Evaluating effectiveness in virtual environments with MR simulation

Doug A. Bowman, Cheryl Stinson, Eric D. Ragan, Siroberto Scerbo	Tobias Höllerer, Cha Lee	Ryan P. McMahan	Regis Kopper
Virginia Tech	University of California, Santa Barbara	University of Texas at Dallas	University of Florida
Blacksburg, Virginia	Santa Barbara, California	Dallas, Texas	Gainesville, Florida
[bowman, cstinson, eragan, scerbo]@vt.edu	[holl, chalee21]@cs.ucsb.edu	rymcmaha@gmail.com	kopper@cise.ufl.edu

ABSTRACT

Both virtual reality (VR) and augmented reality (AR) systems have achieved some success and offer further potential to be used in military training. However, the use of high-end VR and AR remains costly and cumbersome, and the most advanced technologies are not widely deployed in actual training systems. Decision makers need evidence for the effectiveness of such systems in order to justify their use. In particular, it is important to know which display systems (e.g., head-mounted display, CAVE) will provide the best cost/benefit ratio for training, and what display characteristics (e.g., field of view, stereoscopy) are most critical in determining the effectiveness of a VR or AR training system.

The answers to these questions depend on an understanding of the effects of display parameters on task performance and training transfer. Obtaining this knowledge requires empirical studies, but such studies pose significant challenges. Direct comparisons of different displays do not produce generalizable results because the displays differ in many ways. AR studies face the additional issues of unreliable hardware that lacks desirable features and a lack of control of the real-world environment.

Our research addressing these issues is based on two key insights. First, systematically studying the effects of display fidelity using a display simulator, rather than studying actual display technologies, results in more useful and general knowledge. Second, a single simulator, based on a high-end VR system, can be used for displays spanning the mixed reality (MR) continuum, including both VR and AR. In this paper, we discuss the concept of MR simulation, an innovative evaluation methodology that allows for controlled experiments and allows the evaluation of individual components of display fidelity rather than whole systems. We describe our work to validate this methodology and illustrate the use of MR simulation through a number of example experiments.

ABOUT THE AUTHORS

Doug A. Bowman is Professor of Computer Science and Director of the Center for Human-Computer Interaction at Virginia Tech. He received a Ph.D. in Computer Science from the Georgia Institute of Technology.

Cheryl Stinson, **Eric D. Ragan**, and **Siroberto Scerbo** are Ph.D. students in Computer Science at Virginia Tech.

Tobias Höllerer is Professor of Computer Science at the University of California, Santa Barbara. He received a Ph.D. in Computer Science from Columbia University.

Cha Lee is a Ph.D. student in Computer Science at the University of California, Santa Barbara.

Ryan P. McMahan is Assistant Professor of Computer Science at the University of Texas at Dallas. He received a PhD. in Computer Science from Virginia Tech.

Regis Kopper is a post-doctoral fellow in Computer and Information Science and Engineering at the University of Florida. He received a PhD. in Computer Science from Virginia Tech.

Evaluating effectiveness in virtual environments with MR simulation

Doug A. Bowman, Cheryl Stinson, Eric D. Ragan, Siroberto Scerbo	Tobias Höllerer, Cha Lee	Ryan P. McMahan	Regis Kopper
Virginia Tech	University of California, Santa Barbara	University of Texas at Dallas	University of Florida
Blacksburg, Virginia	Santa Barbara, California	Dallas, Texas	Gainesville, Florida
[bowman, cstinson, eragan, scerbo]@vt.edu	[holl, chalee21]@cs.ucsb.edu	rymcmaha@gmail.com	kopper@cise.ufl.edu

INTRODUCTION

Virtual reality (VR) and augmented reality (AR) offer unique experiences to their users. In VR, users are placed into a computer-generated 3D world that can be viewed and navigated in real time (Bowman, Kruijff, LaViola, & Poupyrev, 2005). With high-end VR displays, such as CAVEs and head-mounted displays (HMDs), virtual objects can appear to exist in real 3D space, and the virtual world can appear to surround the user physically. In AR, virtual objects and information are overlaid onto the user's view of the real world (Azuma, 1997), and in the most advanced AR systems (e.g., see-through head-worn displays), these augmentations can appear to become part of the real world.

Both VR and AR systems have achieved some success and offer further potential to be used in military training (Cohn, Schmorrow, Nicholson, Templeman, & Muller, 2003; Office of Technology Assessment, 1994). VR technologies allow trainees to enter a realistic 3D world under full control of the trainers, and can be used for weapons training, tactical training, team communication training, and spatial navigation training, among others. AR technologies can place the trainee in a real-world setting that also includes virtual objects, entities, and/or annotations, providing even higher levels of realism and face-to-face communication with other trainees or trainers.

Despite their success, the use of high-end VR and AR remains costly and cumbersome, and the most advanced technologies are still not widely deployed in actual military training systems. This leads to a number of questions of great practical importance to decision makers:

- For a particular application, will the use of VR or AR be effective?
- When should purely virtual environments be used, and when do augmented physical environments have a greater benefit?
- What VR or AR systems should be used for specific application scenarios? For example, is a desktop game engine sufficient, or should a high-resolution (HMD) be used?
- What display characteristics are most critical in determining the success of a particular application? For example, is it more important to have a wide field of view or to have stereoscopic graphics?

