PinchText: One-Handed Text Entry Technique Combining Pinch Gestures and Hand Positions for Head-Mounted Displays

Haiyan Jiang, Dongdong Weng, Xiaonuo Dongye & Yue Liu

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2022.2115333

PinchText: One-Handed Text Entry Technique Combining Pinch Gestures and Hand Positions for Head-Mounted Displays

Haiyan Jiang, Dongdong Weng, Xiaonuo Dongye, and Yue Liu

Beijing Engineering Research Center of Mixed Reality and Advanced Display, Beijing Institute of Technology, Beijing, China

ABSTRACT

This paper presents PinchText, a mid-air technique with a condensed keys-based keyboard, which combines hand positions and pinch gestures, enabling one-handed text entry for Head-mounted displays (HMDs). Firstly, we conduct Study 1 to collect and analyze the typing data of PinchText with two arm postures and two movement directions, obtaining the range of hand position corresponding to the middle key set. Then, we conduct Study 2, a 6-block experiment, finding that PinchText with Hand-Up Vertical (UpV) and Hand-Down Vertical (DownV) modes could achieve a speed of 12.71 words-per-minute (WPM) and 11.14 WPM respectively with both uncorrected error rates less than 0.5%, which is 71% faster than the index finger pinch-based technique. Finally, Study 3 is conducted to explore the potential of reducing the size of the decoupled visual keyboard of PinchText, verifying that the occlusion of the virtual keyboard can be decreased. Overall, PinchText is an efficient, easy-to-learn, and comfortable text entry technique for HMDs.

1. Introduction

Head-mounted displays (HMDs), platforms for virtual reality (VR) and augmented reality (AR) (Itoh et al., 2021) applications, are increasingly used among the public. As an essential interaction technique for HMDs, text entry is still underdeveloped. User mobility is increasing when using HMDs, requiring new entry methods beyond keyboards and touchscreens. A promising method is to appropriate the thumb-to-finger-based technique (Dash, 2017; Jiang et al., 2019; Prätorius et al., 2014; Wang et al., 2015; Whitmire et al., 2017; Wong et al., 2018). Except for being always available, this technique enables a comfortable and eyes-free text entry by leveraging proprioceptive feedback.

This paper presents PinchText as shown in Figure 1, a one-handed text entry technique that combines pinch gestures and hand positions, enabling word-level text entry in mid-air. A condensed keys-based virtual keyboard used in this technique is divided into three key sets that can be chosen according to the user's hand position. In comparison with the index finger pinch-based technique (*Oculus Quest*, 2020) that requires accurate control of the hand position and angle which could be affected by natural tremors of the hands (König et al., 2009), PinchText only requires hand moves between three large different spaces in the air, where hand jitter does not affect the selection of key sets, and in addition, pinch gestures are accurate, enabling accurate and quick text entry. Additionally, as the virtual keyboard is used as an interactive interface and visual feedback in HMDs (Yildirim & Osborne, 2020), it is usually located in the center of the field of view which may cause occlusion. As the virtual keyboard of PinchText and the hand position are decoupled, it is possible to reduce the size and change the location of the virtual keyboard, decreasing its occlusion, and users could input with the hand being naturally placed vertically, which helps to decrease fatigue.

The first study adopting the Wizard of Oz method is conducted to examine hand movement behaviors in four groups with two postures (Hand-Up and Hand-Down) and two movement directions (Horizontal and Vertical), yielding the range of hand positions corresponding to the middle key sets. Then, the second study is conducted to investigate the text entry performance of the UpV and DownV groups. The results show that after 60-phrase training, the entry speed of the UpV group reaches 12.71 WPM and that of the DownV group reaches 11.09 WPM, while the index finger pinch-based technique with a QWERTY keyboard reaches 7.43 WPM, all with less than 1% error rates. In addition, we find that the alignment of the virtual keyboard with input space (UpV) delivers better performance. A study on the design space for smartglasses graphical menus (Vatavu & Vanderdonckt, 2020) has found the location of the visual menus should be chosen to minimize interference with the visual scene, achieving thus good visual search and good utilization of the screen real estate, and users preferred the top, bottom and side location. In virtual environments (VEs), reducing the size of the virtual keyboard and changing its location instead of in the center of view would be

CONTACT Haiyan Jiang jianghybit@163.com; Dongdong Weng crgj@bit.edu.cn Beijing Engineering Research Center of Mixed Reality and Advanced Display, Beijing Institute of Technology, Beijing, China

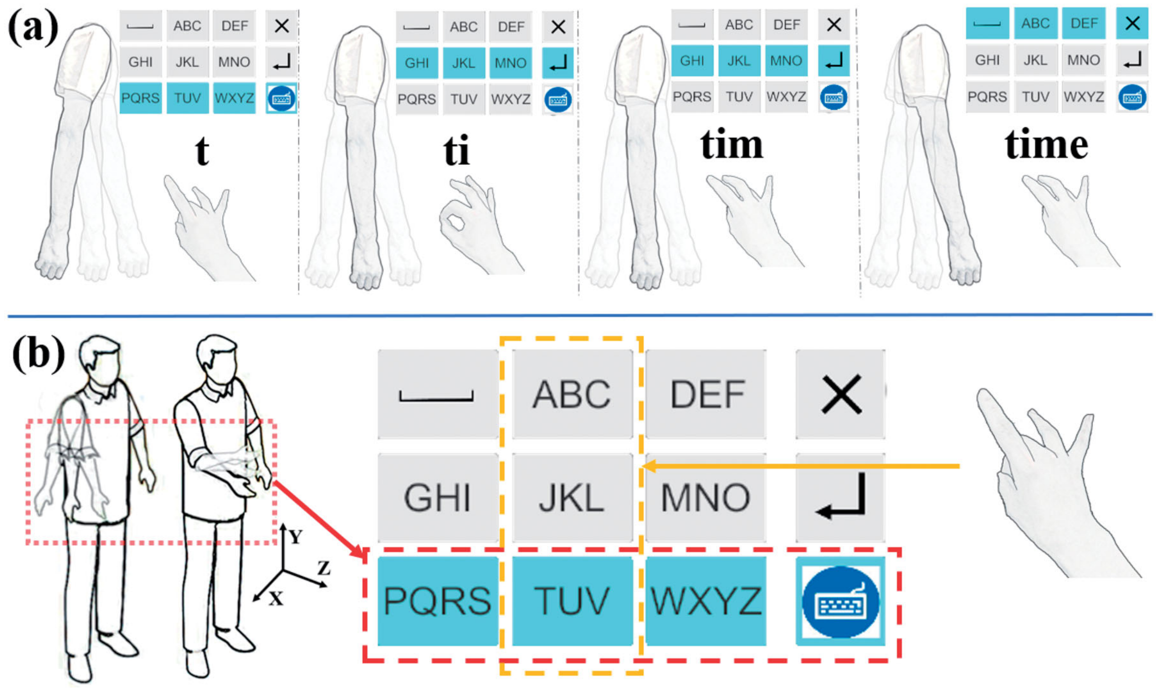

Figure 1. (a) Part of the entry process for "time" using PinchText with the Hand-Down posture. (b) Users are typing using PinchText with Hand-Up and Hand-Down arm postures. Users select the key set with hand movement, which background is blue; then, select one key within the key set with one of four different pinch gestures.

good for viewing of other parts of VEs. The third study therefore examines the possibility of reducing the size of the virtual keyboard which is placed in the lower-left corner of the user's field of view, and the results show that there is no significant difference in the typing performance with different sizes of virtual keyboards, exhibiting the potential of PinchText to decrease occlusion.

In sum, this paper makes contributions as follows: (1) Propose a text entry technique, PinchText, combining hand positions and pinch gestures with a condensed keys-based virtual keyboard according to a set of design considerations in mobile scenarios for HMDs. (2) Conduct three studies for the design and performance evaluation of Pinchtext, verifying the possibility of reducing the occlusion of PinchText in HMDs and giving some insights for the implementation of PinchText.

2. Related work

An overview of research related to our proposed text entry method is presented in this section, including text entry in HMDs, mid-air input, and text entry with multi-letter keys.

2.1. Text entry in HMDs

Text entry is an essential interaction technique for HMDs. Much work about text entry has been done, including physical device-based techniques with keyboards (Hutama et al., 2021; Jiang et al., 2018; Knierim et al., 2018; McGill et al., 2015; Walker et al., 2016; 2017), controllers (Boletsis & Kongsvik, 2019; Chen et al., 2019; Jiang & Weng, 2020; Yu et al., 2018), and touchscreens (Grubert et al., 2018; Gugenheimer et al., 2016; Kim & Kim, 2017; Lu et al., 2017); hands-free speech (Bowman et al., 2002; Pick et al., 2016); head-based (Majaranta et al., 2009; Yu et al., 2017) and gaze-based (Rajanna & Hansen, 2018) techniques; and hand-based techniques (gesture, micro-gesture, and hand trace mid-air input techniques). Aim-and-shoot techniques (*HTC vive*, 2020) and index finger pinch gesture-based techniques (*Oculus Quest*, 2020) are commonly used for text entry in commercial HMDs.

Speicher et al. (Speicher et al., 2018) compared six different text entry techniques: head pointing, controller pointing, controller tapping, freehand, discrete, and continuous cursor, showing that the controller pointing method had good performance in terms of entry speed (15.44 WPM) but it could cause fatigue in long-term use, and the freehand method could input at 9.77 WPM. As well as mid-air input methods, swipe gestures on hardware have been adopted to input text for AR. HoldBoard (Ahn et al., 2017) enabled users to input at 10.2 WPM through 4 h of training. SwipeZone (Grossman et al., 2015) enabled users to input with multistep touch swiping at 8.7 WPM. Moreover, more text entry input techniques for other wearables which could be used in HMDs could be found in reviews (Arif & Mazalek, 2016; Thomas et al., 1998)

Overall, the aim-and-shoot-based text entry technique is commonly used. However, the virtual keyboard occupies a large area, leading to occlusion. If the keyboard size was reduced, accurate pointing would become difficult. In addition, hand jitter appears when pressing the button, causing an obvious Heisenberg effect (Bowman et al., 2002; Wolf et al., 2020), which leads to poor performance. Additionally, too much hand movement may lead to hand fatigue (Speicher et al., 2018). Decreasing occlusion and hand fatigue and increasing input speed are our design goals.

2.2. Mid-air input

Users can input by tapping on virtual keyboards rendered in VEs like tapping on physical keyboards. Hsieh et al. (Hsieh et al., 2016) designed a haptic glove, which could recognize the tap gestures for letter selecting. TiTan (Yeo et al., 2017), Dudley et al.'s work (2018; 2019) enabled users to input in mid-air like tapping on physical keyboards.

Mid-air gestures support character-level or word-level input. For character-level methods, a gesture usually corresponds to a specific character (Sridhar et al., 2015), or the movement trace of the user's hand or fingers in the air directly corresponds to a specific character (Ni et al., 2011; Schick et al., 2012). The entry speed and error rate are affected by the accuracy of recognition algorithms and the speed of hand movement. Word-level methods usually combine the movement trace of the user's hand or fingers or the controller (Chen et al., 2019) with the virtual keyboard, such as in Vulture (Markussen et al., 2014), RotoSwype (Gupta et al., 2019), and HIBEY (Lee et al., 2019).

