The Intuitiveness of Gesture Control with a Mixed Reality Device

Jacob D. Benedict, Jacob D. Guliuzo, & Barbara S. Chaparro
Department of Human Factors and Behavioral Neurobiology
Embry-Riddle Aeronautical University, Daytona Beach, FL

Mixed reality is a new technology that requires users to control a head-mounted device via gestures with their hands. Users of these devices must learn and remember a new way of interacting. It has been shown that creating gestures that resemble movements used to operate touch screens can help with this new transfer. This study investigates how well people learn to use the out-of-the-box gestures for a mixed reality headset, Microsoft HoloLens, after interacting with it for a very short period of time. Performance with the gestures was measured with novices before and after approximately five minutes of practice game play. Participants showed a significant improvement on the gestures to open and position windows and reported them to be easier to do after the short practice. This information could help to create apps or tutorials that help teach these gestures, as well as identifying which gestures are more intuitive to users.

Introduction

Mixed reality is the newest form of virtual environment technology that is emerging in the academic and industry world. Mixed reality (MR) is a middle ground between virtual reality (VR) and augmented reality (AR). VR is typically a head mounted display that fully fills someone's vision. Most VR headsets today, such as the HTC Vive, track the user in space using multiple sensor boxes as well as tracking the controllers used to interact with the headsets. VR creates the most realistic feeling of immersion in the virtual world, however you lose the ability to see or interact with anything other than the device. AR on the other hand simply overlays virtual information in the real or physical world. Google Glass was one of the first widespread forms of AR, this device was similar to a pair of glasses. However, a small transparent screen was placed in the upper right corner of the right eye that would show information. MR merges VR and AR by tracking the user in their environment but projecting the virtual information on the physical world. This allows users to open a virtual projection and place it in one area of their environment. If the user looks away, the window does not follow their field of view, it stays at the specific location and reappears when the user looks back. Commanding the window to follow them can be done as well. One of the most popular mixed reality headsets on the market currently is the Microsoft HoloLens. This headset is being used in a wide variety of domains such as product production lines and the military.

Since MR headsets are becoming more prevalent, it is important to understand how people interact with them. Besides the HoloLens other MR devices that have been released are: Magic Leap One and Acer Mixed Reality. Given the recency of the HoloLens device and the fact that it is still a developer's tool, rather than a consumer product, few published studies examining its usage from a usability perspective exist. Whitlock, Harnner, Brubaker, Kane, & Szafir, (2018) conducted a study to examine differences between gestures, voice, or controller-based operation of a Microsoft HoloLens to complete two tasks, changing a virtual thermostat, and adjusting a virtual security camera. Participants reported the gestures to be the most preferred input method. Gestures, as well as the handheld controllers, were reported as being faster and more accurate, whereas voice controls were reported to be better for tasks that did not require as much accuracy or speed. In another study, gestures were compared against a gaze and button press scheme. An eye tracker in combination with a two-button keyboard was used to create the gaze and button scheme (Canare, Chaparro, & Chaparro, 2018). After completing a sorting task, the gaze and button press scheme was found to be the fastest and resulted in the fewest errors, showing performance similar to a mouse. However, the authors note that this technique is limited for regular use of a computer and that gestures may be more intuitive (Canare, Chaparro, & Chaparro, 2018).

In situations where cognitive load or visual attention might be limited, gestures could be beneficial. Graichen, Graichen, & Krems (2019) conducted a study using a gesture-based interface in a car and compared it to a touchscreen system currently in vehicles. After completing a set of tasks varying in complexity while driving in a simulator, the participants rated the gestures as desired where the touch was neutral. Compared to the touch screen system, the gestures resulted in fewer glances, as well as shorter glances when they did happen, towards the main control panel. As a result, participants reported the gesture-based system was better (Graichen, Graichen, & Krems, 2019). Similarly, Marz, Schwahlen, Geisler, & Kopinski (2016) concluded that gestures are easier to learn and use if they are similar to their touchscreen equivalent. A gesture that closely mimics swiping through pages on a touchscreen is easier for people to use and remember compared to novel gestures that are created to work with a certain interface. On top of being similar to the touch screen version of the gesture, certain gestures are more preferred than others. Gestures that use the full hand or one finger and move either left or right or up and down are more preferred over either static gestures or twisting of the hands (Fariman, Alyamani, Kavakli, & Hamey, 2016)

