Effects of Hand Representations for Typing in Virtual Reality

Jens Grubert* Coburg University of Applied Sciences and Arts
Lukas Witzani† University of Passau
Eyal Ofek‡ Microsoft Research
Michel Pahud§ Microsoft Research
Matthias Kranz¶ University of Passau
Per Ola Kristensson‖ University of Cambridge

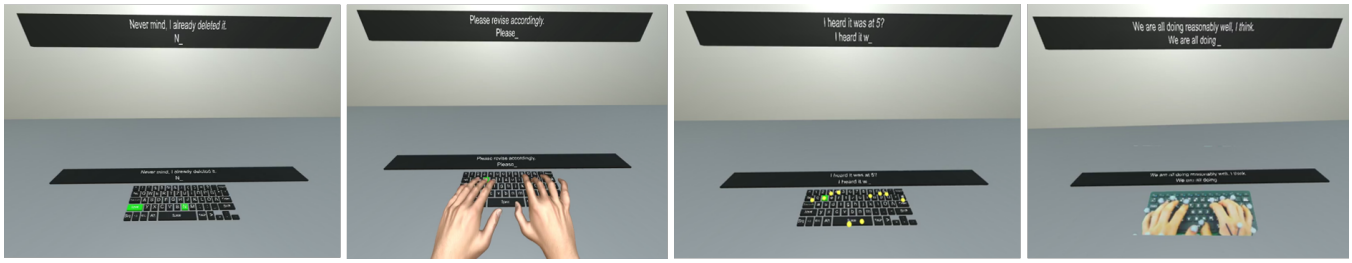

Figure 1: Views on the conditions studied in the experiment on effects of hand representations for typing in VR. From left to right: NoHand, IKHand, Fingertip and VideoHand.

Abstract

Alphanumeric text entry is a challenge for Virtual Reality (VR) applications. VR enables new capabilities, impossible in the real world, such as an unobstructed view of the keyboard, without occlusion by the user's physical hands. Several hand representations have been proposed for typing in VR on standard physical keyboards. However, to date, these hand representations have not been compared regarding their performance and effects on presence for VR text entry. Our work addresses this gap by comparing existing hand representations with minimalistic fingertip visualization. We study the effects of four hand representations (no hand representation, inverse kinematic model, fingertip visualization using spheres and video inlay) on typing in VR using a standard physical keyboard with 24 participants. We found that the fingertip visualization and video inlay both resulted in statistically significant lower text entry error rates compared to no hand or inverse kinematic model representations. We found no statistical differences in text entry speed.

Index Terms: H.5.2: [User Interfaces - Input devices and strategies.]

1 Introduction

Text entry in Virtual Reality (VR) is an important feature for many tasks, such as note taking, messaging and annotation. Existing consumer-grade VR systems, such as HTC Vive, Oculus Rift, Sony PSVR or Samsung's Gear VR, often rely on indirect control of a virtual pointer using hand-held controllers or head or gaze direction. However, these methods are limited in performance and consequently mostly used to enter short texts, such as passwords or names. Further, they require some degree of training due to their unfamiliarity to some users, compared to standard desktop and touchscreen keyboards that users are already familiar and proficient with. In addition, we believe a VR headset coupled with a keyboard can become an enabler for a full portable office in which a user can enjoy a motion-independent robust and immersive virtual office environment (Figure 2). For instance, users of touchdown spaces might find convenient to be able to carry their personalized large office configuration with multiple displays along with them in very tiny spaces.

However, while direct transplantation of standard keyboards to VR is viable, there are critical design parameters that should be investigated since it is plausible they affect performance. In a companion paper [7] we investigate the performance of physical and touch keyboards and physical/virtual co-location for VR text entry.

In this paper we focus specifically on investigating the method for virtually representing a user's hands in VR in several different ways. One possibility is to not reveal the hands at all, or resort to simply visually indicate to the user the actual pressed keys [40].

Figure 2: A vision of a portable virtual office enabled by VR.

*e-mail: jg@jensgrubert.de
†e-mail: lukas.witzani@uni-passau.de
‡e-mail: eyalofek@microsoft.com
§e-mail: mpahud@microsoft.com
¶e-mail: matthias.kranz@uni-passau.de
‖e-mail: pok21@cam.ac.uk

Alternatively, incorporating a video of the user's hand, within a VR world [22] may be the closest representation to physical hands, at a possible cost of breaking the immersiveness of the virtual world. Another common representation that can be used to fit the VR style is to visualize the users' hands using a three-dimensional model, made to fit the scene style. However, depending on the level of sensing of the users finger motions, there may be visual differences between the look of the model and the real hands. Furthermore, the look of the rendered hand, a difference in gender, unnatural interpolation of motions may generate a dissonance between the user and the chosen avatar hands [31]. VR is not limited to physical limitations, and new representations may be proposed that better fit the text entry needs. For example, it is possible to visualize only the user's finger tips and thereby minimize visual clutter, leaving the keyboard mostly visible.

1.1 Contribution

In this paper, we present the results of an experiment, in which we study the effect of four different hand representations: no hand representation, representing the hands via a three dimensional hand model, representing fingertips only, and representing the actual hands via blended video.