Being able to answer these practical questions requires a systematic understanding of the effects of display parameters on user task performance and training transfer. Without knowledge of the effects of the *fidelity* of VR and AR displays (which we have also called *immersion*; Bowman & McMahan, 2007), researchers will not be able to design new displays and applications to improve training effectiveness. Unfortunately, complete systematic knowledge does not yet exist, so developers are forced to guess at the answers to the questions above.

Clearly, obtaining such systematic knowledge of the effects of display parameters requires empirical studies. But such studies also pose significant challenges. Direct comparisons of different displays do not produce generalizable results because the displays differ in many ways. For example, a comparison of task performance with a CAVE and a stereoscopic monitor (e.g., Gruchalla, 2004) may tell us that users perform tasks more quickly in the CAVE, but it cannot tell us why this occurred (field of view? screen size? head tracking?), nor can it tell us what would happen if we used only a single large projection screen. AR studies face the additional issues of unreliable hardware that lacks desirable features (e.g., the real world cannot occlude virtual objects) and a lack of control of the real-world environment (e.g., weather and lighting).

Our research addressing these issues is based on two key insights. First, systematically studying the effects of display fidelity using a display **simulator**, rather than studying actual display technologies, results in more useful and general knowledge. Second, a single simulator, based on a high-end VR system, can be used for displays spanning the mixed reality (MR) continuum (Milgram & Kishino, 1994), including both VR and AR. Figure 1 illustrates this concept.

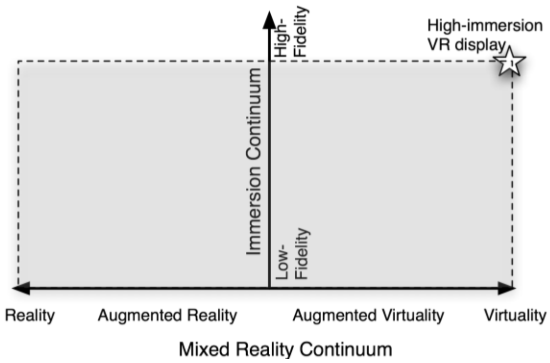

Figure 1. An MR simulator based on a single high-end VR display (upper right) can be used to simulate displays with lower levels of immersion and at different points on the MR continuum (indicated by the shaded box)

BACKGROUND AND RELATED WORK

Before discussing MR simulation and how it can be used to study the effects of various MR system characteristics, we present background information on the concept of fidelity, and discuss current limitations of empirical evaluations of MR systems.

Fidelity and Immersion

Even practitioners familiar with VR are often confused by, or interchangeably use, the terms *immersion* and *presence*. We adopt Slater's definitions (Slater, 2003):

- *Immersion* refers to the objective level of sensory fidelity a VR system provides.
- *Presence* refers to a user's subjective psychological response to a VR system.

Using this definition, a VR system's level of immersion depends only on the system's rendering software and display technology (including all types of sensory displays—visual, auditory, haptic, etc.). To avoid confusion, however, we will substitute the term *display fidelity* for immersion in this paper. Display fidelity is objective and measurable; one system can have a higher level of display fidelity than another. Presence, on the other hand, is an individual and context-dependent user response, related to the experience of "being there."

Display fidelity is not a binary value (although one often hears of "immersive" and "non-immersive" systems). Rather, display fidelity is a continuum – every system has some level of display fidelity, and the highest possible level of display fidelity would be indistinguishable from the real world.

Display fidelity is also a multi-faceted construct. For example, the level of visual display fidelity has many components, including field of view (FOV), field of regard (FOR), display size, resolution, stereoscopy, latency, brightness and contrast, frame rate, and refresh rate. For more detail on the theoretical aspects of display fidelity, see (Bowman & McMahan, 2007).

Different components of display fidelity are important for different training tasks, and our work aims to obtain a set of general results that are tied to the task that was studied in the experiment, not to a particular technology. For instance, if we find empirically that FOR is more important than FOV for a particular task, customers can use that information to choose an HMD (high FOR, low FOV) for that task over a CAVE (medium FOR, medium FOV).

We also extend the concept of display fidelity to apply to other points on the MR continuum. In the context of AR, we can talk about not only the level of display fidelity for the virtual parts of the scene, but also for the real parts of the scene (e.g., the AR display may limit the user's FOV into the real world), and for the relationship between the two (e.g., the registration of the virtual objects to the real scene).

We identify other types of fidelity for MR systems as well. *Interaction fidelity* refers to the degree to which user actions in a system match their real-world counterparts, while *simulation fidelity* refers to the level of realism of the models (e.g., geometric, lighting, or physical) used to produce the virtual environment. We focus mainly on display fidelity in this paper, but touch briefly on interaction fidelity as well.

Effects of Display Fidelity

The level of display fidelity is known to have effects on task performance, user preference, psychological and physiological reaction, and learning in VR (Arthur, 2000; Barfield, Hendrix, & Bystrom, 1997; Robertson, Czerwinski, & Dantzich, 1997; Schulze, Forsberg, Kleppe, Zeleznik, & Laidlaw, 2005; Tan, Gergle, Scupelli, & Pausch, 2003; Ware & Mitchell, 2005). But these results are widely scattered in the literature and may partially depend on the particular display technologies used in the experiments. We are far from a complete understanding of the effects of display fidelity in VR, and know very little about the effects of display fidelity in AR.

