The Role of Latency in the Validity of AR Simulation

Cha Lee*
University of California, Santa Barbara

Scott Bonebrake†
University of California, Santa Barbara

Tobias Höllerer‡
University of California, Santa Barbara

Doug A. Bowman§
Virginia Tech

Abstract

It is extremely challenging to run controlled studies comparing multiple Augmented Reality (AR) systems. We use an AR simulation approach, in which a Virtual Reality (VR) system is used to simulate multiple AR systems. To investigate the validity of this approach, in our first experiment we carefully replicated a well-known study by Ellis et al. using our simulator, obtaining comparable results. We include a discussion on general issues we encountered with replicating a prior study. In our second experiment further exploring the validity of AR simulation, we investigated the effects of simulator latency on the results from experiments conducted in an AR simulator. We found simulator latency to have a significant effect on 3D tracing, however there was no interaction between simulator latency and artificial latency. Based on the results from these two experiments, we conclude that simulator latency is not inconsequential in determining task performance. Simulating visual registration is not sufficient to simulate the overall *perception* of registration errors in an AR system. We also need to keep simulator latency at a minimum. We discuss the impact of these results on the use of the AR simulation approach.

Index Terms: I.3.7 [Three-Dimensional Graphics and Realism]: Virtual Reality—AR Simulation; I.3.6 [Methodology and Techniques]: Device independence—Replication

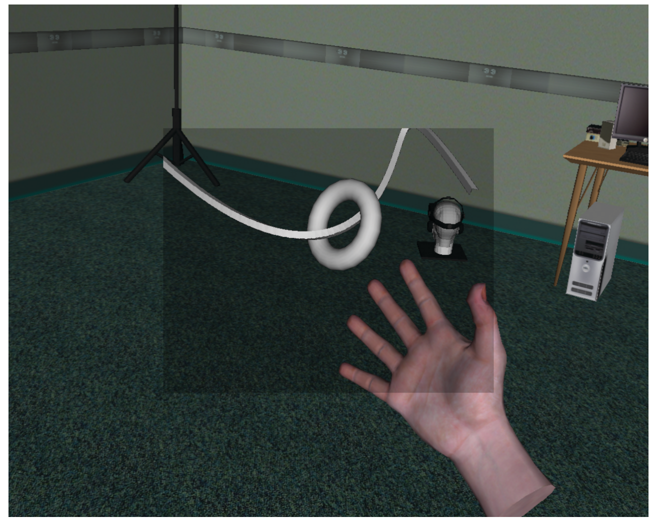

Figure 1: Screenshot of participant's view in Experiment 1. AR window is shown as the dimmer center area where only ring and path are visible.

1 Introduction

The objective level of perceptual fidelity (i.e., immersion [1]) is a fundamental characteristic of Augmented Reality systems. The effects of the level of immersion are not well understood and we need this understanding in order to select hardware and software components that will be effective in real world AR systems. A better understanding of these levels of immersion would also give us a firmer understanding of basic issues regarding the effectiveness of AR. Due to the many options available (hardware and software), direct comparisons of competing technologies are neither scalable nor generalizable. We can avoid these limitations and achieve experimental control by using an "AR simulation" approach, in which a VR system is used to simulate a range of AR systems. An AR simulator, as proposed by Gabbard et al. [4] and Ragan et al. [10], would be capable of simulating multiple hardware configurations, environments at many levels of perceptual fidelity, and controlled and repeatable interactions.

AR is very much dependent on its underlying hardware components (tracking system, graphics system, display, etc). For example, tracking systems can vary greatly in accuracy, refresh rate, and latency depending on the vendor and technology while displays can cross a wide range of field of view, resolution, luminance, and refresh rate. By using a high fidelity VR system as the framework for our AR simulator, we can simulate multiple AR systems. Such a VR system would have high end computing and graphics processing, accurate low-latency tracking, and possess a high perceptual fidelity display. It is easy then to degrade the performance of these components in software to simulate the performance of a variety of systems and levels of immersion. Obviously the range of level of immersion a simulator can simulate is limited to the current state of the art in the technologies used. The range of values associated with resolution, latency, jitter, and luminance are currently limited, but the technologies exist. We can reasonably expect to increase this range as their technologies progress over time. It is more difficult to simulate components such as tactile feedback and true depth cues because these immersion components are more fundamentally difficult to produce. The technologies for creating convincingly lifelike sensations in these areas just don't exist yet. Although limited tactile interaction is feasible, there is no solution for unencumbered and full-body tactile interaction yet. Real-world outdoor scenes are problematic as it is also impossible to simulate the real world with accurate accommodation cues safely. Only true volumetric displays allow accurate accommodation cues, and none provide the visual quality and interaction needed. Even with these restrictions, we feel that many tasks and scenarios in AR can be simulated to a high enough fidelity where experiments become possible in simulation.

But are the results of experiments using AR simulation even valid for real-world AR systems? There are multiple steps required to validate AR simulation. We must analytically compare the level of immersion of our final simulator to real world AR systems so

*e-mail: chalee21@cs.ucsb.edu
†e-mail: egscott2004@gmail.com
‡e-mail: holl@cs.ucsb.edu
§e-mail: bowman@vt.edu

IEEE Virtual Reality 2010
20 - 24 March, Waltham, Massachusetts, USA
978-1-4244-6238-4/10/$26.00 ©2010 IEEE

that these values make sense and are reasonable. Then we need to replicate a small set of experiments from the literature and show that the results from simulation are comparable to the established results. Finally, we need to do direct comparisons between studies run on our simulator and studies with real, practical systems.

