

Insights on visual aid and study design for gesture interaction in limited sensor range Augmented Reality devices

Adam S Williams*
Colorado State University

Francisco Ortega†
Colorado State University

ABSTRACT

Augmented reality (AR) is becoming increasingly available in consumer markets. Most AR devices ship with some form of gesture interactions enabled. Gesture interactions are sometimes made difficult by limitations in the device's sensors and visual field of view. A series of pilot studies comparing different on-screen aids geared towards alerting a user if their gestures were within the range of recognition for a device are presented here. Users have a stated preference for having no borders present after learning the limitations of the system. Before that learning has occurred users prefer to have a solid border shown.

Index Terms: H.5.2 [User Interfaces]: User Centered Evaluation—User Centered Design; H.5.m [Artificial, augmented, and virtual realities]: Evaluation/methodology

1 INTRODUCTION

Augmented reality (AR) devices are becoming increasingly available to consumers. When we look at the current state of the art in these devices several limitations impact a user's initial experiences with them. These limitations include include limited Field of view (FoV) and limited sensor ranges. The Microsoft HoloLens and Magic Leap 1 are some of the best head-mounted optical see-through (HMD) AR displays available today. The Magic Leap FoV is 40 degrees by 30 degrees [1], yet human vision is 210 degrees by 150 degrees (though all of that is not used for binocular vision) [6]. In the device used here (Magic Leap 1) the sensor range for gesture recognition and the FoV are the same.

When considering AR technologies in the long term, we can expect that the sensors and field of view (FoV) in AR devices will improve substantially. However, between now and that time it is important that we learn to work with the current limitations of these devices.

A second reasonable belief around the future of AR is that mid-air gestural interfaces will become the interaction technique of choice for these devices. Some devices are already shipping with gestures as the stock interaction technique (HoloLens 1-2) [2]. The issues of the limited sensor range and field of view become very apparent to a user when they are using gestures in front of the device. A user's gesture could be correct but outside of the sensor range of the device and not recognized.

These limitations pose several problems. Laymen users may not be familiar with the range of recognition in AR-HMDs. Their first instinct may be that the device is not functioning properly. To examine the prevalence of gesture interaction failure due to a user's hand leaving the device FoV we conducted a series of three pilot studies. We hypothesize that users will have high error rates due to that type of failure. We also hypothesize that an on-screen border around the device's sensor range/FoV would reduce that error rate.

*e-mail: adamwil@colostate.edu

†e-mail: fortega@colostate.edu

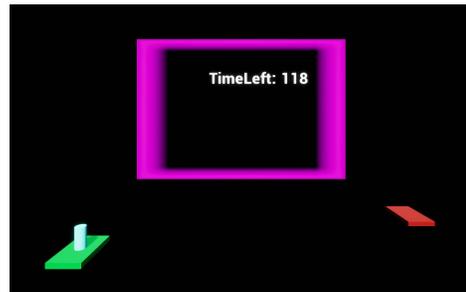


Figure 1: Fully rendered border during study 3 shown in editor

2 STUDY DESIGN

This work shows the results of three small pilot studies. These studies are designed to find out how prevalent dropping objects due to a user's hand leaving an AR-HMD's FoV is. Each of the studies tasked a user with picking up a virtual cylinder about the size of a drinking glass (Figure 1). The gesture interface is designed to match the hand shape someone would use when picking up a physical version of the cylinder. To pick up the cylinder a participant had to grasp it as though they would a physical object. Anything from an open C hand position to a fist was accepted as a grab action. The object is released when the hand is opened or when the hand exits the sensor range.

The NASA Task Load Index (TLX) was administered for each condition [3]. Informal interviews were conducted after the studies. These interviews always included a conversation about which border participants felt helped them with the task and which hurt them as well as an open conversation about the participant's impressions of the borders.

