

# Evaluating the Usability of Microgestures for Text Editing Tasks in Virtual Reality

Xiang Li, Wei He, and Per Ola Kristensson

**Abstract**—As virtual reality (VR) continues to evolve, traditional input methods such as handheld controllers and gesture systems often face challenges with precision, social accessibility, and user fatigue. We introduce microGEXT, a lightweight microgesture-based system designed for text editing in VR without external sensors, which utilizes small, subtle hand movements to reduce physical strain compared to standard gestures. We evaluated microGEXT in three user studies. In Study 1 ( $N = 20$ ), microGEXT reduced overall edit time and fatigue compared to a baseline system. Study 2 ( $N = 20$ ) found that microGEXT performed well in short text selection tasks but was slower for longer text ranges. In Study 3 ( $N = 10$ ), participants found microGEXT intuitive for open-ended information-gathering tasks. Across all studies, microGEXT demonstrated enhanced user experience and reduced physical effort, offering a promising alternative to traditional VR text editing techniques.

**Index Terms**—Microgesture, text editing, text selection, gestural interface, virtual reality, mixed reality

## I. INTRODUCTION

As virtual reality (VR) continues to expand into various fields, the demand for effective input methods has become more critical than ever. Imagine a future where people rely on portable VR work environments, in such settings, prolonged use of controllers or gestures—whether through handheld devices or body movements—poses significant challenges [1], [2], particularly regarding fatigue, such as the well-known “gorilla arm effect” [3], [4]. Furthermore, in confined spaces, such as airplanes [5] or buses [6], using large-scale whole-body movements [7]–[10] or extending the arms [11], [12] for interaction could either disturb others or be entirely impractical due to spatial constraints [13]. These limitations can seriously hinder the effectiveness of VR input methods in real-world scenarios.

The *gorilla arm* effect, common during VR text input, arises from extended arm positions used for virtual keyboard interaction, increasing torque on the shoulder and elbow joints—3.77 times more at the shoulder and twice as much at the elbow compared to relaxed arm positions [3]. Such postures cause fatigue, making tasks like character selection or text editing cumbersome. While large-scale gestures exacerbate this issue, microgestures offer a physically less demanding alternative while retaining the benefits of gesture-based interaction [14], [15]. These subtle, minimal movements reduce physical strain and are well-suited for tasks requiring fine control.

Xiang Li and Per Ola Kristensson are with the Department of Engineering, University of Cambridge, Cambridge, United Kingdom. E-mail: {xl529, pok21}@cam.ac.uk

Wei He is with the Thrust of Urban Governance and Design, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China. E-mail: whe694@connect.hkust-gz.edu.cn

To address these challenges, we introduce *microGEXT*, a microgesture-based system for text editing in VR. Unlike text entry, text editing benefits more from shortcuts [16] that enable efficient structured interactions. Our microGEXT system leverages built-in VR cameras to detect small, ergonomic movements without external hardware, enabling precise control for tasks such as caret navigation and text selection [17] while minimizing fatigue.

We evaluate the usability of microGEXT for VR text editing tasks in three studies. In Study 1 ( $N = 20$ ), participants performed common text editing tasks (e.g., navigation, selection, copy-paste) to compare microGEXT with the baseline. Results showed no significant difference in overall edit times, with microGEXT being faster in some tasks. Participants also reported improved efficiency and satisfaction. Building on this, Study 2 ( $N = 20$ ) focused on microGEXT’s accuracy and speed in precise text range selection, such as highlighting characters or paragraphs. While microGEXT matched the Baseline for shorter selections, it required more time for longer ones but was rated as smoother, less demanding, and less frustrating overall. It was also ranked significantly easier to use, higher in presence, and more preferred, while notably reducing perceived fatigue compared to the Baseline. Finally, Study 3 ( $N = 10$ ) explored microGEXT in open-ended information-gathering tasks, where participants selected, copied, and pasted data between a web browser and a note-taking app in VR. Feedback highlighted microGEXT’s intuitive, efficient, and fatigue-free performance, particularly in enabling seamless task switching during extended sessions.

In summary, our work makes the following contributions:

- We introduce **microGEXT**, a lightweight microgesture-based framework that facilitates precise and efficient text editing in VR, achieving high recognition accuracy using built-in cameras, without external sensors.
- We demonstrate that microGEXT significantly reduces edit times for commands like CUT, DELETE, and SELECT ALL, while lowering physical demand, mental effort, and frustration. It performs well across diverse text range selection tasks, though less effectively for extended ranges, and delivers higher user satisfaction and lower fatigue in structured and open-ended VR text editing scenarios.

## II. RELATED WORK

### A. Text Editing in Real World and Virtual Environments

Text selection and editing in immersive virtual environments (VEs) present unique challenges due to the spatial constraints and limitations of traditional input devices like keyboards and

mice. Song et al. explored these challenges, highlighting the inefficiencies of conventional devices in VEs and proposing controller-based systems as an alternative to improve text selection performance [18]. De Rosa et al.'s Arrow2edit further addressed the spatial interaction difficulties inherent in VEs, focusing on precision-driven tools for text editing [19].

Directional motion-based text selection systems have also gained attention. Xu et al. developed DMove, a system leveraging directional gestures for more efficient text selection in virtual spaces [8]. Wambecke et al. proposed M[eye]cro, integrating eye-tracking with microgestures to enhance user performance in text selection tasks [20]. Similarly, Dudley et al. introduced the VISAR Keyboard, which employs spatial gestures to augment text selection capabilities [21]. Hajika et al. expanded the toolkit for text selection with Radar-Hand, a radar-based gesture tracking system designed for high precision in immersive settings [22]. Kim et al. added tactile feedback to the interaction repertoire with VibAware, a system providing physical sensations to support intuitive and responsive text selection [23].

### B. (Micro)Gestures for Interaction

Gestures have emerged as a promising alternative to conventional input devices, particularly in scenarios where traditional methods are impractical. Sellier et al. demonstrated the natural interaction advantages of gesture interfaces for digital manipulation [24], while Wobbrock et al. highlighted the value of user-defined gestures for tailoring interactions to individual preferences, especially on touch surfaces [25]. Le et al. advanced gesture-based workflows with shortcut gestures for text editing, improving efficiency [16]. Combining modalities, Slambekova et al. found that eye-hand gesture integration enhanced text editing speed and accuracy [26]. Beyond optical tracking, Nandakumar et al. proposed FingerIO, a sonar-based gesture tracking system enabling versatile real-world applications [27].

Microgestures have shown significant potential in immersive environments. Chan et al. studied user preferences for microgestures [14], while Faisandaz et al. demonstrated the effectiveness of eyes-free microgestures in immersive contexts [28]. Sharma et al. introduced grasping microgestures for precise text manipulation [29] and SoloFinger for one-handed text editing on mobile devices [30]. Kandoi et al. emphasized intentional microgestures for tasks requiring high precision [31]. Additional studies have explored novel applications of microgestures. Freeman et al. examined rhythmic microgestures for subtle interactions [32], while Soliman et al.'s FingerInput enabled single-hand microgestures in constrained spaces [33]. Tan et al. developed BikeGesture, which extended gesture control to mobile contexts such as biking [34]. Vatavu et al. introduced iFAD, a multitasking framework utilizing microgestures for real-time editing and navigation [35].

## III. MICROGEXT: A MICROGESTURE RECOGNITION FRAMEWORK FOR TEXT EDITING

Previous studies have highlighted the benefits of gestures and microgestures in immersive interactions, but a fully intu-

itive text editing experience without external cameras, wearables, or specialized hardware remains undeveloped. Many existing solutions are limited by hardware demands or induce fatigue from prolonged mid-air hand use. Our research aims to develop a lightweight VR text editing system that replaces large body movements with subtle microgestures, reducing physical strain. While it may not surpass traditional VR methods with menus and larger gestures, we anticipate comparable efficiency with an improved user experience.

### A. Gesture Dataset

We used the Meta Quest Pro VR headset with the XR Hand package<sup>1</sup> to capture hand skeleton data for gesture recognition. In a user-elicitation study on single-hand microgestures, Chan et al. presented a resulting gesture set and identified prevalent conceptual themes among the elicited gestures. Following this dataset, we included seven distinct gestures along with a *Null* class for a set of text editing tasks in VR (see section III-D).

Ten participants were recruited to interact with a text-editing application, contextualizing each gesture. Before data collection, participants viewed a demonstration video and then performed each gesture 20 times. For static gestures, data was clipped to a standard 2-second duration, while dynamic gestures were clipped to capture complete movement, averaging 5 seconds per gesture. Data was recorded at the Quest Pro's native frame rate of 72 Hz. We also applied Multi-State Gestures for the dynamic *Swipe* gesture. This gesture was segmented into sub-states (0–3), with each frame in a clip of length  $T$  receiving a corresponding sub-state label, enabling more granular gesture phase analysis.

Participants synchronized their swipe with a visual clock in VR, sliding their index finger from 0% to 100% and back within 5 seconds, allowing for automatic sub-state labeling. Smaller sub-state labels denoted movements near the INDEXTIP, while larger labels indicated positions closer to the finger base (INDEXDISTAL). Other gestures were assigned a single sub-state label ‘4.’ This sub-state design improved classification and analysis of gesture phases by enabling finer segmentation of dynamic gestures. Also, the ability to ignore Null gestures is important in preventing false activations when the user is not performing intentional gestures since the gesture recognition system is delimiter-free. We included a *Null* gesture class to capture common hand movements (e.g., resting hands or pinching) that the user may perform unintentionally between deliberate gestures. Participants were asked to perform gestures while ray casting or pinching, and the data was collected, each lasting approximately 2 seconds.

