

Hands-free Multi-Type Character Text Entry in Virtual Reality

Tingjie Wan¹, Rongkai Shi¹, Wenge Xu², Yue Li¹, Katie Atkinson³, Lingyun Yu¹ and Hai-Ning Liang^{1*}

¹Department of Computing, Xi'an Jiaotong-Liverpool University, Suzhou, China.

²DMT Lab, Birmingham City University, Birmingham, UK.

³Department of Computer Science, University of Liverpool, Liverpool, UK.

*Corresponding author(s). E-mail(s): haining.liang@xjtlu.edu.cn;

Abstract

Multi-type characters, including uppercase and lowercase letters, symbols, and numbers, are essential in text entry activities. Although multi-type characters are used in passwords, instant messages, and document composition, there has been limited exploration of multi-character text entry for virtual reality head-mounted displays (VR HMDs). Typically, multi-type character entry requires four kinds of keyboards between which users need to switch. In this research, we explore hands-free approaches for rapid multi-type character entry. Our work explores two efficient and usable hands-free approaches for character selection: eye blinks and dwell. To enable quick switching between keyboards, we leverage the usability and efficiency of continuous head motions in the form of cross-based activation. In a pilot study, we explored the usability and efficiency of four locations of the switch keys, the two hands-free selection mechanisms, and crossing-based switching. In the main experiment, we evaluated four user-inspired layouts designed according to the findings from the pilot study. Results show that both blinking and dwell can work well with crossing-based switching and could lead to a relatively fast text entry rate (5.64 words-per-minute (WPM) with blinking and 5.42 WPM with dwell) with low errors (lower than 3% not corrected error rate (NCER)) for complex 8-digit passwords with upper/lowercase letters, symbols, and numbers. For sentences derived from the Brown Corpus, participants can reach 8.48 WPM with blinking and 7.78 WPM with

dwelling. Overall, as a first exploration, our results show that it is usable and efficient to perform hands-free text entry in VR using either eye blinks or dwell for character selection and crossing for mode switching.

Keywords: Virtual reality; text entry; hands-free interaction; multi-type character entry; mode switching; keyboard layout; user study

1 Introduction

Text entry is indispensable for interactive systems, including virtual reality head-mounted displays (VR HMDs). VR HMDs have been used for professional technical learning and training and are being considered as a platform for the office of the future, remote collaboration, and other scenarios [1–5]. These scenarios require people to enter text as they do with traditional computing devices for daily communication and text composition, such as writing emails or documents using laptops, desktops, or smartphones.

Upper/lowercase letters, symbols, and numbers are essential in daily communication. Meanings of words could change significantly with the presence (or absence) of capitals. For instance, ‘August’ is the eighth month of the year, while ‘august’ is an adjective, which means impressive and respected. Symbols can help people express emotions with emojis (e.g., the frown emoji :- (or happy emoji :-)), which are frequently used in instant messaging. Many of these emoticons have even been cataloged in dictionaries [6]. Numbers can make text messages highly versatile and easily readable, such as dates and amounts. In addition, for passwords to reach an acceptable level of security, they need to use a combination of numbers, upper/lowercase letters, and symbols, as the greater the possible combinations, the lower the risk that a password can be cracked [7]. In some recent VR applications, password entry is needed, for instance, when the users use a VR browser and visit a website that requires logging in to their personal account [8]. Although biometric approaches, such as iris and fingerprint scanning, have been used for identity verification, passwords are one of the most common ways for authentication and, as such, are still necessary because using biometric features for authentication is not often possible and accurate enough, requires additional specialized equipment that can be cumbersome to set up, and is not always reliable [9, 10]. They can also be used as an alternative authentication method as biometric approaches are not always effective or possible [11]. To allow VR users to enter complex sentences or strings like passwords, there is a need to explore text entry techniques beyond lowercase letters only, which has been the primary focus of most text entry techniques. Based on our review of the literature, multi-type characters like passwords have not been the primary focus, with one only exception from [12].

Text entry in VR usually requires the user to hold a handheld controller input [13–17]. However, there are cases where using hands and controllers is

not suitable: (1) users' hands are occupied with other tasks (e.g., surgery training); (2) controllers are not readily available; (3) users with hand/arm-related motor impairments or inefficiencies and who have difficulties with precise input using a handheld controller [18–20]. In these cases, a hands-free technique leveraging other parts of users (like head motions, eye movements or blinks, and dwell for cursor movement or selection confirmation [21–24]) represents a feasible and practical approach. In recent years, hands-free approaches for text entry in VR have been explored to allow users to enter text in VR with good performance and user experience, some emphasizing *cursor movements* and *selection mechanisms* [22, 23, 25] and others on the *keyboard layout* [18, 26]. While these techniques allow fast text entry rates and a positive user experience, their focus has mainly been on lowercase characters only. There is very limited research on developing new text entry techniques that are hands-free and allow inputting different types of characters, including uppercase letters, symbols, and numbers. While it is possible to switch modes directly using text input methods to enable multi-type of characters. As VR supports multiple types of interaction, it is worth exploring whether involving various interaction techniques can lead to better text entry performance for multi-type characters.

Traditional keyboards use switch keys to allow the input of various types of characters. For instance, the QWERTY keyboard uses the 'Caps Lock' key to switch from lowercase to uppercase letters and vice-versa (see Figure 1a) or uses a combination of the 'Shift' key and another key to input one of the two possible characters of the key (e.g., 1 and ! or A and a). These two strategies are used to reduce the size of the keyboard, as having keys to represent all types of characters will make a physical keyboard very large and impractical (and more expensive). While less affected by physical constraints, virtual keyboards follow the physical keyboards with switch keys that allow moving from one keyboard to another to access different types of characters. Given its flexibility, the location of the switch keys in a virtual keyboard could easily be rearranged to maximize performance and user preference while minimizing workload, just like the size and location of character keys [27]. One other aspect that could improve performance is to allow for a fast and continuous mode switching or transition process. Given that we aim to explore a hands-free approach, *crossing-based activation* [28, 29] can be a natural and efficient approach for mode switching (see Figure 1b for an example), especially when head motions are involved, since crossing does not interrupt or break the continuous head/cursor movements, thereby reducing activation time and lowering the motor requirements and movement control [30, 31].

This paper presents a systematic exploration of hands-free text entry in VR involving crossing-based mode switching for multi-type character entry. To our knowledge, this work is the first to explore multi-type character text entry with a virtual keyboard in VR that is entirely hands-free. As with any new text entry technique, the keyboard layout plays a key role because it determines how easy or difficult it is to learn to use it. Thus, we used the most common QWERTY keyboard layout as the foundation for designing the new approach.

Our approach leverages users’ familiarity with the keyboard layout and the concept of switch keys to enable convenient hands-free character selection and smooth mode switching via crossing interaction.

Our work involves a pilot study that explores the impact of four positions of the switch keys, crossing activation, and two hands-free selection mechanisms (dwell and eye blinks, both using head pointing) on performance and user preference for entering complex passwords.

Then, we run another user study to explore the performance of four layouts inspired by feedback from the pilot study. In addition to passwords, the participants also type sentences selected from the Brown Corpus [32]. They are more representative of people’s use in daily life (e.g., “Newark Evening News, March 22, 1961, p.25”) and more complex than the MacKenzie phrase set which is commonly used in text entry studies [33]. The results show that the participants can achieve a relatively fast performance for Brown Corpus sentences (8.48 words-per-minute (WPM) with blinking and 7.78 WPM with dwell), and passwords (5.64 WPM with blinking and 5.42 WPM with dwell). Overall, the results show that our head based hands-free approach with crossing-based switching is a usable and efficient technique for multi-type character entry. These results provide a strong foundation for further research in this area that is of great importance if we have future VR systems that can take the place of current mobile/desktop computers. In short, the following are the main contributions of this work:

- A first exploration of multi-type character text entry with a virtual keyboard in VR that is entirely hands-free;
- A first case implementation of crossing-based mode-switching in a text entry technique in VR;
- Two new metrics (mode-switching time and switch-key movement time) for measuring performance and usability when mode switching is involved; and
- A comparative experiment of user performance and preference of hands-free text entry using two corpora (Brown Corpus sentences and passwords).

2 Related Work

2.1 Keyboard Layout in VR

The QWERTY keyboard is still the most commonly used layout for interactive systems [34], including VR HMDs [35], because (1) users are familiar with it, (2) users are often not willing to invest time in learning new layouts [36, 37], (3) its performance is acceptable in its virtual form [25], and (4) users can easily shift to emerging platforms, such as to smartphones and now VR/AR. Some studies have used a physical keyboard and visualized it in the virtual environment to allow typing in VR [38–41]. Experienced experts’ average typing speed reached 69.172 WPM for sentences that are entirely in lowercase [38], and 41.5 WPM for complex sentences including multiple types of characters [39]. Though this approach could lead to a fast typing speed in general,

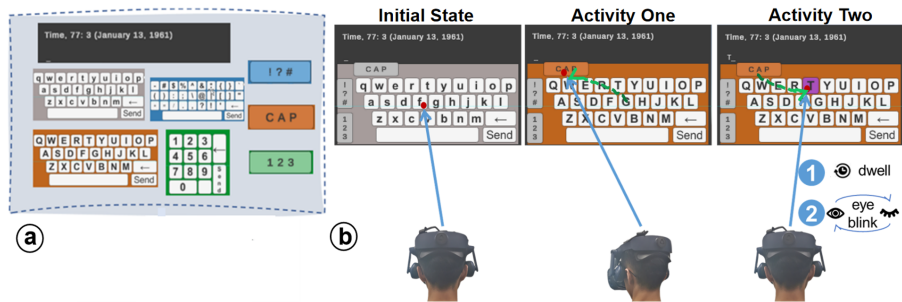


Fig. 1 (a) Inputting a sentence that includes multi-type characters needs four kinds of keyboards and at least three switch keys to move back and forth between these keyboards. (b) *Initial State* - the initial state of the keyboard. *Activity One* - The user switches to uppercase by moving his head to cross the ‘CAP’ switch key to change to the uppercase keyboard (i.e., crossing-based activation); *Activity two* - character selection can be achieved via hands-free text entry techniques (either dwell or eye blinks, both based on head pointing). The red dot on the keyboard layout is the cursor. The blue line indicates the head pointing direction, and the green line represents the movement trajectory of the cursor, both of which are only for demonstration and not shown in real typing scenarios.

it is not convenient or applicable for most users. In contrast, using a virtual form of a QWERTY keyboard can avoid this problem and now has become the most common way for text entry in VR HMD. For example, a drum-like keyboard, which utilizes controllers as drumsticks to “press” the keys via downward movements, leads to a speed of 24.61 WPM [42]. However, there are limitations to a virtual QWERTY keyboard in VR, especially for accessing characters other than lowercase letters. One major limitation is that virtual keyboards group the keys for switching modes in one corner (usually, the lower left side), requiring users to move the virtual pointer to that area to switch to uppercase letters, symbols, numbers, and vice versa. This centralizes the traffic and hand/neck motions, which could cause discomfort and fatigue [43]. Also, having all switch keys in a small area can lead to more false positives.

