Evaluating Display Types for AR Selection and Annotation

Jason Wither* Stephen DiVerdi† Tobias Höllerer‡

Four Eyes Laboratory§
University of California, Santa Barbara, California, USA

Abstract

This paper explores the characteristics of different display devices when used for annotation or selection in augmented reality (AR). We compare three different display types – a head mounted display and two hand held displays. The first hand held display is configured as a magic lens where the user sees what is directly behind the display. The second hand held display is configured to be used at waist level (as you would commonly hold a tablet computer) but the view is still of the scene in front of the user.

Making an annotation or selection in AR requires two distinct tasks by the user. First, the user must find the real (or virtual) object they want to mark. Second, the user must move a cursor to the object's location. We test and compare our three displays with respect to both tasks.

We studied the first part of the task (finding the object) by having users complete a visual search task, looking for a single letter in a set of similar looking letters. Users looked for both real and virtual letters to compare whether the displays have different performance when more of the scene is real versus virtual.

We studied the second part of task by giving users a static cursor at the center of their screen and asking them to select a series of objects. To select each object, the user needs to move the display and line up the cursor with the object they wish to select.

We found that using a hand held display in the magic lens configuration was faster for cursor movement than either of the other two displays. We also found that there was no significant difference between displays in the amount of time it took users to find objects.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques;

Keywords: augmented reality displays, user study, interaction techniques, mobile devices

1 Introduction

Traditionally, augmented reality (AR) applications have primarily used head mounted displays (HMDs) for visual output. There are many reasons for this design choice, one of the primary reasons being that most designers want the AR experience to create a seamless integration between real and virtual worlds. Using a HMD allows for the user's visual field to be completely immersed in the augmented environment making any virtual objects seem present and persistent in the real world. This is particularly important when the application requires constant attention, something that may be needed for a variety of reasons, such as dynamic virtual content.

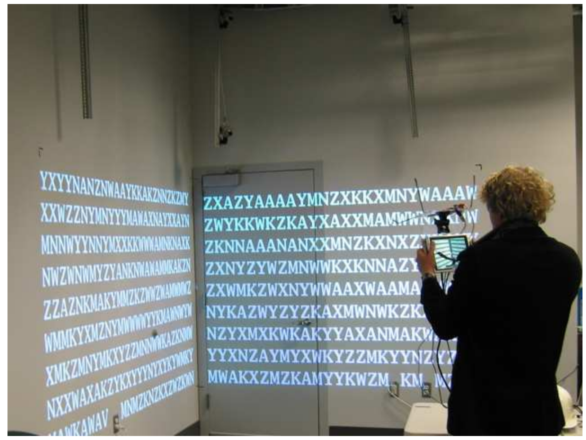

Figure 1: Our study setup. A user is searching for a single *V* among a set of similar looking letters using a magic lens. The display is tracked in 6DOF using infrared LEDs on the device, and cameras mounted to the walls.

Dow et al. [4] have developed an excellent example application in this space, porting a popular desktop computer game, *Façade*, to an AR environment by physically rebuilding the room the game takes place in and having the virtual characters move through that real space.

Recently, there have been an increasingly large number of AR applications that are built to use hand held displays like ultra-mobile computers, PDAs, or cell phones as the primary display. An example application in this space is Wagner et al.'s [23] *Invisible Train* game. This application is a physical model train track that is tracked using PDAs equipped with cameras, and virtual trains are drawn on the real tracks. Users can then interact with the trains using a stylus based interface. The reason most often cited for using hand held displays for applications like this is user acceptance. In the short term it is clear that for AR to be broadly adopted it will need to run on devices similar to those that are already ubiquitous. The commercial market already seems to be headed in the right direction for this to happen, with cameras, accelerometers, and GPS sensors becoming fairly common in many of these devices, and more AR applications being developed for these commodity hand held devices.

While hand held displays are growing in popularity for AR, it is not clear how these displays compare to HMDs for many common AR tasks. There are obviously many differences between the two types of displays, one of the most apparent being the difference in the level of visual immersion each display offers to the user. While using a hand held display, the separation between the virtual and real objects in the scene is much more pronounced, given that the user can simply look up from the display to see which objects exist in the real world. It has not been shown, however, what effect this difference has on task performance. Some virtual reality (VR) research has shown that presence does effect task performance [18], but it has also been shown that hand held displays can be just as

*e-mail: jwither@cs.ucsb.edu
†e-mail: sdiverdi@cs.ucsb.edu
‡e-mail: holl@cs.ucsb.edu
§web: http://ilab.cs.ucsb.edu/

effective for task performance in VR as larger more immersive displays [5][7].

In this paper, we will examine how display choice effects task performance in an AR environment. We have conducted a user study to look at the related tasks of selection of existing objects, and annotation of new objects. These general tasks can be broken up into two conceptual parts. First the user must search for and locate the object they wish to select or annotate, and second they must move some sort of cursor or selection device to that location. One difference that sometimes exists between selection and annotation is that when annotating a new object it is often necessary to assign a distance to that object as well as a direction vector. In the past, we have looked at techniques for completing annotations at a distance, including techniques to determine the distance to the object being annotated [24], but for this study we will only look at making the annotation or selection on the image plane, and assume that there is some other technique like ray casting [2] in place to complete the annotation or selection. We had users search for both real and virtual objects in the study to determine if display choice should be impacted by the amount of virtual content present in the AR scene.

More concretely, this paper compares two different tasks, moving a cursor and visual search, between three displays. The first display device we used is a head mounted display, and the second is a hand held display, which we used in two different configurations. First, users were required to hold the hand held display at approximately waist height and look down at it, like they would a tablet computer. We will call this display configuration the tablet. In this case, the camera was pointed directly off the top of the display (See figure 7). We also looked at holding the hand held display like a magic lens [15] where the camera is pointing directly behind the display, and users must hold the display up at head height, and look through it (See figure 4). We will refer to this configuration as the magic lens.

One goal of this study was to be able to make recommendations for using one display over the other in the application space of annotation and selection. The major findings of our study are:

- When moving a cursor by moving the display, the magic lens was faster than either the HMD or tablet. The tablet also had a higher error rate than either of the other two displays.

- There was no significant difference in time to completion between displays when searching for either real or virtual letters. This result also extends to users who looked directly at the real letters instead of looking through the display.