2.1 Interaction Gesture

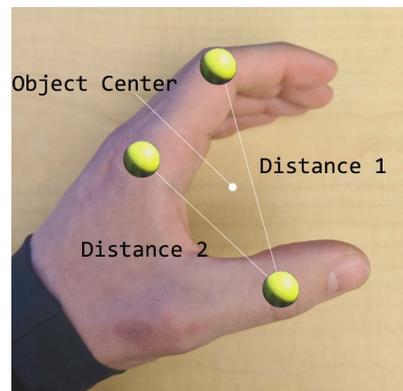


Figure 2: Hand tracking points

The thumb tip, index base, and index second knuckle were tracked for interaction with the cylinder (shown in Figure 2). From testing different tracking locations we found these to be the most sta-

Table 1: Summary

	Study 1	Study 2	Study 3
Movement goal	30 picked up cylinders	30 picked up cylinders	place cylinders until time is out
Participant position	seated	standing	standing
Platform distance	3 feet apart	3 – 3.5 feet apart	3 - 6 feet apart
Platform moving	none	6 inches of movement	none
Time Given	unlimited	unlimited	200 seconds
Number of participants	6	4	4

ble points. To help avoid self-occlusion errors the cylinder stayed grabbed for 1.5 seconds after a tracked point was lost while near the FoV center. This had the drawback of participants needing to open their hand for that long to drop the object intentionally. The movement of tracked positions would only register if it was within half an inch or less of the last used location. After one second of no movement, a large update was allowed. This controlled the noise seen in the tracked points. The angle caused by platforms being placed at elbow height improved the headset camera’s position relative to the participant’s hand which improved tracking accuracy.

The distances between the index and thumb (distance 1 in figure 2), as well as the distance between first index knuckle and thumb (distance 2 in figure 2), were used to determine if someone was using a grab gesture. If any of those distances were equal to or below the diameter of the cylinder a grab attempt was registered. This is similar to the grab gesture approach taken by Ismail et al. however the points tracked were different [4]. We chose not to coach people on how to maximize hand tracking because it would impact how often they let their hand leaves the device FoV. The only instructions for gesturing given to the participant was “grab the cylinder with your hand.” This design taps in the perceived affordances of a cylinder [5]. With this design, none of the users required any help for this gesture and no-one asked how to grab the cylinder. When the tracked points were close to the cylinder (0.5 inches) and a grab was registered the object was picked up and centered in the user’s hand (object center in figure 2). The z-axis of the cylinder was set to be perpendicular to the plane created by the three tracked positions. This caused the cylinder to rotate with a participant’s hand. No yaw rotations were added as the cylinder was the same on all sides.

2.2 Setup

These studies share the same design and task but with three manipulations aimed at making the task more difficult (see Table 1). The main measure of these studies is the count of times a participant’s hand leaves the device FoV when using gesture interactions to grab and move a virtual cylinder. Leaving the FoV when interaction with gestures is referred to as a drop.

The cylinder always appeared on a green 6 inch by 12-inch platform to the left of the user (Figure 1). A second red platform of the same dimensions is to the right of the green platform. The platforms were set at elbow height for each participant. The task was to move the cylinder from the green platform to the red platform. Using their hand to grab it. When the object was successfully placed on the ending platform it was counted as a success. Both a drop and a success trigger the next cylinder to spawn.

2.2.1 Inter-Study Manipulations

The manipulations outlines below are reflected in table 1.

In studies one and two the distance between the platforms is about 1.5* the FoV of the device so that participants had to move their head while moving the cylinder. This equated to about three feet of spacing between platforms when rendered in the headset. The participant was seated during this study. The participants were given 5 practice trials where they were encouraged to become familiar with

this system. After the practice trials, they performed the movement task 10 times in each condition.

In study two the participant was standing. They were given 30 seconds of practice and 60 seconds to move the object as many times as possible. A timer was displayed on the screen (Figure 1). The other difference in study two was that the ending platform moved back and forth between 3 feet to 3.5 feet away over 2 seconds.

