Typing in Mid Air: Assessing One- and Two-Handed Text Input Methods of the Microsoft HoloLens 2

Emily Rickel[(✉)], Kelly Harris, Erika Mandile, Anthony Pagliari, Jessyca L. Derby, and Barbara S. Chaparro

Department of Human Factors and Behavioral Neurobiology, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA
`{rickele,harrik28,mandilel,pagliara,`
`derbyjl}@my.erau.edu, chaparbl@erau.edu`

Abstract. The Microsoft HoloLens 2 is a mixed reality (MR) headset that overlays virtual elements atop a user's view of their physical environment. To input text, the device has the ability to track hands and fingers, allowing for direct interaction with a virtual keyboard. This is an improvement over the HoloLens 1 device, which required head tracking and single-finger air-tapping input. The present study evaluated the performance (speed and accuracy), perceived usability, mental workload, and physical exertion of one-handed and two-handed text entry. A sample of 21 participants (12 male, 9 female) aged 18–32 years typed standardized phrases presented in random order. Typing with two hands was faster and more preferred than one-handed input; however, this input method was also less accurate. Exertion in some body parts was also higher in the two-handed condition. Findings suggest that while two-handed text input was better than one-handed, there is room for improvement to approximate typing on a physical or mobile device keyboard.

Keywords: Human factors · Mixed reality · Mixed reality headset · Microsoft HoloLens 2 · Text input

1 Introduction

Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) are becoming increasingly popular for use in education and other domains [1]. This research focused on the Microsoft HoloLens 2, an MR headset that anchors virtual elements over a user's view of their physical environment, allowing users to complete a variety of tasks while maintaining an awareness of their surroundings. VR gives the user the feeling of "being there" and simulates a realistic environment [2], and AR allows users to see their physical surroundings but overlays virtual elements that move as the user moves, such as in phone applications [3]. MR differs in that virtual elements are anchored in the user's environment, and remain stationary if the user moves, as an object would if it was physically present [3]. As the user moves their head, the virtual elements appear larger or smaller depending on the user's orientation.

The Microsoft HoloLens 2 is diverse in its uses and can launch over 300 applications, including games, business, and productivity applications [4]. Consequently, text entry is a key component for many HoloLens 2 tasks. Prior research on the first version of the device, the Microsoft HoloLens 1, showed text input controlled by head movements and air-tap hand gestures was slow, fatiguing, and frustrating for users [5]. The HoloLens 2 improved upon these concerns with the implementation of functionality to track users' hands and individual fingers, enabling direct interaction with a virtual keyboard. Users can press individual keys using their fingers on one hand or both hands at the same time and receive an audio cue when touching the virtual keyboard that closely mimics that of a physical keyboard.

Although the Microsoft HoloLens 2 provides changes from the first model, there are still challenges facing text input on head-mounted devices. Mid-air keyboards that have the ability to track ten fingers while typing are the "holy grail" of virtual text entry [6]. A study done by Sears [7] showed that participants typing on a traditional QWERTY keyboard averaged 58 words per minute (WPM); almost 10 times faster than the averages calculated in the Microsoft HoloLens 1 study, where speeds averaged from 5.41 WPM for gesture typing and 6.58 WPM for clicker typing [5].

This study aimed to evaluate performance (speed and accuracy) and preference of one-handed and two-handed text entry on the Microsoft HoloLens 2, along with perceived workload, usability, exertion, and eye fatigue.

2   Method

2.1   Participants

The study sample consisted of 21 participants (12 male, 9 female) recruited from a university located in the southeastern United States. Participant ages ranged from 18 to 32 years ($M = 21.62$, $SD = 3.87$). Sixteen participants reported prior use with VR or AR headsets, and four participants reported owning a VR or AR headset. Number of hours for prior VR or AR use ranged from 0 to 50 h ($Mdn = 2$, $IQR = 9.25$). Four participants reported being non-native English speakers. Two participants reported being left-handed. Participants were screened for disabilities or movement problems associated with their hands.

