The Usability of the Microsoft HoloLens for an Augmented Reality Game to Teach Elementary School Children

Brita Munsinger
Department of Computer Science
University of Texas at San Antonio
San Antonio, Texas, USA
brita.munsinger@my.utsa.edu

Greg White
Department of Computer Science
University of Texas at San Antonio
San Antonio, Texas, USA
greg.white@utsa.edu

John Quarles
Department of Computer Science
University of Texas at San Antonio
San Antonio, Texas, USA
john.quarles@utsa.edu

***Abstract*—Our objective in this research is to compare the usability of three distinct head gaze-based selection methods in an Augmented Reality (AR) hidden object game for children: voice recognition, gesture, and physical button (clicker). Prior work on AR applications in STEM education has focused on how it compares with non-AR methods rather than how children respond to different interaction modalities. We investigated the differences between voice, gesture, and clicker based interaction methods based on the metrics of input errors produced and elapsed time to complete the tutorial and game. We found significant differences in input errors between the voice and gesture conditions, and in elapsed tutorial time between the voice and clicker conditions. We hope to apply the results of our study to improve the interface for AR educational games aimed at children, which could pave the way for greater adoption of AR games in schools.**

***Index Terms*—Augmented reality, computer security, gesture recognition, speech recognition, computer science education**

I. Introduction

There is much active research on using augmented reality (AR) for teaching STEM subjects to children. One such study uses a mobile device to add virtual content in the form of games to real world microscopes and microorganisms in science classes [1]. Another study added virtual elements showing air flow to an existing museum display of the physics concept known as Bernoulli's principle [2]. AR has been shown to be an effective way of teaching complex concepts to children and we have chosen to use it in the development of an educational game to teach cybersecurity. However it is not well understood which AR interaction techniques are the most effective for children using AR educational applications on the latest hardware, such as the Microsoft HoloLens.

An ongoing area of research in AR is the efficacy of different means of interaction with virtual objects placed around the real world. One recent study compared user interaction with a 3D visualization in each of three AR systems: a HoloLens, a tablet and a desktop computer [3]. Another recent study compared gesture recognition with voice recognition in a simulation where the user interacts with a virtual dog [4]. We have chosen to work with the Microsoft HoloLens (shown in Figure 1), which offers several modalities of interaction: voice, gesture, and clicker. While the prior work in interaction methods in AR mentioned above has already begun exploring the usability of voice and gesture, neither has done so using children as participants, which is the focus of this paper.

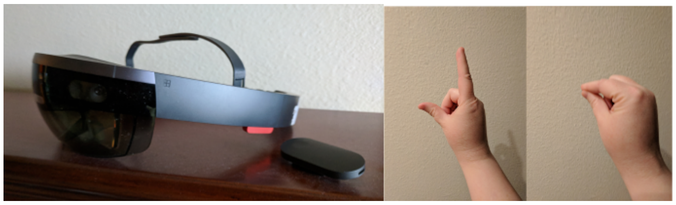

Fig. 1. HoloLens with the clicker and the air tap gesture.

II. Related Work

A. Augmented reality (AR) in STEM education

The paper by da Silva et al. [5] defines best practices for evaluating the educational effectiveness of AR games. They recommend close work with the teacher in designing the curriculum. Radu et al. [6] conducted an assessment study with elementary school teachers to find which mathematical topics would be a good fit for AR applications. They found teachers were enthusiastic about trying the new technology and it was beneficial to involve them in the design process. We also included a teacher in our process for this study. Our study also focuses on using AR to teach children, however we focus on cybersecurity instead of math. Also we study head-mounted rather than hand-held AR devices.

LaPlante et al. [1] developed an AR system to aid in the teaching of life sciences, specifically a mobile game that used light from a smartphone to interact with a light sensitive microorganism. User studies performed by the authors found increased engagement with the lesson and increased interest in pursuing a career in STEM in the future among students surveyed. In Yoon et al. [2], the authors find that the use of AR can aid in understanding of complex scientific concepts. They augmented a museum display with virtual

978-1-7281-4540-2/19/$31.00 ©2019 IEEE

elements depicting air currents moving around a physical ball, illustrating Bernoulli's principle. The authors found that the students' understanding was significantly improved after having interacted with the AR experiment versus those who had not used AR. Previous work such as these two studies demonstrates the value of using AR for teaching, and we would like to extend this into the field of cybersecurity.

B. Gesture, Clicker and Voice-based interaction

Bach et al. [3] compared three different AR systems for visualization, including the Microsoft HoloLens. We are also interested in exploring usability of the HoloLens, however we are comparing several HoloLens interaction methods, rather than focusing on the clicker as in this study. Also, we are studying the use of the HoloLens by children, rather than adult participants, and the impact of interaction method on learning outcomes.

