Evaluating the Performance of Hand-Based Probabilistic Text Input Methods on a Mid-Air Virtual Qwerty Keyboard

John J. Dudley, Jingyao Zheng, Aakar Gupta, Hrvoje Benko,
Matt Longest, Robert Wang, and Per Ola Kristensson

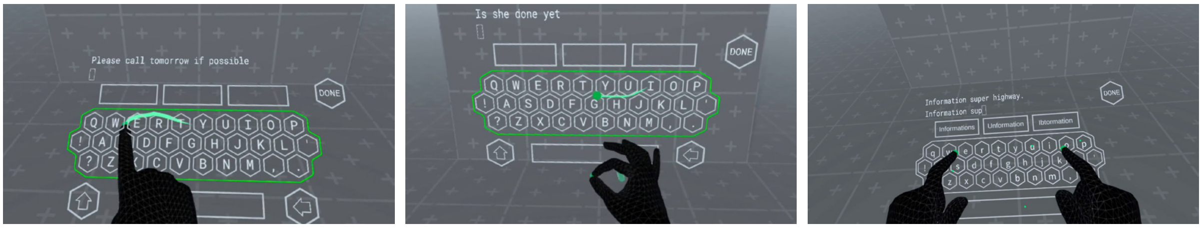


Fig. 1: User's view while typing on the mid-air Qwerty keyboard in VR. Left) shows the *Poke* interaction allowing the user to perform gestures with the tip of their index finger. Middle) shows the *Raycast* interaction allowing the user to perform gestures with a cursor projected from the hand. Right) shows standard touch typing with two index fingers.

Abstract—Integrated hand-tracking on modern virtual reality (VR) headsets can be readily exploited to deliver mid-air virtual input surfaces for text entry. These virtual input surfaces can closely replicate the experience of typing on a Qwerty keyboard on a physical touchscreen, thereby allowing users to leverage their pre-existing typing skills. However, the lack of passive haptic feedback, unconstrained user motion, and potential tracking inaccuracies or observability issues encountered in this interaction setting typically degrades the accuracy of user articulations. We present a comprehensive exploration of error-tolerant probabilistic hand-based input methods to support effective text input on a mid-air virtual Qwerty keyboard. Over three user studies we examine the performance potential of hand-based text input under both gesture and touch typing paradigms. We demonstrate typical entry rates in the range of 20 to 30 wpm and average peak entry rates of 40 to 45 wpm.

Index Terms—Virtual reality, text entry.

1 Introduction

Users are widely accustomed to writing directly with their fingers on both physical keyboards and touchscreens. This type of input is still primarily performed using the Qwerty layout, or similar variants in other languages. Recent advances in the integrated hand tracking capability of modern virtual reality (VR) headsets enable closer replication of the experience of entering text using a physical keyboard layout. In this paper, we evaluate the performance potential and interaction experience of a mid-air Qwerty virtual keyboard supporting direct hand and finger-based input.

The investigation of hand-based input is particularly relevant since a key advantage of delivering an interaction experience that is consistent with users' physical experience of typing on a physical touch surface or keyboard is that it allows users to leverage their already established motor skills and familiarity with the layout. This established familiarity can lessen the time required to achieve acceptable levels of performance in the novel setting. Furthermore, supporting hand-based input represents an essential fallback method when controllers, other input devices, or alternative modalities such as voice are unavailable or inappropriate.

There are two widely supported text input methods available on modern touchscreen devices: *touch typing*, typically supported with auto-correct [17,31,51], and *gesture typing* [30,59]. When touch typing, the user performs discrete presses for each individual character. When gesture typing, the user performs a word-gesture by drawing between each character in the word. In recognition of distinct user preferences, a comprehensive text input system will support flexible use of either method. Both of these input methods can be adapted for use in VR by employing a mid-air virtual input surface.

Practically, however, typing in mid-air on a virtual keyboard introduces new challenges, such as the lack of passive haptic feedback and the inability to rest part of the hand or arm on a physical surface. Mid-air interactions are generally associated with less precise movements [4] compared to similar interactions performed on a physical surface. For this reason, articulations performed by users may be less precise, leading to greater ambiguity in discrete touches and articulated word-gestures. These imprecise movements are further exacerbated by additional noise introduced by tracking errors and/or latency. This presents a technical challenge for the design of an easy-to-use mid-air keyboard. Fortunately this degradation in user input can be mitigated by applying probabilistic input decoding strategies to correctly infer user intent. Text entry methods providing a degree of tolerance to input noise have proven effective in mid-air settings [11]. Gesture typing also provides an inherent degree of tolerance to input noise and so is well-suited to this challenging interaction setting. However, prior work examining mid-air gesture typing in VR has, to date, exclusively explored either projected cursor based interactions or use of controllers. Much of this prior work has also obtained relatively low entry rates (under 20 wpm) for gesture typing in a mid-air setting (see Section 2 for a more in-depth discussion). Our paper seeks to fill this gap in the prior work by evaluating modern hand-based interaction conventions.

This paper examines the performance potential of a mid-air virtual Qwerty keyboard driven by state-of-the-art commodity integrated hand tracking. To provide a platform for this examination, we developed a probabilistic keyboard system for deployment on the Meta Quest to leverage its natively supported articulated hand tracking [21]. The use of integrated hand tracking represents a major point of distinction from most prior work where either tracked controllers or precision marker-

- *John J. Dudley, Jingyao Zheng and Per Ola Kristensson are with the University of Cambridge.*
- *Aakar Gupta, Hrvoje Benko, Matt Longest and Robert Wang are with Reality Labs Research, Meta Inc.*

Manuscript received 25 March 2023; revised 17 June 2023; accepted 7 July 2023. Date of current version 31 October 2023.
This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2023.3320238, provided by the authors.
Digital Object Identifier no. 10.1109/TVCG.2023.3320238

based tracking systems were used. This clear point of distinction is elaborated on in the related work section. Furthermore, our application was developed to run entirely on-device as opposed to relaying to a high-performance computer external to the headset. This sits in contrast to prior work [11, 23, 36, 58] that leverages decoding or recognition systems running on a separate dedicated machine rather than deployed on the typically more processor-constrained headset. Both of these factors serve to provide a much more accurate picture of the true performance potential of hand-based text input on a mid-air Qwerty keyboard given current hardware. We present three user studies evaluating this performance potential using this system with participants representing various levels of VR experience and typing expertise.

Our focus for Study 1 and 2 was on gesture typing as the less-well explored of the two main input methods. In Study 1 we sought to determine what was the most efficient and preferred hand-based input technique for mid-air gesture typing. We compared user performance when gesture typing under two alternative hand-driven keyboard interaction modes: *Poke* and *Raycast*. These two interaction modes were directly derived from established developer guidance[1] for hand-based interaction design in VR. In the *Poke* interaction mode, the user draws with the tip of the index finger by *poking* it within the keyboard plane. Gesture start and stop are indicated by entering and leaving the keyboard plane respectively. In the *Raycast* interaction mode, the user types by controlling a *ray* emanating from a frame attached to their hand. Gesture start and stop is indicated by performing a thumb-index finger pinch. This first study revealed a significant performance advantage to gesture typing with the Poke interaction mode. We also observed a broad spread in entry rates with performance heavily influenced by prior experience with gesture typing.

The significant performance advantage associated with prior experience with gesture typing motivated Study 2 in which we sought to determine the upper-limit performance potential of mid-air gesture typing using the Poke interaction mode. The goal in Study 2 was to assess potential performance by mitigating the effects of learning an unfamiliar interaction technique in an unfamiliar interaction setting. To do this we applied a motor performance saturation protocol based on the approach used by Kristensson and Zhai [30]. This protocol served to short-circuit motor learning so that participants' performance theoretically approached the performance achievable after extensive practice. Our study with 14 participants revealed a rich distribution in the performance potential reached by different participants. At the upper-end of performance we observed one participant (*P2*), who was both very experienced with VR and used a gesture keyboard as their default text entry method on their smartphone, achieve peak entry rates in the range of 48 to 66 wpm. Promisingly too, we saw two participants at the other end of the spectrum, *P3* and *P4*, who were very inexperienced with VR but never used a gesture keyboard achieve peak entry rates in the range of 32 to 63 wpm and 40 to 56 wpm respectively.

Finally, Study 3 compared Poke-based gesture typing against Poke-based touch typing with two index fingers supported by an auto-correct decoder. This study applied a conventional text entry protocol in which participants entered eight blocks of 20 phrases within each condition. We observed that participants' mean entry rates were significantly higher in the touch typing condition but overall preference was equally distributed between the two techniques. Mobile devices support both touch and gesture typing methods to accommodate personal preferences, and our paper provides direct evidence that there is potential value in doing the same on VR devices.

Ultimately, an effective text entry system for use in mixed reality (MR) will support an array of different interaction modes and input methods enabling the user to choose the best option for their current context. The keyboard evaluated in this paper delivers significant flexibility thanks to the use of integrated hand tracking and the robust probabilistic input decoding. At the end of this paper, we outline several core use-cases for text entry in MR and illustrate the developed keyboard's ability to accommodate these use cases.

[1]https://developer.oculus.com/resources/hands-design-interactions/

In summary, this paper offers the following key contributions:

1. We present a probabilistic mid-air virtual Qwerty keyboard system that is deployable on-device and exploits the integrated hand-tracking offered by modern VR HMDs. This keyboard provides an ideal platform for an accurate and realistic evaluation of the performance potential of hand-based text input methods given current device capabilities.