### B. Model Architecture

For our uni-manual gesture recognition study, we adapted the HotGestures multi-task deep learning architecture [36], known for its effectiveness in static and dynamic gesture recognition. As shown in Figure 1, the model processes hand skeleton sequences with advanced spatial and temporal encoding to handle various gesture types. It uses hand skeleton

<sup>1</sup><https://docs.unity3d.com/Packages/com.unity.xr.hands@1.4/manual/>

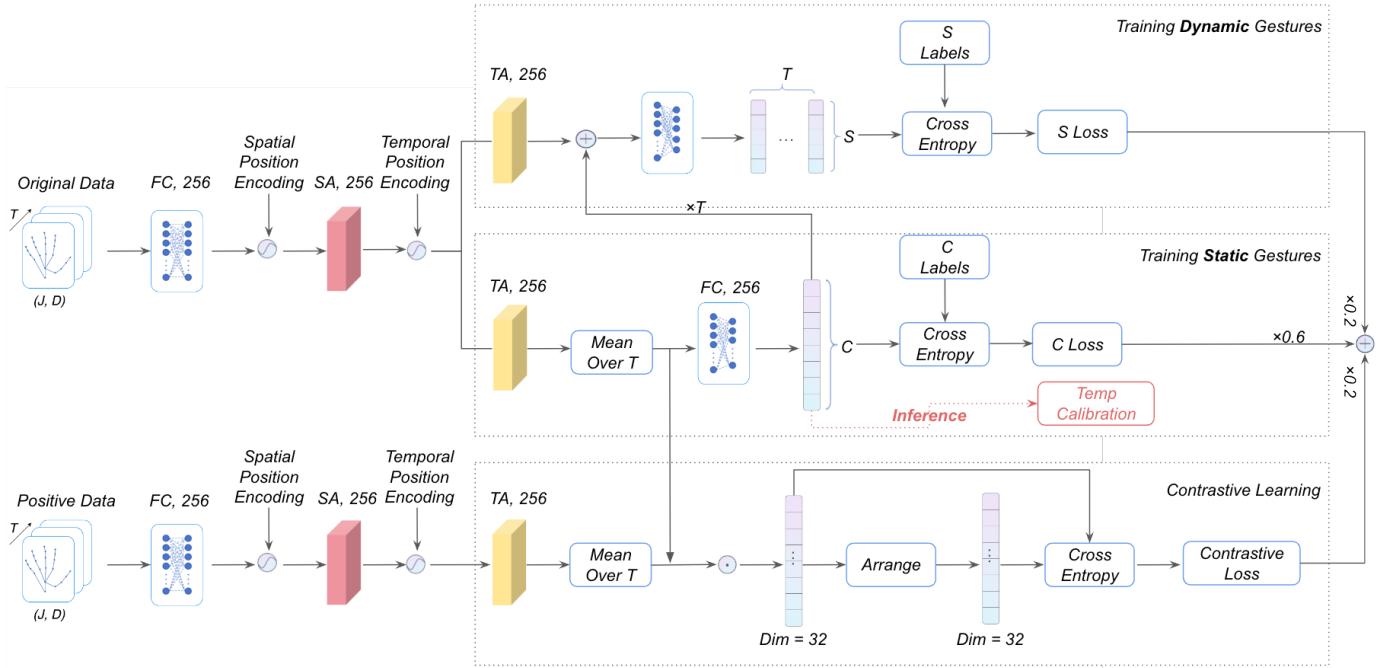


Fig. 1. Overview of the recognition network and framework with contrastive learning. The network consists of Spatial and Temporal Position Encoding, Temporal Attention (TA), and Fully Connected (FC) layers (with ReLU activation function and layer-normalization). The model processes dynamic gestures using a sub-state prediction branch and static gestures using a classification branch. Contrastive learning is applied to enhance the model's ability to distinguish between gestures. The total loss is a weighted combination of dynamic, static, and contrastive losses. Numbers in each block represent the output dimensions. The temperature scaling layer was added before the final prediction output in the calibration process.

graphs with dimensions  $(J, D)$ , where  $J$  is the number of joints, and  $D$  is the feature dimensions per joint. The data passes through a fully connected layer (256 units) for feature extraction, followed by a Spatial Position Encoding (256 units) module for joint relationships and a Temporal Position Encoding module for time-based information. A Temporal Attention (256 units) block highlights key temporal segments, which are crucial for distinguishing dynamic gestures and filtering irrelevant frames in static ones. A temperature scaling layer is utilized before the output to enhance model calibration, improving confidence in predictions [37].

1) *Dynamic Gesture Recognition*: Our model uses a specialized pipeline for dynamic gesture recognition. After the TA layer, the extracted features are concatenated with those from the initial fully connected layer. This combined representation is passed through another fully connected layer (FC, 5), followed by  $T$  parallel branches, where  $T$  is the number of frames in the sequence. Each branch outputs a sub-state probability  $S$ , allowing frame-by-frame analysis of dynamic gestures. The cross-entropy loss for this branch is computed and scaled by a factor of 0.2.

2) *Static Gesture Recognition*: For static gesture recognition, the output of the TA layer is pooled across the temporal dimension using mean pooling, followed by a fully connected layer (FC, 8) to produce the final classification output  $C$ . This branch also applies cross-entropy loss, scaled by 0.6.

3) *Contrastive Learning*: We introduced incorporated contrastive learning to enhance the original model's discriminative capabilities [38]. Positive gesture samples follow a similar pipeline as the static gesture branch, with the addition of a

feature arrangement step before applying a contrastive loss function. This contrastive learning component, weighted by 0.2, significantly helps the model learn more robust features by contrasting similar and dissimilar gesture samples.

Finally, the overall loss for the model is a weighted combination of dynamic gesture loss, static gesture loss, and contrastive loss, ensuring a balanced optimization objective that accommodates the diverse nature of gesture recognition.

### C. Training and Implementation Details

We set the window size to  $T = 20$  and used  $J = 11$  joints (all fingertips, one joint below each fingertip, and the wrist root) as suggested by Song et al. [36]. The feature dimension was set to  $D = 7$ , representing the relative 3D position and 4D quaternion rotation of 10 joints relative to the wrist. For training and evaluation, we selected one participant's data as the test set and used the remaining data for training, employing cross-validation across participant combinations. The training was conducted using PyTorch on a system with an Nvidia GeForce RTX 4090 GPU and an Intel Core i9-13900K CPU.

We used the Adam optimizer with a learning rate of 0.1%, a learning rate scheduler that reduces on a plateau with patience of 5 epochs, and a batch size of 32. Null gestures were included in the training to prevent false positives. The model loss function combined three Cross-Entropy (CE) losses: one for gesture classification, one for sub-state prediction, and one for contrastive learning, represented by  $\mathcal{L}_{\text{class}}$ ,  $\mathcal{L}_{\text{state}}$ , and  $\mathcal{L}_{\text{contrastive}}$ . The loss weights were  $\alpha = 0.6$ ,  $\beta = 0.2$ , and  $\gamma = 0.2$ , corresponding to classification, sub-state, and contrastive tasks, respectively.

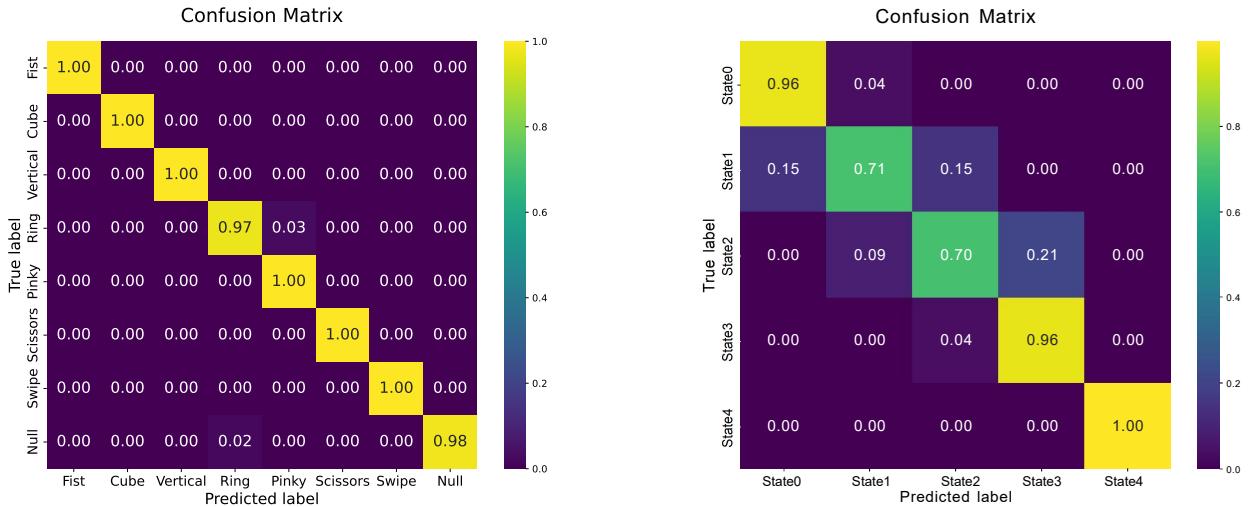


Fig. 2. Class and State Confusion Matrices for gesture recognition performance. The Class Confusion Matrix (left) shows the classification accuracy for each gesture class, with high accuracy across most classes, though some minor misclassifications are observed for gestures like “Null” and “Ring”. The State Confusion Matrix (right) displays the accuracy of sub-state predictions for the “Swipe” gesture, with good performance in most states, though state transitions show some misclassification between adjacent sub-states.

$$\mathcal{L}_{\text{total}} = \alpha \times \mathcal{L}_{\text{class}} + \beta \times \mathcal{L}_{\text{state}} + \gamma \times \mathcal{L}_{\text{contrastive}} \quad (1)$$

1) *Offline Gesture Recognition*: Figure 2 shows the confusion matrices for the window-level class prediction and sub-state prediction on the validation set. For the class prediction, all gesture recognition accuracy is above 97%, indicating the good performance of our recognition model. “Null” and “Ring” gestures have relatively low accuracy since the “Ring” gesture is as easy to detect as the “Pinky” gesture, and the “Null” gesture is similar to the “Cube” gesture. Note that other gestures have 100% accuracy which means the model overfit. For the state prediction, the state ‘4’ has the highest accuracy, and states ‘0’ and ‘3’ are higher than states ‘1-2’. This is expected because class loss weight  $\mathcal{L}_{\text{class}}$  is higher than state loss weight  $\mathcal{L}_{\text{state}}$ , forcing the model to learn better at class prediction and be less sensitive about sub-states. And the 0% and 100% are much easier to recognize than other sub-states.

2) *Online Gesture Recognition*: During the online gesture recognition phase, we implemented a dynamic windowing approach for processing the continuous data stream, rather than relying on a static window size. This allows for real-time gesture detection in virtual reality, where continuous and dynamic input is crucial. A data buffer is maintained, storing the most recent  $T$  frames from the stream. These frames are used as input to predict both the gesture class  $\mathbf{C}$  and the corresponding state sequence  $\mathbf{S}$ .