Our study indicates that while users tend to write comparably fast (with disabled correction options), the different representations lead to significant differences in error rate and preferences. Specifically, inverse kinematics hand model results in lower performance compared to a minimal fingertip visualization, while showing no hands at all results in comparable performance to a inverse kinematics hand model and a minimal fingertip visualization results in comparable performance to a blended video view of the users physical hands.

2 Related Work

There is a large body of work investigating interaction issues in VR ranging from perceptual issues [28] to interaction tasks like navigation, spatial manipulation, system control or symbolic input [2].

Text entry has been extensively researched (see [19, 49, 20, 14, 12] for surveys and overviews). Several strategies have developed in the text entry field to improve performance, in particular optimization [21, 45, 50, 46, 1, 24], decoding (auto-correct) [6, 16, 13, 42, 36] and gesture keyboards [47, 15, 11, 48].

2.1 Text Entry in VR

Relatively few text entry methods have been proposed for VR. Bowman et al. [3] speculate that the reason for this was that symbolic input may seam inappropriate for the immersive VR, and a belief that speech will be the one natural technique for symbolic input. Indeed, when comparing available techniques, they found speech to be the fastest medium for text entry at about 14 words-per-minute (wpm), followed by using a tracked stylus to select characters on a tablet (up to 12 wpm), a specially dedicated glove that could sense a pinch gesture between the thumb and each finger at 6 wpm, and last a commercial chord keyboard which provided 4 wpm. While voice control is becoming a popular input modality [25], it has severe limitations of ambient noise sensitivity, privacy, and possible obtrusiveness in a shared environment [4, 33]. It has also been argued that speech may interfere with the cognitive processes of composing text [32]. Furthermore, while dictation of text may feel natural, it is less so for text editing. Correcting speech recognition errors is also a challenge (see Vertanen [34] for a recent overview).

Prior work has investigated a variety of wearable gloves, where touching between the hand fingers may represent different characters or words, e.g., [3, 9, 17, 26]. Such devices enable a mobile, eyes-free text entry. However, most of them require a considerable learning effort, and may limit the user ability to use other input devices while interacting in VR, such as game controllers.

Yi et al. [44] suggest a system that senses the motion of hands in the air, to simulate typing, claiming a rate of up to 29 wpm, measured not in VR, but while the users could see their hands, and keep them in front of the sensing device. Also, holding the hands in mid-air above some virtual plane is lacking any haptic or tactile sensation feedback and can become tiring quickly and may not fit long typing session.

PalmType [41] uses a proximity sensor to use the non-preferred hand as a typing surface for the index finger of the preferred hand. The authors claim a rate which is better than touchscreen phones. However, the size of the simulated keyboard is limited to a hand size, and is hard to be rendered well in the current HMDs. It is also limiting the interaction to a single finger of a single hand, while the non-preferred hand is occupied as the type surface.

The mainstream mobile phone touchscreen keyboard might be a good text entry option [5]. It is portable, and can generate a relatively high text entry rate [29]. While we do not investigate this small type of touchscreen keyboard in this paper, we do believe it has potential. Currently, a limitation with using a phone is its small size, which combined with the display resolution limitation does not generate a good experience.

2.2 Keyboards for VR

Recent research has investigated the feasibility of typing on a physical full-sized keyboard (hereafter referred to as a desktop keyboard) in VR. An obvious problem is the lack of visual feedback. Without visual feedback users' typing performance degraded substantially. However, by blending video of the user's hands into virtual reality the adverse performance differential significantly reduced [22].

Fundamentally there are three solution strategies for supporting keyboards in VR. First, by providing complete visual feedback by blending the user's hands into virtual reality. Second, by decoding (auto-correcting) the user's typing to compensate for noise induced by the lack of feedback. Third, by investigating hybrid approaches, such as minimal visual feedback, which may or may not require a decoder to compensate for any noise induced by the method.

Walker et al. [39] presented the results of a study of typing on a desktop keyboard with the keyboard either visible or occluded, and while wearing a VR HMD with no keyboard display. They found that the character error rate (CER) was unacceptably high in the HMD condition (7.0% average CER) but could be reduced to an average 3.5% CER using an auto-correcting decoder. A year later, they showed that feedback of a virtual keyboard in VR, showing committed types, can help users correct their hand positions and reduce error rates while typing [40]. They discovered that their participants typed at an average entry rates of 41.2–43.7 words per minute (wpm), with average character error rates of 8.3%–11.8%. These character error rates were reduced to approximately 2.6%-4.0% by auto-correcting the typing using the VelociTap decoder [37]. In contrast, in this paper, we will show that by visualizing users' finger tips while typing, there is no need for an auto-correcting decoder as with the visual feedback users' character error rate is already sufficiently low for both desktop keyboard typing. This provides significant benefits to the user, as auto-correcting decoders, while useful for enabling quick and accurate typing, suffer from the auto-correct trap [42], which reduces entry rates and increases user frustration when the system inadvertently fails to identify the user's intended word.