We have been involved in a large number of studies in recent years that evaluated the effects of level of display fidelity in VR using the simulator approach we describe in this paper (see (Bowman & McMahan, 2007) for a summary of this research and the section below titled "Example MR Simulator Experiments").

Problems with Empirical Studies of MR Systems

Evaluating MR systems with controlled experiments is difficult, and many challenges must be overcome to obtain the desired results. As we noted above, the primary issue is that direct comparisons of different MR systems (e.g., CAVE vs. HMD) do not produce generalizable results because of unavoidable confounds.

From a practical standpoint, such studies are limited to systems that are available to the researchers. Someone interested in the effects of FOV in HMDs is not likely to have several HMDs with different FOVs in her lab, and even if she does, they are likely to differ in other ways (e.g., resolution, weight, brightness). Moreover, such studies are limited to systems that are currently available; proposed future systems cannot be tested.

A problem specific to AR experiments is that we cannot study perfect registration (where virtual objects are perfectly aligned with the real world), because no current tracking systems provide perfect, timely data. This makes any study of the effects of registration error limited, in that "zero error" cannot be one of the conditions. This reveals the inherent impracticalities of attempting to understand problems through the use of a system that is limited by those very problems. Furthermore, when using AR systems, it is not generally possible to isolate different types of errors in order to test their independent effects on a task.

Finally, with respect to outdoor AR systems, it is very challenging to run meaningful generalizable studies outdoors, where quite a few environment parameters (weather, lighting, people's behavior) are beyond the experimenter's control (Livingston et al., 2003; Wither & Höllerer, 2005).

MR SIMULATION

MR simulation can be used to address these limitations. In this section, we describe the implementation of MR simulators and discuss their benefits and limitations.

Implementing MR Simulation

In order to achieve our goal of running controlled experiments on the effects of various components of fidelity, we need an experimental platform (hardware and software) that provides the required level of control. Using actual AR and VR systems would provide a high level of ecological validity (i.e., the results would have direct real world significance), but would not provide good experimental control, since actual AR and VR systems differ in many ways. We instead use high-end VR hardware, and a software framework that allows us to control components of fidelity independently, in order to simulate AR and VR systems. The simulator can display both virtual imagery and "simulated real world" imagery in the case of simulated AR.

The major design issues for the MR simulator are related to the components of fidelity (Bowman & McMahan, 2007) that a simulator user will control to simulate various MR system configurations. We control the components of fidelity separately for the simulated real imagery and for the virtual imagery, so that in mixed reality contexts we can control the relative level of fidelity between the real and virtual parts of the scene. Controllable components for the virtual and simulated real imagery include FOV, FOR, stereoscopy, head-based rendering, resolution, translational/rotational accuracy, latency, jitter, frame rate, and realism of lighting.

Given this design, many interesting conditions can be evaluated. For example, an important issue in AR is visual registration: virtual augmentations do not always appear to be attached to the proper real-world location. We can simulate different levels of registration accuracy by manipulating the translational/rotational accuracy, latency, and jitter components, with lower fidelity levels of these components for the virtual imagery than for the real imagery. The MR simulator can also be used to simulate different actual displays. In the realm of VR, the simulator can be configured to represent, for instance, an HMD (limited FOV but full FOR), a three-wall CAVE (limited FOR but wide FOV), or even a multi-monitor desktop display (non-stereo, several spatially arranged "tiles"). For AR, we can simulate head-worn displays, projected AR, and even handheld displays.

This design, of course, relies on the use of a high-end VR system as the simulator platform. The fidelity characteristics of this VR system determine the maximum level of fidelity that can be achieved by the simulator. In our work, we have primarily made use of two high-end VR systems as simulator hardware, and have planned to use a third system. First, we have used an NVis SX111 HMD (Figure 2), which offers 1280x1024 pixels per eye and a FOV of 102º by 64º. Second, we have used the Duke Immersive Virtual Environment (DiVE) at Duke University (Figure 3).

The DiVE is a six-sided CAVE-like system that offers a full 360º FOR and a resolution of 1050x1050 pixels on each screen, with active stereoscopic graphics and wireless head and wand tracking.

Figure 2. Current MR Simulator platform: NVIS SX111 HMD

Figure 3. Current MR Simulator platforms: Duke Immersive Virtual Environment

Finally, when it is fully operational, we plan to use the UCSB AlloSphere facility. The AlloSphere (Höllerer, Amatriain, & Kuchera-Morin, 2007) consists of a completely surrounding spherical projection screen, approximately 33 feet in diameter, onto which high-resolution projectors can cast a seamless environment map surrounding the user. With a large sweet spot for stereo projection and high-resolution spatial audio rendering through an array of two-way high-gain speakers, the experience turns into a virtual reality of extremely high fidelity and sensory precision.

Benefits of MR Simulation

As we have noted, the most important benefit of the simulator approach is the level of experimental control it provides to the researcher, allowing independent variation of a large number of parameters. This control gives the researcher the flexibility to simulate actual systems or envisioned systems for applied experiments, or to simulate all the different permutations of a set of components for more controlled studies. This latter form of study will provide general results and increase the overall understanding of the effects of fidelity.