To enhance accuracy and minimize false recognitions during real-time detection, we incorporated a finite-state machine (FSM), as suggested by Song et al. [36]. The FSM governs the gesture recognition process through three distinct states. Initially, the FSM is in state  $S1$ , where no gestures are detected. When a gesture is identified with a class probability  $\mathbf{C}$  above a predefined threshold  $\delta$ , the FSM transitions to state  $S2$ . In state  $S2$ , if the same gesture class is consistently detected over  $N$  consecutive frames with probabilities exceeding  $\delta$ , the FSM moves to state  $S3$ , finalizes the gesture recognition, and returns

to  $S1$ . Conversely, if a different gesture class is detected in  $S2$  with a probability greater than  $\delta$ , the FSM stays in  $S2$  without transitioning to  $S3$ . If at any point the model’s predicted class probability drops below  $\delta$ , the FSM immediately reverts to  $S1$ . This system ensures that only gestures with high confidence are processed, effectively reducing false positives and improving the robustness of real-time interaction. We implemented our FSM in Unity and fine-tuned the parameters  $N$  and  $\delta$  based on the results of an in-lab formative study with four participants. During the formative study, they were asked to perform each gesture sequentially to test the speed and accuracy of the model. After completing all the gesture commands, participants were required to give their feedback about the model detection. By experimenting with different parameter values, we found  $N = 10$  and  $\delta = 0.95$  yield the best user experience and lowest error rate.

#### D. Gestures and Features

The microGEXT system utilizes a set of microgestures designed to facilitate intuitive and efficient text editing in virtual environments. These gestures are performed with the dominant hand, while the non-dominant hand controls range selection modes [39]. The following sections detail the interaction flow of our selected microgestures and their corresponding features that we chose as suggested by a user elicitation study [14].

1) *Cut (Scissor Gesture)*: The *Scissor* gesture mimics a cutting motion with the index and middle fingers, activating the CUT function to remove selected text and store it in the clipboard. Users select text, perform the gesture, and the text is cut—providing an intuitive way to delete content.

2) *Copy (Ring Gesture)*: The *Ring* gesture, formed by connecting the index finger and thumb, triggers the COPY function to copy selected text without removing it. This efficient gesture enables copying without external controllers.

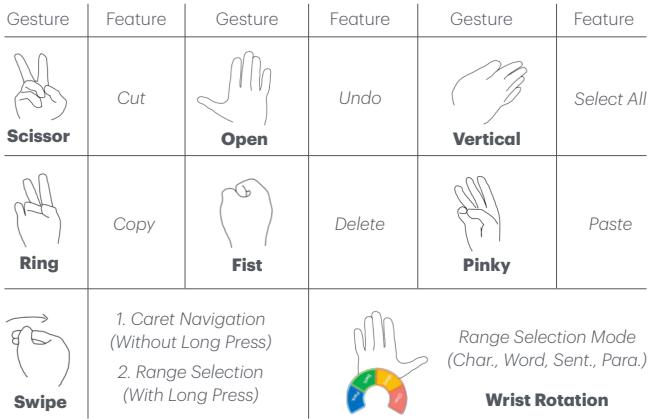


Fig. 3. Our microGEXT’s microgestures for text editing tasks. The right hand (dominant hand) performs gestures such as Scissor (Cut), Ring (Copy), Swipe (Caret Navigation and Range Selection), Open (Undo), Fist (Delete), Vertical (Select All), and Pinky (Paste). The left hand (non-dominant hand) is used for wrist rotation to control the range selection mode, allowing the user to switch between selecting characters, words, sentences, and paragraphs.

### 3) Caret Navigation and Range Selection (Swipe Gesture):

Inspired by DigitSpace [40] and PinchWatch [41], the *Swipe* gesture slides the thumb along the index finger, offering:

- **Caret Navigation:** A simple swipe moves the caret within text for precise positioning.
- **Range Selection:** With a long press, the swipe highlights text, allowing efficient selection.

4) Undo (Open Gesture): The *Open* gesture, represented by an open palm, activates UNDO to reverse the last action, enabling quick corrections.

5) Delete (Fist Gesture): The *Fist* gesture closes the hand into a fist to activate DELETE, removing selected text without copying it to the clipboard. This straightforward gesture facilitates quick text deletion.

6) Select All (Vertical Gesture): The *Vertical* gesture, performed by raising all five fingers, triggers SELECT ALL, allowing users to highlight all text in a document efficiently.

7) Paste (Pinky Gesture): Extending the pinky while folding other fingers triggers the PASTE function, enabling users to quickly insert it at the caret’s position.

8) Range Selection Mode (Wrist Rotation): The left hand controls *Range Selection Mode* through wrist rotation, cycling between character, word, sentence, and paragraph selection [18]. This intuitive interaction adjusts selection granularity without disrupting workflow [42].

## IV. USER STUDY 1: COMPARING MICROGEXT AND BASELINE CONDITIONS FOR TEXT EDITING COMMAND

Study 1 investigates the feasibility of using microgestures for text editing tasks in virtual environments. We conducted a within-subjects user study ( $N = 20$ ), evaluating the impact of microGEXT on usability (e.g., user experience, system usability, and perceived workload) and utility (i.e., edit times and reattempts). The microGEXT was compared to a Baseline condition (see Section IV-A1) to assess overall effectiveness.

While the lightweight framework for microgesture detection is robust, we anticipate that microGEXT might not show

significant advantages over the Baseline condition in terms of speed and accuracy [43], due to challenges with gesture memorability and potential misrecognition (**H1**). However, we expect that participants will prefer using microGEXT, even with occasional but acceptable reattempts, as the minimal physical effort required could significantly reduce fatigue, making it a more comfortable option for extended use (**H2**).

### A. Method

1) *Baseline:* In comparison to the microGEXT system, we selected the text editing tool selection method from the OpenXR<sup>2</sup> package as the Baseline condition, representing a conventional and well-established approach for tool selection in text editing tasks. This Baseline used a menu-based system for text editing, similar to a pop-up text editing menu on a PC, combined with a pinch gesture to confirm the action. The Baseline was adapted from sample implementations within the OpenXR package. Participants could use ray casting to select the desired text editing tool, such as COPY, PASTE, and CUT, and then perform a pinch gesture to confirm the selection. The menu interface design adhered closely to the default OpenXR sample menus, ensuring aesthetic consistency. This Baseline was chosen because it reflects a widely adopted hand interaction technique, allowing users to quickly familiarize themselves with text editing tasks in virtual environments.

2) *Participants and Apparatus:* A total of 20 participants (13 male, and 7 female) were recruited from a local university. The age range of the participants was between 17 and 28 years ( $M = 23.25$ ,  $SD = 2.71$ ). All participants are students and right-handed. None of the participants involved in data collection participated in the user study. In addition to basic demographic information, all participants reported previous VR experience and the familiarity range was between 1 and 7 ( $M = 4.25$ ,  $SD = 1.71$ ). The virtual environment was provided via a Meta Quest Pro VR HMD with hand tracking enabled for interaction. The program was developed in Unity Engine (version 2022.3.3f1c1) with the Open XR Hand package (version 1.4.1). The HMD was connected to a high-performance computer via Quest Link equipped with an Intel i9-13900K CPU, an NVIDIA GeForce RTX 4090 GPU, and 64GB RAM.

3) *Procedure:* Before the experiment, participants signed a consent form and completed a demographic questionnaire. They were then shown a video demonstration introducing each of the seven gestures. Following this, participants viewed video clips of the gestures associated with each tool, as described in the previous section, and were instructed to memorize them as much as possible. Participants were then allowed to practice all eight gestures with feedback provided by the gesture recognition system. No data were recorded during this practice phase, which ensured that participants fully understood how to use both the Baseline and microGEXT systems for text editing tasks. If a participant was unable to complete the practice session successfully, they would not have been allowed to proceed with the formal study or receive compensation. The formal experiment followed, with participants performing tasks under

<sup>2</sup><https://mbucchia.github.io/OpenXR-Toolkit/>

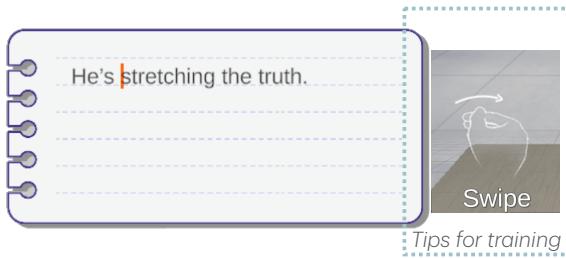


Fig. 4. Screenshot from the training session in Study 1. The gesture instructions, shown in the blue dashed box, provide guidance on how to perform the required gestures (e.g., “Swipe”). These tips were only available during the training session and were hidden during the formal sessions.

two conditions: Baseline and microGEXT. The order of these conditions was counterbalanced across participants using a Latin square design. The experiment consisted of instruction-based random tasks, where participants used text editing tools without selecting specific text snippets. After each condition, participants were given a break and asked to complete a questionnaire assessing their experience. Consequently, the overall study comprised a total of 1,600 trials, calculated as 2 (Condition)  $\times$  8 (Command)  $\times$  5 (Round)  $\times$  20 (Participant). On average, the entire study took approximately 40 minutes per participant, and participants received compensation.

4) *Study Design and Tasks:* In this study, participants followed instructions to use one of eight text editing tools to complete each command (e.g., “Please Paste the Words”, “Please Navigate the Caret”). The order of instructions was randomized within each group. In the Baseline condition, participants used ray casting to select a function and confirmed their choice with a pinch gesture. For caret navigation, they pinched to position the caret and confirmed it with the same gesture. Text selection involved pinching to start, dragging the caret to highlight text, and pinching again to confirm, replicating typical VR text editing techniques. The system automatically selected the relevant text, with participants only needing to execute the editing functions by pinching the corresponding button on the text editing toolbar. In the microGEXT condition, participants used specific gestures to complete commands (see Figure 3). For example, the *Ring* gesture was used to execute the COPY function. Gesture tips were provided during the training session but not in the formal tasks (see Figure 4), although researchers reminded participants of the gestures if needed. Similar to the Baseline, participants only needed to execute the gestures without selecting relevant text.

For caret navigation in microGEXT, participants positioned the caret with a pinch, followed by a *Swipe* gesture to fine-tune placement, and confirmed by pressing their fingers together for two seconds. The same process applied to text selection. The long press gesture was detected automatically, ensuring precise timing to avoid errors. If users made a mistake, they had to reset the task with a pinch gesture. After each command, users received visual and auditory confirmation, and a loading bar signalled the preparation for the next command. Participants were instructed to rest their hands during this time. Each condition consisted of six rounds of eight instructions, with

one training round and five formal rounds.

## B. Results

To capture QUANTITATIVE PERFORMANCE, we measured *Edit Time* for each command and for each round of task, which is the time for each editing attempt, including both successful and reattempted executions; and (2) *Reattempts*, which is the number of times participants had to re-execute a command, either due to microGEXT misrecognizing the gesture or participants incorrectly performing a gesture or selecting the wrong menu item.

