Exploring Controller-based Techniques for Precise and Rapid Text Selection in Virtual Reality

Jianbin Song[*]
School of Advanced Technology
Xi'an Jiaotong-Liverpool University

Rongkai Shi[†]
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Department of Computer Science
University of Liverpool

Yue Li[‡]
Department of Computing
School of Advanced Technology
Xi'an Jiaotong-Liverpool University

BoYu Gao[§]
College of Cyber Security / Guangdong
Institute of Smart Education
Jinan University

Hai-Ning Liang[¶]
Department of Computing
School of Advanced Technology
Xi'an Jiaotong-Liverpool University

2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR) | 979-8-3503-7402-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/VR58804.2024.00047

ABSTRACT

Text selection is a common task in interactive systems. Often, it can be difficult because the letters and words are too small and clustered together to allow precise selection. Compared to traditional 2D interfaces, text selection is more challenging in virtual reality (VR) head-mounted displays (HMDs) because users interact with the immersive 3D space via mid-air interaction, which has higher degrees of freedom but becomes more imprecise and involves a higher workload due to the lack of support from a fixed structure like a desk. There has been limited exploration of techniques that support precise and rapid text selection at the character, word, sentence, or paragraph levels in VR HMDs. To fill this gap, we propose three controller-based text selection methods: *Joystick Movement*, *Depth Movement*, and *Wrist Orientation*. They are evaluated against a baseline selection method via a user study with 24 participants. Results show that the three proposed techniques significantly improved the performance and user experience over the baseline, especially for the selection beyond the character level.

Keywords: Text Selection, Virtual Reality, Head-mounted Display, 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 an essential task in interactive systems. It is normally the first step for follow-up text edition operations, such as copying, cutting, or highlighting. However, precise text selection can be difficult because letters are small targets and are clustered together in small areas. Prior studies have attempted to propose enhanced interaction techniques to overcome this challenge. Much attention is paid to touch-based devices because finger touch brings visual occlusion to the text, making the selection more challenging (e.g., [6, 10, 36]).

With recent advancements in virtual reality (VR) technologies, head-mounted displays (HMDs) have become powerful and can be the next generation of personal portable tools, similar to today's mobile devices, and to some extent as potential alternatives to laptops and desktops. Recent work and advancements have pointed to opportunities of leveraging VR HMDs for office work [8, 27]. Furthermore, there is a trend of supporting remote or collaborative office work in HMDs. Microsoft Office apps, for example, have recently become available on all Meta Quest headsets [35]. Text selection is one indispensable sub-task of these office tasks. Similarly, Apple's Vision Pro HMD has been advertised as a device that can create 'the perfect workspace,' which mixes web browsing, text messaging, and other activities involving interaction with text. Therefore, improving text selection performance and experience in VR HMDs is important.

Selecting a target like a letter in VR HMDs is different from touch-based devices. Unlike directly tapping on the target displayed on interactive surfaces, a typical interaction in VR HMDs is pointing to the target via a virtual ray in the immersive space and indicating the selection by a confirmation action (i.e., raycasting pointing) [22, 42]. In the context of text selection, this pointing-based approach helps minimize the finger occlusion problem. On the other hand, it could be imprecise and unstable because users interact with a 3D space with six degrees of freedom and lack support from a fixed structure like a desk or an interactive surface. Recently, researchers have explored selection approaches in VR [21, 40] and augmented reality (AR) HMDs [20]. They systematically evaluated different HMD-powered input modalities for pointing and selection steps of text selection tasks and provided the first explorations of text selection in HMDs focusing on the most common character-based selection. However, they did not present new techniques.

In daily text selection tasks, users may want to select text longer than a few letters in a word or several words, like for underlining or copying a quote from classic literature or text from a webpage. Such demand may involve a text range across several words, sentences, or paragraphs. Allowing users to select text at different ranges can facilitate text selection performance and experience [10]. This is enabled by an efficient shortcut technique for including text ranges at character, word, sentence, and paragraph levels. Such a technique can be useful for mitigating the imprecision issue in pointing selection and is especially needed for VR HMDs.

To support the varied needs during text selection, this work looks at designing and developing interaction techniques that can select text at the character, word, sentence, and paragraph levels precisely and rapidly in VR HMDs. As a first exploration of this topic, we focused on controller input, which is the most common input modality for currently available VR HMDs. Based on a set of design considerations (see Sect. 3), we proposed three one-handed text selection methods: *Joystick Movement*, *Depth Movement*, and *Wrist Orientation*. They were compared against a baseline technique via a user study ($N = 24$). We evaluated these techniques using tasks with various text lengths and investigated participants' performance at different levels of expertise. Our results showed that the proposed

[*]e-mail: jianbin.song22@student.xjtlu.edu.cn. Joint First Author.
[†]e-mail: rongkai.shi19@student.xjtlu.edu.cn. Joint First Author.
[‡]e-mail: yue.li@xjtlu.edu.cn
[§]e-mail: bygao@jnu.edu.cn
[¶]e-mail: haining.liang@xjtlu.edu.cn. Corresponding Author.

979-8-3503-7402-5/24/$31.00 ©2024 IEEE
DOI 10.1109/VR58804.2024.00047

techniques achieved better performance in text selection tasks and led to enhanced user experience. At the end of the paper, we showed two possible extensions based on the current selection techniques for future implementation and exploration.

This paper's main contribution is twofold: (1) introducing three single-handed controller-based text selection techniques based on four design considerations, and (2) a user study that evaluated their performance and usability against the baseline approach. Our work contributes to a better understanding of how to design efficient and usable techniques for text selection, which is an essential task in the workflow of productivity-based activities involving interaction with text.

2 Related Work

2.1 Text Selection in 2D and 3D Interfaces

Text selection has been well-studied in 2D interactive systems, especially for mobile devices such as smartphones and tablets. The main branch of text selection studies improves the original workflow to provide better text selection performance and experience. For example, Arrow2edit [6] and Press&Slide [3] create shortcut gestures on the virtual keyboard (closer to user's hand) to select a word without moving the finger to the text area for specifying the word. On the other hand, recent studies also improve text selection performance and experience by adopting a secondary input modality to assist hand-based selection (e.g., EyeSayCorrect [43] and GazeButton [26]).

In practical use, users may want to select a text fragment longer than a few characters or words, such as sentences or paragraphs. Researchers have also explored novel techniques to facilitate text selection at different lengths. Le et al. [18] elicited touch-sensitive shortcut gestures for moving the caret and selecting long text snippets quickly and easily. On the other hand, Goguey et al. [10] proposed mode gauges, a touch- and force-based technique to enable seamless switching between text selection modes, which improved discoverability and provided smoother transitions to experts. Recently, Tu et al. [36] proposed Text Pin, a novel pointing-based technique to position the selection handles, which was more efficient than the original dragging handles.

As VR and AR HMDs become more widespread, text selection in 3D environments is gaining attention. Hu et al. [14] evaluated hand- and eye-based techniques caret placement, a critical sub-task for text selection, in AR HMD. They found raycasting was faster compared to other methods for text caret navigation. Ghosh et al. [9] presented EYEditor, a multi-modal text editing tool for smart glasses. In EYEditor, a hand controller with directional buttons is used for text selection. Similarly, Darbar et al. [5] used a smartphone as the pointer to support text selection in AR HMD. Xu et al. [40] conducted a systematic evaluation of text selection techniques in VR HMDs. They compared controller-based, freehand-based, and head-based pointing and selection mechanisms and found that using the head or controller for pointing and pressing the controller button for confirmation led to the best experience. Following this study, Liu et al. [20] compared text selection techniques in AR HMD. Similarly, Meng et al. [21] explored and evaluated hands-free solutions for VR HMD.

Unlike touch-based devices, text selection in VR and AR HMDs no longer happens on a touchscreen, where users can touch it physically and receive tactile (force) feedback. Instead, users have to perform mid-air interaction to select text in 3D spaces. Raycasting pointing, as the most common target-pointing technique in VR/AR, has been used for text selection in prior studies [5, 14, 20, 21, 40]. These studies provide some good foundations but did not focus on specific interaction techniques that can support improved performance and user experience. As such, inspired by this prior research, our work fills an important gap and is about finding concrete techniques that enable precise and rapid text selection for different text lengths in VR HMDs. In short, this work represents a first exploration into establishing such enhanced techniques and is focused on controller-based raycasting methods that meet several design considerations (see Sect. 3) to maximize user performance and experience.