The simulator approach also solves the specific problems discussed above when running experiments comparing specific MR systems. For VR, a simulator running in a high-end surround-screen system could allow evaluation of currently unavailable technologies, such as seamless ultra-wide FOV HMDs. The effectiveness of new system designs can be tested without expensive implementations or additional devices.

The concept of using VR to simulate a complete AR system clearly has several advantages over an actual AR environment. For instance, as mentioned, such an arrangement makes it possible to precisely control the registration of virtual objects, allowing testing of exact levels of registration error. Such an approach even enables the ability to test results of “perfect” registration, which is impossible when using real AR systems (we acknowledge that VR systems also suffer from registration error; see the next section for discussion). The complete registration control also makes it possible to isolate and independently manipulate different types of registration error (e.g., jitter, latency, drift), allowing studies of interactions among the types of error, which actual AR technology does not allow. Simulation can also facilitate the manipulation of other factors of the augmented display, such as field of view or image resolution.

Outdoor AR research would benefit immensely from our simulator approach, since it provides control over factors such as weather, lighting, and people in the scene. As an additional advantage, complete control over what happens in the simulated real environment makes it possible to test a system in a wide variety of use scenarios, including those that might be too difficult, dangerous, or costly to produce in the real world (e.g., AR support for firefighters).

Limitations of MR Simulation

MR simulation also has some limitations. The primary limitation is that the choice of the simulator platform limits the types of systems and levels of fidelity that can be tested; systems with fidelity higher than that of the simulator cannot be evaluated. For example, a six-sided CAVE cannot be simulated with a four-sided CAVE, and a simulation of an outdoor AR system will be limited by the lack of available luminance of the VR display.

Another disadvantage is that the simulation approach does not allow users to physically walk large distances due to size limitations of VR platforms. This issue may require additional consideration if the test system simulates a physically large area and virtual travel techniques might interfere with the investigation.

AR simulation is limited by the fidelity of the real world component in the system. One issue, for example, is the lack of tactile feedback in the simulated real environment. This may not be problematic, however, if the simulation does not require or allow interactions with the simulated physical objects.

Another issue for simulated AR is the tracking error within the VR system, which means that the registration of the simulated real environment cannot be perfect. In modern VR systems, however, the perceived error will be low, and may even be unnoticeable. Although the trackers in any VR system will introduce some degree of latency and jitter, such error usually has low impact because all objects (both virtual and simulated real objects) are affected equally. By contrast, in AR, only the virtual objects exhibit error, resulting in a mismatch between the real and virtual parts of the scene.

VR also presents different depth cues than those experienced in the real world of AR. Even though stereoscopic imagery can offer convergence cues, the current methods used to display virtual objects cannot enable the use of ocular accommodation cues because the objects are always in focus at the depth of the projection screen. Because all objects in VR are virtual, they all provide the same imperfect visual depth cues. In an optical see-through AR environment, on the other hand, while the virtual components suffer from the same types of imperfect cues, the real world objects will provide perfect depth cues. As a result, the distinction between real and virtual objects in a simulated AR environment will differ from the corresponding disparity in an actual AR system.

Though an MR simulation does not provide a perfect representation of an actual MR system, the simulation approach can still provide great benefit to MR research. Additionally, as technological advancements further the realism of virtual reality systems and reduce these limitations, the quality of the simulations will also improve. Finally, many issues with simulation can be mitigated through experimental design.

VALIDITY OF MR SIMULATION

Are the results of experiments using MR simulation valid? Do we obtain the same results as we would with real-world MR systems? To validate MR simulation, we must first analytically compare the level of fidelity of our final simulator to real-world systems to ensure that the simulator can reproduce the fidelity of these systems. We then 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.

AR Replication Study

The goal for our first validation experiment was to replicate an established AR study within our simulator as a step toward validation of AR simulation. Details can be found in (Lee, Bonebrake, Höllerer, & Bowman, 2010). We chose to replicate the second experiment in Ellis et al. (Ellis, Breant, Manges, Jacoby, & Adelstein, 1997), 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.

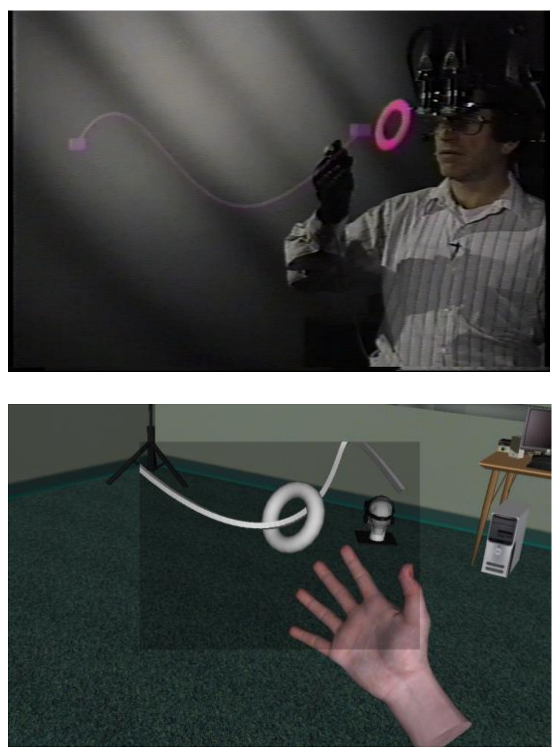

Figure 4. Ellis' original AR latency study (top); replication study run in the MR simulator (bottom)