McGill et al. [22] investigated typing on a physical keyboard in Augmented Virtuality [23]. Specifically, they compared a full keyboard view in reality with a no keyboard condition, a partial and full blending condition. For the blending conditions the authors added a camera view of a partial or full scene into the virtual environment as a billboard without depth cues. They found, that providing a view of the keyboard (partial or full blending) has a positive effect on typing performance. Their implementation is restricted to

typing with a monoscopic view of the keyboard and hands and the visualization of hand movements is bound by the update rate of the employed camera (typically 30 Hz). We study a complementary setup, in which we focus on virtual representations of the keyboard and hands.

Lin et al. [18] investigated the effects of different keyboard representations on user performance and preference for typing in VR but did not study different hand representations in depth.

Grubert et al. [7] investigated the performance of physical and touch keyboards and physical/virtual co-location for VR text entry.

Schwind et al. [31] investigated the effects of gender on different virtual hand representations. The participants had to conduct various tasks, amongst them a typing task with a single sentence. The authors did not report any text entry metrics. For our experiment on hand representations, we reused their androgynous hand model as a compromise between male and female hand depictions.

3 Hand Representation for Typing in VR

In this work, we looked at the ability to enter a substantial amount of text in VR using a standard desktop keyboard setup, which is commonly available and requires little, if any, learning to use in VR. The existing keyboards already have comfortable form factors for two-hand interaction and provide the same haptic feedback in VR as they do in the real world. In contrast to typing in the real world, VR allows the user to be free of physical limitations of this world. For example, if the user's hands occlude a keyboard, it is possible to make their virtual representation transparent so that the user can see the keyboard better. However, it is unclear if this would affect typing ability in VR.

Various hand representations have already been proposed in prior work. However, it remains unclear how those could effect typing performance, effort and presence. Hence, we set out to compare common hand representations using a standard desktop keyboard. Specifically, we compare following hand representations: no hand representation as studied by Walker and Vertanen [40], blended video see-through of the user's hands as proposed by McGill et al. [22], an inverse kinematic hand model as used by Schwind et al. [31] and a minimalistic sphere representation of the user's fingertips. Further, we conjecture that with appropriate visualization, there is no need for an auto-correcting decoder, which reduces the complexity of the typing system.

4 Experiment

To investigate the effect of hand representation on typing in VR we carried out a controlled experiment. We used a within-subjects design with a single independent variable—HandRepresentation–with four levels: no hand representation (NoHand), inverse kinematic hand model (IKHand), fingertip visualization through spheres (Fingertip) and an Augmented Virtuality representation based on chroma keying (VideoHand), see Figure 1.

Since our objective was to evaluate the effect of hand representation on typing, our investigation primarily focused on statistically comparing typing performance metrics across the four conditions and generalizing these differences to the population. We therefore ensured we sampled participants with diverse study backgrounds and. importantly, we did not attempt to subsample participants with similar typing abilities.

Since text entry is a relatively complex task it is important to ensure participants are typing a sufficient amount of text for us to be able to accurately sample their true typing performance. Not doing so would introduce two unwanted sources of error: First, since the difficulty of typing individual phrases varies, inadequate typing inflates noise in the text entry and error rate measurements. Second, since participants are exposed to new unfamiliar conditions, there is inevitably a degree of learning and familiarization within the first few minutes within each condition. For this reason we exposed every participant to 15 minutes of typing in each condition.

In addition, it is also important that the stimulus text models the text participants are likely to type. Such text is known as being *in-domain*. For this reason we used stimulus text from a corpus derived from genuine emails typed on mobile devices [35].

4.1 Method

In the condition NoHand, the participants saw no hand representation at all, but only the text they typed as well as green highlights of keys currently pressed. The condition Fingertip, showed semi-transparent yellow spheres at the finger tips but no other visualization in addition to the highlighted keys. The condition VideoHand used a blended billboard with a live video of the user's hands as well as the physical keyboard. Please note, that in this condition the retroreflective markers where visible as well. While this might degrade the visual experience, the markers where kept as in the other conditions to avoid a potential confound. Also, there was an average end-to-end delay from finger movement to display in VR of 170 ms, which did not result in substantial coordination problems while typing as indicated by pre-tests.

The condition IKHand used an inverse kinematic hand model [31] as well as the key highlights.

The experiment was carried out in a single 130-minute session structured as 5-minute welcoming and introduction, 5 minutes profiling phase, 15 minutes attachment of retroreflective markers, hand rigid bodies and calibration, a 90-minute testing phase (15 minutes per condition + ca. 7-minute breaks and questionnaires in between) and 15 minutes for final questionnaires, interviews and debriefing.

4.2 Participants

We recruited 25 participants from a university campus with diverse study backgrounds. All participants were familiar with QWERTZ desktop keyboard typing and QWERTZ touchscreen keyboard typing, none took part in the previous study. One participant had to be excluded due to logging issues. From the 24 remaining participants (14 female, 10 male, mean age 23.5 years, sd = 2.2, mean height 168.9 cm, sd = 36.7), 15 indicated to have never used a VR HMD before, 4 to have worn a VR HMD once, 2 participants rarely but more than once and 3 participants to wear it occasionally. Seven participants indicated to not play video games, 2 once, 9 rarely, 2 occasionally, 1 frequently and 3 very frequently. Nineteen participants indicated to be highly efficient in typing on a physical keyboard, 3 to be medium efficient and 2 to write with low efficiency on a physical keyboard (we caution against over-interpreting these self-assessed performance indications). Six participants wore contact lenses or glasses. The volunteers have not participated in other VR typing experiments before.