- There was a significant difference in how fast users thought they were able to complete the visual search task between displays. Users thought they were faster while using a HMD when searching for virtual letters, and those users who looked away from the hand held displays when searching for real letters thought that that was significantly faster approach.

The rest of this paper is broken up as follows, in Section 2 we discuss some of the major differences in using a hand held display vs. a head mounted display for an augmented reality application. Section 3 will describe our design, methodology and setup for conducting our user study, and list the hypotheses we created before conducting our study. Our study results will be presented in Section 4, and further discussion will be given in Section 5.

2 Background

Both head mounted displays, and hand held displays have been used for quite some time in augmented reality. HMDs are used far more commonly though for many of the same reasons they are used more frequently in VR. In both cases, a HMD provides a level of visual immersion that is difficult to match with other displays. However, visual immersion means a slightly different thing in AR than VR because in AR many of the things the user is seeing are present in the real world, and could be seen without any display at all. The HMD then adds visual immersion only for the virtual objects in the scene, rather than the entire scene. Another reason for using a HMD is that it leaves the user's hands free to either hold other devices or complete other tasks.

Interaction in AR while using an HMD has been well explored, and there are many different approaches. Billinghurst et al. [1] established that in VR spatial head tracked displays have advantages over screen stabilized approaches for certain tasks, including searching and selection. In our previous work we have studied moving a 3D cursor using tracked head motion by itself as much as possible to leave the user's hands free [25]. Thomas [21] also looked at using head motion for selection and annotation, comparing it to two different techniques based on hand held input devices. Many other people have looked at using hand held devices to control a cursor as well. Zucco et al. [27] have compared different off the shelf pointing devices for use in AR with an HMD. Kaiser et al. [8] have looked at integrating where the user is looking and pointing to select objects in both VR and AR. Others have developed techniques to track the user's hands either with [19], or without extra instrumentation [9] from a HMD mounted camera and use them for input. Many traditional VR techniques for interaction including ray-casting [2], and image plane interaction [13] were also developed primarily to be used with a head mounted display.

Another important issue to consider for any sort of interaction is user comfort. In general it seems that a HMD would be more comfortable for long term application use than a hand held display, but this will depend strongly on the actual display chosen and the type of interactions needed for the application. If a comfortable HMD is chosen and the application requires long continuous use a HMD is likely to be more comfortable. However in many other cases a hand held display will be more comfortable, particularly if the application only requires intermittent use, in which case the hand held display could be lowered to a resting position between uses.

Being able to use the display intermittently and put it down between uses is definitely an advantage of terms of near term user acceptance as well. One case where this has been very clearly illustrated was by Wagner et al. [23] with their *Invisible Train* game. They have had thousands of people play this game, which likely would not have been possible with an HMD based device because of the added bulk, brittleness, and strangeness of a non-main-stream device.

Rekimoto [15] was one of the first to use a hand held display in AR with his NaviCam project in the early 90s. More recently there have been a large number of projects that have been designed to use mobile devices not only as the display, but also to do all the processing. The majority of these projects have focused on how to use the hardware on the device to do tracking, but have also shown the potential of hand held devices in many different domains. Föckler et al. [6] demonstrated using a mobile phone as a museum guide. In their application the phone recognizes objects in the museum and then provides extra information to the user about those objects. Olwal [11] demonstrated another application that made use of a tracked mobile phone, encoding extra information only visible through the phone into maps and other real world objects. Reitmayr and Drummond [14] also used a hand held device for their model based tracking work, but their camera is pointing at an angle between where our cameras point in the tablet and magic lens cases. These projects illustrate some of the potential of using hand held displays for AR, but do not deeply explore the design choices for the device. In many of these papers the camera points directly behind the display, which is necessary in the case of working the commodity hardware like mobile phones, but has not been shown to necessarily be the best choice. Having the camera point up towards the top of the display might also make sense in some applications because the user would become much less fatigued if they could rest the display on a strap around their neck rather than having to hold it up to view the scene through it. For larger displays like a tablet computer instead of a mobile phone, this might be more of an advantage. It is also very interesting to look at the modality of input in each of these cases because it is often very different than established techniques designed for an HMD. Moving the display to either view or interact with the scene is a very commonly used design, and makes sense given the fact that the display is in the user's hands limiting their ability to hold other pointing devices. The *Invisible Train* also used a stylus and touchscreen for input, which is a very different input modality that is not available at all when using a HMD. Because there is no way to have the same interface while using a HMD we chose not to use the stylus for input in our study, even though it might have performed very well.

3 STUDY

We designed our study to look at the two main components of a selection / annotation task independently. The first part of the task that we looked at was cursor movement, or how long it takes for the user to move a cursor to the correct screen coordinates to select something. The second was visual search, or finding a single object out of a set of other similar looking objects. We compared the three different displays for these tasks, a head mounted display, magic lens display, and tablet display.

3.1 Design

3.1.1 Cursor Movement Task

We studied the movement task by having users complete a simple connect the dots puzzle like those commonly found in children's books. The users were shown a numbered series of dots in a recognizable and simple shape in the real world, and asked to virtually connect them by clicking on the next dot in the sequence. The users were also shown a virtual circle around the next dot in the sequence, and had to click within that circle for the dot to be considered selected (see Figures 2 and 3).

The goal of this part of the study was simply to see how fast the user could move the cursor to the correct location, with no extra mental effort required to figure out where the correct location was. Any differences between techniques will therefore show either the mental or physical differences in the user's ability to move the cursor to the correct location. This test was partially inspired by a study conducted by Schulze et al. [17] to determine if display choice impacted users ability to mark spheres in VR. We chose to use a connect the dots pattern rather than to have users click on as many objects as they could within a time limit to remove any effect marking strategy might have on the results.