An exploratory study by the authors with the HoloLens examined user performance with gestures in a simulated office environment (Shelstad et al., 2019). This environment was chosen not only to help simulate what students would be doing naturally, but also to assess the use of the HoloLens as a device to be used for office work in the future. Students were asked to create a simulated work area at a desk with the HoloLens as their primary work device. Participants opened virtual windows for the Microsoft Office Suite products and completed several related tasks such sending emails and creating a PowerPoint. The results from this study showed that participants had some issues with the device gestures and window manipulation. The Net Promoter Score, which shows how likely a person would recommend the product revealed 3 detractors, 3 passive, and 2 promotors. When rating the difficulty of the gestures used to interact with the device, clicking, scrolling, and resizing a window were rated as the most difficult gestures to complete

Current Study

The goal of the current study was to examine user performance with various gestures required to open, move, and manipulate windows using the Microsoft HoloLens. Performance was assessed before and after a short practice period. We hypothesized that the gestures, in general, would become quicker and rated as easier after practice, but we were unsure how ubiquitous this effect would be across gesture types.

Method

Participants

For this experiment 15 students were recruited from a private university; all participants were undergraduate students, ages 18 to 23. All participants had little to no use of a VR or MR headset prior to this study. Participants voluntarily partook in the study and their participation was in accordance with all regulations from the university's Institutional Review Board. Participants completed the study in approximately 30 minutes.

Measures and Equipment

The Microsoft HoloLens first generation with version 10.0.17134.80 (April 2018 Update) was used.

Gestures

Three main gestures used to control windows in the HoloLens environment were examined (Table 1)

Gesture name	Use
Select	Choosing windows, choosing options, selecting something
Drag	Moving windows around. scrolling, resizing
Bloom	Opening the main menu, exiting applications

Table 1. The names of the gestures and the actions they are used to complete.

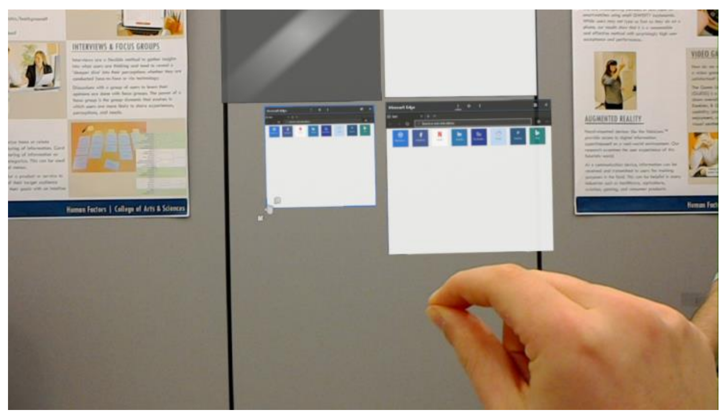

Figure 2. The drag gesture being used to resize a window.

Tutorial

A third-party tutorial was utilized to provide the participants a basic understanding of how to control the HoloLens. This tutorial was chosen over the original Microsoft HoloLens tutorial because it provided a more comprehensive overview of all gestures explored in this study.

Window Manipulations

Participants completed tasks to move and resize multiple windows on the HoloLens using gesture controls. This included

- opening and aligning four windows
- moving the windows into a 2X2 square
- resizing the windows from the corners
- expanding the windows from the edges

These manipulations were evaluated in a pre and post-test method.

Measures

Time on task: The amount of time it took for participants to complete a task. The time was started when the participant started the task and ended when the participant stated they completed it.

Difficulty rating: A subjective rating from 1 (very difficult) to 10 (very easy).

Perceived Usability: The System Usability Scale (SUS) was used to assess the perceived usability of the entire exercise. This scale is a 10-question validated measure that is rated on a five-point scale from strongly disagree to strongly agree. The final calculated score was between 0-100 with 100 being the best available score (Brooke, 1996).

User Tasks

1. Open four windows and place them in a horizontal line.
2. Move the windows into a 2 X 2 square.
3. Resize a window using the corner
4. Resize a window using the bottom
5. Resize a window using the left or right edge

Gameplay

In order to practice the gestures, two games from the HoloLens App Store were used. The first was Tic-Tac-Toe, which used the gesture of selecting as its core method of interaction. The second was a pool game, which used the scroll gesture for the majority of game play. Both of these games were selected because they integrated gestures that are also used for standard window management.

Figure 1. The game of pool used to practice the drag gesture.

