Text Input Performance with a Mixed Reality Head-Mounted Display (HMD)

Jessyca L. Derby, Christopher T. Rarick, & Barbara S. Chaparro
Department of Human Factors and Behavioral Neurobiology
Embry-Riddle Aeronautical University, Daytona Beach, FL

Past research has shown that text input methods must be quick, efficient, and involve little to no learning in order for them to be accepted in a consumer market. The Microsoft HoloLens is a mixed reality head-mounted display (HMD) that requires gestures or a physical clicker for interaction, including text entry. In this study, we investigated performance, perceived usability, and workload of these text input methods. Participants were able to type only 5.41WPM with the gesture method, gave it a poor usability score. The clicker method resulted in a WPM rate of 6.58, was given an “OK” usability score, and lower perceived exertion scores in the bicep and index finger. This study demonstrates the need for further research to develop more optimal text entry methods for MR environments.

INTRODUCTION

Virtual reality (VR), augmented reality (AR), and mixed reality (MR) are emerging technologies that allow individuals to experience something completely unique. VR is commonly recognized as a head-mounted display (HMD) device where the user is immersed in a virtual environment (VE). The user’s environment, objects around them, their interactions, and even themselves can be computerized. AR layers computerized objects, called holograms, on top of the real world. Through the use of devices such as Google Glass, users can see the environment around them as well as the holograms. MR fills in the gap between VR and AR by creating holograms that adapt to the user’s real environment through natural interactions such as occlusion, depth, and physics. MR can be experienced through cellphone devices or even HMDs, such as the Microsoft HoloLens. This device has been described as the work place of the future that can change the way that we work in academic and industrial settings such as engineering, healthcare, and design (Wright, 2018).

In order to understand the general utility of the HoloLens as an office device, we conducted a usability study in which participants were asked to complete several office tasks such as sending an email and creating a PowerPoint. Participants reported that the text entry tasks were the most difficult and reported to be inaccurate, slow, and undesirable for typing long messages (Shelstad et al., 2019). In order for this tool to be accepted in a mass market, text entry methods must be quick, involve little to no learning, and allow for higher efficiency through practice (Zhai & Kristensson, 2012). The HoloLens has the capability to use gestures, voice, and a clicker to interact with a virtual QWERTY keyboard for text entry. Each letter is selected through head movement and inputted by either using a gesture (the user does this by holding up their arm into the 30 degree horizontal and 17.5 degree vertical field of view (FOV) of the HoloLens and then using one finger to complete a tapping motion, resulting in a “click”), by pressing down on the clicker, which essentially acts as a left mouse button, or through dictation and voice recognition. These techniques can be cumbersome, tiring, and not nearly as fast or accurate as text input on other devices using a traditional QWERTY keyboard (Knierim et al., 2018; Ahn, Heo, & Lee, 2017). Participants have been found to type up to 58 WPM on a traditional QWERTY keyboard, 32.5 WPM on QWERTY cell phone keyboards, and 27-34 WPM on swipe or tap smartwatch keyboards (Sears, 1991; Sears et al., 1993; Turner, Chaparro, & He, 2016).

Some researchers, such as Lee et al., have tested typing performance with novel keyboards on the HoloLens (Lee, Lam, Yau, Braud, & Hui, 2019). They trained participants for 8 days on their alphabetical scrolling keyboard, HIBEY, and the default HoloLens QWERTY keyboard. They found that participants achieved 4 WPM without practice and 6.5 WPM with practice on the default keyboard (2019). This was increased to 13 WPM when trained on the HIBEY keyboard (Lee et al., 2019). Other research has shown comparable text entry rates for single joysticks at 6.5 WPM (dual joysticks increase this rate to 7 WPM) and navigating virtual keyboards using a television remote which has been observed as 5.5 WPM initially and 7 WPM with practice (Wilson & Agrawala, 2006; Perrinet et al., 2011).

An alternative text input method is speech-to-text. Speech as a text entry method can be extremely effective, especially when compared to smartphone keyboards (Ruan, Wobbrock, Liou, Ng, & Landay, 2016). According to Ruan et al., speech input was about three times faster when compared to typing on a mobile phone keyboard (161 WPM vs. 53.5 WPM). Participants in our preliminary study also reported that they would rather use voice than traditional text entry methods with the HoloLens. However, one stipulation of using voice as an input method is that not all situations allow for its use. For example, it could be difficult to use voice to text while in a loud office setting. This ambient noise sensitivity is a severe limitation (Grubert et al., 2018). Other issues with voice recognition include lack of privacy, security, and confidentiality in social settings (Ruan, Wobbrock, Liou, Ng, & Landay, 2016).

