Performance Envelopes of Virtual Keyboard Text Input Strategies in Virtual Reality

John J. Dudley*
University of Cambridge

Hrvoje Benko†
Facebook Reality Labs

Daniel Wigdor‡
University of Toronto

Per Ola Kristensson§
University of Cambridge

ABSTRACT

Virtual and Augmented Reality deliver engaging interaction experiences that can transport and extend the capabilities of the user. To ensure these paradigms are more broadly usable and effective, however, it is necessary to also deliver many of the conventional functions of a smartphone or personal computer. It remains unclear how conventional input tasks, such as text entry, can best be translated into virtual and augmented reality. In this paper we examine the performance potential of four alternative text entry strategies in virtual reality (VR). These four strategies are selected to provide full coverage of two fundamental design dimensions: i) physical surface association; and ii) number of engaged fingers. Specifically, we examine typing with index fingers on a surface and in mid-air and typing using all ten fingers on a surface and in mid-air. The central objective is to evaluate the human performance potential of these four typing strategies without being constrained by current tracking and statistical text decoding limitations. To this end we introduce an auto-correction simulator that uses knowledge of the stimulus to emulate statistical text decoding within constrained experimental parameters and use high-precision motion tracking hardware to visualise and detect fingertip interactions. We find that alignment of the virtual keyboard with a physical surface delivers significantly faster entry rates over a mid-air keyboard. Also, users overwhelmingly fail to effectively engage all ten fingers in mid-air typing, resulting in slower entry rates and higher error rates compared to just using two index fingers. In addition to identifying the envelopes of human performance for the four strategies investigated, we also provide a detailed analysis of the underlying features that distinguish each strategy in terms of its performance and behaviour.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality

1 INTRODUCTION

Text entry is a fundamental human-computer interaction task [2]. Even in novel interaction environments, such as those enabled by Virtual and Augmented Reality (VR/AR), text entry is an essential feature for synchronous and asynchronous communication, annotation and documentation. The delivery of seamlessly integrated and efficient text entry methods can potentially improve engagement and sense of presence by avoiding the need to switch between input devices or platforms. However, how to best deliver a productive and enjoyable method for entering text in such environments remains an open research question.

Recent advances in speech recognition have increased the popularity of voice transcription as a text entry method. Speech entry rates are fast and recent technical advances mean that accuracy rates

*e-mail: jjd50@cam.ac.uk
†e-mail: benko@fb.com
‡e-mail: daniel@dgp.toronto.edu
§e-mail: pok21@cam.ac.uk

are also comparable with conventional text entry methods [26]. However, privacy considerations and ambient noise mean that speech-to-text is not always viable. In reality, voice and touch-based text entry are complementary. Ultimately, a robust text entry solution for AR/VR will likely be delivered through a range of different and complementary input methods. Delivering a touch-based text input method that is familiar to users and leverages existing typing skills is therefore a desirable feature in AR/VR.

This paper reports on an exploratory study examining the human performance envelopes, that is, the feasible range of text entry rates and error rates, of four alternative touch-based typing configurations in VR. We report results from a controlled experiment with 24 participants that examines two fundamental design parameters: 1) aligning the keyboard with a physical surface compared to having the keyboard float in mid-air; and 2) typing with all ten fingers compared to just the two index fingers. This investigation thus concentrates on two fundamental factors likely to reflect the different circumstances of use of a virtual keyboard.

The central objective is to understand the empirical human performance potential of particular text entry strategies, independent of current device and software limitations. Therefore, we are motivated to minimise tracking and statistical text decoding performance as factors in the experiment, as current state-of-the-art tracking and statistical text decoding performance would effectively result in an artificial ceiling effect on text entry rates. To address this concern, our VR typing setup uses precision finger tracking provided by an OptiTrack motion capture system and robust auto-corrections delivered through a simulated statistical text decoding strategy (based on relaxed point-based matching [15]). The focus on VR over AR is also motivated by efforts to control for confounding variables in the experiment. However, many of the investigated principles are anticipated to be directly relatable across target display environments (see Section 9 for further discussion).

In addition to investigating potential entry and error rates, the recording of precision fingertip tracking data allows us to examine more subtle micro metrics of performance and behaviour. For example, we examine: touch accuracy variation over the layout; variation of mistypes over the layout; press depth, duration and velocity; as well as hand and finger usage proportions. These micro metrics assist in refining our understanding of touch-based typing requirements in two important ways. First, understanding the behaviour of the fastest typists helps formulate reasonable minimum requirements for tracking fidelity. Second, understanding what behaviours yield high entry rates and low error rates can inform the design of the layout and interactions in order to guide users towards more optimal typing behaviour.

The primary contributions of this paper are: 1) a quantitative evaluation of the performance potential of four feasible touch-based keyboard text input strategies for VR covering two key design dimensions; and 2) a provisional set of indicative micro metrics of performance and behaviour that inform the design of a fully functional keyboard. In highlighting the above contributions, we begin by first reviewing the related work on touch-based typing in virtual and augmented reality. We then describe the system and apparatus used in the controlled experiment before detailing the experimental protocol. The key results of the experiment are highlighted and then

qualified and discussed. Finally, we summarise the main results and revisit the contributions in the context of the broader objective of delivering a productive and enjoyable text entry system tailored to VR/AR.

2 RELATED WORK

In this section we examine the literature relevant to enabling productive text input in VR. The research in this area is particularly interesting given the very broad range of strategies explored. Early work in this area experimented with handwritten notes (e.g. [24]) and audio annotations (e.g. [10, 28]). The potential of glove-sensed hand gestures (e.g [16, 25]) has also been widely explored. Bowman et al. [1] investigated the relative merits of these and other approaches by examining speech, glove, pen and chording keyboard approaches in a single experiment: entry rate results were 13 wpm, 6 wpm, 10 wpm and 4 wpm respectively. Speech-to-text has advanced significantly over the past decade and now provides a viable and widely implemented input strategy for head-mounted displays (HMDs). We therefore do not give further focus to speech-based text entry research.

To help compartmentalise these various approaches and contextualise their relative advantages and disadvantages, we apply the following categorisation: virtual Qwerty keyboards; non-Qwerty layouts; and input device/glove based approaches. Also relevant to this study is work which examines the more rudimentary behaviours of how people type and these are examined at the end of this section.

The familiarity of the standard Qwerty layout strongly motivates its use in AR/VR settings. The significant challenge becomes how to effectively capture input on that layout. ARKB [17] describes an early implementation for vision based tracking of fingertips enabling multi-finger typing in AR. Tracking accuracy and latency were noted to be major challenges to usability given the technology limitations at the time. Leveraging significant technology advancements, ATK [35] makes use of the LeapMotion to demonstrate a full 10 finger mid-air keyboard supported by a probabilistic decoder. Participants achieved 29 wpm after one hour of practice although stimulus phrases were selected to ensure only words in the known vocabulary were included. VISAR [4] also leverage probabilistic decoding, in an approach derived from Vertanen et al. [31], and demonstrates single-finger mid-air text input specifically tailored for AR HMDs. After various refinements, including the provision of error-tolerant word predictions, the touch-based approach yielded a mean entry rate of 17.8 wpm. Although focussing on interaction with large wall displays, Markussen et al. [20, 21] examine both discrete and gesture-based approaches for mid-air text entry.

The challenges of delivering robust touch-based interaction with a virtual keyboard has also motivated the investigation of alternative articulation strategies. Yu et al. [36] compare tap selection on a gamepad, gaze-dwell and gaze-gesture articulation strategies for typing in VR: entry rates achieved were 10.6, 15.6, and 19.0 wpm respectively. With further refinement of the gaze-gesture approach, participants were able to reach an average entry rate of 24.7 wpm when typing the same 10 phrases repeatedly.