Thumb-to-finger micro gestures are usually used for mid-air input. Some methods map the QWERTY keyboard (Wang et al., 2015; Whitmire et al., 2017), T9 keyboard (Prätorius et al., 2014; Wong et al., 2018) or a keyboard arranged in alphabetical order (Dash, 2017) to different areas of the fingers of one hand or both hands, and text entry can be carried out by leveraging the action of thumb-to-finger touching. Some adopt new input methods, such as the two-step input method (Jiang et al., 2019) and the Huffman base-4 method (Bajer et al., 2012). However, the comfort level of the thumb touching different areas of the fingers varies (Huang et al., 2016; Jiang et al., 2019). In TipText (Xu et al., 2019) and BiTipText (Xu et al., 2020), an optimized keyboard was located on the first segment of the index fingers, where users used micro thumb-to-fingertip gestures for text entry.

Pinch gesture-based text entry techniques usually use data gloves or other sensors (Dementyev & Paradiso, 2014; Way & Paradiso, 2014) to detect the pinch gesture. Bowman et al. (Bowman et al., 2002) mapped the QWERTY keyboard to the fingertips, and users first rotated their hand to choose one line of the QWERTY keyboard and then selected one character using a pinch gesture, achieving a speed of 6.34 WPM. KITTY (Mehring et al., 2004) used a similar method, but it switched the keyboard lines by using the other fingertips to tap three contacts on different areas of the thumb. Argot (Peshock et al., 2014) was a wearable one-handed keyboard glove, condensing letters into eight keys like the T9 keyboard and distributing them and the other seven function keys to the fingertips, fingernails, and the sides of the other four fingers. Users touched these locations with their thumb for text entry. Users could use pinch gestures with both hands to select letter groups on a QWERTY keyboard with PinchType (Fashimpaur et al., 2020).

2.3. Text entry with multi-letter keys

Multiletter keys are usually adopted to address the problem of limited space (such as in mobile phones and smartwatches) (Frey et al., 2011; Goldstein et al., 1999; Green et al., 2004; Qin et al., 2018). UniWatch (Poirier & Belatar, 2016), containing a three-key keypad, could input at 9.84 WPM. The 1 Line Keyboard (Li et al., 2011) condensed the three rows of character keys in the QWERTY layout into a single line with eight keys for portable touchscreens. Komninos et al. (Komninos & Dunlop, 2014) proposed an ambiguous keyboard with six keys containing three to six letters each, and the entry speed reached 8 WPM. HiPad (Jiang & Weng, 2020) condensed 26 letters into 6 keys in a circular layout, achieving the speed of 13.57 WPM. James et al. (James & Reischel, 2001) studied text entry with the T9 keyboard on mobile phones, where users could type at 9 WPM. FingerT9 (Wong et al., 2018) was a technique for same-side-hand text entry on smartwatches combining a T9 layout and thumb-to-finger interaction, achieving an entry speed of 5.42 WPM.

2.4. Summary

In summary, mid-air text entry methods are suitable for the input in mobile scenarios, and many one-handed thumb-to-finger-based techniques are proposed for HMDs. As users could accurately touch their finger segments with their thumbs with eyes-free (Huang et al., 2016) and thumb-to-fingertip (pinch gesture) interactions have the highest comfort level in all thumb-to-finger interactions (Huang et al., 2016; Jiang et al., 2019; Lee et al., 2019), pinch gestures could support comfortable eyes-free text entry. However, the number of pinch gestures is limited, making it impossible to input text only using pinch gestures. Moreover, hand positions and condensed keys could expand the input space. Therefore, our goal is to explore a pinch gesture-based text entry technique with condensed keys, enabling word-level text entry by combining the hand positions.

3. Pinchtext design

3.1. Design considerations

3.1.1. Efficiency

Thumb-to-finger pinch gestures are often used for text entry. By taking advantage of human perception, people can complete pinch gestures precisely with eyes-free. In the index finger pinch-based techniques, users need to control the hand position and angle precisely to select one of 26 keys where the area of each key is relatively small, which may make the accurate and quick selection become difficult. PinchText only requires users to move their hands in three large spaces where relatively rough hand displacement enables accurate and quick selection of the key set. By combining pinch gestures and hand positions with a condensed keys-based virtual keyboard, users could input text accurately and quickly.

3.1.2. Comfort

When users input text in the air with common gesture-based text entry techniques, long-time input may cause

"Gorilla arm" (Hincapié-Ramos et al., 2014)—fatigue in the upper limbs. PinchText enables users input when their hands are placed naturally and vertically, benefiting the comfortable long-time use. In addition, people would feel more comfortable when touching fingertips than other areas with thumb-to-finger methods (Huang et al., 2016; Jiang et al., 2019; Lee et al., 2019).

3.1.3. Usability

Users would stand or walk in many VR/AR applications. For some common auxiliary devices for text entry, such as controllers, holding of devices may affect the interaction with other virtual objects and there are cases users cannot access the devices. The gesture-based interaction method is an alternate method. And, one-handed interaction may be suitable for where users' hands are not both available, for example, one hand is carrying a bag.

In the index finger pinch-based text entry method, the hand position and the key in the virtual keyboard are coupled, where the virtual keyboard would cause occlusion. For PinchText, the hand position and the virtual keyboard are decoupled, where the relative hand positions are used for the selection of key sets, while the virtual keyboard is used only for visual feedback and can be placed in any place of the field of view, as well as its size can be reduced, improving the usability.

3.1.4. Acceptability

In public, noticeable interaction methods would reduce users' willingness to use. The use of devices may be affected by perceptive social acceptability (Profita et al., 2013). Different factors including appearance, social status, and cultural conventions would affect that (Rico & Brewster, 2010). Interaction does not appear that the user is interacting with any technology (Toney et al., 2003) and using visuals to convey intent (Profita et al., 2013) benefit to minimize social consequences. Dobbelstein et al. (2015) found that when it came to longer interaction for up to 10 s, users would be more comfortable with a more natural interaction method in terms of perceived social acceptance in public. Serrano et al. (Serrano et al., 2014) compared the social acceptability of a hand-to-face input method for smart glasses. Participants showed concern in using the gestures in front of strangers or in public places.

PinchText could be used with hands being placed naturally and vertically, and the input space of Pinchtext is small, which two help to reduce obtrusiveness.

3.2. Pinchtext using postures

PinchText can be used with two different arm postures: *Hand-Up* and *Hand-Down* (Figure 1). In addition, according to the arm movement habits, we consider two hand movement directions: *Horizontal* and *Vertical*. So, there are four modes:

3.2.1 Hand-Up horizontal (UpH)

The user's hand is placed above the waist in front of the body. The user moves the hand left and right to select the corresponding key set, as shown in Figure 2(a).

3.2.2. Hand-Up vertical (UpV)

The user's hand is placed above the waist in front of the body. The user moves the hand up and down to select the corresponding key set, as shown in Figure 2(b).

3.3.3. Hand-Down horizontal (DownH)

The user's hand is placed naturally and vertically. The user moves the hand left and right to select the corresponding key set, as shown in Figure 2(c).

3.3.4. Hand-Down vertical (DownV)

The user's hand is placed naturally and vertically. The user moves the hand back and forth to select the corresponding key set, as shown in Figure 2(d).

3.3. Input mechanism

PinchText is a word-level text entry technique, and the workflow of that is shown in Figure 3.

3.3.1. Input space initialization

When the user touches the side area of the index finger (Side-key) over 2 s, the position of the hand at this moment is regarded as the initial position wherever the user's hand is placed. The user can initialize the hand position anytime if he/she wants to change the input space by the same action, even during the process of input.

3.3.2. Key selection

The position of the user's hand decides the key set users choose. The different pinch gestures are used to choose specific keys in the chosen key set (Figure 2).

In the UpH and DownH groups, when the user's hand is in the initial position, the middle key set is chosen, including the "GHI," "JKL," "MNO" and "Enter" keys. At this time, the thumb-to-index finger, thumb-to-middle finger, thumb-to-ring finger, and thumb-to-pinky pinch gestures are used to choose "GHI," "JKL," "MNO," and "Enter" keys respectively. When the user moves the hand upward to the position above the initial position in the UpH groups or forward to the position ahead of the initial position in the DownH groups, the upper key set is chosen and four different finger pinches are used to choose one key in the upper key set, including the "Space," "ABC," "DEF" and "Delete" keys; when the user moves the hand down to the position below the initial position in the UpH group or backward to the position behind the initial position in the DownH group, the lower key set is chosen and four different finger pinches are used to choose one key in the lower key set including the "PQRS," "TUPV," "WXYZ" and "Change" keys.

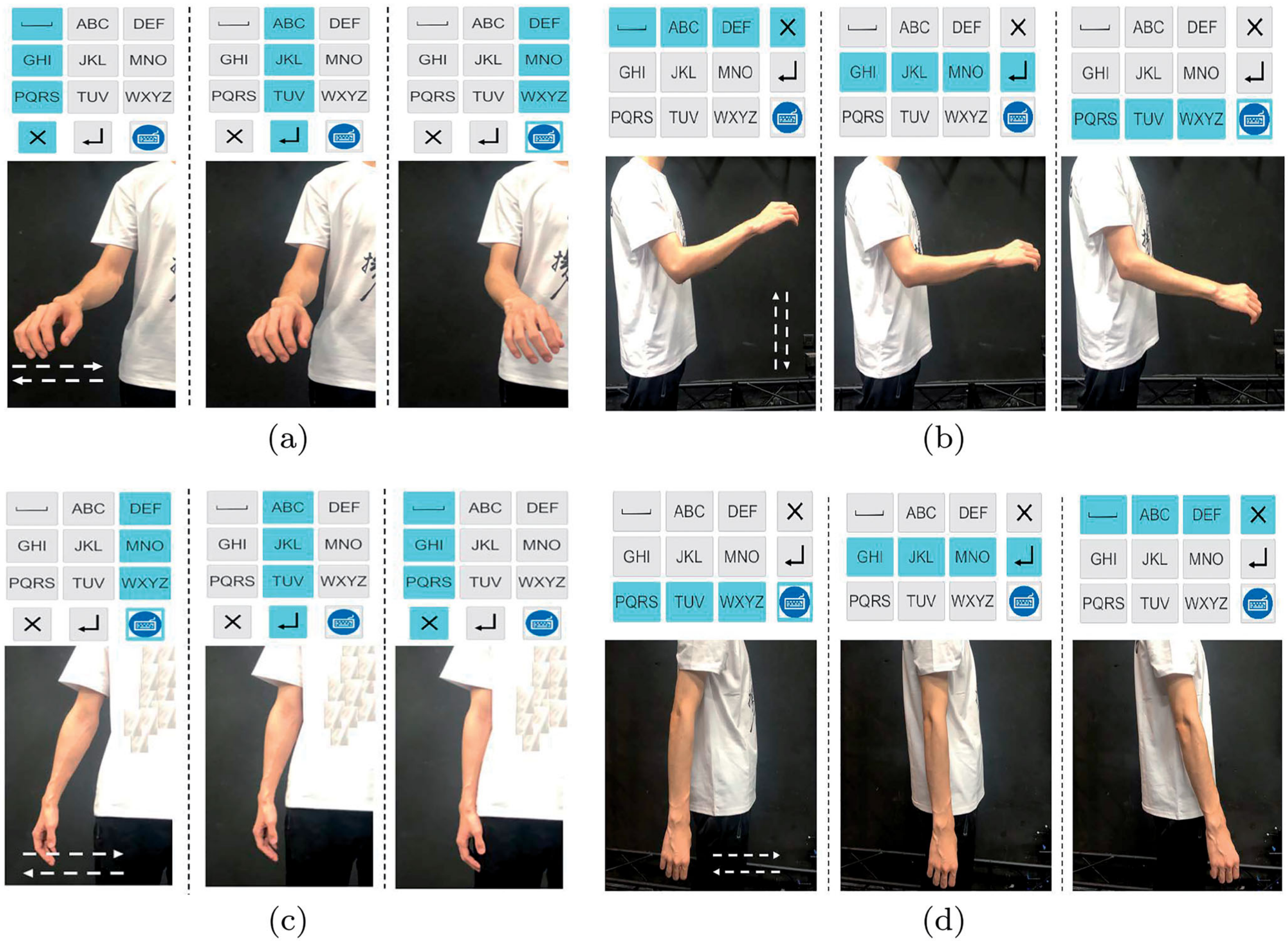

Figure 2. The white dotted arrows indicate the hand movement direction. When the hand moves in the air, the user can select different key sets (blue background). (a) Moving left and right in the UpH group; (b) Moving up and down in the UpV group; (c) Moving left and right in the DownH group; (d)Moving back and forth in the DownV group.