The HoloLens currently has over 300 applications available to use including games, holographic demos, communication apps such as Skype, and office apps like

Google Chrome, Email, and Microsoft Office Preview. Many of these apps require text entry for basic interaction. Without text input techniques that are fast, efficient, and easy to learn it could become frustrating to use all of the features the HoloLens has to offer. This study assessed the current state of HoloLens text input techniques to establish a baseline level of performance and accuracy. This baseline serves as a reference point for future input methods that may be developed for the HoloLens and similar devices.

Current Study

The current study assesses the usability, performance, and user preference for two text input methods (Gesture or Clicker) using the Microsoft HoloLens. We hypothesized that participants were going to prefer the clicker input method over the gesture method because it would be more efficient and less physically demanding.

METHOD

Participants

Participants included 17 college students (eleven males and six females) with ages ranging between 18 to 24. Eleven participants stated that they had used VR or AR before. This experience was very minimal (Mdn = 1 hour, IQR = 1 hour).

Experimental Design

A repeated-measures experimental design was used for this study. Both qualitative and quantitative data was collected and all participants were asked to input text using both the clicker and gesture methods. Participants were randomly assigned to either use the clicker or gesture method first. The independent variable for this study was the input method used (either the clicker or gesture method) and the dependent variables included: typing speed, accuracy, perceived workload, perceived exertion, perceived eye strain, perceived usability, and preference.

Measures

Typing Speed and Accuracy. Words per minute (WPM), adjusted words per minute (AdjWPM), and word error rate (WER) were calculated to evaluate participants' typing performance. AdjWPM was calculated by multiplying the percentage of errors made in a phrase by the calculated WPM of that same phrase. WER was further defined by the type of error made: substitution, insertion, and omission error rate. Substitution errors occurred when participants replaced a word that was in the original phrase with a different word. Insertion errors occurred when participants added an extra word that was not in the original phrase. Omission errors occurred when participants did not type a word that was in the original phrase.

Perceived Workload. The multidimensional NASA Task Load Index (NASA TLX-R) scores were used to evaluate participants' subjective workload and performance (Hart & Staveland, 1988). The NASA TLX-R is a 6-item questionnaire that asks participants to rate each item 21-point scale. Each individual item is related to one of six dimensions: physical demand, mental demand, temporal demand, performance, effort, and frustration. A higher score signifies that the participant had a more demanding experience or worse perceived performance.

Perceived Exertion. The Borg Category Ratio Scale (Borg CR10) was used to evaluate the perceived exertion after using each input method (Borg, 1998). This scale asks participants to rate their current level of exertion from "nothing at all" (0) to "extremely strong" (10) or even "absolute maximum" which can be rated as a 12, 13, or even higher. This was paired with an upper-body map and participants were asked to rate 25 areas on this scale.

Perceived Eye Strain. Six five-point Likert-scale questions were given to participants. These questions asked participants to rate their ability to concentrate, the ease of reading text, text clarity, physical fatigue, mental fatigue, and level of eye strain from very high to very low.

Perceived Usability. The System Usability Scale (SUS) was used to evaluate the participant's perceived usability of each input method (Brooke, 1996). The SUS is a standardized 10-item questionnaire that asks participants to rate each question on a scale of 1-5, from strongly disagree to strongly agree. This is calculated to a final score between 0-100, and can be compared to an adjective rating scale that ranges from worst imaginable (a score between 0 and 25) to best imaginable (a score of 100) (Brooke, 1996; Bangor, Kortum, & Miller, 2009). Some of the questions of the SUS were modified by changing the subject of the question from "system" to "input method".

Preference and Recommendations. Participants were interviewed at the end of the study for their recommendations and preference for each text entry technique. Participants were asked to report which input method allowed for more quick and/or accurate input. In addition to this, they rated their preference for each input method on a 51-point scale (from 0 to 50). Participants were also prompted to give their own feedback about each of the input methods.

Materials

Microsoft HoloLens. The Microsoft HoloLens headset is a mixed reality, wireless, head-mounted display that was released for developers in 2016. Holograms are projected to the HoloLens lenses and is tracked onto a location in the real world. This creates the illusion that the user is viewing the 3D hologram in the real world. Version 10.0.17134.80 (April 2018 Update) was used for this study. Figure 1 depicts the HoloLens and clicker.