One popular alternative layout to QWERTY is the circular design [13, 17, 18]. Placing characters in a circular format and using a crossing selection style could outperform the traditional QWERTY keyboard [18]. [44] proposed a T9-like 3×3 layout. Users need to press the key of the intended character multiple times to finish the input. Other layouts, such as the 12-Key keyboard [45, 46] and cubic arrangement [47], have also been explored. While some of these layouts are shown to be practical, their focus is on lowercase letters without taking into account the issues involved when typing other character types. Also, all of them require an initial learning effort, which users do not prefer [36, 37]. Therefore, for our proposed approach, the QWERTY keyboard layout is used to avoid additional learning effort due to unfamiliar design elements.

2.2 Hands-free Selection in VR

Given various common scenarios where users’ hands are unavailable for interaction, researchers have investigated hands-free approaches to meet the different

demands. Object selection is one of the most important interaction aspects in VR, which is also one of the basic units of a text entry task. Prior hands-free selection studies have focused on voice-, eye-, and head-based approaches [48]. A voice-based approach may require users to say the name of the target objects (e.g., [49]). On the other hand, with eye- and head-based approaches, the interaction procedure generally involves two steps: point and confirm [48]. Users first control their eyes or head to move the cursor targeting the object and trigger confirmation by an action. The confirmation action could be pressing a button on controller [50] (though this is not entirely hands-free), blinking eyes [22], dwelling on the target for a period of time [22, 51, 52], or neck gestures [22].

Crossing-based selection has been proposed for target selection, initially for 2D UIs [28] and recently for VR [29]. Unlike dwell and eye blinks where users are required to stop the pointer over an object of interest when making a selection, crossing requires users to move the pointer beyond the target boundary to select it [28], which reduces selection time and the requirements for movement control [30, 31]. Although crossing suffers in performance and usability when distracting objects surround the intended target (e.g., keys in a keyboard) [53], it works well when there are no distractors. A recent study shows that crossing can substitute raycast-based pointing in object selection in VR with a shorter or similar time performance plus a higher or similar accuracy [29]. Some VR hands-free text entry techniques also utilized crossing-based selection. For example, EyeSwipe [54] uses gaze-crossing paths for text entry. On the other hand, GestureType [25] and RingText [18] use head motions to move the pointer to cross the key regions on a QWERTY keyboard and a circular keyboard, respectively. Results showed that it outperforms dwell in VR text entry [18]. Given the benefits of crossing (continuous nature, efficiency for targets without distractors, good performance for relatively large objects), we use crossing for mode switching, as it lends itself quite well for such a hands-free dynamic task and features of switch keys. As our results show, crossing is very efficient for mode switching and is acceptable by VR users.

2.3 Hands-free Text Entry in VR/AR

Despite the increasing popularity of hands-free text entry techniques for VR systems, most such studies have focused on lowercase letters. Table 1 summarizes hands-free text entry methods that have been developed for VR HMDs. Speech/Voice was excluded as its several significant disadvantages: (1) it requires to be operated in a relatively quiet environment [55]-, (2) it may not be socially acceptable [37], (3) it is not suitable for people with a non-native accent, and (4) it could lead to privacy concerns (e.g., when entering passwords) [18]. These issues prevent speech-based text entry methods from being used in many places, like offices, libraries, and universities. Our survey only led to one paper [23] that considered the need for multi-types character, but even this one has not conducted a user study with mode switching. However, as mentioned, uppercase alphabet letters, symbols, and numbers are

all essential in people’s daily text entry activities (e.g., entering passwords or instant messages which often come with text-based emoticons like a :) smiley). Password-based authentication is currently the main way to authenticate a user [56]. For the best security, setting a longer password with 8 characters or more of various types is recommended [57–59]. As such, entering them requires switching modes. Likewise, instant messages use emoticons composed of various letters/symbols because emoticons play an important social role and are used to compensate or imitate facial expressions when face-to-face communication is not possible [60, 61]. In short, it is important to have an efficient and usable text entry approach with low workloads for VR HMDs that include symbols, uppercase alphabets, and numbers.

Table 1 Summary of hands-free text entry techniques in VR. Note that performance is based on entering lowercase characters only.

Layout (Reference)	Pointing Method	Selection Method	Mode Switching	WPM ^a
QWERTY [16]	Head	Dwell	No	10.20
QWERTY [25]	Head	Dwell	No	10.59
QWERTY [22]	Head	Eye Blink; Dwell; Neck Motion	No	13.47; 11.18; 11.65
QWERTY [62]	Head	Dwell	No	19.98
Circular [18]	Head	Crossing	No	12.27
40-target [23]	Eye	Steady-state visual evoked potentials	Yes, but limited	N/A
QWERTY [63]	Head	Neck Motion	No	10.76

^aIn this column, we reported WPM in the last experiment session mentioned in the reference.

Candidate approaches for key selection that are hands-free and work well with head motions for cursor movement are based on eye blinks, a head dwell time, and neck motions (see Table 1). [22] tested these three types and found that eye blinks led to the fastest speed and highest user preference while neck motions led to low performance and high workload. As Table 1 shows, dwell has also been consistently used in hands-free techniques for text selection and has good performance and user preference. This work focuses on dwell and eye blinks (or blinking) for character selection, given their excellent performance and usability.

3 Keyboard Design and Evaluation Metrics

To design a hands-free virtual keyboard that supports multi-type characters in VR, we identified three design factors: (1) *keyboard layout*, (2) *key-selection*

mechanism, and (3) *mode-switching mechanism*. This section discusses our considerations towards these design factors to propose an efficient, easy-to-learn, and usable text entry technique. In addition, we introduce extended evaluation metrics that afford to measure multi-character input.

3.1 Layout, Size, and Position

The virtual keyboard used is based on a QWERTY layout to minimize any need to learn a new layout design and allow us to focus on mode switching and selection.

The keyboard is placed 50cm away from the center of the user’s view (see Figure 3a). The keyboard size is 36cm×15cm, and the size of each key is 2.8cm×2.8cm. The last row of the keyboard is used to show the space key/bar and the ‘Send’ key for moving to the following phrase. On top of the ‘Send’ key is the backspace key (‘←’).

3.2 Hands-free Key-Selection Mechanism

As mentioned in Section 2.3, dwell and eye blinking are chosen for character selection. Both selection mechanisms use head pointing, as the throughput and effective target widths of the head pointing are higher than eye gaze pointing [51]. The cursor controlled by users’ head movements is a red circle with a size of 1cm×1cm. Dwell allows users to type by hovering the cursor over a key for a predefined time (i.e., a dwell time).

After several pre-tests, we set the dwell time to be 300ms since it represents a suitable trade-off between speed and avoiding unintentional selections in our design. An issue with dwell is that users may continue to dwell on the same key after a selection while searching for the next key. To avoid this, we set a 600ms gap between the same key activation. It is reset if the cursor moves more than 1.4cm (a half key). A key is enlarged and its color is changed to purple inform users of its selection. Blinking, on the other hand, lets users type using eye blinks. Blinking of both eyes is chosen because a recent paper shows that it leads to much higher accuracy and comfort for character selection than using either eye alone [22]. We also set a 300ms time threshold for blinking (i.e., eye-close time). A 300ms eye closure time can help prevent inadvertent selections because it is longer than people’s spontaneous blinking time, which typically lasts around 100ms [64].

We also explored other approaches reported in other papers, using gestures in particular. For example, EyeSwipe [54] uses a gaze and lets users select the first and last characters of a word and gesture through the other characters. Candidate words are then shown for users to select. GestureType [25] is another approach that uses head motions and can lead to good performance in VR. However, it is not hands-free since controller buttons are still needed to indicate the start/end of the gesture. These are word-based approaches where the system predicts the possible word(s) based on users’ input and provides

suggested words but these approaches are not suitable because they cannot work with passwords and words with symbols and numbers.

3.3 Hands-free Mode Switching for Multi-character Input

We run a pre-pilot test with 16 participants to see if crossing, dwell, and eye blinks can serve as a mode-switching mechanism. We evaluated the performance of the three mechanisms using the metrics described in the following section, Section 3.4. The results showed that crossing outperformed dwell and blinking for mode switching and was also ranked higher in usability. Therefore, we chose mode-switching using crossing to allow transitions between lowercase letters, uppercase letters, symbols, and numbers (see Figure 1b). In addition, our data showed that it was more practical and natural to have the lowercase keyboard as default and enable a quick way to return to it, as lowercase letters are more frequently used. One way was for users to move the cursor anywhere outside the keyboard area, which was also shown to be efficient and usable. We adopted this quick switch method in our keyboard design.

As mentioned, the switch keys in virtual keyboards are usually placed in the lower left corner, which can be inefficient and error-prone for mode switching via a hands-free approach. With the switching mechanism determined for our keyboard design, we first wanted to explore and evaluate the influence of the positions of the switch keys, especially to see what position can better support crossing-based mode-switching that is in hands-free.

3.4 Evaluation Metrics

Text entry speed and error rate are two common metrics. Speed is measured in words-per-minute (WPM) [65], with a word defined as five consecutive letters including upper and lower cases, numbers, symbols, and spaces. Error rate is calculated based on the standard character level typing metrics, where the total error rate (TER) = not corrected error rate (NCER) + corrected error rate (CER) [66, 67].