4.3 Apparatus and Materials

Stimulus sentences were drawn from the mobile email phrase set [35], which provides a set of text entry stimulus sentences that have been verified to be both externally valid and easy to memorize for participants. Participants were shown stimulus phrases randomly drawn from the set. An OptiTrack Flex 13 outside-in tracking system was used for spatial tracking of finger tips and the HMD, again with a mean spatial accuracy of 0.2 mm. An Oculus Rift DK2 was used as HMD. A Logitech C910 camera (resolution 640x480, 30Hz) was mounted in front of the HMD and external lighting as well as a green keying background was installed to enable the VideoHand condition, see Figure 3. In addition to fiducials at the finger tips, another rigid-body fiducial was mounted at the back of the hands to support the IKHand condition.

The physical keyboard was a CSL wireless keyboard with physical dimensions of (width $\times$ height) ($w \times h$): 272×92 mm and key dimensions of 15×14 mm, see Figure 4.

Figure 3: Left: Setup with green keying, tracking system and external lighting. Top right: HMD with external camera and tracking fiducial. Bottom right: fiducials used for tracking the hands in the IKHand condition.

4.4 Calibration Data Collection

The calibration phase consisted of four parts: text entry profiling, interpupillary distance (IPD) calibration, finger tip calibration and finger length calibration. During the text entry profiling phase, participants were asked to copy prompted sentences using a desktop keyboard. Stimulus phrases were shown to the participants one at a time. Participants were asked to type them as quickly and as accurately as possible. Participants typed stimulus phrases for 5 min using a desktop keyboard. The IPD was determined with a ruler and then used for setting the correct camera distance for stereo rendering.

For finding out the IPD Participants were asked to sit upright and parallel to the experimenter holding a ruler below their eyes. With the experimenter having one eye closed, participants were asked to look straight into the open eye. The ruler was adjusted till its origin crossed the line between the open eye and the participants corresponding eye. The experimenter closed both eyes and opened the other one. Participants now looked again into the open eye while having read out their IPD by the experimenter.

This method is a fast and simple way to find out the IPD and was chosen to decrease the total time the participants need to wear the HMD. The IPD was then transferred to the experiment software.

For finger tracking, individual retroreflective markers were attached to the nails of participants using double-sided adhesive tape, see Figure 3, bottom right. The finger calibration aimed at determining the offset between the tracked 3D position of each finger tip and its corresponding nail-attached marker. To this end, the participants were asked to hit three soft buttons of decreasing size (large: 54 × 68 mm, medium: 35 × 50 mm, small: 15 × 15 mm) on a Nexus 10 touch surface. Initially, the virtual finger tips were shown at the registered 3D positions of the retroreflective markers. On touchdown, the virtual finger tips were transformed by the offset between the 3D coordinate of the touch point and the retroreflective marker. The final positions of the virtual finger tips were averaged across three measurements. Then the participants verified that they could actually hit targeted keys using their virtual finger tip. If necessary, the process was repeated. This calibration procedure was conducted for each finger individually.

Figure 4: Physical keyboard used in the experiment.

Additionally, rigid bodies holding four retroreflective markers were attached onto the back of each of the participant's hands. For each participant the hand's rigid body rotation was reset in the tracking system's software. Also, the information given by the hands' rigid bodies and the fingertips' markers was used to adapt the finger lengths of the inverse kinematic model to the participant's finger lengths. This ensured that the virtual fingertips were positioned near the real ones. Before each condition, the participants verified that they could actually hit targeted keys using their virtual finger tip. If necessary, parts of the calibration process were repeated.

4.5 Procedure

The order of the conditions was balanced across participants. In either condition, participants were shown a series of stimulus sentences. For an individual stimulus sentence, participants were asked to type it as quickly and as accurately as possible. Participants typed stimulus sentences for 15 minutes in each condition. The conditions were separated by a 5-minute break, in which participants filled out a the SSQ simulator questionnaire [10], the NASA TLX questionnaire [8], the IPQ [27] spatial presence questionnaire and the Flow-Short-Scale [30].

Please note, that participants were not allowed to use the backspace key to correct errors. This was done in line with the suggestion by Walker and Vertanen [40] to avoid excessive use of correction in the NoHand condition.

4.6 Results

Statistical significance tests for entry rate, error rate and time to first keypress (log-transformed) were carried out using General Linear Model (GLM) repeated measures analysis of variance (RM-ANOVA) with Holm-Bonferroni adjustments for multiple comparisons at an initial significance level $\alpha = 0.05$. Effect sizes for the GLM (η_p^2) are specified whenever they are available. All GLM analyses were checked for appropriateness against the dataset. We used GLM RM-ANOVA since it was appropriate for the dataset and provided more statistical power than non-parametric tests.

Statistical significance tests for ratings and preferences were carried out using the non-parametric Friedman's tests coupled with Holm-Bonferroni adjusted post-hoc analyses with Wilcoxon signed-rank tests.