HoloLens Clicker. The clicker comes with the HoloLens and can be used to interact with holograms in place of gesturing. It is used similarly to a mouse. The user can move their head to focus on a hologram and press down on the clicker to interact with it. The clicker is depicted on the left in figure 1.

Figure 1. Microsoft HoloLens and Clicker

Text Input Task. The text input task included a subset of MacKenzie's standardized 500 phrase set. (MacKenzie & Soukoreff, 2003). Participants were asked to speak the phrase aloud and type it into a text box using their assigned input method. Participants also clicked on a "start" and "finish" button to record their typing speed. They completed a training (5 phrases) and evaluation (15 phases) for the two input methods. Participants completed a total of 40 phrases.

Procedure

Participants were recruited through the university's online research pool. After giving their informed consent, participants completed a HoloLens tutorial and were randomly assigned to enter in text using either the clicker or gesture input method. They were asked to reproduce 20 phrases from the MacKenzie phrase set using their assigned input method. This included five practice phrases and 15 test phrases that would later be evaluated. The participants did not have any time constraints.

Upon completion of the text input task, participants were asked to fill out questionnaires (the SUS, NASA-TLX-R, BORG CR10, and eye strain) , and then completed the task again with new phrase sets and the other input method. Participants were then asked to fill out the questionnaires again based on the second input method and were asked about their preferences and opinions about each input method. The study took approximately 90-120 minutes to complete.

RESULTS

To assess the differences between groups, paired samples t-tests were used (assumptions were satisfied for these parametric tests). Whenever it was necessary, Bonferroni corrections were used to correct for familywise error.

Typing Speed

There was a significant difference in average WPM between the gesture method (M= 5.41, SD = 0.888) and the clicker method (M = 6.58, SD = 0.751), t(12)= -3.971, p = .002, *d* = 0.1. There was also a significant difference in average AdjWPM between the gesture method (M=5.28, SD = 0.865) and the clicker method (M=6.46, SD=0.753), t(12)= -3.881, p = .002, *d* = 0.1.

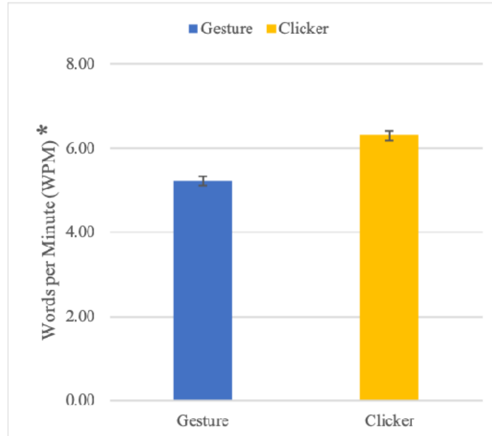

Figure 2. Typing Speed (WPM) Comparison

Accuracy

There was not a significant difference in average WER between the gesture method (M= 0.123, SD = 0.115) and the clicker method (M = 0.104, SD = 0.126), t(12)= .551, p=.591, *d* = .093; nor was there a significant difference in any of the types of errors made, p>.05.

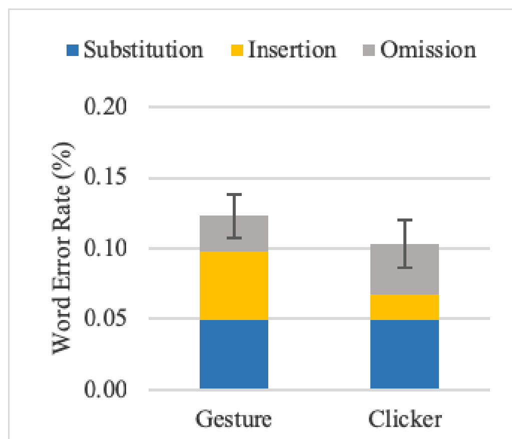

Figure 3. Accuracy (WER) Comparison

Perceived Workload and Exertion

No significant differences were found p > .008 in the subscales of the NASA TLX-R. However, there were trends that showed that the clicker method was perceived as less physical demanding, effortful, and frustrating than the gesture method.