Non-Qwerty layouts have received attention as a way to mitigate restricted input and/or visual space on AR/VR HMDs. For example, Palmtype [33] re-appropriates the palm as a display and interaction surface for a virtual keyboard in AR. This approach builds on the more general body of research demonstrating the benefits of passive haptic feedback for interactions in virtual environments [13, 18]. Both Grossman et al. [7] and Yu et al. [37] examine simplified input strategies that accommodate the limited interaction surface on smart glasses. Other exotic layouts and interaction methods include: a 12 key keyboard with selections made by a combination of taps and slide gestures [23]; and a radial layout rotated using a controller [38]. Eliminating the need for a layout altogether, AirStroke [22] allows users to input characters in mid-air using the Graffiti alphabet. Such

approaches are, however, clearly rate limited, but AirStroke [22] applies a clever strategy of allowing the non-gesturing hand to select word predictions.

Finally, hand held-input devices (e.g. Twiddler [19]) and gloves (e.g. [16, 25]) offer a potential avenue for delivering text input functionality in AR and VR. Several of the studies previously mentioned use game controllers as an alternative means for articulation. While such approaches may be appropriate in certain circumstances and applications, a significant downside is that they encumber the user. Further, users' existing typing skills and keyboard layout awareness are not easily leveraged in these approaches.

More general efforts to better understand and exploit typing performance and behaviour in novel input arrangements also inform this study. Findlater and Wobbrock [6] examine the potential for adaptive keyboard layouts in 10 finger touchscreen typing that update based on observed patterns of behaviour. We also take influence from Sridhar et al. [27] who take a considered approach to understanding dexterity as a precursor to building a mid-air finger articulation based text input system.

In summary, the literature offers a somewhat confusing landscape of different strategies for supporting text entry in AR/VR. It can be difficult to understand the raw potential of these various approaches given the different experimental protocol choices and technical limitations that inevitably colour these results. This factor is, in part, what motivates us to examine high-level design choices using an experimental protocol that is inherently optimistic in determining envelopes of human performance but robustly supports relative comparison between the techniques under investigation within the same experiment. There are clearly many factors which ultimately determine the entry rate potential of a particular input strategy in practical use. Rather than pursuing and demonstrating a 'practical' text entry system at this juncture, we instead take an exploratory approach that will inform subsequent design efforts.

3 APPROACH

This study has three key objectives. These are:

1. Determine the human performance potential of alternative text entry strategies for AR/VR.

2. Capture hand and finger tracking data representative of typical typing behaviour.

3. Mine the recorded tracking data to identify implications for the design and development of a fully functional keyboard and input system tailored to these strategies.

Objectives 1 and 2 above are pursued in parallel. To ensure we examine performance 'potential' and that user typing behaviour is representative of a properly functioning virtual keyboard, an express decision was made to test an 'ideal' system where conventional tracking and statistical text decoder limitations are removed. Therefore, we employ precise marker-based tracking (OptiTrack) and introduce a simulated auto-correction strategy.

Clearly the elected approach yields an optimistic assessment given that currently available low-cost head-mounted or remote finger tracking technology cannot achieve the accuracy levels of an OptiTrack setup. Furthermore, the effectiveness of our simulated auto-corrections may exceed the performance of a conventional statistical text decoder naively applied. Nevertheless, the approach does effectively inform the development of next-generation text entry methods for AR/VR by: i) allowing us to determine which strategies are worthy of practical examination under conventional device limitations; and ii) highlighting technical requirements for tracking and statistical text decoder components to enable high levels of typing performance. The pursuit of objective 3 above informs our understanding of this second point. This analytical approach is inspired in part by prior work performed by Feit et al. [5] and Dhakal et al. [3].
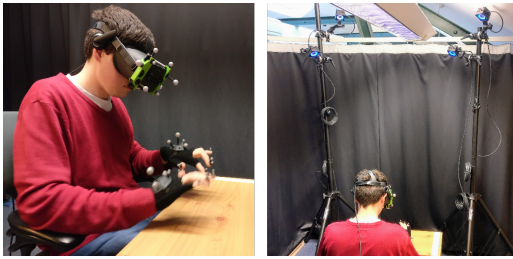
Figure 1: User shown typing in mid-air with HMD, hand and fingertip tracking markers (left). OptiTrack camera setup for precision marker tracking (right).
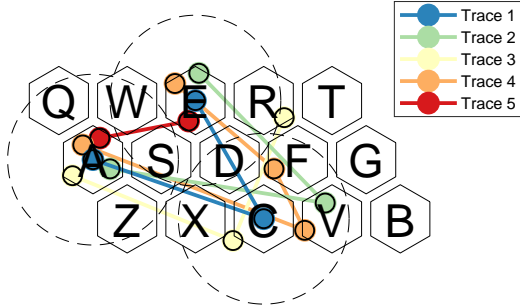

Figure 2: Five example observation sequences (traces) for typing the word 'ACE'. Trace 1 is the target sequence (ideal observations) where the centre of every key is hit. Trace 2 is a good observation sequence in that all observations are within the tolerance of the targets, even though 'V' is actually struck instead of 'C'. Trace 3 is a bad observation sequence since the last observation is outside the tolerance for the target 'E'; this is a substitution error (substitution edit required). Trace 4 is a bad observation sequence since four points are observed; this is an insertion error (deletion edit required). Trace 5 is a bad observation sequence since only two points are observed; this is an omission error (insertion edit required).

4 TEST BED FOR HIGH PERFORMANCE TEXT ENTRY IN VR

In preparation for the controlled user experiment, we developed a test bed for examining text entry strategies delivering high-precision finger tracking and the illusion of robust auto-corrections. These two main system components, in addition to the virtual environment in which they are embedded, are described in detail in the following sections.

4.1 Finger Tracking

Precision (sub-millimetre) fingertip tracking is provided by an Op-tiTrack motion capture system (using Prime 13 cameras). A rigid markerset is attached to the back of each hand to provide position and orientation tracking. Individual markers are then temporarily attached to participant fingertips (on the fingernail). The HMD is also tracked using a separate rigid-body markerset. The experimental setup is shown in Figure 1. The position and orientation of each palm is coarsely represented by the purple rectangular prisms shown in Figure 4. The fingertip positions are represented by purple spheres.

4.2 Simulated Auto-Correction

The behaviour of a standard error correcting statistical text decoder is approximately replicated by performing point-based matching. This approach is introduced by Kristensson and Zhai [15]. The point-based matching procedure determines the number of substitutions, insertions or deletions required to align the observation sequence with the target sequence. Importantly, however, it is possible to apply
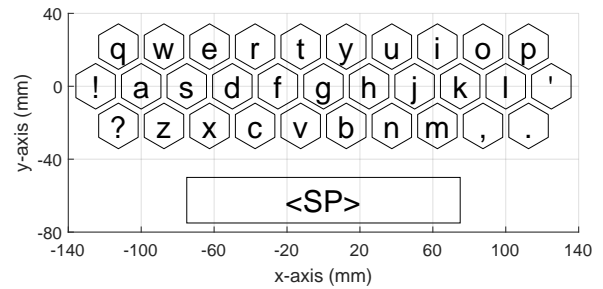

Figure 3: The keyboard layout used in the experiment. Note the reduced set of punctuation.

Figure 4: The keyboard, hands and work environment as viewed in the VR headset.

a tolerance on what is considered a successful match. Based on pilot studies we identified a suitable tolerance of $2.5\times$ the nominal key radius. Several example traces illustrating this approach are presented in Figure 2.

It is important to note that this approach only works because participants must type known preset stimulus phrases. To mimic the behaviour of an auto-correcting decoder, the known words in the stimulus phrase are supplied to the simulated auto-correction component. The latest observation points are sent to the decoder upon particular input events, e.g. space and punctuation (other than apostrophe). If at least 80% of the observation sequence matches the target sequence for a given word in the stimulus phrase, it will be substituted as an auto-correction. Clearly this penalises shorter words, however, such is also the behaviour of a standard statistical text decoder given limited observation points. Once a word in the current stimulus phrase is substituted, it is removed from the list used to evaluate subsequent observation sequences.

4.3 Virtual Environment and Keyboard

A virtual work environment was constructed to provide a thematically relevant context for the text entry task. This environment featured a simple wooden desk against a painted wall. The virtual keyboard and work desk are visible in Figure 4. The surface of the virtual table was aligned with the surface of a physical table in the experiment space. The table can be seen in Figure 1.