In study three the participant was standing and was asked to do the task as many times as possible in 200 seconds. The end platform location was spawned randomly between 3 to 7 feet to the left of the starting platform and 0 to 4 feet in front of the starting platform. The participants were given 90 seconds to practice. The increase in the time given was due to the task now involving movement to and from the ending location and a search component.

The manipulations in study design were to decrease performance and better measure the prevalence of a user’s hand leaving the FoV of the device. A summary of study designs can be found in table 1. An example of a participant interacting in study 3, the task mid-session, and the task near completion can be found in figure 3.

2.2.2 Conditions

Three conditions were tested in each study. A solid border visually indicating the gesture recognizer range, no border, and a border that became brighter as the user’s hand approaches the edge of the recognition area (fading border). The solid border is shown in figure 1. The fading border was the same asset but had the opacity set between 0% and 100%. The exact opacity setting was the relative distance of the user’s hand from the center of the sensor FoV to the edge as a percentage. For example, if a user’s hand was halfway between the center of the screen and the edge of the FoV the border opacity was set to 50% opacity. The border was fully opaque when a participant’s hand center (Figure 2) was in the screen center.

2.2.3 Condition Order

In the first study, we alternated the presentation order of conditions to account for the learning rate. Participants from study one consistently said that the solid border condition made the task and FoV limitations immediately apparent. This was reflected in the data (expanded on in the results section). No participant that started with the solid border dropped the cylinder. Due to that feedback, studies two and three used fading border or the no border condition as the first or second condition. The solid border condition was always last.

2.3 Participants

This study consisted of 15 participants (9 male, 6 female). The participants ranged in age from 31 to 21 ($\mu = 25.83, \sigma = 9.65$). Most of the participants were undergraduate students (10), 2 were graduate. The others were not in university studies. One participant reported limited AR experience (< 1 hour). Two participants reported playing 5 hours or less of virtual reality games per week. Both said it was *beat-saber* which does not involve any hand tracking. All participants reported normal or corrected to normal vision and were right-handed. Study 1 used 6 participants, study 2 and 3 used 4 each (Table 1). Over all the number of times an object was dropped



Figure 3: Study setup, participant left, mid-task view from device middle, end of task view from device right

due to self-occlusion tracking errors was limited to 1-2 times per participant. One participant in study 1 had (15/30) of these errors and was removed from the analysis. All participants reported being right handed.

2.4 System

These studies were conducted using the Magic Leap 1 optical see-through augmented reality headset [1]. The development was done on a windows 10 professional computer with an Nvidia RTX 2080 TI and an Intel i9-9900k 3.6 GHz processor. These studies were developed in the Magic Leap Unreal Engine release 4.23.0.

3 RESULTS

When examining all the studies together the fading border and the no border conditions experienced the most drops when presented first (15/28, 30/37). Participants on average dropped the cylinder 38.94% of the time with no border, 29.47% with a fading border, and 22.11% with a solid border. This trend reflects that people did perform better with a border. This effect is impacted by learning rate and condition order which is discussed below. Overall of the studies the no border condition had the highest number of drops (37), followed by the fading border (28), then the solid border (21). The solid border had no drops when presented first.

Most participants initially used their left hand for the experiment then switched to their right hand. This is most likely due to the placement of the cylinder near their left hand. Participants typically had more successful placements in the solid or fade conditions. This is shown in figure 4. Note that in study one the maximum placements per condition were fixed at 10 for all participants. Study 3 had the longest times given for each condition which may contribute to the larger disparities in placements between borders.

Study 1 by design had 10 attempts per condition. An attempt is any successful pick up of the cylinder, regardless of placement. Study 2 had an average of 13.08 attempts and study 3 had 21.58 attempts per condition. Most participants across studies 2 and 3 had nearly the same number of attempts across all conditions (within 2). In studies 2 and 3 when there were uneven numbers of attempts in a border condition it was in the fade or solid condition. One participant had more attempts in the no and solid and one had more tired in the fade and no conditions.