In this work we begin to address the validity of AR simulation. We first present our results on replicating a well known AR experiment, and in a follow-up study, we investigate the effects of simulator latency on a 3D tracing experiment performed in AR simulation.

2 Related Work

Much work has been done on comparing multiple VR and AR systems, but these findings are hard to generalize since they are dependent on the overall systems used in those experiments. Pausch et al [9] studied visual search between a VR system and a desktop system. Pausch et al. found that non-head-tracked users in VR took significantly more time to determine if a virtual target existed in the room when absent. The display used in this particular study was a HMD mounted on a ceiling post. The VR scenario allowed the users to freely look around the room by rotating their head. The non-VR scenario restricted users' head movements. Interestingly Robertson et al. [11] found that this finding did not extend to VR on a desktop display. In a similar task in AR, Wither et al [14] evaluated different display devices for selection and annotation tasks in AR. The authors compared a HMD, hand held magic lens display, and a tablet display held at waist level. Wither found that using a hand held display in magic lens configuration was faster for cursor movement than either the HMD or tablet display and that there was no significant differences in the visual search task between the displays. The authors had hypothesized that visual search of virtual letters should have been faster with the HMD because of its higher level of immersion, yet the results were inconclusive even though users qualitatively agreed with this hypothesis. These different conclusions on similar tasks highlight the difficulty in generalizing results based on real AR/VR systems. The hardware and level of immersion of the systems were too different. Perhaps other differences in the display may have affected the results and the main real conclusion is that these findings are dependent on their respective scenarios.

Using VR to simulate different levels of immersion, while controlling other components of immersion, has been done before with promising results. McMahan et al [8] studied the effect of field of regard, stereoscopy, and different interaction techniques on object manipulation. By using a three sided CAVE [2], the authors were able to control the level of field of regard by enabling or disabling multiple walls (thereby decreasing and increasing the field of regard of the display). McMahan and colleagues used a single VR display to simulate multiple VR systems with differing levels of immersion, while maintaining experimental control. While field of regard and stereoscopy were not a significant factors for any metric, the 3D interaction techniques had a significant impact on manipulation time and the number of clutches. The results from this experiment are more general since the unknowns were controlled.

In a similar fashion, one can use a VR display to simulate different real world AR systems. In Gabbard et al [4], the authors presented an AR-within-VR concept which placed the AR user in an immersive VR environment (the CAVE). The AR user wore an optical, see-through display and observed graphics registered within VR which represented the real world. The goal for Gabbard and colleagues was to be able to do outdoor AR experiments without the difficulty of actually being outside. The unpredictable nature of weather, and wear-and-tear on both equipment and people make outdoor AR experiments an arduous task. Using the CAVE, the authors wanted to control both ambient and background lighting while maintaining perfect registration, which was impossible with a real AR system. Although it turned out that it was not possible to control ambient lighting while maintaining realistic outdoor illumination levels, it was found to be well suited for controlled AR experiments at night, dawn, dusk, or indoor AR. In Kim et al [5], a desktop VR system was used to simulate an AR heads-up windshield display. A wide-screen TV, contact-less gaze tracker, and wheel joystick set was used as the simulation framework. A qualitative and quantitative evaluation of the simulated display showed a significant reduction in navigation errors and distraction-related measures compared to a typical in-car navigation display for elderly drivers. This study was only possible through simulation since the technology for the actual display did not yet exist. Performing real studies would have been much more difficult even if the display existed due to the nature of the task. In this particular example, AR simulation was the only option.

Recently there has been work to explore AR simulation in depth. In Ragan et al. [10], AR simulation was used to investigate the effects of registration error on task performance for a generic task involving precise motor control for AR object manipulation. A four sided CAVE, an InterSense IS-900 tracking system, and ARToolKit marker props were used to simulate a AR system. Ragan found that both jitter and latency affected tracing performance, with jitter having the larger effect. This result could not have been easily achieved outside of simulation since it required the isolation of jitter and latency from other components of an AR system. Ventura et al. [13] examined the effect of varying levels of field of view and reliability of head-tracking sensors in a target following task. The authors simulated an AR system with X-ray vision within a VR based simulator. They found both field of view and tracker reliability to have significant effects on tracking objects. Their setup avoided the feasibility and control problems of tracking many real people in a live experiment.

Although AR simulation has been proposed and used for multiple experiments, the validation of the concept of AR simulation has not been addressed. Our work is a step toward this goal. Preliminary results of our replication study have been presented as a poster at ISMAR 2009 [6].