In addition to speed and error rate, we propose that two additional metrics are important when mode switching is involved. There are four modes: lowercase (default), uppercase, numbers, and symbols. These four modes form 12 possible transitions between any two of them and as we show the direction of the transitions matters. We can group these transitions according to the target mode, which would lead to four categories: switch-to-uppercase (CAP), switch-to-lowercase (LOW), switch-to-numbers (NUM), and switch-to-symbols (SYM). Accordingly, in each sentence, these two additional metrics can be measured:

- *Mode-switching time*: the duration for doing a mode switch when switching from the current keyboard layout to another. That is, the time for moving from the just-triggered character key to the mode-switching key being crossed (i.e., triggered). We involve an average mode-switching time for each category of transitions (i.e., aforementioned CAP, LOW, NUM, and SYM),

and an average mode-switching time considering all types of switching for a sentence.

- *Switch-key movement time*: the average duration from the completion of a mode switch to the input of the next character minus the time to trigger an input, which is 300ms (the trigger time for both dwell and blinking key-selection mechanisms). In other words, we removed the time for confirmation of selection to get the “true” time cost of cursor movement after switching the mode.

An example can be seen in Figure 2, where a user aims to type a character ‘N’. To do this, the user starts from the default lowercase letter layout and controls the cursor to cross the ‘CAP’ key. This duration is the mode-switching time and can be categorized as switch-to-uppercase (or CAP for short). The keyboard responds to the mode switch and shows the uppercase letters. The user then navigates to the location of ‘N’ and confirms the selection. This duration is the switch-key movement time.

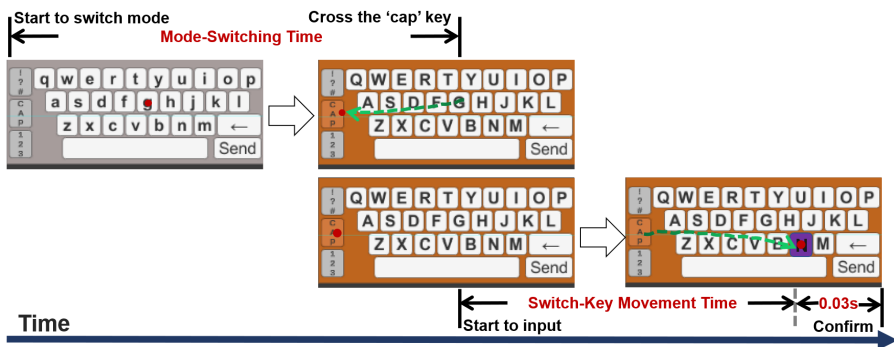


Fig. 2 An example of locating a capital letter ‘N’ from the default lowercase letter layout. The whole time duration can be divided into *mode-switching time* and *switch-key movement time*.

4 Pilot study

The pilot study explores the impact of the four positions of the switch keys (see Figure 3) on text entry performance with two selection mechanisms (blinking and dwell) for typing complex passwords.

4.1 Participants and Apparatus

We recruited 16 participants (8 females, 8 males; aged from 21 to 23, $M = 21.5$, $SD = 0.73$) from a local university. All were non-native English speakers and had normal or corrected-to-normal vision. No participants reported simulator sickness during or after the study.

We used an HTC VIVE Pro Eye HMD with a resolution of 1440×1600 pixels per eye, a 110° Field of View, and 90 Hz refresh rate. It was connected

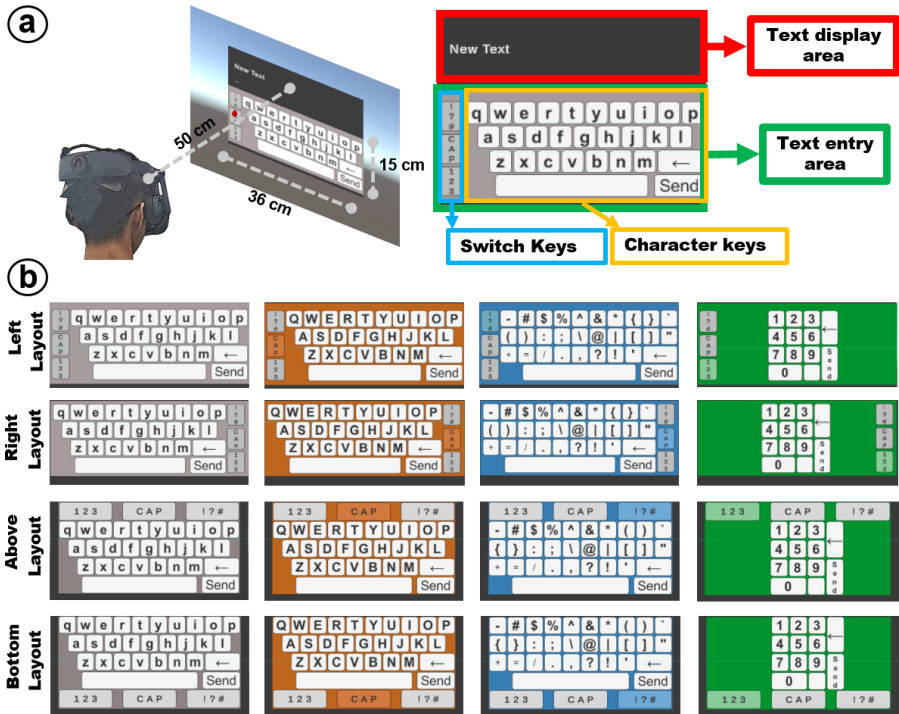


Fig. 3 (a) The keyboard in the user’s view and partitioning of keyboard functions. Four Keyboard layouts with the three switch keys placed on the (b) left-, right-, above-, and down-side of character keys. After a switch is made, the background color changes to the color of the mode: orange, blue, and green represent uppercase letters, symbols, and numbers, respectively. The default color is grey for lowercase letters.

to a Windows 10 PC with an i7-7700k CPU and a GTX 1080 GPU. The application used was implemented in Unity3D (version 2021.1) with the SteamVR Unity plugin (version 1.19.7) and VIVE Eye and Facial Tracking SDK (version 1.3.3.0). Participants were seated throughout the whole experiment.

4.2 Materials

To enable the evaluation of text entry performance for multi-type characters, we used the task of typing complex passwords. The passwords are 8-digit strings composed of randomly generated characters following password security rules [68] and must contain four types of characters and a maximum of two consecutive characters of the same type (Gy7V+KQ is one example password). All passwords using this corpus allow us to make sure all types of characters are involved in representing one of the most challenging typing tasks.

4.3 Experimental Design and Procedure

To minimize any impact of cross-learning effects and fatigue, we used a 2×4 mixed-subjects design with *Technique* as the between-subjects variable (blinking and dwell) and *Keyboard Layout* as the within-subjects variable (left, right, above, and bottom layout). An equal number of participants were assigned to each *Technique* group; that is, eight participants in each group with a gender-balanced distribution. Participants experienced all four layouts. The experiment consisted of four sessions corresponding to the four layouts.

For each session, the participants were asked to first transcribe 5 passwords as practice, then 10 passwords as formal trials for evaluation. We requested our participants to enter as fast and as accurately as possible. To minimize fatigue bias, they had 3-minute breaks between sessions but could rest longer if requested. We randomized the order of the layouts using the Latin Square design and followed the same order for the two *Technique* groups. After completing all sessions, the participants were required to join a semi-structured interview to collect their feedback and suggestions regarding (1) their preference for the four layouts; (2) the preferred switch key locations according to their experience in the experiment and daily usage habits—specifically, whether the three switch keys need to be separated in different locations, and the possible layouts after separation; and (3) improvements on the text entry approach. The experiment lasted around 30 minutes for each participant. In total, we collected (8 participants for the blinking condition + 8 participants for the dwell condition) \times 4 keyboard layouts \times 10 recorded repetitions = 640 sentences.

4.4 Results

We used SPSS 26 for data analysis. We excluded 11 sentences (out of 640 sentences or $\sim 1.72\%$) that the participants were not able to complete. Shapiro-Wilk tests and Q-Q plots indicated that only text entry speeds of both blinking and dwell groups were normally distributed ($p > .05$). We thus applied two-way mixed ANOVAs for it. For non-normal data, we applied Friedman test for *Keyboard Layout* and Mann-Whitney U test for *Technique*. Bonferroni correction was used for post-hoc pairwise comparisons. For interviews, we first transcribed the data and then applied content analysis [69].

4.4.1 Text Entry Speed and Error Rate

ANOVAs revealed significant effects of *Keyboard Layout* ($F_{3,90} = 6.762, p < .001$), but not of *Technique* ($p > .05$) on speed. Post-hoc comparisons indicated that the text entry speed of the left layout (blinking: $M = 5.49, SD = 1.21$; dwell: $M = 5.62, SD = 0.99$) was significantly faster than the other three layouts (see Figure 4a).

Figure 4b and c show the results of the Mann-Whitney U tests. They show that *Technique* had a significant effect on TER ($U = 1534.500, p = .014$). Friedman test results did not yield any significant effect of *Keyboard Layout*

on TER and NCER with blinking ($p > .05$) but showed a significant difference with dwell (TER: $\chi^2(3) = 16.198, p = .001$, NCER: $\chi^2(3) = 10.339, p = .016$). Post-hoc tests showed significant differences in two pairs in TER and only one pair in NCER.