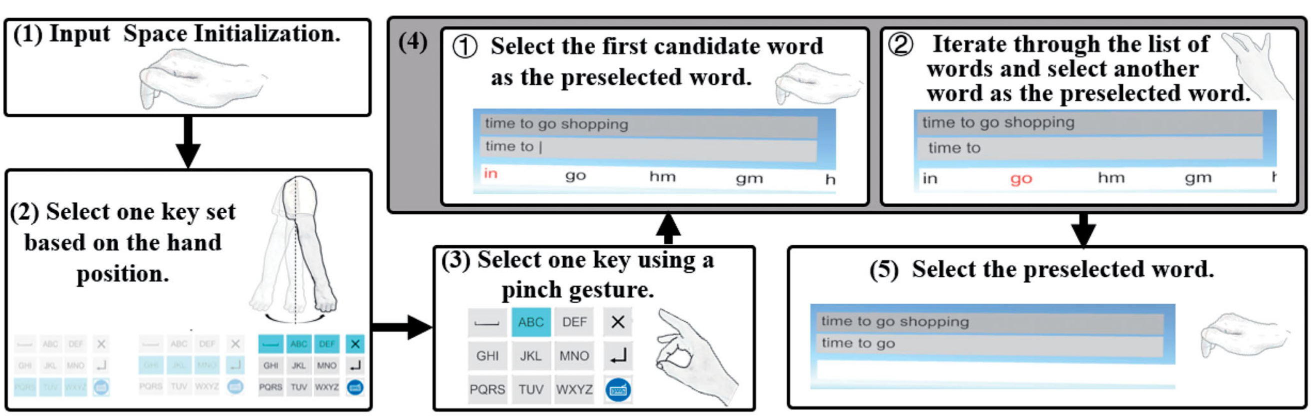

Figure 3. Workflow of Pinchtext. First, users initialize the input space by touching Side-key over 2 s and they can change the input space anytime by the same action if needed. Once the input space is determined, users can select one key set by moving their hand. After that, users select one key within the key sets by one of four pinch gestures. When a key sequence is chosen, users can touch Side-key to select the first candidate word as the preselected word. And users can iterate through the list to choose the next or last word as the preselected word. The color of the preselected word turns red as visual feedback. If users touch Side-key again, the preselected word is selected. The function of the Side-key changes according to touch time.

In the UpV and DownV groups, the selection mechanism is similar to UpH and DownH groups. When the user's hand is in the initial position, the middle key set is chosen, including the "ABC," "JKL," "TUPV" and "Enter" keys, which can be chosen by the thumb-to-index finger, thumb-to-middle finger, thumb-to-ring finger, and thumb-to-pinky pinch gestures. When the user moves the hand to the position to the left of the initial position, the left key set is chosen, including "Space," "GHI," "PQRS" and "Delete" keys. When the user moves the hand to the position to the

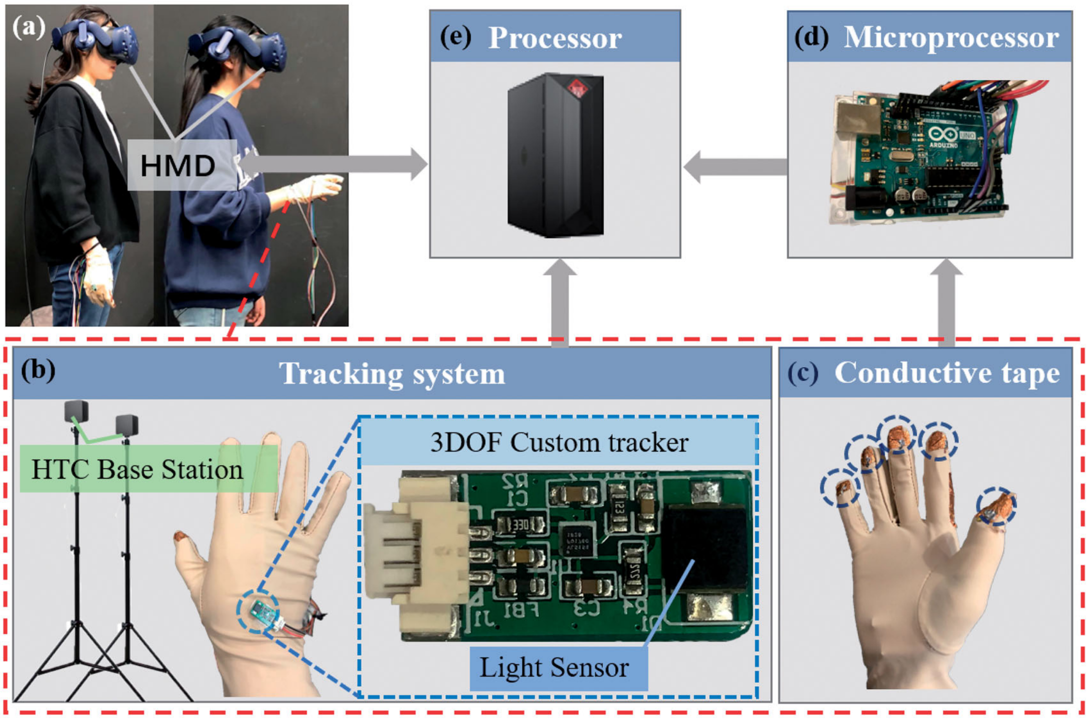

Figure 4. (a) Users are typing using PinchText with Hand-Down and Hand-Up postures. (b) The tracking system used in the prototype tracks the hand position which is transmitted to the processor (e). (c) The conductive tapes are attached to the glove, and when the tape on the thumb fingertip touches any other tape, a touch signal is produced and transmitted to the microprocessor (d) to get the corresponding pinch gesture signal. (e) A processor used to process all signals and data.

right of the initial position, the right key set is chosen, including the "DEL," "MNO," "WXYZ" and "Change" keys. One key is chosen by one of four pinch gestures.

The "Change" key is used to change the keyboard to other layouts, such as a numeric keypad layout or a symbol keyboard layout.

3.3.3. Word selection

When a key sequence is chosen, the disambiguation word list is displayed below the input box (Figure 3(4)), and at this time, the color of all words is black. After finishing a word sequence input, the user can touch Side-key to select the first candidate word as the preselected word. And users can iterate through the list to choose the next or last word as the preselected word. The color of the preselected word turns red as visual feedback. If the user touches Side-key again, the preselected word is selected. In other words, the user can directly touch Side-key twice to enter the first candidate word most of the time. After one word is input, a "space" is added automatically. However, if the word contains only one character, a "space" is not added. Thus, out-of-vocabulary (OOV) words can be input.

3.3.3.1. Next word. The next word can be chosen as the preselected word by the thumb-to-index finger pinch gesture. *Last word*: The last word can be chosen as the preselected word by the thumb-to-ring finger pinch gesture. *Delete*: If there is a preselected word, choosing the "Delete" key makes the user can continue to input the key sequence. If there is no preselected word but a key sequence, choosing the "Delete" deletes the last key of the key sequence. If there is no key sequence, the last word entered is deleted when choosing "Delete".

Since each key is associated with multiple letters, it is necessary to adopt word disambiguation for successful typing. We adopt the disambiguation algorithm in (Jiang & Weng, 2020).

3.4. Hardware

The prototype is implemented by using conductive tape made of copper placed on the five fingertips (Figure 4(c)) and one on the side of the index finger. When the thumb touches any other fingertip or the side of the index finger, a touch signal is produced and transmitted to the microprocessor. Then, through processing by the microprocessor, the corresponding pinch gesture signal is produced and transmitted to the processor.

A tracking system is used to track the position of the opisthenar as the hand position (Figure 4(b)). The tracking system of the prototype is from our lab, with 1 mm accuracy, the working principle of which is similar to HTC VIVE Lighthouse[1]. Other peers could use OptiTrack[2] or other commercial products to track the position of the hand.

4. Study 1: Investigating hand movement behaviors

In four modes, the optimal range for the user's hand movement is unknown. So, we conducted Study 1 with the Wizard of OZ method to investigate the user's hand

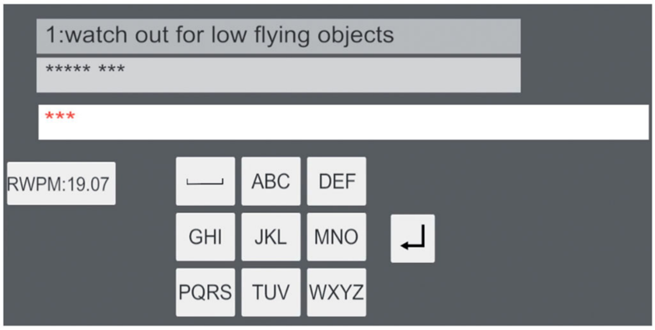

Figure 5. The input feedback interface in Study 1. RWPM stands for the real-time input speed.

movement range of selecting different key sets under four modes. Specifically, participants were told to move their hand position according to the rules in each group to choose different key sets but did not restrict the range of hand movement. Only asterisk feedback was provided during typing, which was similar to the previous work (Azenkot & Zhai, 2012; Findlater et al., 2011), and there was no row or column "highlighting" based on the hand movement, as shown in Figure 5. There was only one candidate word. Therefore, it could simulate the typing effect of the user's ideal hand movement range to obtain the best values. This study also focused on whether different postures and hand movement directions would affect the user's input performance, fatigue, and preferences.

4.1. Participants

16 right-handed participants (8 females) aged 21–32 ($M = 24.43$) in our college were recruited to participate in this study. All participants input using their right hand in this study.