2. Study 1 ($n = 16$) finds that direct articulation with the index finger (mean entry rate of 17.7 wpm) significantly outperforms projected cursor interaction (mean entry rate of 11.9 wpm) by 48.7% for mid-air gesture typing.

3. Study 2 ($n = 14$) demonstrates the performance potential of mid-air gesture typing in VR—select participants already familiar with VR and gesture typing reached peak entry rates of 29 to 66 wpm with no errors.

4. Study 3 ($n = 16$) finds that two-finger touch typing outperforms gesture typing but that both input methods can deliver effective entry rates: a mean entry rate of 21.5 wpm for gesture typing and 25.6 wpm for touch typing.

2 Related Work

The broader principles for text entry methods on modern AR and VR HMDs are still coalescing. Dube and Arif [7] review the various techniques that have been examined as of 2019. One solution is to just use a physical keyboard while embedded in VR [18, 52], however, this approach is not well suited to on-the-go use cases or when locating the keyboard would be disruptive to the current immersive experience. Another solution is to support voice-to-text [1], but again this approach may be inappropriate at certain times due to ambient conditions or for privacy reasons. To support comprehensive text input functionality on an HMD it is therefore necessary to provide an input method that can complement and bridge between intermittent use of a physical keyboard or voice. A wide range of strategies have been explored to bridge this gap and it is not possible to cover all related work in this section. We constrain our scope to investigations of mid-air text input given the focus of our paper. The following subsections split the related work according to their focus on either mid-air gesture typing or mid-air touch typing.

2.1 Mid-Air Gesture Typing

The word-gesture keyboard [24, 30, 59] has emerged as one of the dominant text entry methods on modern touchscreen devices. Its success stems from the ease with which it can be learned [24, 59, 60] and the high entry rates that can be achieved [24, 30, 40]. By design, the word-gesture keyboard provides good tolerance to inaccuracy in users' traces when combined with high quality language models. It is no surprise then that the word-gesture keyboard has been successfully ported for use in augmented and virtual reality. Indeed, the Microsoft HoloLens 2 supports gesture typing on its system keyboard.

The word-gesture keyboard was originally developed as an input method for typing on personal digital assistant (PDA), Tablet PC or mobile phone with a stylus [24, 30, 59]. The technique was then easily transferred to exploit the subsequent popular adoption of smartphones with capacitive touchscreens in the mid to late 2000s [25, 60, 61]. Its popular adoption has been attributed to the high entry rates achievable and the relatively shallow learning curve [24, 26]. These same traits mean that word-gesture typing is also likely to become a popular input method in AR and VR, and indeed various studies [6, 19, 42–44, 53, 55] have examined use of a word-gesture keyboard in an immersive setting using various interaction techniques. The entry rates typically achieved in this setting are in the range of 20 to 40 wpm for top performing participants or expert users [6, 36, 55, 58]. These entry rates sit at the lower range of those typically achieved on a smartphone by experienced gesture typists, which according to the comprehensive study by Leiva et al. [33] involving data from 909 users ranges from 50 wpm for expert users to 40 wpm for those who never use a gesture keyboard. The difference in performance between experienced and non-experienced

Table 1: Summary of mean entry rates and expert or top performer entry rates reported for mid-air word-gesture keyboards.

Study	Interaction	Entry Rate (wpm)	Expert / Top Performance (wpm)
Markussen et al. [36]	Projected Cursor - Hand	28.1	-
Yu et al. [58]	Projected Cursor - Head Orientation	24.7	39.0
Chen et al. [6]	Projected Cursor - Controller	16.4	35.4
Xu et al. [54]	Projected Cursor - Controller	13.7	-
Benoit et al. [5]	Projected Cursor - Hand	15.8*	-
Yanagihara et al. [55]	Direct Touch - Controller	-	21.0
Kern et al. [23]	Direct Touch - Controller	17.5	-
Speiss et al. [46]	Direct Touch - Controller	12.9	-
Lu et al. [34]	Projected Cursor - Head	9.8	-
Wang et al. [53]	Projected Cursor - Hand	6**	-
Dube et al. [9]	Indirect Cursor - Finger Mouse	11.3	16.5
Henderson et al. [22]	Indirect Cursor - Touchscreen	13.2	-
Gupta et al. [19]	Indirect Cursor - Ring	14.8	-

*Entry rate based on entry of a single word. **Approximated from plot.

users of gesture typing observed by Leiva et al. [33] highlights a major challenge in assessing the performance potential of different word-gesture keyboard implementations in VR. Given the still relatively modest adoption of VR, it can be very difficult to recruit participants who are both practiced in gesture typing and familiar with using VR. It is reasonable to expect that transferring existing typing skills to VR will be a new experience for most experiment participants but even more daunting if they have no prior VR experience or have never used a gesture keyboard.

Prior to work focusing on AR and VR, Vulture [36] demonstrated mid-air gesture typing to support text input on large wall displays. Vulture relied on precision hand tracking and users gestured by moving the hand to control an indirect cursor shown on the large wall display. Participants using Vulture with this mid-air interaction mode achieved an average entry rate of 28 wpm [36]. It is worth noting that this was the performance achieved in the final block of nine blocks where participants typed the same six phrases, with only 15 unique words, twice in each block. Yu et al. [58] leverage the orientation tracking capability of the HMD to perform gesture typing with a projected cursor based on head orientation. Using the head orientation-based word-gesture typing method, participants averaged 24.7 wpm and Yu et al. [58] report that the best performing participant averaged 39 wpm in their final session. These entry rates are from the final session in a protocol where the same 10 phrases, with only 39 unique words, were entered in eight sessions. Xu et al. [54] evaluated a range of projected cursor methods for gesture typing on an AR HMD including head-orientation-based, hand-based and controller-based. The controller projection achieved the highest average entry rate at 13.7 wpm.

Multiple studies [6, 23, 46, 55] have demonstrated word-gesture typing with tracked VR controllers but using different interaction schemes. Chen et al. [6] use the tracked controller position to command an indirect cursor while Yanagihara et al. [55] use a pointer rigidly offset from the controller to directly 'hit' keys. Chen et al. [6] report on a pilot study in which participants were able to average 16.4 wpm and an expert user could reach 35.4 wpm. Yanagihara et al. [55] only report on the average entry rate (21.0 wpm) achieved by the first author representing an expert user. Both Kern et al. [23] and Spiess et al. [46] examine using controllers to directly draw word gestures on the keyboard plane, obtaining mean entry rates of 17.4 wpm and 12.9 wpm respectively.

Table 1 summarizes the entry rates reported in prior studies investigating mid-air word-gesture typing. This tables was assembled by performing a Scopus search with the following query: *(TITLE-ABS-Key("gesture") OR TITLE-ABS-Key("sw*pe")) AND (TITLE-ABS-KEY("text input") OR TITLE-ABS-KEY("text entry")) AND (TITLE-ABS-KEY("mid-air") OR TITLE-ABS-KEY("virtual reality") OR TITLE-ABS-KEY("augmented reality") OR TITLE-ABS-KEY("HMD*"))*. On 25 March 2023, this search returned 76 results. Filtering on papers that could be accessed, actually involved word-gesture typing and also presented an evaluation of entry rates yielded the 13 papers contained in Table 1. Where available we include both mean entry rates achieved, as well as expert or top performer entry rates. It is important to use caution when comparing entry rates from different studies. A wide range of factors, such as different protocols, phrase sets and recruitment strategies, can influence performance findings. Nevertheless, we suggest that peak rates provide a useful point of comparison given that they inherently reflect factors intrinsic to the system, such as recognition accuracy and the efficiency of the underlying interaction technique.

Notably, Table 1 reveals that there is no prior work examining word-gesture typing based on direct hand-based interaction, that is, where a user can gesture in mid-air using the tip of their index finger as they would on their smartphone. Direct hand interaction leveraging hand or finger tracking for text entry has been explored for 'touch' based input [10,11,41] but not for entering word-gestures. We address this gap by examining word-gesture typing based on direct finger interaction.

2.2 Mid-Air Touch Typing

The VISAR keyboard [11] was developed to conduct one of the first studies of mid-air touch typing using integrated HMD hand-tracking. In a study performed with the Microsoft HoloLens 1, which provided only coarse position tracking of the hand (no finger tracking), Dudley et al. [11] demonstrated how a robust input decoder and a virtualized input surface can be combined to support mid-air entry on a virtual Qwerty keyboard. Users could only use one finger and the mean entry rate achieved in the most refined version of the system was 17.8 wpm. Speicher et al. [45] compared six different selection-based methods in VR variously leveraging head orientation, tracked controllers, and tracked hands. Speicher et al. [45] found that discrete key selection using rays emitted from two tracked controllers achieved the highest entry rates (averaging 15.4 wpm) and was most favored by participants. A unique approach proposed by Lee and Kim [32] employed additional buttons incorporated into the controller. The user could then 'touch' on of these buttons to indicate selection of a specific key within the sub-region of the keyboard currently highlighted by a ray projected from the controller. Frutos-Pascual et al. [14] examined various aspects of the design of a mid-air AR keyboard but focused primarily on the efficiency of individual character selections and did not report entry rates. Adhikary and Vertanen [1] focused on the combined use of a mid-air keyboard with speech input in a VR setting but did report performance for a no-speech condition. With hand tracking provided by a Leap Motion attached to the VR HMD, Adhikary and Vertanen [1] obtained a mean entry rate of 11.1 wpm on their mid-air touch-based Qwerty keyboard. In a different study, Adhikary and Vertanen [2] again used a Leap Motion for hand tracking and compared two-finger typing on a standard and split mid-air Qwerty keyboard, observing mean entry rates of approximately 16 wpm and 15 wpm respectively.