Study 1 and study 3 had comparable numbers of dropped cylinders (35, 42 in order). Study 2 had 9 drops in total, 8 of which were from one participant. This could be caused by the moving platform making this task more interesting/immersive. The immersion may have caused participants to be more careful with their actions. The breakdown of the percentage of attempts that ended with a dropped cylinder by the study design can be seen in figure 5. In study 3 the no border condition has the highest percentage of failed attempts (23.53%). In this study, the no border and fading border conditions were both presented first an equal number of times. The relatively similar number of failed attempts in study 1 indicated that within

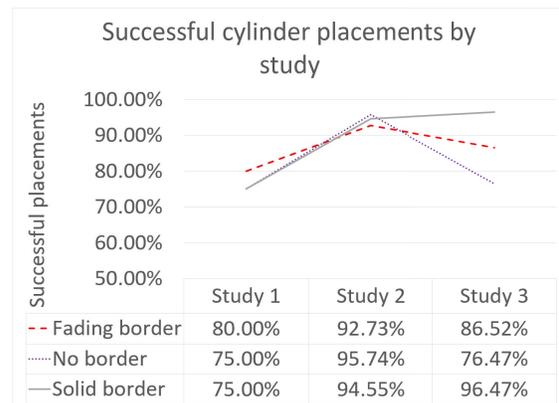


Figure 4: Successful cylinder placements by study (note y axis starts at 50%)

the 10 attempts per condition participants did not fully learn the task. When given more time (studies 2 and 3) participants showed more learning effects. This is most drastically shown in study 3 where the solid border condition (presented last) only has a failure rate of 3.53% (Figure 5).

4 LEARNING RATE

Users quickly learned the bounds of the system immediately upon seeing the solid border. The fading border did not show as strong of a learning effect. In the no border condition people typically realized that the best strategy was to hold your hand directly in front of them. This strategy was encountered in every participant at some point for studies 1 and 2. In study 3 participants tended to hold the object lower than the center and allowed more movement towards the edges of the screen.

In study 2 everyone did very well possibly due to increased interest with the moving platform adding a game nature to the task. Study 2 had the lowest number of attempts in the no border condition (47, mean attempts: 52). This supports our hypothesis that participants took study 2 more seriously and thus acted more deliberately. In study 3 the learning curve from condition to condition is lowered from 29 drops total in the first condition to 5 drops in each of the next two condition.

5 INTERVIEWS

After each study, we asked participants if any border made the task easier, if any made it more difficult, and for any comments they had. Most participants (10/14) said that they liked not having a border but would initially want a border to help familiarize themselves with the bounds of the system. Typically, these participants said the solid border was the best for learning the system bounds. After that

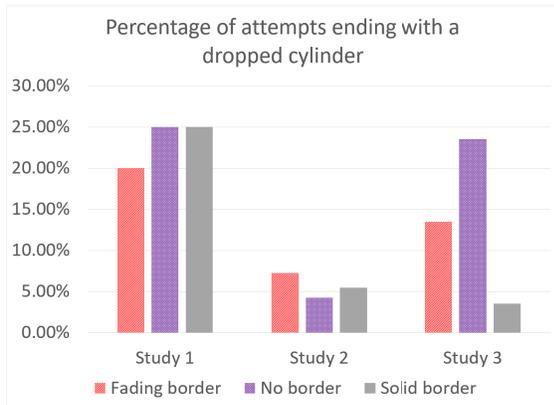


Figure 5: Difficulty by drops

familiarization was done, they would want the border to be removed. Reviews were mixed between the faded and solid border. People that liked the solid border (7/14) said it was the most easily accessible information on bounds but also said it was distracting. When the faded border was liked (2/14) it was due to not being shown when using the device within the appropriate bounds. Three participants were entirely unaware there was a boarder in the fading condition. These were participants that were presented with the fading border last.