4.2. Apparatus and materials

The experiment was conducted on a computer with the Intel Core i7 processor and the NVIDIA GTX 1080Ti graphics card, and the software was implemented with C# in Unity 2019.4.10f1. The experiment software was displayed on an external 23" LCD monitor (DELA0FD DELL SP2318H) set to 1920 × 1080 resolution.

4.3. Experimental design and procedure

The whole experiment lasted approximately 50 min for each participant. Before the experiment, participants were asked to fill in a pre-experiment questionnaire to gather their demographic information. The transcript method (Gupta et al., 2019) was adopted to evaluate the effect of different layouts. The transcript phrases were randomly generated from the MacKenzie phrase set (MacKenzie & Soukoreff, 2003).

The study used a 2 × 2 within-subject design with two factors, constituting four groups as mentioned in 3.2: UpH, UpV, DownH, and DownV. Participants completed four sessions of text entry, each corresponding to one group. The order of groups was fully counterbalanced with a Latin square design. Participants completed the experiment standing in front of the monitor. Before each session, participants were required to complete a training session consisting of 3 phrases. They could rest between any two sessions. After each session, participants were asked to fill in a post-experiment questionnaire to gather subjective feedback. For the whole experiment, 2 (posture) × 2 (direction) × 9 (phrase) × 16 (participant) = 576 timed trials were collected.

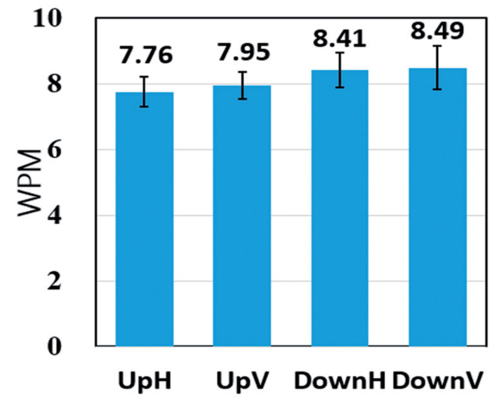

Figure 6. Average entry speed in Study 1.

4.4. Results

A GLM Repeated Measures ANOVA with Greenhouse-Geisser correction or Huynh-Feldt if necessary, ANOVA and the paired-T test were used to analyze the entry speed and error rate in Study 1, 2, and 3. A Wilcoxon signed-rank test is used to analyze the subjective feedback data in Study 1.

4.4.1. Entry speed

Entry speed was measured in the standard words-per-minute (WPM) metric following Mackenzie (MacKenzie, 2002). As shown in Figure 6, the average entry speed of four groups: DownV > DownH > UpV > UpH. However, there was no significant difference between each two groups.

4.4.2. Subjective results

Users' subjective ratings were collected using a 5-point Likert-scale questionnaire including perceived entry speed, perceived accuracy, hand fatigue, and overall preference, where all responses were assigned a score between 1 and 5, with 5 being the most positive and 1 being the most negative response. The results are shown in Figure 7.

4.4.2.1. Perceived speed. The score of the UpV group was significantly higher than the UpH group ($Z = -2.077$, $p = 0.038$), and the score of the DownH group was significantly higher than the UpH group ($Z = -2.124$, $p = 0.034$), while there was no significant difference between DownH and DownV groups ($Z = -0.676$, $p = 0.499$) and between UpV and DownV groups ($Z = -0.051$, $p = 0.959$).

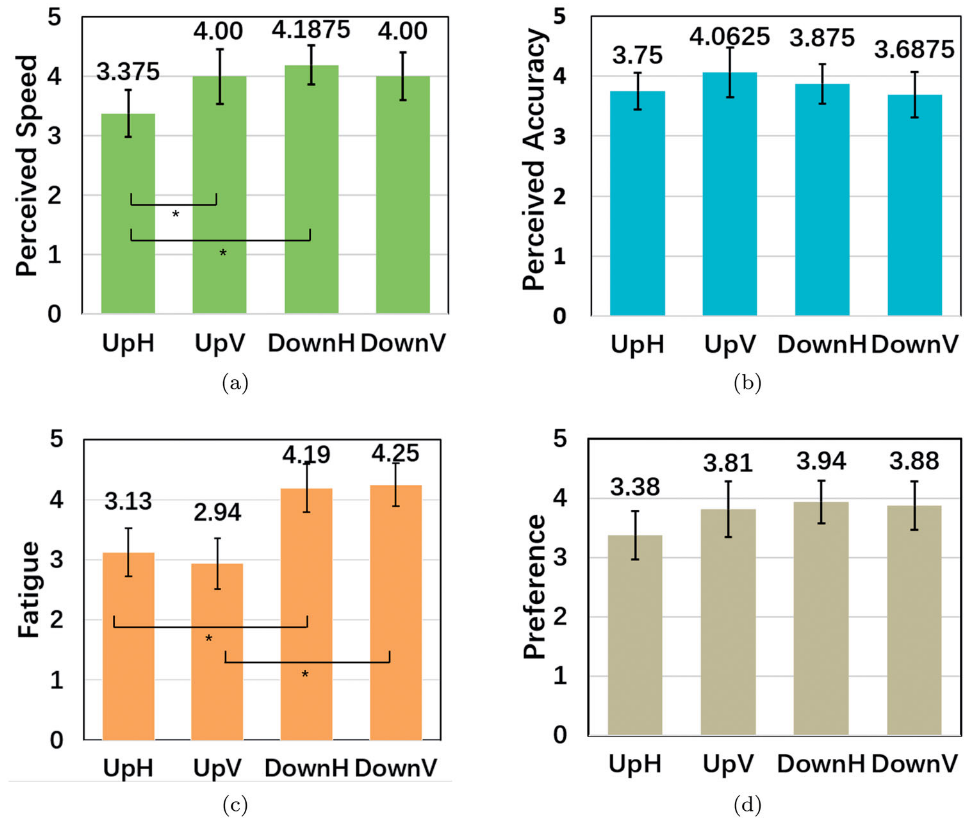

Figure 7. (a) Average perceived entry speed; (b) Average perceived accuracy; (c) Average fatigue; (d) Average preference of four groups. ($* : p < 0.05$).

Table 1. The average distances between the middle set and the other two sets.

	Posture(unit:mm)			
Euclidean distance to Middle set	UpV	UpH	DownV	DownH
Left (Top or Front) set	$M = 94.07$ $SD = 11.81$	$M = 102.83$ $SD = 17.28$	$M = 159.79$ $SD = 35.27$	$M = 88.85$ $SD = 11.24$
Right (Bottom or Back) set	$M = 103.39$ $SD = 17.77$	$M = 105.40$ $SD = 17.41$	$M = 176.76$ $SD = 36.19$	$M = 98.25$ $SD = 17.55$

4.4.2.2. Fatigue. For the hand position, the scores of two Hand-Down groups were significantly higher than the scores of two Hand-Up groups ($Z = -3.172$, $p = 0.002$). The score of DownV group was significantly higher than that of UpV group ($Z = -2.504$, $p = 0.012$) and the score of the DownH group was significantly higher than that of the UpH group ($Z = -2.014$, $p = 0.044$).

4.4.2.3. Perceived accuracy and preference. No significant difference was found.

4.4.3. Hand movement range

The spread of size for touch points of each group was analyzed, helping to set the hand movement range of each key set in four modes. Table 1 shows the average distances between the center of the middle set and the center of the other two sets.

4.5. Results discussion

The results showed that four groups had similar text entry potential. In terms of fatigue, the fatigue level of Hand-Down groups was significantly lower than that of Hand-Up groups. But users mentioned that the input space and display interface in Hand-Up was more consistent than in Hand-Down. The results did not show much difference in subjective perception speed, accuracy, and preference.

The user's hand movement range was similar in the UpV, UpH, and DownH groups, where the range from the middle key set to the other two sets was about 100 mm, while that of the DownV group was larger, which might be related to the hand movement habits and the inertia would produce a relatively large moving distance.

Both Hand-Up and Hand-Down postures should be supported since they were suitable for different situations. Overall, in terms of perceived speed, the score of the UpV group was higher than that of the UpH group, while there was no significant difference in other respects. In addition, the back and forth movement of the arm was more in line with the usual hand swing movement in the DownV group and there was no difference in performance between the DownV and DownH groups. So, we further investigated the typing performance of PinchText with UpV and DownV modes.

5. Study 2: Evaluating the typing performance of PinchText

The purpose of this study was to explore the text entry performance and the learning curve of PinchText. In this study, the UpV and DownV groups were chosen to explore. Participants input text using the mechanism described in Section 3. Besides, a common pinch-based technique with a QWERTY virtual keyboard, where input 26 letters directly combining hand position with the thumb-to-index finger pinch gesture, like the input method of Oculus Quest (*Oculus Quest*, 2020) was chosen for comparison.

5.1. Participants

24 new right-handed participants (17 males) aged 21–31 ($M = 23.48$) were recruited to participate in this study. They all have no text entry experience in HMDs. Participants typed using the right hand in the VE with HTC VIVE Pro.

5.2. Apparatus and materials

The experiment was conducted on a computer with the Intel Core i7 processor and the NVIDIA GTX 1080Ti graphics card, and the software was implemented with C# in Unity 2019.4.10f1. Our virtual environments were displayed in the HTV VIVE PRO which allowed participants to be completely immersed in the virtual world.

5.3. Experimental design and procedure

In the UpV and DownV groups, the virtual keyboard was placed 5 m away in the center of the field of view, and the virtual keyboard occupied a field of view of 55° (horizontal)*32° (vertical). According to the results of Study 1, the range of the middle key was set at 100 mm and the ranges of the other two sets were infinite, so the range of hand movement was not a fixed value for different users. In the QWERTY group, it occupied a field of view of 72° (horizontal)*32° (vertical). In real environments, the hand control area corresponding to each key of the QWERTY keyboard was 3 cm*4cm, and participants needed to move their hands in a 30 cm*12 cm 2 D plane. When the hand space initialization was completed, the hand position at this moment was corresponding to the key "G". The user's hand position was used to select one key, and the index finger pinch gesture was used to confirm the selection.

The experiments followed a between-subject design with eight participants each in the UpV, DownV, and QWERTY groups. Each participant performed two blocks continuously each day in three consecutive days. Each block included 1 training phrase and 10 test phrases. They could rest between two blocks. The whole experiment each day lasted approximately 30 min for each participant. After one day experiment, participants filled a post-experiment questionnaire to gather their demographic information and subjective feedback. The NASA-TLX (Hart, 2006) on a 7-point Likert scale was used in this study. For the whole experiment, 3 (day) × 2 (block) ×10 (phrase) × 24 (participant) = 1440 trials were collected.

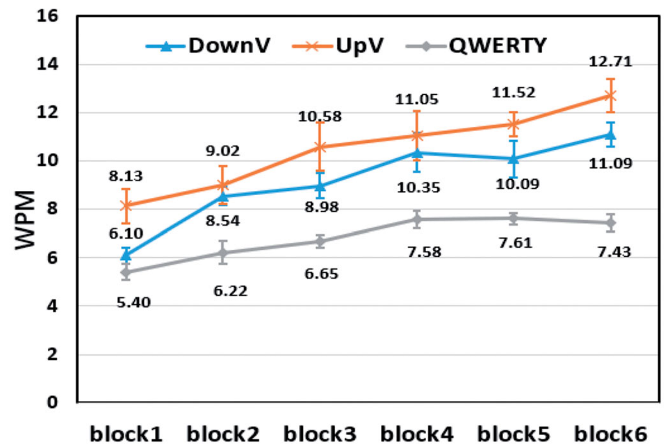

Figure 8. The average entry speed of six blocks of the three groups. Error bars indicate ±1 standard errors.