2.2   Experimental Design

A repeated-measures experimental design was utilized for this study. Qualitative and quantitative data was collected through this study. Participants were asked to input text using two conditions: 1) one-handed (i.e., using their dominant hand), and 2) two-handed. The conditions were presented in a counterbalanced order. The independent variable was the input method being used (one-handed or two-handed) and the dependent variables included: typing speed, typing accuracy, perceived workload, perceived exertion, perceived eye strain, perceived usability, and preference.

2.3   Measures

Text Input Speed and Accuracy. Words per minute (WPM), adjusted words per minute (AdjWPM), and word error rate (WER) were calculated to evaluate the impact that each text input method had on performance. Measures for text input accuracy (WER) were examined by type of error made: substitution, insertion, and omission. Substitution errors occurred participants completely replaced a word with another word. Insertion errors were marked when participants typed an additional word not already a part of the phrase. Omission errors occurred when participants excluded a word from the given phrase.

Perceived Workload. The NASA Task Load Index (NASA-TLX-R) is a 6-item questionnaire that determines participants' subjective workload and perceived performance [8]. Each statement in the questionnaire represents one of the six dimensions: physical demand, mental demand, temporal demand, performance, effort, and frustration. Participants rated each statement on a 21-point scale. A higher rating signifies that the participant perceived the task as being more demanding or that they performed poorly.

Perceived Usability. The System Usability Scale (SUS) was used to gain insight into the participants' perceived usability of each text input method [9]. The SUS is a standardized 10-item questionnaire. Participants rated each question on a scale of 1 (strongly disagree) to 5 (strongly agree). A final score between 0–100 was calculated and placed on an adjective rating scale that ranges from worst imaginable (a score of 0 to 25) to best imaginable (a score of 100) [9, 10]. The questions on the SUS were modified to fit the subject of the study, therefore "system" was changed to "input method".

Perceived Exertion. The Borg Category Ratio Scale (Borg CR10) was used to evaluate the participants' perceived exertion with each text input method [11]. Participants were presented with an upper-body map consisting of 33 areas that were rated based on an exertion scale starting from "nothing at all" (0) to "extremely strong" (10) or even "absolute maximum" which can be rated as a 12, 13, or higher. If participants rated a specific area of their body above "moderate" (3), they were asked to explain their rating.

Perceived Eye Fatigue. A 6-item questionnaire was used to assess the participants' ability to concentrate, their ease of reading text, text clarity, physical fatigue, mental fatigue, and level of eye strain. The questions were rated on a 5-point likert-scale. Higher scores indicated an easier ability to read text, higher satisfaction with text clarity, higher ability to concentrate, and lower levels of fatigue and eye strain.

Preference. Participants were asked at the end of the study to indicate their preferences on the text input methods. Participants were asked to rate the text input methods on a preference scale (from 0 - Least Preferred to 50 - Most Preferred) independently of one another and to explain their rating. Participants were also asked to indicate with which input method they believed they typed the fastest and most accurately.

2.4   Materials

Microsoft HoloLens 2. The Microsoft HoloLens 2 is a wireless MR headset that was first released in 2019. The headset uses spatial mapping technology to create a three-dimensional model of the user's physical environment and to display digital content that users can manipulate through hand tracking, eye tracking, and voice commands [12]. The software version used for this study was Windows Holographic for Business, operating system (OS) build 10.0.19041.1154.

Phrases. Participants were presented with pre-selected phrases of text that originated from a subset of MacKenzie & Soukoreff's [13] standardized set of 500 phrases using Qualtrics, an online survey platform (see Fig. 1). These phrases were designed to evaluate text entry techniques and are characterized as being moderate in length, easy to remember, and representative of the English language.

Fig. 1. Participants' view of the text input task, which consisted of using a virtual keyboard to type pre-selected phrases of text using Qualtrics, an online survey platform.

2.5   Procedure

Participants were recruited from a university located in the southeastern United States. After participants provided their consent to participate in the study, they completed a demographic questionnaire. Participants were then fitted with the Microsoft HoloLens 2 headset, given a brief tutorial on how to use the device, and prompted to complete the device eye calibration procedure.