6 NASA TLX

Table 2: Average NASA TLX scores by condition

	No border	Faded border	Solid border
Study 1	31.94	34.17	32.36
Study 2	33.33	38.33	26.94
Study 3	39.58	38.54	36.25
Mean	34.62	36.47	32.31
SD	14.52	15.68	11.96

Table 3: Average NASA TLX frustration scores by condition

	No border	Faded border	Solid border
Study 1	22.5	26.67	28.33
Study 2	18.33	20	18.33
Study 3	33.75	38.75	36.25
Mean	25	28.857	28.46
SD	16.05	23.63	20.97

The overall NASA TLX perceived workload scores across studies are shown in table 2. In study 1 the no border condition has the lowest score. In both studies 2 and 3, the solid border has the lowest perceived workload. The solid border also has the lowest standard deviation. This could be attributed to it being the last condition presented. When only looking at the frustration scores (Table 3) the solid border and faded borders have higher scores than no border. The frustration metrics align with what participants reported in their interviews. Even while the no border condition was most often the first condition encountered participants felt that is was the least frustrating.

7 DISCUSSION

Users of augmented reality devices quickly learn how to work within the limitations of a devices FoV, even without any explicit instructions. Having a border results in quicker learning (lowest results to begin with). That said most participants (13/14) mentioned that a

solid border is distracting or detrimental and while it is helpful for quickly learning the limitations of the FoV, it should be removed eventually. Most participants either did not notice (3/14) or did not care for (7/14) the boarder that faded.

8 GUIDELINES

When developing tasks in AR a boarder can quickly show users where they need to position their hands for the device to recognize them. This border is suggested to be removed after the user has learned the bounds of the system. The border that faded was disliked by most participants. Many cited it as being distracting. We recommend using no border or a solid border if needed to avoid this.

If the task becomes more interesting users seem to focus on the task performing it more slowly. This focus leads to increased performance despite also having increased task difficulty.

9 FUTURE WORK

We found that the task itself was easy enough that participants did fairly well very early on. With that evidence and the feedback from study one we intentionally presented the solid border condition last in studies 2 and 3. This choice was made to help ensure that there could be errors in the first two conditions. That said it could have inflated the number of successes found in the solid border condition. A future study with all permutations of condition order presented to participants would help find out how much that choice impacted our findings.

Each of these three studies was done with a different set of participants. That allowed for use to minimize learning the task before starting an experiment. While we expect that study 2 was more difficult than study 1, performance in study 2 increased. With no participant completing each study we can not be sure how much the gamified nature of study 2 contributed to this performance increase. A within-subjects experiment with each participant doing each study would help determine this effect.

10 CONCLUSION

This paper presents a process of iterative study design throughout three small user studies. Each study was slightly more difficult than the last. These studies aimed to find out how much on-screen visual aids in the way of borders impact user performance at an interaction task. We found that not having a border was preferred to having one. That said most participants reported that initially having a solid border to help them learn the recognition area of an HMD would be helpful. Maintaining that border past that point increased user frustration with the task. We also present a gesture design for grabbing and translating virtual objects in augmented reality that has no learning curve.

There is a Demo available for this experiment.

REFERENCES

- [1] Learn: Magic leap. <https://developer.magicleap.com/learn/guides/field-of-view>.
- [2] Microsoft hololens: Mixed reality technology for business. <https://www.microsoft.com/en-us/hololens>.
- [3] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., *Advances in Psychology*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, Jan. 1988.
- [4] A. W. Ismail, M. Billinghamurst, M. S. Sunar, and C. S. Yusof. Designing an augmented reality multimodal interface for 6DOF manipulation techniques. In *Intelligent Systems and Applications*, pp. 309–322. Springer International Publishing, 2019.
- [5] D. A. Norman. *The psychology of everyday things*. Basic books, 1988.
- [6] H. M. Traquair. An introduction to clinical perimetry, chpt. 1. *London: Henry Kimpton*, pp. 4–5, 1938.