4.6.1 Entry Rate and Time to First Keypress

Entry rate was measured in wpm, with a word defined as five consecutive characters, including spaces; see Figure 5, first row, for a graphical summary. The entry rate was 36.1 wpm (sd = 18.1) for NoHand, 34.4 wpm (sd = 17.0) for IKHand, 36.4 wpm (sd = 15.3) for Fingertip and 38.7 wpm (sd = 13.6) for VideoHand, see Figure 5, first row. The difference in entry rate was not significant ($F_{3,69} = 2.550$, $\eta_p^2 = 0.1$, $p = 0.063$).

As a calibration point only, we also measured the entry rate ratio between the individual conditions and the profiling phase (Figure 5, second row). On average, NoHand resulted in a 76% entry rate compared to profiling, IKHand 72%, Fingertip 78% and VideoHand 84%. Note that we intentionally did not control for typing proficiency when recruiting participants and it is therefore not meaningful to calculate statistical significance. We note that that the typing ratios are fairly high, indicating that the majority of the participants' typing abilities were preserved in VR. We conjecture

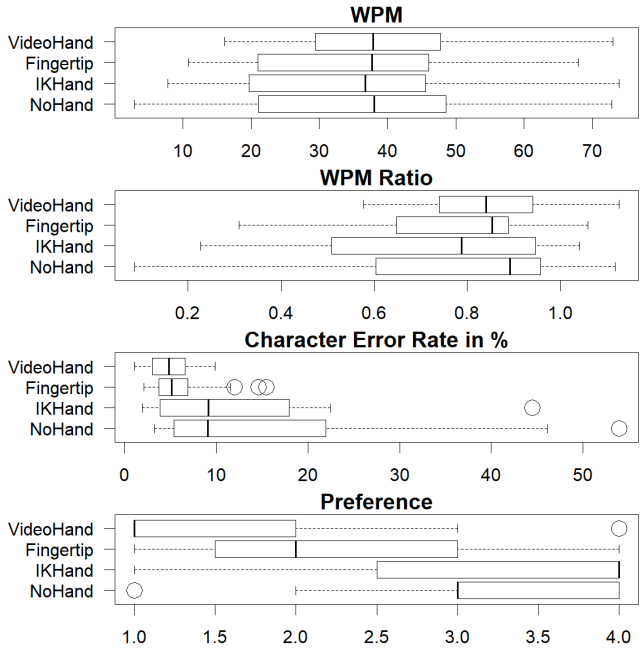

Figure 5: From top to bottom: Text entry rate in words per minute. Text entry ratio between condition and profiling phase. Character error rate. Preference (1 = maximum preference best, 4 = minimum preference).

this is due to participants being explicitly instructed not to perform corrections.

In addition we investigated the time to first keypress, a metric first suggested by McGill et al. [22] to get an indication of the time it takes participants to orient themselves before typing the sentence. The mean time to first keypress was 1.69 seconds (sd = 1.89) for NoHand, 1.57 seconds (sd = 1.42) for IKHand, 1.23 seconds (sd = 0.59) for Fingertip and 1.22 seconds (sd = 0.59) for VideoHand. A repeated measures analysis of variance on the log-transformed durations revealed that the differences were not statistically significant ($F_{3,69} = 1.951, \eta_p^2 = 0.078, p = 0.130$).

Since the number of participants ($n = 24$) was relatively high and the effect sizes are very low it is plausible there is no difference between the conditions for entry rate or time to first keypress.

4.6.2 Error Rate

Error rate was measured as character error rate (CER). CER is the minimum number of character-level insertion, deletion and substitution operations required to transform the response text into the stimulus text, divided by the number of characters in the stimulus text. The character error rate (CER) was 15.2% (sd = 14.2) for NoHand, 11.5% (sd = 9.6) for IKHand, 6.3% (sd = 3.7) for Fingertip and 5.1% (sd = 2.5) for VideoHand, see Figure 5, third row. An omnibus test revealed significance ($F_{3,69} = 9.029$, $\eta_p^2 = 0.282$, $p < 0.001$). Holm-Bonferroni adjusted post-hoc testing revealed that there was no significant difference between NoHand and IKHand (adjusted p-value < 0.025). There was also no significant difference between Fingertip and VideoHand. However, there were significant differences between NoHand and both Fingertip and VideoHand and between IKHand and both Fingertip and VideoHand (adjusted $p < 0.025$).

In other words, using no hand representation or an inverse kinematic hand model both resulted in a similar high CER. Using fingertip visualization with spheres or video see-through of the hand resulted in a similar relatively low CER. There is no statistically difference in CER between using fingertip visualization with spheres or video see-through of the hand.

4.6.3 NASA-TLX, Simulator Sickness and Spatial Presence

The overall median NASA-TLX rating was 57.9 for NoHand, 51.2 for IKHand, 47.5 for Fingertip and 43.8 for VideoHand. A Friedman's test revealed an overall significant difference ($\chi^2(3) = 10.399$, $p < 0.05$). Holm-Bonferroni corrected post hoc analyses with Wilcoxon signed-rank tests revealed that the difference between NoHand and VideoHand was significant ($Z = -2.615$, $p < 0.01$). No other pairwise differences were significant.