A full Qwerty virtual keyboard was designed with all 26 characters and a reduced set of punctuation (,'?!.). Keys are placed with compact tessellation, with each key having an apparent diameter and separation of approximately 25 mm. The top row of keys (Q–P) therefore has an apparent width of 250 mm making it roughly 30% wider than the top row of a typical physical keyboard (190 mm). The two-dimensional keyboard layout is illustrated in Figure 3.

Keyboard touch events are generated when a spherical collider attached at the fingertip marker location first intersects with the keyboard detection plane. The collider attached at each fingertip marker site has a fixed size since we do not perform online association
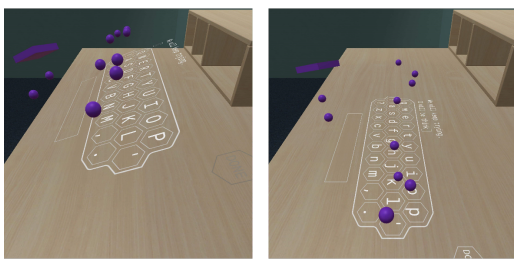
Figure 5: Keyboard in mid-air (left) and aligned with table (right).

of markers to specific fingertips. Note that to generate subsequent touches with the same fingertip, the collider must completely leave and re-enter the detection plane. A simple visual animation at the touch point, synchronised with an audible click sound, provides feedback indicating a generated touch event.

The stimulus phrase is shown in the top row above the keyboard. Entered text is shown immediately below this. For the purpose of the experimental task, a *DONE* key is included for users to press when their entry is complete. The interface layout experienced by participants can be seen in Figure 4. Figure 5 illustrates the positioning of the keyboard in the mid-air and surface-aligned conditions.

Importantly there is no backspace or delete key. As described later, participants were given no opportunity to correct errors.

5 EXPERIMENT PROTOCOL

The experiment required participants to complete a text transcription task. This task was performed in the following four typing conditions:

- MA2: Mid-air, two (index) fingers only

- SUR2: Aligned with physical surface, two (index) fingers only

- MA10: Mid-air, all fingers

- SUR10: Aligned with physical surface, all fingers

After obtaining ethics approval for the study, we recruited 24 participants (10 female, 14 male, median age = 25) through opportunity sampling. Special effort was taken to ensure diversity in the field of study/work of participants. The condition order was fully balanced to address potential learning effects (i.e. no two participants experienced the same order of conditions). The experiment was split into two sessions, with each session examining two of the four conditions. Participants were required to perform these sessions on separate days but with no more than two days break between sessions. A single session would typically run for between 1.5 and 2 hours, resulting in a total experiment time of between 3 and 4 hours. As part of the participant briefing, participants were instructed to notify the researcher if they experienced any VR induced nausea so that the experiment could be suspended. Note that this situation did not arise.

The experiment controlled for posture by enforcing a seated position. In addition, participants were not permitted to rest any part of their hand or arm on the table in the mid-air conditions, but were free to do so in the surface aligned conditions.

Within each condition, participants were presented with 10 practice sentences and 160 test sentences. During the practice sentences, participants were encouraged to attempt different typing strategies and to develop an understanding of the keyboard behaviour.

The 160 test sentences were split into four blocks of 40 sentences with the opportunity for a short break between each. Stimulus sentences were taken from the extended Enron mobile message dataset [29] and filtered based on phrases containing four words or
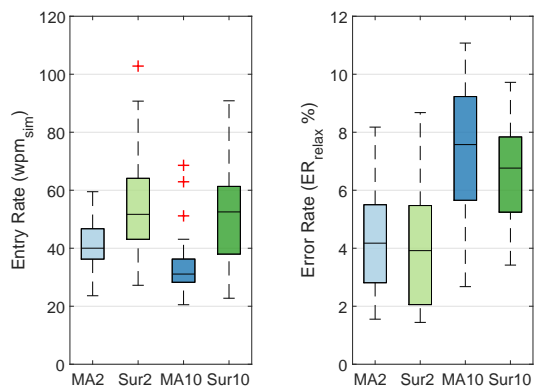

Figure 6: Boxplots of participant mean entry rate (left) and relaxed error rate (right). In this and subsequent boxplots, red crosses indicate outliers based on $q_{1/3} \pm 1.5 \times (q3 - q1)$.
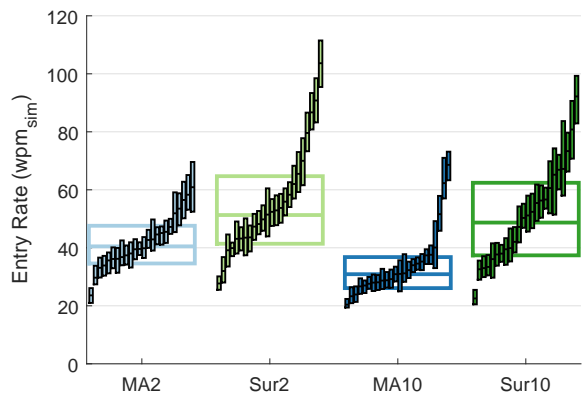

Figure 7: Plot shows individual participant q_1 (first quartile), median and q_3 (third quartile) entry rates sorted by median entry rate to help better illustrate the structure of the distribution. Note that the plot only includes entries where the error rate is below 10%. Aggregate q_1, median and q_3 across all participants for an individual condition are also shown as the outer bars (coloured lines with white fill).

more, and 40 characters or less. Stimulus phrases were selected from this subset without replacement, such that participants never saw the same sentence twice. In summary, a total of $(24p \times 4c \times 160s)$ 15,360 test entries were captured in this experiment.

To remove error correction time as a confounding factor in the experiment, no backspace or deletion functionality was provided by the keyboard. Participants were instructed to type as accurately as possible, but in the event of an error, to continue typing as if no mistake had been made.

6 RESULTS

The results of the controlled experiment are summarised in this section. We begin by examining the human performance potential of the four conditions in terms of entry and error rates. We then examine the various micro metrics that yield a greater understanding of underlying factors that explain user performance and behaviour. Finally, we review the participants' qualitative feedback and general observations of typing behaviour in VR.

6.1 Performance Potential

The key metrics describing performance in text entry are entry and error rate. The standard metric for entry rate is words per minute (wpm), that is, number of words entered divided by time taken. In practice, the numerator is an effective word count where a nominal word length of five 'keystrokes' is used (including spaces). Therefore, we use the entered phrase length minus one (since we start timing from the first touch) dived by five. To highlight the distinction between the standard assessment of entry rate and our investigation incorporating simulated auto-corrections we introduce the measure, wpm_{sim}.

Error rate is typically reported as Character Error Rate (CER), which is the minimum number of character insertion, deletion and substitution operations that transform the stimulus text into the response text, divided by the length of the response text. However, given the behaviour of the simulated auto-corrections it is more appropriate to report error rates in terms of their geometric trace match. Therefore, we report the relaxed geometric match error rate as ER_{relax}. ER_{relax} reflects the number of required edits normalised by the length of the observation sequence. The numerator is the count of substitutions, insertions or deletions required to align the observation sequence with the target sequence given a tolerance of $2.5\times$ the nominal key radius on each target key (this is consistent with the simulated auto-correction procedure outlined in Section 4.2). The denominator is simply the length of the observation sequence.

The entry and error rate results for all captured entries are summarised in Figure 6. Entry rates are observably higher in the on-surface conditions (SUR2, mean = 55.6 wpm_{sim} and SUR10, mean = 51.6 wpm_{sim}) than in the mid-air conditions (MA2, mean = 42.1 wpm_{sim} and MA10, mean = 34.5 wpm_{sim}). Using a repeated measures analysis of variance we find a significant effect for the keyboard test condition on entry rate ($F_{3,23} = 29.370$, $\eta_p^2 = 0.561$, $p < 0.05$). Using an initial significance level of $\alpha = 0.05$ and performing multiple comparisons with a Bonferroni correction (note that all subsequent reported multiple comparisons use this same procedure) shows a significant difference between all conditions except for between SUR2 and SUR10. This result suggests that physical surface alignment is an important factor in producing high entry rates. Section 6.2 will later examine the lower-level features that relate the presence of a physical surface to typing performance.