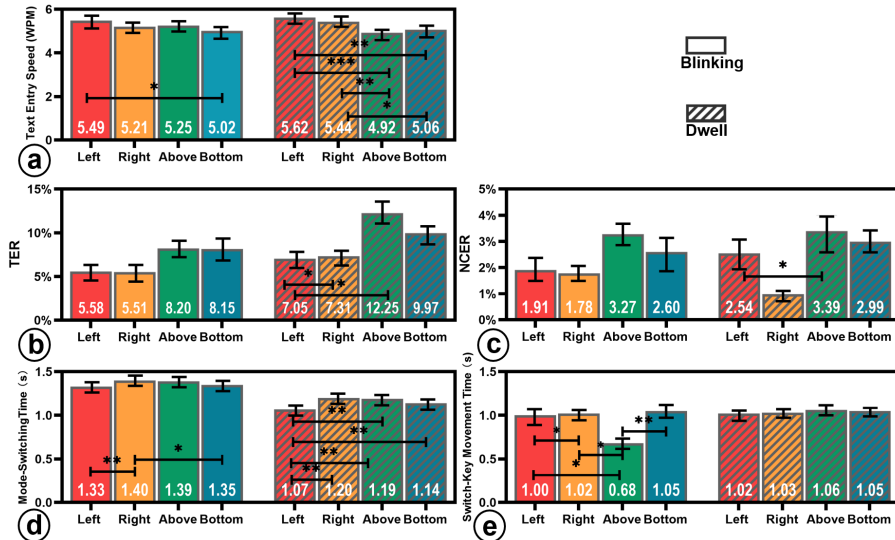


Fig. 4 (a) Mean text entry speed, (b) mean TER, (c) mean NCER, (d) mean mode switching times, and (e) mean switch-key movement times of the four keyboard layouts. Error bars represent 95% confidence intervals. ‘***’, ‘**’, and ‘*’ represent a ‘.001’, ‘.01’, and ‘.05’ significance level, respectively. The same marking scheme is used in the other figures, too.

4.4.2 Mode-Switching and Switch-Key Movement Time

Friedman tests revealed significant differences for *Keyboard Layout* on mode-switching time in each group (blinking: $\chi^2(3) = 6.879, p = .024$; dwell: $\chi^2(3) = 10.522, p = .003$) and on switch-key movement time in the blinking group ($\chi^2(3) = 8.243, p = .006$). Post-hoc tests showed some significant differences for both mode-switching and switch-key movement time, as shown in Figure 4d and e, respectively. Table 2 summarizes the significant results of the four categories of transitions (as discussed in Section 3.4) in the four layouts (left, right, above, and bottom layout) with the two techniques (blinking and dwell).

4.4.3 Interview

The participants preferred the top and left layouts the most since they would not affect their gaze on the text display area because they could still see the text area when moving up or turning left. In addition, the left layout was

Table 2 Friedman test results of mode-switching time among the four types of switching. LOW, CAP, NUM, SYM mean switch-to-lowercases, switch-to-uppercases, switch-to-numbers, and switch-to-symbols, respectively (as described in Section 3.4). p values derived from post-hoc tests are reported in these columns, ‘-’ means no significant difference.

<i>Technique</i>	<i>Keyboard Layout</i>	χ^2	p	LOW	LOW	LOW	CAP	CAP	SYM
				vs. CAP	vs. SYM	vs. NUM	vs. SYM	vs. NUM	vs. NUM
Blinking	Left	23.485	.000	.001	.000	.000	-	-	-
	Right	12.113	.005	-	-	.000	-	.006	-
	Above	18.046	.000	-	.032	.013	.000	.000	-
	Bottom	28.587	.000	.000	.000	.000	.021	-	.002
Dwell	Left	6.429	.023	.042	.017	.007	-	-	-
	Right	15.325	.001	-	.000	.012	.000	-	-
	Above	18.521	.003	.003	.047	-	.000	.000	-
	Bottom	21.548	.000	-	.000	.000	.000	.000	-

preferred because it was more aligned with the physical keyboard and the traditional keyboard used in mobile phones. They also said that character keys could still be glanced at when turning left or right, but not when moving up or down. In the interview, we used *sides* to mean the four areas around the character input area. Using one side means placing all three switch keys together in one area (e.g., all on the left as in the left layout). Using two sides means having the three keys in two areas (Figure 5). Finally, three sides involve three areas. Over half of the participants recommended designs that involve two sides ($N = 9$), followed by three sides ($N = 5$). The least preferred design involves a switch key placed on one side of the keyboard ($N = 2$). None of the two-side designs involved the top-bottom combination. These findings can be summed up into three factors influencing users’ performance and preference: (1) familiarity with typing on the QWERTY keyboard, (2) text display position (at the top-left of the keyboard) where they needed to look frequently, (3) physiological ergonomics to allow them to see the text area easily when making head movements, such as turning left and right.

All participants agreed that crossing for mode switching was efficient and easy to do. It was easy to make switches and recover from an erroneous switch (e.g., “*just make a quick pass through to the correct key*”). As recovering from an incorrect activation of delete and space keys was difficult, six participants suggested using crossing for their activation as well. However, after trying this, we found that it could complicate the text entry process and bring a high risk of false activation because switching is not suitable for objects that are close to each other like keys on a keyboard.

4.5 Discussion

4.5.1 Text Entry Speed and Error Rate

In general, the left layout led to the fastest performance (5.49WPM with blinking; 5.62WPM with dwell). As stated by the participants, one reason could be

that the position of the switch keys is similar to a standard keyboard, and easy to turn to and back from when making switches. Another reason could be that it is aligned with people’s reading habits (left-to-right and top-to-bottom).

Text entry speeds are in line with, and to a large extent better than, the only previous VR study we found that involved passwords [12]. Their participants achieved 3.82–6.57WPM but the passwords they used were simpler (between 5-10 characters), and only 50% were randomly generated (the others were more like memorable patterns).

Unlike previous research [22], we found dwell led to higher TER than blinking, particularly for the above layout. As we observed in our interview data, one possible reason is that the participants made wrong selections with dwell for the keys that are next to the switch keys when switching modes or looking at the text display area. This could be improved by adjusting the use of top space for switches.

4.5.2 Mode-Switching and Switch-Key Movement Time

From Table 2, one can see that differences are concentrated in the switch between the lowercase mode and the other modes. This finding lends strong support for the design of the shortcut to access lowercase letters quickly—as long as the users leave the area of the character keys they would switch to it (as the default mode to return to). Placing the switch keys on the left side led to the best typing performance and user preference. Except for the left layout, the other three layouts led to differences in switching time between CAP vs. SYM or CAP vs. SUM with both blinking and dwell. These results indicate a strong argument to have switch keys on the left, and the need to consider re-configuring the position of the switch key to access uppercase letters fast and conveniently.

For blinking, switch-key movement time showed significant differences among the four keyboard layouts but not for dwell (see Figure 4e). This could be explained by the location of the switch keys in relation to the character keys and text box. We observed that when more character keys are near the switch keys, and the switch keys are closer to the text box, the participants could locate the next character key faster. Thus, the switch-key movement time in the above layout for blinking was the lowest (see Figure 4e). However, for dwell, the more character keys there are near the switch keys, the more likely an unintended dwell activation can occur when the participants are searching for a character (see TER for dwell group in Figure 4b), which increased the difficulty of entering the next character after switching modes. Thus, we can observe that for the above layout, even if most character keys are near the switch keys and the switch keys are close to the text box, the switch-key movement time was not reduced.

In addition, blinking led to a slightly, non-significant better performance with lower TER than dwell. However, some participants in the blinking group commented that blinking can cause discomfort to their eyes, which was not mentioned by the participants in the dwell group. This was understandable

given that for all our participants, it was their first time typing using eye blinks and over a prolonged period. In that sense, they may consider dwell to be more comfortable. However, [24] found that, in their text entry technique for AR using eye blinks, after some practice time and familiarity, the discomfort level would drop significantly and become negligible. As such, blinking is still a possible selection mechanism to have.

4.6 User-inspired Layouts

From the pilot study, we could summarize the following factors that can affect typing performance and usability: (1) the distance between switch keys and character keys; (2) the distance between switch keys and text display area; (3) the number of character keys near the switch keys; (4) the size of switch key with crossing. Although the larger the size, the easier it is to cross, the relationship between the keys and the display area limits its size. For blinking and dwell, we do not see one technique outperforming the other.

Based on these findings, we designed three new keyboard layouts, as shown in Figure 5. The *left-right* design (Figure 5b) has two switch keys (symbols and numbers) placed on the left side while the uppercase switch key is placed on the right side. Findings from the pilot study show that the left layout led to the best performance. Given that the switch key to uppercase letters is often used, it is placed on a separate side to allow for a larger size. This left-right design meets physiological ergonomics aspects. The *left-above* layout (Figure 5c) is designed following the principle of proximity which indicates that similar or related items should be visually grouped. All three switch keys are placed close to each other on the left side. Finally, the *left-bottom* layout (Figure 5d) has the capital switch key placed on the bottom since it can provide bigger space which makes it easy to switch modes using nodding motions.

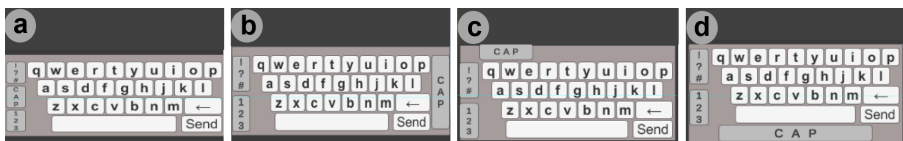


Fig. 5 The four keyboard layouts compared in the main study: the best layout from the pilot study—(a) left-only layout; and three user-inspired keyboard layouts, including (b) left-right layout, (c) left-above layout, and (d) left-bottom layout.

In the pilot study, we compared the four positions of the switch keys on text entry performance. We found that the left layout (we would call it left-only layout hereafter to make it distinguishable from the others) had the best performance and derived three user-inspired layouts. In the main study, we further evaluated these four keyboards. In addition, given that our approach led to a good performance with passwords—unordered and unfamiliar character sequences, we wanted to see how well it could perform for sentences that were more in line with what people would type daily.

5 Main Study

The goal of the experiment is to evaluate the performance of four layouts, including the best layout in the pilot study (Figure 3a) and three user-inspired ones (Figure 3b-d). To test the performance of our design under different text complexity, in addition to passwords, we also included sentences from the Brown Corpus [32], which is a collection of sentence samples of American English that include all types of characters and are more representative of daily text entry containing words, dates, that could be more easily remembered.

5.1 Participants and Apparatus

Twenty-four right-handed participants (12 males; 12 females) between the ages of 19-25 ($M = 22.4, SD = 0.85$) were recruited from the same university campus to participate in this study. We used the same apparatus as in the pilot study.

