 10.1109/TVCG.2020.3023606