We completed this part of the study in two steps. We first wanted to ensure that the difference in the visual display itself was not going to have significant impact on how users perceived the world, so to control for this we had users complete the same connect the dots picture, with the same input device, leaving the display as the only variable. The users were shown the connect the dots pattern that can be seen in Figure 2, a four leaf clover, that fit inside the camera's field of view, and contained 26 dots. The users were then asked to hold the display still and complete the connect the dots using a hand held trackball. Users completed this task using both the HMD and tablet. We used the same pattern for both tests as well as training sessions that were conducted before the test with each display. We were not concerned about a learning effect based on the

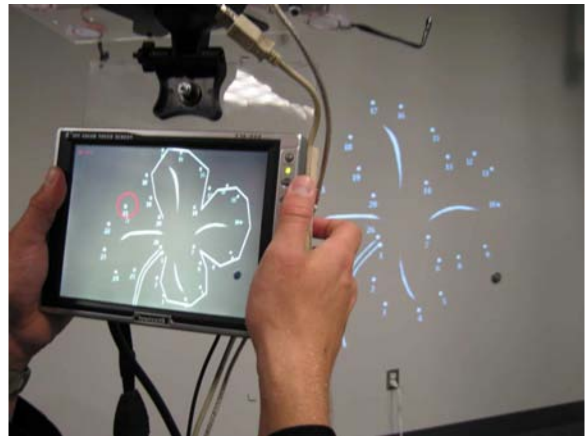

Figure 2: The connect the dots image presented to users when using the hand held trackball to complete the cursor movement task. As can be seen the entire image fits on the display, allowing users to keep the display still, and only use the trackball to complete the task.

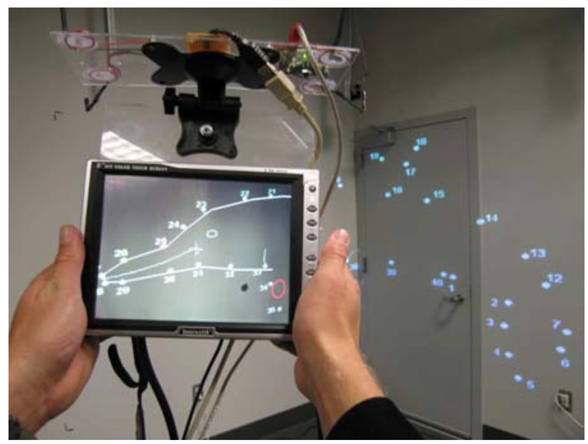

Figure 3: The dolphin image used for the cursor movement task when users were asked to move the display to complete the task. In this case the participant is using the magic lens.

pattern since it was recognizable before the first training session even started.

Once this part of the study was complete we had the user complete a second connect the dots picture by moving the actual displays. This was accomplished by having a cross in the center of the user's field of view act as the cursor. The user had to line this cross up with each dot to select it. While using a HMD, head motion has shown to be an effective means of cursor control [25], and we wanted to test similar techniques using a hand held display to more carefully control our results. Naturally, it is possible that there are better techniques to use with a hand held display. For example, stylus input on a tablet has been shown to be quite fast at object selection [3]. Stylus input is very tablet specific however, and there is not an equivalent technique using a HMD, so we did not include it in our study. Even with the stylus, users would still have to move the display to get the object they wish to select on the screen, so it is still important to see how quickly users can move the actual display. For this part of the study we used a larger connect the dots image of a dolphin containing 40 dots that spanned two walls around a 90 degree corner (See Figure 3). We chose to use two walls for our study so that the objects users were selecting were actually in 3D, instead of just 2D if we had only used one wall. Users were given a chance

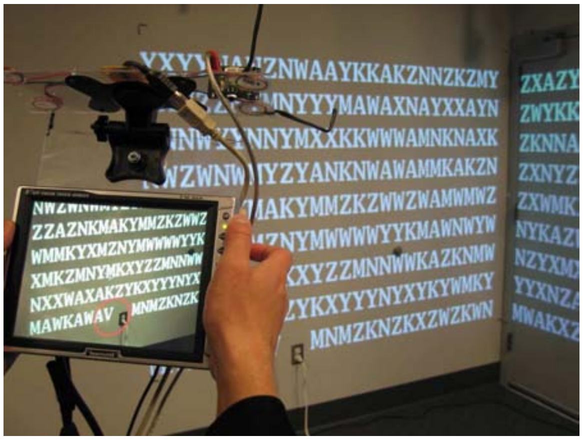

Figure 4: A user's view of world when completing the visual search task for a real letter, in this case a *V*. This user has just found the *V* using the magic lens, and marked it.

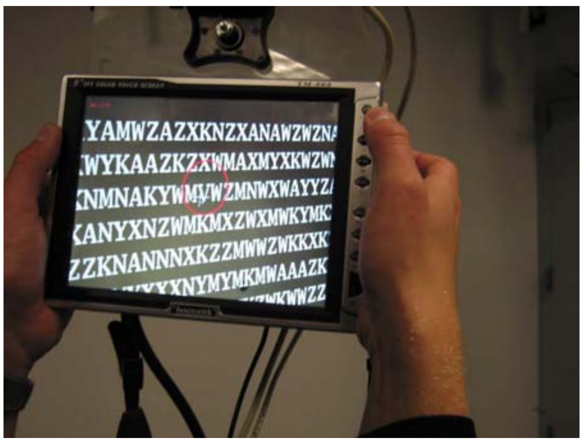

Figure 5: The user's view of the visual search task with virtual letters. As can been seen the letters are only viewable through the display.

to train with each display for input until they were comfortable, and then tested on the same image.

3.1.2 Visual Search Task

The second task we studied was a visual search task. In general visual search is often used to distinguish cognitive issues in search and attention. It has also been used in VR as a test for immersion [12][16]. The task normally involves searching for one or more known objects out of a larger set. The goal of the majority of visual search tasks is to force the user to conduct a limited capacity search [22] requiring them to carefully look through the entire set to find the object they are searching for. The opposite of this would be a search using pre-attentive processes where the target immediately jumps out at the user. An example of this would be if the target was a different color than all of the other objects in the set.

We chose this task to see if the choice of display impacts a user's ability to conduct a visual search. There are several reasons we think that the display might effect visual search. It is possible that one display might allow the user to build a better mental map of where they have already looked than another. Also, if users look away from the hand held displays and directly at the real objects in the scene that might allow them to more quickly complete the search through those objects because they will have a larger field of view than they would looking through a display at the same space.