The Borg CR10 results only included right-handed participants (n=14) because there were not enough left-handed participants to analyze (n=3). Significance was found in the right bicep t(13) = 3.801, p=.002, *d* = -1.119 and right index finger t(13) = 4.989, p < .0001, *d* = -1.168. Figure 5 shows the means and standard deviations for these scores.

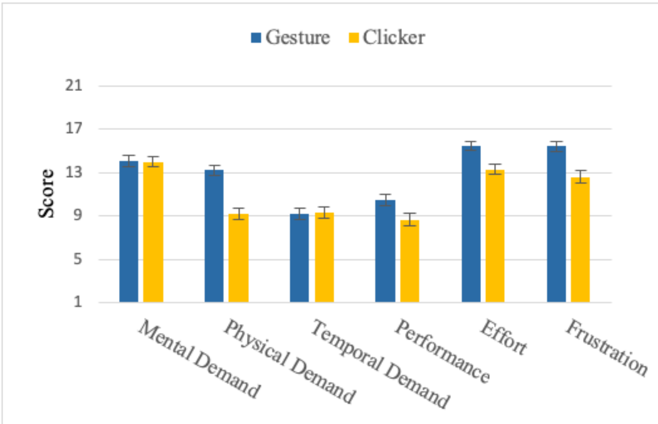

Figure 4. NASA TLX-R Perceived Workload Comparison

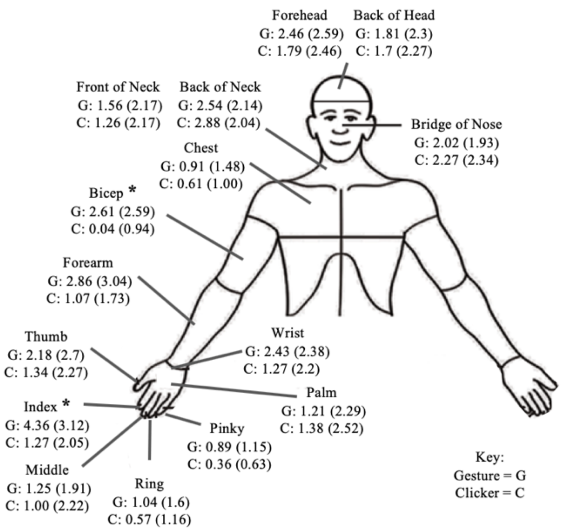

Figure 5. Average Borg CR10 Ratings * p < .03

Perceived Eye Strain

There were no significant differences of participants' levels of eye strain, the ease of reading text, text clarity, ability to concentrate, physical fatigue, or mental fatigue between the gesture or clicker input method $p > .05$.

Perceived Usability

There was a significant difference between the SUS scores of the gesture method (M=45.29, SD = 13.92) and clicker method (M=65, SD=19.02), $t(16)= -4.454$, $p=0.0004$, $d = 1.183$. The gesture method's SUS score was considered "poor" and the clicker method's score was considered "ok" (Bangor, Kortum, & Miller, 2009).

Preference and Recommendations

Both input methods were reported as intuitive and easy to learn, albeit slow and cumbersome. The majority of the participants (15 out of 17) preferred the clicker method over the gesture input because it was perceived as faster, more accurate, and less fatiguing. Participants reported that the current gesture should be changed entirely to ease the exertion they felt in their typing arm after extended use. Some suggested that smaller movements, such as slightly bending the tip of the finger, while others stated that the gesture needs to be redesigned completely to eliminate the need for holding up their arm.

Participants stated that the clicker method could be improved as well. Typically, the user hears a clicking noise through the HoloLens when the clicker is pressed, even if no letter is inputted. Participants reported that this was confusing. They were also confused when they first used the clicker because the "button" to click is represented as a depression on the top of the clicker but the clicking motion was felt underneath the device. This was unexpected at first, but many participants were able to adapt to this feature. Other comments included: the gesture field of view was very small, and head tracking was frustrating, fatiguing, and too sensitive.

DISCUSSION

Overall, participants rated the clicker input method better on almost all of the measures. Participants typed significantly faster with the clicker even though there was no difference in error rates between the clicker and gesture methods. The clicker also had a higher perceived usability score but it was still mediocre at best. These scores showed that improvements could be made to both input methods.