3 Experiment 1: Replication Study

The goal for Experiment 1 was to replicate an established AR study within our simulator as a step toward validation of AR simulation. We chose to replicate the second experiment in Ellis et al. [3], which showed that high-precision path tracing is most sensitive to increasing latency. The experimental design included in the published work was highly detailed which made this particular work desirable for our purposes. While our simulator and Ellis et al's AR system did differ, we attempted to replicate the system performance as closely as possible from the published work and with the aid of Dr. Ellis. Our hypothesis was that we could successfully replicate the results from the original experiment if we were able to replicate the level of immersion present in the original authors' system. We considered the visual fidelity of the display and end-to-end latency to be the two most important immersion components because of the nature of the experiment. In the original experiment, participants were shown a very simple 3D path and ring in gray scale through a head mounted display (HMD). Users were then asked to trace the path with the ring. It was important to restrict the field of view and the resolution to the original hardware used, to fully replicate what participants saw (with respect to the virtual scenes). Although the real world scenery was also shown at this low resolution in our study (unlike the original) we did not consider that as an important component since the objects of interest were all virtual. By restricting the field of view, we also made sure users saw the same amount of the virtual path. Finally, by carefully controlling our own system's end-to-end latency, we could replicate the same overall latency on the virtual objects studied in the original work.

3.1 Task and Environment

The task in the experiment was a 3D tracing task. Participants were shown a gray scale 3D path with a cross-section of 0.5 cm and a length of 76 cm. Attached to the participant's hand was a virtual ring. There were two rings used of inside/outside diameters: 5.08/9.65 cm (large) and 1.78/3.30 cm (small). With no system latency, the ring would be rigidly attached to the participant's hand. The participant was then asked to trace the path from end to end with the inside of the ring, while trying to avoid any collisions between the path and ring. Feedback of a collision was signified by the path blinking brightly and a beeping sound. Each participant was free to move around in a limited area (due to cable constraints). Additional end-to-end latency was then added to the virtual objects in varying amounts for each condition. In Ellis et al. [3] the only virtual content seen by the participant were the ring and path. Since Ellis's experiment was a see-through AR experiment, the ring and path were overlaid over the participant's hand and the lab space at all times. To simulate this correctly, we used a "simulated real hand" to represent the participant's actual hand and a 3D model of the ReCVEB lab at UCSB was used to represent the environment. When additional latency was added, it was only applied to the virtual objects and not the simulated real world objects. In software, we made sure to always render the path and ring on top of the hand and room regardless of orientation, to simulate see-through AR. The display used in Ellis et al. was a CRT haploscope with a 21.4 degree total field of view (19 deg. for each eye). Although the virtual objects could only be seen in this restricted field of view, the haploscope still allowed users an unrestricted view of the real world. Since our HMD was considerably larger in field of view (48 degrees), a transparent AR window was rendered into the participant's view which was equal to 21.4 degrees. The path and ring were only visible within this window as seen in Figure 1 and the simulated real world objects could be seen in the entire 48 degree field of view.

3.2 Apparatus

Our simulator hardware consisted of a Kaiser Proview 60 HMD, WorldViz Precision Position Tracking (PPT) system, two wired Intersense InertiaCube2 orientation sensors, and an Intel Core2 CPU 6600 @ 2.40 GHz with a NVIDIA Quadro FX 4500 video card and running Windows XP. The graphics software used was WorldViz's Vizard development environment, a python scripting development platform using the OpenGL libraries. The two InterSense InertiaCube2 sensors in conjunction with three markers (two for head, one for hand) from the PPT system were used to track the head and hand of the user in 6 degrees of freedom. The simulated real hand model used to represent the actual hand of the participant was a simple unarticulated hand which held the same open pose. Due to tracker constraints we chose not to track the fingers and treated the hand as a rigid object. We surmised this loss of realism was inconsequential for the outcome of this particular study. Other task scenarios might well require more careful modeling and tracking of a user's hand posture.

Base end-to-end latency of our system was measured off-line by using two photo sensors and a two-channel oscilloscope. The first photo sensor was connected to the first channel and detected the start of the tracking phase from a LED marker. The second photo sensor was used to detect a monitor and was attached to the second channel of the oscilloscope. When the rendering application received a valid position target from PPT, a white screen was rendered and was detected by the second photo sensor. The latency was calculated based on the difference of two signals. PPT was modified to specifically search for only three markers and to decrease computation time and vertical sync was also turned off to decrease rendering latency. The off-line end-to-end latency was $\approx$ 50 ms (+/- 5 ms). On-line latency may have been worse, though.

3.3 Study Design

We used a mixed experimental design with repeated measures. The within-subjects independent variables were path type, and end-to-end latency. The between-subjects independent variable was ring size (large, small). Two path types were used: angular and smooth. Angular paths were straight paths which could only bend at 90 degree angles while smooth paths were based on splines and curves. The paths were randomly generated and three paths from each group were selected for each participant randomly. Five different end-to-end latencies (which only affected the display of the virtual path and ring) were used: 50 ms, 100 ms, 200 ms, 300 ms, and 500 ms. This created 30 different conditions for each ring size. These conditions were randomly ordered into a block of trials and each participant repeated the same block three times (order of trials being randomized at each iteration). The dependent measure in this experiment was the number of collisions between the ring and the path. Participants were split equally between the two ring types: seven participants used the small ring, and seven the large ring.

3.4 Participants and Procedure

For the experiment we had 14 unpaid volunteers, ages 21 - 33, nine male and five female. A questionnaire was given to each participant beforehand. All participants reported they were comfortable around computers and had limited to extensive experience with 3D games. All participants were able to perceive stereo as verified by a random dot stereogram.