For this task we had users search for a single letter that either occurred once or not at all, in a set of similar looking letters (see figure 4). We used two sets of camouflage letters, one where the dominant feature in the letters is diagonal lines (AKMNVWXYZ) and one where the dominant feature is horizontal and vertical lines (EFHILT). These are the same sets of letters used by Pausch et al. [12] and Robertson et al. [16] in their visual search work in VR. For our study we covered two walls with the letters, using 450 letters overall. In this set the letter the user was searching for had a 75% chance of being present. We tested each display on two different random sets of letters in two conditions, one where the letters were present in the real world (See Figure 4), and a second where the letters were virtual and only visible through the application (See Figure 5). This gave us a total of 12 visual search test cases per user. We permuted the order of both displays and real or virtual data sets for the 12 tests between users. The motivation for having real and virtual cases was twofold. First, since we are studying an AR application there will commonly be both real and virtual objects present at any time that a user may want to interact with. The second motivation was that the HMD is a much more visually immersive display than the hand held display in either configuration. This extra level of immersion might give it an advantage when searching for virtual letters, while the hand held displays might have an advantage when searching for letters in the real world, since users could easily look away from the displays. We made it clear to the users that they could use whatever approach they wanted to search for the letter, and more than half did look away from the hand held display to directly view the letters on the walls. When looking through any of the displays users could see between 25 and 40 percent of the letters depending on the direction they were looking and exactly where they were standing.

3.2 Setup

The study was conducted in a lab environment with the user standing between 2.5 and 3 meters from each display wall. Users stood behind a table for the entire experiment, but were not instructed to stand exactly in any particular place. A small minority of users took half steps in different directions during the study. To track the user we used WorldViz's Precision Position Tracker (PPT) system with a rigid body plug-in allowing us to track orientation as well as position. The HMD we used as well as the hand held display in both positions can be seen in Figures 4, 6, and 7. For both the HMD and hand held display the four arms extending from the display hold infrared LEDs that are tracked in 3D by PPT by using four infrared cameras. We used Point Grey FireFlyMV cameras for both displays with 6mm micro lenses, giving us a horizontal field of view of 35 degrees. With this setup users could see between 25 and 40 percent of the area used for display which was 2.4 by 1.7 meters on each wall with both sides coming together at a seam in the corner. Both of our displays were video see through and ran at 640x480 which was also the native resolution of the cameras. Our applications ran at 25 fps. We used a Sony Glasstron PLM-S700E for our HMD and an Innovatek 868 touch screen display with a diagonal size of 8 inches for our hand held display. In both cases we used an Alienware Area-51 m5500 to run our applications and drive the displays.

We used two projectors to show our real world objects both to make it faster to switch between setups while running the study, and also for repeatability. We projected white on black, so there was no apparent "screen", only the objects the user was expecting to see. The other options for displaying real objects, like having letters printed on large sheets of paper seemed to be prohibitive in the amount of time it would have taken to switch between test cases, particularly in the visual search portion of the study when we were changing setups frequently. We do not feel we lost any

Figure 6: The HMD we in our study. The four arms support infrared LEDs used for 6DOF tracking.

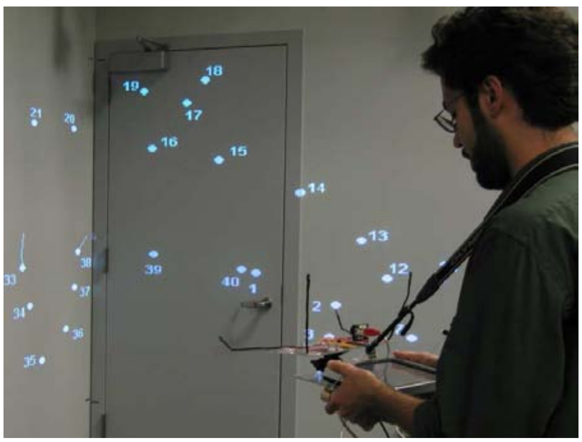

Figure 7: The hand held display used in our study in the tablet configuration. Here a user can be seen completing the cursor movement task.

visual fidelity while looking at the projected objects through the displays. They looked just as clear as any other real objects in the scene. The projectors we used were bright enough and close enough to the walls for images to look very real and solid, even with normal lighting on in the room.

For most of our tests we needed users to click a button to select either the next dot in the connect the dots or the letter that they had searched for. When using the HMD we used the trigger button on the hand held trackball, an ErgoTouch RocketMouse, for this purpose, and for the hand held display we added a button to the back of the display that the user could click while holding the display in either configuration. For the first part of the cursor movement task users used the RocketMouse to move the cursor. Users were not told how to hold the hand held display, but nearly all held it with two hands for the duration of the study.

We also had users fill out a questionnaire during the study. The main purpose of the questionnaire was to ask users how they thought they had performed with each display for each test and how comfortable they were using that display. We had users complete this questionnaire in parts, completing the corresponding part of the questionnaire after each part of the study.

We had 21 users participate in our study, 17 men and 4 women. All were between 22 and 30 years old, and agreed to participate in the study without compensation. All had used some form of hand held display before, and half had used a head mounted display. Half had also previously used a tablet display.

3.3 Hypotheses

We formed the following hypotheses before conducting our study.

Hypotheses for cursor movement task:

1. When using the trackball to move the cursor, the use of the two different displays will not make a significant difference on either the speed or accuracy with which the user will be be able to complete the connect the dots task.

2. When completing the cursor movement task by moving the displays the HMD will perform the best when performance is defined as speed and accuracy. The tablet's performance will suffer due to the extra layer of indirectness between the orientation of the display and the orientation of the camera. The magic lens will also have a poorer performance because of the awkwardness of holding up the display.

Hypotheses for visual search task:

1. When searching for a real letter, users will be able to find the letter more quickly when using the hand held displays than the HMD because they can look away from the display, and by directly looking at the objects have a larger field of view as well as a better spatial sense of where they have already looked.

2. When searching for a virtual letter, the users using the HMD will be able to find the letter more quickly than with the hand held displays because they will have a more immersive view, giving them a better spatial understanding of what they have already looked at and where they still need to look.

4 RESULTS

In this section we present an analysis of the results of both tasks in our study. In our post study questionnaire we found one result that pertains to both parts of the study. We asked users to rate how comfortable they felt each display was to use. The results to this question showed no significant difference (one way repeated measures ANOVA $p = 0.1$). However users did feel that the magic lens was slightly less comfortable, a feeling that would likely be amplified the longer an application was running for. This suggests that a magic lens display could be useful for short tasks or applications where the user may want to look at the virtual objects only sporadically, but not necessarily for applications where long term continuous viewing of virtual objects is needed.