Procedure

Participants first read and signed a consent form. Next, they were given a general overview of the HoloLens device as well as an idea of the kind of tasks they would be completing. Once the participant donned the HoloLens and adjusted properly, they completed the tutorial about the gestures used to open and manipulate windows. After the tutorial was completed the participant was given a pre-test. This test included tasks involving opening, positioning, and resizing windows from the sides and corner. In order to complete these tasks, a combination of different gestures were required. For example, to open a browser window, the participant needed to use the bloom gesture and the select gesture. Participants were asked to rate the difficulty of each gesture. Following the completion of the pre-test, participants played both the tic-tac-toe and pool games. Both of these games were used to help the participants practice certain gestures without specifically training the exact movements. Once the gameplay was complete, participants completed the post-test which was the same tasks used in the pre-test. Lastly, participants took a survey asking general demographic questions relating to previous AR and VR use as well as the SUS.

Results

Paired samples t-tests were conducted to evaluate the differences between the pre and post tests for time on task as well as difficulty ratings.

Time on Task

Task one involved opening four windows and placing them in a horizontal line. There was a significant difference between the pre and post tests on time on task

(Pre *M* = 77.09, *SD* = 30.285, Post *M*= 48.39, *SD* = 24.13), *t*(1,14) = 3.363, *p* = .005, *d* = *1.05*.

The next test compared the time it took to place the four windows into a 2X2 square. There was a significant difference between the pre and post tests on time on task (Pre *M* = 62.41, *SD* = 34.74, Post *M* = 44.7, *SD* = 19.19), *t*(1,14) = 2.592, *p* = .021, *d* =*.66*.

Time to use the gestures to expand the windows from the corner was evaluated for pre to post-test. A significant difference was found when comparing the pre and post times, (Pre *M* = 22.16, *SD* = 14.85, Post *M* = 7.85, *SD* = 4.25), *t*(1,14) = 3.491, *p* = .004, *d* =1.50.

No significant differences were found on the pre and post time on task comparison of expanding the windows from either the sides or the bottom.

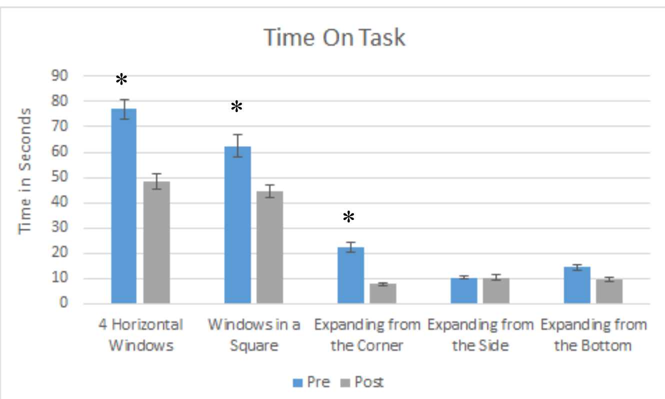

Figure 1. The pre and post time on task difference for each task completed.
⁎ $p < .05$

Perceived Difficulty Compared Pre to Post

Pre and post perceived difficulty of each gesture used to manipulate the windows was recorder. This was measured by using a difficulty rating (1-10 with 1 being very difficult) for each task. The first task involved participants opening four windows and placing them in a line. A rating was given for each gesture used to complete this task. The rating for opening the windows (bloom gesture) was significant between the pre and post group (Pre *M* = 8.5, *SD* = 1.41, Post *M* = 9.2, *SD* = 1.45), *t*(1,14) = -2.320, *p* = .036, *d* = -.52. Placing the windows (Pre *M* = 7.2, *SD* = 1.78, Post *M* = 8.46, *SD* = 1.4), *t*(1,14) = -3.676, *p* = .002, *d* = *-.78*, and moving the windows (Pre *M* = 7.07, *SD* = 1.94, Post *M* =8.71, *SD* = 1.27), *t*(1,14) = -3.735, *p* = .002, *d* = *-1.02* were also found to be easier after practice.

Perceived difficulty on the gestures used to place the four windows into a 2X2 square was collected. Participants reported placing the windows (Pre *M* = 6.07, *SD* = 2.37, Post *M* = 7.4, *SD* = 1.29), *t*(1,14) = -2.697, *p* = .017, *d* =-.73), as well as moving the windows (Pre *M* = 6.2, *SD* = 2.04, Post *M* = 7.67, *SD* =1.67), t(1,14) = -3.290, *p* = .005, *d* = -.79 to be easier after practice.

No significant differences were found on the comparison of pre and post perceived difficulty for expanding the windows from the corner, side, or bottom.