5.4. Results

5.4.1. Entry speed

Text entry speed included the time spent to correct errors. Figure 8 shows the average entry speed of six blocks of the three groups. For the entry speed of the groups: UpV > DownV > QWERTY. Significant difference was found among three groups ($F_{2,21} = 18.712$, $p < 0.000$). The entry speed of six blocks of each group significantly increased (DownV: $F_{1,7} = 41.654$, $p < 0.000$; UpV: $F_{1,7} = 53.235$, $p < 0.000$; QWERTY: $F_{1,7} = 41.448$, $p < 0.000$).

5.4.2. Error rate

The Uncorrected Word Error Rate (UWER), Corrected Key Error Rate (CKER), and Corrected Word Error Rate (CWER) were measured as in the prior word-gesture typing work (Zhu et al., 2018). We calculated the UWER as follows: UWER = (MWD(S, P)× 100)/Len(P), where MWD is the minimum word distance between the transcribed phrase S and target phrase P, and L is the number of words in P. Since participants performed word-level corrections by deleting a word and retyping, CWER is defined as CWER = (WD × 100)/Len(P), where WD is the number of word deletions performed in the phrase. And users could delete the mistaken key when typing the word's key sequence, CKER is defined as CKER= (KD × 100)/Len(K), where KD is the number of key deletions performed in the key sequence of a word. Error rates based on word distances are generally higher than string distance error rates since an error caused due to erring on a single character would lead to the whole word incorrect.

For the UpV and DownV groups, we calculated the UWER, CWER, and CKER, while for the QWERTY group, we only calculated the UWER, CKER.

Figure 9 shows the results. For three types of error rate, no significant difference was found among groups (UWER: $F_{2,21} = 1.487$, $p = 0.249$; CKER: $F_{2,21} = 0.0413$, $p = 0.667$;

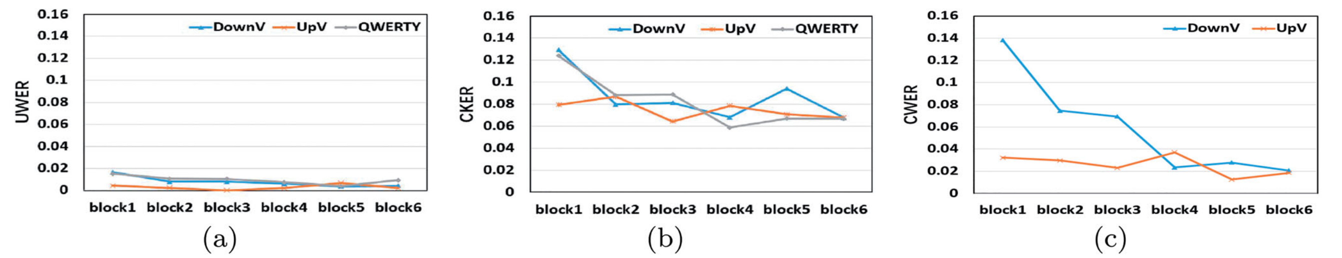

Figure 9. (a) Average uncorrected word error rate (UWER); (b) Average corrected key error rate (CKER) of three groups; (c) Average corrected word error rate (CWER) of the UpV and DownV groups.

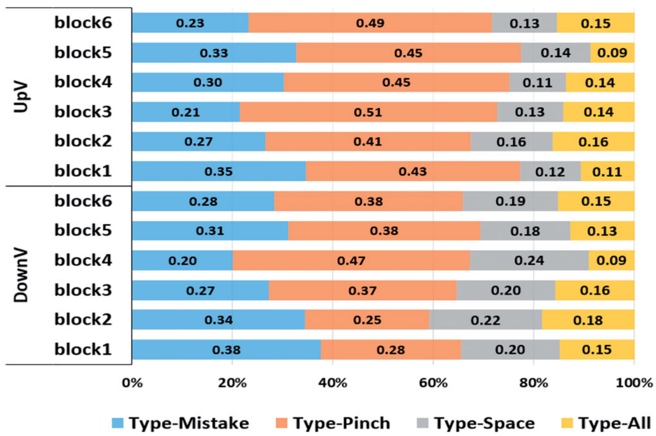

Figure 10. The average error type ratios of each type to all types.

CWER: $F_{2,21} = 0.0413$, $p = 0.667$). For CWER, in the UpV group, no significant difference was found among six blocks ($F_{5,35} = 0.821$, $p = 0.543$) and there was no significant linear trend ($F_{1,7} = 4.027$, $p = 0.085$), while in the DownV group, significant difference was found ($F_{5,35} = 5.621$, $p = 0.01$), and there was a significant linear trend ($F_{1,7} = 16.970$, $p = 0.004$).

5.4.2.1. Error analysis. The error types when participants input key sequences were analyzed. There were four error types: Type-Mistake, where participants input the right key but deleted it; Type-Pinch, where participants chose the right key set, which means they moved their hand to the right position, but they pinched with wrong fingers; Type-Space, where participants chose the wrong key set, which means they did not move their hand to the right position, although they pinched with the right finger; Type-All, where they both chose the key set and pinched with a wrong finger.

Figure 10 shows the average ratio of each type to all types. The error ratio of Type-Pinch was the biggest and followed by that of Type-mistake, while the ratio of Type-Space and Type-All were relatively small. In the future, more attention should be paid to reducing the Type-Pinch mistake, e.g., taking an auto-correction algorithm.

5.4.3. Typing behavior

Figure 1(b) shows a coordinate system in space relative to a person. Figure 11 shows the pinch points in the air of eight keys. For the UpV group, users' hands generally moved along Y-axis, while users' hands barely moved along the X-axis. When users selected the key in the top key set, their hands would move closer to their bodies. For the DownV group, users' hands generally moved along Z-axis, while users' hands barely moved along the X-axis. There is a certain amount of movement along the Y-axis when their hands moved back and forth. The keys in the same key set had a similar pinch point position in the air and that did not seem to be affected by the virtual keyboard where the keys in the same key set had different positions.

We recorded the elapsed time between two key set selections from the start key set to the end key set, as shown in Figure 12. This is in line with our expectations that the elapsed time between the same key set was shorter than that between two adjacent key sets which was shorter than that between the two farthest key sets.

5.4.4. Subjective feedback

Figure 13 shows the average scores of the NASA-TLX. Between each two groups, no difference was yielded in all subscales of each day by the Post hoc Tukey HSD Tests. For the UpV group, the mental demand significantly decreased over three days ($F_{2,14} = 4.20$, $p = 0.037$), while there was no significant in other aspects. For the DownV group, the physical demand ($F_{2,14} = 6.176$, $p = 0.012$) and the effort ($F_{2,14} = 4.586$, $p = 0.029$) significantly decreased, and the performance significantly improved ($F_{2,14} = 3.807$, $p = 0.048$), while no significant difference was found in other aspects. For the QWERTY group, the physical demand ($F_{1.191,8.336} = 7.843$, $p = 0.019$) and the effort ($F_{2,14} = 4.303$, $p = 0.035$) significantly decreased, while no significant difference was found in other aspects.

In word-level text entry, when the users selected the wrong key, they tended to delete the whole sequence. The reason may be that as some participants mentioned "According to the current words, I am not sure whether the input sequence is correct," or "I am accustomed to deleting all sequences and starting over."

6. Study 3: The potential of reducing the size of the virtual keyboards

The virtual keyboard may cause occlusion problems in VEs. For PinchText, the decoupled virtual keyboard as visual feedback could be placed anywhere in the field of view and its size could be reduced to decrease the occlusion.

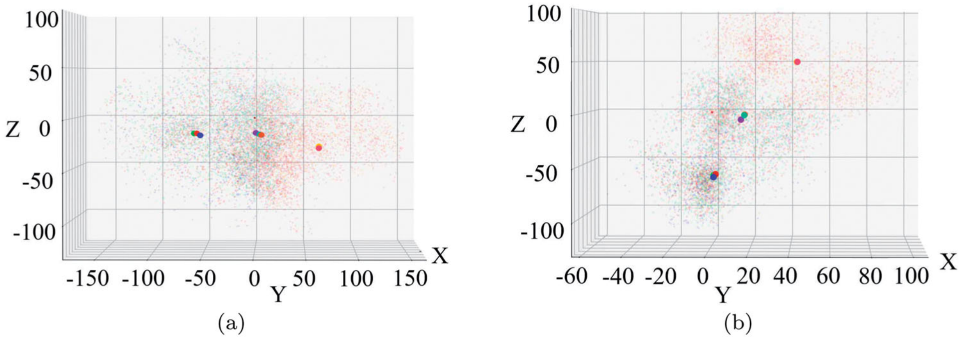

Figure 11. Pinch points corresponding to eight keys from all participants of (a) the UpV group and (b) the DownV group (unit: millimeter). Different colors represent points of different keys. Eight opaque points represent the mean of all points corresponding to each key.

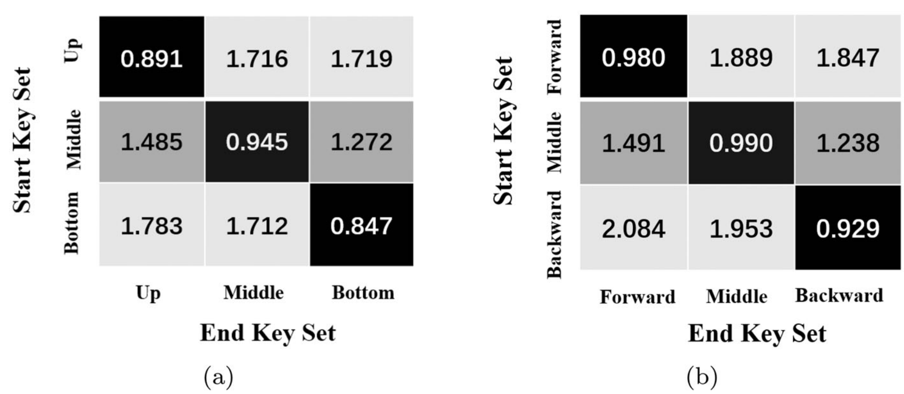

Figure 12. Elapsed time (s) between two key set selections from the start key set to the end key set in (a) UpV and (b) DownV groups.