2.2 Mode Switching

Mode switching can help produce different outputs from the same input [25]. It has been widely studied across various platforms. Early in 1992, Sellen et al. [1] showed that a system-maintained mechanism (e.g., the Caps Lock key) was more error-prone than a user-maintained mode-switching mechanism (e.g., holding down the Shift key) in text editing tasks using a desktop computer. Similarly, Saund and Lank [28] highlighted that letting users specify the 'draw' or 'command' mode prior to a stroke was a usability obstacle. The superiority of 'holding' actions for activating a mode has been further proved in pen- and touch-based interfaces, especially using the non-dominant hand for the action [13, 33, 37].

Researchers have also explored mode-switching in 3D space for different task contexts. For example, Park et al. [24] proposed HandPoseMenu, a hand-posture-based menu system for changing interaction modes. Users can maintain a non-dominant hand gesture to access the corresponding menu (e.g., a color palette) and use the dominant hand to select the menu item and perform main tasks. Song et al. [32] proposed a gesture-based method to enable users to switch the keyboard layers efficiently for mid-air typing in AR HMDs. Users can access the capital letter and special character layers by rotating the wrist or typing with two fingers, which could seamlessly switch back to the default lowercase letter typing with the index finger. Moreover, Wan et al. [39] designed and evaluated controller-based raycasting methods for the same task in VR HMDs. They utilized controller buttons to switch keyboard layers to enable efficient alphanumeric and special character text entry.

In addition to specific task contexts, researchers have also explored mode-switching techniques for general use. Smith et al. [31] proposed five techniques utilizing different inputs for mode switching in AR HMD and found that raising the non-dominant hand and moving the dominant hand to different depths were faster than pressing hardware and virtual buttons and voice control. Surale et al. [34] designed and evaluated bare-hand mid-air mode-switching methods and provided guidance for choosing these methods in VR. They recommended using non-dominant actions for tasks that require accuracy. As for one-handed techniques, they recommended subtle variations of the default pinch gesture, such as rotating the wrist or using another finger. While prior work also explored mode-switching methods with a secondary modality, like eye gaze [15] and head motions [30, 38], this work starts with and only considers controller-based mode-switching methods to facilitate text selection in VR HMD. As results from the user study show, they work well within the text selection selection workflow of the techniques presented in this paper.

2.3 Novice and Expert Use

Novice and expert users have different interaction behaviors. Expert users normally have better performance because they typically use memory-based interactions, such as marking menus, hotkeys, and gestural commands [23]. Graphical menus provide good support for novice users as they are easy and intuitive to learn and use. In contrast, navigating the graphical menus can be slow for expert users as they are familiar with the items and do not need to read them again for location [4]. Thus, supporting a smoother transition to experts is important. One common way is to follow the principle of rehearsal [17], which states that novices should interact with the system in the same way as experts, thereby incidentally learning the expert mode through daily use. Many menu selection techniques, such as Marking Menus [16] and FastTap [11], have followed this principle.

In the context of text selection, novice and expert performance has rarely been discussed. To the best of our knowledge, only Goguey et al. [10] evaluated the discoverability and expert performance of their text selection techniques, but in the context of touch-based devices. In this work, we were aware of the difference between novice and expert use of the techniques and evaluated the performance at different levels of expertise.

3 Design Considerations

To design and develop interaction techniques that enable precise and rapid selection for different text ranges in VR HMDs, we have identified four considerations, as discussed below.

3.1 One-handed vs. Two-handed Interaction

Although using two hands could facilitate more interactions than one hand, we only consider one-handed text selection techniques in this work for two reasons. First, text selection is a simple task that can (and should) be accomplished with one hand with limited effort. Having two hands for such a simple task may complicate the interaction procedure and add unnecessary cognitive load and physical demand. Second, text selection is the beginning of many tasks relevant to text manipulation, like deleting, copying, and modifying the font format. Using one hand could avoid interference in this preparation stage and enable the parallel execution of some text editing tasks that, for instance, could be performed with the other hand.

3.2 User Interface

A clear and understandable user interface (UI) is crucial for interaction techniques. We have two general considerations for better indicating the text selection mode (i.e., whether to select the text at character, word, sentence, or paragraph level). The first is the timing of showing the UI. Completing a text selection task requires the user to control the cursor (caret) from the start to the end position of the content to be selected. The UI should only be activated during this procedure to indicate the mode without impacting selection. We decided not to show the UI when the user is not selecting the text because the UI may block the text. The second consideration is the position of the UI. The cursor and the mode selection UI play the same role in the selection process because of their shared purpose of indicating the selection range. The mode selection UI is designed to center around the cursor, ensuring that it follows the cursor's movement rather than isolated and/or fixed in a position that would require additional attention. Furthermore, a certain interval between the two is retained to minimize occlusion. These considerations are further explained when introducing the techniques in Sect. 4.

3.3 Degrees of Freedom

3D spaces enable interaction with six degrees of freedom, including 3 in translation and 3 in orientation. Normally, text selection happens on the plane where the text or document is demonstrated. The pointing-based text selection procedure can be considered a translational movement on this plane, which the designed technique should not disrupt. More specifically, assuming the text is placed on the x-y plane, the text selection technique should not involve the movement or rotation of the cursor in x- and y-axes.

3.4 Mode Switching Mechanism

The mode-switching techniques should assist the text selection in achieving better precision and speed with minimal cognitive and physical demands. To suit the context of text selection, we identified three important concerns. First, the mode-switching action should be *kinesthetic* [34]—the action should be maintainable while the user is selecting the text because holding the mode could be less error-prone (see Sect. 2.2). Second, the target mode should be achievable directly. Time-based actions or action loops force users to wait or repeat actions, leading to higher cognitive demands for users. Thus, we did not incorporate these actions in our designs. Third, the mode-switching action should be changeable during the text selection. Allowing users to change the mode during the selection procedure affords them an option for correcting an error or fitting their altered intention without redoing a selection.

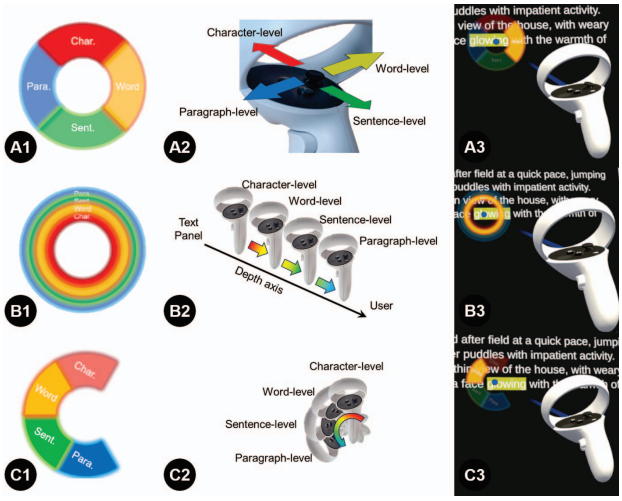

Figure 1: Illustrations of the three techniques: (A1-3) Joystick Movement. (B1-3) Depth Movement. (C1-3) Wrist Orientation. Sub-figures in the first column are illustrations of each technique's user interface. Sub-figures in the second column demonstrate the mode selection mechanism. Sub-figures in the third column show examples of selecting a word with each technique.

4 Technique Design

In this section, we introduce three techniques based on the design considerations. To begin with, we first describe how text is selected by default in current VR HMDs, which also serves as a baseline technique that does not have any text selection shortcuts.