Perceived Difficulty of Gestures

A one-way repeated measure analysis of variance was conducted to understand the perceived difficulty of the gestures during the pretest and again during the post-test. Eight different ratings were recorded, opening, aligning, the original moving, placing in a square, moving in a square, resizing from the corner, resizing from the edge, and resizing from the bottom (Figure 2). There was not a significant difference on perceived difficulty between the gestures *F*(1, 14)= 2.29 , *p*= .135.

For the post-test, the same eight ratings from the pretest were collected again after the game practice period. There was a significant difference between the gestures on perceived difficulty *F*(1, 13)= 4.25 , *p* <.05. Follow up comparisons revealed that opening the windows (*M*= 9.2, *SD*=1.1) was rated significantly easier than placing the windows in a 2 x 2 square (*M*= 7.4, *SD*=1.28), *p* =.034.

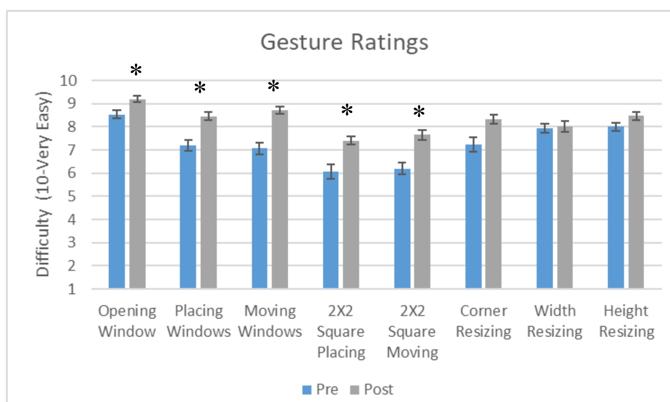

Figure 2. The pre and post perceived difficulty ratings for the gestures. A higher rating equates to the gesture being perceived as easier.
⁎ Indicates a statistically significant result

Perceived Usability

Participants reported the perceived usability (SUS) to be 66.67 (SD = 14.86). According to Bangor, Kortum, & Miller, (2009), this score is considered “OK”. Considering the participants were all first-time users of the new method of interaction, this rating is promising.

Discussion

The results of this study showed that with a short period of practice with the HoloLens headset and simple gestures for window manipulation, users can become significantly faster and perceive the gestures as easier. The gesture that showed the most improvement from pre to post-test was the select gesture. This gesture is one of the most simple to do, but requires practice to understand. One potential reason participants found this gesture so easy is because of how closely it represents clicking a mouse when using a computer. However, participants did tend to try and push their finger forward at first like they were going to touch the item. Opening the windows was the gesture that showed the next greatest improvement. This includes the use of the bloom gesture to open the menu and the select gesture to choose and place the window in the environment. Resizing the windows was rated as the most difficult by participants and showed the least amount of improvement with practice.

These findings provide insight as to what kind of training may be necessary for users to interact with HoloLens applications. With the release of the HoloLens 2 soon approaching, even more emphasis will need to be placed on the gestures. This is because the HoloLens 2 has been announced with an even more interactive gesture control suite (Microsoft, 2019).

The results from the ANOVA on the pre-test perceived difficulty showed that all of the gestures were viewed similar in terms of difficulty. In the post-test ANOVA we found that placing in a square was rated more difficult than the rest of the gestures. From this result we see that this gesture may be perceived as more complicated compared to the others. More time of practicing this gesture may help remedy this, or a redesign for gesture may be a better solution.
Future research in this area could focus on a wider variety of gestures used to control these devices. Also, voice interactions should be examined for scenarios where both of the user’s hands are involved in completing another task.

References

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

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.

Canare, D., Chaparro, B., & Chaparro, A. (2018). Using Gesture, Gaze, and Combination Input Schemes as Alternatives to the Computer Mouse. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 297-301). Sage CA: Los Angeles, CA: SAGE Publications.

Fariman, H. J., Alyamani, H. J., Kavakli, M., & Hamey, L. (2016). Designing a user-defined gesture vocabulary for an in-vehicle climate control system. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (pp. 391-395). ACM.

Graichen, L., Graichen, M., & Krems, J. F. (2019). Evaluation of gesture-based in-vehicle interaction: user experience and the potential to reduce driver distraction. *Human factors*, 0018720818824253.

Microsoft (2019). HoloLens 2. Retrieved from https://www.microsoft.com/en-us/hololens

März, P., Schwahlen, D., Geisler, S., & Kopinski, T. (2016). User expectations on touchless gestures in vehicles. *Mensch und Computer 2016–Workshopband*.

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.

Whitlock, M., Harnner, E., Brubaker, J. R., Kane, S., & Szafir, D. A. (2018). Interacting with Distant Objects in Augmented Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 41-48).