All participants completed both the one-handed and two-handed study conditions in counterbalanced order. Participants typed a total of 40 unique phrases from MacKenzie & Soukoreff's [13] phrase set that were presented in random order. For each condition, participants typed 5 practice phrases and 15 experimental phrases that were later evaluated for text input speed and accuracy. Timing for text input speed was measured by tracking when participants said "start" after they adjusted the virtual keyboard to a comfortable size and position and as they began typing each phrase, and "finish" as they completed inputting each phrase. Participant performance was monitored by a researcher who observed a television screen that showed the participant's view of the task using a screen mirroring device (see Fig. 2). As they typed each phrase, participants were instructed to input text as quickly and as accurately as possible, without using predictive text or abbreviated language (e.g., typing "u" instead of "you"). Participants also were directed not to worry about capitalization or punctuation and were given the option to make corrections, but they were not required to do so. After each condition, participants were asked to complete a series of questionnaires to capture perceived workload (NASA-TLX-R), usability (SUS), exertion (Borg CR10), and eye fatigue. After both conditions were completed, participants rated their preference for one- and two-handed input methods independently of one another and provided suggestions for improving text input using the Microsoft HoloLens 2. The study took approximately 60–90 min per participant.

Fig. 2. Layout of experimental set-up, showing the participant completing the text entry task while wearing the HoloLens 2 (left) as the researcher (right) observes the participant's performance using the television screen and screen-mirroring device (top).

3   Results

Paired samples t-tests were conducted to compare text input speed and accuracy, as well as perceived workload, usability, exertion, eye fatigue, and preference between one- and two-handed text input.

3.1   Text Input Speed and Accuracy

There was a statistically significant difference in typing speed between the one-handed (M = 12.07, SD = 1.78) and two-handed (M = 13.91, SD = 2.62) conditions, $t(20)$ = −3.43, P = 0.003 (two-tailed), D = −0.75. Participants typed faster with the Microsoft HoloLens 2 keyboard when using two hands compared to using one hand (see Fig. 3).
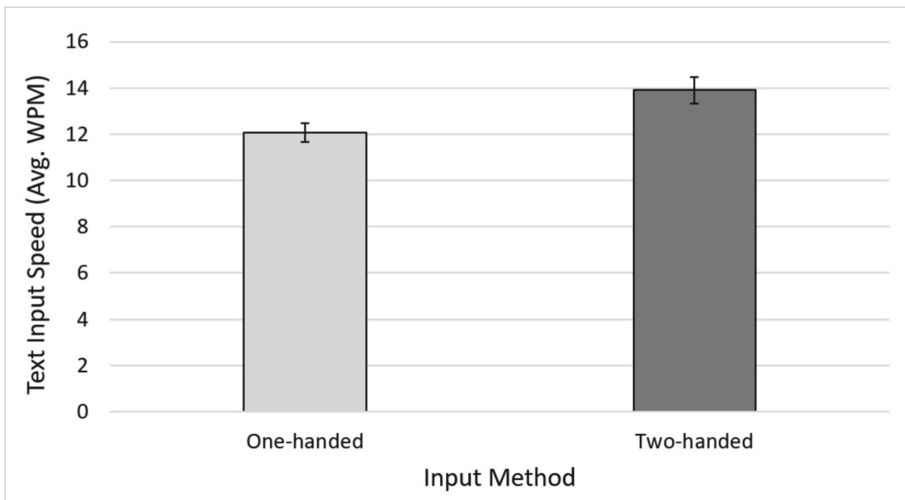

Fig. 3. Comparison of text input speed (average words per minute) between one-handed and two-handed text input. Error bars represent ±1 standard error.

Word error rate (WER), substitution error rate (SER), insertion error rate (IER), and omission error rate (OER) were calculated to assess text input accuracy (see Fig. 4).