We also collected QUALITATIVE FEEDBACK, including: (1) *User Experience*, measured using the short version of the User Experience Questionnaire (UEQ-S) [44]; (2) *Perceived Workload*, evaluated using the NASA-TLX [45]; and (3) *System Usability*, measured using the System Usability Scale (SUS) [46]. We used the Shapiro–Wilks tests and Q–Q plots to check the normality distribution of the data. For normally distributed data, we used Welch’s t-test for two-level comparisons, while we used the Mann–Whitney U test for non-normally distributed data for two-level comparisons. To avoid distorting the statistical analysis, we removed outlier data points with an absolute Z-Score greater than 3.

### 1) Quantitative Performance:

a) *Edit Time for Commands:* We compared edit times between the Baseline and microGEXT conditions across various text editing tasks. For CARET NAVIGATION, RANGE SELECTION, PASTE, COPY, and UNDO, no significant differences were observed ( $p > 0.05$ ), indicating comparable performance between the two conditions.

In contrast, significant improvements were found in the CUT, DELETE, and SELECT ALL tasks. For CUT, microGEXT was faster ( $M = 4.95$ ,  $SD = 1.20$ ) compared to the Baseline ( $M = 5.73$ ,  $SD = 0.91$ ), showing a significant difference ( $U = 293.0$ ,  $p = 0.0123$ ,  $r = 0.398$ ). Similarly, for DELETE, microGEXT ( $M = 4.99$ ,  $SD = 1.28$ ) outperformed the Baseline ( $M = 7.24$ ,  $SD = 2.02$ ), with a significant effect ( $U = 332.0$ ,  $p = 0.0004$ ,  $r = 0.565$ ). Lastly, for SELECT ALL, microGEXT ( $M = 4.75$ ,  $SD = 1.55$ ) was faster than the Baseline ( $M = 6.19$ ,  $SD = 1.46$ ), showing significance ( $U = 318.0$ ,  $p = 0.0015$ ,  $r = 0.505$ ).

b) *Overall Edit Time:* Across all tasks, the microGEXT condition had a significantly shorter average edit time ( $M = 6.58$ ,  $SD = 0.78$ ) compared to the Baseline ( $M = 7.72$ ,  $SD = 1.72$ ), as revealed by a Mann–Whitney U test ( $U = 280.0$ ,  $p = 0.0315$ ,  $r = 0.342$ ).

c) *Reattempts:* Participants required significantly more reattempts in the microGEXT condition ( $M = 2.25$ ,  $SD = 1.69$ ) than in the Baseline ( $M = 0.32$ ,  $SD = 0.67$ ), with a strong effect size ( $U = 46.5$ ,  $p < 0.0001$ ,  $r = -0.657$ ).

### 2) Qualitative Feedback:

a) *User Experience Questionnaire (UEQ):* The UEQ-Short results assessed Pragmatic Quality, Hedonic Quality, and Overall User Experience. For Pragmatic Quality, no significant difference was found between the Baseline ( $M = 1.4$ ,  $SD = 1.12$ ) and microGEXT ( $M = 1.35$ ,  $SD = 0.99$ ;  $U = 212.50$ ,  $p = 0.7439$ ,  $r = 0.063$ ). However, for Hedonic

TABLE I  
COMPARISON OF EDIT TIMES, REATTEMPT COUNTS, UEQ, NASA-TLX,  
AND SUS BETWEEN BASELINE AND MICROGEXT. SIGNIFICANT  
DIFFERENCES ARE MARKED AS “\*”, “\*\*”, “\*\*\*”, AND “\*\*\*\*” FOR  
 $p < 0.05$ ,  $p < 0.01$ ,  $p < 0.001$ , AND  $p < 0.0001$ , RESPECTIVELY. BETTER  
RESULTS ARE INDICATED WITH (↑) AND WORSE RESULTS WITH (↓).

Measure	Baseline (M ± SD)	microGEXT (M ± SD)
<b>Average Edit Times by Task</b>		
CARET NAVIGATION	9.52 ± 3.60	9.69 ± 2.02
RANGE SELECTION	17.11 ± 6.82	13.75 ± 2.48
CUT	5.73 ± 0.91 (↓)	4.95 ± 1.20 (*) (↑)
PASTE	6.25 ± 2.04	5.80 ± 1.95
COPY	5.38 ± 0.93	5.43 ± 1.49
UNDO	6.31 ± 1.58	5.68 ± 1.20
DELETE	7.24 ± 2.02 (↓)	4.99 ± 1.28 (****) (↑)
SELECT ALL	6.19 ± 1.46 (↓)	4.75 ± 1.55 (**) (↑)
<b>Overall Edit Time</b>	7.72 ± 1.72 (↓)	6.58 ± 0.78 (*) (↑)
<b>Reattempts</b>		
	0.32 ± 0.67 (↑)	2.25 ± 1.69 (****) (↓)
<b>UEQ</b>		
Pragmatic Quality	1.4 ± 1.12	1.35 ± 0.99
Hedonic Quality	-0.05 ± 1.47 (↓)	1.96 ± 0.70 (****) (↑)
Overall Experience	0.675 ± 1.12 (↓)	1.66 ± 0.70 (**) (↑)
<b>NASA-TLX</b>		
Mental Demand	3.55 ± 1.73	3.85 ± 1.98
Physical Demand	3.95 ± 1.47	3.1 ± 1.80
Temporal Demand	2.85 ± 1.42	3.15 ± 1.69
Performance	3.7 ± 1.45	3.0 ± 1.45
Effort	3.95 ± 1.57	4.0 ± 1.56
Frustration	2.65 ± 1.57	2.25 ± 1.21
<b>SUS</b>		
Overall Usability	69.38 ± 17.26 (↑)	64.88 ± 18.58 (↓)

Quality, microGEXT significantly outperformed the Baseline ( $M = 1.96$ ,  $SD = 0.70$  vs.  $M = -0.05$ ,  $SD = 1.47$ ;  $U = 40.50$ ,  $p < 0.0001$ ,  $r = 0.798$ ). Similarly, for Overall User Experience, microGEXT ( $M = 1.66$ ,  $SD = 0.70$ ) was significantly better than the Baseline ( $M = 0.675$ ,  $SD = 1.12$ ;  $U = 100.00$ ,  $p = 0.0070$ ,  $r = 0.500$ ).

b) **NASA Task Load Index (NASA-TLX)**: No significant differences were found between the Baseline and microGEXT conditions across all six workload subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration ( $p > 0.05$  for all comparisons). For example, Mental Demand scores were similar ( $M = 3.55$ ,  $SD = 1.73$  for Baseline;  $M = 3.85$ ,  $SD = 1.98$  for microGEXT;  $U = 184.50$ ,  $p = 0.6800$ ), as were Physical Demand ( $M = 3.95$ ,  $SD = 1.47$  for Baseline;  $M = 3.1$ ,  $SD = 1.80$  for microGEXT;  $U = 263.50$ ,  $p = 0.0829$ ) and other subscales.

c) **System Usability Scale (SUS)**: The SUS scores, which measure system usability, showed no significant difference between the Baseline ( $M = 69.38$ ,  $SD = 17.26$ ) and microGEXT ( $M = 64.88$ ,  $SD = 18.58$ ;  $U = 227.00$ ,  $p = 0.4728$ ,  $r = 0.135$ ).

d) **Qualitative Comments**: Participants appreciated the system’s gesture recognition. For instance, “*the gesture recognition is very fast and accurate*” [P1]. However, [P4] pointed out that learning the gestures took time and suggested the need for a dedicated gesture memory process to help users familiarize themselves. [P5] described the overall experience as smooth but mentioned, “*micro-adjustments are convenient, but error tolerance is too low*”, especially during fine ad-

justments. [P7] and [P9] also found that the cursor was too sensitive, which sometimes led to mistakes during use. [P12] found the system innovative, stating that “*(microGEXT is) fun and new, providing strong control*”, but [P14] commented that the gestures for actions like copy and paste were “*unintuitive*”, requiring extra effort to memorize. [P13] noted that prolonged use of gestures, particularly swiping, caused hand discomfort.

However, despite these concerns, many participants [P5, P18, P19] found that after becoming familiar with the system, microGEXT significantly reduced hand fatigue, saying “...*after getting used to it, microGEXT is less tiring*” [P19], making it more efficient than traditional methods, like Baseline.

## V. USER STUDY 2: EVALUATING MICROGEXT FOR PRECISE AND RAPID TEXT RANGE SELECTION

We anticipated that microGEXT would offer significant advantages over the Baseline condition for short text range selections, as it requires only a minimal swipe along the finger to complete the task. In contrast, the Baseline required participants to carefully maintain a pinch gesture until the caret matched the target position, which could be more cumbersome; however, for longer text range selections, we expected microGEXT to have a higher time cost due to its two-step interaction process—first requiring participants to rotate their wrist to change the range selection mode, followed by the swipe gesture (H3). Again, we expected that microGEXT would achieve higher subjective preference scores (H4).

### A. Method

1) **Baseline**: In the Baseline condition, users could select text character-by-character during the selection process. To begin, they controlled a ray to point at the starting position of the desired text. After positioning the ray, they pressed and held the trigger button on the controller, moving the ray to extend the selection over the intended text snippets. Once the caret reached the end of the selection, users released the trigger button to complete the selection. To confirm their selection, they pointed the ray at the “Confirm” button located on the left side of the interface and pressed the trigger again to finalize the task. As in many VR systems, a circular cursor appeared on the text panel to indicate the ray’s current position. If users made an error during the selection, however, they were required to restart the task from the beginning, repeating the entire process. This interaction method demanded high precision and was prone to errors, often resulting in frequent task restarts.

2) **Participants and Apparatus**: All previous participants ( $N = 20$ ) from Study 1 took part in User Study 2 in the same room, following the completion of the final questionnaire in Study 1 and a 5-minute break.

3) **Study Design and Tasks**: We followed the same six tasks as described by Song et al. [18], which were initially adapted from Goguey et al. [47]. Participants were required to select a target text snippet in VR as fast and accurately as possible using both conditions: Baseline and microGEXT. The target text snippets varied from six lengths:

- **Four Characters (Four Char.)**: Selecting 4 characters from a 10-character word.

- *Four Words*: Selecting 4 consecutive words.
- *Sentence (Sent.)*: Selecting a complete sentence.
- *Two Sentences (Two Sent.)*: Selecting 2 consecutive sentences.
- *Paragraph (Para.)*: Selecting an entire paragraph.
- *Two Paragraphs (Two Para.)*: Selecting 2 consecutive paragraphs.

