Opportunities and Challenges of Text Input in Portable Virtual Reality

Pascal Knierim, Thomas Kosch,
Johannes Groschopp, Albrecht Schmidt

LMU Munich, Munich, Germany
{firstname.lastname}@ifi.lmu.de

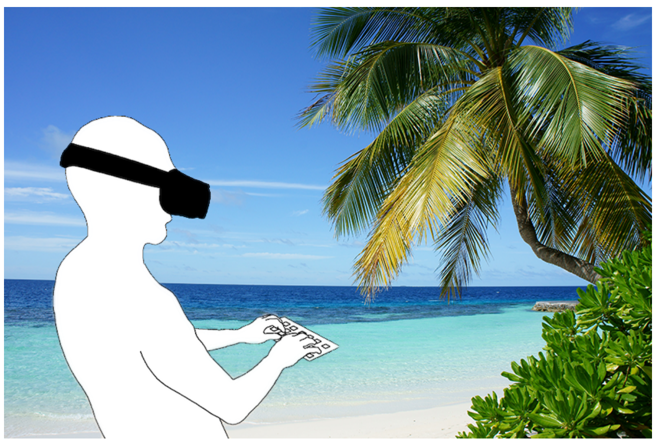

Figure 1: User copy editing text in a relaxing virtual world provided by a portable HMD setup.

CHI'20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.3382920

Abstract

Text input in virtual reality is not widespread outside of labs, although being increasingly researched. Current setups require powerful components that are expensive or not portable, hence preventing effective in-the-wild use. Latest technological advances enable portable mixed reality experiences on smartphones. In this work, we propose a portable low-fidelity solution for text input in mixed reality on a physical keyboard that employs accessible off-the-shelf components. Through a user study with 24 participants, we show that our prototype leads to a significantly higher text input performance compared to soft keyboards. However, it falls behind on copy editing compared to soft keyboards. Qualitative inquiries revealed that participants enjoyed the ample display space and perceived the accompanied privacy as beneficial. Finally, we conclude with challenges and future research that builds upon the presented findings.

Author Keywords

Virtual Reality; Mixed Reality; Text Entry; Copy Editing; Physical Keyboard; Portable Virtual Reality

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI); Mixed / augmented reality;** Haptic devices;

Figure 2: Top: User typing with our mixed reality apparatus.
Bottom: The mixed reality environment with keyboard video texture and large floating display.

Introduction and Background

Virtual Reality (VR) has experienced a substantial growth of interest over the past years due to the availability of inexpensive headsets and powerful workstations. Today, there is a wide variety of headsets available that use different techniques to enable VR and target different application areas, including gaming or entertainment. At the same time, the performance of today's smartphones has considerably increased. Inserted into a VR viewer, smartphones are capable of presenting interactive VR, Augmented Reality (AR), or Mixed Reality (MR) environments. For the realization of the vision of a virtual office where users can work and collaborate [11], potential interactions with computing systems are essential. Previous research showed that text input is possible while being immersed in VR [8, 10, 14]. However, this requires stationary hardware, complex calibration processes, and specialized hardware components.

Grubert et al. [4] pointed out that text editing requires interaction techniques that are fast and precise. Minimizing the performance gap between a laptop and a VR setup when conducting office work is a crucial challenge. We argued [8] that none of the previously proposed solutions for text input in VR enable high text throughput from real-world typing. To date, we still require convincing input and output modalities for the success of virtual offices.

In this work, we present a low-fidelity apparatus that allows for calibration-free text input and copy editing on a physical keyboard while being immersed in a Virtual Environment (VE). Our apparatus consists of off-the-shelf components, such as a smartphone, VR viewer, and a wireless keyboard. Hence, it is fully portable and ready for use in in-the-wild scenarios. The keyboard and the user's hands are dynamically blended into the VE, allowing for comfortable text input. We focused on a simple smartphone-based setup to identify the minimal requirements that make MR truly accessible. Based on the results of our user study (N=24), we find that text input and editing using a MR setup in combination with a physical keyboard as an input device is more efficient compared to sole touch interaction on smartphones. We conclude that the haptic MR setup compensates for the small screen. Furthermore, we find that frequent touch errors were induced by the fat-finger problem [12] while typing on the smartphone. Finally, the virtual screen provides more space than a smartphone display, hence performing copy editing tasks in MR might be more efficient compared to an external keyboard and the smartphone as the display.