The overall median nausea score rating was 2.0 for NoHand (oculo-motor: 7.0), 2.0 for IKHand (oculo-motor: 8.0), 1.0 for Fingertip (oculo-motor: 7.5) and 1.0 for VideoHand (oculo-motor: 6.5). There was no significant difference (Friedman's test; $\chi^2(3) = 4.472$, $p = 0.215$).

For spatial presence, the median rating on a 7-item Likert scale was 3.5 for NoHand, 4.1 for IKHand, 3.8 for Fingertip and 4.0 for VideoHand. The differences for the overall rating were not statistically significant (Friedman's test; $\chi^2(3) = 7.268$, $p = 0.064$). However, using Friedman's test again, we found significant differences in the sub-scales Experienced Realism ($\chi^2(3) = 8.442$, $p < 0.05$) and Spatial Presence ($\chi^2(3) = 9.846$, $p < 0.05$). Involvement was not significant ($\chi^2(3) = 0.872, p = 0.872$). The only significant pairwise differences were between NoHand and all other conditions (Wilcoxon signed-rank tests).

4.6.4 Preferences and Open Comments

On a scale from 1 (best) to 4 (worst) the median preference rating for NoHand was 3, 4 for IKHand, 2 for Fingertip and 1 for VideoHand, see Figure 5, bottom row. Friedman's test revealed an overall significant difference ($\chi^2(3) = 30.150$, $p < 0.0001$). Post-hoc analysis with Wilcoxon signed-rank tests and Holm-Bonferroni correction revealed there were no significance differences in preference between NoHand and IKHand and similarly no significant difference in preference between Fingertip and VideoHand. However, the difference in preference between {NoHand, IKHand} and {Fingertip, VideoHand} were significant ($p < 0.005$).

Participants also commented about the reasons for their ratings. For the NoHand condition, two participants mentioned that it was hard or strenuous to orient themselves on the keyboard. However, three participants mentioned that showing no hand representation at all helped them to concentrate on writing. For IKHand, six participants mentioned that it was hard to orient on the keyboard, four mentioned that the hand was occluding too much of the keyboard, six participants experienced the hand as confusing or distracting. Two participants mentioned that the hands felt real, while two explicitly stated that they felt unreal with one stating that she would "prefer a more female version with nail polish". For Fingertip, six participants mentioned the feeling of accurate positioning, six that the spheres help to orient and do not occlude the keyboard substantially. One stated that the representation was "so abstract that it helped to orient better than the virtual hand model", one liked the "playful effect" of the spheres. However, one also mentioned that it felt "hard to orient if no finger is attached". For VideoHand, one participant highlighted the accurate depiction of hand positions, one that she felt in control, six mentioned that they prefer the representation as they are used to such a depiction of their hands, with one mentioning it would be the "least eerie choice". However, four participants also mentioned that the letters were hard to read due to the blurry depiction of the keyboard.

5 Discussion

Our research investigated the effects of different representations of the user's hands on text entry in VR. As a baseline, we used a video of the user's hands, composited in the virtual world as proposed by McGill et al. [22]. In our experiment, which presented a sparse VR

environment with minimalistic representations of a desk and a wall, it proved to be a useful representation. There could be cases where this representation could interfere with the virtual content rendering and may break immersiveness of the virtual experience. For example, in a futuristic settings, the user may be represented by an avatar wearing a spacesuit, while her hand will be represented by an everyday hands video. Even in a more "down-to-earth" setting, such as a virtual meeting or a virtual office, the video hand represents the conditions in the user's real environment and not in the VR set. Illumination, style, and for re-projection of the keyboard even the view direction of the video will not fit the VR world. Also, while using video hands does not require an external tracking system but merely a RGB camera attached to the HMD (or integrated, as in HTC Vive), we as well as McGill et al. [22] needed to instrument the environment to enable robust chroma keying. For unprepared environments getting a robust image mask for both the user hands and the keyboard can be challenging (e.g. a dark office or home environment). While one could simply show the whole camera view, this potentially could reduce immersion further.

Our study indicated that the studied minimalistic representation of finger tips has comparable performance to a video inlay. It has a high input rate with low error rate, and strong preference by the participants of our user study. We could imagine that this or other abstract representations could be beneficial over a video inlay for generic VR scenarios that aim at a high user presence. However, this should be investigated thoroughly in future work.

A representation of the hands by a full 3D model, on the other hand, was found to lack in these areas. Its error rate was significantly higher, in fact, showing no significant differences to depicting no hands at all. One possible reason for the high error rate may be the low visibility of the keyboard behind the hands. Another may be due to any differences that may occur between the visible motion of the model and the actual motion of the user's hands. Since tracking of the user's hand is based many times on partial data (position is known only at recognizable markers on the finger in our case, or near recognizable features when computer vision is used), a fit of a model is used to interpolate the full motion and look of the hands. Specifically, while in our experiment the accuracy of finger tip positions in both IKHand and Fingertip conditions where the same, the other joints in the IKHand model where interpolated by the inverse kinematics model. Hence, they might show larger deviations from their physical counterparts. Any resulting difference in the model motion or look and the physical hands, may generate mistakes by the user, or even dissonance between the user and the avatar hands due to uncanny valley effects. As a result, our participants chose this representation as their least preferable.