The HMD was automatically calibrated and registered to each user with a constant position error of less than two cm. Previous work [12] has found evidence that users are able to adapt to small perturbations in head registration for motor performance in VR tasks. Before beginning the actual experiment, each participant was asked to spend approximately five minutes on a training data set. A moderately easy path was picked for this training phase and no extra end-to-end latency was introduced. The study moderator guided the user until the user was both comfortable completing the path and was completing the path at a consistent pace before proceeding to the actual trials. Once the study began, the study administrator monitored the experiments and ensured each user completed the experiment as intended (no skipping). We did not add any software guarantees for completing the task correctly, since the original experiment did not have them as well. For each trial, the participant would touch the virtual ring to a start object (a virtual box) before attempting to trace the path. Once the path was completed, the participant would touch a nearby red box with the virtual ring to signify the end of the trial. At this point the participant's score was recorded automatically. Participants were allowed to rest between each trial and a mandatory break was enforced after each block of trials.

For more details on the original experimental design and models, please refer to Ellis [3].

3.5 Results

Preliminary analysis of the results showed that users reached asymptotic performance after the first block of trials; thus only the second and third blocks were used in the following analysis. Ellis' significant effects were based on the log of the raw collision score, and to be consistent we also followed this rule. All significant effects, based on a multi-variate ANOVA of the log of the average tracing performance of blocks two and three, are shown in Table 1. These results are comparable to the results from Ellis [3] as shown in Table 2. Unlike the original experiment, our study also showed a significant interaction between path and latency as discussed below.

In addition to comparable statistically significant effects, all of the effects were in the same direction as in the original study. The number of collisions was greater for the angular paths as compared to the smooth paths, the small ring resulted in worse performance, and the number of collisions increased as latency was increased. Although the effects were similar there were some interesting differences in the absolute performance data, as shown in Figures 2 and 3.

The participants' absolute tracing performance is noticeably worse in our experiment. We believe this may have been due to the differences in the collision algorithms used, and the difference in mis-registration between the original experiment and ours. It was unclear from Ellis et al how the original experiment counted collisions during sustained contact between the ring and path. Initially we only counted this type of collision once, but this favored the careless participants in the pilot study we ran. This was most apparent with difficult paths and participants quickly learned (through the audio/visual feedback) that simply sweeping their hand through segments would result in better scores. After several qualitative evaluations, we decided to add collisions for every 200 ms the ring stayed in contact with the path. This increased collisions overall, and penalized the careless participants while keeping the number of collisions manageable.

Ellis et al.'s experiment also showed a visible increase in collisions as system latency increased for the small ring. This effect was less observable and in fact appears to start to level off as if a ceiling effect was occurring in our experiment. We hypothesized that due to the lower clearance for the small ring, users moved their head extremely close to their hand to get a better view. This would occasionally cause small jitter issues, because PPT is a vision-based tracking system which depends on line of sight and is very sensitive to occlusion. Having this jitter issue and the rigorous collision criteria may have made the task too difficult when end-to-end latency was increased. This apparent ceiling effect (which was not observed in the original study) manifests itself in different degrees based on the path type, which is consistent with the observed new interaction between path type and latency.

Another potential cause of the difference in absolute tracing performance is the difference in perceived registration between Ellis' real AR system and our AR simulator. In Ellis' system, users saw their real hand and the virtual ring, with visual mis-registration of the hand and the ring due to latency. In our AR simulator, users saw a simulated real hand and virtual ring, also with visual mis-registration due to latency. However, in our simulator, there is also a second type of mis-registration. Participants saw a simulated real hand which appeared to lag behind the position/movement of their real hand (perceived via the proprioceptive sense) due to the unavoidable latency of the simulator itself. This second mis-registration could have added difficulty to the task, and we investigate this effect further in Experiment 2, as described in Section 4.

3.6 Issues in Replicating Prior Studies

During the course of this work, we learned some valuable lessons with regards to replicating and simulating previous experiments. One important reason for choosing the experiment in Ellis [3] was the very detailed description of the design and analysis the authors provided in the paper. In addition, Dr. Ellis was kind enough to be available for questions. Even with all this, understanding the original setup and design of the experiment was extremely challenging.

Table 1: Significant Effects for Experiment 1

Effect	df	F level	
Ring	1, 12	9.075	P < 0.011
Path	1, 12	25.638	P < 0.001
Latency	4, 48	14.245	P < 0.001
Path x Latency	4, 48	7.484	P < 0.001
Path x Ring x Latency	4, 48	3.348	P < 0.017

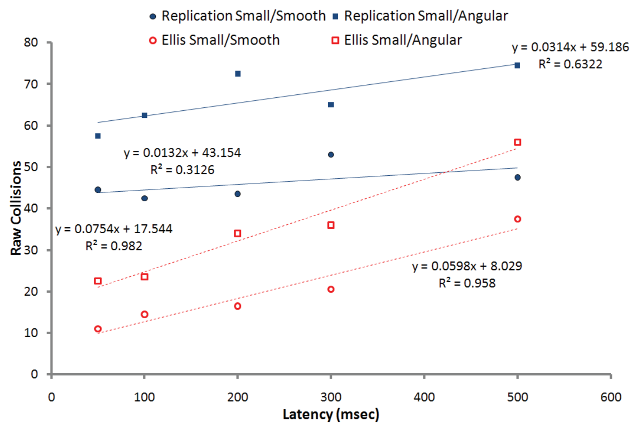

Figure 2: Comparison of trend lines for tracing performance of the **small** ring. Our replication study results are represented by the solid blue data points.