4.1 Baseline (BL)

The baseline selection method (BL) only allows users to select text character-by-character during the selection process, which is also the technique prior VR/AR studies investigated (as shown in Sect. 2.1). To initiate the selection, a user first needs to control the ray and point to the starting position. The user then presses the trigger button on the dominant-hand controller and moves the ray to embrace the target text snippets while holding down the trigger button. Once the caret has been moved to the ending position, the selection procedure ends by releasing the trigger button. Like in most VR applications, a circular cursor is displayed on the text panel to indicate where the ray is pointed to.

4.2 Joystick Movement (JM)

Joystick Movement (JM) maps the tilting direction of the controller joystick to a text selection mode (Fig. 1(A1-3)). Users can change to character-level, word-level, sentence-level, and paragraph-level selection modes by tilting the joystick upward, rightward, downward, and leftward, respectively. Character level is the default mode. Except for tilting the joystick upward, users can select characters without directional input. A pie menu (more specifically, a doughnut-shaped UI) is used to indicate the current selection mode. The interface is divided into four equal sections for the four modes. The UI design follows our design considerations (see Sect. 3). It centers around the cursor, follows the cursor's movements, and its hollow

design prevents occluding the text. The outer circle of the UI has a radius of 0.45m, and the radius of the inner circle is 0.23m, leading to the width of the annulus being 0.22m. To provide explicit visual cues, the activated mode is opaque, and the non-activated modes are semi-transparent.

The minimum selection unit adapts dynamically based on the activated mode. For example, pushing the joystick rightward illuminates the orange area, indicating word-level selection, where each word under the cursor is sequentially selected, as shown in Fig. 1(A3). To select the text based on the minimum unit, a user must maintain the joystick tilted while pressing and holding the trigger button for selection and navigating the cursor. The mode can be activated before the text is selected but must be deactivated after the text selection is confirmed (i.e., the trigger button is released).

4.3 Depth Movement (DM)

The Depth Movement (DM) technique switches the selection mode by moving the controller to a pre-defined region in the depth axis (Fig. 1(B1-3)). DM calculates the distance between the controller and the headset in the depth axis (after calibration). When the distance is greater than 0.32m, users can select the text at the character level. When users move the controller back at a distance between 0.26m and 0.32m, the mode changes to the word level, and at a distance between 0.2m and 0.26m, it switches to the sentence level. Finally, if the distance is smaller than 0.2m, users select the text at the paragraph level. This mechanism is similar to how light spreads out—the farther the light source (DM) is away from a wall (text panel), the larger the illuminated region (selection range) on the wall is. In addition, closer proximity to the user's body requires the user to flex his forearm, making precise pointing selection more challenging. Thus, we assign this range to a larger selection region, which requires less precision. These depth ranges were pre-tested by three volunteers before the formal evaluation.

The UI consists of four annuli, with the cursor at the center. The size of the UI remains the same as JM (outer radius = 0.45m, inner radius = 0.23m). The width of each annulus is 0.055m. The four annuli represent character-level mode (red), word-level mode (yellow), sentence-level mode (green), and paragraph-level mode (blue) from inner to outer. Similar to JM, the activated mode in DM is more opaque, while the non-activated ones are close to transparent. In addition, users need to maintain the controller at the target depth range before releasing the trigger button to confirm the selection.

4.4 Wrist Orientation (WO)

The Wrist Orientation (WO) leverages wrist rotations to switch the selection mode (Fig. 1(C1-3)). WO detects the rotations of the controller in the global z-axis of the virtual world, reflecting the wrist rotations. When the controller is held naturally in a standard posture, the rotation is 0°. To select characters, users maintain the default posture or with a supination angle of up to 20°. To select text beyond the character level, users need to perform wrist pronation. DM switches to a mode with a larger selection range for every 20° pronations. These rotation ranges were pre-tested by the same three users who tested DM. The UI consists of four equal-sized sectors (see Fig. 1(C1)). Each sector is 60°. The size, color code, and visual effects of the UI are the same as those used in JM and DM.

5 Evaluation

A user study was conducted to evaluate the four techniques. We made the following hypotheses.

- **H1.** Compared to the baseline, the three proposed techniques (JM, DM, and WO) would demonstrate superior performance when selecting words, sentences, and paragraphs in VR HMDs.
- **H2.** With empowered selection shortcuts, the three proposed techniques would show high usability and provide better user experiences. The perceived workload for completing text selection tasks in VR would be lower.
- **H3.** We hypothesize that the proposed techniques would be easy to learn and use. They would show superior performance than the baseline regardless of users' expertise.

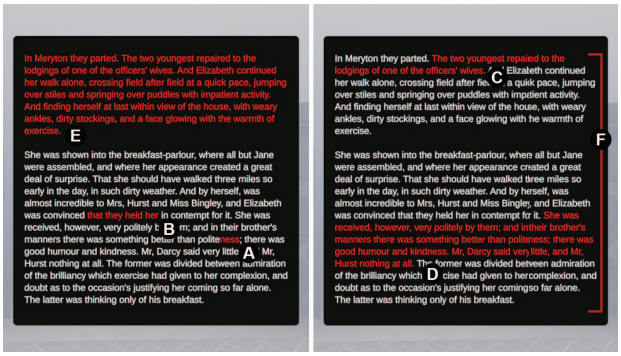

Figure 2: Example tasks used in our experiment. (A) Four characters. (B) Four words. (C) Sentence. (D) Two Sentences. (E) Paragraph. (F) Two Paragraphs.

5.1 Participants and Apparatus

Twenty-four participants (11 females and 13 males) were recruited from a local university. The sample size meets the minimum sample size needed to have enough power to detect an effect (computed via the G*Power tool). They aged between 18 and 26 years old ($M = 22$, $SD = 2.4$). All have normal or corrected-to-normal vision and are right-handed. Ten participants are frequent VR HMD users who use the device more than once per week.

We used a Meta Quest 2 VR HMD to provide the virtual environment. A Quest 2 Controller was used for interaction. The HMD was connected to a high-performance laptop, which powered the experimental program. The laptop has an Intel i7-12700H @ 2.30GHz CPU, an NVIDIA GeForce RTX 3060 Laptop GPU, and 40.0GB RAM. The program was developed in Unity (version 2021.3.20f1c1) with the Oculus Integration package (version 50.0).

5.2 Task

We used similar tasks as described by Goguey et al. [10]. The task required participants to select a target text snippet in VR as fast and accurately as possible. The target text snippets (Task) varied from six lengths:

- *Four characters (Four Char.)*: Selecting 4 letters from a 10-letter word. See Fig. 2(A).
- *Four words*: Selecting 4 adjacent words. See Fig. 2(B)
- *Sentence (Sent.)*: Selecting a full sentence. See Fig. 2(C).
- *Two sentences (Two Sent.)*: Selecting 2 adjacent sentences. See Fig. 2(D).
- *Paragraph (Para.)*: Selecting a full paragraph. See Fig. 2(E).
- *Two paragraphs (Two Para.)*: Selecting two adjacent paragraphs. See Fig. 2(F).

We removed three tasks from Goguey et al.'s task pool: two words and four characters, the middle of a word to the end of a paragraph, and the whole text [10] because the first two are relatively less common in daily uses and the last one, the whole text, is usually assigned to the context menu of the document.

The following parameters about the text panel were mostly based on the guidelines for using text in VR [7] but modified to suit our experimental design. We provided text using a dark grey (#2E2E2E) text panel positioned 3 meters in front and center of the user's vision (fixed once initiated). The panel's size was 300px×300px, and the texts were located in a 280px×280px area in the center. The texts were in white, while the target text snippet was highlighted in red. We used Unity's built-in Liberation Sans font, which is close to the recommended Arial font. The font size was 12, and the line spacing was set to 1.2. The texts were left-aligned. The text material included 199 words in two paragraphs and 21 lines (including a line break). The longest line had 68 characters, and the median was 62 characters. We had pilot tests with three participants (who also tested the techniques but did not participate in the formal experiment) to ensure the texts could be seen clearly. All confirmed they could see the text clearly and comfortably in this setup. In addition, no participants in the formal experiments claimed they had issues seeing and reading the text.

5.3 Design

We used a within-subjects design with TECHNIQUE as the independent variable (JM vs. WO vs. DM vs. BL). The order of TECHNIQUE conditions was fully counter-balanced via a balanced Latin square approach. Within each TECHNIQUE condition, the order of TASK is randomized.