The key difference between the text selection tasks in User Study 1 and User Study 2 for the microGEXT condition was the introduction of a mode-switching mechanism for the left wrist, similar to the approaches used by Song et al. [18] and Song et al. [42]. In the microGEXT condition, users activated the mode switch menu by performing a thumb-up gesture. Once activated, the mode selection process was controlled via wrist rotation, with a vertical ray extending from the palm to guide the selection of the minimum selection unit. The mode aligned with the ray was highlighted in purple, providing visual feedback to the user. To confirm the selection, users relaxed their hands by spreading their four fingers. A visual change in the mode canvas, displayed in front of the user's view, indicated the successful mode switch. The system then dynamically adjusted the minimum selection unit based on the selected mode, allowing for a smooth transition between different interaction states.

The text panel parameters were primarily based on those from Song et al. [18], with adjustments made to fit our experimental design. The panel was sized at 996px × 683px, with the text displayed within a designated area of 896px × 615px. We used the LiberationSans SDF font with white text, and the target text snippet was highlighted in red. The formatted text consisted of four paragraphs, totalling approximately 194 words. We selected a font size of 24 and a line spacing of 7.3, with left alignment. The text spanned 21 lines in total, including line breaks, with the longest line containing around 122 characters and the median line length at approximately 62 characters. To ensure readability, we conducted in-lab pilot tests with four participants, all of whom reported that the text was clear and easy to read, without any readability issues.

*4) Procedure:* It is similar to our first user study; participants were given a tutorial video and informed about the task of user study 2 in the following formal experiment. Then, participants also have one round of training sessions to familiarize themselves with the system. Following this, participants can have a short break before conducting the formal user study. Then, participants completed the formal trials for each condition, following the experimental design described in the previous section. The same questionnaires as Study 1 were given right after the completion of a condition, followed by a short break and a post-studies questionnaire assessing their experience (i.e., ease of use, presence, final rating) and perceived fatigue for both Study 1 and 2. Consequently, the overall study comprised a total of 1,200 trials, calculated as 2 (System) × 6 (Instruction) × 5 (Round) × 20 (Participant). The whole study would take approximately 40 minutes.

## B. Results

In User Study 2, we used similar measures to evaluate the QUANTITATIVE PERFORMANCE and QUALITATIVE FEED-

TABLE II

COMPARISON OF AVERAGE AND OVERALL EDIT TIMES, UEQ, NASA-TLX, AND SUS SCORES BETWEEN BASELINE AND MICROGEXT. SIGNIFICANT DIFFERENCES ARE MARKED AS \*, \*\*, \*\*\*, AND \*\*\*\* FOR  $p < 0.05$ ,  $p < 0.01$ ,  $p < 0.001$ , AND  $p < 0.0001$ , RESPECTIVELY. BETTER RESULTS ARE INDICATED WITH ( $\uparrow$ ) AND WORSE RESULTS WITH ( $\downarrow$ ).

Measure	Baseline (M ± SD)	microGEXT (M ± SD)
<b>Average Edit Times by Task</b>		
FOUR CHAR.	28.04 ± 19.49	19.15 ± 4.43
FOUR WORDS	21.23 ± 16.97	19.58 ± 4.75
SENTENCE	21.24 ± 18.83	15.17 ± 3.21
TWO SENTENCES	16.98 ± 9.59	16.07 ± 3.90
PARAGRAPH	9.96 ± 5.67 (***) ( $\uparrow$ )	14.82 ± 4.64 ( $\downarrow$ )
TWO PARAGRAPHS	10.38 ± 3.94 (****) ( $\uparrow$ )	15.51 ± 3.29 ( $\downarrow$ )
<b>Overall Edit Time</b>	17.85 ± 11.42	16.69 ± 3.08
<b>UEQ Scores</b>		
Pragmatic Quality	0.2625 ± 1.55 ( $\downarrow$ )	1.5 ± 1.04 (**) ( $\uparrow$ )
Hedonic Quality	-0.55 ± 1.49 ( $\downarrow$ )	1.96 ± 0.69 (***) ( $\uparrow$ )
Overall Experience	-0.14375 ± 1.31 ( $\downarrow$ )	1.73 ± 0.69 (***) ( $\uparrow$ )
<b>NASA-TLX Scores</b>		
Mental Demand	4.5 ± 1.88 ( $\downarrow$ )	2.9 ± 1.17 (**) ( $\uparrow$ )
Physical Demand	5.0 ± 1.75 ( $\downarrow$ )	3.65 ± 1.46 (*) ( $\uparrow$ )
Temporal Demand	3.6 ± 1.67	3.15 ± 1.50
Performance	3.55 ± 1.64	2.65 ± 1.50
Effort	4.45 ± 1.76 ( $\downarrow$ )	3.2 ± 1.20 (*) ( $\uparrow$ )
Frustration	4.3 ± 1.81 ( $\downarrow$ )	2.2 ± 1.15 (*) ( $\uparrow$ )
<b>SUS Scores</b>		
Overall Usability	55.88 ± 24.69 ( $\downarrow$ )	70.42 ± 13.77 (*) ( $\uparrow$ )

BACK. Specifically, we collected: (1) Edit Time; (2) User Experience; (3) Perceived Workload; and (4) System Usability.

### 1) Quantitative Performance:

*a) Average Edit Time:* The average edit times for six tasks were compared between the Baseline and microGEXT conditions. For shorter text range selection tasks (FOUR CHAR., FOUR WORDS, SENTENCE, and TWO SENTENCES), no significant differences were observed ( $p > 0.05$ ), with both conditions demonstrating comparable performance. For example, in the FOUR CHAR. task, the Baseline mean was  $M = 28.04$ ,  $SD = 19.49$ , while microGEXT had  $M = 19.15$ ,  $SD = 4.43$ . Similarly, in the SENTENCE task, the Baseline mean was  $M = 21.24$ ,  $SD = 18.83$ , compared to microGEXT ( $M = 15.17$ ,  $SD = 3.21$ ).

In contrast, significant differences were observed for longer text range selection tasks. For the PARAGRAPH task, microGEXT was significantly slower ( $M = 14.82$ ,  $SD = 4.64$ ) compared to the Baseline ( $M = 9.96$ ,  $SD = 5.67$ ;  $U = 62.0$ ,  $p = 0.0002$ ,  $r = -0.590$ ). Similarly, in the TWO PARAGRAPHS task, microGEXT ( $M = 15.51$ ,  $SD = 3.29$ ) was significantly slower than the Baseline ( $M = 10.38$ ,  $SD = 3.94$ ;  $U = 40.0$ ,  $p < 0.0001$ ,  $r = -0.684$ ).

*b) Overall Edit Time:* When comparing the overall edit time across all tasks, the Baseline condition ( $M = 17.85$ ,  $SD = 11.42$ ) and microGEXT ( $M = 16.69$ ,  $SD = 3.08$ ) showed no significant difference ( $U = 132.0$ ,  $p = 0.0679$ ,  $r = -0.291$ ).

## C. Qualitative Feedback

*a) User Experience Questionnaire:* The results from the UEQ-Short for User Study 2 were analyzed using the Mann-Whitney U test. For Pragmatic Quality, there was a significant

difference between the Baseline and microGEXT conditions ( $U = 104.00$ ,  $p = 0.0096$ ,  $r = 0.4800$ ). The Baseline condition had a mean score of 0.2625 ( $SD = 1.55$ ), while microGEXT had a mean score of 1.5 ( $SD = 1.04$ ).

For Hedonic Quality, a highly significant difference was found, with microGEXT outperforming the Baseline ( $U = 33.5$ ,  $p < 0.0001$ ,  $r = 0.8325$ ). The Baseline had a mean of -0.55 ( $SD = 1.49$ ), whereas microGEXT had a much higher mean of 1.96 ( $SD = 0.69$ ).

In terms of the Overall User Experience, microGEXT also showed a significant improvement over the Baseline ( $U = 44.00$ ,  $p < 0.0001$ ,  $r = 0.7800$ ). The Baseline condition had a mean of -0.14375 ( $SD = 1.31$ ), while microGEXT achieved a mean of 1.73 ( $SD = 0.69$ ).

*b) User Experience Questionnaire (UEQ):* The UEQ-Short results for Study 2 revealed significant differences in all dimensions. For Pragmatic Quality, microGEXT ( $M = 1.5$ ,  $SD = 1.04$ ) scored significantly higher than the Baseline ( $M = 0.26$ ,  $SD = 1.55$ ;  $U = 104.00$ ,  $p = 0.0096$ ,  $r = 0.480$ ). Hedonic Quality showed a highly significant improvement with microGEXT ( $M = 1.96$ ,  $SD = 0.69$ ) outperforming the Baseline ( $M = -0.55$ ,  $SD = 1.49$ ;  $U = 33.50$ ,  $p < 0.0001$ ,  $r = 0.833$ ). For Overall User Experience, microGEXT ( $M = 1.73$ ,  $SD = 0.69$ ) also significantly surpassed the Baseline ( $M = -0.14$ ,  $SD = 1.31$ ;  $U = 44.00$ ,  $p < 0.0001$ ,  $r = 0.780$ ).

*c) NASA Task Load Index (NASA-TLX):* The NASA-TLX results indicated significant reductions in workload with microGEXT across several subscales. Mental Demand was significantly lower for microGEXT ( $M = 2.9$ ,  $SD = 1.17$ ) compared to the Baseline ( $M = 4.5$ ,  $SD = 1.88$ ;  $U = 300.00$ ,  $p = 0.0062$ ,  $r = 0.500$ ). Physical Demand also decreased significantly with microGEXT ( $M = 3.65$ ,  $SD = 1.46$ ) versus the Baseline ( $M = 5.0$ ,  $SD = 1.75$ ;  $U = 293.50$ ,  $p = 0.0103$ ,  $r = 0.468$ ). Similarly, Effort was significantly lower for microGEXT ( $M = 3.2$ ,  $SD = 1.20$ ) compared to the Baseline ( $M = 4.45$ ,  $SD = 1.76$ ;  $U = 281.00$ ,  $p = 0.0261$ ,  $r = 0.405$ ), and Frustration was markedly reduced with microGEXT ( $M = 2.2$ ,  $SD = 1.15$ ) versus the Baseline ( $M = 4.3$ ,  $SD = 1.81$ ;  $U = 330.50$ ,  $p = 0.0004$ ,  $r = 0.653$ ).

No significant differences were observed for Temporal Demand ( $M = 3.15$ ,  $SD = 1.50$  for microGEXT vs.  $M = 3.6$ ,  $SD = 1.67$  for Baseline;  $U = 231.00$ ,  $p = 0.3987$ ,  $r = 0.155$ ) or Performance ( $M = 2.65$ ,  $SD = 1.50$  for microGEXT vs.  $M = 3.55$ ,  $SD = 1.64$  for Baseline;  $U = 268.00$ ,  $p = 0.0600$ ,  $r = 0.340$ ).