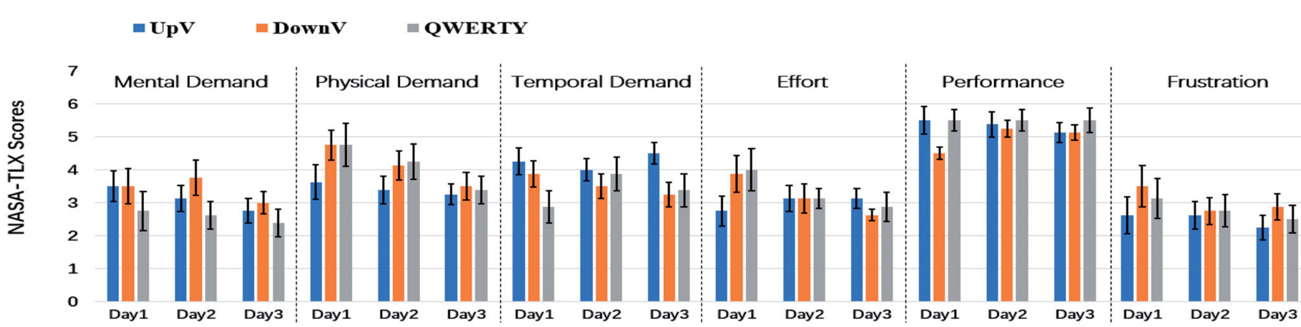

Figure 13. Average scores of the NASA-TLX. Error bars indicate ±1 standard errors.

Therefore, this study explored the effect on the performance of the size of virtual keyboards. In this study, we only investigated the DownV mode since all the effects of virtual keyboards in all modes are similar.

6.1. Participants and apparatus

12 right-handed participants (9 males) aged from 21 to 27 ($M = 24.52$) in study 2 participated in this study. All participants typed using their right hand.

6.2. Apparatus and materials

The experiment was conducted on a computer with the Intel Core i7 processor and the NVIDIA GTX 1080Ti graphics card, and the software was implemented with C# in Unity 2019.4.10f1. Our virtual environments were displayed in the HTV VIVE PRO. Figure 14 shows a VE used in this study.

6.3. Experiment design and procedure

In order to investigate the impact of the position and size of the virtual keyboard on user performance, the virtual keyboard was always placed in the lower-left corner of the user's field of view where the center of the field of view could not be blocked, including three size keyboards: *Large*, *Medium*, and *Small*.

Large: occupied the field of view about 17.79°*10.93°.

Medium: 0.7 times the size of the Large virtual keyboard.

Small: 0.5 times the size of the Large virtual keyboard.

The virtual keyboards were placed at 3 m away, 0.6 m below, and 0.4 m left of the center of the field of view.

Figure 14. The virtual environment used in the study 3.

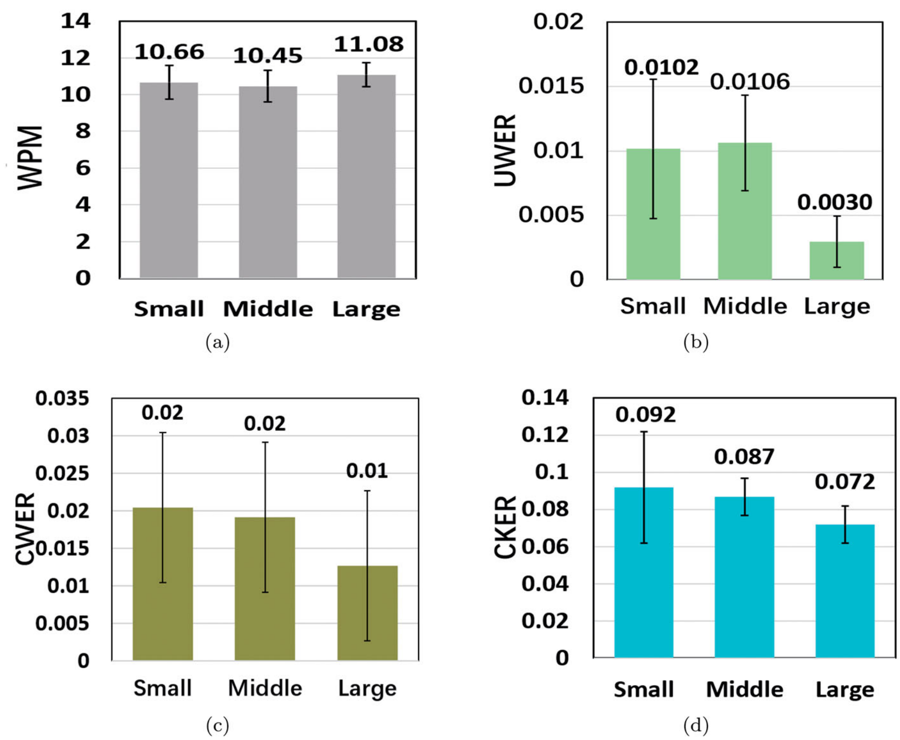

Figure 15. (a) Average entry speed; (b) Average uncorrected word error rate (UWER); (C) Average corrected word error rate (CWER); (d) Average corrected key error rate (CKER). Error bars indicate ±1 standard errors.

The experiments followed a within-subject design with one factor. Each participant performed three modes consecutively at three sessions, each session consisting of three training phrases and 12 test phrases. The order of groups was fully counterbalanced with a Latin square design. They could rest between any two sessions. The whole experiment lasted approximately 30 min for each participant. For the whole experiment, 3 (mode) × 12 (phrase) × 12 (participant) = 432 trials were collected.

6.4. Results

6.4.1. Entry speed

Figure 15(a) shows the average entry speed of three modes. No significant difference was found between any two modes.

6.4.2. Error rate

Figure 15(b–d) show the average UWER, CWER, and CKER of three modes respectively. No significant difference was found between any two modes.

6.4.3. Subjective feedback

In this experiment, the placement of the virtual keyboard was fixed relative to the screen. The user mentioned that when the virtual keyboard is too small, it was difficult to read the letters on the key at first, but they would get used to it after a while. As the entry speed and error rate were not affected, it also verified that the virtual keyboard could only be used as visual feedback.

Some users mentioned that placing it in the lower-left corner would cause discomfort. It was recommended to place it directly below the center. And the virtual keyboard can be placed relative to objects of interest, such as the chat box in the application software. That can be improved in the future.

7. Discussion

In this paper, a new text entry technique combining hand position in the air with pinches was proposed. Firstly, the hand movement behaviors of four different possible modes with this technique were investigated. Then, its potential

Table 2. Performance of PinchText and other existing techniques before and after learning.

Technique	Novice WPM	Expert WPM	Novice ER%	Expert ER%
PinchText (UpV)	8.13	12.71	0.4	0.2
PinchText (DownV)	6.10	11.09	1.6	0.4
PalmType* (Wang et al., 2015)	8.62	10.50	NA	NA
FingerT9 (Wong et al., 2018)	3.72	5.42	10.6	4.65
Quadmetric keyboard (L. H. Lee et al., 2019)	3.53	6.47	14.88	5.46
FingerText (D. Lee et al., 2021)	19.80	31.3	9.0	3.8
Hand Gestural Technique* (Hsieh et al., 2016)	4.34	5.42	6.63	4.98

*Character-level entry techniques, not word-level entry.

entry performance and the possibility of decreasing the occlusion of virtual keyboards were studied. There are some insights from those results.

7.1. Effectiveness of combining hand positions with pinches

Pinches between the thumb and other fingertips are more comfortable than other thumb-to-finger gestures and can work accurately due to the self-perception of human beings. As the number of pinch gestures is limited, however, it is impossible to input 26 and more characters directly.

By combining the hand positions and pinch gestures, word-level text can be input. First, the entry speed of the UpV group could reach 12.71 WPM after a 60-phrase training and keeps the rising trend. PinchText with UpV is 71% faster than the common pinch-based method with a QWERTY keyboard. Table 2 shows the entry speed (WPM) and error rate (ER) comparison with existing text entry techniques based on hands. The novice metrics indicate the performance of the corresponding technique in the first session of evaluation, while expert metrics indicate the best performance of this technique after learning. The table only includes a meta-comparison across different techniques instead of a direct rigorous comparison. However, it helps to contextualize the performance of PinchText by giving standard metrics of WPM and ER. PalmType (Wang et al., 2015) and the hand gestural technique proposed by Hsieh (Hsieh et al., 2016) are character-level techniques, while other techniques are word-level techniques. PalmType (Wang et al., 2015) is evaluated by repeating four words six times and FingerText (Lee et al., 2021) repeating one word seven times to simulate expert performance, while the technique proposed by Hsieh (Hsieh et al., 2016) conducts a 12-phrase typing study and other techniques conducted evaluations with at least 50 phrases typing. FingerText (Lee et al., 2021) enables text entry using intra-hand touches between the thumb and fingers, and the entry speed is more quickly than that of other techniques. However, the system of FingerText uses a set of five fingernails mounted capacitive sensors to track intra-hand touches, which may not be possible or preferred in all scenarios. Through learning, the performances of all techniques are improved in terms of WPM and ER. Compared to those techniques, PinchText has good novice and expert performances. UCER (ER in table) of the UpV group and DownV group are both less than 0.5%. Participants are not English native speakers, which may lead to a decrease in performance for all groups, like the results of (Yu et al., 2018) that native users have better performance of entry speed than non-native users. In addition, in this experiment, users correct errors manually without any correction algorithms in all groups, and they are not familiar with the input methods of VR, which may decrease the performance. It could still demonstrate that the performance of PinchText was better. More algorithms including auto-correction algorithms and dynamically thresholds control algorithm of hand position will help to improve the performance of PinchText.

In study 3, the virtual keyboard is placed in the lower-left corner of the user's field of view without occupying the center of the field of view and the size of it can be adjusted. The results show that the size of the virtual keyboard did not affect the user's input performance. The virtual keyboard and hand position are decoupled, and the virtual keyboard is mainly used as feedback. The user's observation of the virtual keyboard would reduce when users become more familiar with this technique. In addition, users can use peripheral vision to perceive visual feedback. The virtual keyboard can be placed elsewhere in the field of view instead of the center of the view.

7.2. Workload of PinchText

Combining hand position with pinch gesture may cause cognitive load, but the cognitive load would decrease with the increase of the user's familiarity with the input method. In the QWERTY group, users were only required to match the hand position with the key and then pinched to select a letter, so the cognitive load was relatively low. Using different pinch gestures may cause cognitive load and lead to more errors, so the error ratio of Type-Pinch was big. After 3-day training, the mental demand of the three groups tended to be the same. The unfamiliarity may cause high mental demand. But as the time of use increased, the mental demand would decrease. As some participants mentioned," At first I spent a lot of time thinking about which pinch I should do, but gradually I get used to pinching right without thinking about." The work (Gupta et al., 2020) found that on-fingers feedback helps to decrease mental demand, frustration, and effort. For PinchText, the pinch gestures provide feedback when users choose a key, which would be helpful to decrease the mental demand, frustration, and effort.

At the beginning of use, the physical demand of the DownV group was higher than that of the UpV group. But with the increase in familiarity, the physical demand of the DownV group decreased, which was no different from the UpV group. Whether it would continue to decline requires more research. The effort of the DownV group on the last day was already lower than the UpV and QWERTY groups. It could be predicted that users in the DownV group will be less fatigued than the UpV group when typing for a long time.

7.3. Usability

Our study demonstrated that the methods of DownV and DownH have good performance. Compared with the common pinch-based method with a QWERTY keyboard, it seems to be less obtrusive, as mentioned that "the methods of DownV or DownH groups are not so obtrusive." Meanwhile, the user's hand position and the keyboard are decoupled, enabling users to input in mobile scenarios. The pinch gestures and the forward and backward movement of the user's hand which is relative to their body are less affected by the user's position in the world. The one-handed text entry technique enables users to input with only one hand, which is convenient to use in the wild, as mentioned that "I may need to carry a handbag with one hand when I'm outside." Meanwhile, PinchText opens up chances to explore two-hand interaction with two tasks and two-hand cooperation, where one hand performs text entry.