5.1 Limitations and Future Work

Our study focused on specific items in a large design space of how to design text entry systems in VR. For our evaluation, we focused on the scenario of a user sitting in front of a desk doing extensive text entry. One reason was to measure text input rate at its limit. Another reason was the observation that this configuration is still popular by many VR applications that do not require the user to walk. In particular, we can see a great potential of VR as a continuous unlimited VR display, replacing all physical screens in an office environment, supporting both 2D and 3D applications and visualizations. In this scenario, there is need for a robust text entry, which we believe can be filled by current keyboards with additional sensors for hand rendering.

Alternatively, there are many mobile scenarios which could benefit from efficient text entry techniques, also for shorter text sequences. Here, either a handheld or arm-mounted touch screen might serve as a suitable interaction device. In this context, future work should investigate recent mobile text entry techniques for VR, e.g. based on gesturing [43].

Also, we relied on high precision stationary optical tracking system. But even with this system, we did not sense the movement of physical key presses. The display of the fingers as as they move while typing may help people that do not touch type. The use of mobile depth sensors for hand tracking such as the leap motion could be a viable alternative, but their input accuracy for typing would need to be studied.

Further, the four tested hand representations of our study are just four points in a vast possible design space, and we can imagine more to be suggested. For example, it may be that rendering a semi-transparent silhouette may combine the keyboard visibility of the finger tips with some hint of the hand model that may increase realism, but not too much to generate potential uncanny valley dissonance. In this regard, Logitech and Vive recently announced a developer kit for text entry using a physical keyboard, which employs a semi-transparent video inlay of the user's hands [38]. Also, the keyboards and hands were rendered at 1:1 scale in the VR world. However, this size limits the text visibility on the keys in the HMD display. Keyboards and fingers may be scaled up, allowing greater visibility, or even scaled down to create less occlusion. Future work should investigate the effects of this scaling on text entry performance as resolution and sharpness of VR HMDs keep increasing. Finally, we aim at studying the effects of abstract hand representations such as spheres compared to video inlays in more detail. Specifically, we want to measure presence of both representation in various VR scenarios that go beyond a neutral office environment (such as the previously mentioned space scenario).

6 Conclusions

We have studied the effect of different representations of the user's hands on typing performance. We found that a minimalistic representation of the user's fingertips may enhance keyboard visibility and be as performant as viewing a live video of the user hands, while using 3D avatar hands that are fit to match the user hands, may decrese performance as much as not showing any view of the hands at all.

Finally, we believe that VR may expand from the current use of immersive experiences, to a work tool even in the common office, allowing information workers to interact and observe data without the limitation of physical screens. One barrier for such a vision is a robust text entry and editing tool, and we hope this work will be a step in this direction.

Acknowledgments

Per Ola Kristensson was supported by EPSRC (grant number EP/N010558/1). We thank the volunteers in the experiment and the anonymous reviewers for their feedback.

References

[1] X. Bi, B. A. Smith, and S. Zhai. Quasi-qwerty soft keyboard optimization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 283–286. ACM, 2010.

[2] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004.

[3] 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*, volume 46, pages 2154–2158. SAGE Publications, 2002.

[4] D. Dobbelstein, P. Hock, and E. Rukzio. Belt: An unobtrusive touch input device for head-worn displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2135–2138, New York, NY, USA, 2015. ACM.

[5] G. González, J. P. Molina, A. S. García, D. Martínez, and P. González. Evaluation of text input techniques in immersive virtual environments. In *New Trends on Human–Computer Interaction*, pages 109–118. Springer, 2009.

[6] J. Goodman, G. Venolia, K. Steury, and C. Parker. Language modeling for soft keyboards. In *Eighteenth National Conference on Artificial Intelligence*, AAAI '02, pages 419–424, Menlo Park, CA, USA, 2002. American Association for Artificial Intelligence.

[7] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Text entry in immersive head-mounted display-based virtual reality usingstandard keyboards. In *IEEE Virtual Reality (VR) 2018*, page to appear. IEEE, 2018.

[8] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[9] Y.-T. Hsieh, A. Jylhä, V. Orso, L. Gamberini, and G. Jacucci. Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4203–4215, New York, NY, USA, 2016. ACM.

[10] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[11] P. O. Kristensson. *Discrete and Continuous Shape Writing for Text Entry and Control*. PhD thesis, Linköping University, 2007.

[12] P. O. Kristensson. Five challenges for intelligent text entry methods. *AI Magazine*, 30(4):85, 2009.

[13] P. O. Kristensson. Five challenges for intelligent text entry methods. *AI Magazine*, 30(4):85–94, 2009.

[14] P. O. Kristensson. Next-generation text entry. *Computer*, 48(7):84–87, 2015.

[15] 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, pages 43–52, New York, NY, USA, 2004. ACM.

[16] P. O. Kristensson and S. Zhai. Relaxing stylus typing precision by geometric pattern matching. In *IUI '05: Proceedings of the 10th International Conference on Intelligent User Interfaces*, pages 151–158. ACM Press, 2005.

