Typing in Mixed Reality: Does Eye-Tracking Improve Performance?

Cecilia Schmitz
Michigan Technological University
Houghton, MI, USA
cmschmit@mtu.edu

Scott Kuhl
Michigan Technological University
Houghton, MI, USA
kuhl@mtu.edu

Keith Vertanen
Michigan Technological University
Houghton, MI, USA
vertanen@mtu.edu

ABSTRACT

Accuracy and speed are pivotal when it comes to typing. Mixed reality headsets offer users the groundbreaking ability to project virtual objects into the physical world. However, when typing on a virtual keyboard in mixed reality space, users lose the tactile feedback that comes with a physical keyboard, making typing much more difficult. Our goal was to explore the capability of users to type using all ten fingers on a virtual key in mixed reality. We measured user performance when typing with index fingers versus all ten fingers. We also examined the usage of eye-tracking to disable all keys the user wasn't looking at, and the effect it had on improving speed and accuracy. Our findings so far indicate that, while eye-tracking seems to help accuracy, it is not enough to bring 10 finger typing up to the same level of performance as index finger typing.

CCS CONCEPTS

• **Human-centered computing → Text input**; **Augmented reality (AR)**.

ACM Reference Format:
Cecilia Schmitz, Scott Kuhl, and Keith Vertanen. 2023. Typing in Mixed Reality: Does Eye-Tracking Improve Performance?. In *Infinite Loop, 2023, Houghton, MI*. ACM, New York, NY, USA, 3 pages.

1 INTRODUCTION

While mixed reality (MR) devices are still under heavy development, we see it as a possibility that MR will grow to support day-to-day tasks such as writing emails or browsing the internet. However, the current difficulties with text entry in MR may act as a barrier for performing these tasks. Auxiliary input devices such as handheld controllers [2, 11] or physical keyboards [4, 6, 8, 13] can help alleviate the difficulties with text entry, but puts additional strain on users to acquire, setup, and transport these devices in addition to an MR headset. Speech-to-text has also been proposed as a solution [1], but this can raise issues with privacy, and won't work well for difficult to predict text such as usernames or passwords. Some have attempted to reinvent text-entry entirely, for example, measuring only which finger was tapped as opposed to which key, and using that information coupled with predictive technology to guess what the user is typing [5]. While these results are promising, they also carry the burden of adding an additional learning curve onto the usage of MR, potentially raising the barrier to entry surrounding the technology. Our goal is to recreate the traditional QWERTY keyboard in MR such that it is more usable, allowing users to utilize their existing typing skills.

Our goal for this study is twofold. First, we want to explore the capability of users to type using all ten fingers. Most MR text entry solutions do not involve ten finger typing on a virtual keyboard [1, 3, 7]. An interesting method of 10-finger test entry in MR called TapGazer was proposed by researchers Zhenyi He, Christof Lutteroth, and Ken Perlin, however this still required an auxiliary input device to detect which finger was tapping [5]. 10-finger typing akin to how one would type on a physical keyboard has remained an unexplored concept in regards to virtual keyboards. However, our initial experimentation revealed that 10-finger typing on a virtual keyboard was extremely difficult, often leading to accidental key presses, similar to the Midas Touch problem [10] observed when making selections using eye-tracking. This led to our secondary goal: To determine if eye-tracking could improve accuracy when using ten fingers on a virtual keyboard.

Eye-tracking has been used as a supplement or replacement to typing in virtual and mixed reality [7, 9]. Head rotation has also been used as input to develop alternate typing strategies in these environments [11, 14]. We theorized that, due to the lack of tactile feedback on a virtual keyboard, users would be more likely to gaze at the key they intended to press. Therefore, by tracking where the user was looking, we could predict which key they wanted to press and disable other keys, thus increasing the accuracy with which they could type. Eye-tracking has been used to supplement navigation of menus in VR with relative success [10], demonstrating the potential usefulness of eye-tracking in creating user interfaces. This approach is also similar to another approach, where a gaze 'cursor' was used as a supplement to hand-tracking input to determine which key to press [7]. However, the use of dwell time in this model increased the eyestrain experienced by users [7, 10]. It is our hope that by avoiding the use of dwell times in our model, users will be able to type more accurately without experiencing increased eyestrain.

2 USER STUDY

We currently have had 12 participants complete four counterbalanced conditions:

- **Index** — Participants were instructed to use only their index fingers to type. Only the index finger colliders were enabled.
- **IndexEye** — Participants were instructed to use only their index fingers to type, and to look at the keys as they typed. Only the index finger colliders were enabled. Keys outside

Infinite Loop, 2023, Houghton, MI

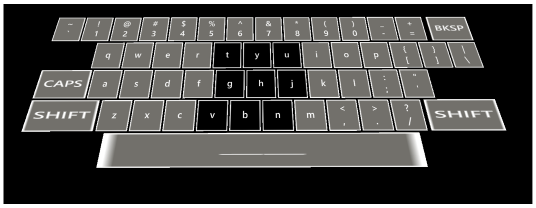

Figure 1: During the IndexEye and TenEye conditions, inactive keys would be grayed out, and only the keys near the users' gaze positions would appear transparent.