We used a similar but extended experimental design as described in [2] to examine the effect of the level of expertise. Participants completed eight trials for each TECHNIQUE × TASK condition. The target text snippet was randomized and used for two consecutive trials. The randomized trials were treated as *Novice Trials*, and the follow-up repeated trials were *Experienced Trials*. Participants were expected to know where the target snippet was in an experienced trial as they just completed the same one in a novice trial. We also adapted the repeat-until-success methodology—wherein if participants failed a trial, they repeated it until they succeeded. This stage resulted in 6 tasks × (4 novice trials + 4 experienced trials) = 48 trials for each TECHNIQUE condition per participant. After participants completed all 48 trials for a TECHNIQUE condition, they had a follow-up stage including 12 trials: 2 trials for each of the six TASK conditions, and the order was fully randomized. We called these trials *Mixed Trials* for testing the occasional use of the technique. Participants needed to trigger the 'Next' button to move to the next trial. This was to indicate a trial's start and reset the hand position. In total, we collected 24 participants × 4 techniques × 6 tasks × (4 novice trials + 4 experienced trials + 2 mixed trials) = 5760 data trials.

5.4 Procedure

Each participant would take approximately 35 minutes to complete the user study, which was divided into four sessions. First, we gave participants a demographic questionnaire to collect their information and provided a brief introduction about the experiment. Second, participants received at least five minutes of training to familiarize themselves with the techniques. They were also informed about the task and procedure in the following formal experiment. Third, participants completed the formal trials for each condition, following the experimental design described in the previous section. Post-session questionnaires were given right after the completion of a condition, followed by a short break. Fourth, once participants completed all conditions, they received a semi-structured interview about their preferences and feedback.

5.5 Measurements

For each trial, we collected the following two types of performance data—total time and number of editions. Total time was measured from when the trial was initiated until the participants selected the target text snippet correctly. We calculated the *Mean Total Time* in novice, expert, and mixed trials for each TASK and TECHNIQUE. Additionally, we counted the number of editions participants made in each trial. A 0 time of edition indicates the participant completed the trial in one attempt. The *Mean Number of Editions* in each condition and each trial type was calculated for analysis.

In addition to the objective measures, we also collected subjective feedback. First, we used a raw *NASA-Task Load Index (NASA-TLX)* questionnaire [12] to measure the perceived workload for completing the task with the technique. It included six sub-scales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. In addition, an Overall workload score was also calculated. Second, we used a positive version of the *System Usability Scale (SUS)* questionnaire [19] to gauge techniques' usability. The final SUS score was calculated based on the ratings of ten questions. Third, a short version of the *User Experience Questionnaire (UEQ-S)* [29], including eight items, was used to measure user experience in terms of Pragmatic Quality, Hedonic Quality, and Overall User Experience. Finally, participants were asked to rank the four techniques based on their preferences and provide reasons at the end of the experiment. We also asked participants about their experience and suggestions for the UI design, mode-switching mechanism, and any other aspects of the techniques.

6 RESULTS

For the objective measures, we first identified and discarded the outliers ($> M + 3SD$) in each condition. We removed 198 trials, which count for 3.44% of the collected trials. Shapiro-Wilk tests and Q-Q plots were used to check the normality of the data. Both have shown that the mean total time and number of editions were not normally distributed. On the other hand, our subjective measures were questionnaire-based measurements. Thus, we used the Friedman tests for all measurements. Post hoc pairwise comparisons were conducted with Wilcoxon tests with Bonferroni corrections.

6.1 Novice Trials

The mean total time and number of editions in novice trials are summarized in Fig. 3(A1-A2). Except for the TASK *Four Char.*, significant differences were found in the remaining five tasks.

For the TASK *Four Words*, Friedman test revealed a significant main effect of TECHNIQUE on number of editions ($\chi^2(3) = 13.058$, $p = .005$, $W = .181$). Post hoc tests showed that WO ($Mdn = 0.5$) had a significantly lower number of editions than BL ($Mdn = 1$, $p = .009$).

For the TASK *Sent.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 22.05$, $p < .001$, $W = .306$) and on number of editions ($\chi^2(3) = 10.673$, $p = .014$, $W = .148$). Post hoc tests showed that JM ($Mdn = 2.902$) and WO ($Mdn = 2.371$) were significantly faster than BL ($Mdn = 3.505$) (both $p < .001$), WO was significantly faster than DM ($Mdn = 3.26$, $p = .029$).

For the TASK *Two Sent.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 18.2$, $p < .001$, $W = .253$) and number of editions ($\chi^2(3) = 8.576$, $p = .035$, $W = .119$). Post hoc tests showed that BL ($Mdn = 4.314$) was significantly slower than JM ($Mdn = 2.881$, $p < .001$), DM ($Mdn = 3.444$, $p = .043$), and WO ($Mdn = 3.255$, $p = .011$).

For the TASK *Para.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 36.35$, $p < .001$, $W = .505$) and number of editions ($\chi^2(3) = 20.237$, $p < .001$, $W = .281$). Post hoc tests showed that DM ($Mdn = 1.614$) and WO ($Mdn = 1.594$) were significantly faster than JM ($Mdn = 2.104$) and BL ($Mdn = 2.907$) (all $p < .001$). Regarding the number of editions, participants had significantly more attempts in BL ($Mdn = 0.5$) than in JM ($Mdn = 0$, $p = .003$), DM ($Mdn = 0$, $p = .013$) and WO ($Mdn = 0$, $p = .007$).

For the TASK *Two Para.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 41.65$, $p < .001$, $W = .578$) and number of editions ($\chi^2(3) = 18.389$, $p < .001$, $W = .255$). We

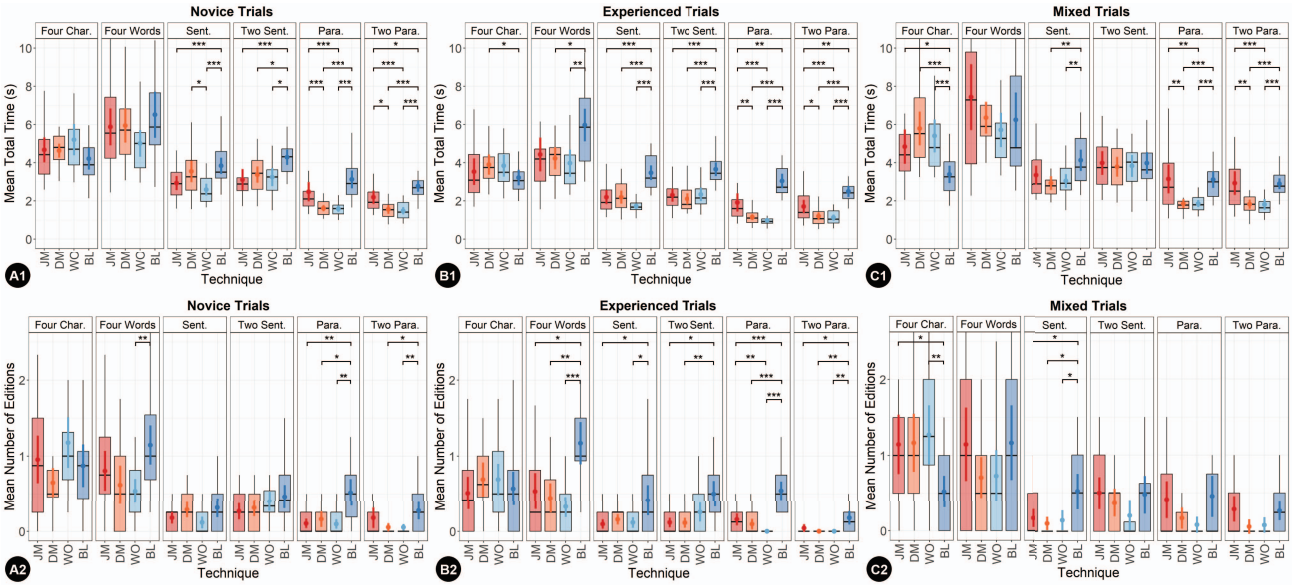

Figure 3: Objective results. (A1) Mean total time for novice trials. (B1) Mean total time for experienced trials. (C1) Mean total time for mixed trials. (A2) Mean number of editions for novice trials. (B2) Mean number of editions for experienced trials. (C2) Mean number of editions for mixed trials. Error bars represent 95% confidence intervals. '*', '**', '***' indicate Bonferroni-adjusted p values at $< .05$, $< .01$, and $< .001$ levels in pairwise comparisons, respectively.

found that BL ($Mdn = 2.69$) took a significantly longer time to complete this task than JM ($Mdn = 1.928$, $p = .048$), DM ($Mdn = 1.542$, $p < .001$), and WO ($Mdn = 1.411$, $p < .001$). In addition, JM took a significantly longer time than DM ($p = .015$) and WO ($p < .001$). Regarding the number of editions, participants had significantly more attempts in BL ($Mdn = 0.25$) than in DM ($Mdn = 0$, $p = .013$) and WO ($Mdn = 0$, $p = .006$).