[17] 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, pages 86–95, New York, NY, USA, 2005. ACM.

[18] J.-W. Lin, P.-H. Han, J.-Y. Lee, Y.-S. Chen, T.-W. Chang, K.-W. Chen, and Y.-P. Hung. Visualizing the keyboard in virtual reality for enhancing immersive experience. In *ACM SIGGRAPH 2017 Posters*, page 35. ACM, 2017.

[19] I. S. MacKenzie and R. W. Soukoreff. Text entry for mobile computing: Models and methods, theory and practice. *Human–Computer Interaction*, 17(2-3):147–198, 2002.

[20] I. S. MacKenzie and K. Tanaka-Ishii. *Text Entry Systems*. Morgan Kauffman, 2007.

[21] I. S. MacKenzie and S. X. Zhang. The design and evaluation of a high-performance soft keyboard. In *CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 25–31, New York, NY, USA, 1999. ACM Press.

[22] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster. A dose of reality: overcoming usability challenges in vr head-mounted displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2143–2152. ACM, 2015.

[23] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(12):1321–1329, 1994.

[24] A. Oulasvirta, A. Reichel, W. Li, Y. Zhang, M. Bachynskyi, K. Vertanen, and P. O. Kristensson. Improving two-thumb text entry on touchscreen devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2765–2774. ACM, 2013.

[25] S. Pick, A. S. Puika, and T. W. Kuhlen. Swifter: Design and evaluation of a speech-based text input metaphor for immersive virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 109–112. IEEE, 2016.

[26] M. Prätorius, U. Burgbacher, D. Valkov, and K. Hinrichs. Sensing thumb-to-finger taps for symbolic input in vr/ar environments. *IEEE computer graphics and applications*, 2015.

[27] H. Regenbrecht and T. Schubert. Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators and virtual environments*, 11(4):425–434, 2002.

[28] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments - a review. *ACM Comput. Surv.*, 46(2):23:1–23:40, Dec. 2013.

[29] 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, pages 679–688, New York, NY, USA, 2015. ACM.

[30] F. Rheinberg, R. Vollmeyer, and S. Engeser. *Die erfassung des flow-erlebens*. na, 2003.

[31] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. These are not my hands!: Effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1577–1582. ACM, 2017.

[32] B. Shneiderman. The limits of speech recognition. *Communications of the ACM*, 43(9):63–65, 2000.

[33] Y.-C. Tung, C.-Y. Hsu, H.-Y. Wang, S. Chyou, J.-W. Lin, P.-J. Wu, A. Valstar, and M. Y. Chen. User-defined game input for smart glasses in public space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3327–3336, New York, NY, USA, 2015. ACM.

[34] K. Vertanen. *Efficient correction interfaces for speech recognition*. PhD thesis, Citeseer, 2009.

[35] 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, pages 295–298, New York, NY, USA, 2011. ACM.

[36] 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, pages 659–668, New York, NY, USA, 2015. ACM.

[37] 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*, pages 659–668. ACM, 2015.

[38] Vive. Introducing the Logitech BRIDGE SDK. `https://blog.vive.com/us/2017/11/02/introducing-the-logitech-bridge-sdk/`, 2017. Last accessed November 20th, 2017.

[39] J. Walker, S. Kuhl, and K. Vertanen. Decoder-assisted typing using an HMD and a physical keyboard. In *CHI 2016 Workshop on Inviscid Text Entry and Beyond*, page unpublished, 2016.

[40] 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*, pages 5457–5461. ACM, 2017.

[41] 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, pages 153–160, New York, NY, USA, 2015. ACM.

[42] D. Weir, H. Pohl, S. Rogers, K. Vertanen, and P. O. Kristensson. Uncertain text entry on mobile devices. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2307–2316, New York, NY, USA, 2014. ACM.

[43] H.-S. Yeo, X.-S. Phang, S. J. Castellucci, P. O. Kristensson, and A. Quigley. Investigating tilt-based gesture keyboard entry for single-handed text entry on large devices. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4194–4202. ACM, 2017.

[44] 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, pages 539–548, New York, NY, USA, 2015. ACM.

[45] S. Zhai, M. Hunter, and B. A. Smith. The metropolis keyboard-an exploration of quantitative techniques for virtual keyboard design. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 119–128. ACM, 2000.

[46] S. Zhai, M. Hunter, and B. A. Smith. Performance optimization of virtual keyboards. *Human–Computer Interaction*, 17(2-3):229–269, 2002.

[47] 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, pages 97–104, New York, NY, USA, 2003. ACM.

[48] S. Zhai and P. O. Kristensson. The word-gesture keyboard: Reimagining keyboard interaction. *Communications of the ACM*, 55(9):91–101, 2012.

[49] S. Zhai, P.-O. Kristensson, and B. A. Smith. In search of effective text input interfaces for off the desktop computing. *Interacting with computers*, 17(3):229–250, 2004.

[50] S. Zhai, A. Sue, and J. Accot. Movement model, hits distribution and learning in virtual keyboarding. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 17–24. ACM, 2002.