Chen et al. [4] built a mobile AR game that uses both gesture and speech as interaction modalities. In the game players can interact with a virtual dog using spoken commands or gestures. In this study, the authors find voice recognition is more accurate than gesture recognition, but gesture is faster than voice. They conclude both are effective methods of interaction. We also plan to compare multiple interaction modalities in our study.

In [7], the authors study the performance of automatic speech recognition systems on the voices of children. Speech recognition engines are generally trained on adult speech. Children's voices are higher than adults due to a shorter vocal tract, and this can confuse speech recognition systems. Also, children tend to make more grammatical errors in speech and their voices change significantly as they grow up. Other work on speech recognition for children [8] finds similar difficulty in adapting systems designed for adults for use with children's speech. In our study we are also concerned with the voice recognition performance of the HoloLens on children's speech. We plan to build upon this previous work studying gesture, clicker and voice-based interaction in AR.

C. Ethics of AR research and children

One other aspect of this research we would like to mention is ethics in AR research with children. Prior work in this area [9] suggests several approaches for handling differences in how children respond to AR versus adults. One of their suggestions is to include an expert in child development on your research team. In this study we included an elementary school teacher with subject matter expertise in STEM. Another concern the authors mention is additional emotional impact upon children due to their developmental stage. This is something we take seriously and have reviewed the content included in our game with our teacher contact with this in mind.

III. Methodology and User Study

A. Experimental platform and game

Our augmented reality (AR) educational game was written in C# and built using the Unity3D game engine and deployed on the Microsoft HoloLens. Spoken cybersecurity information included in the game was recorded by the author. For all conditions, we added custom scripts written in C# to track participant performance in elapsed time and errors made. To develop the cybersecurity information used in our game and ensure it was age-appropriate, we consulted with our participants' teacher, who specializes in teaching STEM subjects such as cybersecurity.

The game itself is a three-dimensional version of a classic hidden object game. The game begins with a tutorial where players can practice selecting objects. The tutorial can be played as many times as you want. While playing the game, participants wander around their classroom looking for computer-generated objects hidden among real-world objects while wearing the HoloLens on their head. Once they have found and selected an object, a text message containing a piece of cybersecurity information is displayed in the air in front of their face and an audio recording of the same message is played. Once they have found all the objects, another message is displayed telling them they have finished. There are icons of the hidden objects displayed at the top of the HoloLens screen so participants can keep track of what they've found.

B. Participants

Our study participants were students from an elementary school in the southwestern United States. The students were all in the fifth grade and ranged in age from 10-13. The ethnic composition of our study population was over 90% Hispanic, with the remainder of the students split between Caucasian and African American. We had a total of 29 participants, 13 males and 16 females.

C. Experimental design and procedure

Before we began our study at the elementary school, we sought and received approval from our university's Institutional Review Board (IRB). As part of the IRB process, we created a paper permission slip that was distributed to all students who wished to participate in our study. Only students who had parental consent to participate and brought back signed permission slips were included in our study.

Prior to bringing in participants, we started the game and placed tutorial and game holograms around the classroom (see Figure 2). The size of the classroom was approximately 30 by 40 feet. There were tables placed along both sides of the classroom and down the middle, which we used to obscure some of the holograms (by placing holograms underneath so they could only be seen at certain angles). We added world anchors in our Unity code to preserve the location of the holograms between instances of the game so each time a new subject played the game the holograms would be in the identical location within the classroom. We also took screen shots of the placement of the holograms in order to verify their locations remained the same across study days. We ran participants over the course of three school days.

The experimental conditions were Voice (using a single spoken word command ("select") to communicate with the

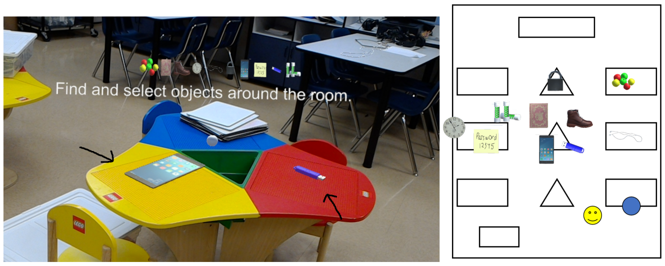

Fig. 2. On the left is a screen shot from the HoloLens showing a virtual smartphone and virtual USB stick drive in game. On the right is a drawing that shows the layout of the classroom in which we ran our study. The rectangles and triangles represent tables and desks within the classroom. The yellow circle shows the starting position of the participant. The blue circle shows the position of the researcher. The small graphics represent the locations of holograms placed in the room.

voice recognition software on the HoloLens), Gesture (using a simple hand gesture to communicate a command to the HoloLens), and Clicker (communication with the HoloLens by pressing a hand-held clicker device). Voice and gesture recognition was active as long as participants wore the HoloLens and no additional steps were needed to initiate recognition. We brought each subject individually into the classroom and gave them instructions about how to use the HoloLens and play the game. Neither their teacher nor any other students were present during the study. Instructions were given from a script. Depending upon which group the subject had been assigned, they would hear about how to speak with the HoloLens, use the air tap gesture, or use the clicker. Prior to putting the HoloLens on the subject's head, we would demonstrate the interaction method they were assigned to use and have them perform it as well to ensure their understanding.