5.2 Experiment Design and Procedure

We used a mixed-subjects design with *Keyboard Layout* and *Corpus* as two within-subjects variables, and *Technique* as the between-subjects variable. That means we had two groups—blinking and dwell because results from the pilot study show that dwell and blinking have equivalent performance but with different advantages and disadvantages. For dwell and blinking, there are two variables: *Keyboard layouts* with four conditions and *Corpus* with two types (Brown Corpus [32] and 8-digit randomly generated passwords). The two corpora represent two levels of difficulty and common text entry scenarios. The Brown Corpus sentences are more representative of typing activities and contain complex sentences, which have a large number of uppercase letters, numbers, and symbols (e.g., “*The Dallas Morning News, February 17, 1961*”). As such, entering these sentences requires frequent mode switches.

Similar to the pilot study, a four-session design was arranged and participants completed text entry tasks for one layout in each session and rested for 5 minutes in between to minimize any feeling of fatigue. In each session, participants needed to complete 2 blocks for one layout. Each block had 10 sentences, five randomly selected from the Brown Corpus and five randomly generated passwords. Before the two blocks, participants were given 4 sentences (2 from the Brown Corpus and 2 passwords) for training to allow them to familiarize themselves with the devices and the techniques. Participants were encouraged to take breaks between blocks and whenever they needed a rest. The order of keyboard layouts was counterbalanced using a Latin Square design. Each session lasted about 15-20 minutes for each participant. Before starting the sessions, participants first filled in a pre-study questionnaire about their demographic information and VR and typing experience. They were then given a brief introduction about the study aims, the text entry methods, and the procedure before signing a consent form to join the experiment. At the end of

the study, we conducted a paired comparison analysis [70] and an unstructured interview. The paired comparison required participants to choose the preferred member of each pair in the six possible pairwise comparisons of the four layouts. Based on this, we calculated the rankings of the layouts. The unstructured interview aimed to collect participants' subjective feelings about the techniques and experiment. In this study, we collected (12 participants for the blinking condition + 12 participants for the dwell condition) \times 4 keyboard layouts \times 2 corpora \times 10 recorded repetitions = 1920 sentences in total.

5.3 Results

We excluded 54 trials (out of the 1920 trials or \sim 2.81%) because of incomplete completions. Shapiro-Wilk tests and Q-Q plots showed text entry speeds had a normal distribution, while TER, NCER, mode-switching time, and switch-key movement time were not normally distributed. Thus, we applied three-way mixed ANOVAs to text entry speeds for blinking and dwell groups and repeated measures ANOVAs for each group. For non-normal data, the Wilcoxon test for *Corpus* and Friedman test for *Keyboard Layout* was used for within-groups and the Mann-Whitney U test for between-groups. Post-hoc pairwise comparisons were used if significant differences were identified. We computed z-scores of participants' ranking data provided in the paired comparison.

5.3.1 Text Entry Speed

Results of RM-ANOVAs showed *Keyboard Layout* had no significant effect on text entry speed ($p > .05$ for both blinking and dwell groups) with passwords. On the other hand, with the Brown Corpus, significant effects of *Keyboard Layout* were found (blinking: $F_{3,33} = 2.575, p = 0.46$; dwell: $F_{3,33} = 5.558, p = .038$). With the Brown Corpus, the left-bottom layout achieved the fastest text entry for the dwell group (M=7.78, SD=1.94) and the left-above reached the slowest text entry speed - 7.17 WPM (SD=1.01), while for the blink group, the highest is the left-above layout (M=8.48,SD=1.85) and the slowest is the left-right layout (M=7.95,SD=1.16). Figure 6a shows a summary of the results.

For the blinking group (Figure 6a), significant differences were found between left-right and left-above ($p = .007$) and left-right and left-bottom layouts ($p = .005$). For the dwell group (Figure 6a), significant differences were found between left-only and left-bottom ($p = .029$), left-right and left-bottom ($p = .031$), and left-above and left-bottom layouts ($p = .023$). *Corpus* has led to significant differences (dwell: $F_{1,11} = 150.343, p < .001$; blinking: $F_{1,11} = 586.585, p < .001$). There was no significant interaction effect ($p > .05$).

Results of three-way ANOVAs revealed significant effects of *Corpus* ($F_{1,22} = 393.364, p < .001$) and *Technique* ($F_{1,22} = 1.311, p = .265$) on text entry speed. There was also an interaction effect between *Technique* and *Corpus* ($F_{1,22} = 10.227, p = .004$).

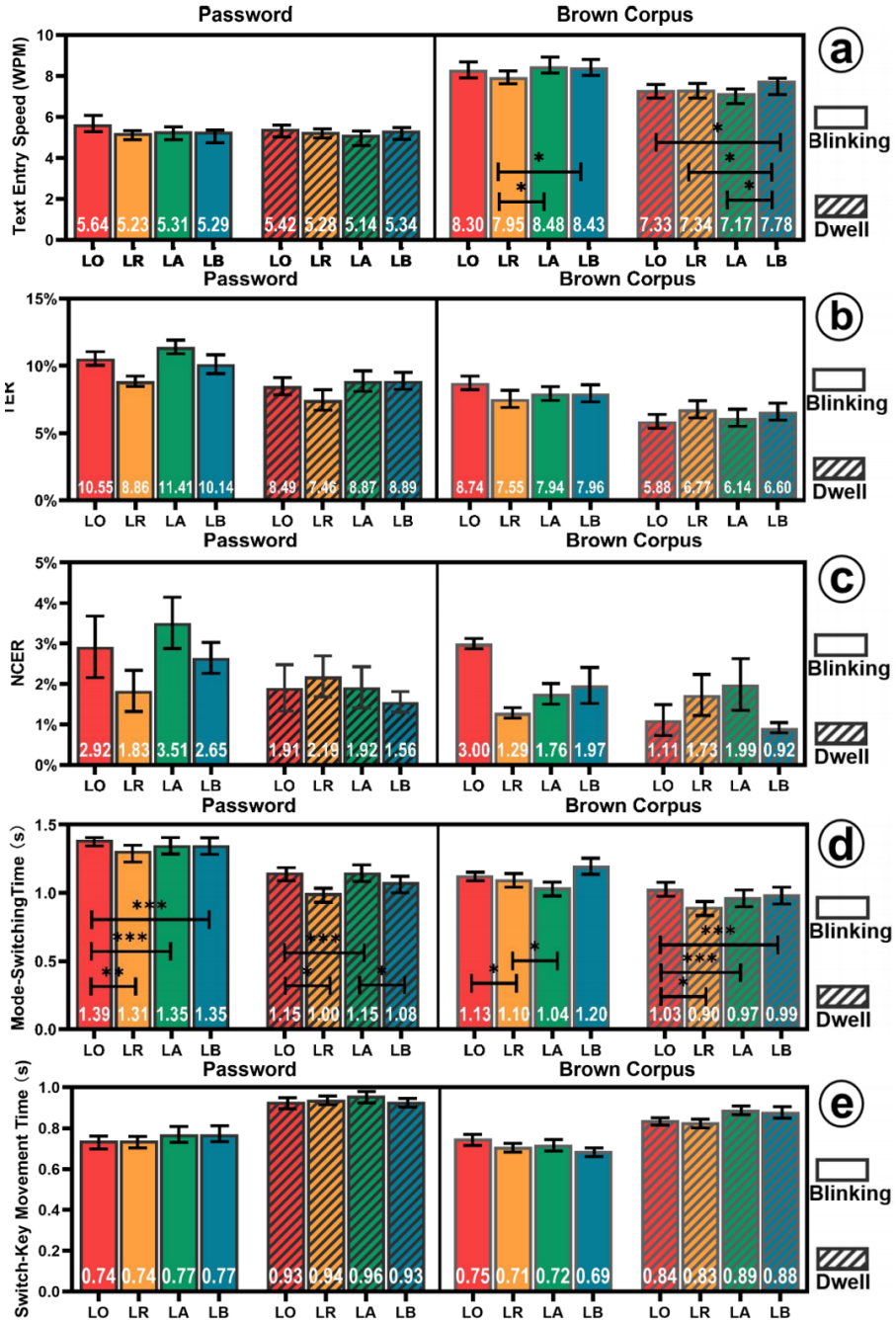


Fig. 6 (a) Mean text entry speed, (b) mean TER, (c) mean NCER, (d) mean mode-switching times, and (e) mean switch-key movement times of the four keyboard layouts. LO, LR, LA, LB are short for left-only, left-right, left-above, and left-bottom layout, respectively.

5.3.2 Error Rate

As shown in Figure 6b-c, Friedman tests indicated no significant effect of *Keyboard Layout* on TER and NCER for blinking and dwell with both corpora ($p > .05$). Wilcoxon tests showed that *Corpus* had a significant effect on TER (blinking: $z = -1.551, p = .021$; dwell: $z = -1.804, p = .015$). Only with dwell, there was a significant difference of *Corpus* on NCER ($z = -4.168, p < .001$; see Figure 6c). Mann-Whitney U tests showed a significant effect of *Technique* on TER ($U = 248.000, p = .018$) and NCER ($U = 462.500, p = .013$) between the two groups.

5.3.3 Mode-Switching Time

Friedman tests revealed a significant effect of *Keyboard Layout* on mode-switching time for both blinking and dwell with passwords (blinking: $\chi^2(3) = 12.107, p = .001$; dwell: $\chi^2(3) = 6.551, p = .044$). Post-hoc tests indicated a significant difference between left-only and left-bottom ($p < .001$), left-only and left-right ($p < .001$), and left-only and left-above layouts ($p = .009$) with blinking. For dwell, the significant differences were found in left-only and left-right ($p = .036$), and left-right and left-above layouts ($p = .026$). With the Brown Corpus, both blinking and dwell led to a significant effect of *Keyboard Layout* on mode-switching time (blinking: $\chi^2(3) = 10.402, p = .008$; dwell: $\chi^2(3) = 18.568, p = .001$). For the blinking group, there was a significant difference between left-only and left-above ($p = .001$), left-only and left-right ($p = .031$), and left-above and left-bottom layouts ($p = .042$) (Figure 6d). For the dwell group, there were three pairs having significant differences: left-only and left-bottom ($p < .001$), left-only and left-right ($p < .001$), and left-only and left-above ($p = .001$) (Figure 6d).