Interestingly and somewhat counter-intuitively, the ten finger conditions (MA10 and SUR10) do not yield significantly faster entry rates than their two finger alternatives. In fact, having ten fingers in mid-air appears to be detrimental to performance. This result correlates with the significantly higher error rates in the ten finger conditions ($F_{3,23} = 31.431$, $\eta_p^2 = 0.577$, $p < 0.05$). The effect is significant between the two and ten finger conditions but not within each. Although the experiment protocol did not enforce corrections, high error rates will typically lead to a negative impact on uncorrected entry rates: users pause to re-evaluate their place in the phrase and/or make more careful and precise movements to avoid further errors.

Figure 7 provides an alternative perspective on the entry rate results. Here the interquartile range is plotted for each participant. Only entries where the error rates was below 10% are included in this plot. Note that within each condition the plot order is sorted based on participant median to better illustrate the distribution. It is interesting to note the clear upper tail effect is more prevalent in certain conditions. This will be examined in more detail later in Section 7 when the metrics of the high and low performing participants are analysed.

6.2 Micro Metrics of Performance and Behaviour

This section examines a collection of lower-level features that are key determinants of typing entry and error rate. These features are subsequently referred to as micro metrics of performance and
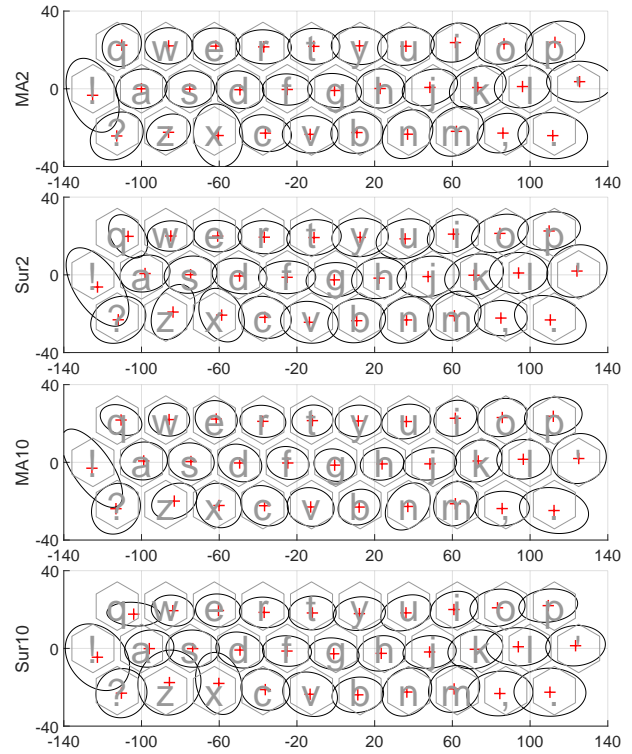

Figure 8: Touch point covariance for each key over the layout represented as ellipses. Ellipses describe the 50% confidence interval.

behaviour.

These features help reveal what aspects of the typing task are most influenced by the different conditions. For example, in the following section we examine the accuracy of touches over the layout. We find that higher accuracy is achieved in the two-finger conditions. Conversely, 10 finger typing yields less accurate touches and this result correlates closely with heightened error-rates identified for conditions MA10 and SUR10 .

6.2.1 Touch Accuracy

Figure 8 provides a summary representation of the touch accuracy variation over the layout. Note that these plots are generated from entries where the error rate was below 10% to ensure reasonable confidence in the realignment of the ideal and observed sequence. The ellipse on each key reflects the centroid and covariance of the touches associated with that key. Recall that the relaxed point-based matching used in delivering the simulated auto-correction behaviour meant that users could touch outside the bounds of the target key and still experience a successful auto-correction provided it was within the distance threshold.

Several interesting observations can be made from Figure 8. First, in all conditions the variance in touch error tends to be higher in the x-axis than in the y-axis. This feature is suggestive of more precise finger articulation (i.e. to switch between key row) than wrist and/or forearm articulation (i.e. to move over the layout laterally). When all touches are collapsed together, the standard deviation in the x-direction is approximately double that in the y-direction.

Second, Figure 8 highlights the fact that touches are more precise at the centre of the keyboard than at the edges. The additional variation in touch error towards the edges generally appears to radiate away from the very centre of the keyboard. One likely interpretation of this result is the fact that typical strategies in standard typing
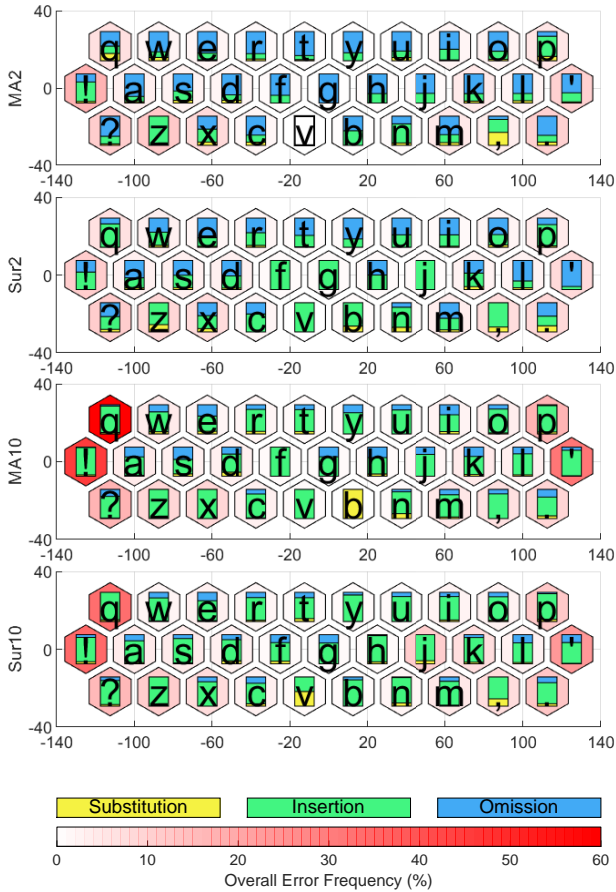
Figure 9: Relative proportions of standard mistypes for each key over the layout. The overall frequencies of key mistypes as a proportion of total presses for that key are represented by the red shading.

Figure 10: Boxplots of participant mean inter-key interval.

Figure 11: An illustrative example of *P22* typing the phrase 'How are things with you?' with two index fingers in SUR2. Purple trace is left index finger, green is right.

involve maintaining the fingers in an approximate 'home' position. We conjecture that moving fingertips away from their 'home' position at high velocity may be introducing this 'smearing' effect on touch error radiating outwards.

6.2.2 Touch Errors: Substitutions, Insertions, Omissions

In this section, the distribution of common typing errors over the keyboard layout is examined. Understanding any relationship between key position and/or typing configuration may inform alternative strategies for addressing such errors. Standard mistypes fall into three categories: substitutions—an incorrect key is pressed; insertions—an additional undesired key is pressed; and omissions—a desired key is not pressed.

Figure 9 illustrates the frequency of the three main mistype categories over the layout for the four conditions. A frequently observed mistype among participants in the ten finger conditions was the pinky finger inadvertently generating key presses at the extreme edges of the layout. This is observable in Figure 9 as a high proportion of insertions on keys *Q!?P'* for conditions MA10 and SUR10.

Another common mistype observed, but less visible in Figure 9, are omissions on commonly doubled characters such as *T*, *L*, and *O*. This error stems from participants failing to raise their finger sufficiently high to exit and re-enter the detection plane. This particular issue is investigated in more detail later in subsection 6.2.7.