Typing in Smartphone-Based Mixed Reality

The fundamental requirement to realize effortless typing on a physical keyboard in mixed reality is to enable the user to localize and reach out for the keyboard and understand the keyboard's location in relation to their fingers [3, 8].

To investigate the effect typing in a low-fidelity portable mixed reality environment, we implemented our apparatus using a Google Pixel 2 XL as the main component. We incorporated the smartphone with the Google Daydream VR viewer to create a head-mounted display (HMD). To enable six degrees of freedom tracking and to capture the environment, the smartphone's inertial measurement unit and camera are used. Since the heat sink of the VR viewer blocks the camera, we drilled a notch into it.

A wireless keyboard is used for text input. A printed visual marker is attached above the keyboard to enable visual tracking of the keyboard. Following the approach of Feiner et al. [2], we use the smartphone's camera during runtime to create a cropped video texture of the keyboard, which is dynamically anchored to the physical position of the key-

Figure 4: Participants typing text during the user study.
Top: *Mixed Reality* condition;
Bottom: *Smartphone + Keyboard* condition

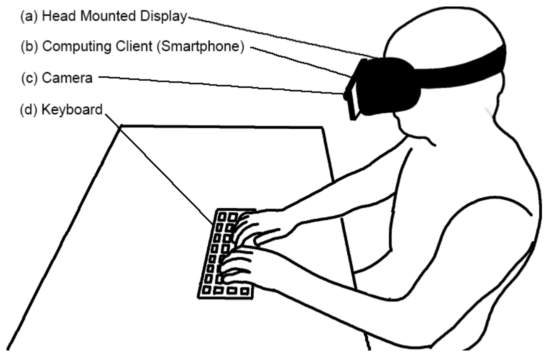

Figure 3: Our mixed reality apparatus for text input comprises a Google Daydream HMD [a], a Google Pixel 2 XL [b], and a wireless keyboard [d].

board within the virtual environment. All components of our apparatus are shown in Figure 3. The virtual environment, including the cropped and arranged video, is demonstrated in Figure 2.

Method

Our mobile apparatus enables users to visually perceive the physical keyboard and their own hands while being immersed in a virtual environment. The objective of the following study is to evaluate the text input and editing performance using a mobile low-fidelity setup in contrast to today's smartphone input. We investigate the overall user experience by assessing system usability scale [1], NASA-TLX [6], and AttrakDiff [7]. We used a 3×1 factorial design with the within-subject variable SETUP. We employed three different levels for SETUP: *Mixed Reality*, *Smartphone + Keyboard*, and *Smartphone*. Both conditions that include the keyboard are shown in Figure 4. The typing performance was measured while employing a physical keyboard using the MR apparatus, the smartphone display, or direct typing using the smartphone soft keyboard.

Subjects

In total, we recruited 24 participants via social media and our university's mailing-list to participate in our user study. The participants (six female) were aged from 19 to 38 (M = 27, SD = 4.66). Five participants were wearing corrective lenses during the study. Participants received either 5 EUR or course credits as compensation for their participation.

Apparatus

The apparatus for this study comprised thee individual setups sharing the same three, but individual combination of components: smartphone, keyboard, and MR HMD. The latter was only facilitated for the *Mixed Reality* condition.

Smartphone The smartphone setup served as a baseline and consisted only of a Google Pixel 2 XL running Android Pie. The smartphone was running our application in portrait mode showing the stimulus and text edit field at the top of the screen and below the google stock soft keyboard.

Smartphone + Keyboard For the second setup, we facilitated an Apple Magic Keyboard, which pairs wireless with the smartphone. This time the smartphone is placed above the keyboard in landscape mode serving as a portable display showing only the stimulus.

Mixed Reality For the mixed reality setup, we used our developed MR apparatus comprising the modified Google Daydream View 2, smartphone, and keyboard. We designed a virtual environment showing a room with a large screen displaying the stimulus. The cropped video of the physical keyboard and hands is displayed within the virtual environment at the corresponding physical location.
All smartphone applications were developed with the Unity game engine 2018.3. For head and keyboard tracking, we employed the Vuforia Engine 7.5.