6.2 Experienced Trials

The mean total time and number of editions in experienced trials are summarized in Fig. 3(B1-B2).

For the TASK *Four Char.*, a Friedman test revealed a significant main effect of TECHNIQUE on time ($\chi^2(3) = 10.65$, $p = .014$, $W = .148$) and post hoc tests revealed that BL ($Mdn = 3.06$) was significantly faster than DM ($Mdn = 3.748$, $p = .029$). No significant differences were found in terms of number of editions.

For the TASK *Four Words*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 10.85$, $p = .013$, $W = .151$) and on number of editions ($\chi^2(3) = 23.153$, $p < .001$, $W = .322$). Regarding total time, participants took significantly longer time using BL ($Mdn = 5.858$) than DM ($Mdn = 4.43$, $p = .026$) and WO ($Mdn = 3.449$, $p = .002$). Moreover, participants also had significantly more times of editions with BL ($Mdn = 1$) than JM ($Mdn = 0.25$, $p = .012$), DM ($Mdn = 0.25$, $p = .008$), and WO ($Mdn = 0.25$, $p < .001$).

For the TASK *Sent.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 28.85$, $p < .001$, $W = .401$) and on number of editions ($\chi^2(3) = 12.291$, $p = .006$, $W = .171$). Post hoc showed that BL performed significantly worse in terms of total time (JM-BL, DM-BL, WO-BL: all $p < .001$; JM: $Mdn = 1.914$, DM: $Mdn = 2.139$, WO: $Mdn = 1.673$, BL: $Mdn = 3.198$) and number of editions (JM-BL: $p = .028$, WO-BL: $p = .024$; BL: $Mdn = 0.25$, JM: $Mdn = 0$, and WO: $Mdn = 0$).

For the TASK *Two Sent.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 28.75$, $p < .001$, $W = .399$) and number of editions ($\chi^2(3) = 15.930$, $p = .001$, $W = .221$). Similar to TASK *Sent.*, BL performed significantly worse in terms of total time (JM-BL, DM-BL, WO-BL: all $p < .001$; JM: $Mdn = 2.193$, DM: $Mdn = 1.817$, WO: $Mdn = 2.155$, BL: $Mdn = 3.444$) and the number of editions (JM-BL: $p = .011$, DM-BL: $p = .007$; BL: $Mdn = 0.5$, JM: $Mdn = 0$, and DM: $Mdn = 0$).

For the TASK *Para.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 49.15$, $p < .001$, $W = .683$) and number of editions ($\chi^2(3) = 46.293$, $p < .001$, $W = .643$). Post hoc tests showed that DM and WO performed better. In terms of total time, BL ($Mdn = 2.717$) was slower than JM ($Mdn = 1.598$, $p = .003$), DM ($Mdn = 1.101$, $p < .001$), and WO ($Mdn = 0.923$, $p < .001$), JM was slower than DM ($p = .005$) and WO ($p < .001$). On the other hand, BL ($Mdn = 0.5$) had a higher number of editions than JM ($Mdn = 0.125$), DM ($Mdn = 0$), and WO ($Mdn = 0$) (all $p < .001$). JM had a higher number of editions than WO ($p = .009$).

For the TASK *Two Para.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 50.6$, $p < .001$, $W = .703$) and number of editions ($\chi^2(3) = 28.317$, $p < .001$, $W = .393$). Post hoc tests showed that BL performed significantly worse than the other techniques in terms of total time (JM-BL: $p = .007$, DM-BL $p < .001$, WO-BL: $p < .001$; JM: $Mdn = 1.398$, DM: $Mdn = 1.073$, WO: $Mdn = 1.047$, BL: $Mdn = 2.424$) and number of editions (JM-BL: $p = .045$, DM-BL $p = .01$, WO-BL: $p = .01$; JM: $Mdn = 0$, DM: $Mdn = 0$, WO: $Mdn = 0$, BL: $Mdn = 0.125$). Furthermore, post hoc tests also showed that JM was significantly slower than DM ($p = .012$) and WO ($p < .001$).

6.3 Mixed Trials

The mean total time and number of editions in mixed trials are summarized in Fig. 3(C1-C2).

For the TASK *Four Char.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 21.15$, $p < .001$, $W = .294$) and number of editions ($\chi^2(3) = 14.443$, $p = .002$, $W = .201$). Post hoc tests showed that participants took a significantly shorter time to complete the task using BL ($Mdn = 3.265$) than JM ($Mdn = 4.411$, $p = .019$), DM ($Mdn = 5.525$, $p < .001$), and WO ($Mdn = 4.79$,

$p < .001$), and a significantly fewer number of editions using BL ($Mdn = 0.5$) than using JM ($Mdn = 1$, $p = .043$) and WO ($Mdn = 1.25$, $p = .005$).

For the TASK *Sent.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 12.85$, $p = .005$, $W = .178$) and on number of editions ($\chi^2(3) = 12.140$, $p = .007$, $W = .169$). BL ($Mdn = 3.753$) was significantly slower than DM ($Mdn = 2.763$, $p = .009$) and WO ($Mdn = 2.904$, $p = .003$). On the other hand, BL ($Mdn = 0.5$) led to a significantly higher number of editions than JM ($Mdn = 0$, $p = .048$), DM ($Mdn = 0$, $p = .012$), and WO ($Mdn = 0$, $p = .032$).

For the TASK *Para.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3, N = 23) = 27.470$, $p < .001$, $W = .398$) and number of editions ($\chi^2(3, N = 23) = 9.246$, $p = .026$, $W = .134$). Post hoc tests showed that DM ($Mdn = 1.756$) and WO ($Mdn = 1.776$) were faster than JM ($Mdn = 2.69$) and BL ($Mdn = 2.98$) (JM-DM: $p = .01$, JM-WO: $p = .005$, DM-BL: $p < .001$, and WO-BL: $p < .001$). While no significant differences were found among the four techniques on the number of editions.[1]

For the TASK *Two Para.*, we found a significant main effect of TECHNIQUE on time ($\chi^2(3) = 33.7$, $p < .001$, $W = .468$) and number of editions ($\chi^2(3) = 14.949$, $p = .002$, $W = .208$). Post hoc results indicate that DM ($Mdn = 1.798$) and WO ($Mdn = 1.616$) led to significantly shorter total time than JM ($Mdn = 2.492$) and BL ($Mdn = 2.749$) (JM-DM: $p = .009$, JM-WO: $p < .001$, DM-BL: $p < .001$, WO-BL: $p < .001$).

For the TASK *Four Words* and TWO SENT., we did not find any significant differences.

6.4 Subjective Measurements

6.4.1 Perceived Workload

Fig. 4(A) shows participants' responses to the NASA-TLX questionnaire. Friedman tests showed that NASA scores in Physical Demand ($\chi^2(3) = 23.423$, $p < .001$, $W = .325$) and Frustration ($\chi^2(3) = 14.035$, $p = .003$, $W = .195$) were significantly different using four techniques. Significant differences were not found in the remaining dimensions and Overall scores ($p > .05$). Post hoc tests showed that DM ($Mdn = 15$) received significantly lower scores in Physical Demand than WO ($Mdn = 35$) and BL ($Mdn = 32.5$) (both $p = .002$). In addition, WO ($Mdn = 25$) received significantly lower scores in Frustration than BL ($Mdn = 37.5$, $p = .014$).