*d) System Usability Scale (SUS):* The SUS results showed a significant improvement in system usability for microGEXT ( $M = 70.42$ ,  $SD = 13.77$ ) compared to the Baseline ( $M = 55.88$ ,  $SD = 24.69$ ;  $U = 104.50$ ,  $p = 0.0101$ ,  $r = 0.478$ ).

*e) Qualitative Comments:* Many appreciated the microGEXT's precision, for example, [P2] noting that "(microGEXT) allows (me) for more precise control of text selection compared to the default system." However, [P10] noted that "the current selection status on the left-hand panel wasn't clearly visible," suggesting it needed better visual clarity.

TABLE III  
COMPARISON OF EASE OF USE, PREFERENCE, PRESENCE, AND PERCEIVED FATIGUE BETWEEN BASELINE AND MICROGEXT CONDITIONS. SIGNIFICANT DIFFERENCES ARE MARKED WITH ‘\*’, ‘\*\*’, AND ‘\*\*\*’, INDICATING SIGNIFICANCE LEVELS AT  $p < 0.05$ ,  $p < 0.01$ , AND  $p < 0.001$ , RESPECTIVELY. BETTER RESULTS ARE INDICATED WITH (↑) AND WORSE RESULTS WITH (↓).

Measure	Baseline ( $M \pm SD$ )	microGEXT ( $M \pm SD$ )
Ease of Use	$4.05 \pm 1.67$ (↓)	$5.60 \pm 0.99$ (***) (↑)
Preference	$4.15 \pm 1.42$ (↓)	$6.00 \pm 0.79$ (****) (↑)
Presence	$4.80 \pm 1.54$ (↓)	$6.10 \pm 1.02$ (**) (↑)
Perceived Fatigue	$12.70 \pm 3.13$ (↓)	$10.60 \pm 2.11$ (*) (↑)

Gesture memorization and sensitivity were common concerns. [P3] remarked, "*the system is easier to use than the default mode, but gestures require practice and are easily forgotten.*" Similarly, [P16] emphasized the need for sensitivity adjustments, as the current settings could lead to occasional misrecognition of gestures.

The left-hand panel was praised for its adaptability, as [P7] highlighted that "...it helps adapt to different text selection scenarios." But some suggested improvements in fluidity [P15] and smoother switching functions noted by [P11], while "*microGEXT worked well for large text blocks, switching between functions could be smoother.*"

#### D. Post-studies Questionnaires

We provided a post-studies questionnaire to evaluate participants' (1) Perceived Fatigue, evaluated by using Borg Rating of Perceived Exertion (RPE) Scale (or Borg 6–20) [48], where 6 means "no exertion at all (rest)" and 20 represents maximal exertion, meaning the person is pushing themselves to their absolute physical limit; (2) Ease of Use; (3) Presence; and (4) Overall Preference, regarding the use of these two systems in both studies. The last three metrics were assessed via a single 7-point Likert scale, where 1 indicated "not at all" and 7 indicated "very much", according to Gugenhimer et al. [49].

*a) Ease of Use:* The microGEXT condition ( $M = 5.60$ ,  $SD = 0.99$ ) was rated significantly easier to use than the Baseline ( $M = 4.05$ ,  $SD = 1.67$ ;  $U = 86.50$ ,  $p = 0.0015$ ,  $r = 0.216$ ).

*b) Preference:* Participants strongly preferred the microGEXT condition ( $M = 6.00$ ,  $SD = 0.79$ ) over the Baseline ( $M = 4.15$ ,  $SD = 1.42$ ;  $U = 52.00$ ,  $p = 0.00004$ ,  $r = 0.130$ ), with a highly significant difference.

*c) Presence:* MicroGEXT provided a stronger sense of presence ( $M = 6.10$ ,  $SD = 1.02$ ) compared to the Baseline ( $M = 4.80$ ,  $SD = 1.54$ ;  $U = 100.50$ ,  $p = 0.0055$ ,  $r = 0.251$ ).

*d) Perceived Fatigue:* Participants reported significantly less perceived fatigue with microGEXT ( $M = 10.60$ ,  $SD = 2.11$ ) than with the Baseline ( $M = 12.70$ ,  $SD = 3.13$ ;  $U = 273.50$ ,  $p = 0.0450$ ,  $r = 0.684$ ).

## VI. USER STUDY 3: EXPLORING MICROGEXT FOR OPEN-ENDED INFORMATION GATHERING IN WEB BROWSING AND NOTE-TAKING

We conducted User Study 3 to allow participants to fully experience microGEXT in an open-ended information-gathering scenario in web browsing and note-taking.

### A. Method

*1) Participants and Apparatus:* We invited all previous participants to participate in User Study 3, conducted at least one day after completing the first two studies. Half of the participants ( $N = 10$ , 5 male, 5 female) from the earlier user studies were willing to participate. The age range of the participants was between 17 and 26 years ( $M = 23.5$ ,  $SD = 2.68$ ), and the VR familiarity range was between 1 and 6 ( $M = 4.4$ ,  $SD = 1.58$ ). The study took place in the same room using the same apparatus as the previous sessions.

*2) Procedure:* Participants in User Study 3 engaged in an open-ended scenario focused on information gathering through web browsing and note-taking, without specific tasks or wrong actions. The study was conducted in a virtual environment, where participants viewed two screens simultaneously: the left screen displayed a web browser with information about various travel-related topics (e.g., mountain locations, opening hours, altitudes, etc.), and the right screen showed a notes application with sample questions (e.g., what is the altitude of the mountain). Participants were required to find the relevant answers from the web browser and then add the proper information into the notes application.

At the beginning of the study, participants were asked to revisit the video tutorial that explained how to use both the microGEXT and Baseline conditions for text editing in browsers and note applications. Following the video, participants are allowed to complete a training session to familiarize themselves with the microGEXT system. The training session involved browsing a website and transferring information to the notes application, mainly by copying and pasting editions. After the training, participants proceeded to the formal round, which consisted of two conditions: one using microGEXT and the other using the Baseline. Each condition involved interacting with a different website to gather and transfer information. After completing the tasks in both conditions, participants filled out the same questionnaires as in the previous studies to assess their experience. The entire study took approximately 30 minutes to complete, and participants received an additional £5 as compensation for their time and effort.

### B. Results

In Study 3, given the smaller group of participants, we focused primarily on subjective questionnaires and semi-structured interviews to gather QUALITATIVE FEEDBACK from participants. Specifically, we collected the feedback across three key areas: (1) User Experience; (2) Perceived Workload; and (3) System Usability. Considering that the scenarios were open-ended and reading speeds varied among individuals, we did not collect edit times during this study.

TABLE IV

COMPARISON OF UEQ, NASA-TLX, AND SUS SCORES BETWEEN BASELINE AND MICROGEXT. SIGNIFICANT DIFFERENCES ARE MARKED WITH ‘\*\*\*’ FOR  $p < 0.01$ . BETTER RESULTS ARE INDICATED WITH (↑) AND WORSE RESULTS WITH (↓).

Measure	Baseline (M ± SD)	microGEXT (M ± SD)
<b>UEQ</b>		
Pragmatic Quality	$1.25 \pm 1.31$	$1.9 \pm 0.59$
Hedonic Quality	$-0.425 \pm 1.31$ (↓)	$1.9 \pm 0.60$ (**) (↑)
Overall Experience	$0.4125 \pm 1.09$ (↓)	$1.9 \pm 0.51$ (**) (↑)
<b>NASA-TLX</b>		
Mental Demand	$3.6 \pm 1.84$	$3.5 \pm 1.58$
Physical Demand	$3.8 \pm 1.23$	$3.6 \pm 1.65$
Temporal Demand	$3.1 \pm 1.10$	$3.3 \pm 1.25$
Performance	$3.55 \pm 1.64$	$2.65 \pm 1.50$
Effort	$4.45 \pm 1.76$	$3.2 \pm 1.20$
Frustration	$4.3 \pm 1.81$	$2.2 \pm 1.15$
<b>SUS</b>		
Overall Usability	$69.0 \pm 21.51$	$74.5 \pm 13.17$

#### 1) Qualitative Feedback:

*a) User Experience Questionnaire (UEQ):* The Pragmatic Quality scores showed no significant difference between the Baseline ( $M = 1.25$ ,  $SD = 1.31$ ) and microGEXT ( $M = 1.9$ ,  $SD = 0.59$ ;  $U = 39.50$ ,  $p = 0.4471$ ,  $r = 0.210$ ). However, significant differences were observed for Hedonic Quality, where microGEXT ( $M = 1.9$ ,  $SD = 0.60$ ) outperformed the Baseline ( $M = -0.425$ ,  $SD = 1.31$ ;  $U = 9.00$ ,  $p = 0.0021$ ,  $r = 0.820$ ). Similarly, Overall User Experience scores were significantly higher for microGEXT ( $M = 1.9$ ,  $SD = 0.51$ ) compared to the Baseline ( $M = 0.4125$ ,  $SD = 1.09$ ;  $U = 14.50$ ,  $p = 0.0080$ ,  $r = 0.710$ ).

*b) NASA Task Load Index (NASA-TLX):* The NASA-TLX results revealed no significant differences between the Baseline and microGEXT conditions across all subscales ( $p > 0.05$ ). For instance, Mental Demand was similar ( $M = 3.6$ ,  $SD = 1.84$  for Baseline;  $M = 3.5$ ,  $SD = 1.58$  for microGEXT;  $U = 49.50$ ,  $p = 1.0000$ ), as were Physical Demand ( $M = 3.8$ ,  $SD = 1.23$  for Baseline;  $M = 3.6$ ,  $SD = 1.65$  for microGEXT;  $U = 56.00$ ,  $p = 0.6700$ ) and Temporal Demand ( $M = 3.1$ ,  $SD = 1.10$  for Baseline;  $M = 3.3$ ,  $SD = 1.25$  for microGEXT;  $U = 44.00$ ,  $p = 0.6682$ ). Performance ( $U = 25.00$ ,  $p = 0.0586$ ), Effort ( $U = 50.50$ ,  $p = 1.0000$ ), and Frustration ( $U = 54.00$ ,  $p = 0.7836$ ) also showed no significant differences.

*c) System Usability Scale (SUS):* The SUS results indicated no significant difference in system usability scores between the Baseline ( $M = 69$ ,  $SD = 21.51$ ) and microGEXT ( $M = 74.5$ ,  $SD = 13.17$ ;  $U = 40.50$ ,  $p = 0.4955$ ,  $r = 0.190$ ).