After putting the HoloLens on the subject's head, they would begin to play the tutorial and game. All participants began the game from the same location at the front of the classroom. They were not given a time limit in which to complete the tutorial or game. They were also not required to locate holograms in any specific order. We kept track of the input errors (voice, gesture and clicker) used by the subject and how long they took to complete the game. There was no time limit imposed on any part of the process. Clicker and gesture input data were recorded automatically, while voice input data was recorded manually during the study. Once they were finished with the game, we took the HoloLens off and recorded the input and timing data collected by our game.

Each subject was assigned to the voice, gesture or clicker group. Our experimental design was between subjects with three groups.

D. Metrics

In our study we collected data on the total number of input errors generated by participants and the time taken to complete the tasks in game. Input errors are the total number of voice commands, air tap gestures or clicker presses that exceed the minimum required number (15) to complete the tutorial and game. Tutorial time is the elapsed time in minutes to complete the pre-game tutorial. Game time is the elapsed time in minutes to complete the game.

TABLE I
MEAN AND (STANDARD ERROR) FOR INPUT ERRORS, TUTORIAL TIME AND GAME TIME

Factor	Input errors	Tutorial time	Game time
Voice	35.75 (15.42)	2.67 (0.78)	4.39 (0.62)
Gesture	4.54 (2.03)	1.85 (0.58)	3.83 (0.61)
Clicker	8.1 (2.21)	0.44 (0.06)	2.98 (0.47)

E. Hypotheses

Based on previous work, such as the limitations of speech recognition for children [7], and the novelty of the tap gesture with respect to novice HoloLens users, we have formulated the following hypotheses about the selection task in a hidden object game.

- Hypothesis 1: There is a significant difference in input errors between voice, clicker and gesture conditions.
- Hypothesis 2: There is a significant difference in time (tutorial time and game time) between voice, clicker and gesture conditions.

IV. Results

See Table I for descriptive statistics on input errors, tutorial time and game time. We first ran the Shapiro-Wilk test on all data to test for normality. We found all p-values to be less than 0.05, indicating non-normal data. We also plotted the data to visually inspect for normality which confirmed the results we found on the Shapiro-Wilk test.

Next we ran a Kruskal-Wallis rank sum test on our data and a post-hoc Kruskal-Dunn test with a Bonferroni adjustment. We found two factors to be significant: input errors (χ^2 = 7.16, p = 0.03) and tutorial time (χ^2 = 8.91, p = 0.01). Time spent completing the game was measured separately from time taken to complete the pre-game tutorial and was not found to be significant. Using the results of our post-hoc Kruskal-Dunn pairwise testing, we found significant differences between the voice (V) and gesture (G) groups on the input errors factor (p = 0.02). Within the tutorial time factor, we found significant differences between the clicker (C) and voice (V) groups (p = 0.01). We also calculated the effect size for input errors and tutorial time over the three factors. While several of the effect sizes we found were large, the most dramatic effects we found were between the clicker and voice (0.5) and clicker and gesture (0.41) for tutorial time, and in the input errors found when comparing the voice and gesture conditions (0.42).

V. Discussion

In our study we examined the number of input errors and elapsed time over three conditions: voice, gesture, and clicker input. We divided elapsed time into portions corresponding to the tutorial and the game itself. Input errors were counted together across the tutorial and game. We had hypothesized that there would be significant differences between the input conditions that would manifest in the factors we measured.

Having analyzed our data, we found significant differences in two of the factors we measured: input errors and tutorial time. We find it interesting that for two of our factors, the total measurement including both tutorial and game was significant, while in the case of timing, only the part corresponding to the tutorial was found to be significant. It is possible that allowing our participants to practice as long as they wanted with the tutorial minimized differences in performance in the game itself. Prior work comparing the HoloLens with tablets and desktops for interaction with an augmented reality environment [3] found that participants' skills using the HoloLens improved with increased exposure. It appears that we may be seeing a similar effect, but we would need to explore this further in future work.