To address the lack of passive haptic feedback when typing on mid-air virtual keyboards, Dube et al. [8] examine the potential benefit of providing mid-air ultrasonic feedback. Dube et al. [8] investigate entry performance with and without ultrasonic feedback during typing and find that such feedback can improve entry rates and reduce error rates. Even with ultrasonic feedback, however, the mean entry rate achieved by participants was still relatively low at 12.3 wpm.

Most prior work examining various aspects of mid-air touch typing to date has leveraged external tracking systems or devices [10, 20, 56]. Dudley et al. [10] offer an informative study of the feasible performance of both mid-air 10 finger and 2 finger touch typing in VR. This study leveraged a precision marker-based tracking system to track participants' fingertips and a simulated input decoding using a Wizard of Oz protocol to measure potential entry rates in VR. This setup served to examine feasible entry rates if unencumbered by (at the time) current hand tracking and input decoding limitations. The mean simulated entry rate for 2-finger and 10-finger mid-air typing in the study by Dudley et al. [10] was 42.1 and 34.5 wpm respectively. An unfortunate issue with 10 finger mid-air typing being the occurrence of spurious touches due to unintentional finger coactivations [13].

With a similar scope to this paper, Kern et al. [23] directly compare typing touch and gesture typing performance in both VR and AR leveraging tracked controllers. Kern et al. [23] found that touch typing using a controller delivered faster entry rates than gesture typing using a controller in both VR and AR (19.5 versus 17.5 wpm in VR and 16.1 versus 15.4 wpm in AR). These results are based on the entry of only ten phrases with a total 59 unique words (two phrases in the practice phase and eight phrases in the test phase) in each condition. Our investigation of direct hand-based interaction with the tracked fingertip delivers new insights beyond that provided by Kern et al. [23]. Hand tracking is less precise than controller-based tracking but does more closely replicate the experience of typing on a physical layout. Further, we evaluate performance in both typing paradigms over 160 phrases with an average of 503 unique words. Our paper is the first to directly compare gesture and touch typing on a mid-air virtual Qwerty keyboard exclusively using integrated hand tracking.

3 Keyboard Design

We built a versatile probabilistic keyboard system supporting both touch and gesture typing. It was developed in Unity and can be readily deployed to AR and VR headsets. In our studies we deploy it to the Meta Quest 2. It is deployed as an Android apk and leverages the integrated hand tracking available on the Meta Quest. No external tracking or processing is required.

The keyboard, as it appears in the VR headset, is shown in Figure 1. Users may interact with the keyboard using either of the two interaction modes previously introduced: *Poke* and *Raycast*. The implemented keyboard supports both gesture typing and touch typing. When gesture typing, users articulate word gestures by either directly *poking* their finger into the keyboard plane or by controlling a *raycast* cursor projected onto the keyboard plane. In the *Poke* interaction mode, gesture start and stop are indicated by entry into and departure from the keyboard plane while in the *Raycast* mode, gesture start and stop are indicated by a thumb-to-index finger pinch and release. Touch typing is performed in a similar manner except that the initial *Poke* touch point or *Raycast* pinch point is treated as a discrete 'touch'.

For *Poke* gesture and touch typing, 'touch' on/off events are raised when a collider attached to the fingertip enters a collider aligned with the keyboard plane. The *Raycast* mode is directly based on Meta Quest developer guidance and utilizes the hand tracking 'PointerPose'[2] as the root Transform for the raycast. *Raycast* and pinch on the same hand preserves the design intent[1] of a consistent *Raycast* interaction across various forms of interface. Audio feedback was provided to the user upon touch and gesture start/stop events (please refer to the video figure for details) in both interaction modes.

The default keyboard size and position were determined through a lab-based pilot study and fixed in world space. The *Raycast* keyboard was 60 cm wide and initially placed 80 cm in front of the user. The *Poke* keyboard was 30 cm wide and initially placed 32.5 cm in front of the user. Participants were encouraged to adjust their seating to ensure comfortable use of the keyboards.

For the purpose of the studies presented in this paper we implemented two different input decoding techniques specifically designed to maximize user performance under each typing method. The full-system latency, from triggering of the decoder to display of the result, is affected by various factors including the length of the word but is typically in the range of 200 to 300 ms for both decoders. For conciseness and given we do not claim any algorithmic contribution in relation to these techniques, we only provide a brief overview of the two implementations.

3.1 Touch Decoder

The keyboard's touch decoder is implemented using a token-passing strategy with both spatial and language model components [15, 51]. The likelihood of any discrete observation being associated with a given key is estimated using a 2D Gaussian. These discrete observations are combined together through token passing to produce a set of potential partial or full word hypotheses. As each new observation is consumed, low probability hypotheses, as determined by a character language model and the spatial model, are removed. Finally, a word language model re-ranks the set of hypotheses based on the prior sentence context. We used the 'tiny' 3-gram word language model made publicly available by Vertanen and Kristensson [48] and a 5-gram character language model produced from Project Gutenberg text.

Input decoding was triggered when the user pressed the space key to deliver an auto-correction functionality comparable to that typically found on modern smartphone keyboards. The most likely hypothesis replaces the word actually entered by the user and the next most likely hypotheses are displayed as selectable alternatives above the keyboard. These alternatives remain available for selection until the next time the input decoder is triggered, as seen Figure 1 (right). If the user's literal entry that was replaced is not among the top-4 hypotheses returned by the decoder, this literal entry is included as one of the selectable alternatives. When touch typing, pressing the backspace button removes the previously entered character.

3.2 Word-Gesture Decoder

The keyboard's gesture decoder is implemented using a geometry-based finite state transducer (FST) [3, 27, 38, 57]. The FST approach offers several key benefits. First, it provides good robustness to imprecise articulation and interaction at different scales. This allows it to handle the typically noisy mid-air gestures performed by users as well as support resizing of the keyboard. The relevance of this capability in VR is discussed further in Section 8. A second key benefit of the FST approach is that it permits seamless handling of both discrete input (i.e., single character at a time typing) and gesture input within the same decoding architecture. For gesture input, observations may be treated as 'in-transit', indicating that the trace is currently transitioning between target keys. When a target key is reached then this observation is consumed as being 'aligned'. For discrete input, all observations are consumed as being 'aligned'.

An FST reads from an input tape and writes to an output tape. A set of multiple alternative hypotheses are maintained at each timestep as each new observation is consumed. For efficiency, a transition graph is constructed from a defined vocabulary list. We refer to Ouyang et al. [38] for a more complete description of the FST for decoding word-gestures. The implementation evaluated in this paper uses a vocabulary of approximately 64 thousand words and leverages a 3-gram language model. Again, we used the 'tiny' 3-gram word language model made publicly available by Vertanen and Kristensson [48] and derive our vocabulary from this language model's unigram.

After performing a word gesture, the top recognition is automatically inserted into the input field and the next three most likely alternatives are displayed above the keyboard. The user can select one of these alternative words to replace the automatically inserted top recognition result. When gesture typing, pressing the backspace button removes the previously entered word.

4 Evaluation Outline

Our primary goal in this work is to examine the potential of a mid-air Qwerty keyboard for VR based exclusively on hand-based interaction. The two dominant input methods on mobile devices are gesture and

[2]https://developer.oculus.com/documentation/unity/unity-handtracking/

touch typing and we seek to understand how these methods are best adapted to a mid-air setting. However, as outlined in Section 2, the literature on hand-based interaction for gesture typing is far less comprehensive than it is for touch typing and so we begin by addressing this specific gap. We therefore sequentially tackle the following three key research questions: **RQ1**) What is the most effective and preferred hand-based interaction mode for a mid-air gesture keyboard? **RQ2**) What is the feasible performance limit for a gesture keyboard in VR? and **RQ3**) How does hand-based gesture typing compare to two-finger touch typing in VR? Our design of the three studies directly relates to these three research questions. The goal in Study 1 is to examine the less well explored input method (gesture typing) and identify the specific effect of interaction mode: *Poke* versus *Raycast*. Study 1 reveals the dominance of *Poke* and so in Study 2 we seek to determine the performance potential of *Poke* gesture typing. Finally, since no prior work has directly compared mid-air hand-based touch and gesture typing, in Study 3 we examine their performance and preference distributions.

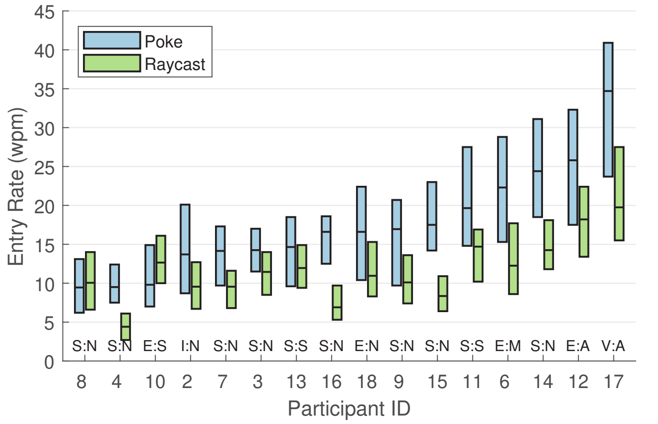


Fig. 2: Inter-quartile range of participant entry rates in Poke and Raycast interaction modes. Participants are ordered along the x-axis based on median entry rate in the Poke condition. Indicated just above the x-axis is the reported level of experience with VR (I: inexperienced, S: some experience, E: experienced, V: very experienced) and frequency of use of gesture typing (N: never, S: sometimes, M: most of the time, A: all the time) in the format $<$*experience level*$>$:$<$*frequency*$>$.