Figure 4 shows the experimental setup used in Ellis' work and the user's view of our simulation through the HMD. We simulated the real AR system by providing two different FOVs (one for the simulated real world and another for the virtual objects), by always rendering virtual content on top of simulated real content regardless of depth, and by adding different amounts of artificial latency to the tracking data to match Ellis' different latency conditions. Despite the work we did to replicate the AR experiment carefully in simulation, there were still differences. The simulated real world was not photorealistic, and our tracker had more jitter in certain conditions. The most important difference, however, was that there was a mismatch between the proprioceptive and visual systems when the user moved his hand. In other words, the user felt his hand moving at one time but saw the virtual hand moving later. This mismatch occurred because our simulator had its own base level of latency.

Our study had similar results to Ellis' experiment. Interested readers can find full details in (Lee, Bonebrake, Höllerer, & Bowman, 2010). We found all of the same significant effects of latency and ring size. However, in absolute terms, performance in our study was worse than in the original experiment. This led us to hypothesize that the simulator's base latency made the task more difficult, so we studied this effect in our next experiment.

Effects of Simulator Latency

To investigate this effect, we ran a second experiment (Lee et al., 2010), in which we separated the end-to-end latency of our first experiment into two components: simulator latency (the unavoidable base latency of the simulator system) and artificial latency (intentionally added latency used to simulate different MR systems). Since we wanted to see how simulator latency could have affected our results in the replication study, we needed to be able to vary this value to evaluate multiple simulator latencies. 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 total simulator latency. Increasing the simulator latency would cause the simulated real world and simulated real hand to lag and would also have an additive effect on the lag of virtual objects.

The task and the levels of artificial latency were the same as those in the replication study. We found that both artificial latency and simulator latency had significant effects on performance. However, we did not find an interaction effect for the two variables, indicating that the effects of artificial and simulator latency are additive. This implies that studies of latency in MR simulators can be valid, in the sense that they will properly demonstrate the effects of artificial latency, despite the fact that performance may be significantly worse overall due to the effects of simulator latency. Full details can be found in (Lee et al., 2010).

We hypothesized that simulator latency might have no effect whatsoever for other tasks. For example, in an AR visual search task, the registration of the virtual content to the real world seems to be the most important factor, and this registration would not be affected by simulator latency. To test this hypothesis, we ran a third experiment comparing task performance using a real AR system to performance in a range of MR simulators with different levels of simulator latency (Lee, Gauglitz, Höllerer, & Bowman, 2012).

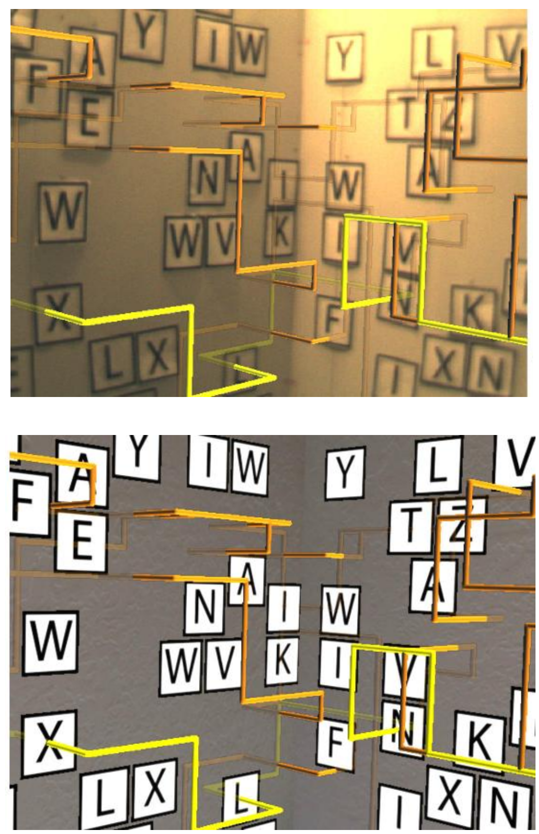

Figure 5. Environment for second simulator latency study: real AR condition (top); simulated AR condition (bottom)

Figure 5 shows the experimental environment and task in both the real AR and simulator conditions. Participants had to follow a virtual pipe as it moved through a room and in and out of the room's walls (xray vision allowed users to see the pipes behind the walls). Each intersection of the pipe with the wall went through a paper card with a letter printed on it, and participants had to call out the sequence of letters as they followed the pipe from beginning to end. The real AR system, based on a video see-through HMD, had a base latency of 48ms, which we simulated using artificial latency in our MR simulator conditions. The simulators had additional simulator latency of 0, 50, and 150ms. We measured the time to perform the task without errors.