*d) Qualitative Comments:* Participants in Study 3 highlighted both the strengths and areas for improvement of microGEXT, particularly in relation to its efficiency for continuous tasks. [P9] noted that “...for consecutive operations like copy-paste, microGEXT allows for quick and seamless completion of both tasks, whereas the default method requires separate actions through the panel, which reduces efficiency.” Once familiar with the system, [P9] found microGEXT to be “simple and efficient for note-taking, provided the gesture

*recognition was accurate.”* [P19] echoed the sentiment of efficiency, remarking that “*once familiar with the system, users can quickly perform the intended functions.*” However, they pointed out that the accuracy of the gesture recognition for locking gestures could be improved, and mentioned that the gesture memorization process “*requires some time, making microGEXT slightly more challenging compared to the default method.*” [P18] appreciated the speed and convenience of microGEXT, saying that it was “*much faster than traditional methods and quite convenient.*” However, they observed that “*(microGEXT) sometimes selects unnecessary spaces when recognizing paragraphs or sentences.*”

## VII. DISCUSSION, LIMITATIONS, AND FUTURE WORK

A key finding across all studies is the significant reduction in physical demand and fatigue when using microGEXT. Traditional VR input methods, especially those that rely on large arm movements, are prone to causing users fatigue over time, as exemplified by the *gorilla arm syndrome* [3], [4]. microGEXT addresses this challenge effectively through small, precise hand movements that reduce the physical strain typically associated with text editing tasks in VR. The NASA-TLX results from Studies 1 and 2 support this, showing that microGEXT consistently lowered physical and mental demand, while also reducing frustration, particularly in tasks requiring high precision like text selection and deletion.

However, it is also clear from the results that microGEXT has its limitations, especially for complex tasks requiring long text selections. For example, in Study 2, we found that microGEXT was significantly slower than the Baseline system for paragraph and two-paragraph text selections. This suggests that while microGEXT excels at handling short, quick interactions, its two-step interaction process (wrist rotation to change modes followed by gesture execution) can introduce delays for more complex operations. This highlights the importance of balancing simplicity and efficiency in gesture design. Future iterations of the system could explore optimizing the interaction flow for long-range selections, possibly through more dynamic or context-sensitive gesture controls.

While the accuracy of the microGEXT recognition system was effective in the studies, we see potential to further improve its performance. The recognition model was trained and calibrated using 12 datasets from 6 subjects, each performing the gestures twice. While a larger and more diverse dataset could improve the model’s ability to generalize to new users, the addition of new gestures would require further data collection and model retraining, which poses a scalability challenge. Future work could also explore the use of automated data augmentation techniques or transfer learning methods to improve recognition accuracy without requiring a vast amount of additional training data.

The current microGEXT system incorporates eight dynamic gestures designed for common text editing tasks. As the number of gestures in the system increases, relying solely on gesture-based interaction becomes increasingly challenging. A larger gesture set makes it more difficult for users to remember and recognize gestures, while also requiring more training

data, thus complicating usability and increasing the demand for system resources. To address this, we see interesting future research in investigating how to build systems that could employ a hybrid interaction model that combines microgestures with other input methods, such as voice commands or contextual menus, allowing users to switch between interaction modes depending on task complexity.

Finally, it would be interesting to study the adoption of microGEXT in deployment studies, which would allow a nuanced understanding of barriers, appropriation moves, and other aspects that emerge through interaction with technologies in everyday life [50].

## VIII. CONCLUSION

This paper introduced microGEXT, a microgesture-based system for text editing in virtual reality (VR), designed to address challenges like user fatigue, precision limitations, and social accessibility. We evaluated its usability, efficiency, and user experience through three user studies.

Study 1 ( $N = 20$ ) compared microGEXT with a Baseline gesture-plus-menu system, showing significant reductions in edit time for commands like CUT, DELETE, and SELECT ALL, along with improved user experience. Study 2 ( $N = 20$ ) focused on text range selection, revealing better performance for shorter tasks and reduced physical and mental demand for longer ones. Study 3 ( $N = 10$ ) highlighted microGEXT’s effectiveness in open-ended tasks, minimizing fatigue and supporting seamless task transitions.

Overall, microGEXT demonstrates how microgesture interaction has the potential to bestow VR input methods with interactions that reduce fatigue and increase user satisfaction while retaining all or most performance of traditional mid-air interfaces. We see promising future work in exploring how to further improve recognition accuracy, expand the gesture set, and study and optimize the system in prolonged usage scenarios in real-world applications. To support further research in this area, we provide our open-source code and anonymized datasets as supplemental materials.

## ACKNOWLEDGMENTS

Xiang Li is supported by the China Scholarship Council (CSC) International Cambridge Scholarship (No. 202208320092). Per Ola Kristensson is supported by the EPSRC (grants EP/S027432/1 and EP/W02456X/1).

## REFERENCES

- [1] J. Grubert, E. Ofek, M. Pahud, and P. O. Kristensson, “The Office of the Future: Virtual, Portable, and Global,” vol. 38, no. 6, pp. 125–133, IEEE Computer Graphics and Applications. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8617763>
- [2] V. Biener, E. Ofek, M. Pahud, P. O. Kristensson, and J. Grubert, “Extended Reality for Knowledge Work in Everyday Environments,” in *Everyday Virtual and Augmented Reality*, A. Simeone, B. Weyers, S. Bialkova, and R. W. Lindeman, Eds. Springer International Publishing, pp. 21–56. [Online]. Available: [https://doi.org/10.1007/978-3-031-05804-2\\_2](https://doi.org/10.1007/978-3-031-05804-2_2)

- [3] E. G. Q. Palmeira, A. Campos, I. A. Moraes, A. G. de Siqueira, and M. G. G. Ferreira, "Quantifying the 'Gorilla Arm' Effect in a Virtual Reality Text Entry Task via Ray-Casting: A Preliminary Single-Subject Study," in *Proceedings of the 25th Symposium on Virtual and Augmented Reality*, ser. SVR '23. New York, NY, USA: Association for Computing Machinery, 2024, p. 274–278. [Online]. Available: <https://doi.org/10.1145/3625008.3625046>
- [4] J. D. Hincapié-Ramos, X. Guo, P. Moghadasi, 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*, ser. CHI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 1063–1072. [Online]. Available: <https://doi.org/10.1145/2556288.2557130>
- [5] J. R. Williamson, M. McGill, and K. Outram, "PlaneVR: Social Acceptability of Virtual Reality for Aeroplane Passengers," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. Association for Computing Machinery, pp. 1–14. [Online]. Available: <https://dl.acm.org/doi/10.1145/3290605.3300310>
- [6] W.-J. Tseng, S. Huron, E. Lecolinet, and J. Gugenheimer, "FingerMapper: Mapping Finger Motions onto Virtual Arms to Enable Safe Virtual Reality Interaction in Confined Spaces," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. Association for Computing Machinery, pp. 1–14. [Online]. Available: <https://dl.acm.org/doi/10.1145/3544548.3580736>
- [7] F. 'Floyd' Mueller, R. Patibanda, R. Byrne, Z. Li, Y. Wang, J. Andres, X. Li, J. Marquez, S. Greuter, J. Duckworth, and J. Marshall, "Limited Control Over the Body as Intriguing Play Design Resource," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. Association for Computing Machinery, pp. 1–16. [Online]. Available: <https://dl.acm.org/doi/10.1145/3411764.3445744>
- [8] 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*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–14. [Online]. Available: <https://dl.acm.org/doi/10.1145/3290605.3300674>
- [9] X. Li, W. He, S. Jin, J. Gugenheimer, P. Hui, H.-N. Liang, and P. O. Kristensson, "Investigating Creation Perspectives and Icon Placement Preferences for On-Body Menus in Virtual Reality," *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. ISS, Oct. 2024. [Online]. Available: <https://doi.org/10.1145/3698136>
- [10] F. F. Mueller, N. Semertzidis, J. Andres, J. Marshall, S. Benford, X. Li, L. Matjeka, and Y. Mehta, "Toward Understanding the Design of Intertwined Human-Computer Integrations," *ACM Trans. Comput.-Hum. Interact.*, vol. 30, no. 5, Sep. 2023. [Online]. Available: <https://doi.org/10.1145/3590766>
- [11] 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)*, Oct. 2022, pp. 131–140, iSSN: 1554-7868. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9995581>
- [12] X. Li, Y. Chen, R. Patibanda, and F. F. Mueller, "vrCAPTCHA: Exploring CAPTCHA Designs in Virtual Reality," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3411763.3451985>
- [13] K.-T. Yang, C.-H. Wang, and L. Chan, "ShareSpace: Facilitating Shared Use of the Physical Space by both VR Head-Mounted Display and External Users," in *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '18. Association for Computing Machinery, pp. 499–509. [Online]. Available: <https://dl.acm.org/doi/10.1145/3242587.3242630>
- [14] E. Chan, T. Seyed, W. Stuerzlinger, X.-D. Yang, and F. Maurer, "User Elicitation on Single-hand Microgestures," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 3403–3414. [Online]. Available: <https://dl.acm.org/doi/10.1145/2858036.2858589>
- [15] W. Xu, H.-N. Liang, Q. He, X. Li, K. Yu, and Y. Chen, "Results and Guidelines From a Repeated-Measures Design Experiment Comparing Standing and Seated Full-Body Gesture-Based Immersive Virtual Reality Exergames: Within-Subjects Evaluation," vol. 8, no. 3, p. e17972, company: JMIR Publications Inc., Toronto, Canada. [Online]. Available: <https://doi.org/10.1145/3290605.3300632>
- Publisher: JMIR Publications Inc., Toronto, Canada. [Online]. Available: <https://games.jmir.org/2020/3/e17972>
- [16] 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 Trans. Comput.-Hum. Interact.*, vol. 27, no. 5, pp. 33:1–33:38, Aug. 2020. [Online]. Available: <https://dl.acm.org/doi/10.1145/3396233>
- [17] 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)*, Oct. 2022, pp. 640–645, iSSN: 2771-1110. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9974482>
- [18] J. Song, R. Shi, Y. Li, B. Gao, and H.-N. Liang, "Exploring Controller-based Techniques for Precise and Rapid Text Selection in Virtual Reality," in *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, Mar. 2024, pp. 244–253, iSSN: 2642-5254. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10494188>
- [19] M. De Rosa, V. Fuccella, G. Costagliola, M. G. Albanese, F. Galasso, and L. Galasso, "Arrow2edit: A Technique for Editing Text on Smartphones," in *Human-Computer Interaction*, M. Kurosu and A. Hashizume, Eds. Springer Nature Switzerland, pp. 416–432. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-031-35596-7\\_27](https://link.springer.com/chapter/10.1007/978-3-031-35596-7_27)
- [20] J. Wambecke, A. Goguey, L. Nigay, L. Dargent, D. Hauret, S. Lafon, and J.-S. L. de Visme, "M[eye]cro: Eye-gaze+Microgestures for Multitasking and Interruptions," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. EICS, pp. 210:1–210:22, May 2021. [Online]. Available: <https://dl.acm.org/doi/10.1145/3461732>
- [21] J. J. Dudley, K. Vertanen, and P. O. Kristensson, "Fast and Precise Touch-Based Text Entry for Head-Mounted Augmented Reality with Variable Occlusion," *ACM Transactions on Computer-Human Interaction*, vol. 25, no. 6, pp. 1–40, Dec. 2018. [Online]. Available: <https://dl.acm.org/doi/10.1145/3232163>
- [22] R. Hajika, T. S. Gunasekaran, C. D. S. Y. Haigh, Y. S. Pai, E. Hayashi, J. Lien, D. Lottridge, and M. Billinghurst, "RadarHand: A Wrist-Worn Radar for On-Skin Touch-Based Proprioceptive Gestures," *ACM Transactions on Computer-Human Interaction*, vol. 31, no. 2, pp. 1–36, Apr. 2024. [Online]. Available: <https://dl.acm.org/doi/10.1145/3617365>
- [23] J. Kim, M. Kim, W. S. Lee, and S. H. Yoon, "VibAware: Context-Aware Tap and Swipe Gestures Using Bio-Acoustic Sensing," in *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*. Sydney NSW Australia: ACM, Oct. 2023, pp. 1–12. [Online]. Available: <https://dl.acm.org/doi/10.1145/3607822.3614544>
- [24] Q. Sellier, A. Sluyters, J. Vanderdonckt, and I. Poncin, "Evaluating Gesture User Interfaces: Quantitative Measures, Qualitative Scales, and Method," *International Journal of Human-Computer Studies*, p. 103242, Feb. 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1071581924000260>
- [25] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, "User-Defined Gestures for Surface Computing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09. Boston MA USA: ACM, Apr. 2009, pp. 1083–1092. [Online]. Available: <https://dl.acm.org/doi/10.1145/1518701.1518866>
- [26] D. Slambekova, R. Bailey, and J. Geigel, "Gaze and Gesture Based Object Manipulation in Virtual Worlds," in *Proceedings of the 18th ACM symposium on Virtual reality software and technology*, ser. VRST '12. New York, NY, USA: Association for Computing Machinery, Dec. 2012, pp. 203–204. [Online]. Available: <https://dl.acm.org/doi/10.1145/2407336.2407380>
- [27] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "FingerIO: Using Active Sonar for Fine-Grained Finger Tracking," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. San Jose California USA: ACM, May 2016, pp. 1515–1525. [Online]. Available: <https://dl.acm.org/doi/10.1145/2858036.2858580>
- [28] G. R. J. Faisandaz, A. Goguey, C. Jouffrais, and L. Nigay, "μGeT: Multimodal Eyes-free Text Selection Technique Combining Touch Interaction and Microgestures," in *Proceedings of the 25th International Conference on Multimodal Interaction*, ser. ICMI '23. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 594–603. [Online]. Available: <https://dl.acm.org/doi/10.1145/3577190.3614131>
- [29] A. Sharma, J. S. Roo, and J. Steimle, "Grasping Microgestures: Eliciting Single-hand Microgestures for Handheld Objects," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–13. [Online]. Available: <https://dl.acm.org/doi/10.1145/3290605.3300632>