5 Study 1: Poke versus Raycast

Study 1 examines the performance of the two dominant interaction modes currently persistent in VR: *Poke* and *Raycast*. We compare the entry and error rate performance when gesture typing for each of these interaction modes in a within-subjects experiment.

5.1 Protocol

Participants performed a standard transcription task, entering 10 practice phrases followed by 50 test phrases in each condition. 120 stimulus phrases were drawn at random from the Enron mobile message dataset [49] after filtering based on phrases containing four words or more, and 40 characters or less. There were 317 unique words across the 120 phrases. Phrase order was randomized for each participant and the order of conditions was counterbalanced. Each condition took approximately one hour to complete.

Participants were instructed to type as quickly and as accurately as possible. They were permitted to use backspace and select from word alternatives. After entering each phrase participants were shown their entry rate in words per minute (wpm) and their character error rate (CER). We use the standard definitions for wpm and CER. First, wpm is defined as the number of words entered divided by the time taken, where the numerator is an effective word count using a nominal word length of five 'characters' including spaces. As a conservative estimate of entry rate, we use the entered phrase length minus one since the entry time is measured from the commencement of the first word gesture to the completion of the final word gesture. Selection of a word alternative at the end of the phrase is treated as a final interaction event and included in the total entry duration. Second, CER is the minimum number of character insertion, deletion and substitution operations required to transform the response text into the stimulus text, divided by the stimulus text length. We report CER as a percentage.

After a block of 10 entries, a dialog appeared and participants were instructed to take a short break. After 30s, the task resumed automatically. The keyboard would be in uppercase mode by default on presentation of a new phrase and participants were instructed that use of the *Shift* key otherwise was optional.

Participants were asked to complete a pre- and post-experiment questionnaire. The pre-experiment questionnaire requested details on age, gender, prior experience using VR, prior experience with word-gesture typing, and self-assessed typing speed on a smartphone. The post-experiment questionnaire asked participants to reflect on their experience of using the mid-air word-gesture keyboard in VR in terms of the degree to which it allowed them to type fast and/or accurately in each condition. Finally, they were asked to indicate their preferred interaction mode. An optional text field was also available at the end of the post-experiment questionnaire to provide participants with an opportunity to express any general feedback or observations.

5.2 Participants

The experiment itself was conducted remotely with the initial phase supervised by a research assistant over video. This methodology was pursued to minimize health risks to participants associated with the COVID-19 pandemic. However, the methodology also provides secondary benefits and insights. First, remote experimentation facilitates the recruitment of participants with demographics more broadly representative than typical campus-based studies. Second, the fact that it is possible to remotely instruct participants in the use of the system serves to highlight the walk-up usability of the developed word-gesture keyboard and test application. Detailed instructions on the experimental protocol and how to use the keyboard were presented at the start of the experiment. We used a set of set of demonstration videos for this purpose. Participants were recruited via our organization's managed participant pool and were compensated for their participation.

18 participants were recruited in total. We excluded two participants from subsequent analysis: *P1* failed to complete the experiment and *P5* exhibited significantly elevated error rates (median CER of 15.5%) in the Raycast condition (all other participants had a median CER of 0.0%). The results reported are therefore based on 16 participants (8 female, 6 male, 1 non-binary, 1 prefer not to say). The ages of the participants ranged from 22 to 63 with a mean of 36.3. This participant sampling reflects a considerably more diverse age group than, for example, the final study of Yu et al. [58] (12 participants aged 20 to 24, mean = 21.1) or Markussen et al. [36] (12 participants aged 18 to 31, mean = 24.4).

5.3 Results

The entry rate performance of the study participants is summarized in Figure 2. This plot shows the inter-quartile range of participant entry rates in each of the Poke and Raycast interaction modes. The participants are ordered along the x-axis by their median entry rate in the Poke condition to help illustrate the distribution of performance.

The mean entry rate in the Poke condition was 17.7 wpm (s.d. 6.3) and 11.9 wpm (s.d. 4.0) in the Raycast condition. A repeated measures analysis of variance shows a significant effect for the interaction mode on entry rate ($F_{1,15} = 37.153$, $\eta_p^2 = 0.712$, $p < 0.001$). Figure 2 reveals that only *P8* and *P10* achieved higher median entry rates in the Raycast condition. This result suggests a pronounced entry rate performance advantage for Poke over Raycast, with the mean entry rate in the Poke condition being 48.7% higher than the Raycast condition.

Also represented in Figure 2 just above the x-axis is the participants' reported frequency of using gesture typing. Only two participants (*P12*,*P17*) reported using gesture typing all the time as their standard text input method on their smartphone. Notably these two participants also achieved the highest median entry rates. One participant (*P6*) reported using gesture typing most of the time and they achieved the

Table 2: Error-free peak entry rates in wpm achieved by participants for each phrase (Phr. 1 to 5). Also shown are participants' level of prior experience with VR (VR Exp.), normal frequency of using a word-gesture keyboard on their personal smartphone (WGK Use), self-assessed normal entry rate on a smartphone (ER), and self-rating of speed (Speed) and accuracy (Acc.) achieved in the post-experiment questionnaire. Please refer to the supplementary material linked in Acknowledgments for full text of questions and available responses. In the 'WGK Use' column, Never* indicates that a participant has never used a word-gesture keyboard before. Top entry rates achieved for each phrase are indicated in bold. Entries equal to or exceeding 40 wpm are shaded in red. Entries equal to or less than 30 wpm are shaded in blue.

P	VR Exp.	WGK Use	ER	Phr. 1	Phr. 2	Phr. 3	Phr. 4	Phr. 5	Speed	Acc.
1	Exp.	Never	Slow	29.2	39.3	44.8	41.3	33.3	Slow	Inacc.
2	Very exp.	All the time	Very fast	**49.2**	50.7	**66.0**	**48.1**	**59.6**	Very fast	About avg.
3	Very inexp.	Never*	About avg.	40.6	35.9	63.3	32.0	38.9	About avg.	About avg.
4	Very inexp.	Never	About avg.	42.8	40.0	55.8	41.4	48.8	About avg.	About avg.
5	Very inexp.	Never	Very fast	22.6	19.0	34.8	26.7	18.3	Slow	Inacc.
6	Some exp.	Most of the time	Fast	24.8	36.0	44.2	18.3	22.9	Slow	Very inacc.
7	Some exp.	Never	Fast	26.1	17.9	32.3	23.9	26.7	Slow	About avg.
8	Some exp.	Some of the time	About avg.	27.1	26.4	39.9	25.9	34.0	Slow	About avg.
9	Some exp.	Never*	Fast	48.5	45.7	55.2	46.3	39.7	Fast	About avg.
10	Exp.	All the time	Very fast	29.5	**56.6**	**66.0**	35.2	58.0	About avg.	Inacc.
11	Inexp.	Never*	About avg.	28.4	31.4	31.1	33.0	29.3	About avg.	About avg.
12	Some exp.	Never	Fast	34.6	34.7	37.9	39.9	35.6	About avg.	About avg.
13	Some exp.	Some of the time	Fast	39.0	36.7	38.0	37.2	34.6	About avg.	Inacc.
14	Some exp.	Never	Fast	23.9	21.8	43.0	31.7	18.8	About avg.	About avg.

fourth highest median entry rate in the participant group. Three participants reported using gesture typing sometimes and the remaining 10 participants reported never using gesture typing. Among the 10 participants who reported never using gesture typing, three (*P4*,*P9*,*P14*) also indicated that they had never even used a gesture keyboard. Despite this lack of prior exposure, it is notable then that *P14* achieved the third highest median entry rate. As a point of comparison with gesture typing on a smartphone, Reyal et al. [40] conducted a study with 12 participants (3 familiar with a gesture keyboard) who achieved a mean entry rate of 30.6 wpm in the fifth of five 10 minute typing sessions and a 95% confidence interval of 27.2 to 34.0 wpm.

All participants achieved a median CER of 0.0% in both conditions indicating that the backspace and alternative word selection functionality enabled very accurate transcription. The mean CER for the participant group was 1.9% (s.d. 1.6) in the Poke condition and 1.5% (s.d. 1.2) in the Raycast condition. This difference was not significant.

5.3.1 Interaction Preference

In the post-experiment questionnaire, participants were asked to indicate their preferred interaction mode. 13 of 16 participants preferred Poke over Raycast. The factors that participants highlighted for this preference of Poke over Raycast included greater comfort/less fatiguing (*P4*,*P11*,*P15*), similarity to typing on a smartphone (*P3*,*P12*), reduced sensitivity (*P14*,*P17*) and speed of execution (P16). Notable written comments regarding the Poke method included, "Felt more natural/similar to other gesturing keyboards." (*P12*) and "Poke seemed more tangible since I was actually 'touching' the keyboard." (*P17*). Of the three participants who preferred the Raycast method, the factors listed as contributing to this preference were greater accuracy (*P13*,*P18*) and greater reliability in gesture execution (*P10*).

6 Study 2: Assessing Peak Potential Entry Rates

Study 2 seeks to understand the human performance potential of the word-gesture keyboard in a VR setting. Inspired by Kristensson and Zhai [30] in their early studies of stylus-based word-gesture keyboards, we utilize a less conventional experimental protocol aimed at assessing peak performance potential by saturating motor memory. This is achieved by requiring participants to transcribe the same stimulus phrase 40 times in a row. The results produced by such a protocol yield rich distributions indicative of performance potential after extensive practice with the keyboard. Mutasim et al. [37] recently demonstrated that a repeated entry protocol offers a viable way to estimate expert performance for text input on novel and unfamiliar keyboards.