The results of this study (details in (Lee, Gauglitz, Höllerer, & Bowman, 2012)) showed that the MR simulator conditions were not significantly different in performance from the real AR condition, and could be considered statistically equivalent based on a threshold of one standard deviation on either side of the mean performance of the real AR condition. However, the real AR condition did have the worst performance in absolute terms. We conclude that simulator latency does not have a significant effect on performance in visual path following, and that it is likely that results obtained from the simulator are equivalent to those obtained with the real AR system. This is evidence for the validity of the MR simulation approach.

Future Validation Studies

We are currently planning two additional studies investigating the validity of MR simulation. First, we will investigate the possible effects of visual realism (simulation fidelity) on results in MR simulator experiments. As we noted in the replication study, the simulated real world used in MR simulators may not be visually realistic in terms of the quality of the model, textures, lighting, and shadows. Even if the simulated real world does not play a major role in the experimental task, could this difference in sensory stimuli have an influence on the results? To address this question, we are building three virtual models of a real-world location, using different levels of visual realism. We will ask users to perform a task in a simulated AR system using these models as the simulated real world, and compare those results to one another and to results obtained with a real AR system in the physical world.

The second planned study will examine the claim that we can simulate various MR displays using a single MR simulator platform. We are developing a visual search task in a cluttered virtual environment, and will ask users to perform this task in a four-wall CAVE display and a simulated four-wall CAVE displayed in a high-end HMD. This study will help us understand how far we can take the display simulation idea, even when there are obvious differences between the simulator platform and the simulated display (e.g., ergonomics, accommodation distance, quality of stereoscopy).

EXAMPLE MR SIMULATOR EXPERIMENTS

We conclude by describing a few of the experiments we have run so far using the MR simulator approach.

Procedure Learning Experiment

Researchers have proposed that display fidelity could have advantages for tasks involving abstract mental activities, such as conceptual learning; however, there are few empirical results that support this idea. We hypothesized that higher levels of display fidelity would benefit such tasks if the mental activity can be mapped to objects or locations in a 3D environment. To investigate this hypothesis, we performed an experiment in which participants memorized procedures in a virtual environment and then attempted to recall those procedures. See (Ragan, Sowndararajan, Kopper, & Bowman, 2010) for complete details.

We aimed to understand the effects of three components of display fidelity (FOV, FOR, and software FOV—the FOV of the virtual camera) on learning. To study these components independently, all conditions used an MR simulator running in a four-wall CAVE. FOV was varied by using “blinders” attached to clear lab glasses; FOR was varied by using either one screen or all four screens; and software FOV was varied by modifying the viewing parameters in software—we tested software FOVs that were matched and unmatched to the physical FOV of the display.

Users were asked to watch a procedure that was presented in a virtual environment, rehearse that procedure verbally with help from the experimenter, and then demonstrate their learning of the procedure by verbally stating its steps without help.

Results demonstrated that a matched software FOV, a higher FOV, and a higher FOR all contributed to more effective memorization. The best performance was achieved with a matched SFOV and either a high FOV or a high FOR, or both. In addition, our experiment demonstrated that memorization in a virtual environment could be transferred to the real world. The results suggest that, for procedure memorization tasks, increasing the level of display fidelity even to moderate levels, such as those found in HMDs and display walls, can improve performance significantly compared to lower levels of display fidelity. Complete results can be found in (Ragan, Sowndararajan, Kopper, & Bowman, 2010).

First-Person Shooter Studies

Another set of MR simulator studies focused on the combined effects of display fidelity and interaction fidelity (recall that this describes the similarity of interaction techniques to real-world actions) for the popular "first-person shooter" (FPS) style of games. We chose this style because of its demanding interaction requirements, variety of user tasks (including travel, maneuvering, visual search, aiming, and firing), and relevance to serious applications such as military training. The studies used the Duke University DiVE system described above.

In the first study (McMahan, Bowman, Zielinksi, & Brady, 2012), we wanted to explore the general effects of interaction fidelity and display fidelity, and find out whether one influenced the other. Thus, we designed two levels of each variable, representing "low" and "high" fidelity. The combination of the two low-fidelity conditions was similar to a typical home gaming setup, while the combination of the two high-fidelity conditions represented a highly immersive VR setup. The other two conditions were mixtures of these.

The low interaction fidelity condition used a typical mouse and keyboard interface for FPS games, with the mouse being used to turn, aim, and fire, and the keyboard to travel through the virtual world. The high interaction fidelity condition (the "natural" interface) used a tracked handheld controller for direct aiming and firing, and a technique called the "human joystick" for travel. In the human joystick technique, the user would stand in the center of the DiVE (the mat visible on the floor in figure 6), and then physically step in the direction she wanted to travel, with movement starting once she stepped outside a small circular area, and the speed of movement proportional to the distance from the center. Although this technique is not highly natural, it has higher interaction fidelity than the mouse and keyboard technique due to its use of physical leg movements with direction mapped directly to the environment. More natural travel techniques were not practical in the limited space of the DiVE.

The low display fidelity condition used a single screen of the DiVE without stereo. It therefore also required a method for rotating the view, so we provided a technique that turned the viewpoint when the cursor was near the edge of the screen. The high display fidelity condition used all six screens of the DiVE with stereoscopic graphics enabled, so users could turn physically to view the environment in different directions. This meant that for the mouse and keyboard conditions, users had to be able to turn the mouse and keyboard with them; we placed the devices on a turntable for this purpose. Figure 6 shows a user in the high display fidelity, high interaction fidelity condition.