4.1 Cursor Movement Task

As described, we had users perform two similar tasks to test cursor movement. In the first task users held the display still and moved the cursor with a hand held trackball. Our hypothesis was that using the two different displays would not effect task performance if the control device used to complete the task was constant between the two.

We found that there was no significant difference in the amount of time it took users to complete the connect the dots task. It took users 31.5 seconds to complete the task while viewing the scene through the HMD, and 33.9 seconds while viewing the scene through the hand held display. Those times correspond to 1.5 seconds and 1.62 seconds per successful click respectively. A paired t-test showed no significant difference between these times (p =

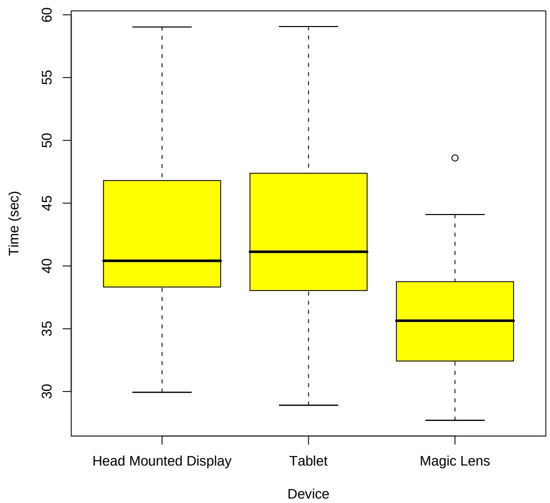

Figure 8: Results for the average time users took to complete the cursor movement task with each of the three displays. In this and all our plots, the dark line in the middle of the box is the median, the box itself extends plus and minus one quartile, and the whiskers extent to the fifth and ninety fifth percentiles. Any outliers are represented as empty circles.

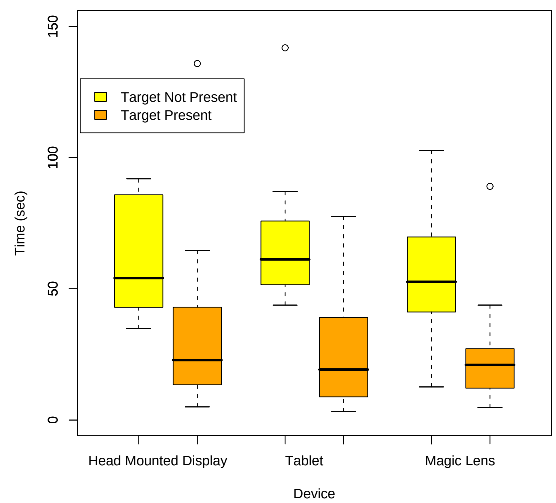

Figure 9: Results from the visual search task when the letters were visible in the real world.

0.2). We did find that users made significantly more errors (clicks not on the next dot), while using the hand held display (1.95 errors) than the HMD (0.86 errors) (paired t-test p = 0.017).

We were surprised by this result, since our hypothesis was that the two displays would be the same in time and accuracy. We think the discrepancy can be explained by the fact that some users had a difficult time holding both the trackball and the display. These users still tried to hold the display with two hands which made it more awkward to use the trackball at the same time. We tried to control for this by attaching a neck strap to the display, but some users did not want to use it and instead held all the weight of the display in their hands, which was more difficult to do with one hand. Other users also used both hands to hold the trackball when using the HMD, feeling that they could hold it steadier that way. In general, we feel that this difference was much more likely due to ergonomic issues of using the trackball and hand held display simultaneously, something that would not normally be done, than a difference in how the scene was perceived through the two displays.

For the second cursor movement task we had users move the display itself to change the location of the cursor which was statically placed in the center of the field of view. Our hypothesis was that the HMD would perform the best, while the performance of the magic lens would suffer because of ergonomics, and the tablet would suffer because of the extra mental rotation users would need to perform to go from screen coordinates to camera coordinates.

We found that the magic lens was actually significantly faster than the other two displays at completing the connect the dots task (one way repeated measures ANOVA $p < 0.0001$) as can be seen in Figure 8. Users completed the 40 dot picture in 36.2 seconds (0.91 seconds per dot) with the magic lens, compared to 42.7 seconds (1.07 seconds per dot) with the HMD, and 42.2 seconds (1.06 seconds per dot) with the tablet.

There was also a significant difference in the number of errors that users made while completing the task (one way repeated measures ANOVA p = 0.019). The HMD and magic lens had the fewest errors, with 3.24 and 3.62 erroneous clicks respectively, while the tablet had 5.05 errors per trial.

These results give us a rough ordering of the three displays for our task. The magic lens was the fastest and had a small number of errors, the HMD was not as fast but also had a small number of errors, and the tablet was slow and had a large number of errors. We think the reasons why the magic lens performed the best are because it is a very intuitive and direct interaction, and people's hands are generally very dexterous. The intuitiveness of the display was particularly shown by one user in our study who made shooting noises every time he clicked while completing the task. We feel that the HMD was not as fast for two related reasons. Head motion is generally augmented with eye motion, something that users were not able to make use of for this technique. Because of this users were not as used to having to move their head for every annotation. Also, head motion does not generally need to be as precise as it did for this study because eye motion can compensate for not turning one's head completely to look directly at each object.

The tablet was likely more error prone because of the extra mental rotation users had to perform when moving the display. Some users particularly had trouble remembering immediately which way to move the display to move the cursor up and down. This extra level of mental effort likely reduced users precision with the display leading to the larger number of bad clicks.