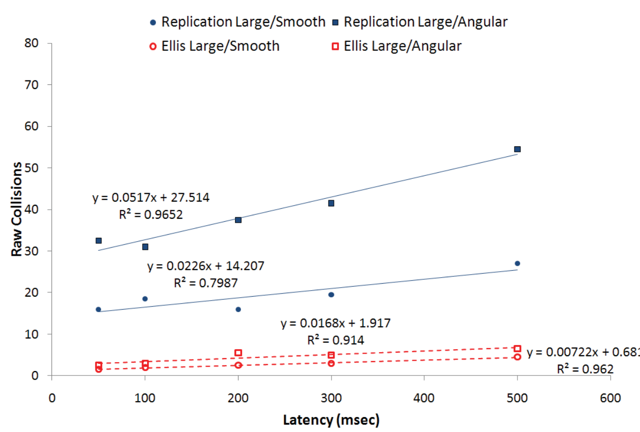

Figure 3: Comparison of trend lines for tracing performance of the **large** ring. Our replication study results are represented by the solid blue data points.

Since we did not have access to the original paths used by the original authors, we could not be sure if our models were absolutely correct. The models we generated based on the definitions in the original work had a fairly high variance in difficulty. After several attempts, we decided to manually remove the extremely hard paths since some of them were almost impossible to complete with a reasonable level of accuracy. Still, we are not certain that our paths were similar in difficulty to the ones used in Ellis' experiment. In general, replication would be facilitated considerably if the original experimenters provided public access to the models and environment used. Descriptions in publications, no matter how detailed, are never as useful as the original content. Having the original content would remove one level of uncertainty when comparing results. To practice what we preach, we are providing all data on the described experiments in this paper at this project's web page [7].

The second most demanding issue we faced was in determining the correct collision criteria. In the first iteration of Experiment 1, we made the decision to only count a sustained contact between the path and ring as a single collision. It quickly became apparent that this would greatly favor the careless participants. Since we were sampling for collisions at every frame, it was not feasible to count that particular type of contact as a collision every single frame because it became annoying to participants quickly due to the feedback mechanisms. Since it was not possible to get the actual code, we used a video recording of the original experiment and our own evaluations to determine a sampling frequency. Although this seemed the optimal solution for both precision in collisions and user response, it created an uncertainty in the raw data. In general, then, if the original experimenters provided public access to the source code used to implement the experiment, replication would be much simpler and less uncertain.

In summary, it is very challenging to replicate AR experiments given the current state of reporting experiments in publications. It will not always be possible to overcome differences in hardware, but if source code and 3D models were made publicly available, this would greatly ease the replication process. It would be extremely hard if not impossible to repeat any experiment without very detailed notes or the guidance of the original authors. This highlights the importance and need for even more detailed reports on experiments in the community. Although this may not be feasible within a conference or journal paper format, this information is invaluable to the repeatability of these experiments.

3.7 Summary

In Experiment 1, we replicated a prior study as closely as possible and obtained similar, although not identical, results. We discussed reasons for the differences between our results and Ellis [3], and we conclude that this study supports the validity of AR simulation but does not constitute a proof of validity. In general, we want to know the situations an experiment using an AR simulator can be trusted to provide valid results applicable to real AR systems. We want to know what characteristics of the simulator itself might have an unintended effect on results. For Experiment 1, there was an additional latency effect which was not present in Ellis' experiment: the latency applied to the simulated real world. What effect might this latency have had on the results of our experiments? We examined this in Experiment 2.

4 Experiment 2: Simulator Latency vs. Artificial Latency

To investigate this effect, we separated the end-to-end latency of Experiment 1 into two components: simulator latency and artificial latency. It is simulator latency which makes AR simulators inherently different from real AR systems. This subtle but important point must be well understood. A real see-through AR system would not exhibit any latency for the real world scenery. A video-see-though AR system would have a small but non-zero latency due to the video delay on the real world scenery. In an AR simulator all parts of the scene, including the simulated real world, are subject to the base latency of the simulator.

Table 2: Significant Effects for Original Ellis Experiment

Effect	df	F level	
Ring	1, 12	112.7	$P < 0.001$
Path	1, 12	46.2	$P < 0.001$
Latency	4, 48	31.8	$P < 0.001$
Path x Ring x Latency	4, 48	3.8	$P < 0.009$

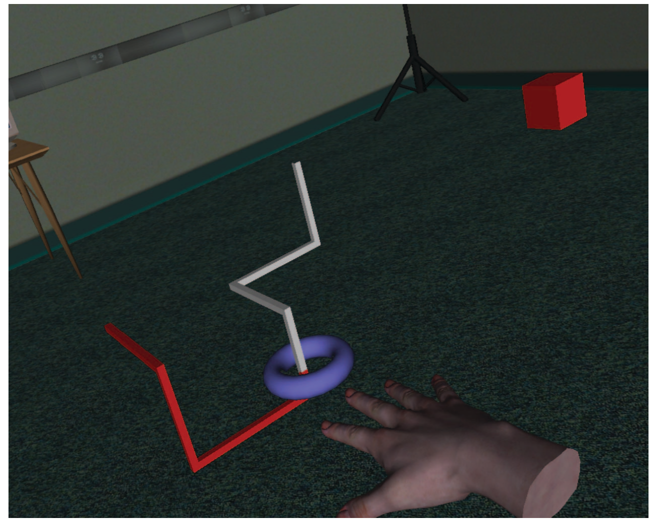

Figure 4: Screenshot of Experiment 2. As can be seen, the 3D path changes color as the ring completes each part.