- [30] A. Sharma, M. A. Hedderich, D. Bhardwaj, B. Fruchard, J. McIntosh, A. S. Nittala, D. Klakow, D. Ashbrook, and J. Steimle, "SoloFinger: Robust Microgestures while Grasping Everyday Objects," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–15. [Online]. Available: <https://dl.acm.org/doi/10.1145/3411764.3445197>
- [31] C. Kandoi, C. Jung, S. Mannan, H. Vanderhoeven, Q. Meisman, N. Krishnaswamy, and N. Blanchard, "Intentional Microgesture Recognition for Extended Human-Computer Interaction," in *Human-Computer Interaction*, ser. Lecture Notes in Computer Science, M. Kurosu and A. Hashizume, Eds. Cham: Springer Nature Switzerland, 2023, pp. 499–518. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-031-35596-7\\_32](https://link.springer.com/chapter/10.1007/978-3-031-35596-7_32)
- [32] E. Freeman, G. Griffiths, and S. A. Brewster, "Rhythmic Micro-Gestures: Discreet Interaction On-the-go," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ser. ICMI '17. New York, NY, USA: Association for Computing Machinery, Nov. 2017, pp. 115–119. [Online]. Available: <https://dl.acm.org/doi/10.1145/3136755.3136815>
- [33] M. Soliman, F. Mueller, L. Hegemann, J. S. Roo, C. Theobalt, and J. Steimle, "FingerInput: Capturing Expressive Single-Hand Thumb-to-Finger Microgestures," in *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ser. ISS '18. New York, NY, USA: Association for Computing Machinery, Nov. 2018, pp. 177–187. [Online]. Available: <https://dl.acm.org/doi/10.1145/3279778.3279799>
- [34] Y. Tan, S. H. Yoon, and K. Ramani, "BikeGesture: User Elicitation and Performance of Micro Hand Gesture as Input for Cycling," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '17. New York, NY, USA: Association for Computing Machinery, May 2017, pp. 2147–2154. [Online]. Available: <https://dl.acm.org/doi/10.1145/3027063.3053075>
- [35] R.-D. Vatavu, "iFAD Gestures: Understanding Users' Gesture Input Performance with Index-Finger Augmentation Devices," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–17. [Online]. Available: <https://dl.acm.org/doi/10.1145/3544548.3580928>
- [36] Z. Song, J. J. Dudley, and P. O. Kristensson, "HotGestures: Complementing Command Selection and Use with Delimiter-Free Gesture-Based Shortcuts in Virtual Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 11, pp. 4600–4610, Nov. 2023, iEEE Transactions on Visualization and Computer Graphics. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10269004>
- [37] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks," in *Proceedings of the 34th International Conference on Machine Learning*. PMLR, pp. 1321–1330, ISSN: 2640-3498. [Online]. Available: <https://proceedings.mlr.press/v70/guo17a.html>
- [38] Y. Tian, C. Sun, B. Poole, D. Krishnan, C. Schmid, and P. Isola, "What makes for good views for contrastive learning?" in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS '20. Red Hook, NY, USA: Curran Associates Inc., 2020. [Online]. Available: <https://dl.acm.org/doi/abs/10.5555/3495724.3496297>
- [39] X. Li, J.-D. Wang, J. J. Dudley, and P. O. Kristensson, "Swarm Manipulation in Virtual Reality," in *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*, ser. SUI '23. Association for Computing Machinery, pp. 1–11. [Online]. Available: <https://dl.acm.org/doi/10.1145/3607822.3614519>
- [40] D.-Y. Huang, L. Chan, S. Yang, F. Wang, R.-H. Liang, D.-N. Yang, Y.-P. Hung, and B.-Y. Chen, "DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. San Jose California USA: ACM, May 2016, pp. 1526–1537. [Online]. Available: <https://dl.acm.org/doi/10.1145/2858036.2858483>
- [41] C. Loclair, S. Gustafson, and P. Baudisch, "PinchWatch: A Wearable Device for One-Handed Microinteractions," 2010. [Online]. Available: <https://api.semanticscholar.org/CorpusID:112927411>
- [42] 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)*, Oct. 2022, pp. 864–871, iSSN: 1554-7868. [Online]. Available: <https://ieeexplore.ieee.org/document/9995649>
- [43] P. O. Kristensson and K. Vertanen, "The Inviscid Text Entry Rate and Its Application as a Grand Goal for Mobile Text Entry," in *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, ser. MobileHCI '14. Association for Computing Machinery, pp. 335–338. [Online]. Available: <https://dl.acm.org/doi/10.1145/2628363.2628405>
- [44] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)," publisher: UNIR. [Online]. Available: <https://hdl.handle.net/11441/107084>
- [45] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006, publisher: SAGE Publications Inc. [Online]. Available: <https://doi.org/10.1177/154193120605000909>
- [46] J. Brooke, "Sus: A 'Quick and Dirty' Usability Scale," 1996. [Online]. Available: <https://api.semanticscholar.org/CorpusID:107686571>
- [47] 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*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3173574.3174051>
- [48] N. Williams, "The Borg Rating of Perceived Exertion (RPE) scale," vol. 67, no. 5, pp. 404–405. [Online]. Available: <https://doi.org/10.1093/ocmed/kqx063>
- [49] J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio, "ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. Association for Computing Machinery, pp. 4021–4033. [Online]. Available: <https://dl.acm.org/doi/10.1145/3025453.3025683>
- [50] V. Biener, S. Kalamkar, N. Nouri, E. Ofek, M. Pahud, J. J. Dudley, J. Hu, P. O. Kristensson, M. Weerasinghe, K. C. Pucihar, M. Kljun, S. Streuber, and J. Grubert, "Quantifying the Effects of Working in VR for One Week," vol. 28, no. 11, pp. 3810–3820, IEEE Transactions on Visualization and Computer Graphics. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9872089>



**Xiang Li** is a PhD student in the Intelligent Interactive Systems group at the University of Cambridge and a Student Fellow at The Leverhulme Centre for the Future of Intelligence. He obtained his dual BSc degrees from the University of Liverpool and Xi'an Jiaotong-Liverpool University in 2022. Previously, he worked at Monash University, Carnegie Mellon University, the Institute Polytechnique de Paris (Télécom Paris), and the Hong Kong University of Science and Technology (Guangzhou).



**Wei He** is a PhD student in Urban Governance and Design at the Hong Kong University of Science and Technology (Guangzhou). Before this, he worked as a research assistant in the Department of Industrial and Systems Engineering at The Hong Kong Polytechnic University and at the Society Hub of the Hong Kong University of Science and Technology (Guangzhou) in 2023. He earned his BSc in Vehicle Engineering from Hunan University in 2022.



**Per Ola Kristensson** is a Professor of Interactive Systems Engineering in the Department of Engineering at the University of Cambridge and a Fellow of Trinity College. He is also a co-founder and co-director of the Centre for Human-Inspired Artificial Intelligence at the University of Cambridge. He is an Associate Editor of ACM Transactions on Computer-Human Interaction and ACM Transactions on Interactive Intelligent Systems and serves as a Steering Committee Member for ACM CHI.