4.2 Visual Search Task

Our visual search task was divided into two parts. In one part the users searched for letters that existed in the real world, and in the other part the letters were only present virtually in the application. During this part of the test users were directed to mark the letter if it existed and tell the study administrator if they thought the letter was not present. In our analysis of the case when the search letter was present we have only included results from trials where the user found the letter without first claiming that it was not present. Including these extra cases does not change the significance of any of our results, as they are largely outliers. Generally when a user scanned the entire space and decided a letter was not present they had a more difficult time finding it on a second pass than they would have a different letter in a different location in a new test. This is largely because users did not change their search strategy between

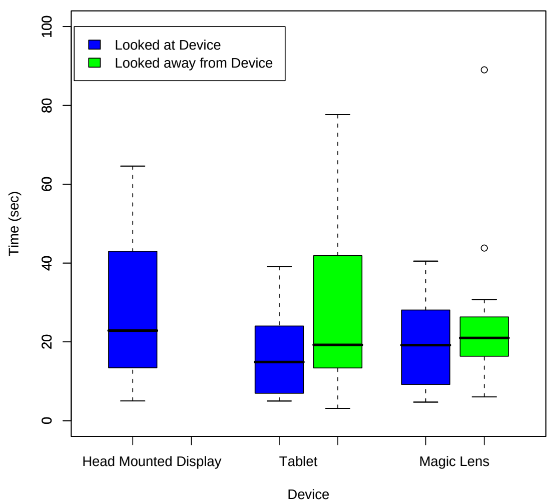

Figure 10: Results from the visual search task when the letters were visible in the real world and the target letter was present. Results are split between users who looked up from the display and those who did not.

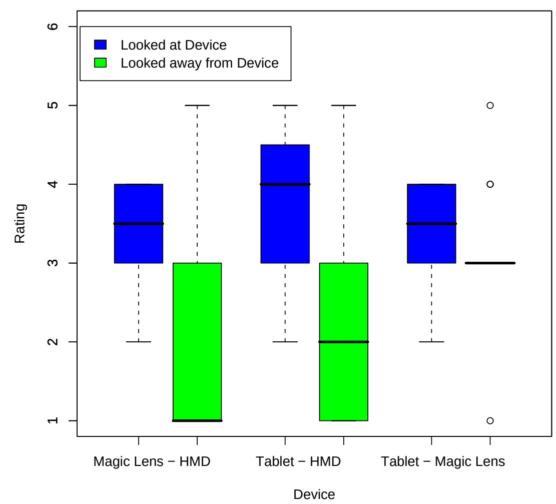

Figure 11: Users responses to the post study questionnaire when asked on a 1 to 5 scale which display they thought they were faster with. A score of 1 would indicate they thought they were much faster with the left display of each pair, and a score of 5 would indicate they thought they were faster with the display on the right. Results are again split between users who looked up from the display, and those who did not.

passes through the letters, and if their strategy caused them to miss the target letter the first time they were also likely to miss it the second time.

We will first discuss the case where the letters were present in the real world. In this case we found no significant difference in time to finding a present letter between displays (one way ANOVA p = 0.47). Users using the HMD, magic lens, and tablet took 30.2, 23.3, and 24.9 seconds respectively to find the hidden letter. In all cases the standard deviation was quite large (between 15 and 30 seconds) due to the target letters being randomly spread throughout the search space. In some cases that meant users would see the target almost immediately, while in other cases they would have to scan through nearly the entire space before finding the target letter. We also found no significant difference between displays when searching for letters that were not present (ANOVA p = 0.4). In this case users reported that they thought the letter was not present in 59, 54.2, and 70.3 seconds for the HMD, magic lens, and tablet respectively. One interesting thing to note though is that these times are roughly double those of when the users did find the letter. This indicates that users did very little rescanning before deciding the letter was not present. This seems appropriate for our task since the space in which the letters was present was very simple, allowing users to easily know where they had already looked, and where they had not. Both results for when the target was present and not present can be seen together in Figure 9.

We also looked at results to see if there was any difference in performance between those users who chose to look away from the hand held display and directly at the letters on the wall, and those who did all of their scanning by looking at the display. 13 of our 21 users looked up from the display and directly at the letters on the wall when that was possible. We found from a two way ANOVA that there was no significant difference in the time it took to find a present letter between those who looked up and those who did not. This division can be seen in Figure 10. We are quite surprised by this result, as we had hypothesized that it would be much faster to look directly at the letters than it would be to look through the display. Users of our study were similarly surprised. Using a two way repeated measures ANOVA there was a significant interaction between which display users thought was the fastest to find the letters with, and whether they looked up from the display (p = 0.013). Users who did look up thought they had performed faster with either of the hand held displays, since they looked away from both, while users who did not look away from the display thought they had performed faster with the HMD. This result can also be seen visually in Figure 11. Another possible indication of a difference between looking up and not looking up comes from how frequently users said the target letter was not present, when it was in fact present. Those users who did not look away from the hand held display gave this false negative answer 7 out of 25 times (28%), while users who did look away from the display only did 3 out of 38 times (7.9%). The number of times that this happened in either case is small enough that it is difficult to tell if there is any significance to it. If it was significant it might show that users who looked away from the display were more likely to look through the entire space, while users who looked at the display might have inadvertently skipped parts of it.

We also had users search through letters that were only virtually present. The motivation for this was to see if some displays were better for viewing virtual objects with, while others were better for viewing objects in the real world. However, we found no significance between times in searching for letters that were present (one way ANOVA p = 0.73). On average users took 30.4, 35.4, and 31.4 seconds to find the present letter using the HMD, magic lens, and tablet respectively. When the letters were not present there was again no significance in how long users took to decide that the target letter was not present (one way ANOVA p = 0.11). In this case the averages times the users took to decide this were 65.3, 70.8, and 95.8 seconds respectively, in the same order given above. These times can also be seen in Figure 12. This result again disagreed with our hypothesis, since we thought users would be faster when using the HMD because of the more immersive view of the virtual objects. Interestingly users also thought this. When asked which display they thought they performed fastest with users picked

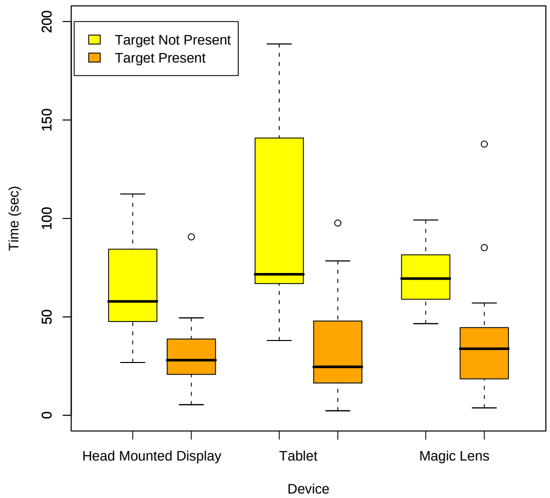

Figure 12: Results from the visual search task when the letters were visible only virtually.