We define this base latency as *simulator latency* and it consists of the tracker latency, compute time, render time, and display time. In our simulator, this amounted to 50 ms of latency. Since we wanted to see how this could have affected our results from Experiment 1, we needed to be able to vary this value to evaluate multiple simulator latencies. As shown in Figure 5 we achieved this by simply adding an amount of simulator delay to the base end-to-end latency of our simulator. All simulated real objects would then incur a delay equivalent to the new simulator latency sum. Increasing the simulator latency would cause the simulated real world and simulated real hand to lag and swim and also have an additive effect on the virtual objects.

Artificial latency is the latency difference between virtual objects and the real world. An AR simulator uses artificial latency to simulate the end-to-end latency of the real world AR system. The virtual objects would only incur a latency cost equivalent to the base end-to-end latency of that particular real-world AR system. In this experiment virtual objects incur a latency cost equivalent to the sum of the artificial latency (which is nonzero) and the simulator latency (which is minimally the base system latency). We hypothesized that simulator latency would have a smaller effect on the task in Experiment 1, because we felt that the visual mis-registration between the simulated real hand and the virtual ring (caused by artificial latency) was the main factor influencing performance.

4.1 Task and Environment

The overall task remained the same as Experiment 1, except for a few changes. Participants were asked to trace a 3D path with a virtual ring attached to their simulated real hand. Based on user feedback and our own observations from Experiment 1, we modified the software to provide more visual aids to the participants. Since our interest was in the effect of simulator latency, we only used the angular path since these were more interesting. They required the user to make sharper turns with the ring and would show the effects of latency much more. The size of the paths were kept the same from the first experiment, but the paths were subdivided into centimeter long segments and visually showed the user which segments had been completed by turning red (as seen in Figure 4). Participants had to start at a particular end and could not skip sections of the path during tracing. Only the large ring was used to minimize the jitter problem related to occlusion we had noticed earlier in our tracker. The ring was also colored purple to distinguish it from the gray path, since participants in Experiment 1 had noted some difficulty of visually detecting collisions when both objects were gray scale. The same lab model was also used to represent the real world environment. Lastly we removed the restriction on field of view and allowed the AR window to encompass the entire physical view such that the virtual objects were always perceivable within the HMD. For each task there was a different amount of simulator latency and artificial latency. The total end-to-end latency of the virtual objects was equivalent to the sum of the simulator latency and artificial latency. The total end-to-end latency of the simulated real objects was just the simulator latency.

4.2 Apparatus

The same hardware and software setup was used from Experiment 1, other than a newer InterSense Cube2 tracker replacing one of the cubes used in Experiment 1.

4.3 Study Design

This experiment was a within-subjects, repeated measures user study with two independent variables: simulator latency and artificial latency. Simulator latency was set at 50 ms, 100 ms, and 150 ms. Artificial latency was also set at 50 ms, 100 ms, 150 ms. The dependent variable was once again the number of collisions. With three different simulator latencies, three different artificial latencies, and three different angular paths there were a total of 27 different conditions. This created a block of 27 trials. Each participant performed the same block of trials three times (order of trials randomly generated each time).

4.4 Participants and Procedure

For this experiment we had 13 participants, ten males and three females who all received modest monetary compensation for their time participating in the study. The same questionnaire from Experiment 1 was given to these participants. The users' ages ranged from 19 to 36 years old. All users were able to perceive stereo as verified by a random dot stereo gram test. All users were comfortable around computers and had some experience with 3D games.

All participants were fitted with the HMD and hand device and a calibration step was performed. Each participant was first asked to practice on a selected path and with no artificial latency and no added simulator latency for five minutes. The study administrator guided the participant until the participant felt familiar with the task and was completing the training path in a consistent amount of time before beginning the actual trials. During the trials users were allowed to rest between trials and a mandatory rest was enforced after each block.

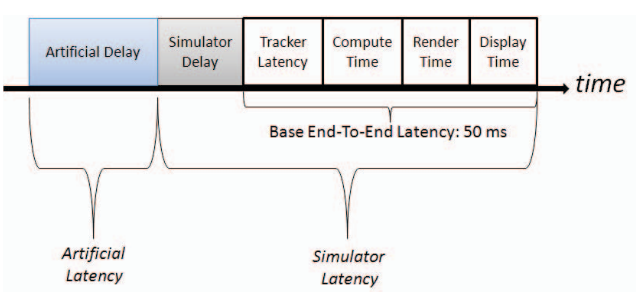

Figure 5: Simulator Latency and Artificial Latency in our simulator. Our simulator had a base latency of ≈ 50 ms (tracker latency + compute time + render time + display time). Additional simulator delay is added to this base latency to simulate a wider range of simulators. Artificial delay is added to create additional latency on virtual objects.