7.4. Limits and future work

This work proposed a one-handed text entry technique combining the hand positions with pinch gestures. Using autocorrection algorithms will help improve input efficiency. In addition, there are many aspects that can be improved.

First, in the experiments, as the hands were not visualized, some users expressed their willingness to see their hands, which may affect the text entry and needs further study. Second, in the experiments, it could be observed that some users wanted to use the rotation of the wrist to choose the key set, which seems to be natural. Wrist rotation can be used for the PinchText technique. Third, PinchText can be used in mobile scenes. As we only conducted static lab studies, the performances of PinchText using in different mobile scenes need more studies. Fourth, when inputting passwords, names, etc., it is necessary to input single characters. More exploration of the OOV input performance is needed. Fifth, as we want to investigate the ideal hand movement range, and the participants are not familiar with HMDs, which may introduce inaccuracies, we conduct the first study with an LCD screen where participants feel more comfortable. However, that causes a potential decoupling between the first and second user studies. It would be better to investigate the ideal hand movement range with participants who are used to using HMDs. Sixth, according to the elapsed time between two key selections 5.4.3 and the frequency of letters used in words, we can optimize the keyboard layout to improve the input speed. Finally, the prototype of PinchText was implemented by using conductive tape made of copper to detect the user's pinch gestures. In the future, other devices could be used to detect gestures instead of wearing gloves, such as the method similar to (Soliman et al., 2018). Thus, users could input using bare hands.

8. Conclusion

This paper proposes a new text entry technique, PinchText, which combines the hand position in the air and pinch gestures, enabling one-handed text entry for VR, AR, and MR. This technique could be used in both Hand-Up and Hand-Down arm postures. Thereinto, users can input when the hand is naturally placed vertically, which helps to decrease fatigue in long-time use. First, Study 1 investigates the users' typing behavior of four methods with PinchText, which determines the hand position range of middle key sets. Then, Study 2 examines the UpV and DownV groups' typing performance and learning curves and compares them with the common pinch-based with QWERTY keyboard. Also, we analyzed the error types and typing behaviors, giving some insights for the further implementation of PinchText. Finally, the virtual keyboard is placed in the lower-left in the field of view instead of the center and explored if the size of the virtual keyboard would affect the entry efficiency. The results show that users can achieve 12.71 WPM in UpV group and 11.14 WPM in the DownV group both with less than 0.5% uncorrected error rate after 60-phrase training and PinchText is 71% faster than the common pinch-based technique with QWERTY keyboard; virtual keyboard as visual feedback for users could be placed elsewhere instead of the center of the field of view and its size could be reduced, helping decrease the occlusion. Overall, PinchText is an efficient and comfortable one-handed text entry technique for mobile scenes in HMDs.

Notes

1. https://www.vive.com/us/accessory/base-station/
2. https://www.optitrack.com/

Disclosure statement

No potential competing interest was reported by the author(s).

Funding

This work was supported by the National Key Research and Development Program of China [No. 2018YFB1005002] and the National Natural Science Foundation of China [No. 62072036] and the 111 Project [B18005].

References

Ahn, S., Heo, S., & Lee, G. (2017). Typing on a smartwatch for smart glasses. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (pp. 201–209). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3132272.3134136

Arif, A. S., & Mazalek, A. (2016). A survey of text entry techniques for smartwatches. In *Proceedings, part II, of the 18th international conference on human-computer interaction. Interaction platforms and techniques - volume 9732* (pp. 255–267). Springer-Verlag.

Azenkot, S., & Zhai, S. (2012). Touch behavior with different postures on soft smartphone keyboards. In *Proceedings of the 14th International Conference on Human-Computer Interaction With Mobile Devices and Services* (pp. 251–260). Association for Computing Machinery. https://doi.org/10.1145/2371574.2371612

Bajer, B., MacKenzie, I. S., & Baljko, M. (2012). Huffman base-4 text entry glove (h4 teg). In *2012 16th International Symposium on Wearable Computers* (pp. 41–47). USA: IEEE Computer Society.

Boletsis, C., & Kongsvik, S. (2019). Text input in virtual reality: A preliminary evaluation of the drum-like vr keyboard. *Technologies*, *7*(2), 31. https://doi.org/10.3390/technologies7020031

Bowman, D. A., Wingrave, C. A., Campbell, J. M., Ly, V. Q., & Rhoton, C. J. (2002). Novel uses of pinch gloves tm for virtual environment interaction techniques. *Virtual Reality*, *6*(3), 122–129. https://doi.org/10.1007/s100550200013

Chen, S., Wang, J., Guerra, S., Mittal, N., & Prakkamakul, S. (2019). Exploring word-gesture text entry techniques in virtual reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3290607.3312762

Dash, S. (2017). *Bluetap – the ultimate virtual-reality (vr) keyboard*. https://medium.com/eunoia-i-o/bluetap-the-ultimate-virtual-reality-vr-keyboard-77f1e3d57d6f Library Catalog: medium.com

Dementyev, A., & Paradiso, J. A. (2014). Wristflex: Low-power gesture input with wrist-worn pressure sensors. In *Proceedings of the 27th Annual Acm Symposium on User Interface Software and Technology* (pp. 161–166). Association for Computing Machinery.

Dobbelstein, D., Hock, P., & Rukzio, E. (2015). Belt: An unobtrusive touch input device for head-worn displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2135–2138). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702450

Dudley, J. J., Vertanen, K., & Kristensson, P. O. (2018). 12) Fast and precise touch-based text entry for head-mounted augmented reality with variable occlusion. *ACM Transactions on Computer-Human Interaction*, *25*(6), 1–40. https://doi.org/10.1145/3232163

Dudley, J. J., Benko, H., Wigdor, D., & Kristensson, P. O. (2019). Performance envelopes of virtual keyboard text input strategies in virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 289–300). IEEE Computer Society. https://doi.org/10.1109/ISMAR.2019.00027

Fashimpaur, J., Kin, K., & Longest, M. (2020). Pinchtype: Text entry for virtual and augmented reality using comfortable thumb to fingertip pinches. In *Extended Abstracts of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3334480.3382888

Findlater, L., Wobbrock, J. O., & Wigdor, D. (2011). Typing on flat glass: Examining ten-finger expert typing patterns on touch surfaces. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 2453–2462). Association for Computing Machinery. https://doi.org/10.1145/1978942.1979301

Frey, B., Southern, C., & Romero, M. (2011). Brailletouch: Mobile texting for the visually impaired. In C. Stephanidis (Ed.), *Universal access in human-computer interaction. Context diversity* (pp. 19–25). Springer Berlin Heidelberg.

Goldstein, M., Book, R., Alsiö, G., & Tessa, S. (1999). Non-keyboard qwerty touch typing: A portable input interface for the mobile user. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 32–39). Association for Computing Machinery. https://doi.org/10.1145/302979.302984

Green, N., Kruger, J., Faldu, C., & St. Amant, R. (2004). A reduced qwerty keyboard for mobile text entry. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1429–1432). Association for Computing Machinery. https://doi.org/10.1145/985921.986082

Grossman, T., Chen, X. A., & Fitzmaurice, G. (2015). Typing on glasses: Adapting text entry to smart eyewear. In *Proceedings of the 17th International Conference on Human-Computer Interaction With Mobile Devices and Services* (pp. 144–152). Association for Computing Machinery. https://doi.org/10.1145/2785830.2785867

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

Gugenheimer, J., Dobbelstein, D., Winkler, C., Haas, G., & Rukzio, E. (2016). Facetouch: Enabling touch interaction in display fixed UIs for mobile virtual reality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology, 10* (pp. 49–60).

Gupta, A., Ji, C., Yeo, H.-S., Quigley, A., & Vogel, D. (2019). Rotoswype: Word-gesture typing using a ring. In *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems* (pp. 1–12). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300244

Gupta, A., Samad, M., Kin, K., Kristensson, P., & Benko, H. (2020). Investigating Remote Tactile Feedback for Mid-Air Text-Entry in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (350–360). Los Alamitos, CA, USA: IEEE Computer Society. https://doi.org/10.1109/ISMAR50242.2020.00062

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

Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., & Irani, P. (2014). Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 1063–1072). Association for Computing Machinery. https://doi.org/10.1145/2556288.2557130

Hsieh, Y.-T., Jylhä, A., Orso, V., Gamberini, L., & Jacucci, G. (2016). Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses. In *Proceedings of the 2016 Chi Conference on Human Factors in Computing Systems* (pp. 4203–4215). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858436

HTC vive (2020). *Website*. https://www.vive.com/cn/

Huang, D.-Y., Chan, L., Yang, S., Wang, F., Liang, R.-H., & Yang, D.-N. (2016). 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* (pp. 1526–1537). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858483

Hutama, W., Harashima, H., Ishikawa, H., & Manabe, H. (2021). Hmk: Head-mountedkeyboard for text input in virtual or augmented reality. In *The Adjunct Publication of the 34th Annual Acm Symposium on User Interface Software and Technology* (pp. 115–117). Association for Computing Machinery. https://doi.org/10.1145/3474349.3480195

Itoh, Y., Langlotz, T., Sutton, J., & Plopski, A. (2021). Towards indistinguishable augmented reality: A survey on optical see-through head-mounted displays. *ACM Computing Surveys*, *54*(6), 1–36. https://doi.org/10.1145/3453157

James, C. L., & Reischel, K. M. (2001). Text input for mobile devices: Comparing model prediction to actual performance. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 365–371). Association for Computing Machinery. https://doi.org/10.1145/365024.365300

Jiang, H., & Weng, D. (2020). Hipad: Text entry for head-mounted displays using circular touchpad. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 692–703). IEEE Computer Society. https://doi.org/10.1109/VR46266.2020.00092

Jiang, H., Weng, D., Zhang, Z., Bao, Y., Jia, Y., & Nie, M. (2018). Hikeyb: High-efficiency mixed reality system for text entry. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (pp. 132–137). IEEE Computer Society. https://doi.org/10.1109/ISMAR-Adjunct.2018.00051

Jiang, H., Weng, D., Zhang, Z., & Chen, F. (2019). Hifinger: One-handed text entry technique for virtual environments based on touches between fingers. *Sensors*, *19*(14), 3063. https://doi.org/10.3390/s19143063

Kim, Y. R., & Kim, G. J. (2017). Hovr-type: Smartphone as a typing interface in VR using hovering. In *2017 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 200–203). IEEE Computer Society. https://doi.org/10.1109/ICCE.2017.7889285

Knierim, P., Schwind, V., Feit, A. M., Nieuwenhuizen, F., & Henze, N. (2018). Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp.1–9). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3173574.3173919

Komninos, A., & Dunlop, M. (2014). Text input on a smart watch. *IEEE Pervasive Computing*, *13*(4), 50–58. https://doi.org/10.1109/MPRV.2014.77

König, W. A., Gerken, J., Dierdorf, S., & Reiterer, H. (2009). Adaptive pointing – design and evaluation of a precision enhancing technique for absolute pointing devices. In T. Gross (Eds.), *Human-Computer Interaction – Interact 2009* (pp. 658–671). Springer Berlin Heidelberg.