Word Error Rate (WER). There was a statistically significant difference in WER between the one-handed (M = 0.05, SD = 0.06) and two-handed (M = 0.08, SD = 0.09) conditions, $t(20)$ = −2.70, p = 0.014 (two-tailed), d = −0.59. Participants made more word errors while typing with two hands compared to one hand.

Substitution Error Rate (SER). There was a statistically significant difference in SER between the one-handed (M = 0.04, SD = 0.05) and two-handed (M = 0.06, SD = 0.08) conditions, $t(20)$ = −2.40, p = 0.026 (two-tailed), d = −0.52. Participants made more substitution errors while typing with two hands compared to one hand.

Insertion Error Rate (IER). There was no statistically significant difference in IER between the one-handed and two-handed conditions.

Omission Error Rate (OER). There was a statistically significant difference in OER between the one-handed ($M = 0.004$, $SD = 0.007$) and two-handed ($M = 0.009$, $SD = 0.01$) conditions, $t(20) = -2.35$, $p = 0.029$ (two-tailed), $d = -0.51$. Participants made more omission errors while typing with two hands compared to one hand.
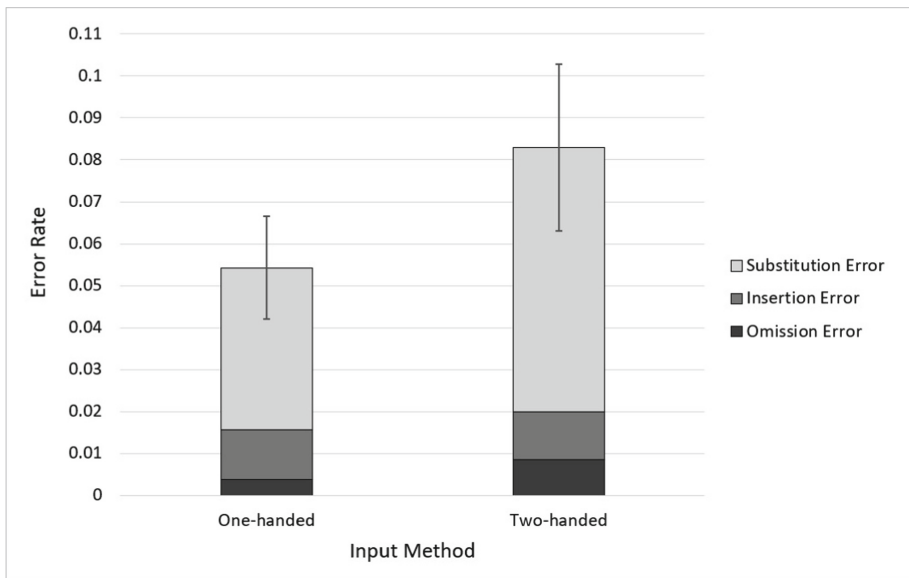

Fig. 4. Comparison of text input accuracy (word error rate). Error bars represent ± 1 standard error.

To summarize, participants typed significantly faster using two hands compared to one hand; however, participants made significantly more errors using two hands.

3.2   Perceived Workload

There was a statistically significant difference in mental demand scores between the one-handed ($M = 6.52$, $SD = 4.46$) and two-handed ($M = 8.52$, $SD = 4.90$) conditions, $t(20) = -2.49$, $p = 0.022$ (two-tailed), $d = -0.54$. Participants reported higher mental demand when typing with two hands compared to one hand.

There was a statistically significant difference in physical demand scores between the one-handed ($M = 10.95$, $SD = 5.04$) and two-handed ($M = 8.95$, $SD = 5.04$) conditions, $t(20) = 2.28$, $p = 0.033$ (two-tailed), $d = -0.50$. Participants reported lower physical demand when typing with two hands compared to one hand.