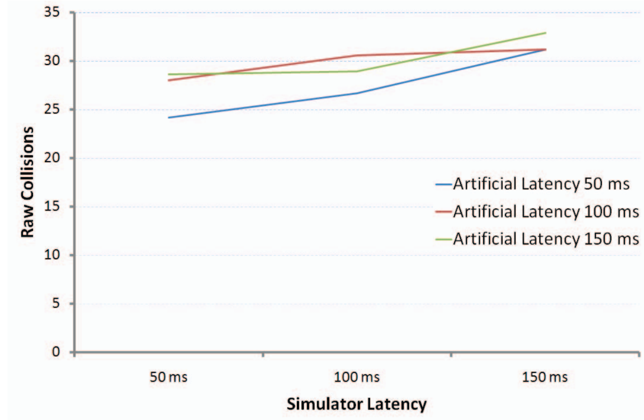

Figure 6: A plot of the effects of artificial latency and simulator latency on tracing performance. Artificial latency is represented by the three lines of 50 ms, 100 ms, and 150 ms. Simulator latency is represented as the X axis at 50 ms, 100 ms, and 150 ms.

4.5 Results

Preliminary analysis showed no significant differences between the three blocks of trials, so the number of collisions for each condition from the three blocks of trials were averaged, and the log of the average was used to get a single score and a multi-variate ANOVA was used to obtain statistically significant effects. The significant effects can be seen in Table 3. Both artificial latency and simulator latency were found to be statistically significant, however there was no interaction between the two effects.

The trends of the effects of simulator and artificial latency on the *raw number of collisions* are also plotted on Figure 6. As can be seen in the figure, the number of collisions increased when both artificial latency and simulator latency increased. For all values of artificial latency (50 ms, 100 ms, 150 ms), an increase in the simulator latency also resulted in an increase in the number of collisions.

A post-hoc analysis (Tukey multiple comparison of means with a 95 percent family-wise confidence level) was used to determine which pairs of conditions were significantly different. These results are shown in Table 4 along with the pairwise comparisons of the different simulator latencies and artificial latencies. The comparisons of artificial latency and simulator latency suggests that the effect of artificial latency is not significant between 100 ms and 150 ms and that the effect of simulator latency is not significant between 50 ms and 100 ms. It is harder to make any claims about the combined effect of simulator and artificial latency. The pairwise comparison of the combined artificial/simulator latency conditions shows that the only condition with any significant difference to other conditions was when both latencies were at 50 ms. This makes sense since this was the lowest latency (both artificial and simulator) condition and was the easiest condition for this task.

Our hypothesis that simulator latency would not have as big an effect on this task as artificial latency was shown to be false. There

Table 3: Significant Effects for Experiment 2

Effect	df	F level	P Value
Artificial Latency	2, 342	4.1659	$P < 0.017$
Simulator Latency	2, 342	10.8073	$P < 0.001$

was no significant interaction between artificial latency and simulator latency, suggesting that the effects of simulator latency and artificial latency may be additive for this particular task. Based on this we are confident that the results from Experiment 1 are still valid despite the effects of simulator latency. The effects of our simulator latency may have contributed to the absolute level of performance, but the task is still dependent on artificial latency. What this also tells us is that it is not sufficient to just simulate correct registration between the simulated real world scenery and the virtual scenery in an AR simulator. The mis-registration between the actual real world and the simulated real world is significant for this task and must be considered for other tasks in AR.

5 Discussion of AR Simulation

From Experiment 1 we know an experiment performed in an AR simulator can produce comparable results to the same experiment on a real AR system, if the real AR system is carefully simulated. This is a promising result and is a step toward the validation of AR simulation in general. Although the inherent simulator latency may have affected the results of the experiment, it did not interact with the effect of artificial latency in Experiment 2. This finding strongly suggests that the results from Experiment 1 are still valid despite the effect of simulator latency. Simulator latency was an additive effect and did indeed increase collisions, but it did not change the effect of artificial latency. Experiment 2 also showed that AR simulation with non-zero simulator latency is not equivalent to actual AR systems. The differences between the simulator and actual AR system can have significant effects on experiments run on an AR simulator. We found that simulator latency is not inconsequential in determining task performance in an AR simulator, and that just simulating correct visual registration is not sufficient. This does not simulate the overall perception of registration errors.

We have only looked at a single task at this point, so we do not know whether simulator latency will still have significant effects on performance of other tasks. We hypothesize that simulator latency will not affects tasks that rely on visual registration, such as identifying the real object to which a virtual label is attached. We do not know the effects of other differences between AR simulators and actual AR systems. Thus we continue to perform studies on this type, so that we learn about the effects of different simulator characteristics and therefore can use AR simulation in appropriate situations; namely in situations where it has high validity and provides large benefits.