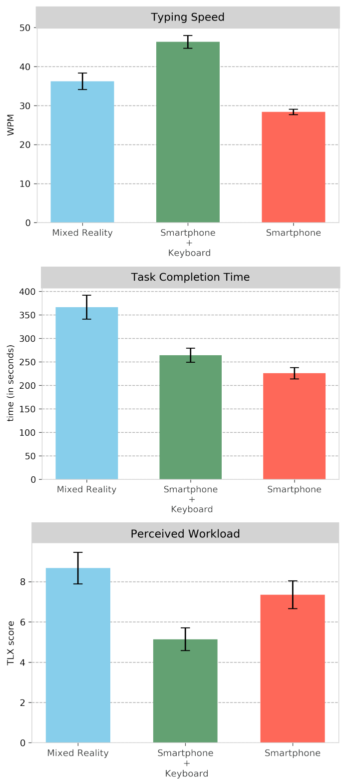

Figure 5: Mean values of words per minute (text input task), task completion time (copy edit task), and NASA-TLX score (both tasks) for each condition. Error bars show the standard error of the mean (SE).

Task

In the user study, participants had to accomplish two simple tasks. First, a simple text input task and second a copy editing task requiring to remove spelling error and adding or removing words. Participants started in a resting position with their hands placed next to the keyboard or smartphone. While being in this pose, a 3-second countdown elapsed on the smartphone or virtual display, indicating the start of either the text input or copy editing task.

Text Input For the text input task, a random sentence from the MacKenzie and Soukoreff [9] phrase set was displayed. Participants were asked to enter the phrase as fast and accurately as possible. Participants could correct errors during input but were also allowed to confirm inaccurate or incomplete phrases. With the enter key, participants confirmed the input, and the next phrase was displayed. For each condition, participants performed three sets of ten phrases. The task was the same for all conditions.

Copy Editing For the copy editing task, the participants had to review and correct three different texts. Each text consisted of 12 modified sentences from the MacKenzie and Soukoreff [9] phrase set. The required corrections were indicated between the lines highlighted in green. Participants were asked to edit as fast as possible all corrections. Except for the Mixed Reality condition, the edit cursor could be placed by touching the screen or with the arrow keys of the keyboard. We compensate for potential complexity differences by counterbalancing the prepared texts across all conditions.

Procedure

After welcoming the participants, we asked them to sign the consent form and explained the apparatus as well as the course of the study. Afterward, we asked participants to put on the HMD to adjust it to the head for the best visual results. Before starting with the typing task, participants were asked to get familiar with the virtual environment and get used to the tracking and visualization of the keyboard. After finishing both tasks (input and copy edit), participants had to fill out the RAW NASA-TLX [6], the AttrakDiff, and the System Usability Scale (SUS) [1] questionnaire. This procedure was subsequently repeated for all conditions. The first set of ten phrases at the start of each condition was a practice set to familiarize the participant with the apparatus. We did not include this set in our analysis. SETUP was presented in a counterbalanced order using a full Latin square to prevent sequence effects. After finishing the third iteration, we conducted a short semi-structured interview and asked for comments about their performance, user experience, and personal preference. Including the debriefing, participants completed the study between 60 to 90 minutes.

Results

We conducted multiple one-way repeated measure analyses of variance (RM-ANOVA) in order to reveal statistically significant effects of the within-subjects variables SETUP. All significance levels are set to $\alpha = .05$.

Words Per Minute (WPM)

For the text input task, participants entered a total of 2160 sentences. Since we discarded the first ten sentences of each participant, only 1440 sentences were used for analyses. We used the logged keystrokes to calculate the WPM by dividing the length of the final input by the time required to input the presented phrase [13]. The calculated WPM provides a measure for the average typing performance. We found a significant effect of SETUP on the typing speed, $F(1.23, 28.28) = 31.22$, $p < .001$. Furthermore, post hoc tests revealed a significant difference between the conditions *Smartphone + Keyboard* and *Smartphone* (M = 17.97,

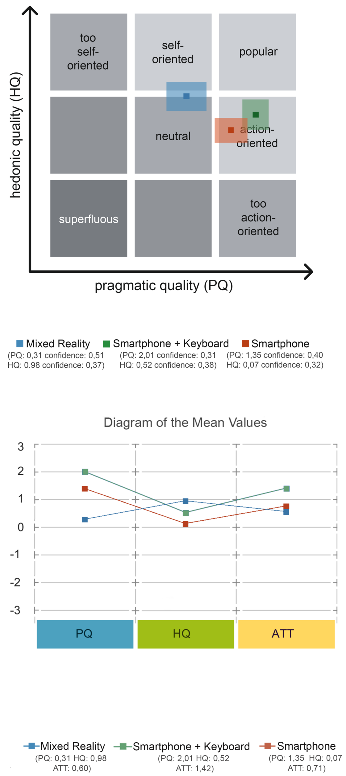

Figure 6: Diagrams from the AttrakDiff questionnaire revealing the characteristics including the pragmatic quality (PQ) and hedonic quality (HQ) (top diagram) and the mean values of the dimensions (bottom diagram).