There was no statistically significant difference between one-handed and two-handed conditions for temporal demand, performance, effort, and frustration subscales (see Fig. 5).
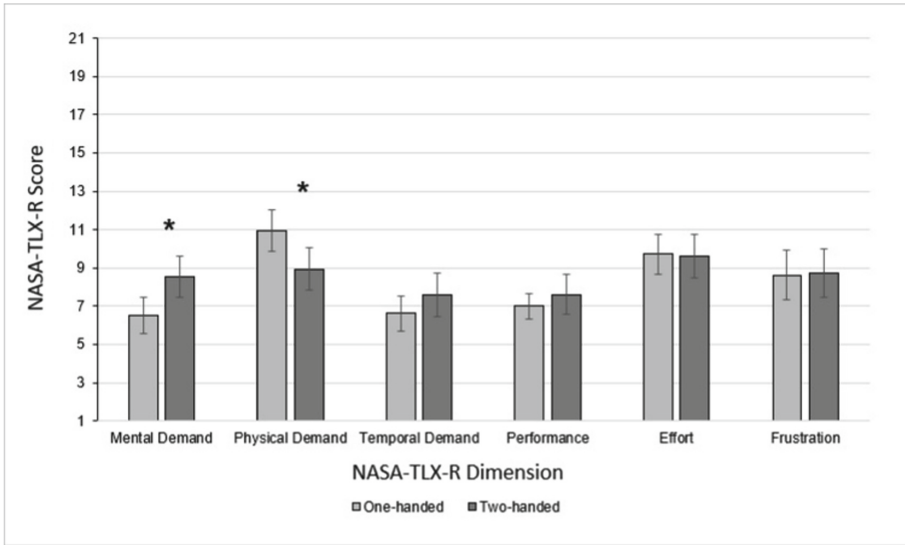

Fig. 5. Comparison of perceived workload (NASA-TLX-R) between one-handed and two-handed text input. Error bars represent ± 1 standard error. * indicates $p < 0.05$.

3.3   Perceived Usability

There was no statistically significant difference in perceived usability between the one-handed ($M = 62.02$, $SD = 19.02$) and two-handed conditions ($M = 65.13$, $SD = 17.39$). Participants perceived the usability of one- and two-handed typing to be similar and falling within the adjective rating scale of "ok" (see Fig. 6).
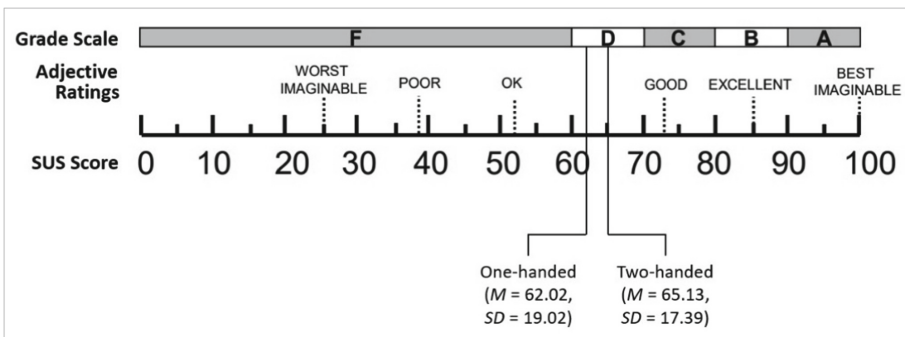

Fig. 6. Comparison of perceived usability (System Usability Scale). Figure adapted from [10].

3.4 Perceived Exertion

Participants reported significantly greater exertion in their left hand and left index finger when typing with two hands than with one hand (palm of left hand: $t(19) = -2.33$, $p = 0.031$ (two-tailed), $d = -0.52$; back of left hand: $t(20) = -2.13$, $p = 0.046$ (two-tailed), $d = -0.47$; left index finger: $t(20) = -2.69$, $p = 0.014$ (two-tailed), $d = -0.59$). In general, perceived exertion was minimal across all body parts.

3.5 Perceived Eye Fatigue

There was no statistically significant difference in reported eye strain ratings between one-handed and two-handed typing conditions for ease of reading text, text clarity, ability to concentrate, physical fatigue, mental fatigue, or level of eyestrain.