Participants were placed in an FPS game that required them to navigate several rooms with varying shapes, sizes, and obstacles, destroying "bots" (enemies) along the way. We measured performance metrics such as completion time, shooting accuracy, and damage taken. We also used questionnaires to ask participants about their sense of presence, engagement with the game, and opinions of interface usability.

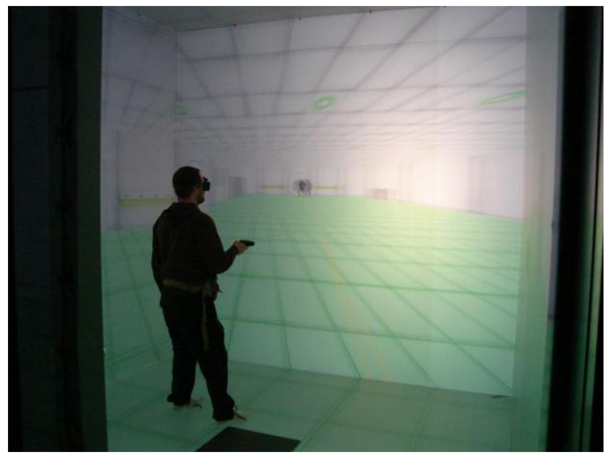

Figure 6. FPS experiment in the DiVE

Performance results (see (McMahan, Bowman, Zielinksi, & Brady, 2012) for complete details) were strongly in favor of two conditions: the condition with low display fidelity and low interaction fidelity, and the condition with high display fidelity and high interaction fidelity. These two conditions are representative of traditional gaming setups and high-end VR setups that simulate the real world as closely as possible. The other two combinations were unfamiliar to users (despite the fact that they were instructed on each combination and practiced it before completing the trials for that condition); these mismatched conditions resulted in poor performance. Thus, the primary lesson from this study was that familiarity, rather than interaction fidelity or display fidelity alone, may be the best predictor of performance and usability.

To explore these effects in a deeper way, we conducted follow-up studies (to be described in detail in a future publication) that allowed us to assess individual aspects of display and interaction fidelity and their influence on the component tasks of an FPS game – long-distance travel, maneuvering (short movements to adjust the viewpoint or avoid an obstacle), searching for enemies, aiming, and firing. We found that high levels of FOR were generally beneficial to performance when using high-fidelity interaction techniques, and that the highest-fidelity interaction techniques improved performance on tasks like aiming and firing.

Visual Scanning Studies

Finally, we are using the MR simulation methodology to examine the effects of fidelity on the effectiveness of military training systems. We have chosen *visual scanning*, in which a warfighter carefully looks at the surrounding environment to detect threats such as snipers or IEDs, as a representative task that might be trained in VR. Our task scenario (Figure 7) involves riding in a vehicle down an urban street and scanning one side of the street (buildings, side streets, roofs, alleys) for threats.

Figure 7. Urban environment used in the visual scanning experiments

We aim to determine how different levels of fidelity impact the effectiveness of such VR training systems, with the goal of producing guidelines that will help the military design future VR trainers. With the MR simulator approach, we can compare different training system configurations using a single VR system.

The first study of this sort examined the effects of amplified head rotations on visual scanning performance (Kopper, Stinson, & Bowman, 2011). Many training systems do not have a 360º FOR, but may still wish to allow trainees to move their heads naturally to turn the virtual camera. In this case, amplifying head rotations can allow 360º of virtual turning with a smaller amount of physical turning. We found that amplification was difficult for users to detect, but high amplification levels (3x) could degrade performance in a counting task during visual scanning.

We recently completed a study examining the effects of FOV and scene complexity (the amount of visual content and detail in the environment) on training effectiveness for a visual scanning task. We measured not only performance, but also how well participants learned a visual scanning strategy we taught them.

We found that participants who trained with higher scene complexity exhibited a better use of the proper strategy when they were assessed in a realistic environment. Lower scene complexity during training may allow trainees to be lazy, resulting in sub-optimal strategies. This underscores the importance of visual realism for training, since participants with more simplistic visuals were less successful in learning the scanning strategies. We also found that performance (threat identification) during training was not necessarily a good predictor of performance during assessment, reinforcing the need to examine strategy learning instead of performance alone. This experiment will be fully described in an upcoming publication.

CONCLUSIONS AND FUTURE WORK

It is critical for the VR and AR research communities to understand the fundamental effects of display and interface characteristics. It is equally critical for practitioners to be able to choose appropriate VR and AR systems that maximize benefit and minimize cost. Both of these require knowledge that can only come from empirical studies of MR systems, but comparing MR systems is fraught with challenges. In this paper, we have presented our MR simulation methodology, which allows for controlled experiments, requires only a single high-end VR system, and allows researchers to study individual components of display and interaction fidelity rather than whole systems.

In the future, we plan to use the simulator approach to study other regions of the MR continuum, such as displays that present only real-world data (e.g., teleconferencing systems). We also hope to simulate other aspects of display systems, such as their ergonomic characteristics, and other types of sensory displays, such as auditory or haptic displays. Finally, we plan to develop a standardized MR simulator software platform, which will allow rapid configuration of experiments and the simulation of a wide range of system components.