We restricted participants to using the *Poke* interaction mode only in line with the performance advantage and user preference findings in Study 1. Five phrases were drawn at random from the Enron mobile message dataset [49] after first filtering based on phrases containing four words or more, and 40 characters or less. The five phrases used were: Phrase 1, "We need a process to deal with this"; Phrase 2, "If he wants it"; Phrase 3, "That was so sweet"; Phrase 4, "Any news from Redmond"; and Phrase 5, "Could you see where this stands".

As described above, the study protocol asked participants to enter the same phrase 40 times in a row. With five different phrases, each participant performed a total of 200 phrase entries. The order of phrases was randomized for each participant. The total duration of the study was approximately one hour. To assess peak entry rates we take the highest error-free entry rate achieved by each participant for each phrase. Use of backspace and alternative word selections was permitted. Participants were instructed to maximize their entry rate by building up their speed as their familiarity with the phrase increased.

6.1 Participants

As in Study 1, participants were recruited through the a remote participant panel. A total of 17 participants were recruited and participated in the study. Three participants failed to register error-free entries on one or more phrases indicating non-compliance with experiment instructions and were therefore excluded from subsequent analysis.[3] The 14 remaining participants (8 female, 6 male) ranged in age from 22 to 65 and had a mean age of 40.

6.2 Results

Table 2 summarizes the peak error-free entry rates achieved by each participant across the five phrases. Also listed in Table 2 are individual participant responses to the pre-experiment questions relating to prior VR experience, frequency of use of word-gesture typing, and self-assessed typing speed on a smartphone. In summary, the participant sample reflects a rich variation in familiarity with gesture typing and general prior experience of VR.

Table 2 reveals a spectrum of entry rates achieved by participants. The performance of *P2* stands out with the top entry rate in four of the five phrases recorded by this participant. *P10* also reached very high entry rates, recording the top entry rate for Phrase 2 and equaling *P2* on Phrase 3. Notably both *P2* and *P10* were the only two participants to respond that they use the gesture typing method on their smartphone 'All the time' in the pre-experiment questionnaire. *P2* was 'Very experienced' with using VR while *P10* was 'Experienced'. We therefore conjecture that both prior word-gesture keyboard experience and prior VR experience were key enablers in allowing these two participants to type in the approximate range of 50 to 60 wpm.

P6, *P8* and *P13* were the remaining three participants who indicated some regular usage of the gesture typing method on their smartphone.

[3] One participant failed to enter any phrases correctly. Another failed to enter Phrases 1 and 5 correctly and the final participant just failed on Phrase 5.

P8 responded that they use gesture typing ‘Most of the time’ while *P8* and *P13* use gesture typing ‘Some of the time’. All three participants responded that they have ‘Some experience’ using VR. Despite being regular or semi-regular users of gesture typing, the peak entry rates achieved by these three participants are largely consistent with other non-gesture-typing participants. This result sits in contrast to the peak entry rates achieved by *P3* and *P4* who were both ‘Very inexperienced’ with VR and ‘Never’ used gesture typing on their smartphone. Nevertheless, *P3* and *P4* reached peak entry rates of 32.0 to 63.6 wpm and 40.0 to 55.8 wpm respectively across the five phrases. This observation is promising for the walk-up usability of mid-air hand-based word-gesture typing in VR, a critical factor in the adoption of any new keyboard interaction mode. A further subset of participants (*P5*,*P7* and *P14*) did not exceed 30 wpm in more than one or two of the five phrases. Nevertheless, peak entry rates in an approximate range of 20 to 30 wpm are comparable with expert/top performer entry rates reported in the prior work listed in Table 1. This is despite the fact this subset of participants represents the lower end of the spectrum of performance in our study and indicated that they ‘Never’ use gesture typing on their smartphone. Averaging over all participants in Table 2, the mean peak entry rate achieved for Phrases 1 to 5 respectively were: 33.3 wpm, 35.2 wpm, 46.6 wpm 34.4 wpm and 35.6 wpm.

The mixed results captured in Table 2 highlight the challenge of conducting user studies examining the less-widely adopted gesture input method in combination with the novel VR setting. The peak-performance protocol does, however, reveal the spectrum of performance that can be expected both after a degree of practice and after extensive use on a smartphone. Our results show that entry rates in excess of 50 and 60 wpm are achievable on a mid-air gesture keyboard. As a point of comparison, Leiva et al. [33] report that experienced gesture typists average around 50 wpm on a smartphone. This suggests that the performance gap for mid-air text entry with respect to text entry on a smartphone may be smaller than widely observed to date.

6.2.1 Perceived Speed and Accuracy

The individual participant responses in the post-experiment questionnaire to the question, “Using the gesture keyboard in VR, I was able to type:” in terms of both speed and accuracy are listed in the final two columns of Table 2. This self-assessment of performance suggests that perceived entry rates tend to be ‘About average’ or ‘Slow’ which is not unexpected given the main point of comparison for most participants is likely to be typing on a smartphone or physical keyboard. In terms of an envisaged text entry use-case in VR requiring the entry of a short message to a friend or typing brief in-game notes, slightly slower entry rates may be tolerable if the interaction modality maintains immersion and does not require locating secondary input devices. The responses summarized in the final column of Table 2 regarding perceived accuracy also indicate that input accuracy is about average to inaccurate. Given the nature of the protocol, encouraging participants to maximize their entry rate, it is also reasonable to expect that participants perceive a generally negative trade-off in terms of accuracy.

6.2.2 Qualitative Feedback

Participants were given the opportunity to provide any general observations or feedback at the end of the post-experiment questionnaire. Two participants (*P3*, *P4*) commented on the long duration of the study, with one suggesting that their performance may have deteriorated due to tiredness. Multiple participants (*P6*, *P7*, *P10*, *P13*, *P14*) commented on the difficulty of typing accurately with two participants specifically observing that shorter words were easier to type than longer words and another participant commenting that presented word alternatives were often not useful. Finally, *P10* commented that, “Gesture keyboard had a pretty shallow learning curve,” while *P12* commented that using the keyboard, “Took a while to get used to it.” These various participant comments highlight clear opportunities for improvement.

7 Study 3: Comparing Gesture and Touch Typing

Study 3 compares the performance of participants gesture typing and two-finger touch typing on the mid-air Qwerty keyboard.

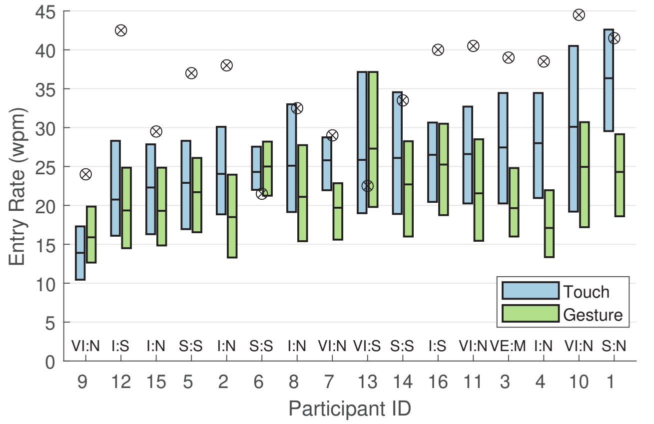


Fig. 3: Inter-quartile range of participant entry rates in touch and gesture conditions. Participants are ordered along the *x*-axis based on median entry rate in the touch typing condition. Indicated just above the *x*-axis is the level of experience with VR (VI: very inexperienced, I: inexperienced, S: some experience, E: experienced, VE: very experienced) and frequency of use of gesture typing (N: never, S: sometimes, M: most of the time, A: all the time) in the format <*experience level*>:<*frequency*>. The ‘⊗’ symbol indicates the participant’s mean entry rate in the smartphone typing speed test.

7.1 Protocol

This study was conducted in the laboratory and so reflects tighter control on instructions given to participants as well as the local environment in which the study was performed. At the beginning of the study, participants completed the same short demographics and prior-experience questionnaire from Studies 1 and 2. We also asked participants to complete a one-minute typing speed test using both a physical keyboard and their own smartphone. This typing test used a subset of phrases randomly drawn from the full set of phrases used in the main study.

For the study proper, participants were asked to transcribe 170 phrases in each condition. The first 10 phrases were designated as practice while the remaining 160 phrases were split into 8 blocks of 20 phrases. The stimulus phrases were randomly drawn from the MacKenzie phrase set [35] such that each participant entered 340 unique phrases over the full study. We used this phrase set as we desired non-overlapping phrases with the previous studies and we required at least 340 unique phrases. The 170 phrases in each condition contained on average 503 unique words. Five words in the phrase set were out of vocabulary for both decoders.

After completing the second session of the study, the physical keyboard and smartphone typing tests were repeated. Finally, participants responded to a post-experiment questionnaire in which they rated each condition in terms of its speed, accuracy and comfort. Participants were also asked to indicate an overall preference between the two conditions and share any general feedback. Each condition took between 60 and 90 minutes to complete (including introduction and breaks). The two conditions were completed in separate sessions in an attempt to control for fatigue. We enforced a minimum break between sessions of 4 hours and a maximum break of 2 days. Participants received an Amazon voucher as a token of appreciation for their participation.

7.2 Participants

We recruited 16 participants (7 females, 9 males) through convenience sampling at our local campus. Participants ranged in age from 20 to 26 with a mean of 21.6. All participants except one were right-handed. The keyboard application defaults to right hand use for gesture typing but can be configured for left hand use with a settings button.