When we looked more deeply at pairwise comparisons within our data, we found several more significant results. There was a significant difference between the number of input errors performed by our participants in the voice and gesture conditions. This did not carry over to the clicker condition. This means we can accept our hypothesis H1. Prior work comparing voice and gesture inputs on other platforms [4] found that both modalities worked well for interacting with virtual objects, which is in contrast to our findings. Perhaps this has to do with our use of the HoloLens and its built in speech recognition rather than the Leap Motion and Google Speech API used in [4]. We thought that both the clicker and hand gestures might feel unnatural to children compared with speaking, so it is surprising that the differences between voice and gesture are significant while that between voice and clicker is not. Due to differences in how we measured inputs for voice, gesture and clicker, we feel it is premature to draw broad conclusions from this result. However, we feel further work will help illuminate this finding.

Our next significant pairwise comparison is in the tutorial time factor. In this case, the difference between the voice and clicker conditions was significant. We can accept our hypothesis H2. This is an interesting divergence from the previous result. Again, since we had anticipated that both clicker and gesture might be more difficult to master, it is surprising that the difference between voice and clicker is significant while the difference between voice and gesture is not. We already know that our participants had difficulties learning how to speak with the HoloLens, and this is consistent with previous work on speech recognition in children [7]. This result indicates that our participants learned how to use the clicker much more quickly. Time spent in the tutorial and game were measured identically within the game program itself over all three interaction conditions, so we feel this is our strongest finding of the three.

VI. Conclusions and Future Work

In our study we found several significant results regarding the usability of three interaction techniques on the Microsoft HoloLens. Our study focused on voice, gesture and clicker interaction methods in an augmented reality (AR) educational game for teaching children. One of the takeaways from this study is that voice recognition for children's speech continues to be a challenging problem. Even if the voice recognition works well for adults, as it does in the Microsoft HoloLens, that will not necessarily translate to it working well for children. We also found that voice-based interaction with our game produced many more input errors than gesture-based interaction, and that time needed to complete the tutorial in our game was significantly longer in voice-based interaction than clicker-based interaction. These were surprising results given that it would seem more natural for children to interact with a game using their voice rather than learning a new hand gesture or using extra hardware such as a clicker.

There are two avenues we would like to explore in future work. The first is whether voice interaction with the HoloLens would improve with the use of a different speech recognition system or some form of training with children's voices. Second, we would like to extend the educational game using augmented reality (AR) on the HoloLens to other STEM subjects. We look forward to continuing to improve the interactive experience of children with augmented reality.

VII. Acknowledgment

The authors wish to thank Cerina Zamora of Spicewood Park Elementary. This work was supported by grants from the National Science Foundation (DGE-1419367, IIS-1648949, IIS-1350995)

References

[1] C. LaPlante, M. Nolin, and T. Saulnier, "Playscope: augmented microscopy as a tool to increase stem engagement," in *Proceedings of the 12th International Conference on the Foundations of Digital Games*. ACM, 2017, p. 63.

[2] S. Yoon, E. Anderson, J. Lin, and K. Elinich, "How augmented reality enables conceptual understanding of challenging science content," *Journal of Educational Technology & Society*, vol. 20, no. 1, p. 156, 2017.

[3] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister, "The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality?" *IEEE transactions on visualization and computer graphics*, vol. 24, no. 1, pp. 457–467, 2018.

[4] Z. Chen, J. Li, Y. Hua, R. Shen, and A. Basu, "Multimodal interaction in augmented reality," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017, pp. 206–209.

[5] M. M. da Silva, R. A. Roberto, V. Teichrieb, and P. S. Cavalcante, "Towards the development of guidelines for educational evaluation of augmented reality tools," in *IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR)*. IEEE, 2016, pp. 17–21.

[6] I. Radu, B. McCarthy, and Y. Kao, "Discovering educational augmented reality math applications by prototyping with elementary-school teachers," in *2016 IEEE Virtual Reality (VR)*. IEEE, 2016, pp. 271–272.

[7] J. Kennedy, S. Lemaignan, C. Montassier, P. Lavalade, B. Irfan, F. Papadopoulos, E. Senft, and T. Belpaeme, "Child speech recognition in human-robot interaction: evaluations and recommendations," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 82–90.

[8] P. Vogt, M. De Haas, C. De Jong, P. Baxter, and E. Krahmer, "Child-robot interactions for second language tutoring to preschool children," *Frontiers in human neuroscience*, vol. 11, p. 73, 2017.

[9] E. Southgate, S. P. Smith, and J. Scevak, "Asking ethical questions in research using immersive virtual and augmented reality technologies with children and youth," in *2017 IEEE Virtual Reality (VR)*. IEEE, 2017, pp. 12–18.