Wilcoxon tests indicated a significant effect of *Corpus* on mode-switching time with blinking ($z = -7.339, p < .001$) and dwell ($z = 4.992, p < .001$). Mann-Whitney U test revealed *Technique* (i.e., the between-subjects variable) significantly affects mode-switching time between the two groups ($U = 4596.000, p < .001$).

Table 3 shows the pairwise comparison results of the mode-switching time. Similar to the pilot study results (see Table 2), the significant differences are primarily concentrated in the switch between the lowercase mode and the other modes. Figure 7 shows the mean mode-switching time of the four layouts with dwell and blinking.

5.3.4 Switch-Key Movement Time

Friedman tests identified no significant differences in switch-key movement time among *Keyboard Layout* with blinking and dwell using two corpora ($p > .05$). Wilcoxon tests showed that *Corpus* significantly affected switch-key movement time (blinking: $z = -7.322, p < .001$; dwell: $z = -4.814, p < .001$).

Table 3 Friedman test results of mode-switching time among different types of switching in the main study. In the *Keyboard Layout* column, LO, LR, LA, LB are short for left-only, left-right, left-above, and left-bottom layout, respectively. LOW, CAP, NUM, SYM mean switch-to-lowercases, switch-to-upercases, switch-to-numbers, and switch-to-symbols, respectively (as described in Section 3.4). p values derived from post-hoc tests are reported in these columns, ‘-’ means no significant difference.

<i>Corpus</i> - <i>Technique</i>	<i>Keyboard</i> <i>Layout</i>	χ^2	p	LOW	LOW	LOW	CAP	CAP	SYM	
				vs. CAP	vs. SYM	vs. NUM	vs. SYM	vs. NUM	vs. NUM	
Password	LO	24.587	.000	.003	-	.000	.032	.021	.001	
	LR	131.152	.000	.000	.000	.000	.000	.000	-	
	Blinking	LA	50.152	.000	.004	.000	.000	.025	.000	-
		LB	89.562	.000	.000	.000	.000	.001	-	.001
Password	LO	62.126	.000	.007	-	.000	-	.000	.000	
	LR	114.254	.000	.000	.000	.000	.000	-	.000	
	Dwell	LA	80.526	.000	.000	.000	.000	-	.000	.001
		LB	85.325	.000	.000	.000	.000	.000	-	.000
Brown	LO	10.258	.008	.008	.014	.001	-	-	-	
	Corpus	LR	50.652	.000	.000	.000	.025	.001	-	-
		LA	21.212	.000	.000	.001	.000	-	-	-
	Blinking	LB	62.254	.000	.000	.000	-	-	-	
Brown	LO	5.667	.047	.013	.012	.037	-	-	-	
	Corpus	LR	42.125	.000	.000	.001	-	-	-	
		LA	12.457	.001	.001	.000	-	-	-	
	Dwell	LB	16.757	.001	.000	.002	.000	-	-	

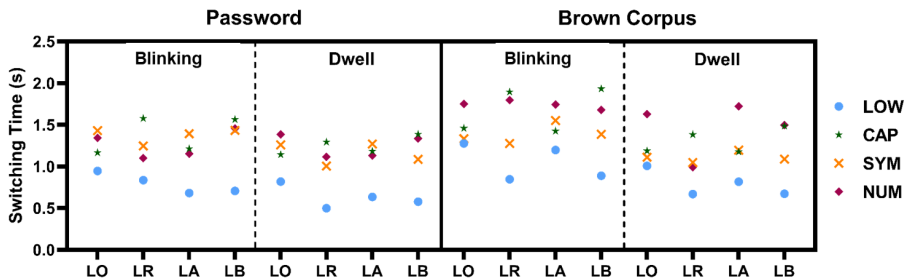


Fig. 7 Mean mode-switching time among four types of switching, with blinking and dwell and two corpora. The labels are the same as in Table 3.

Mann-Whitney U tests identified a significant difference in switch-key movement time when using blinking and dwell ($U = 5873.500, p < .001$). Figure 6e shows a summary of the results.

5.3.5 Paired Comparison of Preferred Layout

The paired comparison results were transcribed into a frequency matrix, then being normalized and evaluated [70].¹ From most preferred to least preferred,

¹The higher the scale value z is, the more favorable the compared member is.

participants' ranking preference of the layouts was left-above ($z = 0.13$), left-only ($z = 0.1$), left-right ($z = -0.11$), and left-bottom ($z = -0.12$) for the blinking group. While for the dwell group it was left-above ($z = 0.08$), left-only ($z = 0$), left-bottom ($z = -0.02$), and left-right ($z = -0.1$). As can be seen, left-above layout was the most preferred layout rated by participants regardless of the selection mechanisms.

6 Discussion

6.1 Text Entry Speed

Overall, all four layouts have led to relatively high performance, especially with the Brown Corpus sentences. With blinking, the user-inspired left-above layout achieved the best results (8.48WPM), while with dwell, the user-inspired left-bottom layout had the best performance (7.78WPM). Text entry speeds for passwords were similar to the pilot study results, which was expected. Interestingly, the left-only layout was still the best for passwords, which supports our earlier observation about the need for participants to keep looking back and forth to the text display area to check the current password. Switch keys placed as close to the text area as possible while keeping low unintended activation helped improve their performance.

Results for the Brown Corpus suggest that blinking has led to significantly faster text entry speed than dwell (at 8.48WPM with the left-above keyboard), supporting previous results from [22]. The Brown Corpus sentences, while complex, are still easier to remember and require fewer mode switches compared to randomly generated passwords. These two features allow participants to enter text quickly. Typing is slowed down when the linguistic structure of the presented text is degraded [71]. These also enlighten us that different text complexity can comprehensively reflect typing performance.

Having said this, our results are in line with a previous VR study with password entry tasks [12], where participants were required to type passwords (half-familiar simple ones and half randomly generated ones, between 5-10 characters long) and were able to achieve 3.82-6.57 WPM. As such, with our design, participants achieved a relatively fast speed for passwords that were more complex and unfamiliar to our participants. All our passwords were 8-digit strings composed of randomly generated characters following password security rules to make them complicated to guess and hack (e.g., Gy7V+KQ). Similarly, while participants' performance is lower than what has been reported in other hands-free techniques (see Table 1), the sentences that they had to enter are more difficult and complex. Before this research, to the best of our knowledge, all hands-free techniques involved lowercase letters only and used sentences from the MacKenzie phrase set [33]. Given the complexity of the Brown Corpus sentences, our approach can be considered efficient and usable for multi-type character entry that does not require additional sensing/input devices and is entirely hands-free.

6.2 Error Rate

The error rates of the three user-inspired layouts were not significantly different from the left-only layout (the best performing one in the pilot study). The Brown Corpus had lower TER with both blinking and dwell. The significant difference of *Corpus* using dwell was stronger than blinking. This is because, with dwell, participants had more errors when a text was more complex since they would pause when they needed to think about the next action. For NCER, only dwell showed differences between the two corpora, which shows that text complexity does not restrict the participants' willingness to correct errors with blinking but not with dwell.

On the whole, the error rate is acceptable and relatively low ([12] has $\sim 3.5\%$) since the mean error rate of 1-3% for 8-digital passwords means that there is only 1 uncorrected character among 4-12 password phrase attempts (usually, 5 attempts are allowed in commercial applications). In addition, the passwords we set were quite complex and difficult to remember, and participants were unfamiliar with them. The error rate should be lower when they enter their own familiar or simpler passwords.

6.3 Mode-Switching and Switch-Key Movement Time

Our results show that the user-inspired layouts were better at reducing mode-switching time. The three layouts arranging the switch keys on two sides with larger key sizes allowed for bigger areas available for crossing.

Dwell showed better results than blinking in mode-switching time in two corpora. This finding is the opposite of the results of text entry speed. The mode-switching time with dwell was significantly shorter than blinking, even though crossing was used for both. Because dwell and crossing only require head movements without additional user actions, the two can supplement each other well. However, blinking needs an extra conscious effort (eye blinks), and users need to rotate trigger actions frequently when typing and do mode switches, which can increase users' workload. This explains the reason for the shorter switch-key movement time with dwell.

Comparing the mode-switching time of four layouts in Table 3 and Figure 7, the three optimized layouts significantly reduced the time of switching to lowercase characters (i.e., LOW). The left-right layout reduced the switching time to lowercase letters the most, but also resulted in a significant increase in the switching time to uppercase letters. On the other hand, the left-above layout reduced the switching time to lowercase letters while ensuring that the performance to switch to uppercase letters was not compromised. The reason behind this is that with the left-right layout, the uppercase switch is on the right and far from where the text is displayed, leaving the upper part of the input area available for switching to lowercase letters.

In short, the results show that the left-above layout performed better at mode switching and supported the entry of Brown Corpus-like sentences well. In general, compared to dwell, blinking seems more suitable as a hands-free

approach as it leads to good performance with lower errors and lowers the use of head motions. It is also easier to correct mistakes with blinking. While there is a factor of eye fatigue, results from [24] show that as users become familiar with it, eye strains become negligible.

7 Summary of Main Findings and Lessons Derived from the Experiments

Based on the results, we can make the following four key lessons for hands-free multi-character text entry in VR:

1. Crossing activation is suitable for hands-free mode switching and can complement other hands-free selection mechanisms;
2. Eye blinking is a suitable hands-free selection mechanism for multi-type character text entry;
3. As multi-type character text entry typically involves more lowercase letters, a feature allowing for quick access to them is helpful, like in our case of having them as default; and
4. The location of the key switches is important. The left side of the keyboard and closer to the text display area are preferable.

8 Limitations and Future Work

This study has some limitations, which can serve as possible directions for future work. First, our results show that the distance between the switch keys and the text input box and the size of the key affects text entry performance. We did not include these variables in our experiment. These factors can be explored in greater detail in the future. Second, we used two corpora and while they cover various levels of complexity, we have not considered other possible scenarios (e.g., capital letters only for some words or emotion icons). Future work can explore other cases where mode switching can be helpful and necessary for typing tasks. In addition, our work is a first and provides a solid starting point of multi-type character and could extend to other populations (e.g., impaired and elderly users) as part of our future work. Third, we pre-tested and used a 300ms time threshold for both key-selection mechanisms across the studies. A different time threshold could lead to varied results, particularly for objective measurements. It was not the focus of this study but we would like to evaluate it in future work. Finally, our focus is on hands-free approaches given their benefits presented in the introduction (see Section 1). However, because there is also limited research on hand-supported multi-type character entry in VR, it will be worthy of exploring hand-based techniques and approaches, which could open further possibilities to make text entry in VR HMDs more aligned with other types of interactive systems like smartphones and desktop/laptop computers.