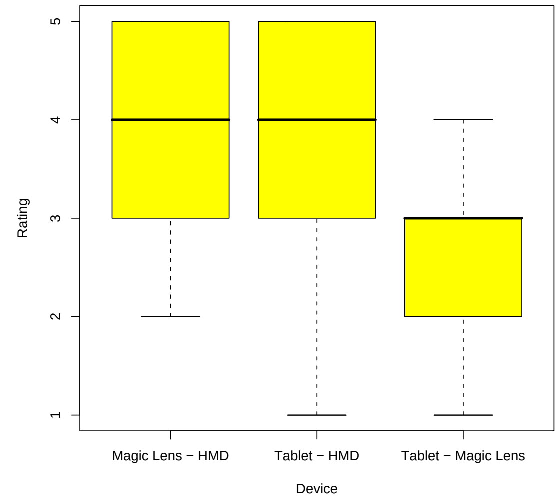

Figure 13: Questionnaire results from when users were asked which device they thought they had performed better with when only virtual letters were present.

the HMD over either of the other two (one way repeated measures ANOVA p = 0.006) as can be seen in Figure 13.

5 Discussion

While we have looked only at two common but specific task, we feel that our results shed some light on the more general differences between using a hand held display and head mounted display in augmented reality.

For the cursor movement task we found that the magic lens performed the best, followed by the HMD, and then the tablet. We feel that this result can be interpreted in a number of ways. First it clearly shows that using display motion as an interface for a hand held display works very well. On the other hand the HMD was not much worse, and has the advantage of not fatiguing the user nearly as quickly, perhaps making it a better choice if the application being developed requires more long term attention. Fatigue did not manifest itself in our study however because all of our tasks were relatively short.

There should obviously be many factors other than performance in a cursor movement task that should be taken into account when choosing which display to use in a particular application, including the audience and type of application. We feel that all three displays performed well enough to be recommended for different applications. For instance, even the tablet could be recommended in the case of a heavier display that users would not be able to hold up for as long, especially since a stylus based interface could be more easily be used with it. It could be ideal for an application where users make a small number of annotations and then attach extra content to those annotations, a task that the stylus would likely excel in. It also did not perform a large amount worse in our study showing that with more use, users could likely become familiar with the extra mental rotation needed to go from the screen coordinate system to the camera coordinate system.

The visual search task that we used in our study has been used in VR by Pausch et al. [12] and others to try to determine differing amounts of presence in the virtual environment between displays. We do not feel that we can make a similar claim of presence in AR because presence and immersion do not have nearly as clear of definitions in AR as they do in VR [26]. There have been a small number of attempts to compare the amount of presence felt in an AR environment to a similar VR environment. Dow et al. [4] found that in their fully immersive, interactive story based AR application users had a higher sense of presence than in a desktop based VR version of the same application. Tang et al. [20] tried to elicit similar responses from users using the ITC-Sense of Presence Inventory, a standard presence questionnaire. They were unable to find any significant differences between the AR and VR versions of their application however, likely because the questionnaire was not designed with AR in mind. There has been no attempt to quantify presence in AR to the level where it might be possible to distinguish different levels of presence in different displays.

One result that we were very interested in was if it would make a difference if users looked up from the display when searching for a target letter in the visual search task. We found that there was no difference in the time it took them to find the target letter, although there was perhaps some difference in how the space was perceived since a larger percentage of users thought a present letter was not present when looking through the hand held displays. This was likely because users had a much larger area that they could see the letters in when looking away from the display than looking at the small hand held screen. There have been studies in both AR [10] and VR [7] looking at whether the size of the screen the user is looking at effects task performance, but those studies have produced mixed results. Our study does little to clarify this since the task performance was the same whether users looked at the small display screen or directly at the wall, while at the same time users who looked directly at the wall thought they were much faster. This contradictory result is possibly because users were able to see the objects better without looking through the hand held display, but because there was a high level of task involvement in both cases that was not the dominant factor in how fast they were able to find the target letter. Users who did not use the display to find the target letter were also slowed somewhat by having to find the target, then move the display, and find it a second time while looking through the display. The fact that users thought they did better when looking away from a hand held display when searching for real letters, and thought they did better with the HMD when searching for virtual letters might suggest that display choice should also depend on the level of virtual content in the scene. If there is more virtual content, a more immersive HMD might be a better choice.

6 Conclusions and Future Work

We have conducted a study comparing techniques used for selection and annotation in augmented reality between three different displays. To compare techniques common to selection and annotation we broke these tasks up into two smaller parts. The first step that needs to be done for any selection or annotation is to find the object to be selected or annotated. Once the user has determined what they would like to select or annotate they then need to move a cursor to that location to complete the task. We have looked at these two steps independently, first looking at how long it takes to move a cursor while using different displays by having the user select a large number of objects. Second we looked at how well users can find an object with each display by presenting them with a complex visual scene where the object they wish to select is difficult to find, forcing them to perform an effortful, limited-capacity search. We had users look for both real and virtual objects to determine if some displays would perform better in one case than the other.

We found several significant results in our study. First we found in the cursor movement portion of the study that the magic lens was the fastest of all the displays we tested. The HMD and tablet had approximately the same speed, but users made fewer errors with the HMD. This validates using a hand held display for AR, particularly if the task involves a large amount of cursor movement. While a HMD may have many other advantages in some cases, this study has shown that for annotation and selection using a magic lens may be more suitable than a HMD. It performed faster for the cursor movement portion of the study and no worse than the others during the visual search part of the study. This result also shows that more visually immersive displays do not directly correlate to better task performance, something that has also been found in VR [7].

In the visual search portion of our study we had two particularly interesting results. We found it quite surprising that there was no significant difference in either the virtual or real case in task performance between the different displays. In spite of that, users had strong feelings that they had performed better with particular displays. They favored the hand held displays that they could look away from when searching for real objects, and the HMD when searching for virtual objects. These results suggest to us that there were other factors involved that caused the task performance to be so similar. Perhaps the user's high task attention overwhelmed the the smaller differences caused by visual difference between displays.