6.4.2 Overall Usability

Results of a Friedman test showed that there was no significant difference in the SUS scores among four techniques ($\chi^2(3) = 2.33$, $p = .506$, $W = .032$). The medians (Mdn) of SUS scores for each technique are 70 (JM), 70 (WO), 75 (DM), and 65 (BL).

6.4.3 User Experience

Friedman test revealed a significant main effect on Pragmatic Quality ($\chi^2(3) = 8.495$, $p = .036$, $W = .117$), Hedonic Quality ($\chi^2(3) = 38.959$, $p < .001$, $W = .541$), and Overall User Experience ($\chi^2(3) = 38.460$, $p < .001$, $W = .534$) among TECHNIQUE conditions. Post hoc tests showed that BL ($Mdn = -1.625$) was rated lower than JM ($Mdn = 1.75$), WO ($Mdn = 1.875$), and DM ($Mdn = 1.5$) on Hedonic Quality (all $p < .001$). Similarly, BL ($Mdn = -0.6875$) was also rated lower than JM ($Mdn = 1.625$), WO ($Mdn = 1.5$), and DM ($Mdn = 1.4375$) in terms of Overall User Experience (all $p < .001$). Regarding the Pragmatic Quality, post hoc tests did not show any significant differences. UEQ results are summarized in Fig. 4(B).

[1]For the TASK *Para.* in mixed trials, we found P3's performances using WO were all identified as outliers and removed. To balance the sample size for Friedman tests, we removed P3 for the analysis in TASK *Para.* in mixed trials (i.e., $N = 23$).

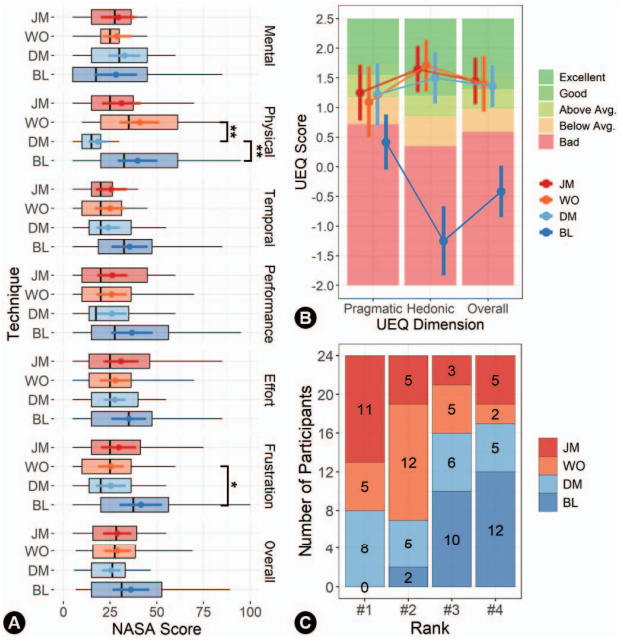

Figure 4: Subjective results. (A) NASA scores. (B) UEQ scores. (C) Participants' rankings. In (A) and (B), error bars represent 95% confidence intervals, '*' and '**' indicate Bonferroni-adjusted p values at $< .05$ and $< .01$ levels in pairwise comparisons, respectively.

6.4.4 User Preferences and Comments on Techniques

Fig. 4(C) shows participants' ranking of the four text selection techniques. Most participants ($N = 22$, 91.67%) disliked BL and ranked them in third or last place. 11 participants (45.83%) ranked JM as the most favored technique, and 12 participants (50%) ranked WO as the second place.

Overall, participants provided positive feedback for the techniques' UI, particularly for WO. Most participants found our UI design logical and provided strong support during the selection. Some participants raised their concerns about the interfaces and suggested improvements. P9, P11, P12, and P21 felt JM required relatively higher costs of learning and memory because they found it difficult to remember the mapping between the joystick's direction and mode. On the other hand, P14, P19, and P24 desired a visual representation of the distance boundaries between each selection mode during the selection process for DM. However, we did not prioritize this approach because it may involve many visual elements and could potentially introduce additional issues, such as occlusion.

For JM, three participants (P5, P10, and P23) felt this technique could be more usable with a system-maintained mechanism (i.e., toggle and lock a mode before text selection) because they felt holding the joystick direction while performing the selection caused physical exertion. As for DM and WO, while these two techniques have already significantly improved the selection performance, participants expected to be able to customize the ranges of each mode to achieve better performance.

7 DISCUSSION

H1 regarding text selection performance is largely supported. In terms of completion time, the proposed techniques were faster than BL (the baseline technique) in sentence-level and paragraph-level tasks. It was not expected that better performance would be observed in selecting characters because the interactions were the same among the techniques. DM and WO were faster than BL in word-level tasks

in experienced trials but not in novice trials. During the experiment, we observed that some participants stuck to the default selection approach when they first met the randomly assigned target snippet in novice trials but transitioned to using the technique in the following experienced trials. This implies that the benefits of using a new technique for new simple tasks might not overcome the cost of recalling its use. This is also reflected in mixed trials where different selection tasks were randomly given. Among the three proposed techniques, JM was slower than DM and WO when selecting one or two paragraphs. With DM or WO, changing the modes maps its spatial relationships. Participants could shift to the paragraph-level selection mode by doing the 'extreme' arm bending or wrist rotation gestures, which may cost less thinking than JM. We explain more regarding this aspect when discussing **H3** in the following sections. On the other hand, the number of editions was considerably low in word-level, sentence-level, and paragraph-level tasks across the techniques and trial types. The proposed techniques performed better in experienced trials but not in novice and mixed trials. One possible reason is that participants felt using BL was more costly, and thus, they were more focused when using BL to avoid mistakes and avoid redoing a trial.

Our **H2** was about techniques' usability and user experience. It was only supported in terms of user experience. Regarding perceived workload, only WO and DM showed significantly lower physical demand compared to BL. As mentioned in the previous section (Sect. 6.4.4), some participants felt tilting the joystick while performing the selection caused extra exertion on their thumb. Some other participants, though not arguing JM directly, expressed their favor of whole-hand gestures. However, this preference is not a shared identity, as seen from Fig. 4(C). P1, a frequent VR HMD user, mentioned that he felt JM was the most natural interaction for him and liked JM the most because it was very similar to the operations in VR games he had played. An unexpected result is that participants felt the proposed techniques significantly improved user experience (according to UEQ scores) but not usability (according to SUS scores). We hypothesize that this is because the proposed techniques are more task-specific, while BL, the default pointing selection technique, has more general use and has literally no learning cost. In addition, although we designed different trial types, it was the first time for participants to use the proposed techniques. We envision that users will find the proposed techniques have higher usability with more use.

The third hypothesis (**H3**) regarding user performance in different expertise was not supported, especially for occasional use (mixed trials). The main cause of this result could be limited practice time given in the user study. In the future, we plan to let participants practice more in the long run to evaluate the techniques further (discussed more in Sect. 9). For the proposed techniques, each of the four selection modes is associated with a unique color. This association provides an intuitive link between the modes and the colors. With direct visual feedback, participants can receive immediate responses during the novice stage. However, as participants became more familiar with the operational logic, they gained more proficiency. With JM, participants could only establish a connection between the four directions of manipulating the joystick with four selection modes, while with DM and WO, participants could associate the selection modes with spatial characters—the distance between the controller and the text panel for DM and the wrist rotation for WO. These explicit gestures, coupled with the accumulation of spatial and visual memory, enabled participants to become increasingly adept at recognizing mode changes, thereby leading to better performance.

Overall, results from the user study show that the three proposed techniques, Joystick Movement, Depth Movement, and Wrist Orientation, can facilitate precise and rapid text selection in VR HMDs and help users select text in different lengths cost-effectively. Some participants have expressed a desire for our techniques to be applied in practical use, which indicates that participants have a strong positive outlook on our designs and believe they hold significant potential for real applications.