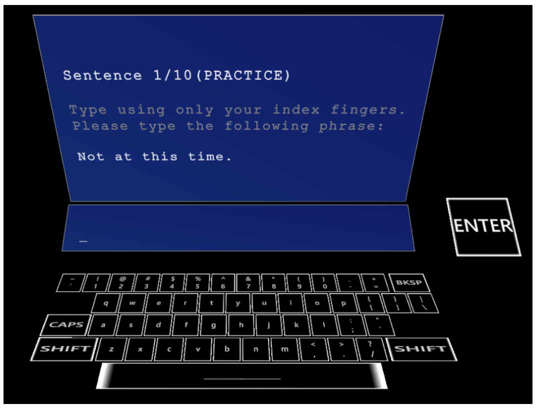

Figure 2: The virtual keyboard used during the experiment.

of the radius around the participants' gaze location would be grayed out and unable to be pressed (Figure 1.

- **Ten** — Participants were instructed to use all ten fingers to type. The colliders on all ten fingers were enabled.
- **TenEye** — Participants were instructed to use all ten fingers to type, and to look at the keys as they typed. The colliders on all ten fingers were enabled. Keys outside of the radius around participants' gaze location would be grayed out and unable to be pressed (Figure 1).

Participants used a Microsoft MR HoloLens 2 headset while seated on a chair. We spawned in the virtual keyboard in front of the participant and had participants adjust the height of the chair such that their hands could hover comfortably over the keys. All conditions would visually highlight keys and play an audio sound to signal that the key was pressed. The IndexEye and TenEye conditions provided additional feedback by graying out keys that were unable to be pressed. We used a full QWERTY keyboard including numbers, symbols, shift, caps lock, and backspace. The size of the keyboard was approximately 35 cm × 12 cm, with each letter key being approximately 2 cm × 2 cm. The keyboard was deterministic and did not provide auto-correct or word predictions. The full keyboard layout can be seen in Figure 2 Participants saw their current text above the keyboard and could correct any errors using backspace. A key was triggered when a participant's finger pushed a key downward past a threshold. In the event that multiple keys were pressed at the same time, only the furthest pressed key would be registered as pressed.

During the one-hour study, participants completed an initial survey and then typed sentences in each of the four conditions. Between each condition, participants took a two-minute break and completed a survey about the previous condition. Participants also completed a final survey. At the start of each condition, participants typed two practice sentences and eight evaluation sentences. We used sentences from the "mem1-5" set from the mobile Enron dataset [12]. Sentences contained upper and lowercase letters, punctuation, and numbers. Participants could not receive the same sentence twice during a given condition, but there was a chance sentences could be repeated over the course of the study. Participants typed with both hands. During the IndexEye and TenEye conditions, participants could only press keys within an approximate 1-key radius of their detected gaze location. All other keys were grayed out and could not be pressed.

3 RESULTS

We measured entry rate in words per minute (WPM), with a word being five characters including space. We timed from the first key press until the enter button press. Average entry rates were 11.0, 9.8, 7.1, and 7.4 WPM in Index, IndexEye, Ten, and TenEye respectively (Figure 3). We also measured what percentage of the time users were looking at keys during the Index and Ten conditions, such that the keys would have been active in the IndexEye and TenEye conditions. Users looked at the keys on an average of 83.3% of key presses during the IndexEye condition and 84.2% of key presses during the TenEye condition.

We also measured the number of backspace presses per output character. This averaged out to be 0.10, 0.05, 0.30, and 0.14 for Index, IndexEye, Ten, and TenEye respectively (Figure 4.

Users also filled out questionnaires after completing each condition. Users were asked to rank on a scale from 1–7 how quickly they felt they could type, how accurately they felt they could type, and how physically straining they found typing to be (including eyestrain). Lower numbers represented less speed/accuracy and greater physical strain, and higher numbers represented more speed/accuracy and less physical strain. The Index condition had average scores of 3.8 for speed, 5.2 for accuracy, and 4.3 for physical strain. The IndexEye condition had average scores of 4.0 for speed, 5.7 for accuracy, and 4.3 for physical strain. The Ten condition had averages scores of 3.0 for speed, 3.8 for accuracy, and 3.7 for physical strain. The TenEye condition had average scores of 3.1 for speed, 4.5 for accuracy, and 4.4 for physical strain.

In the final questionnaire, participants were asked to rank the conditions. Seven of the twelve participants ranked Index Finger Only first. The second most preferred condition, by four participants, was Index Finger ET.