The most obvious distinction between the 2 finger conditions and the ten finger conditions is the dominant mistype being omissions for 2 fingers and insertions for ten fingers. This result is consistent with
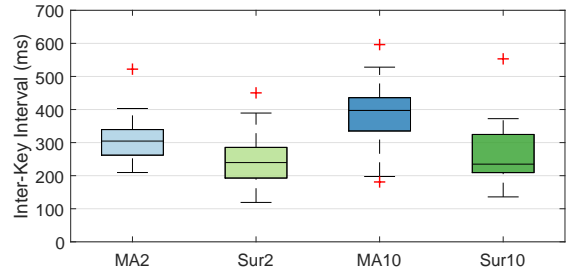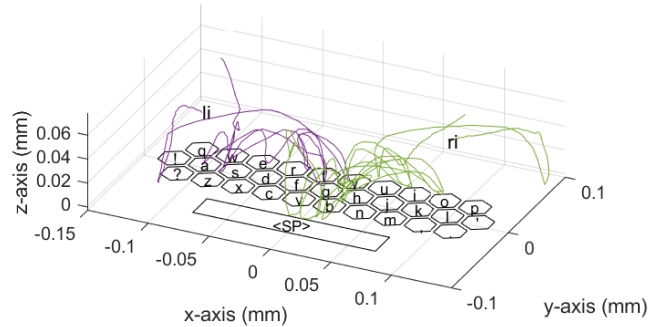
the higher error rates observed and general difficulty participants had in avoiding inadvertent touches with other fingers.

6.2.3 Inter-key interval (IKI)

The inter-key interval (IKI) metric reflects the time between key presses. It therefore correlates closely with entry rate. Figure 10 summarises the IKI in each of the four test conditions.

Repeated measures analysis of variance shows the test condition effect to be significant ($F_{3,23} = 48.318$, $\eta_p^2 = 0.678$, $p < 0.05$). The differences are significant between MA2 and all other conditions and MA10 and all other conditions. In other words, significantly faster IKIs were observed in the on-surface conditions (with interquartile ranges of approximately 200 to 300 ms). This is consistent with the faster entry rate results for these conditions. More time taken between key presses for MA2 (median of approximately 300 ms) and MA10 (median of approximately 400 ms) is correspondingly a significant contributor to the slower entry rates for these mid-air conditions.

It is likely that this additional time taken to transition between keys in the mid-air conditions is in part a result of the longer trajectory followed by the fingers. For on-surface typing the height of the fingertip is comparatively simple to regulate given the potential to rest the palm of the hand on the physical surface. By contrast, mid-air typing involves more challenging depth regulation given the lack of a fixed surface reference plane. The implication of this difference on press depth is examined later in subsection 6.2.6.

6.2.4 Fingertip Trajectory

The captured tracking data enables the examination of lower-level features describing the fingertip trajectories in executing the typing task. Figure 11 provides an illustrative plot of the path traced by the tip of each index finger in *P22*'s execution of the phrase,
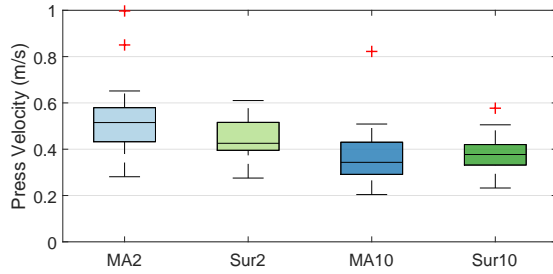
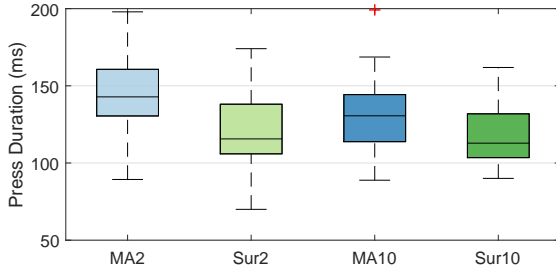Figure 12: Boxplots of participant mean press velocity.

Figure 13: Boxplots of participant mean press duration.

Figure 14: The z-offset resolved into the keyboard frame for the trace shown in Figure 11. The depth, frequency and duration of touches can be easily observed. LI: left index (middle), RI: right index (bottom).

Figure 15: Boxplots of participant mean press depth.

'How are things with you?' Figure 11 highlights the complex co-ordinated movement of fingers while typing. A key objective of such analysis is identifying features that might help discriminate between re-positioning (i.e. preparing for a key press) and striking (i.e. executing a key press) motion of the finger. To this end, we examine mean fingertip velocity as the touch event is first initiated in all typing conditions.

Figure 12 summarises the participant mean press velocity in each condition. A significant effect of test condition on fingertip velocity at touch time is observed ($F_{3,23} = 22.383$, $\eta_p^2 = 0.493$, $p < 0.05$). The differences are significant between MA2 and all other conditions and SUR2 and all other conditions. This result highlights the fact that the fingertip is travelling significantly faster when touches are generated in the 2 finger conditions than in the 10 finger conditions. This result is intuitive when considering the fact that when only two fingers are available, the motion between target keys must be faster to maintain a given entry rate. The significant difference between MA2 and SUR2 is likely a consequence of the absence of the physical limit and therefore no penalty (i.e. potentially painful striking of the surface with the fingertip) on high speed touches.

6.2.5 Press Duration

The press duration is the period of time spent inside the detection plane when executing a key press. Figure 14 illustrates the z-component (in the direction out of the keyboard plane) of the fingertip trace resolved into the keyboard frame. This z-position resolved in the keyboard frame enables simple analysis of press duration.

Figure 13 summarises the participant mean press durations for each of the test conditions. Shorter presses were observed in the surface-aligned conditions (SUR2, mean = 120.3 ms and SUR10, mean = 118.9 ms) than the mid-air conditions (MA2, mean = 142.9 ms and MA10, mean = 128.7 ms). The effect of the test condition is observed to be significant ($F_{3,23} = 9.017$, $\eta_p^2 = 0.282$, $p < 0.05$). Performing multiple comparisons, a significant difference is observed bet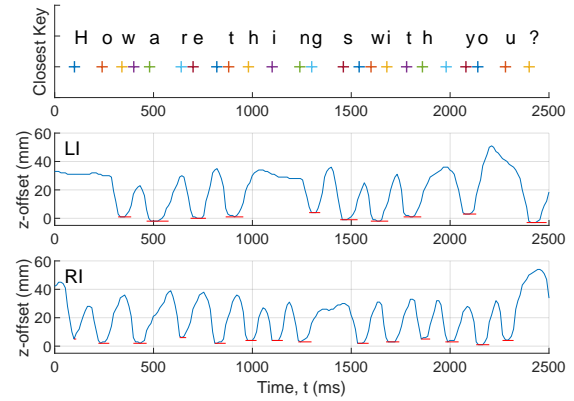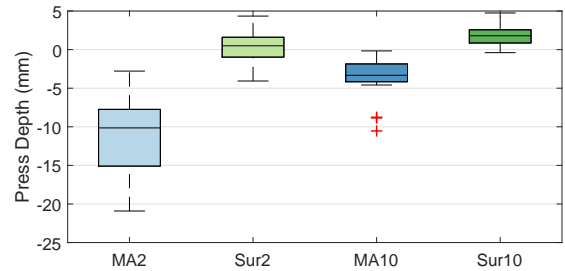ween MA2 and all other conditions. In other words, the presses in the MA2 condition last significantly longer than those in the two surface-aligned conditions as well as the ten finger mid-air condition. It is likely that the longer period spent within the detection plane is a consequence of deeper travel as examined in the following subsection.

6.2.6 Press Depth

The press depth is the maximum distance past the detection plane travelled by the finger. This measure is observable in Figure 14 as the local minimum in the z-offset at each of the touch events.

The mean press depth in each condition is summarised in Figure 15. Clearly the press depth is physically constrained in the surface-aligned conditions. Recall, however, that touch events are raised based on a simple collision detection between a collider attached at the fingertip marker location and the keyboard plane. Since real time association of fingertips is not performed, the same fingertip collider size is used for all fingers. For this reason as well as other potential sources of minor variation (e.g. marker attachment location, finger sizing, finger orientation while pressing), it is possible for touch events to occur before the physical limit is reached. As can be observed in Figure 15, these inadvertent touches are clearly more prevalent in the SUR10 condition.