7.3 Results

The entry rate performance of participants is summarized in Figure 3. Over the full set of 160 phrases, the mean entry rate was 25.6 wpm in

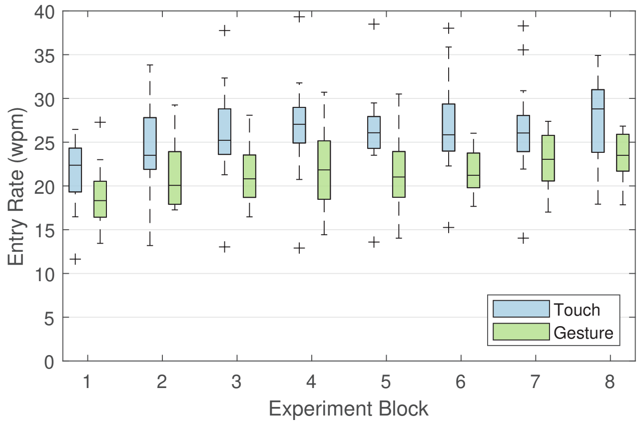


Fig. 4: Boxplot of participant entry rates in touch and gesture typing conditions for each experimental block.

the touch typing condition and 21.5 wpm in the gesture typing condition. A repeated measures analysis of variance shows a significant effect for typing method on entry rate ($F_{1,15} = 20.186$, $\eta_p^2 = 0.574$, $p < 0.001$). The mean character error rate for the full test phrase set was 2.25% and 1.66% for touch and gesture typing respectively. No significant effect was found for typing method on character error rate.

The general result observable in Figure 3 is that the majority of participants performed better when touch typing than gesture typing. *P9*, *P6* and *P13* were marginally faster in terms of median entry rate when gesture typing. Notably none of the 16 participants indicated that they used gesture-typing as their default input method on their smartphone. Also shown in Figure 3 is participants' mean entry rate on the two one-minute typing tests completed using their personal smartphone. *P6* and *P13* both achieved median entry rates above their mean smartphone typing speed. Four more participants (*P8*, *P7*, *P14* and *P1*) had a mean smartphone typing speed approximately within the inter-quartile range of their mid-air touch typing entry rates.

Figure 4 presents boxplots of participants' mean entry rate in each of the eight experimental blocks. The trend observable in Figure 4 suggests that during blocks 1 and 2 participants are still gaining familiarity with the keyboard and interaction mode. By block 4, the entry rate performance begins to plateau. In the final experimental block, the mean entry rate in each condition was 27.6 wpm and 23.4 wpm for touch and gesture respectively. A repeated measures analysis of variance again shows a significant effect for typing method on entry rate ($F_{1,15} = 13.996$, $\eta_p^2 = 0.483$, $p = 0.002$). The mean character error rate for the final block of 20 test phrases was 1.65% and 2.12% for touch and gesture typing respectively. This difference was not significant.

Much prior work makes little effort to support error correction and this tends to elevate reported entry rates at the expense of higher character error rates. In our protocol, we encouraged participants to "type as quickly and as accurately as possible" and demonstrated in study instructions how to correct errors, match stimulus capitalization and include sentence terminating punctuation. We observed that participants sometimes expended significant effort and time when correcting errors leading to reduced entry rates. To enable comparisons with prior work where less emphasis has been placed on correcting errors, or where no error correction functionality was available, we computed participants' mean entry rate from the final experiment block for the subset of phrases where the backspace key was never pressed. This filtering represents the subset of phrases entered by participants where they performed no error correction, although they may have selected word alternatives. The participants' mean entry rates for this subset of phrases were 33.8 wpm and 30.6 wpm for touch and gesture respectively. The effect of typing method on entry rate was significant ($F_{1,15} = 6.680$, $\eta_p^2 = 0.308$, $p = 0.021$). The character error rate for this same subset remained relatively low despite the lack of error correction and difference between conditions was not significant: 1.90% for touch typing and 1.59% for gesture typing.

As an additional point of comparison with the results presented in Study 2, the error-free peak entry rates averaged over all participants were 44.9 wpm and 39.2 wpm for touch and gesture typing respectively. This suggests that with extensive practice, participants can reasonably achieve entry rates well in excess of 20 wpm. The highest peak entry rate for touch typing (56.9 wpm) and the highest peak entry rate for gesture typing (50.4 wpm) were both achieved by *P10*.

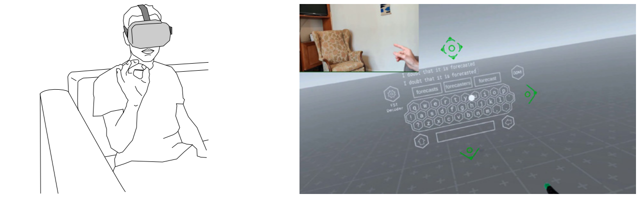
7.3.1 Perceived Speed, Accuracy and Comfort

The participants' median responses to the post-experiment questionnaire regarding speed, accuracy and comfort are now briefly summarized. Ratings were captured on a scale from *Very <Slow/Inaccurate/Uncomfortable>* (1) to *Very <Fast/Accurate/Comfortable>* (5). Participants generally perceived touch typing to be faster (median of 4 for touch typing versus 3 for gesture typing) but slightly less accurate than gesture typing (median of 3.5 for touch typing versus 4 for gesture typing). Neither of these differences were significant based on a Wilcoxon signed rank test. Both typing methods were rated similarly (both median of 4) in terms of comfort. In response to, "My preferred interaction mode for typing in VR was:", eight participants indicated touch and eight indicated gesture typing.

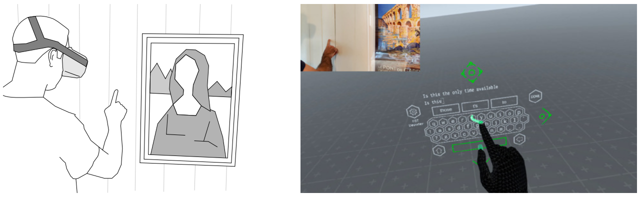
8 A Complete Mid-Air Virtual Qwerty Keyboard

Text entry methods that achieve mainstream success tend to share two common traits: i) good performance; and ii) similarity, and hence immediate efficacy, to existing methods [26]. Our results from the three studies suggest that the implemented mid-air Qwerty keyboard satisfies these two requirements. In Study 2 we found that a subset of participants experienced in VR are able to achieve entry rates in the range of 29 to 66 wpm and the *Poke* interaction mode made possible by the integrated hand tracking enables an experience closely resembling gesture typing on a capacitive touchscreen with an index finger. Study 3 confirmed the performance potential of the gesture input method for mid-air text input and established that the similarly familiar act of touch typing on a touchscreen can also be readily ported for use on immersive HMDs. We suggest, however, that an effective text entry method for mixed reality (MR) devices introduces several additional requirements due to the various contexts of use encountered in immersive settings. An effective text entry method for MR should ideally: i) be seamlessly engaged with the currently active interaction mode; ii) support appropriation of physical surfaces to improve comfort; and iii) be sufficiently flexible to accommodate physical space constraints.

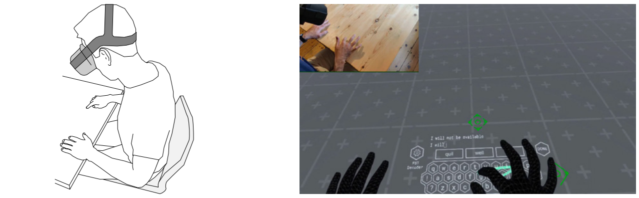
Figure 5 and the accompanying video figure illustrate how the probabilistic mid-air keyboard system is capable of satisfying these additional requirements. First, the keyboard supports the two dominant modes of interaction afforded by integrated hand tracking: *Poke* and *Raycast*. A user interacting with other virtual objects and interface elements using a *Raycast* interaction technique can seamlessly leverage this same technique to enter text, as illustrated in Figure 5a. Further, robust input decoding means that the keyboard layout can be placed at a depth consistent with other interface elements to maintain a consistent interaction experience. Second, the keyboard supports alignment with a physical surface to enable, for example, the appropriation of a wall or table to provide passive haptic feedback as illustrated in Figures 5b and 5c. This ensures a consistent experience if the main MR application currently engaged with also involves surface-based interactions. In addition, leveraging physical surfaces when available is likely to improve comfort. The keyboard supports flexible sizing to enable use in confined spaces, such as an airplane, car or other vehicle as illustrated in Figure 5d. This functionality is enabled by the input decoding system, which provides an effective degree of tolerance to different layout sizes. Finally, the agnostic design of the interactions and decoding plugins allows the probabilistic keyboard system to be readily ported to other platforms and devices. Figure 5e shows the same keyboard system running on-device as a UWP application on the Microsoft HoloLens 2 AR headset, supporting all of the same capabilities and interactions. These various additional features of the keyboard and the use-cases they enable are readily supported thanks to the integrated hand-tracking and cable-free form of modern MR headsets.



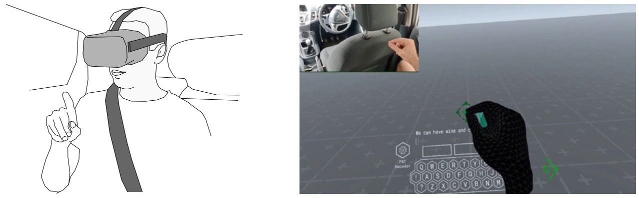
(a) *Raycast* gesture-typing on distant enlarged keyboard.