SE = 2.18, with $p < .001$), between *Smartphone + Keyboard* and *Mixed Reality* (M = 10.14, SE = 1.37, with $p < .001$) and between *Smartphone* and *Mixed Reality* (M = -7.82, SE = 2.99, with $p = .046$).

Error Rate

Besides the WPM, the typing and editing performance can also be expressed through the *Error Rate*. We calculated the ratio of the length of the input and the minimum number of insertions, deletions, or substitutions that are needed to transform the presented text into the transcribed on [9]. The results neither show a significant effect of SETUP on the *Error Rate* for the typing task, $F(1.81, 41.68) = 1.109$, $p = .339$ nor for the copy editing task $F(1.93, 44.40) = .702$, $p = .496$. Besides, we calculated the *Corrected Error Rate* [13], which represents the effort put into correcting errors. We found no significant effect of SETUP regarding the number of corrections, $F(1.37, 31.41) = 0.301$, $p = .658$.

Task Completion Time (TCT)

For the copy editing task, we measured the TCT as a performance indicator. We measured from the very first keypress to the confirmation keypress for each text. We found a significant main effect of SETUP on the TCTs of the copy editing task, $F(2, 46) = 25.86$, $p < .001$. A post-hoc tests revealed significant differences between *Smartphone + Keyboard* and *Mixed Reality* (M = -102.41, E = 21.93, with $p < .001$), between *Smartphone* and *Mixed Reality* (M = -140.60, SE = 20.92, with $p < .001$), but no significant effect between *Smartphone + Keyboard* and *Smartphone* ($p = .120$).

Task Load Index

We assessed the raw score of the NASA-TLX [5], representing the perceived subjective workload the participants had while inputting or copy editing text. We found a significant main effect of SETUP on the perceived workload, $F(1.61, 36.93) = 13.83$, $p < .001$. Post-hoc tests revealed a significant difference between *Smartphone + Keyboard* and *Smartphone* (M = -2.22, SE = 0.52, with $p < .001$), between *Smartphone + Keyboard* and *Mixed Reality* (M = -3.54, SE = 0.63, with $p < .001$), but no significant effect between *Smartphone* and *Mixed Reality* ($p = .393$).

System Usability Scale (SUS)

To receive an indication of the overall usability of our apparatus, we assessed the SUS [1]. We found a significant effect of SETUP, $F(1.74, 39.97) = 32.70$, $p < .001$. Post-hoc tests revealed a significant difference between the conditions *Smartphone + Keyboard* and *Mixed Reality* (M = 23.16, SE = 3.29, with $p < .001$), between *Smartphone* and *Mixed Reality* (M = 19.27, SE = 3.51, with $p < .001$), but no significant difference between *Smartphone + Keyboard* and *Smartphone* ($p = .295$).

AttrakDiff

To gain further insights into the perceived user experience, we used the AttrakDiff questionnaire, which accesses the user experience divided into pragmatic and hedonic quality. Participants rated the system by ranking word pairs of different dimensions. The results are shown in Figure 6. The top diagram classifies the apparatus into character areas (i.e., *self-oriented* or *action-oriented*). The bottom diagram shows the mean values of the dimensions of AttrakDiff. The results show that the *Mixed Reality* setup has the highest hedonic quality, but the lowest pragmatic quality. According to the diagram, the characteristics of the apparatus is not unambiguous and lies between the areas *neutral* and *self-oriented*. The other two setups, *Smartphone + Keyboard* and *Smartphone* lie in the characteristics area of *action-oriented*, thus were rated more practical.

Personal Preferences and Qualitative Results

After conducting the user study, we asked the participants regarding their preferred SETUP and to provide additional qualitative feedback. Participants ranked *Smartphone + Keyboard* as the best solution for portable text input and editing, followed by the *Mixed Reality*, which is directly followed by the *Smartphone* setup. Participants endorsed the great display-space and privacy in MR, however, complained about occasional orientation problems due to the limited field of view of the HMD.