A repeated measures analysis of variance shows the test condition to be a significant effect ($F_{3,23} = 99.461$, $\eta_p^2 = 0.812$, $p < 0.05$). The difference between all conditions is significant except for between the two surface-aligned conditions. For the mid-air conditions, the depth of touch is considerably larger in MA2 than MA10. One interpretation of this result is that when only the index finger is engaged there is no penalty for deep movements into the detection plane. By contrast, when all fingers are engaged the user must be
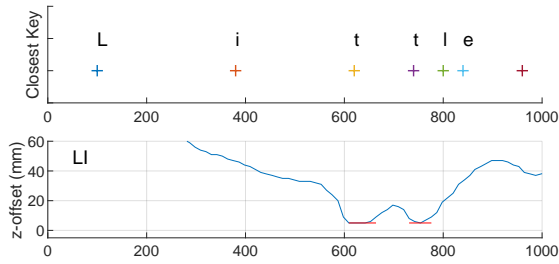
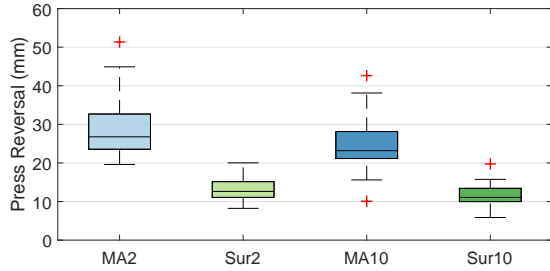Figure 16: Illustration of a double tap executed with the left index finger while typing the word 'little'.

Figure 17: Boxplots of participant mean press reversal.

Figure 18: Mean participant hand and finger usage rate (%).

conscious of not moving other fingers into the detection plane. The deeper penetration into the detection plane for the mid-air conditions has a corresponding impact on press duration as highlighted in the previous subsection.

6.2.7 Press Reversal

As discussed in subsection 6.2.2, a commonly observed mistype was an omission of repeated characters. This motivates examination of the trajectory followed by the finger in such circumstances. In particular, it is useful to determine what distance users will typically lift their fingers in order to indicate a 'press-and-release'. The press reversal therefore describes the minimum distance travelled to generate a double-tap of a repeated key. Figure 16 provides a helpful illustration of this motion. Here the user generates two presses on 't' in order to type 'little' and the press reversal here was 12 mm.

Figure 17 summarises the distribution of mean press reversal distances across the four test conditions. The effect of test condition is significant ($F_{3,23} = 63.887$, $\eta_p^2 = 0.735$, $p < 0.05$). Performing multiple comparisons, the difference is significant between all conditions expect for between the two surface aligned conditions. The significantly shorter press reversals observed in the surface aligned conditions is likely a reflection of the higher degree of control that can be exercised when the palm of the hand is resting on a physical surface. Press reversal distances are highest in the MA2 condition which is consistent with the generally more pronounced movements observed in this conditions and also reflected in the analysis of press velocity and depth.

6.2.8 Hand and Finger Usage

The usage proportions for each hand and finger help give a sense of what typing behaviours are promoted by each of the typing conditions. Clearly, the two finger conditions constrain participants to type with index fingers only, yet understanding right/left dominance can be informative. More relevant, however, is the extent to which participants are able to fully exploit the full complement of ten fingers.
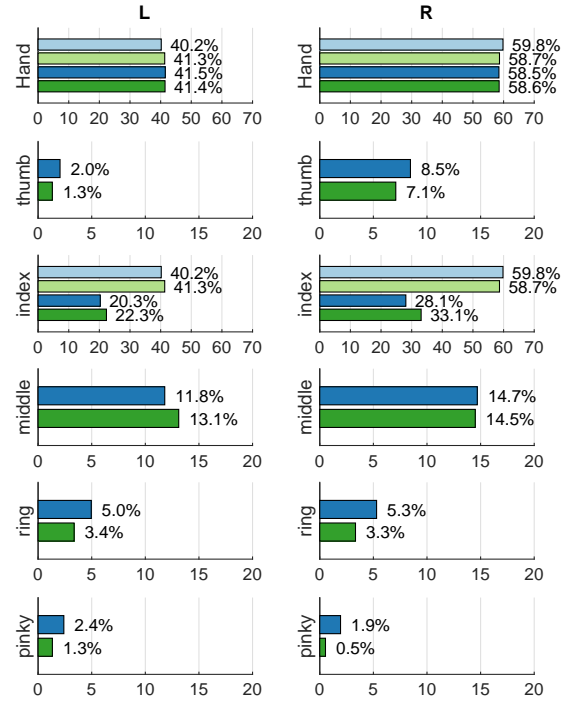
Recall that tracking was performed with passive markers and so there was no real time association of fingertips with markers. Nevertheless, such an association is relatively simple to apply in post-processing given the recorded left/right hand poses.

Figure 18 summarises the usage percentages across each of the test conditions for each touch event. Recall that the two finger conditions used only the index fingers so the hand usage percentage is the same as the finger usage. The index fingers are also dominant in the ten finger conditions, followed by the middle fingers then right thumb (used for space).

The usage percentage of the ring and pinky fingers is higher in MA10 than SUR10. Referring back to the common mistype results presented in subsection 6.2, however, it is likely that this additional involvement of the outer fingers is actually a reflection of inadvertent insertions. Otherwise, the usage distribution in the 10 finger conditions is remarkably similar.

6.3 Qualitative Feedback

After each experimental session, participants completed a short survey asking them to reflect on their experience with the typing conditions. Three statements examined the participant's perception of their speed ('How quickly were you able to type in this condition?'), accuracy ('How accurately were you able to type in this condition?') and comfort ('How comfortable was typing in this condition?') while performing the typing task. Responses were collected on a five-point Likert scale (1: negative; 5: positive). The median response of the 24 participants to these statements are summarised in Table 2.

It is interesting to observe that these median results for speed and comfort correlate well with the recorded entry and error rates. The two surface-aligned conditions (SUR2 and SUR10) received a median rating of 4 for speed and these were also the two fastest conditions in terms of entry rate. The 3 rating for condition MA2 and 2 rating for condition MA10 are also consistent with the entry rate trend observed in the quantitative data. Similarly, the accuracy

Table 1: Comparison of mean performance and behavioural measures for top and bottom performing users. Bold values indicate a significant difference based on an independent two-sample *t*-test at a 5% significance level.

Metric	MA2			SUR2			MA10			SUR10		
	Bot-6	Top-6	Diff.	Bot-6	Top-6	Diff.	Bot-6	Top-6	Diff.	Bot-6	Top-6	Diff.
wpm_{sim}	32.6	54.0	**65.8%**	41.2	82.0	**99.1%**	25.1	49.3	**96.8%**	35.0	69.5	**98.4%**
ER_{relax}	3.7	6.0	**62.9%**	4.0	4.4	10.7%	7.3	8.3	14.4%	7.8	6.7	-14.1%
IKI (ms)	387.8	232.0	**-40.2%**	314.7	154.9	**-50.8%**	495.7	270.6	**-45.4%**	369.1	188.5	**-48.9%**
Press Vel. (m/s)	0.49	0.62	26.5%	0.37	0.53	**42.6%**	0.30	0.48	**61.6%**	0.32	0.43	**32.3%**
Press Dur. (ms)	167.9	130.3	**-22.4%**	131.6	97.0	**-26.3%**	139.5	101.9	**-26.9%**	133.2	101.7	**-23.7%**
Press Depth (mm)	-14.0	-11.9	-14.8%	0.4	0.1	-71.5%	-2.8	-4.1	45.6%	1.7	1.6	-2.7%
Press Rev. (mm)	29.2	28.0	-4.1%	12.2	12.7	4.5%	23.1	25.2	9.2%	11.1	11.0	-1.0%

rating of 4 for the two finger conditions (MA2 and SUR2) is consistent with the lower error rates observed in these conditions. The marginally higher error rates in condition MA10 compared with condition SUR10 is also consistent with the accuracy ratings of 2 and 3 respectively.

In terms of comfort, the on-table conditions were both perceived positively (median ratings of 4). By contrast the perception of comfort for two fingers in mid-air was neutral (3) and negative (2) for ten fingers in mid-air. This result is to be expected given the additional effort required in maintaining the hand cantilevered in space when typing in mid-air.