ACKNOWLEDGEMENTS

The authors would like to thank the Office of Naval Research, Code 30, for its generous support of this research. Thanks also to Rachael Brady at Duke and Nicholas Polys at Virginia Tech for the use of their facilities for MR Simulator experiments.

REFERENCES

Arthur, K. (2000). *Effects of Field of View on Performance with Head-Mounted Displays.* (Doctoral dissertation), University of North Carolina, Chapel Hill, NC.

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355-385.

Barfield, W., Hendrix, C., & Bystrom, K. (1997). *Visualizing the Structure of Virtual Objects Using Head Tracked Stereoscopic Displays*. In Proceedings of Virtual Reality Annual International Symposium, 114-120.

Bowman, D., Kruijff, E., LaViola, J., & Poupyrev, I. (2005). *3D User Interfaces: Theory and Practice*. Boston: Addison-Wesley.

Bowman, D., & McMahan, R. (2007). Virtual Reality: How Much Immersion is Enough? *IEEE Computer, 40*(7), 36-43.

Cohn, J., Schmorrow, D., Nicholson, D., Templeman, J., & Muller, P. (2003). *Virtual Technologies and Environments for Expeditionary Warfare Training*. In Proceedings of NATO Human Factors and Medicine Symposium on Advanced Technologies for Military Training, Genoa, Italy.

Ellis, S. R., Breant, F., Manges, B., Jacoby, R., & Adelstein, B. D. (1997). *Factors Influencing Operator Interaction with Virtual Objects Viewed via Head-Mounted See-Through Displays: Viewing Conditions and Rendering Latency*. In Proceedings of Virtual Reality Annual International Symposium, 138-145.

Gruchalla, K. (2004). *Immersive Well-Path Editing: Investigating the Added Value of Immersion*. In Proceedings of IEEE Virtual Reality, Chicago, 157-164.

Höllerer, T., Amatriain, X., & Kuchera-Morin, J. (2007). *The Allosphere: a Large-Scale Immersive Surround-View Instrument*. In Proceedings of Emerging Display Technologies Workshop (EDT), San Diego, CA.

Kopper, R., Stinson, C., & Bowman, D. (2011). *Towards an Understanding of the Effects of Amplified Head Rotations*. In Proceedings of Workshop on Perceptual Illusions in Virtual Environments, Singapore, 10-15.

Lee, C., Bonebrake, S., Höllerer, T., & Bowman, D. (2010). *The Role of Latency in the Validity of AR Simulation*. In Proceedings of IEEE Virtual Reality, Boston, 11-18.

Lee, C., Gauglitz, S., Höllerer, T., & Bowman, D. (2012). *Examining the Equivalence of Simulated and Real AR on a Visual Following and Identification Task*. In Proceedings of IEEE Virtual Reality, Costa Mesa, California, 77-78.

Livingston, M. A., II, J. E. S., Gabbard, J. L., Höllerer, T. H., Hix, D., Julier, S. J., . . . Brown, D. (2003). *Resolving Multiple Occluded Layers in Augmented Reality*. In Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR), Tokyo, Japan, 56-65.

McMahan, R., Bowman, D., Zielinksi, D., & Brady, R. (2012). Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game. *IEEE Transactions on Visualization and Computer Graphics*, *18*(4), 626-633.

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IECE Transactions on Information and Systems, E77-D*(12), 1321-1329.

Office of Technology Assessment, U. S. C. (1994). *Virtual Reality and Technologies for Combat Simulation - Background Paper*. (OTA-BP-ISS-136). Washington, DC: U.S. Government Printing Office.

Ragan, E., Sowndararajan, A., Kopper, R., & Bowman, D. (2010). The Effects of Higher Levels of Immersion on Procedure Memorization Performance and Implications for Educational Virtual Environments. *Presence: Teleoperators & Virtual Environments*, *19*(6), 527-543. doi: 10.1162/pres_a_00016

Robertson, G., Czerwinski, M., & Dantzich, M. v. (1997). *Immersion in desktop virtual reality*. In Proceedings of ACM Symposium on User Interface Software and Technology, Banff, Alberta, Canada, 11 - 19.

Schulze, J. P., Forsberg, A. S., Kleppe, A., Zeleznik, R. C., & Laidlaw, D. H. (2005). *Characterizing the Effect of Level of Immersion on a 3D Marking Task*. In Proceedings of HCI International, Las Vegas.

Slater, M. (2003). A Note on Presence Terminology. *Presence-Connect, 3*.

Tan, D. S., Gergle, D., Scupelli, P., & Pausch, R. (2003). *With similar visual angles, larger displays improve spatial performance*. In Proceedings of SIGCHI Conference on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA, 217 - 224.

Ware, C., & Mitchell, P. (2005). *Reevaluating Stereo and Motion Cues for Visualizing Graphs in Three Dimensions*. In Proceedings of Symposium on Applied Perception in Graphics and Visualization, 51-58.

Wither, J., & Höllerer, T. (2005). *Pictorial Depth Cues for Augmented Reality*. In Proceedings of Ninth IEEE International Symposium on Wearable Computers, Osaka, Japan, 92-99.