Discussion and Limitations

Considering text input, we found that our mixed reality apparatus led to significant higher words per minute compared to soft keyboard input. Results did not show significant changes in the error rates of the typed text. Further analysis revealed that the slightly higher workload and lower usability caused by the HMD was mainly compensated through the support of the physical keyboard. For copy editing texts, the mixed reality led to a significantly higher task completion time (TCT) compared to both the smartphones soft keyboard and the smartphone and keyboard combination. Further, the analysis revealed that participants benefit from the large virtual display space but got thwarted by the lacking opportunity to quickly navigating the text (e.g., touch or mouse). Adding mouse support or alternative methods to place the cursor quickly might have yielded different results considering the TCT. The analysis of additional qualitative feedback unfolded that participants overall enjoyed our apparatus. They envisaged working in mixed reality and highlighted the larger display area and the possibility to collaborate in future scenarios. We argue that optimizing the setup and further improve the interaction modalities is necessary. Improved positioning of the keyboard visualizations and multimodal input for copy editing are relevant parameters to improve portable mixed reality text entry.

Conclusions

In this paper, we investigated a portable low-fidelity solution for text input in mixed realities. Our off-the-shelf apparatus comprises a smartphone, a virtual reality viewer, and a wireless keyboard. In a user study with 24 participants, we compared state-of-the-art smartphone soft keyboards to physical keyboard input and our mixed reality approach. We compared typing performance, error rate, task completion time, subjective workload, overall usability, and user experience.

The results show that participants have significantly higher input speeds when being immersed in mixed reality compared to smartphone input, while error rates remain low. In contrast, copy editing required considerably more time to complete, but participants enjoyed interacting with the large virtual display.

Already today, the combination of portable virtual reality viewer and current smartphones allow us to have virtual mobile offices. We believe that future portable mixed reality systems can fully support us by simulating well-known but highly flexible virtual environments while being on the move.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research as part of the project Begreifen (Grant No. 16SV7527) and KoBeLU (Grant No. 16SV7599K).

REFERENCES

[1] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (Nov 1995). `https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale`

[2] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the world: 2 D windows for 3 D augmented reality. In *ACM Symposium on User Interface Software and Technology*. 145–155.

[3] Anna Maria Feit, Daryl Weir, and Antti Oulasvirta. 2016. 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*. ACM Press. DOI: http://dx.doi.org/10.1145/2858036.2858233

[4] Jens Grubert, Eyal Ofek, Michel Pahud, and Per Ola Kristensson. 2018. The Office of the Future: Virtual, Portable, and Global. *IEEE Computer Graphics and Applications* 38, 6 (Nov 2018), 125–133. DOI: http://dx.doi.org/10.1109/mcg.2018.2875609

[5] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (Oct 2006), 904–908. DOI: http://dx.doi.org/10.1177/154193120605000909

[6] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. DOI: http://dx.doi.org/https://doi.org/10.1016/S0166-4115(08)62386-9

[7] Marc Hassenzahl, Michael Burmester, and Franz Koller. 1998. AttrakDiff work model and measuring tool. (1998). http://attrakdiff.de/index-en.html

[8] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press. DOI:http://dx.doi.org/10.1145/3173574.3173919

[9] I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 754–755. DOI: http://dx.doi.org/10.1145/765891.765971

[10] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, 2143–2152. DOI: http://dx.doi.org/10.1145/2702123.2702382

[11] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. 1998. The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '98)*. Association for Computing Machinery, New York, NY, USA, 179–188. DOI: http://dx.doi.org/10.1145/280814.280861

[12] Katie A. Siek, Yvonne Rogers, and Kay H. Connelly. 2005. Fat Finger Worries: How Older and Younger Users Physically Interact with PDAs. In *Human-Computer Interaction - INTERACT 2005*, Maria Francesca Costabile and Fabio Paternò (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 267–280.

[13] R. William Soukoreff and I. Scott MacKenzie. 2003. Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 113–120. DOI: http://dx.doi.org/10.1145/642611.642632

[14] James Walker, Bochao Li, Keith Vertanen, and Scott Kuhl. 2017. Efficient Typing on a Visually Occluded Physical Keyboard. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5457–5461. DOI: http://dx.doi.org/10.1145/3025453.3025783