3.6 Preference

Participants reported their preference for each text input method (one-handed, two-handed) on a scale from 0 – Least Preferred to 50 – Most Preferred. There was a statistically significant difference in preference ratings between the one-handed ($M = 19.90$, $SD = 10.97$) and two-handed ($M = 32.81$, $SD = 11.94$) conditions, $t(20) = -3.54$, $p = 0.002$ (two-tailed), $d = -0.77$. Participants preferred inputting text with two hands compared to one hand (see Fig. 7).
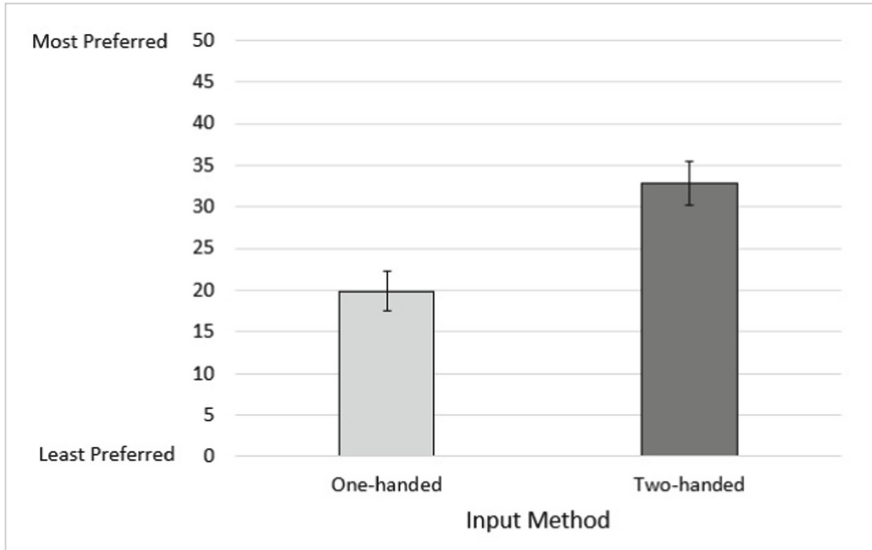

Fig. 7. Comparison of reported preference between one-handed and two-handed text input. Error bars represent ±1 standard error.

Several participants liked the auditory feedback provided by the system when they clicked a key, as some relied on this feature to determine whether the system recognized each keystroke. Several participants, however, reported they would like to use more than their index finger on each hand and recommended that the system accommodate the ability to type with all of their fingers. They also did not like how the keyboard reset its size and position between each phrase, and recommended that there be an option to save the size and position when users close the keyboard and then open it again within a short period of time. In addition, participants reported difficulty in typing double letters (e.g., "ee") and suggested this could be improved by reducing the system's lag after clicking a key.

4   Discussion

Overall, participants typed faster using the two-handed input method and preferred this method to one-handed input. Interestingly, the two-handed input method was more prone to error. Participants often stated that they sometimes accidentally clicked wrong keys because they were typing faster with two hands and perhaps being less precise. Additionally, participants occasionally attempted to use more than one finger on a single hand to type, which resulted in accidental touches of keys.

Many participants commented that the keyboard should allow detection of all fingers, making it more realistic to typing on a physical keyboard. Participants commented that having the keyboard fixed at a downward angle, similar to a physical keyboard, could prevent users from reaching out to type, reducing the amount of exertion for upper extremities.

The two-handed method had a higher perceived usability score, however it is only considered to be "ok" [10], meaning there is room for improvement. Overall, participants' preference for the two-handed method may suggest that they may be willing to sacrifice accuracy for speed and less exertion both mentally and physically. Participants reported that typing using two hands required a higher mental demand ($M = 8.52$, $SD = 4.89$) than typing with just one hand ($M = 6.52$, $SD = 4.46$). Many participants indicated that they had less accuracy typing two-handed. One participant thought it was easier to concentrate on the text input task when typing one-handed, whereas typing two-handed was distracting since they would often get confused about which hand was typing which letter. Another participant suggested that typing two-handed is something they had to get used to because they had to be aware of their hand placement. This higher degree of concentration may have contributed to the higher level of mental demand in two-handed typing.