After completing both experimental sessions, participants were asked to select their preferred test condition. Note that participants were not made aware of their quantitative performance at any stage. The preference counts were 1, 14, 0 and 9 for each of MA2, SUR2, MA10 and SUR10 respectively. This result indicates a clear preference for the virtual keyboard plane being aligned with a physical surface. The subjective preference for the surface aligned conditions is also consistent with the quantitative entry rate results obtained. The most preferred condition overall was SUR2. This demonstrates good correlation between the subjective experience and quantitative performance given that SUR2 also yielded the highest mean entry rate and lowest mean error rate.

7 INDICATORS OF HIGH AND LOW PERFORMANCE

In this section we examine the key metrics that most clearly distinguish high performers from low performers in the typing task. To this end, we compare indicators of performance for the top six participants to the bottom six participants. Such an analysis can help highlight what constitutes 'ideal' typing behaviour.

The top and bottom six participants are each selected based on their mean entry rate across all conditions. Table 1 summarises the entry and error rates for these two groups in each condition and revisits several of the core micro metrics introduced and examined in subsection 6.2.

As expected based on the group selection strategy, the top performers achieve significantly faster entry rates in all conditions. The top performers do, however, exhibit higher error rates in the MA2, SUR2 and MA10 conditions, with significantly higher error rates in the MA2 condition. This reflects a common performance tradeoff of speed for accuracy.

Table 2: Median response in post session survey. Likert scale from 1 to 5. Questions asked participants to reflect on speed, accuracy and comfort in the condition.

Aspect	MA2	SUR2	MA10	SUR10
speed	3	4	2	4
accuracy	3.5	4	2	3
comfort	3	4	2	4

The micro metrics presented in Table 1 highlight how the top group is generally faster in their movements. Table 1 suggests that the performance difference between the groups largely stems from shorter inter-key intervals and shorter press durations.

8 DISCUSSION

This study highlights the complex nature of novel text entry system design. At the conception of this reported experiment, it was hypothesised that participant performance in mid-air with ten fingers would match, if not exceed, two finger performance. The results clearly indicate that the opposite is the case. The comparatively similar performance between the two surface-aligned conditions suggests that it may be hard for people to visually attend to more than two fingers on a virtual keyboard. There is precedence in this result with Kin et al. [11] determining that novice users employing two fingers (one per hand) can perform as well or better than when using 10 fingers in a multitarget selection task.

The micro metrics introduced and examined in subsection 6.2 form an attempt to shed light on the factors that dictate performance. Another perspective on this analysis comes from a brief consideration of the physiology of the hand and how this relates to typing. The physiology of the human hand means that movement of the middle, ring and pinky fingers can be difficult to decouple. The resistance provided in a physical keyboard is sufficient to prevent such coupled motion from generating insertions errors, however, there is clearly no resistance provided by a virtual keyboard in mid-air. Particularly problematic in ten-finger mid-air typing is the inability for users to decouple hand motion from fingertip motion. For example, a user may intend to strike a key with their middle finger and move their wrist to do so but without a corresponding retraction of the ring and pinky finger this motion is likely to yield three distinct intersections with the detection plane. Intelligently addressing such errors represents a particularly difficult discrimination and inference challenge.

At this point it is also worth reflecting on the implications of the experiment protocol for the human performance envelopes that have been identified. Clearly a transcription typing task is very distinct from text composition. Vertanen and Kristensson [30] found that entry rates dropped by between approximately 15 and 35% depending on the nature of the composition task. When composing text other factors may dictate what features of a text entry system are preferable. Furthermore, the experiment exposes participants to extended periods of one-phrase-at-a-time text entry. This may not be representative of typical text entry use cases in VR. Despite best efforts to control for learning and exhaustion effects by fully balancing condition order, individual quantitative and qualitative results may inevitably be influenced by these factors.

Also of interest is a general sense of how the performance envelopes obtained compare with other studies conducted in this space. Walker et al. [32] and Grubert et al. [9] both examine the use of a physical keyboard with an HMD. Grubert et al. [9] found that

entry rates on a physical keyboard were approximately 50% slower when wearing an HMD (with virtual representations of fingertips and keyboard) than when not wearing one. Participants in the study performed by Walker et al. [32] experienced only a marginal drop in performance when wearing an HMD but were supported by a probabilistic decoder.

8.1 Implications for a Functional Keyboard

In this section we return to the third objective of this study: informing the design of a functional keyboard tailored for use in VR. We review several examples of the way in which high-fidelity performance and behavioural data can inform the design of a fully functional keyboard system.

First, as highlighted in Section 7, the behaviour of top performers informs tracking accuracy and touch detection threshold target levels. For example, if the sub-group of top performers is able to type in the vicinity of of 80 to 100 wpm this implies a minimum detection threshold to make this feasible. For example, if an intersection based approach such as the one used in this study is used, the period between tracking position updates must be at least several times smaller than the typical press duration in order to avoid frequent failed detections.

Second, understanding the error distribution over the keyboard layout can inform likelihood estimates, that is, $P(key|touch_{x,y})$, in a probabilistic auto-correction strategy [34]. Similarly, understanding hand/finger specific performance can likewise inform the likelihood estimate, that is, $P(key|hand/finger,touch_{x,y})$. Such modulation of the confidence in particular touches might, for example, help to address the frequent insertion errors highlighted in subsection 6.2.2 at the edges of the keyboard layout associated with the pinky fingers.

Third, understanding common errors can inform layout refinement and/or alternative 'touch' detection strategies. For example, we observed frequent double-character omissions due to a failure to leave and re-enter the detection plane. Our analysis in subsections 6.2.4 and 6.2.7 provide some preliminary insight into how such intent might be discriminated. For example, it may be feasible to detect the intent of a repeated character when a press reversal above a set threshold occurs while still inside the detection plane.

9 LIMITATIONS AND FUTURE WORK

There are several important limitations of this study and aspects of AR/VR typing requiring more detailed investigation. While the ultimate goal of this research is to develop highly efficient and easy to learn text entry methods for use with AR and VR HMDs, today's display and tracking technology necessitates an experiment conducted in VR. Current commercially available AR HMDs suffer from tracking, resolution and field-of-view limitations that were predicted to have a confounding effect on our investigation of raw performance potential. Nevertheless, many of the results obtained and behaviours observed are likely common to both deployments. Key differences, such as the effect of being able to see one's own physical hands as opposed to a virtual representation, require further investigation.

It is also important to highlight that this study examines text entry without enforcing or requiring error correction. Clearly this fact means that the entry rate envelopes of human performance identified are optimistic. A fully functional text entry system for AR/VR must provide a means to perform error corrections and the best strategy for achieving this also requires future investigation. Also related to this point is the reduced keyboard layout used in this study. Again, a fully functional keyboard is likely to require the full complement of punctuation, numerals and support case modification.

For experimental purposes, participants in this study were confined to a seated posture. For any practical text entry system in AR/VR, however, a range of postures must ideally be supported. Posture may clearly have a strong influence on the performance

and enjoyment of a given text entry strategy, and understanding this sensitivity remains as future work.

It is also important to highlight how certain design choices in the development of the test bed might necessitate caution in the generalisation of the results obtained. In the implementation examined, fingertips are represented as spheres instead of virtual hands. The effect of this choice has been examined by Grubert et al. [8] and Knierim et al. [12]. Grubert et al. [8] find that representing the fingertips alone can perform as well as live video of the user's hands with the added benefit of minimising keyboard occlusion. Note too that this presents a key distinction from potential performance in AR where users can see their own hands. Similarly, the choice of keyboard sizing and placement as well as placement of the input field may influence performance. For example, placement of the input field immediately above the keyboard potentially promotes focusing on keys rather than falling back on learned touch-typing skills. Likewise, some behaviours are potentially specific to keyboard layout, sizing and placement. The effects of these various design choices require future investigation.

A further avenue of future work is the expansion of the range of potential text entry strategies evaluated in the test bed. It is important to avoid design fixation and limiting investigation to those methods that are minor variations on conventional text entry strategies. For example, the virtualisation of the keyboard enables many novel input strategies such as split keyboards and/or keyboards that are fixed relative to certain joints. This does, however, expose a well-known factor in text entry research: the significant time investment required to learn a fundamentally new text entry strategy—and the fact that historically users have been unwilling to adopt a text entry method that demands upfront learning investment [14].