(b) *Poke* gesture-typing with keyboard aligned to a physical wall.



(c) *Poke* gesture-typing with keyboard aligned to a physical desk.



(d) *Poke* gesture-typing in a confined space.

(e) *Poke* touch typing in augmented reality.

Fig. 5: Common text entry use cases in mixed reality (MR) supported by the probabilistic mid-air Qwerty keyboard system.

9 Discussion

Study 1 revealed a distinct performance advantage to using Poke over Raycast. Poke was also preferred by 13 of the 16 participants in the study. The entry rates observed in Study 1 were lower than some related work but are based on a considerably more age diverse sample and a more challenging protocol (317 unique words versus the 15 and 39 unique words of Markussen et al. [36] and Yu et al. [58] respectively). Our evaluation with a remote user panel also likely excluded the typical performance biases encountered when recruiting participants locally from engineering or computer science programs. This study therefore offers a useful empirical benchmark reflecting the distribution of VR text entry performance likely to be seen in a broader sample of users.

The results of the peak-performance assessment from Study 2 suggest that entry rates in the range of 50 to 60 wpm are achievable on a mid-air word-gesture keyboard. This observation represents a distinct delta from previously reported performance using a word-gesture input method in a VR or AR setting. We hypothesize that the efficiency with which the two participants who use gesture typing by default on their smartphone were able to transplant their skills to the novel VR setting stems in part from the *Poke* interaction mode that closely resembles the motor task of typing on a smartphone. This affords good walk-up usability, as demonstrated by several participants who never used gesture typing but were nevertheless readily able to complete the study.

Finally, Study 3 revealed that two-finger typing generally supported higher entry rates than gesture typing but that both methods are effective and acceptable to users. This highlights the benefits of supporting both common interaction modes in a comprehensive text input system. In Study 3 we observed a substantial entry rate reduction on phrases where participants performed error correction by using the backspace key. This motivates future work on streamlining the error correction process which to date has received limited attention in the literature.

More broadly, our three evaluations inform the future design of mid-air virtual Qwerty keyboards. We demonstrate that both touch and gesture typing provide acceptable (equally distributed preference in Study 3) and productive (mean >20 wpm in Study 3) input methods for typical text entry use cases likely to be encountered in VR or AR. We also highlight the significant effect that prior experience with VR and gesture typing can have on performance. From a design point of view, however, we illustrate through our peak performance analysis that the effective implementation of a probabilistic keyboard system, such as the one presented in this paper, imposes no obvious constraints on users being able to achieve entry rates in excess of 50 to 60 wpm.

The key limitations of this work and avenues for future work include the following: (1) the diversity and number of participants in the study; (2) the potential performance impact of employing a composition task [16, 29, 50] rather than a transcription task; (3) the choice of phrase set in the studies [28, 39, 47]; and (4) the impact of the language model and any personalization [12].

10 Conclusions

In this paper we have presented a versatile probabilistic mid-air virtual Qwerty keyboard system for touch and gesture typing that is deployable on-device and exploits the integrated hand-tracking offered by modern VR HMDs. Using this system we have demonstrated that high text entry rates are feasible using a mid-air Qwerty keyboard in VR with integrated hand tracking. Our three user studies have revealed a rich spectrum of performance influenced by prior experience with VR and prior expertise. In Study 1 we found a substantial entry rate advantage for the Poke interaction mode over Raycast for gesture typing. Although mean entry rates were not particularly high for the age and experience diverse participant group, the three participants who regularly use gesture typing on their smartphone achieved median entry rates in the range of 25 to 35 wpm. Furthermore, one participant who had never previously used a gesture keyboard was able to achieve a median entry rate of 25 wpm. Two participants in Study 2 that self-rated as 'Experienced' and 'Very Experienced' in VR and who also use word-gesture typing as their default text input method on their smartphone achieved peak entry rates in the range of 29 to 66 wpm. In Study 3 we found that in the final test block, participants achieved mean entry rates for gesture and touch typing of 23.4 wpm and 27.6 wpm respectively. Excluding phrases from this same block where participants spent time correcting errors, mean entry rates were 30.6 wpm and 33.8 wpm for gesture and touch typing respectively.

This is the first investigation evaluating and comparing the performance potential of touch and gesture based text input on a mid-air Qwerty keyboard driven exclusively by integrated hand tracking. In addition to these informative empirical benchmarks, we also demonstrate various features of the developed mid-air keyboard as they pertain to different use-cases encountered in VR. Given the findings of this study we envisage that hand-based touch and gesture typing will become widely used input methods in VR thanks to their efficiency and good tolerance to articulation and tracking induced noise.

Acknowledgments

This work was supported by Reality Labs Research, Meta Inc. Supporting data for this publication is available at https://doi.org/10.17863/CAM.97198.

REFERENCES

[1] J. Adhikary and K. Vertanen. Text Entry in Virtual Environments using Speech and a Midair Keyboard. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2648–2658, May 2021. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2021.3067776 2, 3

[2] J. Adhikary and K. Vertanen. Typing on Midair Virtual Keyboards: Exploring Visual Designs and Interaction Styles. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen, eds., *Human-Computer Interaction – INTERACT 2021*, pp. 132–151. Springer International Publishing, Cham, 2021. 3

[3] O. Alsharif, T. Ouyang, F. Beaufays, S. Zhai, T. Breuel, and J. Schalkwyk. Long short term memory neural network for keyboard gesture decoding. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2076–2080, 2015. doi: 10.1109/ICASSP.2015.7178336 4

[4] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. Fitzmaurice. Experimental Evaluation of Sketching on Surfaces in VR. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 5643–5654. ACM, New York, NY, USA, 2017. doi: 10.1145/3025453.3025474 1

[5] G. Benoit, G. M. Poor, and A. Jude. Bimanual Word Gesture Keyboards for Mid-Air Gestures. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, p. 1500–1507. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3027063.3053137 3

[6] S. Chen, J. Wang, S. Guerra, N. Mittal, and S. Prakkamakul. 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, May 2019. doi: 10.1145/3290607.3312762 2, 3

[7] T. J. Dube and A. S. Arif. Text Entry in Virtual Reality: A Comprehensive Review of the Literature. In M. Kurosu, ed., *Human-Computer Interaction. Recognition and Interaction Technologies*, Lecture Notes in Computer Science, pp. 419–437. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-22643-5_33 2

[8] T. J. Dube and A. S. Arif. Ultrasonic Keyboard: A Mid-Air Virtual Qwerty with Ultrasonic Feedback for Virtual Reality. In *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3569009.3573117 3

[9] T. J. Dube, K. Johnson, and A. S. Arif. Shapeshifter: Gesture Typing in Virtual Reality with a Force-Based Digital Thimble. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491101.3519679 3

[10] J. J. Dudley, H. Benko, D. Wigdor, and P. O. Kristensson. 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, Oct. 2019. doi: 10.1109/ISMAR.2019.00027 3, 4

[11] J. J. Dudley, K. Vertanen, and P. O. Kristensson. Fast and Precise Touch-Based Text Entry for Head-Mounted Augmented Reality with Variable Occlusion. *ACM Transactions on Computer-Human Interaction*, 25(6):30:1–30:40, Dec. 2018. doi: 10.1145/3232163 1, 2, 3

[12] A. Fowler, K. Partridge, C. Chelba, X. Bi, T. Ouyang, and S. Zhai. Effects of Language Modeling and Its Personalization on Touchscreen Typing Performance. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, p. 649–658. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2702123.2702503 9

[13] C. R. Foy, J. J. Dudley, A. Gupta, H. Benko, and P. O. Kristensson. Understanding, Detecting and Mitigating the Effects of Coactivations in Ten-Finger Mid-Air Typing in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445671 4

[14] M. Frutos-Pascual, C. Gale, J. M. Harrison, C. Creed, and I. Williams. Character Input in Augmented Reality: An Evaluation of Keyboard Position and Interaction Visualisation for Head-Mounted Displays. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen, eds., *Human-Computer Interaction – INTERACT 2021*, pp. 480–501. Springer International Publishing, Cham, 2021. 3

[15] D. Gaines, J. Dudley, P. O. Kristensson, and K. Vertanen. Statistical Keyboard Decoding. In J. H. Williamson, A. Oulasvirta, P. O. Kristensson, and N. Banovic, eds., *Bayesian Methods for Interaction and Design*, pp. 188–211. Cambridge University Press, 2022. doi: 10.1017/9781108874830.009 4

[16] D. Gaines, P. O. Kristensson, and K. Vertanen. Enhancing the Composition Task in Text Entry Studies: Eliciting Difficult Text and Improving Error Rate Calculation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445199 9

[17] J. Goodman, G. Venolia, K. Steury, and C. Parker. Language Modeling for Soft Keyboards. In *Proceedings of the 7th International Conference on Intelligent User Interfaces*, IUI '02, p. 194–195. Association for Computing Machinery, New York, NY, USA, 2002. doi: 10.1145/502716.502753 1

[18] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Text Entry in Immersive Head-Mounted Display-Based Virtual Reality Using Standard Keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 159–166, Mar. 2018. doi: 10.1109/VR.2018.8446059 2

[19] A. Gupta, C. Ji, H.-S. Yeo, A. Quigley, and D. Vogel. RotoSwype: Word-Gesture Typing Using a Ring. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300244 2, 3

[20] A. Gupta, M. Samad, K. Kin, P. O. Kristensson, and H. Benko. Investigating Remote Tactile Feedback for Mid-Air Text-Entry in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 350–360, Nov. 2020. ISSN: 1554-7868. doi: 10.1109/ISMAR50242.2020.00062 4