Most participants stated that, overall, they preferred the clicker method over the gesture method even though it was not more accurate. They suggested that the gesture interaction should be redesigned to alleviate the need to hold up their arm while typing. An alternative improvement could be to increase the FOV of the HoloLens. Currently, because of the small FOV, participants had to hold their arm up at eye level while gesturing. If this interaction area was larger, users could relax their arm more while typing. The recent announcement of the HoloLens 2 promises a FOV twice the size (Wong, 2019).

Participants also suggested improvements for selecting the letters on the keyboard. Currently, head tracking is used to select letters. Participants stated that this was tiresome and too sensitive, causing them to become frustrated. Suggestions included: creating a sensitivity setting similar to those available for a computer mouse or gaming joystick, changing the sensitivity of the headset, or eliminating the head movement entirely by introducing eye tracking or a trackpad for the clicker. Future research should be conducted on performance and usability of text entry after the suggested improvements have been implemented. Voice as an input method should also be investigated in future research, as it has been shown to be preferred and more usable than the current gesture and clicker methods.

REFERENCES

Ahn, S., Heo, S., & Lee, G. (2017, October). *Typing on a Smartwatch for Smart Glasses*. In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, Brighton, United Kingdom.

Borg, G. (1998). *Borg's Perceived Exertion And Pain Scales*. Champaign, IL: Human Kinetics.

Bangor, A., Kortum, P., T., & Miller., J., T. (2009). Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114-123.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

Grubert, J., Witzani, L., Ofek, E., Pahud, M., Kranz, M., & Kristensson, P. O. (2018, March). *Text Entry in Immersive Head-Mounted Display-based Virtual Reality using Standard Keyboards.* Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces, Reutlingen, Germany.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.

Knierim, P., Schwind, V., Feit, A. M., Nieuwenhuizen, F., & Henze, N., (2018, April). *Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands.* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montréal, QC, Canada.

Lee, L. H., Lam, K. Y., Yau, Y. P., Braud, T., & Hui, P. (2019, March). *HIBEY: Hide the keyboard in augmented reality.* Paper presented at the 17th annual meeting of the IEEE International Conference on Pervasive Computing and Communications, Kyoto, Japan.

MacKenzie, I., S., & R., W., Soukoreff. (2003, April). Phrase Sets for Evaluating Text Entry Techniques. In *CHI'03 extended abstracts on Human factors in computing systems,* 754-755.

Perrinet, J., Pañeda, X. G., Cabrero, S., Melendi, D., García, R., & García. (2011). Evaluation of Virtual Keyboards for Interactive Digital Television Applications. *International Journal of Human-Computer Interaction 27*(8), 703-728.

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. (2016). Speech is 3x faster than typing for English and Mandarin text entry on mobile devices. *arXiv preprint arXiv:1608.07323*.

Sears, A. (1991). Improving Touchscreen Keyboards: Design issues and a comparison with other devices. *Interacting with Computers, 3*(3).

Sears, A., Revis, D., Sawtski, J., Crittenden, R., & Shneiderman, B. (1993). Investigation touchscreen typing: The effect of keyboard size on typing speed. *Behaviour & Information Technology, 12,* 17 - 22.

Shelstad, W. J., Benedict, J. D., Smith, J. K., Momo, T., Guliuzo, J. D., Reuss, N. R., Courson, A., & Chaparro, B. S. (February, 2019). *User Experience Evaluation of the Microsoft HoloLens for Student-Related Tasks.* Poster presented at the Human Factors and Applied Psychology Student Conference, Orlando, FL.

Turner, C. J., Chaparro, B. S., & He, J. (2016). Text input on a smartwatch QWERTY keyboard: tap vs. trace. *International Journal of Human-Computer Interaction, 33*(2), 143-150.

Wilson, A. D., & Agrawala, M. (April, 2006). *Text Entry Using a Dual Joystick Game Controller.* Proceedings of the 2006 CHI conference, Montréal, Québec, Canada.

Wong, R. (2019, February, 24). Microsoft Hololens 2 ushers in the next generation of augmented reality. *Mashable.* Retrieved from https://mashable.com/article/microsoft-hololens-2-mwc-2019/#SOVugEG7K8q6

Wright, N. (2018, November 27). How Microsoft HoloLens could change the way we work [Blog post]. Retrieved from https://www.nigelfrank.com/blog/how-microsoft-hololens-could-change-the-way-we-work/

Zhai, S., & Kristensson, P. O. (2012). The word-gesture keyboard: reimagining keyboard interaction. *Communications of the ACM, 55*(9), 91-101.