Table 4: Tukey Post Hoc Analysis for Experiment 2

Pairwise Comparison	P Adj.
Artificial 150 ms and Simulator 150 ms vs Artificial 50 ms and Simulator 50 ms	$P < 0.001$
Artificial 100 ms and Simulator 150 ms vs Artificial 50 ms and Simulator 50 ms	$P < 0.003$
Artificial 50 ms and Simulator 150 ms vs Artificial 50 ms and Simulator 50 ms	$P < 0.010$
Artificial 100 ms and Simulator 100 ms vs Artificial 50 ms and Simulator 50 ms	$P < 0.016$
Simulator 150 vs Simulator 50	$P < 0.001$
Simulator 150 vs Simulator 100	$P < 0.001$
Artificial 100 vs Artificial 50	$P < 0.048$
Artificial 150 vs Artificial 50	$P < 0.035$

6 Conclusions and Future Work

In this paper we replicated a prior study in Experiment 1 with comparable results. We found the same statistically significant effects and performance trends were in the same direction. We hypothesize that the differences were due to tracker differences, collision algorithm differences, and simulator latency. In Experiment 2 we investigated the effects of simulator latency on simulated-AR based experiments. Our findings suggest that simulator latency is an additive effect on artificial latency within the context of this task. Even though simulator latency was a significant effect, there was no interaction between simulator latency and artificial latency. This result reinforces our belief that the results from Experiment 1 are indeed valid. We have also included a discussion on the need for detailed and publicly available experiments for replication and a discussion on the implications of this work on AR simulation in general.

This work is part of a larger goal to create a simulation framework capable of simulating experiments from the entire range of the mixed reality continuum with fairly high fidelity. For the long term goal we would like to use the AlloSphere, a high-fidelity virtual environment and computing system, to conduct controlled studies investigating levels of immersion components in AR. In the short term we are planning on continuing to investigate other components which could make AR simulation valid.

Future projects include investigating the effects of simulator latency on different tasks (visual search, etc). Another project looks at different potentially problematic aspects of an AR simulator, such as accommodation cues. Optical see-through AR provides correct accommodation cues for the real world but not for the virtual objects. A simulator would have incorrect cues for both simulated real world and virtual objects. Due to this, an AR simulator could potentially produce different results.

Acknowledgements

This research was partially supported by ONR grant N00014-09-1-1113, and NSF CAREER grant IIS-0747520. We would like to thank Dr. Stephen R. Ellis for his time and invaluable guidance in replicating the original study. We would also like to acknowledge Masaki Miyanohara and the ReCVEB lab at UCSB for providing the virtual model of their lab.

References

[1] D. A. Bowman and R. P. McMahan. Virtual reality: How much immersion is enough? *Computer*, 40(7):36–43, 2007.

[2] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The cave: audio visual experience automatic virtual environment. *Commun. ACM*, 35(6):64–72, 1992.

[3] S. R. Ellis, F. Breant, B. Manges, R. Jacoby, and B. D. Adelstein. Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. In *VRAIS '97: Proceedings of the 1997 Virtual Reality Annual International Symposium (VRAIS '97)*, page 138, Washington, DC, USA, 1997. IEEE Computer Society.

[4] J. L. Gabbard, J. E. Swan, II, and D. Hix. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence: Teleoper. Virtual Environ.*, 15(1):16–32, 2006.

[5] S. Kim and A. K. Dey. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *CHI*

'09: Proceedings of the 27th international conference on Human factors in computing systems, pages 133–142, New York, NY, USA, 2009. ACM.

[6] C. Lee, S. Bonebrake, T. Höllerer, and D. A. Bowman. A replication study testing the validity of ar simulation in vr for controlled experiments. *Proceedings of International Symposium on Mixed and Augmented Reality*, 7, 2009.

[7] http://cs.ucsb.edu/ chalee21/Projects/VR2010/.

[8] R. P. McMahan, D. Gorton, J. Gresock, W. McConnell, and D. A. Bowman. Separating the effects of level of immersion and 3d interaction techniques. In *VRST '06: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 108–111, New York, NY, USA, 2006. ACM.

[9] R. Pausch, D. Proffitt, and G. Williams. Quantifying immersion in virtual reality. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 13–18, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[10] E. Ragan, C. Wilkes, D. A. Bowman, and T. Höllerer. Simulation of augmented reality systems in purely virtual environments. *Virtual Reality Conference, IEEE*, 0:287–288, 2009.

[11] G. Robertson, M. Czerwinski, and M. van Dantzich. Immersion in desktop virtual reality. In *UIST '97: Proceedings of the 10th annual ACM symposium on User interface software and technology*, pages 11–19, New York, NY, USA, 1997. ACM.

[12] D. W. Sprague, B. A. Po, and K. S. Booth. The importance of accurate VR head registration on skilled motor performance. In *GI '06: Proceedings of Graphics Interface 2006*, pages 131–137, Toronto, Ont., Canada, Canada, 2006. Canadian Information Processing Society.

[13] J. Ventura, M. Jang, T. Crain, T. Höllerer, and D. Bowman. Evaluating the effects of tracker reliability and field of view on a target following task in augmented reality. In *Virtual Reality Software and Technology, 2009. 16th ACM Symposium on*.

[14] J. Wither, S. DiVerdi, and T. Hollerer. Evaluating display types for ar selection and annotation. In *ISMAR '07: Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–4, Washington, DC, USA, 2007. IEEE Computer Society.