Lee, L. H., Lam, K. Y., Li, T., Braud, T., Su, X., & Hui, P. (2019). Quadmetric optimized thumb-to-finger interaction for force assisted one-handed text entry on mobile headsets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(3), 1–27. https://doi.org/10.1145/3351252

Lee, L. H., Yung Lam, K., Yau, Y. P., Braud, T., Hui, P. (2019). Hibey: Hide the keyboard in augmented reality. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom)* (pp. 1–10). IEEE Computer Society.

Lee, D., Kim, J., & Oakley, I. (2021). Fingertext: Exploring and optimizing performance for wearable, mobile and one-handed typing. In *Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445106

Li, F. C. Y., Guy, R. T., Yatani, K., & Truong, K. N. (2011). The 1line keyboard: A qwerty layout in a single line. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (pp. 461–470). Association for Computing Machinery. https://doi.org/10.1145/2047196.2047257

Lu, Y., Yu, C., Yi, X., Shi, Y., & Zhao, S. (2017). Blindtype: Eyes-free text entry on handheld touchpad by leveraging thumb's muscle memory. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(2), 1–24. https://doi.org/10.1145/3090083

MacKenzie, I. S. (2002). *A note on calculating text entry speed* [Unpublished work]. http://www.yorku.ca/mack/RN-TextEntrySpeed.html

MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems* (pp. 754–755). Association for Computing Machinery. https://doi.org/10.1145/765891.765971

Majaranta, P., Ahola, U.-K., & Špakov, O. (2009). Fast gaze typing with an adjustable dwell time. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 357–360). Association for Computing Machinery. https://doi.org/10.1145/1518701.1518758

Markussen, A., Jakobsen, M. R., & Hornbæk, K. (2014). Vulture: A mid-air word-gesture keyboard. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 1073–1082). Association for Computing Machinery. https://doi.org/10.1145/2556288.2556964

McGill, M., Boland, D., Murray-Smith, R., & Brewster, S. (2015). A dose of reality: Overcoming usability challenges in vr head-mounted displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2143–2152). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702382

Mehring, C., Kuester, F., Singh, K. D., & Chen, M. (2004). Kitty: Keyboard independent touch typing in VR. In *IEEE Virtual Reality 2004* (pp. 243–244). IEEE Computer Society. https://doi.org/10.1109/VR.2004.1310090

Ni, T., Bowman, D., & North, C. (2011). Airstroke: Bringing unistroke text entry to freehand gesture interfaces. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 2473–2476). Association for Computing Machinery. https://doi.org/10.1145/1978942.1979303

Oculus Quest (2020). *Website*. https://www.oculus.com/quest/

Peshock, A., Duvall, J., & Dunne, L. E. (2014). Argot: A wearable one-handed keyboard glove. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program* (pp. 87–92). Association for Computing Machinery. https://doi.org/10.1145/2641248.2641266

Pick, S., Puika, A. S., & Kuhlen, T. W. (2016). Swifter: Design and evaluation of a speech-based text input metaphor for immersive virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 109–112). https://doi.org/10.1109/3DUI.2016.7460039

Poirier, F., & Belatar, M. (2016). Uniwatch: A soft keyboard for text entry on smartwatches using 3 keys. In M. Kurosu (Ed.), *Human-computer interaction. interaction platforms and techniques* (pp. 341–349). Springer International Publishing.

Prätorius, M., Valkov, D., Burgbacher, U., & Hinrichs, K. (2014). Digitap: An eyes-free VR/AR symbolic input device. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (pp. 9–18). Association for Computing Machinery. https://doi.org/10.1145/2671015.2671029

Profita, H. P., Clawson, J., Gilliland, S., Zeagler, C., Starner, T., Budd, J., & Luen Do, E. Y. (2013). Don't mind me touching my wrist: A case study of interacting with on-body technology in public. In *Proceedings of the 2013 International Symposium on Wearable Computers* (pp. 89–96). Association for Computing Machinery. https://doi.org/10.1145/2493988.2494331

Qin, R., Zhu, S., Lin, Y.-H., Ko, Y.-J., & Bi, X. (2018). Optimal-t9: An optimized t9-like keyboard for small touchscreen devices. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (pp. 137–146). Association for Computing Machinery. https://doi.org/10.1145/3279778.3279786

Rajanna, V., & Hansen, J. P. (2018). Gaze typing in virtual reality: Impact of keyboard design, selection method, and motion. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. Association for Computing Machinery. https://doi.org/10.1145/3204493.3204541

Rico, J., & Brewster, S. (2010). Usable gestures for mobile interfaces: Evaluating social acceptability. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 887–896). Association for Computing Machinery. https://doi.org/10.1145/1753326.1753458

Schick, A., Morlock, D., Amma, C., Schultz, T., & Stiefelhagen, R. (2012). Vision-based handwriting recognition for unrestricted text input in mid-air. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (pp. 217–220). Association for Computing Machinery. https://doi.org/10.1145/2388676.2388719

Serrano, M., Ens, B. M., & Irani, P. P. (2014). Exploring the use of hand-to-face input for interacting with head-worn displays. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 3181–3190). Association for Computing Machinery. https://doi.org/10.1145/2556288.2556984

Soliman, M., Mueller, F., Hegemann, L., Roo, J. S., Theobalt, C., & Steimle, J. (2018). Fingerinput: Capturing expressive single-hand thumb-to-finger microgestures. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (pp. 177–187). Association for Computing Machinery. https://doi.org/10.1145/3279778.3279799

Speicher, M., Feit, A. M., Ziegler, P., & Krüger, A. (2018). Selection-based text entry in virtual reality. In *Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3174221

Sridhar, S., Feit, A. M., Theobalt, C., & Oulasvirta, A. (2015). Investigating the dexterity of multi-finger input for mid-air text entry. In *Proceedings of the 33rd Annual ACM Conference on*

Human Factors in Computing Systems (pp. 3643–3652). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702136

Thomas, B., Tyerman, S., & Grimmer, K. (1998). Evaluation of text input mechanisms for wearable computers. *Virtual Reality*, *3*(3), 187–199. https://doi.org/10.1007/BF01408561

Toney, A., Mulley, B., Thomas, B. H., & Piekarski, W. (2003). Social weight: Designing to minimise the social consequences arising from technology use by the mobile professional. *Personal and Ubiquitous Computing*, *7*(5), 309–320. https://doi.org/10.1007/s00779-003-0245-8

Vatavu, R.-D., & Vanderdonckt, J. (2020). Design space and users' preferences for smartglasses graphical menus: A vignette study. In *19th International Conference on Mobile and Ubiquitous Multimedia* (pp. 1–12). Association for Computing Machinery. https://doi.org/10.1145/3428361.3428467

Walker, J., Kuhl, S., & Vertanen, K. (2016). Decoder-assisted typing using an HMD and a physical keyboard. In *CHI '16: Extended abstracts of the the ACM conference on human factors in computing systems*. Association for Computing Machinery.

Walker, J., Li, B., Vertanen, K., & Kuhl, S. (2017). Efficient typing on a visually occluded physical keyboard. In *Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems* (pp. 5457–5461). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025783

Wang, C.-Y., Chu, W.-C., Chiu, P.-T., Hsiu, M.-C., Chiang, Y.-H., & Chen, M. Y. (2015). Palmtype: Using palms as keyboards for smart glasses. In *Proceedings of the 17th International Conference on Human-Computer Interaction With Mobile Devices and Services* (pp. 153–160). Association for Computing Machinery. https://doi.org/10.1145/2785830.2785886

Way, D., Paradiso, J. (2014). A usability user study concerning free-hand microgesture and wrist-worn sensors. In *2014 11th International Conference on Wearable and Implantable Body Sensor Networks* (pp. 138–142).

Whitmire, E., Jain, M., Jain, D., Nelson, G., Karkar, R., Patel, S., & Goel, M. (2017). Digitouch: Reconfigurable thumb-to-finger input and text entry on head-mounted displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–21. https://doi.org/10.1145/3130978

Wolf, D., Gugenheimer, J., Combosch, M., & Rukzio, E. (2020). Understanding the Heisenberg effect of spatial interaction: A selection induced error for spatially tracked input devices. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376876

Wong, P. C., Zhu, K., & Fu, H. (2018). Fingert9: Leveraging thumb-to-finger interaction for same-side-hand text entry on smartwatches. In *Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173752

Xu, Z., Chen, W., Zhao, D., Luo, J., Wu, T.-Y., Gong, J., Yin, S., Zhai, J., & Yang, X. -D. (2020). Bitiptext: Bimanual eyes-free text entry on a fingertip keyboard. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376306

Xu, Z., Wong, P. C., Gong, J., Wu, T.-Y., Nittala, A. S., Bi, X., Steimle, J., Fu, H., Zhu, K., & Yang, X.-D. (2019). Tiptext: Eyes-free text entry on a fingertip keyboard. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (pp. 883–899). Association for Computing Machinery. https://doi.org/10.1145/3332165.3347865

Yeo, H.-S., Phang, X.-S., Ha, T., Woo, W., & Quigley, A. (2017). Titan: Exploring midair text entry using freehand input. In *Proceedings of the 2017 Chi Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3041–3049). Association for Computing Machinery. https://doi.org/10.1145/3027063.3053228

Yildirim, C., & Osborne, E. (2020). Text entry in virtual reality: A comparison of 2d and 3d keyboard layouts. In *HCI International 2020 – Late Breaking Papers: Virtual and Augmented Reality: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings* (pp. 450–460). Springer-Verlag.

Yu, C., Gu, Y., Yang, Z., Yi, X., Luo, H., & Shi, Y. (2017). Tap, dwell or gesture? Exploring head-based text entry techniques for hmds. In *Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems* (pp. 4479–4488). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025964

Yu, D., Fan, K., Zhang, H., Monteiro, D., Xu, W., & Liang, H. (2018). Pizzatext: Text entry for virtual reality systems using dual thumbsticks. *IEEE Transactions on Visualization and Computer Graphics*, *24*(11), 2927–2935.

Zhu, S., Luo, T., Bi, X., & Zhai, S. (2018). Typing on an invisible keyboard. In *Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3174013

About the Authors

Haiyan Jiang is a Ph.D. candidate at School of Optics and Photonics, Beijing Institute of Technology, China. Her research interests include virtual reality, mixed reality, and human-computer interaction (HCI). She is especially interested in deep learning-based HCI and virtual hand generation.

Dongdong Weng received the Ph.D. degree in optical engineering from the Beijing Institute of Technology, Beijing, China, in 2006. He is currently a Professor of Optical Engineering with the School of Optics and Photonics, Beijing Institute of Technology. His research interests include virtual reality, mixed reality, human-computer interaction, and digital human.

Xiaonuo DongYe is a graduate student in the School of Optics and Photonics, Beijing Institute of Technology, China. His primary research interests are in mixed reality and display systems. In particular, his research investigates human-computer physical interaction in virtual reality.

Yue Liu is a professor of Optical Engineering and Software at Beijing Institute of Technology, China. His research interests include computer vision, human computer interaction, virtual and augmented reality, evaluation of 3D display devices etc.