9 Conclusion

Multi-type characters, i.e., the combinations of uppercase and lowercase letters, symbols, and numbers, are indispensable for daily text entry activities. This paper presented a first exploration of multi-type character text entry with a virtual keyboard in virtual reality (VR) that is entirely hands-free. We combined a crossing-based mode-switching mechanism with two hands-free selection mechanisms (eye blinks and dwell) and integrated them into iteratively designed keyboard layouts. Two experiments were run to examine the performance of several keyboard layouts, especially the switch keys' locations, using complex 8-digit passwords and sentences from Brown Corpus which include uppercase and lowercase letters, symbols, and numbers, and are more representative of sentences people type. Results show that our combination of crossing-based and selection mechanisms and proposed keyboard layouts represent efficient, accurate, and usable text entry approaches for multi-type character entry in VR and serve as the foundation for further in this area.

Acknowledgement(s)

The authors want thank the participants who volunteered their time to do the experiments. We also thank the reviewers for the insightful comments and detailed suggestions that have us improve our paper.

Ethics Declaration

The study is categorized as Low-Risk Research (LRR), conducted according to the guidelines regulating LRR experiments, and approved by the University Ethics Committee at Xi'an Jiaotong- Liverpool University.

Disclosure statement

The authors declare that there are no conflicts of interest.

Availability of data and materials

Data associated with this research can be made available upon reasonable request to the corresponding author.

Author Contributions

All authors contributed to this work.

Funding

This research was funded in part by Xi'an Jiaotong-Liverpool University (XJTLU) Key Program Special Fund (#KSF-A-03), the National Science

Foundation of China (#62272396; #62207022), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (#22KJB520038), and XJTU Research Development Fund (#RDF-17-01-54).

References

- [1] Dube, T.J., Arif, A.S.: Text entry in virtual reality: A comprehensive review of the literature. In: Kurosu, M. (ed.) *Human-Computer Interaction. Recognition and Interaction Technologies*, pp. 419–437. Springer, Cham (2019)
- [2] Grubert, J., Ofek, E., Pahud, M., Kristensson, P.O.: The office of the future: Virtual, portable, and global. *IEEE Computer Graphics and Applications* **38**(6), 125–133 (2018). <https://doi.org/10.1109/MCG.2018.2875609>
- [3] Wiederhold, B.K., Riva, G.: Virtual reality therapy: Emerging topics and future challenges. *Cyberpsychology, Behavior, and Social Networking* **22**(1), 3–6 (2019). <https://doi.org/10.1089/cyber.2018.29136.bkw>. PMID: 30649958
- [4] Serrano, B., Botella, C., Wiederhold, B.K., Baños, R.M.: In: Rizzo, A.S., Bouchard, S. (eds.) *Virtual Reality and Anxiety Disorders Treatment: Evolution and Future Perspectives*, pp. 47–84. Springer, New York, NY (2019). https://doi.org/10.1007/978-1-4939-9482-3_3
- [5] Biener, V., Gesslein, T., Schneider, D., Kawala, F., Otte, A., Kristensson, P.O., Pahud, M., Ofek, E., Campos, C., Kljun, M., *et al.*: Povrpoint: Authoring presentations in mobile virtual reality. *IEEE Transactions on Visualization and Computer Graphics* **28**(5), 2069–2079 (2022). <https://doi.org/10.1109/TVCG.2022.3150474>
- [6] Dresner, E., Herring, S.C.: Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force. *Communication Theory* **20**(3), 249–268 (2010). <https://doi.org/10.1111/j.1468-2885.2010.01362.x>
- [7] Ma, J., Yang, W., Luo, M., Li, N.: A study of probabilistic password models. In: *2014 IEEE Symposium on Security and Privacy*, pp. 689–704 (2014). <https://doi.org/10.1109/SP.2014.50>
- [8] George, C., Khamis, M., von Zeszschwitz, E., Burger, M., Schmidt, H., Alt, F., Hussmann, H.: Seamless and secure vr : Adapting and evaluating established authentication systems for virtual reality. In: *Network and Distributed System Security Symposium (NDSS 2017)* (2017). <https://doi.org/10.14722/usec.2017.23028>
- [9] Olade, I., Liang, H.-N., Fleming, C.: A review of multimodal facial

- biometric authentication methods in mobile devices and their application in head mounted displays. In: 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), pp. 1997–2004 (2018). <https://doi.org/10.1109/SmartWorld.2018.00334>
- [10] Luo, S., Nguyen, A., Song, C., Lin, F., Xu, W., Yan, Z.: Oculock: Exploring human visual system for authentication in virtual reality head-mounted display. In: 2020 Network and Distributed System Security Symposium (NDSS) (2020)
- [11] Barkadehi, M.H., Nilashi, M., Ibrahim, O., Zakeri Fardi, A., Samad, S.: Authentication systems: A literature review and classification. *Telematics and Informatics* **35**(5), 1491–1511 (2018). <https://doi.org/10.1016/j.tele.2018.03.018>
- [12] Schneider, D., Otte, A., Gesslein, T., Gagel, P., Kuth, B., Damlakhi, M.S., Dietz, O., Ofek, E., Pahud, M., Kristensson, P.O., Müller, J., Grubert, J.: Reconfiguration: Reconfiguring physical keyboards in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* **25**(11), 3190–3201 (2019). <https://doi.org/10.1109/TVCG.2019.2932239>
- [13] Jiang, H., Weng, D.: 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 (2020). <https://doi.org/10.1109/VR46266.2020.00092>
- [14] Boletsis, C., Kongsvik, S.: Controller-based text-input techniques for virtual reality: An empirical comparison. *International Journal of Virtual Reality* **19**(3), 2–15 (2019). <https://doi.org/10.20870/IJVR.2019.19.3.2917>
- [15] Chen, S., Wang, J., Guerra, S., Mittal, N., Prakkamakul, S.: Exploring word-gesture text entry techniques in virtual reality. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. CHI EA '19, pp. 1–6. Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3290607.3312762>
- [16] Speicher, M., Feit, A.M., Ziegler, P., Krüger, A.: Selection-based text entry in virtual reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, pp. 1–13. Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3173574.3174221>

- [17] Yu, D., Fan, K., Zhang, H., Monteiro, D., Xu, W., Liang, H.-N.: Pizza-text: Text entry for virtual reality systems using dual thumbsticks. *IEEE Transactions on Visualization and Computer Graphics* **24**(11), 2927–2935 (2018). <https://doi.org/10.1109/TVCG.2018.2868581>
- [18] Xu, W., Liang, H.-N., Zhao, Y., Zhang, T., Yu, D., Monteiro, D., Yue, Y.: Ringtext: Dwell-free and hands-free text entry for mobile head-mounted displays using head motions. *IEEE Transactions on Visualization and Computer Graphics* **25**(5), 1991–2001 (2019). <https://doi.org/10.1109/TVCG.2019.2898736>
- [19] Meng, X., Xu, W., Liang, H.-N.: An exploration of hands-free text selection for virtual reality head-mounted displays. In: 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 74–81 (2022). <https://doi.org/10.1109/ISMAR55827.2022.00021>
- [20] Li, Z., Qin, Z., Luo, Y., Pan, Y., Liang, H.-N.: Exploring the design space for hands-free robot dog interaction via augmented reality. In: 2023 9th International Conference on Virtual Reality (ICVR), pp. 288–295 (2023)
- [21] Yan, Y., Yu, C., Yi, X., Shi, Y.: Headgesture: Hands-free input approach leveraging head movements for hmd devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2**(4) (2018). <https://doi.org/10.1145/3287076>
- [22] Lu, X., Yu, D., Liang, H.-N., Xu, W., Chen, Y., Li, X., Hasan, K.: Exploration of hands-free text entry techniques for virtual reality. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 344–349 (2020). <https://doi.org/10.1109/ISMAR50242.2020.00061>
- [23] Ma, X., Yao, Z., Wang, Y., Pei, W., Chen, H.: Combining brain-computer interface and eye tracking for high-speed text entry in virtual reality. In: 23rd International Conference on Intelligent User Interfaces. IUI '18, pp. 263–267. Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3172944.3172988>
- [24] Lu, X., Yu, D., Liang, H.-N., Goncalves, J.: Itext: Hands-free text entry on an imaginary keyboard for augmented reality systems. In: The 34th Annual ACM Symposium on User Interface Software and Technology. UIST '21, pp. 815–825. Association for Computing Machinery, New York, NY, USA (2021). <https://doi.org/10.1145/3472749.3474788>
- [25] Yu, C., Gu, Y., Yang, Z., Yi, X., Luo, H., Shi, Y.: Tap, dwell or gesture? exploring head-based text entry techniques for hmds. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, pp. 4479–4488. Association for Computing Machinery, New York, NY, USA (2017). <https://doi.org/10.1145/3025453.3025964>