4 DISCUSSION AND CONCLUSION

We found Index Finger Only had the highest WPM, making it the quickest keyboard to type on. Eye-tracking improved participant WPM when participants used all ten fingers, but the improvement was negligible with a difference of only 0.3 WPM. Eye-tracking improved participant accuracy scores in both the index finger and ten finger conditions without lowering physical strain scores. Additionally, the number of backspace presses per output character

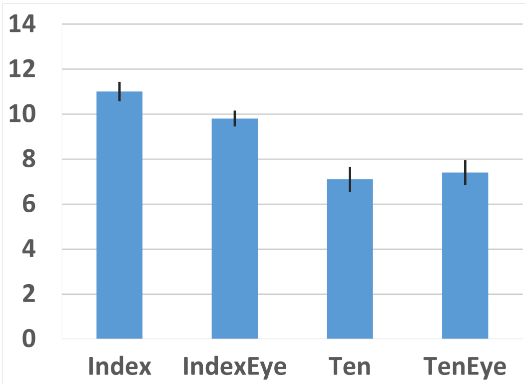

Figure 3: Average WPM by condition. Error bars represent standard error.

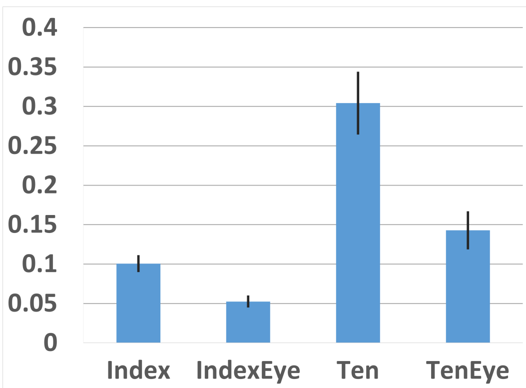

Figure 4: Average number of backspace presses per output character, separated by condition. Error bars represent standard error.

decreased from Index to IndexEye and Ten to TenEye, further supporting that eye-tracking improves accuracy when typing. This supports our hypothesis that, due to the lack of tactile feedback on virtual keyboards, users' gazes will naturally be drawn to the key they intend to press. Therefore, our implementation of eye-tracking can improve performance without hindering users. However, WPM was still lower in the eye-tracking conditions despite users perceiving and results showing increased accuracy. The accuracy improvement was also not reflected in participant preference, with the overwhelming majority showing a preference for the non-eye-tracking index finger keyboard.

ACKNOWLEDGMENTS

This material is based upon work supported by the NSF under Grant No. IIS-1909089.

REFERENCES

[1] Jiban Adhikary and Keith Vertanen. 2021. Text Entry in Virtual Environments using Speech and a Midair Keyboard. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2648–2658. https://doi.org/10.1109/TVCG.2021.3067776

[2] Costas Boletsis and Stian Kongsvik. 2019. Controller-based Text-input Techniques for Virtual Reality: An Empirical Comparison. *International Journal of Virtual Reality* 19, 3 (Oct. 2019), 2–15. https://doi.org/10.20870/IJVR.2019.19.3.2917

[3] John J. Dudley, Keith Vertanen, and Per Ola Kristensson. 2018. Fast and Precise Touch-Based Text Entry for Head-Mounted Augmented Reality with Variable Occlusion. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 6, Article 30 (12 2018), 40 pages. https://doi.org/10.1145/3232163

[4] Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Ola Kristensson. 2018. 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)*. 159–166.

[5] Zhenyi He, Christof Lutteroth, and Ken Perlin. 2022. TapGazer: Text Entry with Finger Tapping and Gaze-directed Word Selection. *Conference on Human Factors in Computing Systems - Proceedings*, Article 337 (2022), 16 pages. https://doi.org/10.1145/3491102.3501838

[6] 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*. 1–9.

[7] Mathias N. Lystbæk, Ken Pfeuffer, Jens Emil Sloth Grønbæk, and Hans Gellersen. 2022. Exploring Gaze for Assisting Freehand Selection-based Text Entry in AR. *Proceedings of the ACM on Human-Computer Interaction* 6, ETRA, Article 141 (May 2022), 16 pages. https://doi.org/10.1145/3530882

[8] 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 SIGCHI Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 2143–2152. https://doi.acm.org/10.1145/2702123.2702382

[9] Diogo Pedrosa, Maria Da Graça Pimentel, Amy Wright, and Khai Truong. 2015. Filteryedping: Design Challenges and User Performance of Dwell-Free Eye Typing. *ACM Transactions on Accessible Computing* 6, 1, Article 3 (2015), 37 pages. https://doi.org/10.1145/2724728

[10] Ken Pfeuffer, Lukas Mecke, Sarah Delgado Rodriguez, Mariam Hassib, Hannah Maier, and Florian Alt. 2020. Empirical Evaluation of Gaze-enhanced Menus in Virtual Reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST)*. https://dl.acm.org/doi/pdf/10.1145/3385956.3418962

[11] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based Text Entry in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[12] Keith Vertanen and Per Ola Kristensson. 2011. A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails. In *MobileHCI '11: Proceedings of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden).

[13] James Walker, Bochao Li, Keith Vertanen, and Scott Kuhl. 2017. Efficient Typing on a Visually Occluded Physical Keyboard. In *CHI '17: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, CO). 5457–5461.

[14] Chun Yu, Yizheng Gu, Zhican Tang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. Tap, dwell or gesture?: Exploring head-based text entry techniques for HMDs. *Conference on Human Factors in Computing Systems - Proceedings* (2017), 4479–2288. https://doi.org/10.1145/3025453.3025964