Results in this study demonstrate speeds twice that reported by Derby et al. [5] for HoloLens 1 text input ($M = 5.41$ WPM, $SD = 0.89$, for gesture method; $M = 6.58$ WPM, $SD = 0.75$, for clicker method). While this comparison shows marked improvement in typing speeds for the HoloLens 2, typing speeds are not yet comparable to that of a physical keyboard (see Table 1).

Table 1. Comparison of different typing speeds across device types.

Device	Method	Average WPM
Microsoft HoloLens 1 [5]	Gesture	5.41
	Clicker	6.58
Microsoft HoloLens 2	One-handed	12.07
	Two-handed	13.91
Touchscreen Keyboard [7]	Two-handed	25
Physical Keyboard [7]	Two-handed	58

4.1 Limitations and Future Research

There were some limitations involved in this study. Our sample only included college-aged students. This limited generalizability and therefore it may have been beneficial to have a more diverse and robust sample of participants. Another limitation to this study was the system would sometimes automatically correct text input mistakes, which potentially could have affected error rates if incorrect words were substituted, submitted, and evaluated. Additionally, participants were instructed to not use speech-to-type or predictive text while typing. By restricting participants in this way, the process of typing may not have been representative of how an everyday user may type with the device.

In the future, research examining text input performance of the HoloLens 2 with other populations should be conducted. Other HoloLens 2 text input methods (e.g., voice-to-text, swiping gesture) should also be investigated, as well as scenarios that are more representative how an everyday user may input text with the headset. Additionally, text input performance should be continuously evaluated as new iterations of the HoloLens are developed. Improvements to future versions of the HoloLens are expected to focus on three key areas: improvement in immersion, improvement in comfort and social acceptability, and increasing the value of what can be accomplished using the headset [14]. These modifications could change the efficiency and effectiveness of typing using the MR headset, as well as increase consumer acceptance for a variety of applications and use cases.

References

1. Milman, N.B.: Defining and conceptualizing mixed reality, augmented reality, and virtual reality. Distance Learning **15**(2), 55–58 (2018)
2. Zheng, J.M., Chan, K.W., Gibson, I.: Virtual reality. IEEE Potentials **17**(2), 20–23 (1998). https://doi.org/10.1109/45.666641
3. Brigham, T.J.: Reality check: basics of augmented, virtual, and mixed reality. Med. Ref. Ser. Q. **36**(2), 171–178 (2017)
4. Browse All HoloLens Apps: https://www.microsoft.com/en-us/store/collections/hlgettingstarted/hololens. last accessed 25 Jan 2022
5. Derby, J.L., Rarick, C.T., Chaparro, B.S.: Text input performance with a mixed reality head-mounted display (HDM). Hum. Fac. Ergono. Soc. Ann. Meet. **63**, 1476–1480 (2019)

6. Dudley, J.J., Vertanen, K., Kristensson, P.O.: Fast and precise touch-based text entry for head- mounted augmented reality with variable occlusion. ACM Trans. Comp.-Hum. Interac. **25**(6), 1–40 (2018)

7. Sears, A.: Improving touchscreen keyboards: design issues and a comparison with other devices. Interac. Comp. **3**(3), 253–269 (1991)

8. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. Human Mental Workload, 139–183 (1988)

9. Brooke, J.: SUS - a quick and dirty usability scale. Usabi. Eval. Indus. 189–194 (1996)

10. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Stud. **4**(3), 114–123 (2009)

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

12. Microsoft: https://www.microsoft.com/en-us/hololens/hardware . last accessed 23 Jan 2022

13. MacKenzie, I.S., Soukoreff, R.W.: Phrase sets for evaluating text entry techniques. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems, pp. 754–755. Association for Computing Machinery, New York, NY, USA (2003)

14. Microsoft is working on Hololens 3: Consumer version. https://mspoweruser.com/microsoft-is-working-on-hololens-3-consumer-version/. last accessed 25 Jan 2022