10 CONCLUSIONS

This paper describes an empirical investigation of two fundamental design choices for text input in VR: the number of fingers engaged and whether the virtual keyboard is aligned with a physical surface or floating in mid-air. We find that aligning the keyboard with a physical surface yields significantly higher entry rates, with greater user comfort. Contrary to expectations, the results suggest that users struggle to effectively leverage the availability of all ten fingers. In fact, when typing in mid-air, the availability of more fingers appears to be detrimental to performance. Nevertheless, the choice between surface or mid-air typing may be dictated by the user's circumstance and so it is useful to understand the anticipated envelopes of human performance of these different but complementary strategies.

In addition to identifying the envelopes of human performance for the four strategies investigated we also provided a detailed analysis of the underlying features that distinguish each strategy in terms of its performance and behaviour. These insights in turn inform the design of a fully functional text entry system, including its tracking characteristics and statistical text decoder design. It is important to highlight that the conditions examined and distinctions made in their analysis are not a reflection of a desire to find a single 'best' input strategy for VR. Rather, it is hoped that a better understanding of the influence of various design decisions and underlying performance and behavioural indicators will ultimately yield a flexible and adaptable text entry system suitable for a variety of use-contexts.

REFERENCES

[1] D. A. Bowman, C. J. Rhoton, and M. S. Pinho. Text input techniques for immersive virtual environments: An empirical comparison. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, pp. 2154–2158. SAGE Publications, 2002.

[2] S. K. Card, T. P. Moran, and A. Newell. *The psychology of human-computer interaction*. Lawrence Erlbaum Associates, 1983.

[3] V. Dhakal, A. M. Feit, P. O. Kristensson, and A. Oulasvirta. Observations on Typing from 136 Million Keystrokes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 646:1–646:12. ACM, New York, NY, USA, 2018. event-place: Montreal QC, Canada. doi: 10.1145/3173574.3174220

[4] 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 (TOCHI)*, 25(6):30:1–30:40, Dec. 2018. doi: 10.1145/3232163

[5] A. M. Feit, D. Weir, and A. Oulasvirta. How We Type: Movement Strategies and Performance in Everyday Typing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pp. 4262–4273. ACM, New York, NY, USA, 2016. event-place: San Jose, California, USA. doi: 10.1145/2858036.2858233

[6] L. Findlater and J. Wobbrock. Personalized Input: Improving Ten-finger Touchscreen Typing Through Automatic Adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 815–824. ACM, New York, NY, USA, 2012. event-place: Austin, Texas, USA. doi: 10.1145/2207676.2208520

[7] T. Grossman, X. A. Chen, and G. Fitzmaurice. Typing on Glasses: Adapting Text Entry to Smart Eyewear. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pp. 144–152. ACM, New York, NY, USA, 2015. doi: 10.1145/2785830.2785867

[8] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Effects of Hand Representations for Typing in Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 151–158, Mar. 2018. doi: 10.1109/VR.2018.8446250

[9] 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

[10] R. Harmon, W. Patterson, W. Ribarsky, and J. Bolter. The virtual annotation system. In *Virtual Reality Annual International Symposium, 1996., Proceedings of the IEEE 1996*, pp. 239–245. IEEE, 1996.

[11] K. Kin, M. Agrawala, and T. DeRose. Determining the Benefits of Direct-touch, Bimanual, and Multifinger Input on a Multitouch Workstation. In *Proceedings of Graphics Interface 2009*, GI '09, pp. 119–124. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 2009. event-place: Kelowna, British Columbia, Canada.

[12] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze. 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. 345:1–345:9. ACM, New York, NY, USA, 2018. event-place: Montreal QC, Canada. doi: 10.1145/3173574.3173919

[13] L. Kohli and M. Whitton. The Haptic Hand: Providing User Interface Feedback with the Non-dominant Hand in Virtual Environments. In *Proceedings of Graphics Interface 2005*, GI '05, pp. 1–8. Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2005. event-place: Victoria, British Columbia.

[14] P. O. Kristensson. Next-Generation Text Entry. *Computer*, 48(7):84–87, 2015.

[15] 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, pp. 151–158. ACM, New York, NY, USA, 2005. doi: 10.1145/1040830.1040867

[16] F. Kuester, M. Chen, M. E. Phair, and C. Mehring. Towards Keyboard Independent Touch Typing in VR. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '05,

pp. 86–95. ACM, New York, NY, USA, 2005. doi: 10.1145/1101616.1101635

[17] M. Lee and W. Woo. ARKB: 3d vision-based Augmented Reality Keyboard. In *ICAT*, 2003.

[18] R. W. Lindeman. *Bimanual Interaction, Passive-haptic Feedback, 3D Widget Representation, and Simulated Surface Constraints for Interaction in Immersive Virtual Environments*. PhD Thesis, The George Washington University, 1999.

[19] K. Lyons, T. Starner, D. Plaisted, J. Fusia, A. Lyons, A. Drew, and E. W. Looney. Twiddler Typing: One-handed Chording Text Entry for Mobile Phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pp. 671–678. ACM, New York, NY, USA, 2004. event-place: Vienna, Austria. doi: 10.1145/985692.985777

[20] A. Markussen, M. R. Jakobsen, and K. Hornbaek. Selection-Based Mid-Air Text Entry on Large Displays. In P. Kotz, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, eds., *Human-Computer Interaction INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part I*, pp. 401–418. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-40483-2_28

[21] 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

[22] T. Ni, D. Bowman, and C. North. AirStroke: Bringing Unistroke Text Entry to Freehand Gesture Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 2473–2476. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1979303

[23] T. Ogitani, Y. Arahori, Y. Shinyama, and K. Gondow. 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, May 2018. doi: 10.1109/AINA.2018.00059

[24] I. Poupyrev, N. Tomokazu, and S. Weghorst. Virtual Notepad: handwriting in immersive VR. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, pp. 126–132. IEEE, 1998.

[25] R. Rosenberg and M. Slater. The chording glove: a glove-based text input device. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 29(2):186–191, 1999.

[26] S. Ruan, J. O. Wobbrock, K. Liou, A. Ng, and J. A. Landay. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4):159:1–159:23, Jan. 2018. doi: 10.1145/3161187

[27] S. Sridhar, A. M. Feit, C. Theobalt, and A. Oulasvirta. Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 3643–3652. ACM, New York, NY, USA, 2015. doi: 10.1145/2702123.2702136

[28] J. C. Verlinden, J. D. Bolter, and C. van der Mast. Virtual Annotation: Verbal Communication in Virtual Reality. Technical Report GIT-GVU-93-40, Georgia Institute of Technology, 1993.

[29] 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*, pp. 295–298. ACM, 2011.

[30] K. Vertanen and P. O. Kristensson. Complementing text entry evaluations with a composition task. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(2):8, 2014.

[31] 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*, pp. 659–668. ACM, 2015.

[32] 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. event-place: Denver, Colorado, USA. doi: 10.1145/3025453.3025783

[33] C.-Y. Wang, W.-C. Chu, P.-T. Chiu, M.-C. Hsiu, Y.-H. Chiang, and M. Y. Chen. PalmType: Using Palms As Keyboards for Smart Glasses. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pp. 153–160. ACM, New York, NY, USA, 2015. doi: 10.1145/2785830. 2785886

[34] D. Weir, H. Pohl, S. Rogers, K. Vertanen, and P. O. Kristensson. Uncertain Text Entry on Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 2307–2316. ACM, New York, NY, USA, 2014. doi: 10.1145/2556288. 2557412

[35] 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 & Technology*, UIST '15, pp. 539–548. ACM, New York, NY, USA, 2015. doi: 10.1145/2807442.2807504

[36] 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

[37] C. Yu, K. Sun, M. Zhong, X. Li, P. Zhao, and Y. Shi. One-Dimensional Handwriting: Inputting Letters and Words on Smart Glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 71–82. ACM, 2016.

[38] D. Yu, K. Fan, H. Zhang, D. Monteiro, W. Xu, and H. Liang. PizzaText: Text Entry for Virtual Reality Systems Using Dual Thumbsticks. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2927–2935, Nov. 2018. doi: 10.1109/TVCG.2018.2868581