Presence has been defined and redefined for years in VR, but clearly those definitions do not apply directly to AR. In future work we are very interested in trying to formulate more concretely what presence means in AR, and if the difference in visual immersion from different displays could effect the user's level of presence in an AR application.

References

[1] Mark Billinghurst, Jerry Bowskill, Nick Dyer, and Jason Morphett. Spatial information displays on a wearable computer. *IEEE Computer Graphics and Applications*, 18(6):24–31, November/December 1998.

[2] Douglas A. Bowman. *Interaction techniques for common tasks in immersive virtual environments: design, evaluation, and application*. PhD thesis, Georgia Institute of Technology, 1999. Adviser-Larry F. Hodges.

[3] Alan Chamberlain and Roy Kalawsky. A comparative investigation into two pointing systems for use with wearable computers while mobile. In *ISWC '04: Proceedings of the Eighth International Symposium on Wearable Computers (ISWC'04)*, pages 110–117, Washington, DC, USA, 2004. IEEE Computer Society.

[4] Steven Dow, Manish Mehta, Ellie Harmon, Blair MacIntyre, and Michael Mateas. Presence and engagement in an interactive drama. In *CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1475–1484, New York, NY, USA, 2007. ACM Press.

[5] George Fitzmaurice, Shumin Zhai, and Mark Chignell. Virtual reality for palmtop computers. *ACM Transactions on Information Systems*, 11(3):197–218, 1993.

[6] Paul Föckler, Thomas Zeidler, Benjamin Brombach, Erich Bruns, and Oliver Bimber. Phoneguide: Museum guidance supported by on-device object recognition on mobile phones. In *MUM '05: Proceedings of the 4th International Conference on Mobile and Ubiquitous Multimedia*, pages 3–10, New York, NY, USA, 2005. ACM Press.

[7] Jane Hwang, Jaehoon Jung, and Gerard Jounghyun Kim. Hand-held virtual reality: A feasibility study. In *VRST '06: Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pages 356–363, New York, NY, USA, 2006. ACM Press.

[8] Ed Kaiser, Alex Olwal, David McGee, Hrvoje Benko, Andrea Corradini, Xiaoguang Li, Phil Cohen, and Steven Feiner. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In *ICMI '03: Proceedings of the 5th International Conference on Multimodal Interfaces*, pages 12–19, New York, NY, USA, 2003. ACM Press.

[9] Matthias Kölsch, Ryan Bane, Tobias Höllerer, and Matthew Turk. Touching the visualized invisible: Wearable ar with a multimodal interface. *IEEE Computer Graphics and Applications*, 26(3):62–71, 2006.

[10] Ji-Young Oh and Hong Hua. User evaluations on form factors of tangible magic lenses. In *ISMAR '06: Proceedings of the IEEE/ACM Internation Symposium on Mixed and Augmented Reality*, pages 23–32, 2006.

[11] Alex Olwal. Lightsense: Enabling spatially aware handheld interaction devices. In *IMSAR '06: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 119–122, 2006.

[12] Randy Pausch, Dennis Proffi tt, and George 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.

[13] Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. Image plane interaction techniques in 3d immersive environments. In *SI3D '97: Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 39–47., New York, NY, USA, 1997. ACM Press.

[14] Gerhard Reitmayr and Tom Drummond. Going out: Robust model-based tracking for outdoor augmented reality. In *IMSAR '06: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 109–118, 2006.

[15] Jun Rekimoto. The magnifying glass approach to augmented reality systems. In *ICAT/VRST '95: Proceedings of the International Conference on Artificial Reality and Tele-Existence and Conference on Virtual Reality Software and Technology*, pages 123–132, 1995.

[16] George Robertson, Mary Czerwinski, and Maarten 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 Press.

[17] Jürgen Schulze, Andrew Forsberg, Kleppe, Robert Zeleznik, and David Laidlaw. Characterizing the effect of level of immersion on a 3D marking task. In *Proceedings of HCI International*, Las Vegas, NV, 2005.

[18] Mel Slater, Vasilis Linakis, Martin Usoh, and Rob Kooper. Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. In Mark Green, editor, *ACM Virtual Reality Software and Technology*, pages 163–172, 1996.

[19] Ross Smith, Wayne Piekarski, and Grant Wigley. Hand tracking for low powered mobile ar user interfaces. In *Proceedings of the 6th Australasian User Interface Conference*, Newcastle, NSW, 2005.

[20] Arthur Tang, Frank Biocca, and Lynette Lim. Comparing differences in presence during social interaction in augmented reality versus virtual reality environments: An exploratory study. In *Proceedings of the Seventh Annual International Workshop on Presence*, 2004.

[21] Bruce Thomas. Evaluation of three input techniques for selection and annotation of physical objects through an augmented reality view. In *ISMAR '06: Proceedings of the IEEE/ACM Internation Symposium on Mixed and Augmented Reality*, pages 33–36, 2006.

[22] Anne Treisman. Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8:194–214, 1982.

[23] Daniel Wagner, Thomas Pintaric, Florian Ledermann, and Dieter Schmalstieg. Towards massively multi-user augmented reality on handheld devices. In *Proceedings of Third International Conference on Pervasive Computing*, 2005.

[24] Jason Wither, Stephen DiVerdi, and Tobias Höllerer. Using aerial photographs for improved mobile ar annotation. In *ISMAR '06: Proceedings of the IEEE/ACM Internation Symposium on Mixed and Augmented Reality*, pages 159–162, 2006.

[25] Jason Wither and Tobias Höllerer. Evaluating techniques for interaction at a distance. In *ISWC '04: Proceedings of the Eighth International Symposium on Wearable Computers (ISWC'04)*, pages 124–127, Washington, DC, USA, 2004. IEEE Computer Society.

[26] Bob Witmer and Michael Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Envrionments*, 7(3):225 – 240, 1998.

[27] Joanne Zucco, Bruce Thomas, and Karen Grimmer. Evaluation of three wearable computer pointing devices for selection tasks. In *ISWC '05: Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, pages 178–185, Washington, DC, USA, 2005. IEEE Computer Society

ACKNOWLEDGMENTS

This research was in part funded by a grant from NSF IGERT in Interactive Digital Multimedia #DGE-0221713 and a research contract with the Korea Institute of Science and Technology (KIST) through the Tangible Space Initiative Project..