[21] S. Han, B. Liu, R. Cabezas, C. D. Twigg, P. Zhang, J. Petkau, T.-H. Yu, C.-J. Tai, M. Akbay, Z. Wang, A. Nitzan, G. Dong, Y. Ye, L. Tao, C. Wan, and R. Wang. MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality. *ACM Trans. Graph.*, 39(4), July 2020. Place: New York, NY, USA Publisher: Association for Computing Machinery. doi: 10.1145/3386569.3392452 1

[22] J. Henderson, J. Ceha, and E. Lank. STAT: Subtle Typing Around the Thigh for Head-Mounted Displays. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3379503.3403549 3

[23] F. Kern, F. Niebling, and M. E. Latoschik. Text input for non-stationary xr workspaces: Investigating tap and word-gesture keyboards in virtual and augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2658–2669, 2023. doi: 10.1109/TVCG.2023.3247098 2, 3, 4

[24] P. O. Kristensson. *Discrete and Continuous Shape Writing for Text Entry and Control*. PhD Thesis, Institutionen för datavetenskap, 2007. 2

[25] P. O. Kristensson. Five Challenges for Intelligent Text Entry Methods. *AI Magazine*, 30(4):85–85, 2009. 2

[26] P. O. Kristensson. Next-Generation Text Entry. *Computer*, 48(7):84–87, July 2015. Conference Name: Computer. doi: 10.1109/MC.2015.185 2, 8

[27] P. O. Kristensson. Statistical Language Processing for Text Entry. In A. Oulasvirta, P. O. Kristensson, X. Bi, and A. Howes, eds., *Computational Interaction*. Oxford University Press, Oxford, UK, 2018. 4

[28] P. O. Kristensson and K. Vertanen. Performance Comparisons of Phrase Sets and Presentation Styles for Text Entry Evaluations. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, IUI '12, p. 29–32. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2166966.2166972 9

[29] P. O. Kristensson and K. Vertanen. The Inviscid Text Entry Rate and Its Application as a Grand Goal for Mobile Text Entry. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, MobileHCI '14, p. 335–338. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2628363.2628405 9

[30] P.-O. Kristensson and S. Zhai. SHARK^2: A Large Vocabulary Shorthand Writing System for Pen-Based Computers. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, p. 43–52. Association for Computing Machinery, New York, NY, USA, 2004. doi: 10.1145/1029632.1029640 1, 2, 6

[31] P.-O. Kristensson and S. Zhai. Relaxing Stylus Typing Precision by Geometric Pattern Matching. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, IUI '05, p. 151–158. Asso-

ciation for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1040830.1040867 1

[32] Y. Lee and G. J. Kim. Vitty: Virtual Touch Typing Interface with Added Finger Buttons. In S. Lackey and J. Chen, eds., *Virtual, Augmented and Mixed Reality*, pp. 111–119. Springer International Publishing, Cham, 2017. 3

[33] L. A. Leiva, S. Kim, W. Cui, X. Bi, and A. Oulasvirta. How we swipe: A large-scale shape-writing dataset and empirical findings. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, MobileHCI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3447526.3472059 2, 3, 7

[34] X. Lu, D. Yu, H.-N. Liang, and J. Goncalves. IText: Hands-Free Text Entry on an Imaginary Keyboard for Augmented Reality Systems. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, p. 815–825. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474788 3

[35] I. S. MacKenzie and R. W. Soukoreff. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, p. 754–755. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/765891.765971 7

[36] A. Markussen, M. R. Jakobsen, and K. Hornbaek. Vulture: A Mid-air Word-gesture Keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 1073–1082. ACM, New York, NY, USA, 2014. doi: 10.1145/2556288.2556964 2, 3, 5, 9

[37] A. K. Mutasim, A. U. Batmaz, M. Hudhud Mughrabi, and W. Stuerzlinger. Does Repeatedly Typing the Same Phrase Provide a Good Estimate of Expert Text Entry Performance? In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544549.3585647 6

[38] T. Ouyang, D. Rybach, F. Beaufays, and M. Riley. Mobile Keyboard Input Decoding with Finite-State Transducers. *CoRR*, abs/1704.03987, 2017. _eprint: 1704.03987. 4

[39] T. Paek and B.-J. P. Hsu. Sampling Representative Phrase Sets for Text Entry Experiments: A Procedure and Public Resource. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 2477–2480. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1979304 9

[40] S. Reyal, S. Zhai, and P. O. Kristensson. Performance and User Experience of Touchscreen and Gesture Keyboards in a Lab Setting and in the Wild. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, p. 679–688. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2702123.2702597 2, 6

[41] M. Richardson, M. Durasoff, and R. Wang. Decoding Surface Touch Typing from Hand-Tracking. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, pp. 686–696. Association for Computing Machinery, New York, NY, USA, Oct. 2020. doi: 10.1145/3379337.3415816 3

[42] J. Shen, J. Dudley, and P. O. Kristensson. Simulating Realistic Human Motion Trajectories of Mid-Air Gesture Typing. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 393–402, 2021. doi: 10.1109/ISMAR52148.2021.00056 2

[43] J. Shen, J. Dudley, and P. O. Kristensson. Fast and Robust Mid-Air Gesture Typing for AR Headsets Using 3D Trajectory Decoding. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 2

[44] J. Shen, J. Hu, J. J. Dudley, and P. O. Kristensson. Personalization of a Mid-Air Gesture Keyboard using Multi-Objective Bayesian Optimization. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 702–710, 2022. doi: 10.1109/ISMAR55827.2022.00088 2

[45] M. Speicher, A. M. Feit, P. Ziegler, and A. Krüger. Selection-based text entry in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3174221 3

[46] F. Spiess, P. Weber, and H. Schuldt. Direct Interaction Word-Gesture Text Input in Virtual Reality. In *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 140–143, 2022. doi: 10.1109/AIVR56993.2022.00028 3

[47] K. Vertanen, D. Gaines, C. Fletcher, A. M. Stanage, R. Watling, and P. O. Kristensson. Velociwatch: Designing and evaluating a virtual keyboard for the input of challenging text. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300821 9

[48] K. Vertanen and P. Kristensson. Mining, analyzing, and modeling text written on mobile devices. *Natural Language Engineering*, 27(1):1–33, 2021. Publisher: Cambridge University Press. doi: 10.1017/S1351324919000548 4

[49] K. Vertanen and P. O. Kristensson. A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pp. 295–298. Association for Computing Machinery, New York, NY, USA, Aug. 2011. doi: 10.1145/2037373.2037418 5, 6

[50] K. Vertanen and P. O. Kristensson. Complementing Text Entry Evaluations with a Composition Task. *ACM Trans. Comput.-Hum. Interact.*, 21(2), feb 2014. doi: 10.1145/2555691 9

[51] K. Vertanen, H. Memmi, J. Emge, S. Reyal, and P. O. Kristensson. VelociTap: Investigating Fast Mobile Text Entry Using Sentence-Based Decoding of Touchscreen Keyboard Input. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, p. 659–668. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2702123.2702135 1, 4

[52] J. Walker, B. Li, K. Vertanen, and S. Kuhl. Efficient Typing on a Visually Occluded Physical Keyboard. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 5457–5461. ACM, New York, NY, USA, 2017. doi: 10.1145/3025453.3025783 2

[53] Y. Wang, Y. Wang, J. Chen, Y. Wang, J. Yang, T. Jiang, and J. He. Investigating the Performance of Gesture-Based Input for Mid-Air Text Entry in a Virtual Environment: A Comparison of Hand-Up versus Hand-Down Postures. *Sensors*, 21(5):1582, Jan. 2021. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute. doi: 10.3390/s21051582 2, 3

[54] W. Xu, H.-N. Liang, A. He, and Z. Wang. Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 279–288, 2019. doi: 10.1109/ISMAR.2019.00026 3

[55] N. Yanagihara, B. Shizuki, and S. Takahashi. Text Entry Method for Immersive Virtual Environments Using Curved Keyboard. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, pp. 1–2. Association for Computing Machinery, New York, NY, USA, Nov. 2019. doi: 10.1145/3359996.3365026 2, 3

[56] X. Yi, C. Yu, M. Zhang, S. Gao, K. Sun, and Y. Shi. ATK: Enabling Ten-Finger Freehand Typing in Air Based on 3D Hand Tracking Data. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, UIST '15, p. 539–548. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2807442.2807504 4

[57] S. J. Young, N. H. Russell, and J. H. S. Thornton. Token Passing: a Simple Conceptual Model for Connected Recognition Systems. 1989. Publisher: University of Cambridge, Dept. of Engineering. 4

[58] C. Yu, Y. Gu, Z. Yang, X. Yi, H. Luo, and Y. Shi. Tap, Dwell or Gesture?: Exploring Head-Based Text Entry Techniques for HMDs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 4479–4488. ACM, New York, NY, USA, 2017. doi: 10.1145/3025453.3025964 2, 3, 5, 9

[59] S. Zhai and P.-O. Kristensson. Shorthand Writing on Stylus Keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, p. 97–104. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/642611.642630 1, 2

[60] S. Zhai and P. O. Kristensson. The Word-gesture Keyboard: Reimagining Keyboard Interaction. *Communications of the ACM*, 55(9):91–101, Sept. 2012. doi: 10.1145/2330667.2330689 2

[61] S. Zhai, P. O. Kristensson, P. Gong, M. Greiner, S. A. Peng, L. M. Liu, and A. Dunnigan. Shapewriter on the Iphone: From the Laboratory to the Real World. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, p. 2667–2670. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1520340.1520380 2