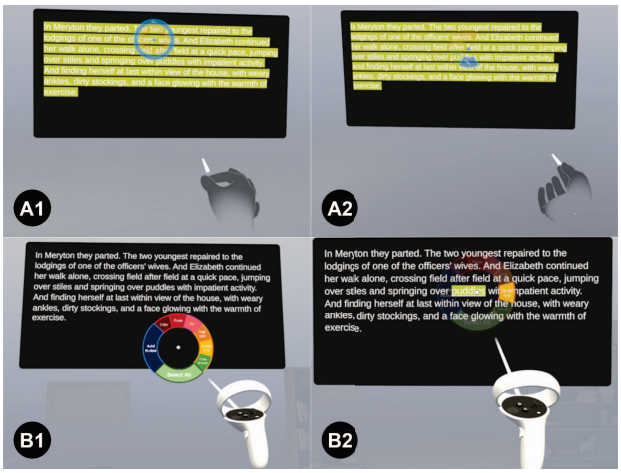

Figure 5: Two Example Extensions. (A1-2) The first example shows freehand versions derived from (A1) DM and (A2) WO. (B1-2) The second example showcases follow-up text edition operations after selecting the text using JM. In (B2), a user, after selecting a text fragment, underlined it.

8 Example Extensions

In this section, we present two mock-up extension applications illustrating possible future developments based on the proposed techniques.

8.1 Freehand Text Selection

DM and WO are two techniques leveraging hand orientation and movement, and thus, they can be used without handheld controllers. Current state-of-the-art VR headsets, such as Meta Quest 2/3 and PICO 4, support 6 degrees of freedom hand-tracking via built-in inside-out cameras. We implemented freehand versions of DM and WO with Quest 2's hand-tracking module, as shown in Fig. 5(A1-A2). However, during the development, we noticed that hand-tracking technologies have certain limitations that may affect text selection performance and experience. We found the hand tracking was less precise or even lost when hands were occluded or out of the tracking space, which was relatively small. Nonetheless, this extension demonstrates the possibilities for freehand interaction, opening up new avenues for exploration in this field. With the arrival of better headsets, such as the Apple Pro Vision, that promise more flawless hand-tracking and spatial interaction capabilities, freehand text selection, such as the one shown in this example, could be more feasible.

8.2 Follow-up Text Editing Operations

Text selection is typically the first step in text editing workflows. In Fig. 5(B1-2), we showcase commonly used functions in the form of a context menu, including three clipboard options (copy, paste, and cut, at the top), highlight options (highlight, underline, and strike-through, on the right), select all, and add notes. We implemented these functions as an extension of the original JM technique via controller input. They are available when text selection is completed to provide users with efficient text editing operations that are seamlessly integrated with text selection. The emergence of these features can broaden the application scope to more practical uses.

9 Limitations and Future Work

We identified four limitations of this research, which could serve as directions for further work.

First, all participants in our user study were right-handed users. The performance of left-handed users could be further investigated. For future studies, we plan to have a broader and more diverse participant pool and investigate individual differences in text selection, such as hand dominance, gender, and prior experience with VR. In addition, we would deliver the demographic questionnaire at the end of experiments to limit stereotype threat in future user studies.

Second, given the complex experimental design with six selection tasks and three task types, we could not provide participants with more training and experimental trials. The performance boundary, especially the upper limit, is still unknown. We plan to conduct long-term user studies to investigate the learning curve and examine user performance at different expertise with more reinforced, long-term use.

Third, we removed two composite tasks involving character-level selection and longer-segment selection from the task set proposed by Goguey et al. [10]. This is because such tasks are not so common in users' daily activities. Nevertheless, the current design of the proposed techniques can complete such tasks by a character-to-character selection approach, if needed. Furthermore, Goguey et al.'s task set is not exhaustive, and there could possibly be other selection combinations that will be discovered in the future that are specific to VR as users start using it more regularly in their daily routines. In the future, we plan to extend our work, which will consider a broader range of selection actions supported by other types of input modalities, such as barehand and hands-free.

Finally, as a first in-depth exploration of new interaction techniques to assist text selection in VR HMDs, we primarily focused on visual feedback during the selection procedure. Recent work has shown that feedback modalities affect user performance at different stages of the text entry process in VR [41]. Therefore, in the future, we want to evaluate the effects of multimodal feedback, such as visual, auditory, and haptic feedback, on text selection performance and provide guidelines for designing VR text selection techniques that could achieve the best user performance and experience.

With advancements in pass-through technologies, VR HMDs are becoming more functional and allow improved blending between the immersive virtual world and the real world. In the future, our techniques could be optimized and tested in a broader range of scenarios, catering to public environments, collaborative work, or a walking task context.

10 Conclusion

Virtual reality (VR) is a platform that supports 3D interaction with increasing promise. Text selection is a highly representative task with a composite process involving pointing, translation, and selection. However, performing text selection precisely and rapidly in 3D VR environments still faces many challenges compared to 2D interactive surfaces (like tablets). In this work, we deeply recognized the complexity by listing the considerations and proposed three controller-based methods: *Joystick Movement*, *Depth Movement*, and *Wrist Orientation* techniques to assist text selection in VR head-mounted displays (HMDs), which were aimed at helping to select text precisely and rapidly at word, sentence, and paragraph levels. We evaluated these techniques with 24 participants across three simulated levels of expertise and six different tasks. Our results demonstrated that, to a great extent, the proposed techniques significantly enhanced the performance and user experience compared to a baseline technique. Our work lays the optimizations of the default raycasting technique and accumulates design knowledge of text selection techniques for VR HMDs.

Acknowledgments

The authors thank the participants who volunteered their time. We also thank the reviewers whose insightful comments and suggestions helped improve our paper. This work was partially funded by the Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (#SZS2022004), the National Natural Science Foundation of China (#62272396; #62207022; #62372212), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (#22KJB520038), and the National Key R&D Program of China (#2022YFC3303604).

References

[1] G. P. K. Abigail J. Sellen and W. A. Buxton. The prevention of mode errors through sensory feedback. *Human–Computer Interaction*, 7(2):141–164, 1992. doi: 10.1207/s15327051hci0702_1

[2] S. Ahn, S. Santosa, M. Parent, D. Wigdor, T. Grossman, and M. Giordano. StickyPie: A gaze-based, scale-invariant marking menu optimized for AR/VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. ACM, New York, NY, USA, 2021. doi: 10.1145/3411764.3445297

[3] T. Ando, T. Isomoto, B. Shizuki, and S. Takahashi. One-handed rapid text selection and command execution method for smartphones. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pp. 1–6. ACM, New York, NY, USA, 2019. doi: 10.1145/3290607.3312850

[4] A. Cockburn, C. Gutwin, J. Scarr, and S. Malacria. Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys*, 47(2), Nov. 2014. doi: 10.1145/2659796

[5] R. Darbar, J. Odicio-Vilchez, T. Lainé, A. Prouzeau, and M. Hachet. Text selection in AR-HMD using a smartphone as an input device. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 526–527. IEEE, New York, NY, USA, 2021. doi: 10.1109/VRW52623.2021.00145

[6] M. De Rosa, V. Fuccella, G. Costagliola, M. G. Albanese, F. Galasso, and L. Galasso. Arrow2edit: A technique for editing text on smartphones. In M. Kurosu and A. Hashizume, eds., *Human-Computer Interaction*, pp. 416–432. Springer Nature Switzerland, Cham, 2023. doi: 10.1007/978-3-031-35596-7_27

[7] 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. ACM, New York, NY, USA, 2018. doi: 10.1145/3170427.3188695

[8] N. Fereydooni and B. N. Walker. Virtual reality as a remote workspace platform: Opportunities and challenges. Available online at: https://www.microsoft.com/en-us/research/publication/virtual-reality-as-a-remote-workspace-platform-opportunities-and-challenges/, last accessed on 31.01.2024, August 2020.