- [26] Rajanna, V., Hansen, J.P.: 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. ETRA '18. Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3204493.3204541>
- [27] Dube, T.J., Arif, A.S.: Impact of key shape and dimension on text entry in virtual reality. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. CHI EA '20, pp. 1–10. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3334480.3382882>
- [28] Accot, J., Zhai, S.: More than dotting the i's — foundations for crossing-based interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '02, pp. 73–80. Association for Computing Machinery, New York, NY, USA (2002). <https://doi.org/10.1145/503376.503390>
- [29] Tu, H., Huang, S., Yuan, J., Ren, X., Tian, F.: Crossing-based selection with virtual reality head-mounted displays. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, pp. 1–14. Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3290605.3300848>
- [30] Pavlovych, A., Stuerzlinger, W.: The tradeoff between spatial jitter and latency in pointing tasks. In: Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems. EICS '09, pp. 187–196. Association for Computing Machinery, New York, NY, USA (2009). <https://doi.org/10.1145/1570433.1570469>
- [31] Cockburn, A., Firth, A.: Improving the acquisition of small targets. In: O'Neill, E., Palanque, P., Johnson, P. (eds.) *People and Computers XVII — Designing for Society*, pp. 181–196. Springer, London (2004)
- [32] Francis, W.N., Kucera, H.: Brown corpus manual. *Letters to the Editor* 5(2), 7 (1979)
- [33] MacKenzie, I.S., Soukoreff, R.W.: Phrase sets for evaluating text entry techniques. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems. CHI EA '03, pp. 754–755. Association for Computing Machinery, New York, NY, USA (2003). <https://doi.org/10.1145/765891.765971>
- [34] Noyes, J.: The qwerty keyboard: a review. *International Journal of Man-Machine Studies* 18(3), 265–281 (1983). [https://doi.org/10.1016/S0020-7373\(83\)80010-8](https://doi.org/10.1016/S0020-7373(83)80010-8)

- [35] Li, Y., Sarcar, S., Zheng, Y., Ren, X.: Exploring text revision with backspace and caret in virtual reality. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21. Association for Computing Machinery, New York, NY, USA (2021). <https://doi.org/10.1145/3411764.3445474>
- [36] Bi, X., Smith, B.A., Zhai, S.: Quasi-qwerty soft keyboard optimization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '10, pp. 283–286. Association for Computing Machinery, New York, NY, USA (2010). <https://doi.org/10.1145/1753326.1753367>
- [37] Lee, L.H., Braud, T., Lam, K.Y., Yau, Y.P., Hui, P.: From seen to unseen: Designing keyboard-less interfaces for text entry on the constrained screen real estate of augmented reality headsets. *Pervasive and Mobile Computing* **64**, 101148 (2020). <https://doi.org/10.1016/j.pmcj.2020.101148>
- [38] Knierim, P., Schwind, V., Feit, A.M., Nieuwenhuizen, F., Henze, N.: 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. CHI '18, pp. 1–9. Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3173574.3173919>
- [39] Pham, D.-M., Stuerzlinger, W.: Hawkey: Efficient and versatile text entry for virtual reality. In: 25th ACM Symposium on Virtual Reality Software and Technology. VRST '19. Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3359996.3364265>
- [40] Otte, A., Menzner, T., Gesslein, T., Gagel, P., Schneider, D., Grubert, J.: Towards utilizing touch-sensitive physical keyboards for text entry in virtual reality. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 1729–1732 (2019). <https://doi.org/10.1109/VR.2019.8797740>
- [41] Grubert, J., Witzani, L., Ofek, E., Pahud, M., Kranz, M., Kristensson, P.O.: Effects of hand representations for typing in virtual reality. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 151–158 (2018). <https://doi.org/10.1109/VR.2018.8446250>
- [42] Boletsis, C., Kongsvik, S.: Text input in virtual reality: A preliminary evaluation of the drum-like vr keyboard. *Technologies* **7**(2) (2019). <https://doi.org/10.3390/technologies7020031>
- [43] Ciobanu, O., Gavati, C., Cozmei, R.: The keyboard remains the least ergonomically designed computer device. In: 2015 E-Health and Bioengineering Conference (EHB), pp. 1–4 (2015). <https://doi.org/10.1109/EHB>

2015.7391585

- [44] Min, K.: Text input tool for immersive vr based on 3×3 screen cells. In: Proceedings of the 5th International Conference on Convergence and Hybrid Information Technology. ICHIT'11, pp. 778–786. Springer, Berlin, Heidelberg (2011)
- [45] Prätorius, M., Burgbacher, U., Valkov, D., Hinrichs, K.: Sensing thumb-to-finger taps for symbolic input in vr/ar environments. *IEEE Computer Graphics and Applications* **35**(5), 42–54 (2015). <https://doi.org/10.1109/MCG.2015.106>
- [46] Ogitani, T., Arahori, Y., Shinyama, Y., Gondow, K.: Space saving text input method for head mounted display with virtual 12-key keyboard. In: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), pp. 342–349 (2018). <https://doi.org/10.1109/AINA.2018.00059>
- [47] Yanagihara, N., Shizuki, B.: Cubic keyboard for virtual reality. In: Proceedings of the Symposium on Spatial User Interaction. SUI '18, p. 170. Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3267782.3274687>
- [48] Monteiro, P., Gonçalves, G., Coelho, H., Melo, M., Bessa, M.: Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics* **27**(5), 2702–2713 (2021). <https://doi.org/10.1109/TVCG.2021.3067687>
- [49] Chabot, S., Drozdal, J., Zhou, Y., Su, H., Braasch, J.: Language learning in a cognitive and immersive environment using contextualized panoramic imagery. In: International Conference on Human-Computer Interaction, pp. 202–209 (2019). Springer
- [50] Qian, Y.Y., Teather, R.J.: The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality. In: Proceedings of the 5th Symposium on Spatial User Interaction. SUI '17, pp. 91–98. Association for Computing Machinery, New York, NY, USA (2017). <https://doi.org/10.1145/3131277.3132182>
- [51] Minakata, K., Hansen, J.P., MacKenzie, I.S., Bækgaard, P., Rajanna, V.: Pointing by gaze, head, and foot in a head-mounted display. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. ETRA '19. Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3317956.3318150>
- [52] Mardanbegi, D., Pfeiffer, T.: Eyemrkt: A toolkit for developing eye gaze interactive applications in virtual and augmented reality. In: Proceedings

- of the 11th ACM Symposium on Eye Tracking Research & Applications. ETRA '19. Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3317956.3318155>
- [53] Tu, H., Huang, J., Liang, H.-N., Skarbez, R., Tian, F., Duh, H.B.-L.: Distractor effects on crossing-based interaction. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21. Association for Computing Machinery, New York, NY, USA (2021). <https://doi.org/10.1145/3411764.3445340>
- [54] Kurauchi, A., Feng, W., Joshi, A., Morimoto, C., Betke, M.: Eyeswipe: Dwell-free text entry using gaze paths. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, pp. 1952–1956. Association for Computing Machinery, New York, NY, USA (2016). <https://doi.org/10.1145/2858036.2858335>
- [55] Grubert, J., Witzani, L., Ofek, E., Pahud, M., Kranz, M., Kristensson, P.O.: Text entry in immersive head-mounted display-based virtual reality using standard keyboards. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 159–166 (2018). <https://doi.org/10.1109/VR.2018.8446059>
- [56] Herley, C., Van Oorschot, P.: A research agenda acknowledging the persistence of passwords. *IEEE Security Privacy* **10**(1), 28–36 (2012). <https://doi.org/10.1109/MSP.2011.150>
- [57] Payton, L.: Memory for passwords: The effects of varying number, type, and composition. *Psi Chi Journal of Undergraduate Research* **15**(4) (2010)
- [58] Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F.: Encountering stronger password requirements: User attitudes and behaviors. In: Proceedings of the Sixth Symposium on Usable Privacy and Security. SOUPS '10. Association for Computing Machinery, New York, NY, USA (2010). <https://doi.org/10.1145/1837110.1837113>
- [59] Proctor, R.W., Lien, M.-C., Vu, K.-P.L., Schultz, E.E., Salvendy, G.: Improving computer security for authentication of users: Influence of proactive password restrictions. *Behavior Research Methods, Instruments, & Computers* **34**(2), 163–169 (2002)
- [60] Garrison, A., Remley, D., Thomas, P., Wierszewski, E.: Conventional faces: Emoticons in instant messaging discourse. *Computers and Composition* **28**(2), 112–125 (2011)
- [61] Park, J., Baek, Y.M., Cha, M.: Cross-cultural comparison of nonverbal

- cues in emoticons on twitter: Evidence from big data analysis. *Journal of Communication* **64**(2), 333–354 (2014)
- [62] Majaranta, P., Ahola, U.-K., Špakov, O.: Fast Gaze Typing with an Adjustable Dwell Time, pp. 357–360. Association for Computing Machinery, New York, NY, USA (2009). <https://doi.org/10.1145/1518701.1518758>
- [63] Lu, X., Yu, D., Liang, H.-N., Feng, X., Xu, W.: Depthtext: Leveraging head movements towards the depth dimension for hands-free text entry in mobile virtual reality systems. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 1060–1061 (2019). <https://doi.org/10.1109/VR.2019.8797901>. IEEE
- [64] Królak, A., Strumiłło, P.: Eye-blink detection system for human–computer interaction. *Universal Access in the Information Society* **11**, 409–419 (2012). <https://doi.org/10.1007/s10209-011-0256-6>
- [65] Yamada, H.: A Historical Study of Typewriters and Typing Methods: from the Position of Planning Japanese Parallels. VII vol. 13, pp. 1547–1556 (1981)
- [66] Soukoreff, R.W., MacKenzie, I.S.: Metrics for text entry research: An evaluation of msd and kspc, and a new unified error metric. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '03, pp. 113–120. Association for Computing Machinery, New York, NY, USA (2003). <https://doi.org/10.1145/642611.642632>
- [67] Soukoreff, R.W., MacKenzie, I.S.: Measuring errors in text entry tasks: An application of the levenshtein string distance statistic. In: CHI '01 Extended Abstracts on Human Factors in Computing Systems. CHI EA '01, pp. 319–320. Association for Computing Machinery, New York, NY, USA (2001). <https://doi.org/10.1145/634067.634256>
- [68] Shay, R., Komanduri, S., Durity, A.L., Huh, P.S., Mazurek, M.L., Segreti, S.M., Ur, B., Bauer, L., Christin, N., Cranor, L.F.: Designing password policies for strength and usability. *ACM Trans. Inf. Syst. Secur.* **18**(4) (2016). <https://doi.org/10.1145/2891411>
- [69] Stemler, S.: An overview of content analysis. *Practical assessment, research, and evaluation* **7**(1), 17 (2000)
- [70] Cattelan, M.: Models for Paired Comparison Data: A Review with Emphasis on Dependent Data. *Statistical Science* **27**(3), 412–433 (2012). <https://doi.org/10.1214/12-STS396>

- [71] Salthouse, T.A.: Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological bulletin* **99**(3), 303 (1986)