[9] 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*, CHI '20, pp. 1–13. ACM, New York, NY, USA, 2020. doi: 10.1145/3313831.3376173

[10] A. Goguey, S. Malacria, and C. Gutwin. Improving discoverability and expert performance in force-sensitive text selection for touch devices with mode gauges. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–12. ACM, New York, NY, USA, 2018. doi: 10.1145/3173574.3174051

[11] C. Gutwin, A. Cockburn, J. Scarr, S. Malacria, and S. C. Olson. Faster command selection on tablets with FastTap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 2617–2626. ACM, New York, NY, USA, 2014. doi: 10.1145/2556288.2557136

[12] S. G. Hart. Nasa-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi: 10.1177/154193120605000909

[13] K. Hinckley, F. Guimbretiere, P. Baudisch, R. Sarin, M. Agrawala, and E. Cutrell. The Springboard: Multiple modes in one spring-loaded control. In *Proceedings of the SIGCHI Conference on Human Factors*

in Computing Systems, CHI '06, pp. 181–190. ACM, New York, NY, USA, 2006. doi: 10.1145/1124772.1124801

[14] J. Hu, J. J. Dudley, and P. O. Kristensson. An evaluation of caret navigation methods for text editing in augmented reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 640–645. IEEE, New York, NY, USA, 2022. doi: 10.1109/ISMAR-Adjunct57072.2022.00132

[15] Y. Hu Fleischhauer, H. B. Surale, F. Alt, and K. Pfeuffer. Gaze-based mode-switching to enhance interaction with menus on tablets. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, ETRA '23. ACM, New York, NY, USA, 2023. doi: 10.1145/3588015.3588409

[16] G. Kurtenbach and W. Buxton. User learning and performance with marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pp. 258–264. ACM, New York, NY, USA, 1994. doi: 10.1145/191666.191759

[17] G. P. Kurtenbach. *The design and evaluation of marking menus.* PhD thesis, University of Toronto, Canada, 1993.

[18] H. V. Le, S. Mayer, M. Weiß, J. Vogelsang, H. Weingärtner, and N. Henze. Shortcut gestures for mobile text editing on fully touch sensitive smartphones. *ACM Transactions on Computer-Human Interaction*, 27(5), Aug. 2020. doi: 10.1145/3396233

[19] J. R. Lewis. The system usability scale: Past, present, and future. *International Journal of Human–Computer Interaction*, 34(7):577–590, 2018. doi: 10.1080/10447318.2018.1455307

[20] X. Liu, X. Meng, B. Spittle, W. Xu, B. Gao, and H.-N. Liang. Exploring text selection in augmented reality systems. In *Proceedings of the 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '22. ACM, New York, NY, USA, 2023. doi: 10.1145/3574131.3574459

[21] X. Meng, W. Xu, and H.-N. Liang. An exploration of hands-free text selection for virtual reality head-mounted displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 74–81. IEEE, New York, NY, USA, 2022. doi: 10.1109/ISMAR55827.2022.00021

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

[23] D. L. Odell, R. C. Davis, A. Smith, and P. K. Wright. Toolglasses, marking menus, and hotkeys: A comparison of one and two-handed command selection techniques. In *Proceedings of Graphics Interface 2004*, GI '04, pp. 17–24. Canadian Human-Computer Communications Society, Waterloo, CAN, 2004.

[24] C. Park, H. Cho, S. Park, Y.-S. Yoon, and S.-U. Jung. HandPoseMenu: Hand posture-based virtual menus for changing interaction mode in 3D space. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, ISS '19, pp. 361–366. ACM, New York, NY, USA, 2019. doi: 10.1145/3343055.3360752

[25] J. Raskin. *The Humane Interface: New directions for designing interactive systems*. Addison-Wesley Professional, 2000.

[26] R. Rivu, Y. Abdrabou, K. Pfeuffer, M. Hassib, and F. Alt. Gaze'N'Touch: Enhancing text selection on mobile devices using gaze. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pp. 1–8. ACM, New York, NY, USA, 2020. doi: 10.1145/3334480.3382802

[27] A. Ruvimova, J. Kim, T. Fritz, M. Hancock, and D. C. Shepherd. "Transport Me Away": Fostering flow in open offices through virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. ACM, New York, NY, USA, 2020. doi: 10.1145/3313831.3376724

[28] E. Saund and E. Lank. Stylus input and editing without prior selection of mode. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST '03, pp. 213–216. ACM, New York, NY, USA, 2003. doi: 10.1145/964696.964720

[29] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6):103–108, 2017. doi: 10.9781/ijimai.2017.09.001

[30] R. Shi, N. Zhu, H.-N. Liang, and S. Zhao. Exploring head-based mode-switching in virtual reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 118–127. IEEE, New York, NY, USA, 2021. doi: 10.1109/ISMAR52148.2021.00026

[31] J. Smith, I. Wang, J. Woodward, and J. Ruiz. Experimental analysis of single mode switching techniques in augmented reality. In *Proceedings of the 45th Graphics Interface Conference*, pp. 1–8, 2019. doi: 10.20380/GI2019.20

[32] Z. Song, J. J. Dudley, and P. O. Kristensson. Efficient special character entry on a virtual keyboard by hand gesture-based mode switching. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 864–871. IEEE, New York, NY, USA, 2022. doi: 10.1109/ISMAR55827.2022.00105

[33] H. B. Surale, F. Matulic, and D. Vogel. Experimental analysis of mode switching techniques in touch-based user interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 3267–3280. ACM, New York, NY, USA, 2017. doi: 10.1145/3025453.3025865

[34] H. B. Surale, F. Matulic, and D. Vogel. Experimental analysis of barehand mid-air mode-switching techniques in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 1–14. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300426

[35] N. Sutrich. Microsoft Office is finally available on Meta Quest headsets. Android Central, December 2023. Available online at: https://www.androidcentral.com/gaming/virtual-reality/microsoft-office-is-finally-available-on-meta-quest-headsets, last accessed on 31.01.2024.

[36] H. Tu, B. Gao, H. Wu, and F. Lyu. Text Pin: Improving text selection with mode-augmented handles on touchscreen mobile devices. *International Journal of Human-Computer Studies*, 175:103028, 2023. doi: 10.1016/j.ijhcs.2023.103028

[37] H. Tu, X.-D. Yang, F. Wang, F. Tian, and X. Ren. Mode switching techniques through pen and device profiles. In *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction*, APCHI '12, pp. 169–176. ACM, New York, NY, USA, 2012. doi: 10.1145/2350046.2350081

[38] T. Wan, R. Shi, W. Xu, Y. Li, K. Atkinson, L. Yu, and H.-N. Liang. Hands-free multi-type character text entry in virtual reality. *Virtual Reality*, 28(8):1–19, 2024. doi: 10.1007/s10055-023-00902-z

[39] T. Wan, Y. Wei, R. Shi, J. Shen, P. O. Kristensson, K. Atkinson, and H.-N. Liang. Design and evaluation of controller-based raycasting methods for efficient alphanumeric and special character entry in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2024. doi: 10.1109/TVCG.2024.3349428

[40] W. Xu, X. Meng, K. Yu, S. Sarcar, and H.-N. Liang. Evaluation of text selection techniques in virtual reality head-mounted displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 131–140. IEEE, New York, NY, USA, 2022. doi: 10.1109/ISMAR55827.2022.00027

[41] C. Yildirim. Point and Select: Effects of multimodal feedback on text entry performance in virtual reality. *International Journal of Human-Computer Interaction*, pp. 1–15, 2022. doi: 10.1080/10447318.2022.2107330

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

[43] M. Zhao, H. Huang, Z. Li, R. Liu, W. Cui, K. Toshniwal, A. Goel, A. Wang, X. Zhao, S. Rashidian, F. Baig, K. Phi, S. Zhai, I. Ramakrishnan, F. Wang, and X. Bi. EyeSayCorrect: Eye gaze and voice based hands-free text correction for mobile devices. In *27th International Conference on Intelligent User Interfaces*, IUI '22, pp. 470–482. ACM, New York, NY, USA, 2022. doi: 10.1145/3490099.3511103